



# REPORT DOCUMENTATION PAGE

Form Approved  
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

<b>1. REPORT DATE</b> Revised		<b>2. REPORT TYPE</b> Final		<b>3. DATES COVERED</b> 01 Oct 2001 - 1 Feb 2012	
<b>4. TITLE AND SUBTITLE</b>  Comprehensive Reproductive System Care Program – Clinical Breast Care Project (CRSCP-CBCP)				<b>5a. CONTRACT NUMBER</b>	
				<b>5b. GRANT NUMBER</b> W81XWH-05-2-0053	
				<b>5c. PROGRAM ELEMENT NUMBER</b>	
<b>6. AUTHOR(S)</b>  Craig D. Shriver, M.D., FACS, COL MC USA Ms. Lee Bronfman  E-Mail: cshriver@hmfj.org				<b>5d. PROJECT NUMBER</b>	
				<b>5e. TASK NUMBER</b>	
				<b>5f. WORK UNIT NUMBER</b>	
<b>7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)</b>  The Henry M. Jackson Foundation for the Advancement of Military, Inc. Bethesda, Maryland 20817				<b>8. PERFORMING ORGANIZATION REPORT NUMBER</b>	
<b>9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)</b> U.S. Army Medical Research and Materiel Command Fort Detrick, Maryland 21702-5012				<b>10. SPONSOR/MONITOR'S ACRONYM(S)</b>	
				<b>11. SPONSOR/MONITOR'S REPORT NUMBER(S)</b>	
<b>12. DISTRIBUTION / AVAILABILITY STATEMENT</b> Approved for Public Release; Distribution Unlimited					
<b>13. SUPPLEMENTARY NOTES</b>					
<b>14. ABSTRACT</b>  The Windber/Walter Reed Clinical Breast Care Project will help lead the way in the 21st century in the crusade against breast disorders. The project will utilize a multidisciplinary approach as the standard of care for treating breast diseases and breast cancer. This multidisciplinary model integrates prevention, screening, diagnosis, treatment and continuing care, but the project is further unique in the incorporation of advances in risk reduction, informatics, tissue banking and research. These efforts focus on decreasing the morbidity and mortality of breast cancer among American women.					
<b>15. SUBJECT TERMS</b> Tissue Banking, Biomedical Informatics, Focused Research, Translational Research, Genomics, Proteomics, Risk Reduction, Comprehensive Breast Care					
<b>16. SECURITY CLASSIFICATION OF:</b>			<b>17. LIMITATION OF ABSTRACT</b>	<b>18. NUMBER OF PAGES</b>	<b>19a. NAME OF RESPONSIBLE PERSON</b> USAMRMC
<b>a. REPORT</b> U	<b>b. ABSTRACT</b> U	<b>c. THIS PAGE</b> U			<b>19b. TELEPHONE NUMBER</b> (include area code)
			UU	FI JI	

## Table of Contents

<b>Cover.....</b>	<b>1</b>
<b>SF 298.....</b>	<b>2</b>
<b>Table of Contents.....</b>	<b>3</b>
<b>Introduction.....</b>	<b>4</b>
<b>Body.....</b>	<b>6</b>
<b>Pillar specific scientific plans and methods.....</b>	<b>9</b>
<b>Key Research Accomplishments.....</b>	<b>41</b>
<b>Reportable Outcomes.....</b>	<b>44</b>
<b>Conclusions.....</b>	<b>46</b>
<b>References.....</b>	<b>50</b>
<b>Appendices.....</b>	<b>50</b>
<b>Attachments.....</b>	<b>50</b>

**Comprehensive Reproductive System Care Program—Clinical  
Breast Care Project (CRSCP-CBCP)  
Final Report**

**I. INTRODUCTION:**

The Clinical Breast Care Project (CBCP) is the outcome of the initial FY00 and subsequent Congressional appropriations, and consists of an extensive collaborative effort between Windber Medical Center (Windber, PA – 12<sup>th</sup> Congressional District of the late Honorable John P. Murtha) and Walter Reed Army Medical Center, (now the Walter Reed National Military Medical Center) with funding management by the Henry M. Jackson Foundation for the Advancement of Military Medicine.

We believe that there are three broad areas where the BC-TRCOE (Breast Cancer – Translational Research Center of Excellence) stands poised to make major contributions to breast cancer research and its translation into clinical practice. These areas include the identification of molecular profiles of disease with high clinical relevance, deepening our understanding of the genetic risk of breast disease and the enhancement of our understanding of breast tumor biology. These three themes are supported by the five pillars of the BC-TRCOE. There is no doubt that our understanding of the biology of Breast Cancer in all of its various forms and manifestations remains incomplete. We believe that our high-value repository of biospecimens, our strong biomedical informatics infrastructure and our research base with strong internal and external collaborations puts us in an excellent position to make contributions to the understanding of breast disease that will have impact on the quality of life for breast cancer patients and their families.

**Clinically Relevant Molecular Profiling:** This is a cross-cutting theme with clinical, risk assessment and basic research components. The primary focus of this theme is to evaluate the utility of existing molecular profiles that have relevance to risk assessment, diagnosis, prognosis and therapy in a clinical setting and to discover new profiles that can be evaluated in the clinic. Projects within this theme have well defined translational goals. The development of comprehensive and highly informative molecular profiles will be a foundation for the development and delivery of personalized/individualized medicine. A variety of research modalities will be used to identify these profiles including immunohistochemistry, gene and protein expression analysis and genetic profiling including Next Generation DNA Sequencing. Two major new initiatives are outlined below one involving the development and testing of clinically relevant immunohistochemical profiles for disease stratification and therapeutic guidance and the other using complete genomics sequencing of tumor and matched normal DNA to develop clinically relevant profiles that could aid in disease diagnosis, prognosis and therapy selection.

**Genetic Risk:** The rapid developments of high throughput genotyping and genomic sequencing of individuals has reminded the research community of the power of family

studies in the assessment of genetic risk. Evaluating family risk and translating that into individual risk is the primary goal of this theme. There is both clear clinical relevance and a strong basic research component to this theme. Understanding the underlying biology of observed racial disparities in disease prevalence, presentation and outcome will also be a major part of this effort. The interaction of the theme with the Risk Reduction pillar of the BC-TRCOE and the number of projects outlined below that deal with research into the basis of the observed racial disparities in breast cancer morbidity and mortality point out the relevance of this theme to the overall goals of the BC-TRCOE.

**Tumor Biology:** A unique combination of resources and expertise put the BC-TRCOE in a strong position to further our understanding of the basic biology of breast disease including breast cancer. Many of the projects outlined in the Focused Research pillar address basic problems associated with tumor heterogeneity. The tumor microenvironment and stromal interactions, metastasis and recurrence, as well as the role of cancer stem cells and tumor evolution affecting the efficacy of treatment are emphasized. We firmly believe that a robust understanding of breast tumor biology is a key to the successful translation of the research performed at the BC-TRCOE to the clinic.

**Ultimate Goal of this project:** Decrease morbidity and mortality of breast cancer among American Women. Through the interlacing of the five pillars, the Clinical Breast Care Project (CBCP) will help lead the crusade against breast disorders.

- Develop a comprehensive breast care center/system that enables health care providers with a multidisciplinary team approach to work toward a common goal.
- Empower women with breast cancer and other breast disorders with the decision-making tools and environment to enhance quality of life and to meet psychosocial needs of the patients and their families.
- To continue support and grow a world-class biorepository of biospecimens that enable research into diseases of the breast.
- Develop research facilities that drive world-class high-throughput translational research.
- Develop an integrated computational and biomedical informatics infrastructure with an integrated data warehouse that forms the foundation for analysis of research findings leading to new and actionable knowledge related to diseases of the breast.
- Empower the clinical staff with a physician decision support system incorporating our evolving understanding of breast cancer and other breast diseases from research both within and outside the BC-TRCOE.

The five pillars of the CBCP are: (1) Risk Reduction (2) Focused Research (Genomics and Proteomics); (3) Tissue banking; (4) Biomedical informatics; and (5) Clinical Care.

## II. BODY

### Pillar Specific Goals and Objectives:

#### *1. Breast Cancer Risk Reduction:*

- Identify the population of patients at above average risk for the development of breast cancer.
- Decrease this identified population's rate of breast cancer development.
- Analyze potential cost differential in the prevention of breast cancer development.
- Incorporation of newly identified markers of breast cancer risk into the assessment of breast cancer risk.
- Identify patients from families that might harbor mutations in the BRCA1 or BRCA2 genes and offer testing to identify these mutations
- Identify families with unexplained high frequencies of breast cancer as potential research subjects

#### *2. Biorepository:*

- Collect and store a broad spectrum of biospecimens from every patient undergoing a breast biopsy and/or breast surgery at the former WRAMC, (Walter Reed Army Medical Center) now WRNMMC, (Walter Reed National Military Medical Center) WMC (Windber Medical Center), AAMC (Anne Arundel Medical Center), and our affiliated hospitals, who consent to participate in BC-TRCOE IRB-approved protocols.
- Collect and store biospecimens (blood) from women who are free of breast disease who consent to participate in BC-TRCOE IRB-approved protocols to act as controls.
- Utilize the power of this extensive biorepository as a major resource for breast disease research.
- Leverage the BC-TRCOE biorepository to maximize the utilization of the repository, with BC-TRCOE leadership approval, for the overall benefit of breast cancer patients and research, as able and appropriate.
- Participate in national/international projects that can benefit from resources of the BC-TRCOE biorepository.

#### *3. Focused Research (including: Genomics and Proteomics Research):*

**Genomics:** Utilize high-throughput and translational research in a unique Discovery Science environment to include but not be limited to:

- DNA analysis with genotyping studies, Copy Number Variation (CNV), gene sequencing and whole genome sequencing
- RNA/cDNA micro arrays, to identify expression level differentials across the entire spectrum of breast disease and from cancer specimens of all stages and

types, as well as the accompanying lymph nodes and metastatic deposits, and blood.

- Measure epigenetic changes associated with disease
- Study molecular differences between breast tumors from African American and Caucasian women as the identification of such differences will allow for the development of more effective therapies that will improve outcomes in African American women with breast cancer.
- Perform whole genome DNA sequencing on DNA from 40 or more cases of breast cancer.
- Using state-of-the-art 3D cell culture techniques and modern approaches to the study of cancer cell biology, study the mechanisms of cell invasion, migration and ultimately metastasis in breast cancer cell lines.
- Use our unique collection of breast cancer biospecimens to characterize microRNA (miRNA) expression in breast cancer progression and metastasis.
- Identify genetic changes in low- and high-grade breast tumors to improve our understanding of the evolutionary process of breast cancer and to identify a protein signature that can discriminate low from high-grade breast tumors, allowing for more accurate diagnosis and risk assessment.
- Use our unique collection of breast cancer biospecimens to characterize molecular signatures that can differentiate primary breast tumors with and without metastatic potential, as well as between primary tumors and subsequent metastases.
- Improve our understanding of the molecular changes associated with HER2 amplification and over-expression to allow for more precise diagnosis of HER2+ patients and development of customized treatment options in patients with HER2+ breast cancer.
- Study the role of matrix metalloproteinases in breast cancer with the goal of developing diagnostic and prognostic marker of breast cancer based on expression of MMPs and polymorphisms in MMPs.
- Identify molecular alterations in the breast tumor microenvironment that contribute to tumorigenesis and which may lead to improved methods of breast cancer prevention and treatment.
- Use our unique collection of breast cancer biospecimens to study angiogenesis and lymphogenesis in different grades of DCIS and IDC.

**Proteomics:** Through collaboration with world class partners, utilize high-throughput and translational research in a unique Discovery Science environment to include but not be limited to:

- Identify protein signatures associated with the development and progression of pre-malignant breast disease to improve our understanding of the biologic processes involved in early breast disease development and progression and to drive the development of personalized therapeutics for breast disease.
- Mass spectrometer pattern analysis and protein identification.
- Identify protein expression level differentials across the entire spectrum of breast disease and cancer specimens of all stages and types, as well as the

- accompanying lymph nodes and metastatic deposits, serum and blood.
- Use Accurate Mass Tag (AMT) technology to assess protein expression changes in tumor tissues.
  - Search for novel protein biomarkers, individual or pattern.
  - Store all this expression data in a data warehouse where it can then be utilized for biologic pathway development and in-silico biology research for hypothesis-driven research.
  - Store all this expression data in a data warehouse where it can then be utilized for biologic pathway development and in-silico biology research for hypothesis-driven research.

#### **Collaborative Research:**

- Conduct quantitative analysis of therapy relevant proteins by immunohistochemistry within various subclasses of breast cancer to provide better patient selection into clinical trials for targeted and combination therapies.
- Perform affiliated translational laboratory research in support of the main expression profiling and biomarker discovery goals of the BC-TRCOE research laboratories.
- Develop alliances with other research organizations and entities and carry out project-supported research in support of same. Currently we have established collaborations with the Pacific Northwest National Laboratory, Vanderbilt University, The Institute for Systems Biology, the Thomas Jefferson University, the MGR Global, etc.
- Collaborate with the NCI/NHGRI TCGA (The Cancer Genome Atlas) project to study the genomics of breast cancer.

#### ***4. Biomedical Informatics:***

- Develop a comprehensive QA program and aid in SOP development for data collection and generation to ensure acquisition of high quality of data.
- Develop and support a robust laboratory information management system to ensure proper tracking of data acquisition.
- Develop and implement a clinically relevant and laboratory research-linked prospective, longitudinal computerized data warehouse to support translational research and ultimately support physician decision making
- Develop an analytical system including developing specific algorithms for integrative data analysis and mining, and deploying existing applications and algorithms to ensure execution of data analysis, mining, and modeling.
- Develop a breast knowledgebase to support clinical and research activities in BC-TRCOE.
- Develop other needed infrastructure to support the activities in all other BC-TRCOE pillars.
- Incorporate the rapidly growing public genomic and proteomic datasets related to breast cancer into our data warehouse to be able to mine the

combined data sets for the generation of new hypotheses regarding breast cancer development, progression and treatment.

### ***5. Clinical Care:***

- Provide state of the art clinical care and treatment of patients seen in the BC-TRCOE at Walter Reed National Military Medical Center Bethesda.
- Decrease the negative psychological impact on the patient of having an evaluation or treatment intervention for breast disease by utilizing objective measurement instruments to longitudinally assess the patient's psychological response to evaluation and intervention, and base modifications of these procedures on those results.
- Create and maintain an environment (medical, physical, psychological) conducive to the multiple needs of the patient undergoing breast disease evaluation / treatment.
- Recruit patients into the various BC-TRCOE protocols to obtain the clinical data and biospecimens needed to meet the BC-TRCOE translational research goals.

## **III. PILLAR SPECIFIC SCIENTIFIC PLANS AND METHODS**

### **Comprehensive Reproductive System Care Program—Clinica Breast Care Project (CRSCP-CBCP) Statement of Work**

Task 1: Identify and counsel no less than 100 patients at high risk for of breast cancer, and employ risk reduction strategies.

Task 2: Accrue over 500 patients annually to the “core” BC-TRCOE protocols through consenting patients in the main BC-TRCOE clinical sites.

Task 3: Acquire through consented protocol acquisitions, over 5,000 specimens annually (neoplastic and non-neoplastic breast tissues and tumors, lymph nodes, metastatic deposits, blood and its components, bone marrow) on patients with all types of breast diseases and cancer.

Task 4: Bank these biospecimens in the BC-TRCOE Biorepository as the substrate for all molecular analyses carried out in BC-TRCOE labs, as outlined in the BC-TRCOE Core Protocols. Utilize this repository as the basis for intramural and extramural collaborations for secondary usage research.

Task 5: Perform focused research as outlined below on the biospecimens and clinical data collected under the BC-TRCOE Core protocols including global expression analysis of the DNA, RNA, and Protein features and including targeted research into genomic analysis of Stages I, II, and III breast cancer,

DCIS, LCIS, and pre-malignant neoplasia. Present findings in peer-reviewed national meetings and publications.

Task 6: Perform whole genome DNA sequencing on DNA from 40 cases of breast cancer.

Task 7: Develop and support a robust laboratory information management system to ensure proper tracking of data acquisition and a clinically relevant and laboratory research-linked prospective, longitudinal computerized data warehouse to support translational research and ultimately support physician decision making.

Task 8: Develop an analytical system for integrative data analysis and mining, and develop a breast knowledgebase to support clinical and research activities in BC-TRCOE.

### ***BC-TRCOE PILLAR #1 BREAST CANCER RISK REDUCTION:***

#### **OBJECTIVES:**

- a. To collect data on all female patients 18 and older who formerly presented to the General Surgery Clinic at Walter Reed Army Medical Center and now present to the Breast Care and Research Center at Walter Reed National Military Medical Center and to the Joyce Murtha Breast Care Center in Windber, PA (as of 16 April 2009) who are found to be at an increased or elevated risk for developing breast cancer.
- b. To utilize this database to analyze the diagnosis, and treatment outcomes for patients found to be at an increased risk for developing breast cancer. Analysis will include but not be limited to: risk factors for developing breast cancer, effectiveness of various modalities of treatment, and actual risk of developing cancer.

#### **BACKGROUND:**

Each year, approximately 200,000 women in the United States are diagnosed with breast cancer, and one in nine American women will develop breast cancer in her lifetime. But hereditary breast cancer - caused by a mutant gene passed from parents to their children - is rare. Estimates of the incidence of hereditary breast cancer range from between 5 to 10 percent to as many as 27 percent of all breast cancers.

In 1994, the first gene associated with breast cancer — BRCA1 (for BREast CANcer1) was identified on chromosome 17. A year later, a second gene associated with breast cancer — BRCA2 — was discovered on chromosome 13. When individuals carry a mutated form of either BRCA1 or BRCA2, they have an increased risk of developing breast or ovarian cancer at some point in their lives. Children of parents with a BRCA1 or BRCA2 mutation have a 50 percent chance of inheriting the gene mutation.

In 1995 and 1996, studies of DNA samples revealed that Ashkenazi (Eastern European) Jews are 10 times more likely to have mutations in BRCA1 and BRCA 2 genes than the general population. Approximately 2.65 percent of the Ashkenazi Jewish population has a mutation in these genes, while only 0.2 percent of the general population carry these mutations.

**MEDICAL APPLICATION:**

Not all hereditary breast cancers are caused by BRCA1 and BRCA2. In fact, researchers now believe that at least half of hereditary breast cancers are not linked to these genes. Scientists also now think that these remaining cases of hereditary breast cancer are not caused by another single, unidentified gene, but rather by many genes, each accounting for a small fraction of breast cancers.

Hereditary breast cancer is suspected when there is a strong family history of breast cancer: occurrences of the disease in at least three first or second-degree relatives (sisters, mothers, aunts). Currently the only tests available are DNA tests to determine whether an individual in such a high-risk family has a genetic mutation in the BRCA1 or BRCA2. Genetic counselors can help individuals and families make decisions regarding testing. For those who do test positive for the BRCA1 or BRCA2 gene, surveillance (mammography and clinical breast exams) can help detect the disease at an early stage. A woman who tests positive can also consider taking the drug tamoxifen, which has been found to reduce the risk of developing breast cancer by almost 50 percent in women at high risk. Clinical trials are now under way to determine whether another drug, raloxifene, is also effective in preventing breast cancer.

The field of oncology/surgical oncology is an ever-changing one with new developments in both diagnosis and treatment. We propose to collect data from all patients in the WRNMMC Breast Care and Research Center determined to be at an elevated risk for developing breast cancer in order to assess risk factors in this population for developing the disease and track outcomes of preventive and therapeutic interventions. Analysis of outcome will include comparison of various treatment modalities/regimens with regard to efficacy, risks for failure, complications, and overall morbidity/mortality/survival. The patient population of WRNMMC can provide a significant number of patients to compare/contrast our findings with those of our civilian counterparts, specifically the Joyce Murtha Breast Care Center. The database will also allow us to analyze breast cancer risk data to provide scientific-based evidence that will guide the general surgeon and medical oncologist in optimal care of the patient at an elevated risk for developing breast cancer.

**PLAN:**

The risk Reduction Clinic at WRNMMC and at Joyce Murtha Breast Care Center (JMBCC) is a multi-disciplinary program designed to identify, counsel and manage women at high risk for breast cancer. Patients receive an in-depth personal and family health history by a world renowned medical oncologist.

Current research shows there are risk factors that may influence the development of breast cancer. Identifying people with these risk factors and implementing closer surveillance and risk reduction techniques may detect cancer earlier. Earlier detection of breast cancer leads to better prognosis and outcomes. Calculations of risk are based on computer models extensively validated as accurate in identifying women at high risk.

Subjects: All female patients age 18 and older, seen in the WRAMC Comprehensive Breast Center at a high risk for developing breast cancer. Patients will have consented to WU # 01-20006, Tissue and Blood Library Establishment for Molecular, Biochemical, and Histologic Study of Breast Disease or WU # 01-20007, Creation of Blood Library for the Analysis of Blood for Molecular Changes Associated with Breast Disease and Breast Cancer Development. The core questionnaire completed by the patients contains the NCI Breast Cancer Risk Assessment Tool used to determine a patient's risk factor. Please see ATTACHMENTS 4,6 and 8.

Study Design and Methodology: Patients being seen in the WRAMC Comprehensive Breast Center at WRAMC or at the JMBCC in Windber, PA will be assessed for their risk of developing breast cancer by their history of LCIS or ADH or by applying the NCI Breast Cancer Risk Assessment Tool. Identified high-risk patients will be referred to the CBCP Risk Reduction Clinic. Patients are confirmed to meet the inclusion criteria and consented to one of two core protocols. Information collected will include the data contained on the enclosed database forms and may include data from previous clinic visits. All applicable patients will be followed indefinitely according to the applicable protocol.

If patients are referred for genetic testing, as per the American Society of Clinical Oncology, counseling involves the following eleven points:

1. Information on the specific test being performed
2. Implications of positive and negative results
3. Options for estimation without genetic testing
4. Risk of passing a mutation to a child
5. Technical accuracy of the test
6. Possibility that the test will not be informative
7. Fees involved in testing
8. Risk of psychological distress
9. Risk of insurance or employer discrimination
10. Confidentiality issues
11. Options for medical surveillance and screening following testing

It has been observed that healthy female relatives of individuals with ovarian or breast cancer tend to exaggerate their risk of incurring either form of cancer and, thus, accurate risk assessment is essential to quality genetic counseling for breast cancer. Breast cancer genetic counseling serves the goal of helping women to analyze their own and their relatives' risk of developing breast cancer.

In BRCA screening, genetic counselors offer services in compiling family histories, personalizing epidemiology, and, more recently, conducting genetic testing to empower healthy women of families stricken by breast cancer to alter their lifestyles and healthcare to ensure avoidance or early detection of breast cancer. In addition, breast cancer victims and healthy members of a single family can enable accurate screening of female family members by obtaining a sequence of their BRCA genes. Detection of a familial BRCA mutation in individuals outside of the Ashkenazi Jewish population requires time-consuming genetic analysis of a large number of affected and unaffected family members in order to identify the specific BRCA mutation for a particular family.

Testing positive or negative for a BRCA mutation is simply a risk assessment, not a certainty of experiencing or avoiding, respectively, breast cancer. Individuals with a BRCA mutation have an 80% risk of developing breast cancer by age 80. Therefore, 20% of BRCA mutation carriers never develop breast cancer. A first-degree relative of a carrier who tests negative for the mutation has the same breast cancer risk as women of the general population, namely 11%.

Dr. Raymond Weiss, Medical Oncologist has seen 229 patients at the Joyce Murtha Breast Care Center (Risk Reduction program began 16 April 2009) and a total of 2359 patients at WRAMC and WRNMMC.

## ***BC-TRCOE PILLAR #2 BIOREPOSITORY***

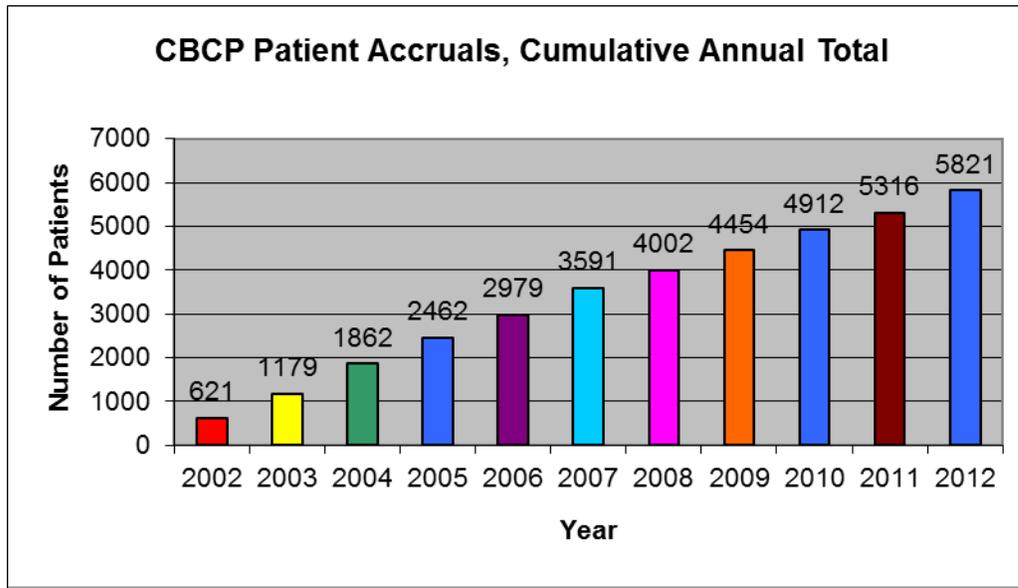
### **BACKGROUND**

Although there have been remarkable improvements in breast cancer diagnosis and management, most of the complex molecular mechanisms associated with the onset, progression and/or severity of breast cancer are still not well understood. As part of the BC-TRCOE we carry out molecular, biochemical and histological analysis of breast tissue and/or blood and blood components from breast cancer patients to provide insights into the molecular mechanisms that may be relevant in the development of breast cancer and breast diseases. To achieve this aim, a large supply and a wide variety of good quality tissue samples are needed. Unfortunately, good quality donor breast tissue is extremely scarce and when available is often not backed by a comprehensive medical history and/or is not a good representation of the target population or study area. The non-availability of a steady and consistent supply of good quality tissue limits the systematic analysis of tissues and negatively impacts the generation of biologically useful information in research laboratories and by extension negatively impacts new findings that benefit clinical practice. The objective of this project is therefore the acquisition and banking of breast tissue, lymph nodes, serum/plasma and other blood derivatives from informed and consenting donors.

Since the inception of the Clinical Breast Care Project the Biorepository Pillar has been critical to the success of the project. As we move forward into the establishment of the

BC-TRCOE it is important to look at the success of the biorepository and to understand the firm foundation that it has laid for building the Center of Excellence.

Figure BB-1 shows the cumulative patient accrual into the CBCP protocols since 2002. These patients, who have been recruited and consented into the CBCP protocols at WRAMC, WRNMMC, AAMC, JMBCC and other participating CBCP clinical intake sites are the foundations of the translational research that has



**Figure BB -1. Cumulative patient accruals into CBCP protocols since 2002.**

occurred within the CBCP and which will continue in the BC-TRCOE. From these patients we have collected and stored in our biorepository over 50,821 biospecimens (Figure BB-2). These specimens represent a broad spectrum of tissues, blood and blood products (Figure BB-3) that are not only a unique and valuable resource for the BC-TRCOE but are also the substrates for our translational research program. Along with the biospecimens that have been collected from CBCP participants, each consented patient also provides nearly 800 field of demographic, medical, life and family history data as well as complete pathology data on donated tissues. The collection of tissues that the BC-TRCOE inherits from the CBCP is even more valuable because of this rich annotation. Patients have been recruited from a number of partnering clinical intake sites over the history of the CBCP (Figure BB-4). At the start of the BC-TRCOE the active partners are WRAMC, the Joyce Murtha Breast Care Center in Windber, PA, and the Anne Arundel Medical Center in Annapolis, MD.

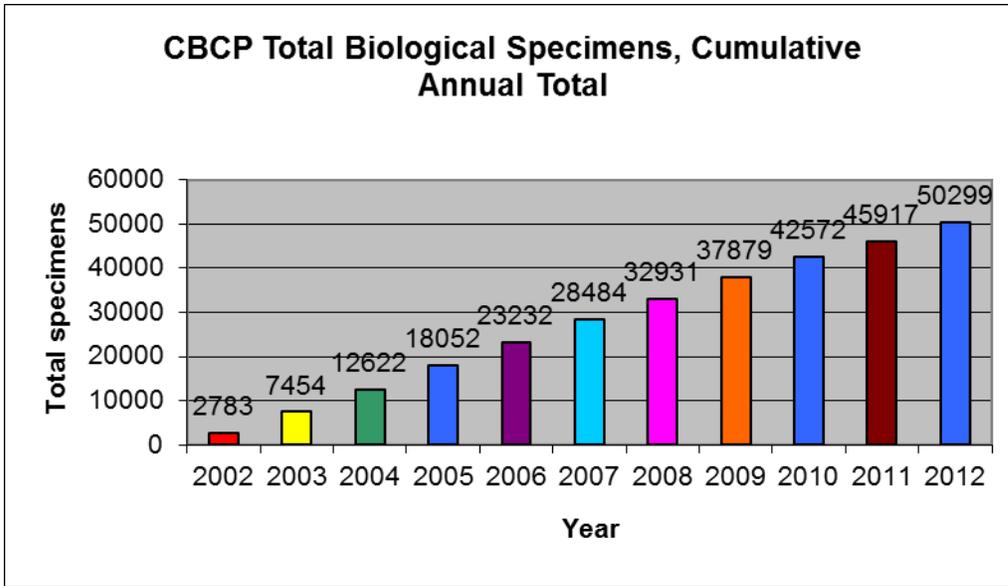


Figure BB-2. Total biospecimens collected and banked by the biorepository.

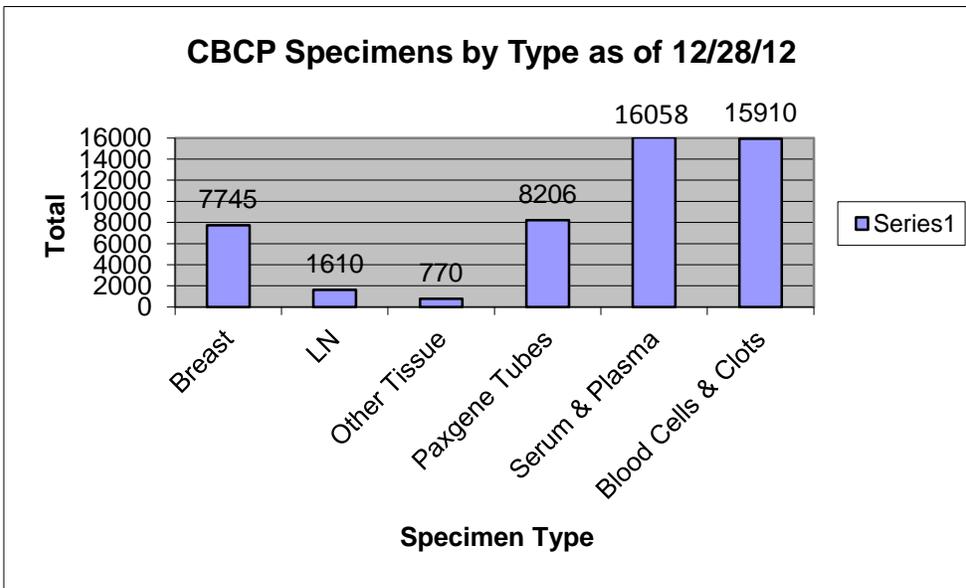
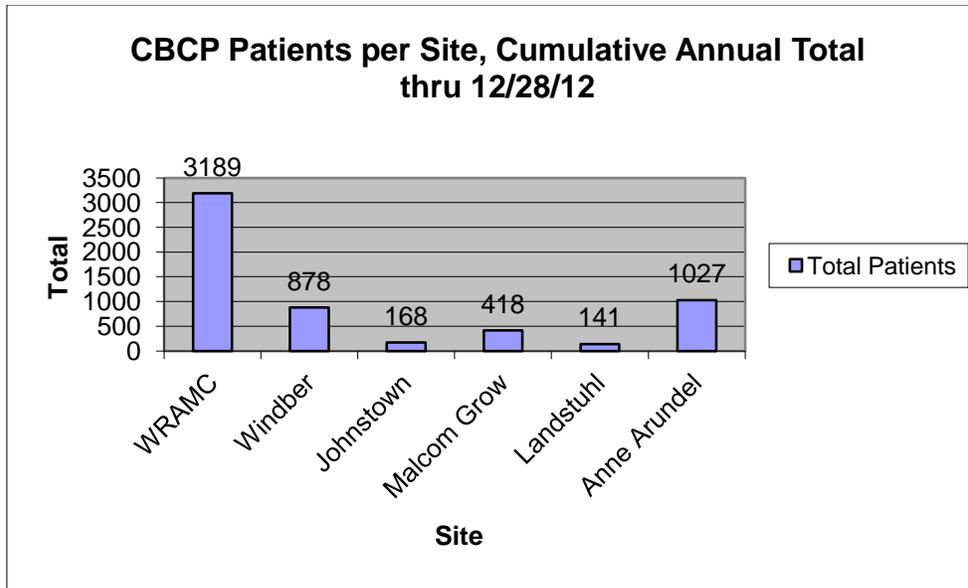


Figure BB-3. The numbers and types of biospecimens collected by the CBCP



**Figure BB-4. Numbers of patient recruited to CBCP protocols at various partner sites.**

The tissue banking pillar was established at both sites in collaboration and entails acquisitions, storage, and movement amongst the sites for research purposes of tissue garnered from all breast surgeries being performed at both locations. The robust IRB-approved protocol that enables this pillar is unique in four critical aspects: It is a tissue usage protocol, not a repository one; It is hypothesis-generating, not hypothesis-driven research; It allows for patients to pre-consent for secondary, future uses of the tissues in presently-unknown research; It contains a unique fail-safe mechanism to protect the complete diagnostic integrity of all samples.

A Core activities of the CBCP is the collection and banking of human biological specimens for research. Tissue specimens are collected from a variety of donors under different conditions. The goal is to collect good quality specimens that will provide products such as DNA, RNA and proteins for research. To achieve this goal, the Tissue Banking arm of the CBCP has incorporated the science of Biospecimen Research into its activities. We have performed research studies to determine how temperature changes and time of collection to processing affect the quality of breast specimens collected for research. These findings have been presented at national conferences and have been published in peer reviewed journals. Our results contributed to the shaping of the NCI Office of Biorepositories & Biospecimen Research (OBBR)'s Best Practices and their proposed guide lines for specimen collection and processing for tissue banks.

### ***BC-TRCOE PILLAR # 3 Focused Research including Genomics and Proteomics***

The research pillar of the BC-TRCOE focuses on the translational research program involving the clinical programs at Walter Reed Army Medical Center's Breast Center and

the Joyce Murtha Breast Care Center and the genomic and proteomic analysis carried out at the Windber Research Institute.

The following is a list of the projects that make up the Focused Research Pillar of the BC-TRCOE. Major new initiatives include a major project that will generate complete genomic DNA sequence from up to 40 breast cancer cases. Another new initiative will utilize immunohistochemical to generate clinically relevant profiles of breast tumors to better stratify the disease in terms of prognosis and treatment options.

### **Estimating the Prognostic Value of Gene Expression Data from Primary Tumors in Predicting Lymph Node Status**

Despite tremendous advances in early detection and treatment, breast cancer remains the second leading cause of cancer-related death in women. While a number of clinical parameters have been identified, axillary lymph node status remains the most important prognostic factor in predicting disease outcome. Previous studies in our laboratory using allelic imbalance (AI) to detect chromosomal changes have shown that there are high levels of discordance of AI between primary breast tumors and corresponding axillary lymph node tumors (Ellsworth et al. 2005). In 2009, we published a RNA-based signature of metastasis of 51 genes that differ significantly between primary breast and matched metastatic lymph node tumors.

We have recently completed a study evaluating gene expression patterns in 74 primary breast tumors with and without metastasis and have submitted for publication in a special issue of International Journal of Breast Cancer. While an effective signature was not identified, the reasons why a global signature was not detected are intriguing, including genetic predisposition to metastasis. We have generated 100K SNP data from white women with invasive breast cancer and without breast cancer to look for genetic differences that may suggest host susceptibility, not to development of invasive tumors but for propensity to successfully metastasize. In addition, in collaboration with the Mass Spectrometry Research Center at Vanderbilt University, we are identifying protein expression patterns in primary tumors from patients with and without metastasis, protein differences in primary vs. metastatic lymph node tumors and between disease-free lymph nodes from patients with and without lymph node metastases.

### **Allelic Imbalance Studies in Breast Cancer**

The allelic imbalance panel we developed in 2002 has been used often to evaluate genetic changes in all stages of breast tissue (disease-free through metastatic). We have completed all studies using this panel. In 2010, we published the results of our AI evaluation in non-neoplastic diseases, concluding that ADH and CCH are in fact, genomically naive and that the multitude of studies that have evaluated pre-neoplastic lesions from breasts with invasive breast tumor cannot represent the status of pure early lesions.

Our last study will be to evaluate patterns of allelic imbalance in the context of outcome. We have identified a number of chromosomal regions either present at significantly higher (suggestive of metastasis) or lower (protective effect) in tumors from patients who have died of disease.

### **Genomic differences in African and Caucasian American Women with Breast Cancer**

Breast cancer incidence in women in the United States varies among different ethnic groups. The highest incidence occurs in Caucasian women; however, mortality is highest among African American women. Furthermore, African American women are more likely to be diagnosed at a younger age and with more advanced stages of the disease. These differences in outcome between Caucasian and African American women are frequently attributed to disparities in health care access and socioeconomic status. However, molecular factors may also be involved as indicated by the differences in tumor biology between African American and Caucasian women. Tumors in African American women tend to be larger and more aggressive with high-grade nuclear atypia and lymphocytic infiltrate. Tumors from African American women are also more likely to be triple negative (ER-, PR-, and HER2-) making them refractory to treatment with current targeted therapies such as tamoxifen and herceptin. Initial profiling of the African American women enrolled in the CBCP was published in 2007, identifying both demographic and molecular differences between CCW and AW. Recently we generated gene expression data from both tumors and non-malignant tissues from AAW and CW. 18 genes were differentially expressed in tumors and 13 in non-malignant breast specimens, including PSPHL, SOS1 and CRYBB2, which are differentially expressed in both tumors and non-malignant tissues.

### **Breast Tumor Microenvironment**

Alterations in cell signaling, angiogenesis, cell migration, motility, and proliferation establish the tumor microenvironment as a dynamic entity contributing to breast tumorigenesis. These alterations can be contributed to epigenetic, morphological, and genetic changes. Research in the past decade has increased our understanding of how the changing tumor microenvironment influences tumor development, and this is proving to be of significant importance in the progression of breast cancer research. An improved understanding of the molecular changes occurring in the breast tumor microenvironment may have significant clinical ramifications as well as provide an increased understanding of breast tumor etiology. Continued investigation into this research could provide insights into tumor development and progression, leading to meaningful advances in breast cancer. In 2009, we published an extensive review paper in Expert Review of Molecular Diagnostics weighing in on the debate as to whether morphologically normal breast stroma adjacent and distant from the tumor has chromosomal changes.

Copy number evaluation is being performed on fresh frozen genomic DNA samples from women with breast cancer, and a variety of pathological classifications using Affymetrix 500K SNP chips. Copy number and LOH analysis is being performed using Affymetrix

genotyping console. Preliminary data were presented at the AACR meeting in April. Recently, we received a grant from CDMRP to evaluate gene expression differences in adipose close to and distant from breast tumors as well as adipose from non-malignant breasts to improve our understanding of how this long ignored component of the breast stroma contributes to tumorigenesis.

### **Combined Genomics and Proteomic Analysis of Metastatic Potential in Breast Cancers**

We are studying the metastatic potential of breast cancers with collaborators at the Pacific Northwest National Laboratory (PNNL). A total of 76 estrogen receptor positive (ER+) tumor samples (20 LN-PM; 20 LN+PM; 19LN-PRM; 13 LN+PRM) have been collected by laser capture microdissection from histological tissue preparations for comparing the proteomic and genomic profiles from these four categories or groups. These samples underwent genomic analysis at WRI and are undergoing proteomic analysis at PNNL.

WRI and PNNL have begun the exchange of data on the differentially expressed genes discovered by each group in their independent analysis (genomics- WRI, proteomics- PNNL). With this data exchange we are developing our strategies to integrate these data sets.

A proteomics methods paper has been drafted from the initial LC/MS proteomics methods optimization research on OCT embedded tissues and the manuscript has been submitted to the Journal of Proteome Research.

### **Proteomics (CBCP, WRI) report Y2010 Kumar Kolli**

#### **Body**

Exploration/Identification of biomarkers through differential proteomics on retrospectively collected clinical breast cancer samples has been the major focus of proteomics research core facility at WRI and we have continued our efforts in Y 2010 to adapt the emerging protein microarray technology in the proteomics lab for high throughput profiling of serum samples. In this report, we summarize preliminary results on profiling of a panel of cytokines from breast cancer serum samples using antibody microarrays along with the serum profiling using GeLC/MS approach to evaluate the degradation protein products in the serum as markers to determine the invasiveness of breast cancer.

## **Profiling of breast cancer sera by cytokine antibody array to identify early disease markers**

### **Introduction**

Breast cancer is the most common cancer and the second leading cause of cancer death in women in the U.S with more than 40,000 mortalities each year. Early detection remains the most promising approach to improve the long-term survival of cancer patients. One area of research interest to explore the early markers for cancer is profiling serum cytokines. These proteins play an important role in innate immunity, apoptosis, angiogenesis and differentiation. The expressions of individual cytokines are altered in cancers and the interplay between immune and tumor cells further alters their expressions and measuring cytokine levels in serum might be useful in determining the specific stage and grade of tumors. With the emerging protein microarray based technologies, the discovery-driven translational research on cancer related clinical samples has drawn a considerable attention to rapidly profile the protein expressions using chip based protein microarrays for biomarker discovery in the area of disease diagnostics and prognostics. In the present study, we are exploring breast cancer serum samples to identify the early markers for cancer detection using cytokine antibody arrays.

### **Methods**

Serum samples from 24 postmenopausal subjects which were evenly divided between benign and invasive cases were probed with glass slide based antibody array (AAH-CYT-G2000-8; RayBiotech, Inc) to profile 174 cytokines simultaneously from each serum sample (30uL). The serum samples were prepared according to the protocol provided by vendor and profiled based on the principles of sandwich type immunoassay and the cytokine signals were read on a GenePix 4000B fluorescent scanner using Cy3 channel. Spot alignment was performed with GenePix Pro 4.0 where signal and background detection levels for each spot were defined as median intensity. Background subtraction and positive control normalization were done with RayBio Analysis Tool while average ratio and independent two-tailed t-test analyses were performed in Microsoft Excel to find the abundance difference of each cytokine between invasive and benign groups.

### **Results and Conclusion**

The differential analysis of cytokine profiles from benign and invasive cases has shown only moderate expression changes for several cytokines. The two cytokines found to be elevated in the invasive group were MMP-9 and lowered expressions for MIP-1-delta, MIF, leptin R, SDF-1beta, TRAIL R4, prolactin, and LIF were observed in invasive group. Profiling of additional samples is under way, while extending the study to include other stages and node positive serum samples has also been planned. As the node positive and node negative BC tissue proteomics/genomics data were available, the cytokine profiling for the matched serum samples has been planned and this analysis will expand the scope of understanding the molecular mechanisms that drive breast cancer progression and invasion. Prediction models will be built from these microarray datasets to correctly identify the breast cancer cases.

## **Evaluating protein degradation products as serum biomarkers to determine the invasiveness of breast carcinoma**

### Introduction

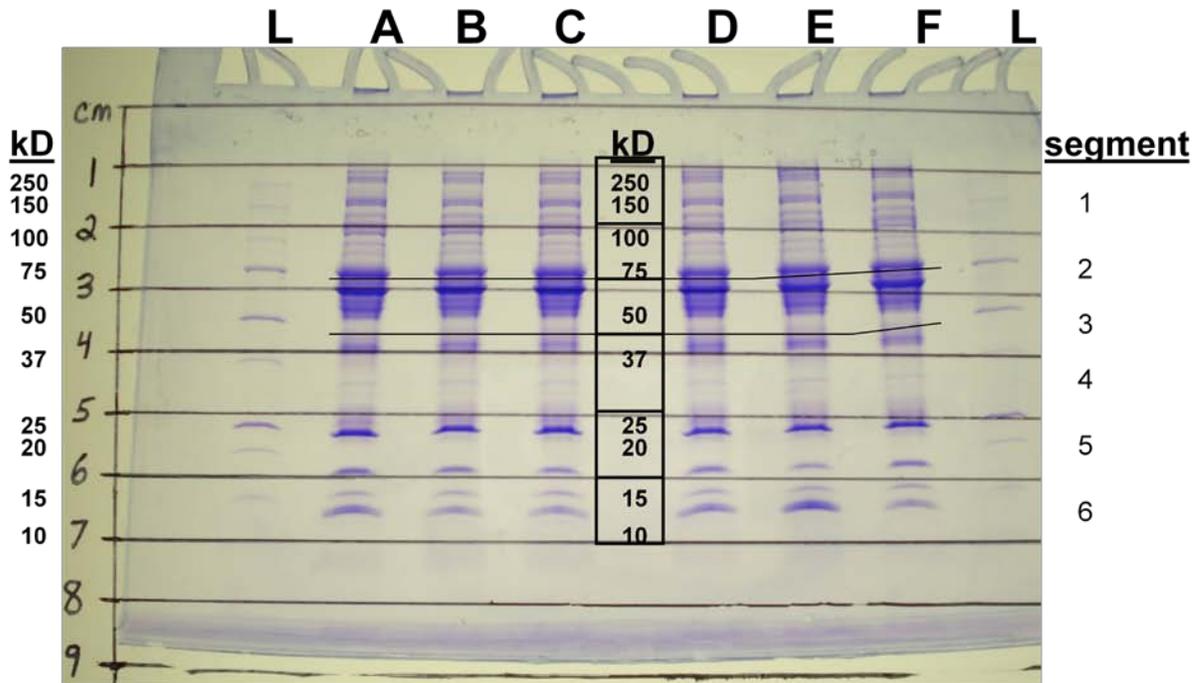
The process by which cancer spreads from its site of origin to other tissues is a complex series of steps. The first of these involves the cancer cells freeing themselves from the primary tumor by degrading the proteins that make up the surrounding extracellular matrix (ECM), which separates the tumor from adjoining tissues. Degradation of the ECM is accomplished by the tumor cells secreting a variety of proteases, including the matrix metalloproteases (MMPs) and Plasminogen activators (PAs). Our hypothesis is that these secreted enzymes leak into the blood stream where they degrade serum proteins.

### Methods

The breast cancer serum (20  $\mu$ L) samples (20 benign and 20 invasive, stage I) were processed with Enchant multi affinity protein separation kit to remove the high abundant proteins. The depleted flow through fractions of serum samples from the respective groups were pooled to prepare three pooled samples for each group and the samples were separated on 1D gel (fig1). The gel lanes were cut into segments (fig1) and proteolytically digested, and the extracted peptides were run in duplicates on a LTQ-FT MS. MS/MS were searched against a database of human proteins with Sequest, and identifications were statistically validated at a 1% protein false discovery rate using the Provalt algorithm in ProteoIQ (BioInquire, Athens, GA). The proteoIQ data mining tools were used to identify the protein degradation in the samples by assembling the data from each segment.

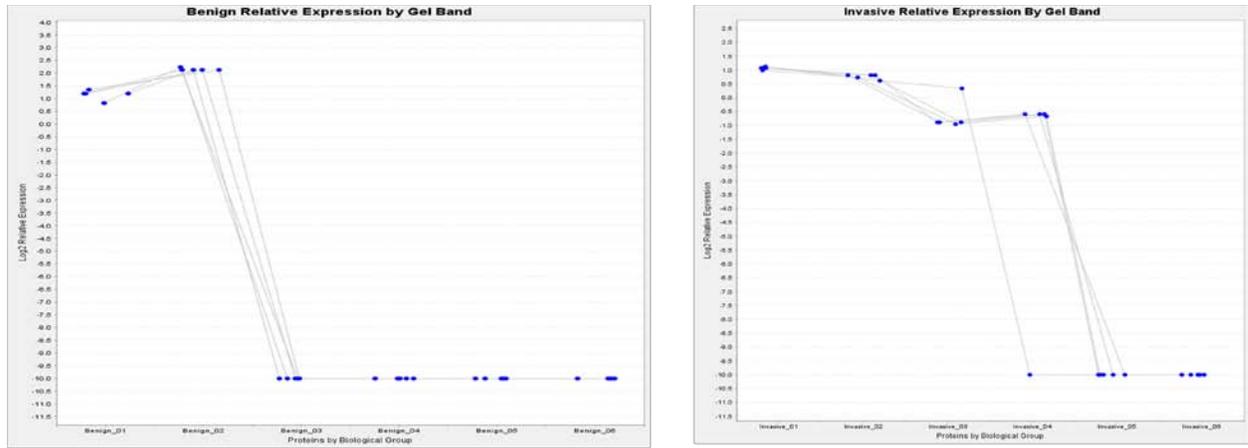
### Results and conclusion

Based on our enzyme degradation hypothesis, we predict that the level of these fragments and their extent of degradation will be indicative of tumor invasiveness and thus capable of serving as serum biomarkers for this condition. To this end, we have applied a GeLC-MS approach for this study where intact proteins are separated via SDS gel electrophoresis prior to trypsin digestion and LC-MS/MS analysis. This sample-flow allows for the identification of proteins whose experimental MW (determined by SDS gel) differs significantly from their calculated MWs. This process has allowed us to identify several protein degradation products and other wise-abnormally modified proteins which may be related indirectly to tumor invasiveness. The presence of degraded inter-a-trypsin inhibitors can be clearly seen (fig2) as the presence of peptides from these proteins being detected across 4 gel slices in the invasive samples. In the benign samples, peptides are only detected in 2 of the gel slices (fig2). The preliminary results from this pooled sample study are encouraging and we are planning to apply this workflow on breast cancer samples individually to correlate the invasive characteristics of primary tumors to the extent of protein degradation products in the cancer serum samples.



ABC = Benign pooled samples  
 DEF = Invasive pooled samples  
 Each lane contains 30ug of protein.

Figure1. Separation of pooled serum samples on 1D gel and image shows the gel segments used for LC/MS/MS analysis.



Benign

Invasive

Figure2. peptides of inter-a-trypsin inhibitor identified from different gel segments in invasive sample.

### Collaborative Research

The proteomics group at WRI established collaborative research relationships with Scientists at University of Pittsburgh and Wistar Institute to develop the predictive models on the proteomics datasets generated from serum samples and to develop novel concepts of sub-categorizing the breast cancers (triple negative) through isoform level gene expression and predicting each sub-category by the status of isoform specific gene regulatory networks and this will be addressed by combination of innovative computational modeling approaches and use of NextGen sequencing technologies. These collaborative research partnerships resulted in developing proposals to apply for NIH and DOD grants in Y 2010 and the abstracts for the proposals were given below.

### Bayesian Rule Learning for Disease Prediction and Biomarker Discovery (Biomedical Informatics, University of Pittsburgh; Dr. Vanathi Gopalakrishnan)

**The problem:** High-throughput biomedical data from biomarker profiling studies aimed at early detection of diseases like breast cancer and lung cancer are accumulating rapidly. Although many popular machine learning methods have been utilized for analysis of such high-dimensional datasets, no single method has consistently outperformed others. Moreover, scientists have the need to simultaneously address two related tasks: disease prediction and biomarker discovery, using the same sets of data and tools. One way, as undertaken in this project, to address this need is to find the most accurate classifier for the disease from a given set of profiles and present the discriminative markers used in that model to the scientist for further verification. The large space of possible models coupled with the small sample size of the data make it hard to accurately estimate predictive accuracy.

**The solution:** This project will develop, evaluate and refine novel Bayesian Rule Learning (BRL) methods that are algorithmically efficient, result in parsimonious models and accurately estimate predictive uncertainty from sparse biomedical datasets. BRL methods utilize a Bayesian score to evaluate rule models, thereby quantifying the uncertainty in the validity of the rule itself. This novel technique that combines the mathematical rigor of Bayesian network learning with rule-based modeling opens up a hitherto underexplored area of fundamental research in informatics involving such hybrid methodologies. Rules enable modular representation of knowledge and collaboration with scientists, as it is easier to present the model and extract markers both visually and computationally. Rule-based inference is also simpler and more tractable. The Bayesian approach enables prior knowledge to be incorporated and evaluated in a continual fashion with a human in the loop. The latter is very important for refinement of both tools and models.

**The specific aims:** This project will test the hypothesis that the BRL methods developed and extended herein produce more accurate and parsimonious models for disease state prediction than other state-of-the-art machine learning methods. This project evaluates BRL methods and models using existing proteomic datasets for three diverse diseases – rare, neurodegenerative Amyotrophic Lateral Sclerosis (ALS), and the two most common cancers in the world, lung and breast cancers. Experimental verification will be performed using a new set of retrospectively collected breast cancer sera samples to evaluate model generalizability.

**The significance:** This project will produce: (1) a novel biomedical data mining tool for analyzing data from biomarker profiling studies of any disease, (2) methodological insights into the applicability of this tool and current machine learning methods for such tasks, and (3) new data for research on the early detection of breast cancer. It has potential to help develop new diagnostic tests for early detection of ALS, lung and breast cancers and lays a firm foundation for building modeling frameworks that can incorporate both prior knowledge and data to provide the technological capability for combining evidence from multiple, heterogeneous sources.

***Isoform specific miRNA:mRNA regulatory networks in breast cancers*** (Molecular and Cellular Oncogenesis, Wistar Institute; Dr. Ramana Davuluri)

Triple-negative breast cancer (TNBC), lacking estrogen and progesterone receptor expression and *HER2* amplification, accounts for approximately 15-20% of breast cancer diagnoses. TNBCs comprise a heterogeneous group of very aggressive tumors with poor prognosis. Some TNBC tumors are similar to basal-like breast cancers (BLBC), which are characterized by elevated expression of basal cytokeratins (CK5/6, CK14), myoepithelial marker (p63) and HER1. However, the similarity between BLBCs and TNBCs is only partial as defined by high-throughput gene expression analyses. The lack in complete similarity may in part be explained by both BLBC and TNBC comprising entities that in themselves are heterogeneous in their molecular profile.

Recent mRNA-seq data suggest that more than 90% of the human genes, including many tissue specific and developmentally regulated genes, produce multiple transcript

isoforms, through alternative splicing and alternative usage of transcription initiation and/or termination. Notably, there is growing evidence linking aberrant use of alternative mRNA isoforms and cancer formation: several oncogenes and tumor-suppressor genes (e.g. LEF1, TP53, TP63, TP73, HNF4A, RASSF1, and BCL2L1) are already known to have multiple promoters and alternative splice-forms, and moreover, it is known that the aberrant use of one isoform over another in some of these genes is directly linked to cancerous cell growth. However, nearly after a decade from the completion of the human genome draft sequence, we still assume “gene” as the basic physical and functional unit in the genome (Hayden Nature 2010). We argue that “gene/transcript isoform” should be the basic functional unit and biologists should look at the differential regulation of gene isoforms rather than that of genes between normal and disease conditions.

Similarly, non-coding RNAs (ncRNAs) have emerged as key regulators at both transcriptional and post-transcriptional levels. The discovery of small ncRNAs of size 20-31 nucleotides has revolutionized thinking about the control of gene expression. Although in many cases the target of small RNAs and their mechanism is still not clear, it is certain that many of these small RNAs play important role in several biological processes including development timing, cell differentiation, cell proliferation, cell death, metabolic control, transposon silencing and antiviral defense. Several classes of small ncRNAs have been discovered such as small interfering RNAs (siRNA), micro RNAs (miRNAs) and Piwi-interacting RNAs (piRNAs). The best understood among the small ncRNA classes is miRNAs. MiRNA modulates post-transcriptional gene regulation by base pairing to either full or partial complementary sequences primarily in the 3' untranslated region (UTR), but also inside the coding sequence (CDS) and the 5' UTR, of the corresponding messenger RNAs (mRNAs). The miRNA:mRNA interactions affect mRNA stability and translation. In normal cells the miRNA:mRNA interactions need to be very precise in order to regulate specific gene isoforms post-transcriptionally. A miRNA can target a specific mRNA isoform but not the other isoforms of the same gene (due to inclusion/skipping of exons/exonic regions), leading to balanced expression of mRNA isoforms of same gene in a given cell type. Therefore, it is critical to address the problem of miRNA:mRNA interactions at isoform level rather than at gene level. We propose a novel idea that miRNA:mRNA interactions are disrupted more at isoform level than at gene level leading to imbalanced expression of gene isoforms in cancer cells. We, therefore, hypothesize that the heterogeneity in TNBCs can be characterized by genome-wide combined signatures of mRNA expression at isoform level and expression of miRNAs. And, consequently different sub-categories of TNBCs can be predicted by the status of isoform-specific miRNA-mRNA interactions involving tumor suppressor or oncogenic proteins, such as p53 family members.

We will use integrative biology approach by performing mRNA-Seq and small RNA-Seq experiments, using Illumina GA massive parallel sequencing, and efficiently predict the miRNA:mRNA interactions by data-mining of the mRNA-seq and small-RNA seq data to define homogeneous sub-categories of TNBCs.

This challenging problem will be addressed by combination of innovative computational modeling approaches and use of NextGen sequencing technologies. We will:

- (i) develop a knowledgebase of mammalian gene isoforms, expression patterns, and miRNA:mRNA interactions at isoform level by integrating cross-species genomic

- data and high-throughput transcriptome data, including recent massive parallel sequencing data from the cancer genome atlas (TCGA) data portal;
- (ii) conduct mRNA-seq to identify and quantify full-length poly-A transcript isoforms and small RNA-seq to quantify miRNAs and other small RNAs (e.g. miRNAs, piRNAs and siRNAs) in 40 TNBC primary tissue samples and 20 ER positive breast cancer tissue samples;
  - (iii) perform bioinformatics and statistical analysis of mRNA-seq and small RNA-seq data to identify major homogenous sub-categories of TNBCs;
  - (iv) perform integrative analysis of mRNA-seq and small RNA-seq data, and develop novel computational methods to model the combinatorial interaction of miRNAs in targeting the protein coding genes at isoform level, and predict different sub-categories of breast cancer based on the status of the miRNA:mRNA networks that are specific to each sub-category.

The proposed work will not only pave a new way to think about identification of novel cancer subtypes, but will also lead the research effort to understand the molecular mechanisms involved in the pathogenesis of different subtypes of TNBCs. Understanding the isoform-specific miRNA:mRNA networks is clearly central if we are to develop rational and mechanistic therapies for complex genetic diseases, such as breast cancer, that arise from the accumulated contributions of many gene–gene interactions and environmental factors. It is anticipated that this research will lead to most effective treatment, matched to their cancer profile, eventually improving the breast cancer survival rates.

### **Studying Cancer Genomes**

The focus of the BC-TRCOE at WRI is deepening our understanding of the genetic risk of breast disease through DNA sequencing. The rapid developments of high throughput genotyping and genomic sequencing of individuals has reminded the research community of the power of family studies in the assessment of genetic risk. Evaluating family risk and translating that into individual risk is the primary goal of this theme. There is both clear clinical relevance and a strong basic research component to this theme. Understanding the underlying biology of observed racial disparities in disease prevalence, presentation and outcome will also be a major part of this effort. The interaction of the theme with the Risk Reduction pillar of the BC-TRCOE and the number of projects outlined below that deal with research into the basis of the observed racial disparities in breast cancer morbidity and mortality point out the relevance of this theme to the overall goals of the BC-TRCOE.

Two developments have occurred in the last decade that are about to fundamentally alter the understanding and treatment of cancer, including breast cancer. First, the elucidation of the sequence of the human genome has created a framework to identify variation in the genomes of many individuals and to associate that variation with susceptibility to various diseases. For diseases like cancer, a collection of diseases that are the result of somatic mutations, it gives the opportunity to identify the constellation of mutations that give rise to individual tumors and to elucidate the etiology of the disease, its prognosis and to determine the vulnerabilities the tumors and therefore suggest potential treatments. The

second development has been the emergence of “next generation” DNA sequencing technologies that have reduced the cost of sequencing an individual human genome from hundreds of millions of dollars to a few thousand dollars. We are rapidly approaching the time when an individual’s complete DNA sequence will be part of their medical record which will create a context in which their life style choices can be evaluated to give a clearer assessment of their risk of developing various cancers. When an individual is diagnosed with cancer, determining the DNA sequence of the tumor will be part of the normal clinical workup and it will be critical to designing an appropriate course of treatment.

There are two National/International efforts underway to sequence cancer genomes. The International Cancer Genome Consortium (<http://www.icgc.org/home/>) is made up of teams from 10 countries with the goals of sequencing the genomes of 50 different cancer types and/or subtypes. In the United States, The Cancer Genome Atlas (TCGA) project (<http://cancergenome.nih.gov/>) is a joint project between the National Cancer Institute (NCI) and the National Human Genome Research Institute (NHGRI) which is investing \$275 million to chart the genetic changes and perturbation to complex pathways in more than 20 types of cancer. The TCGA has completed its pilot phase which focused on glioblastoma, lung and ovarian cancers and which will be discussed below. TCGA is moving into a phase that will sequence genomes from a number of different tumor types and which will perform other high-throughput analyses on samples from these tumors in order to help understand and elucidate the molecular, pathway and physiological changes associated with cancers. As will be discussed below the BC-TRCOE is already part of the TCGA working group for breast cancer and will be supplying cases for this portion of the project.

While the BC-TRCOE is participating in the NCI/NHGRI/TCGA project and will be submitting between 100 to 150 breast cancer cases to this effort, we also believe that we are in a position to make significant contributions to our understanding of breast cancer by subjecting a number of cases to complete genome sequencing independently of the TCGA project. A problem that we see with the TCGA project as it is currently envisioned is that not all of the cases will undergo complete genome sequencing. This is because when the project was initially conceived the cost of sequencing a complete human genome was prohibitive. Instead the plan was to sequence the “exome”, the part of the genome covered by the exons of protein coding genes. Though this solution was reasonable in 2005 at the inception of this project, the rapid changes in technology over the past 7 years requires a new evaluation of the cost-benefit ratio of studying tumor genomes by complete genome sequencing. There are a number of solutions to sequencing individual genomes that has brought the cost per genome to ~\$10,000 or less. Since interpreting a cancer genome requires comparing it to the patient’s non-cancer genome, each case requires the sequencing of two complete human genomes making the cost per case somewhere in the neighborhood of \$20,000.

A number of types of analysis are enabled by complete genomic sequencing. Mutational profiles allow examination of the types and extent of deletion, insertions and amplification associated with various cancers. Rearrangements, especially those

associated with gene fusions can be determined. Point mutations can be cataloged and, by looking at the sequence of multiple tumors of the same type, mutations that are driving the oncogenic process can be identified as can mutations in genes that are potential therapeutic targets. Access to the mutational spectrum of tumors can help to elucidate the mechanisms of carcinogenesis and to identify various DNA repair pathways that are involved in tumor development and progression. Finally because for each case the normal (eg. from blood) DNA will also be sequenced it will be possible by examining the right set of cases to identify new genes involved in cancer susceptibility.

The main research focus for the WRI portion of the BC-TRCOE has been the characterization of breast tumors by the analysis of the somatic mutations in these tumors as revealed by their DNA sequence. We have two major projects which are generating and interpreting DNA sequence data. These are 1.) various internally initiated projects using BC-TRCOE cases to address basic breast cancer problems and for which the DNA sequencing is being performed by Complete Genomics and 2.) our participation in the NCI/NHGRI project, The Cancer Genome Atlas (TCGA). The BC-TRCOE currently has nearly 100 cases, out of a current total of about 800, that are going through the TCGA pipeline and for which DNA sequence (exome), gene expression, copy number variation, DNA methylation status, micro RNA status and Reverse Phase Protein Array (RPPA) data are being generated. In addition to being a tissue source site for the TCGA project, several WRI investigators are working to analyze the TCGA data as part of the Data Analysis Working Group. These projects will be discussed below.

The first project that we have undertaken with BC-TRCOE cases that will be sequenced by Complete Genomics, is to examine cases where a single tumor has distinct regions that, by immunohistochemistry (IHC), are of different molecular subtypes. The question is whether these regions represent independent occurrences of breast cancer or whether the regions represent different developmental pathways that were followed by different parts of the same tumor. By isolating these different regions of these tumors and determining their DNA sequence we should be able to distinguish between these two possibilities.

To do these experiments it is necessary to separate the different subtypes by laser capture microdissection. This is somewhat laborious given that in order to obtain a whole genome sequence of the tumor material requires ~7 micrograms of DNA. We have now completed the preparation of DNA from 2 tumors that have two different subtypes by IHC and from one homogenous tumor. These samples along with matched DNA from blood from the same patients have been sent to Complete Genomics for whole genome sequencing. These samples have all passed the Complete Genomics QA/QC process and are now in the production sequencing pipeline. Data for three cases (8 samples) have been generated and are undergoing analysis.

The second project involves our participation in the NCI/NHGRI TCGA project. We have been shipping biospecimens to the TCGA (The Cancer Genome Atlas) since late 2010. We have sent 97 cases (tumor and normal samples) of which 86 have passed the QA/QC processes of the project. We have also collected clinical data (enrollment,

pharmaceutical therapy and radiation therapy data) on these cases and have entered the information in OpenClinica. These cases along with approximately 700 other samples are undergoing analysis. WRI scientists have been part of the team analyzing the exome sequences, copy number variations, gene expression, micro RNA content, DNA methylation analysis and the results of reverse phase protein arrays from these samples. We have also played a very active role in providing QA/QC for the very diverse clinical information that is being collected on the cases undergoing analysis by TCGA. This data are complex and different tissue source sites, if which there are about 20, have some variability in the way that they record and report these data. Because these data will become part of a public data repository it is important to develop methods to assure uniformity in the data. An example of the type of analysis that has been done for this data “cleaning” project is attached in Appendix 1. We have also participated in the preparation of a manuscript describing the initial results of the breast cancer TCGA project which has been published in *Nature*.

### **Historical perspective - FY2005**

**Performing targeted research into genomic analysis of Stages I, II, and III breast cancer, DCIS, LCIS and pre-malignant neoplasia, and presenting findings at national meetings and in peer-reviewed publications during FY05.** CBCP has been tremendously successful in this regard this year as well. We have verified that important genetic changes occur across multiple chromosomal regions as breast epithelium transitions from non-neoplastic into neoplastic types. We have identified these changes and published the specific markers for chromosomal loci that are modified as a requirement for, or as a result of, the malignant transformation process. We are deciding now as an organization whether to seek patent protection for this panel of markers, which may be consider a diagnostic and prognostic marker panel for breast cancer and breast cancer development.

### **Perform mass spectrometry fingerprinting of 200 sera samples from patients with diagnosis of breast diseases and analyze for distinct patterns based on disease state.**

Using over 600 CBCP sera samples, we have undertaken a set of proof of principle experiments to show there is sufficient information in serum to diagnose the presence or absence of breast cancer. We hypothesize that there are molecular patterns in serum that are unique to breast cancer. These patterns are composed of multiple (3 or more) molecules, which by themselves are not informative, but when taken together, and measured relative to each other, are informative.

Our approach is to profile sera using high-resolution mass spectrometry, and then apply a novel, patented pattern recognition algorithm to these data, to identify patterns that can distinguish the sera from women with breast cancer, from women without breast cancer.

In 2005 we presented a poster at ASCO reporting our first results. Using sera from women representing 67 cases of *in situ* carcinoma, 195 invasive carcinoma, and 350 normal/benign breast conditions, we identified several informative serum patterns with

sensitivities and specificities in the range of 80 – 90% on a blinded validation set. One profile, consisting of 10 serum features, yielded 98.5% specificity (95% CI 95.2 – 99.6%) and 90.3% sensitivity (95% CI 82.4 – 95%) on a Testing set of 196 non-malignant sera and 103 invasive sera. When this pattern was challenged with a Blinded Validation set of 54 non-malignant and 41 invasive sera we observed 92% specificity (95% CI 74.5 – 93.6%) and 86% sensitivity (95% CI 76.4 – 97.8%).

With such studies there are two significant concerns – sample set bias and overfitting of data. Recently, using an expanded set of sera now representing 454 women with invasive cancers, we performed an extensive analysis of the data to minimize sample set bias, and implemented an algorithm to monitor for overfitting of data. We identified several serum patterns that distinguished between invasive cancer and normal/benign, non-neoplastic conditions. Performance was 85% sensitivity (standard deviation  $\pm$  7%) and 82% sensitivity (standard deviation  $\pm$ 10%). While a little lower than the first study, these results are considerably more robust.

These unpublished proof of principle results show that serum does contain sufficient information to distinguish between the presence and absence of breast cancer. As our serum set increases in size, and our algorithm experiences a greater extent of the variability of the disease within the population, we anticipate the performance should improve significantly.

**NB. After 2010 we discontinued in house proteomic research but continued proteomics research with outside collaborators.**

### **Historical perspective - FY 2009**

**Performing targeted research into genomic analysis of Stages I, II, and III breast cancer, DCIS, LCIS and pre-malignant neoplasia, and presenting findings at national meetings and in peer-reviewed publications the period of July 1, 2007 through June 30, 2008.** Performing targeted research into genomic analysis of Stages I, II, and III breast cancer, DCIS, LCIS and pre-malignant neoplasia, and presenting findings at national meetings and in peer-reviewed publications the period of July 1, 2007 through June 30, 2008. CBCP continues to have tremendous success in this regard. We have verified that important genetic changes occur across multiple chromosomal regions as breast epithelium transitions from non-neoplastic into neoplastic states. We have identified these changes and published the specific markers for chromosomal loci that we identified as having changed during this transition. Whether or not any of these changes are causal in the malignant transformation process remains to be determined. We are considering, as an organization, whether to seek patent protection for this panel of markers, which may be considered a diagnostic and prognostic marker panel for breast cancer and breast cancer development.

**Perform mass spectrometry fingerprinting of 200 sera samples from patients with diagnosis of breast diseases and analyze for distinct patterns based on disease state.** Perform mass spectrometry fingerprinting of 200 sera samples from patients with

diagnosis of breast diseases and analyze for distinct patterns based on disease state. The activities of proteomics facility for the FY07 were mainly centered on the data analysis of Maldi-TOF profiles generated from the acetone processed BC serum samples, planning and implementation of new serum processing workflows to couple them to Maldi-TOF and LC/MS/MS analyses and optimization of workflow for tissue proteomics in collaboration with the PNNL group.

We have applied a simple differential precipitation procedure using acetone to enrich the low molecular wt peptides/proteins from breast cancer serum samples for MALDI-TOF profiling and utilized the MS profile patterns to develop a model that discriminates the two disease groups (benign and invasive) involved in the study. A total of 64 serum samples, evenly divided between benign and invasive cases, were processed by acetone in triplicates and each replicate sample was analyzed twice on MALDI-TOF instrument. The assessment of MS profiles and clinical covariate data yielded 31 benign and 30 invasive cancer cases for comparative analysis. To determine the discriminative features that distinguish benign and invasive cases, both univariate and multivariate analyses were performed on the features. The univariate analysis revealed several features were differentially expressed between benign and invasive BC cases. The unsupervised hierarchical clustering analyses indicate that there were some subgroups among the cancer and benign groups. The discriminative potential of peak clusters was analyzed using a Classification and Regression Tree (CART) algorithm implemented in Ciphergen ProteinChip™ Biomarker Pattern software. CART modeling selects a set of predictors and their interactions that optimally predict the outcome of the dependent variables. Samples were split into 70% training sets and 30% test sets. The menopausal status (categorical variable) was balanced between the benign and invasive groups in the training sets. The average performance of all the predictive models that were built using CART was 50%. The distributions of menopausal status, age, and HRT history for the subjects examined in this study could be discriminated, with statistical significance, between case and control groups. Only the menopausal status was able to be accommodated in the forming of training sets for multivariate analysis. The inability to balance for clinical variables may contribute to inconsistency and high ACE in classification model performance.

We have started a new workflow that supports the parallel processing (96 well plate format) of serum samples using Enchant multi-protein affinity separation kit (Pall Life Sciences) for MaldiTOF and LC/MS/MS profiling. The preliminary analysis using couple of test serum samples demonstrated good feature rich MALDI-TOF profiles from the processed samples using this workflow. We are planning to couple downstream protein fractionation workflow to the Enchant processed samples for good protein coverage by LC/MS/MS analysis. Alongside this workflow, we are also optimizing the single tube tissue sample preparation for shotgun proteomics of BC tissue samples.

## Reportable outcomes

(Conference abstracts)

Ron Orlando, Craig Shriver, Bernard Seth, David Kirchner, Richard Mural, James Atwood, Kyle Jones and **V.S. Kumar Kolli**. Evaluating protein degradation products as serum biomarkers to determine the invasiveness of breast carcinoma. American Society for Mass Spectrometry (ASMS), Salt Lake City, UT, May 23-27, 2010.

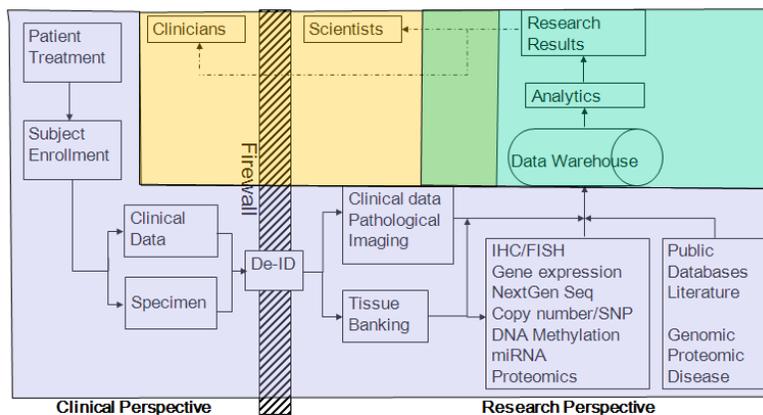
David Kirchner, Richard Katenhusen, Tapan Maiti, Craig Shriver, Richard Mural, **V.S. Kumar Kolli**. Shotgun proteomics of breast cancer sera to identify the markers for early cancer detection. American Society for Mass Spectrometry (ASMS), Salt Lake City, UT, May 23-27, 2010.

**NB. After 2010 we discontinued in house proteomic research but continued proteomics research with outside collaborators.**

### *BC-TRCOE PILLAR # 4 - Biomedical Informatics*

#### The Data Warehouse for Translational Research

Biomedical Informatics: Biomedical Informatics focuses on the management and



utilization of biomedical information, and our view of its role in translational research is shown on the left figure. From the data flow point of view, it involves data collection and generation, where data tracking is needed (light blue). Cleaned data are then centralized in the

data warehouse (light green), and subject to data analysis and mining for knowledge generation which is then presented to research scientists and clinicians to complete the two-way cycle of translational research (light yellow which partially overlaps with the data warehouse section).

We have developed additional workflow utilities for improving the quality and use of the data in the Data Warehouse for Translational Research (DW4TR).

- i) A new utility, the Data Correction Utility using the InforSense analytical workflow to correct data errors in the DW4TR, has been developed and deployed. This application was completely developed by the Biomedical

- Informatics staff. Currently it is in full use and continue to be improved based on user feedback.
- ii) Several IHC data fields have been revised to reflect the new needs including the final status of HER2 based on IHC and FISH, and loading of the newly generated data with improved granularity of Ki67 and p53 for invasive breast cancer samples is in progress.
  - iii) The sample selection utility using temporal information is in use and improvements have been made based on user feedbacks.
  - iv) We have established regular procedure and routine for data loading to the DW4TR: data are extracted from CLWS, applying the data processed by DCU, and then loaded into the DW.
  - v) After the acquisition of InforSense by IDBS there are substantial personnel changes. IDBS is committed to continue to focus on translational research. Several teleconferences have been held. In July, Dr. Hai Hu visited InforSense—now part of IDBS, in Burlington, MA, to review the development of a microarray data analysis workflow utility and to start the strategic planning of the future relationship between the two organizations. In November, several IDBS leaders visited WRI. We are planning on an additional strategic meeting late January.

#### **LIMS development of the clinical component**

- i) A cosmetic revision was done on the CLWS. The pre-printing rule was changed: additional questions in the questionnaire can now be pre-printed out in the CLWS for follow-up patients per request from the clinical staff.
- ii) The development of the new modules for clinical/tissue components in collaboration with GenoLogics was discontinued due to the directional problem of GenoLogics. The company apologized, and we successfully negotiated for a full refund of the pre-payment for the development. As a result, we plan to develop a system to replace the CLWS in-house.
- iii) Our plan is to develop the LIMS in modules, aiming to replace the CLWS that is currently in use but already obsolete.
- iv) With ProLogic serving as a development subcontractor, we have developed a new tool for completing and tracking pathology checklist and sample attribute sheet using TAB PC technology. This was accompanied with a drastic revision of the current Pathology Checklist led by the CBCP Head Pathologist with close teamwork involving the biomedical informatics group here, the ProLogic team, and the MDR Global leader. This Pathology Checklist tablet data capturing system development with is now done with the first phase development. Additional revision will be needed before this can be integrated with the new LIMS
- v) The Komen Promise Grant won by a 5-member consortium led by Dr. Hallgeir Rui of Thomas Jefferson, has been in execution since January 2010. It is a 5-year grant aiming to screen the expression of 250 drug targets across 5000 breast cancer tissues. Dr. Craig Shriver of WRAMC, Mr. Albert Kovatich of MDR Global, Dr. John Eberhardt III of DecisionQ, and Dr. Hai Hu of WRI are the other 4 leaders of the consortium. Execution of the Komen

Grant has a positive impact on CBCP. The WRI Biomedical Informatics team is responsible for the biomedical informatics infrastructure including the development of a LIMS and expansion of the DW4TR to support the project, and we will also support data analysis and research project design. The development of the LIMS is underway, which is coordinated with the development of the LIMS for CBCP CLWS replacement. One important component of the LIMS, the data entry form for the clinicopathologic data of the breast cancer cases, has been developed and is currently in use, which is useful for the development of the BC-TRCOE outcome data form.

- vi) One experienced developer was hired on December 20, to focus on LIMS development for both this project and the Komen Promise Grant.

### **IT Infrastructure and Support**

Historically, WRI IT support was outsourced, and from its inception the WRI IT infrastructure grew organically. We have reached a point where the whole system needs to be completely redesigned to lay the foundation for future research and support activities of the next phase of the research at WRI. With the help of professional consultants and a professional IT infrastructure company, we have completely redesigned our IT infrastructure. The new system applies virtual machine technology, with 4 servers hosting all the applications that were acquired through the 10 years of WRI's history. The servers have 32 GB memory each that are shareable, with a total of 36 TB shareable storage that can be further expanded as needed. The new network is supported by Level 3 with a speed of 1 Gbps

Effective January of 2011, Biomedical Informatics will provide WRI IT support as a newly added responsibility. For this purpose, a new IT Manager and an IT Support Specialist have been hired on December 20 and November 15, respectively. The original person in charge of the infrastructure re-structuring is no longer with WRI effective November 11. This change of personnel suspended the virtualization of the applications which has now been resumed. The status of the main tasks are shown below:

- i) Procurement of all hardware complete
- ii) Hardware installation complete, which includes network hardware, servers, and virtual storage
- iii) Network hardware configuration complete
- iv) Cut-off of WRI from the MANVT network complete and now Level 3 providing internet connectivity for WRI
- v) Virtualization of current physical servers – on going

### **SOP for clinical data entry (core questionnaire) have been reviewed and updated**

- i) In collaboration with WRAMC, WRI team combined two SOPs (one for nurses and one for data entry) into single document. During this process, a number of discrepancies between these documents has been corrected and all changes to SOP have been highlighted
- ii) WRI team visited WRAMC and presented new SOP document to the clinical data staff. During the presentation, additional clarifications and corrections were made.

- iii) The WRI team also reviewed, clarified, and updated the SOP with the JMBCC nurses.
- iv) The SOP is now in a final version—but we are maintaining it as a ‘live document’ so new issues coming out from the data collection/QA/data entry process will be incorporated.

### **Historical perspective - FY 2005**

**Continuing software development of the CLWS (Clinical Laboratory Workflow System) and its further deployment into the clinical/research arms of the CRSCP-CBCP.** During the last year, a standard operating procedure (SOP) for data entry and quality assurance (QA) of the questionnaire data has been established. The current QA measures for clinical data include, visual inspection by the data entry clerk, double blind-data entry, and a computer program checking applying all the QA rules that have been developed to examine the data integrity within and between data fields. A QA issue tracking (QAIT) system has also been developed and implemented, which dramatically improved the communications on QA issues between the WRI data entry group and the Walter Reed questionnaire QA team.

With the belief that electronic data capturing will enhance clinical data collection and precision, WRI has taken the initiative in developing a prototype tablet application using the Pathology Checklist as the first example following a decision made at the last CBCP Offsite meeting. This prototype is under further evaluation at both WRI and Walter Reed before a full-blown development starts.

The data warehouse development has further evolved to provide data and analytical support to CBCP research, which is done by collaborating with leading data management and analysis companies such as InforSense and Concentia Digital. An On-Line Analytical Processing (OLAP) tool has been developed to allow for easy access and stratification of hundreds of data elements across thousands of subjects enrolled in CBCP. A ‘patient view’ tool has been developed to allow for exploration of warehoused data from an individual patient. An application prototype has been developed to enable users to access clinical or experimental images at ease with a built-in robust data-element filtering capability. All these translational research-enabling applications will be migrated to and further developed on a newly designed patient-centric, object-oriented data model with a temporal dimension. We expect that this newly designed data model, which is in the process of being implemented, will dramatically enhance our translational research capability when compared to the old questionnaire-based data structure.

### **Historical perspective - FY 2009**

The patient-centric modularly structured clinical data model for the data warehouse continues to be developed and improved by incorporation of additional questionnaires and data from other clinical and translational projects such as the Gynecological Disease Program. A comparison between our data model with existing models and systems such as SNOMED, MeSH, HL7 has been done, and a comparison with caBIG VCDE has been

started. Such comparisons indicated that our data model is unique in supporting clinical and translational research. Issues like how to integrate the information captured from different questionnaires where similar questions may be asked in different ways are being worked on. The interface has been dramatically improved to make it more convenient to use and more powerful for users to interrogate the data from the data warehouse. A manuscript on this development is being prepared.

Currently this data warehouse is playing a pivotal role in supporting research in WRI and WRAMC. We use it for patient cohort formation, specimen selection, and breast cancer risk factor assessment. A number of collaborative projects are under development to make the best use of the data in this data warehouse. With a proof-of-principle pilot project completed, next we will focus on developing molecular study data models in the data warehouse to host the large amount of genomic and proteomic study data.

### **LIMS (Laboratory Information Management System)**

We have identified the LIMS product from Genologics as the system for WRI to adopt. There are two modules commercially available from the company, Geneus for genomic studies, and Proteus for proteomic studies. The company is also developing biomedical informatics modules for clinical data and specimen management. We have worked out with the company that WRI will have early access to the new modules in development.

For implementation of the Geneus and Proteus, we are holding meetings twice a week with Genologics for architecting and planning. As of now, the detailed implementation plan has been developed, a two-server hardware structure has been determined and ordered, a decision on the database is being finalized. In addition, a mock installation has been completed. The plan is to use these two modules to track the data generated in the lab, and at the same time to use the Clinical Laboratory Workflow System, CLWS to keep track of the clinical activities and specimens. Note that the latter is from the previous LIMS from Cimarron software, co-developed by WRI and Cimarron. When the biomedical informatics modules from Genologics are available, all the data tracking will be performed from the backbone structure of Genologics.

In the current structure, the clinical data is directly feeding into the data warehouse. When the Geneus and Proteus are functional, the laboratory data will be fed into the data warehouse as well. This way, the LIMS is for transactional activities, and the data warehouse is for reporting and research purposes. We believe that with this two-separate-system structure, the need for data collection and generation and storage and analysis we will be completely satisfied.

### **Other current active research and development projects**

- **QA measures of the CBCP clinical data:** a set of QA measures have been developed as a collaborative effort between WRI and WRAMC. This includes visual inspection of the completed questionnaires, double blinded data entry, and application of an automated computer program to enforce a metrics of QA rules. A tool called QA Issue Tracking system (QAIT, see next paragraph) is used to

track QA issues. A manuscript on these QA measures and results is being finalized.

- **QAIT:** A QA Issue Tracking system has been developed to facilitate communications between data entry at WRI and the clinical data collection/QA team at WRAMC. This web-based system provides centralized management of QA issues with role-based access to related information. It has greatly facilitated the QA process and enhanced overall quality of the CBCP clinical data. A manuscript on this project has been submitted to JAMIA and is currently under revision.
- **Use of the CBCP clinical data warehouse as a research environment for breast cancer risk factor assessment:** The data warehouse and the On-Line Analytical Processing tool as the interface have proven to be a powerful tool in supporting scientific research at WRI and WRAMC. We have tested the idea that the CBCP clinical data warehouse can serve as a research environment for breast cancer risk factor assessment, and have presented the preliminary data at SABCS 2007. Current the draft manuscript has been prepared and is being reviewed by the co-authors.
- **BIOBASE analysis of the differentially expressed genes in normal breast tissues from African American and European American women:** WRI lab scientist Dr. Lori Field has previously reported identification of the set of these differentially expressed genes at AACR. Using the same data set we have performed an in silico transcription factor analysis to explore the possible underlying mechanism why such genes were differentially expressed. The preliminary results have been presented at SABCS 2007 and currently a manuscript is being prepared.
- **Epidemiological Study of Traditional Breast Cancer Risk Factors and Histologically Confirmed Common Breast Pathologies:** A manuscript has been submitted to The Breast Journal and here is the abstract:  
We have previously reported that proliferative non-atypical breast disease co-occurs with breast cancer according to non-random patterns. We have speculated that this phenomenon is indicative of a permissive environment for unusual cell proliferation. Elucidation of the risk factors for proliferative breast disease may yield important insight into the early development of micro-environments supportive of breast cancer development. In this report we examine common breast cancer risk factors in association with histologically confirmed common proliferative breast diseases. The study population of 1181 women was drawn from the Clinical Breast Cancer Project Study. *In situ* breast cancer, atypical breast disease, and benign proliferative lesions (sclerosing adenosis, hyperplasia of the usual type, and columnar cell lesions) individually and in combination have similar risk factors to those observed in breast cancer – age, age at menopause, and obesity/never exercising. Age is a significant risk factor for all of these lesions in premenopausal women; age at menopause and obesity or lack of exercise significantly increases the risk of many of these lesions in postmenopausal women. The co-current

presence of at least two benign lesions is associated with similar statistically significant risk factors for *in situ* and atypical breast pathologies in pre-menopausal and post-menopausal women.

***BC-TRCOE PILLAR # 5 Clinical Care:***

The objectives of the Clinical Care Pillar are to:

- a) Decrease the negative psychological impact on the patient of having an evaluation or treatment intervention for breast disease by utilizing objective measurement instruments to longitudinally assess the patient's psychological response to evaluation and intervention, and base modifications of these procedures on those results.
- b) Create and maintain an environment (medical, physical, psychological) conducive to the multiple needs of the patient undergoing breast disease evaluation / treatment.
- c) Recruit patients into the various BC-TRCOE protocols to obtain the clinical data and biospecimens needed to meet the BC-TRCOE's translational research goals.

This pillar of the BC-TRCOE is the foundation upon which all the success of and project rests. Without patients enrolled in our biospecimen repository protocols, there would be no translational research center of excellence. These patients come from the clinical care environment. Since its inception in 2000, the CBCP (now the BC-TRCOE) has had as a priority, the development and staffing of the core clinical centers at Walter Reed Army Medical Center, Walter Reed National Military Medical Center, the Joyce Murtha Breast Care Center in Windber, PA and at our newest site, the Pat and Lesly Sajack Breast Center at Anne Arundel Medical Center in Annapolis, Maryland. Under the direction of Lorraine Tafra, MD more than 500 newly diagnosed cases of breast cancer are seen at AAMC each year.

At each center the staff is dually trained as clinical/research providers, to seamlessly integrate the need for a strong research focus in the clinical center with the requirement to provide state-of-the-art clinical care to the patients.

The reputation of the BC-TRCOE is that of an exceptional translational research project with very possibly the world's most pristine collection of breast tissue. This has resulted in a number of well respected medical centers expressing interest in joining us as research partners.

The care of our patients is provided by Physicians, Advance Practice Nurses (Nurse Practitioners) and certified Nurse Navigators with all personnel having as their prime job description, the research aspects of the BC-TRCOE.

Both the former Walter Reed Army Medical Center and the Walter Reed National Military Medical Center in Bethesda, MD have built state of the art comprehensive breast care centers with women's imaging co-located with the breast care center. The

Joyce Murtha Breast Care Center in Windber, PA is a beautiful facility that rivals the most highly regarded breast centers in the country. Our partners at Anne Arundel have opened a brand new cutting-edge breast center. They have added 4 additional exam rooms, a bariatric exam room with a two procedure rooms, a counselor's office, conference room, patient library, multiple consult rooms to meet with patients & families, and a clinical trials room. At the new Walter Reed National Military Medical Center in Bethesda, Maryland much input was provided by the staff of the former National Naval Medical Center and the Walter Reed Army Medical Center to ensure a state of the art, cutting edge breast center, co-located with breast imaging, was the result. The Breast Center has a procedure room and recovery room enabling surgery within the center. The Breast Imaging Center has a designated Aurora Breast MRI machine. The merging of the Army and Navy Breast Centers has been an overwhelming success with the clinical care provided primarily by the former NNMC staff and the research component provided by the former WRAMC staff. The Breast Care and Translational Research Center of Excellence received a 3 year full accreditation by the NAPBC (National Accreditation Program for Breast Centers) on September 11, 2012 and the surveyor made the following comments:

" this is a model breast center and exemplary program.....at every level, there was a palpable pride and service mentality... The Breast Care and Research Center at Walter Reed is truly a cut above the rest."

At the new Walter Reed National Military Medical Center, the former National Naval Medical Center and Walter Reed Army Medical Center have combined to become the Breast Care and Imaging Translational Research Center of Excellence. We anticipate seeing between 5,000 and 6,000 patients per year and diagnose approximately 200 new breast cancers per year.

**The MULTIDISCIPLINARY CONFERENCE** occurs every Thursday. Newly diagnosed breast cancer patients and their significant other(s) are assigned a examination room where they meet the various providers that comprise the interdisciplinary breast care team. Providers include a breast surgeon, a medical oncologist, a radiation oncologist, a psychologist, nurse navigators/case managers, a physical therapist, and a plastic surgeon. Each specialty has the opportunity to meet privately and evaluate each patient. The benefit of this approach is that instead of having individual appointments spread over several days or weeks, the patient need only schedule one day to see all the various providers.

This allows us to educate, facilitate and coordinate a comprehensive breast treatment plan for the patient that maximizes treatment options and streamlines patient care in a patient-focused environment.

It also allows us to discuss the various research protocols with patients and, if they agree, obtain informed written consent and complete, with the assistance of a research nurse, an extensive clinical questionnaire that captures more than 500 fields of clinical data.

After morning activities, the breast care team assembles and conducts an interdisciplinary conference to discuss each patient's case, resulting in a comprehensive treatment plan built on a team consensus. The results of the conference are then reviewed with the patient/significant other and time is provided to clarify and ask questions.

In an article entitled "*Hidden in Plain View – Integration of Effective Patient Partnerships with Evidence Based Medicine in the Military Health System –The Walter Reed Army Medical Center Clinical Breast Care Project*" the MULTIDISCIPLINARY CONFERENCE (former TEMPLATE day at WRAMC) is described by the following quote "This method of physician communication and consensus avoids conflicting messages to the patient and allows for the best evidence-based approach. The literature supports the notion that a group decision is superior to sequential individual ones. Staff and patient satisfaction, stability of staff retention, and continuous improvement attitude, creates optimal outcomes in a safe, high quality, and supportive, attractive physical environment. As Residency Director, clearly COL Shriver leads by example and has an impact on physicians during their graduate medical education by experiencing how idealized care can be operationalized in a military setting. This model of care should certainly be considered as the National Capital Area moves forward with the merger of Walter Reed Army Medical Center with the National Naval Medical Center and becomes the Walter Reed National Military Medical Center (WRNMMC) in 2011."

### **Psycho-Social Oncology Services:**

The WRNMMCB has 2 designated psycho-oncology staff members (1 psychologist, 1 social worker) to assess all newly diagnosed breast cancer patients. The yearly rates average about 110 new patients. These staff members also evaluate and counsel other cancer patients throughout the hospital system.

Services provided for the breast center:

- Psycho-social assessment including NCCN distress scale for all new breast cancer patients
- Individual, couples, family counseling related to diagnosis and other stressors. Telephone support also available for those for whom transportation is an issue
- referral to community resources as needed
- multi-discipline collaboration and coordination of care

Groups offered:

- Breast Cancer support group meets 2x/month 1.5hr each-open to all stages and ages occasional guest speakers
- Breast Cancer yoga class meets 2x/month 1hr each session-all stages, ages, and abilities
- Young women's (ages 20-40) Cancer (all cancers) educational support group 8 week sessions 1.5hr ea.

### III. KEY RESEARCH ACCOMPLISHMENTS

- Continue to build the world's largest and best-characterized biospecimen repository of breast disease specimens, now numbering over 50,000 total specimens from more than 5800 enrolled research patients.
- Identified, characterized, and verified the chromosomal changes that occur during the progression of breast tissue from benign to early malignant disease.
- Developed an albumin-depletion methodology for the proteomics analysis of serum samples that was critical to our ability to find important low-abundance proteins and peptides that are found in the bloodstream of patients with breast cancer (i.e., an important step in developing a "breast cancer blood test").
- Implemented the working version of the database and data warehouse system that CBCP has been developing for four years, that integrates the clinical, molecular, pathologic, and biorepository aspects of CBCP translational research, along with robust query capability and analysis tools.
- Gene expression difference have been found between African American women and European American women that may lead to insights into the differences in breast cancer severity seen between these populations.
- Continued studies of low-abundance proteins and peptides that are found in the bloodstream of patients with breast cancer (i.e., an important step in developing a "breast cancer blood test")
- Further enhanced the use of the database and data warehouse system that CBCP has developed for last five years, to integrate the clinical, molecular, pathologic, and biorepository aspects of CBCP translational research. Its robust query capability and analysis tools have assisted in stratifying patients populations for our studies.
- Other developments in the Biomedical informatics core have included development of a new patient centric data model, new tools for microarray data QA, MS data protein peak detection and alignment and an analysis of breast disease co-occurrence.
- Gene expression analysis of primary (node negative) versus primary (node positive) revealed a 70 gene signature distinguishing two types of tumors.
- Gene expression of blood RNA from 100 patients with invasive breast cancer compared to blood RNA from 100 disease free controls has identified several candidate markers of breast cancer detection.
- Specimens continue to be genotyped with a panel of 26 markers that give insight into the stage and prognosis of the tumors.
- A collaboration with Vanderbilt University to examine proteomics differences associated with lymph node metastasis continues and has resulted in the discovery of patterns of protein expression that may distinguish tumors that has lymph node metastasis from those that do not.
- We are currently performing targeted research aimed at identifying blood based biomarkers for the early detection of breast cancer. For this research activity, we are utilizing a multiplex assay platform (xMAP) to analyze a panel of biomarkers made up of matrix metalloproteinase 1, 2 and 9 (known mediators of the extracellular matrix modification) and growth factors. Such a

combination will provide positive predictive values that approach 100% compared to the earlier studies that utilized only the matrix metalloproteinases with predictive values of 80%.

- A new collaboration with Pacific Northwest Laboratory (PNNL) has been initiated to use proteomic analysis of breast cancer tissue to identify protein biocarkers for metastasis and disease progression.
- Further enhancements have been made to the database and data warehouse system that CBCP has developed for last five years, to integrate the clinical, molecular, pathologic, and biorepository aspects of CBCP translational research. The data warehouse and the On-Line Analytical Processing tool as the interface have proven to be a powerful tool in supporting scientific research at WRI and WRAMC and the underlying patient centric data model is being modified to support other types of disease.
- Gene expression difference have been found between African American women and Caucasian American women that may lead to insights into the differences in breast cancer severity seen between these populations. Recently we generated gene expression data from both tumors and non-malignant tissues from AAW and CW. 18 genes were differentially expressed in tumors and 13 genes in non-malignant breast specimens, including PSPHL, SOS1 and CRYBB2, which are differentially expressed in both tumors and non-malignant tissues.
- The Biomedical informatics core has continued development of the patient centric data model, enhanced tools for microarray data QA, and further analysis of breast disease co-occurrence.
- In 2010, we published the results of our Allelic Imbalance evaluation in non-neoplastic diseases, concluding that ADH and CCH are in fact, genomically naive and that the multitude of studies that have evaluated pre-neoplastic lesions from breasts with invasive breast tumor cannot represent the status of pure early lesions.
- Copy number evaluation is being performed on fresh frozen genomic DNA samples from women with breast cancer, and a variety of pathological classifications using Affymetrix 500K SNP chips. Copy number and LOH analysis is being performed using Affymetrix genotyping console. Preliminary data were presented at the AACR meeting in April.
- In collaboration with Pacific Northwest Laboratory (PNNL) using the previously developed Accurate Mass Tag data base for proteins expressed in breast tissue we have examined breast tumors for the proteins that are markers of metastasis to lymph nodes.

### **IIIa. COMPLEMENTARY ACTIVITIES SUPPORTED BY SEPARATE FUNDING**

- **The Komen Promise Grant:**

Work continues on a Komen Promise grant awarded to Hallgier Rui of Thomas Jefferson University in 2009. On behalf of the Clinical Breast Care Project at Walter Reed Army Medical Center, the Henry M. Jackson Foundation is committed to participate as a consortium collaborator on this grant submission: **“Therapy-relevant Stratification of Breast Cancer Patients: Integrating Pathology and Biomarker Analyses”** As for the progress, Aim 1 of the grant involves the block and clinical data collection efforts and most of our work has been directed at this activity. The CBCP planned to submit approx. 500 cases. To date, we have collected blocks on 344 cases and have completed the clinical data forms on 237 cases. Of these 237 forms, we have performed a Quality Control check on 122 of them and sent them to the bioinformatics team for data entry testing.

Aim 2 of the grant involves participation in the Phase I/II clinical trial “Phase I/II trial of AMG 706 in Combination with Ixabepilone and Capecitabine in Women with Locally Recurrent or Metastatic Breast Cancer”. Clinical trial preparation, consent form and IRB/safety board approval efforts are being led by co-PI Edith Mitchell. As stated in the previous annual report, Amgen halted access to Motesanib, since then we have been working with Abbot Laboratories because they approved the use of Veliparib (ABT-888). We have just received word that the clinical trial is now approved at Thomas Jefferson University. We will submit the protocol here at WRNMMC for IRB approval, meet with LTC Christopher Gallagher in Medical Oncology who will assign a research nurse to coordinate this clinical trial and TJU will provide the needed training. Simultaneously, Research nurse Deborah Nielsen will work with TJU to obtain required FDA documents for the IND, organize protocol, submit the protocol for IRB approval at Anne Arundel Medical center. TJU will then provide Deborah the training needed to implement this clinical trial. The Phase II clinical trial is titled “A Randomized phase II trial to determine the impact on pathologic complete response with the addition of Carboplatin or Carboplatin and Veliparib to standard chemotherapy in the neoadjuvant treatment of stage IIB and III triple-negative breast cancer.”

From the CBCP perspective, the next few months will also be directed at collecting blocks on the remaining 237 cases (approx.) and completing the completion and QC of all data forms.

- **The Cancer Genome Atlas project for NCI:**  
The Cancer Genome Atlas-Breast Cancer project team recently published a milestone research article in Nature (on-line Sep. 23, 2012; in print, Oct. 4, 2012, <http://www.nature.com/nature/journal/v490/n7418/full/nature11412.html> entitled “Comprehensive molecular portraits of human breast tumors”. The publication was accompanied with a press release from NIH and drew intensive attention from major news media. Three components of the Clinical Breast Care Project, Walter Reed National Military Medical Center

(WRNMMC), Windber Research Institute (WRI), and MDR Global, participated in this study. We functioned as a breast cancer tissue supplier, and we also participated in data analysis.

As a “Tissue Supplier” for the project, CBCP supplied over 10% of all the breast cancer cases used in the study. Among the 16 tissue suppliers, CBCP tissues were of the highest quality with a passing rate of 89.6%, and three batches passed at 100% which the director of the TCGA Program Office claimed as “amazing”. CBCP tissues were procured following stringent Standard Operating Procedures and QA programs, and banked at the Tissue Banking facility at WRI following another stringent set of SOP and QA programs. The quality of CBCP tissues, is attested to in this TCGA project.

The Biomedical Informatics group at WRI led the clinical data QA of the TCGA-BC project, and developed methods to examine and enhance the quality of data fields that are important to the study. We also performed genomic data analysis. As part of the “Disease Working Group” for the project, we enforced the ASCO/CAP guidelines for HER2 clinical calls, and used additional molecular data to supplement such calls where needed. We authored Section II of the Supplementary Methods of the paper.

The CBCP authors in this paper are: Craig D. Shriver, Jeffrey A. Hooke, and Leigh Campbell from WRNMMC; Richard J. Mural, Hai Hu, Stella B. Somiari, Caroline Larson, Brenda Deyarmin, Leonid Kvecher, Yaqin Chen, and Chunqing Luo from WRI; and Albert J. Kovatich from MDR Global.

The CBCP has shipped cases. All 110 have been QC'd and 102 cases passed. That's a 92.7% pass rate which is outstanding, far better than many prestigious cancer centers and a testimony to the quality of our biobanking pillar of this program.

#### **IV. REPORTABLE OUTCOMES**

**The number of subjects recruited to Core CBCP Research Protocols for the period 21 April 2005 to 30 November 2012 are as follows:**

**Clinical Breast Care Project Walter Reed Army Medical Center/Walter Reed National Military Medical Center**

- **Enrolled – 1,664 subjects**
- **Blood sample collections – 12,788**
- **Tissue sample collections - 3,871**

**The Joyce Murtha Breast Care Center in Windber, Pennsylvania**

- **Enrolled - 732 subjects**
- **Blood sample collections - 8,012**
- **Tissue sample collections - 474**

**Anne Arundel Medical Center in Annapolis Maryland**

- **Enrolled - 1005 subjects**
- **Blood sample collections - 4,816**
- **Tissue sample collections - 2,613**

**TOTAL Enrollment – 3,401 subjects**

**TOTAL Blood Sample Collections – 25,616**

**TOTAL Tissue Sample Collections – 6,958**

**The number of subjects patients seen by Raymond Weiss, Medical Oncologist who may be at high risk for breast Cancer for the period 21 April 2005 to 30 November 2012 are as follows:**

**Joyce Murtha Breast Care Center - 229 (Risk Reduction program began 16 April 2009)**

**WALTER REED ARMY MEDICAL CENTER/WALTER REED MILITARY MEDICAL CENTER - 2359**

**CBCP OFFSITE MEETINGS**

The CBCP conducted the 6<sup>th</sup>, 7<sup>th</sup>, 8<sup>th</sup>, 9<sup>th</sup>, 10<sup>th</sup> and 11<sup>th</sup> Annual Offsite meetings during this period of time. Offsite meetings began on Sunday evenings with dinner and concluded Wednesday mornings at noon. The annual offsite meetings present a forum for sharing the current state of research in the project.

Presentations were given by scientists from Windber Research Institute, Staff from Walter Reed, collaborators working on projects and noted guest speakers. The meeting also provided an opportunity for the Scientific Advisory Board and Steering Committee to meet and provide feedback to project leadership. The 3 day offsite meeting began with a presentation by a breast cancer survivor to remind us that the patient with breast cancer is always at the center of all we do, all we study and all we learn. Please see ATTACHMENTS 12 and 13 for an agenda and letter of invitation.

## V. CONCLUSIONS

Breast cancer is the most common non skin-related malignancy among women in the western world. It accounts for one-third of all cancers diagnosed. Age is the single most important risk factor for the development of breast cancer, as incidence and mortality both increase with age. However, a significant number of breast cancers are diagnosed among young women. Each year, over 10,000 new breast cancer cases are detected in women under the age of 40. Over 90% of these occur among women aged 30-39 years and 8 women per 10,000 in this age group die from breast cancer every year. Breast cancer is the single leading cause of death in women aged 40-49 years. Despite the low absolute risk of breast cancer in women under 40 years of age, the incidence is increasing in this age group. The incidence in younger women is probably underestimated based on the current understanding of the biology of breast cancer. Given the doubling time of most breast carcinomas a one centimeter breast cancer is estimated to have been in situ for a period of four to six years. Cancers identified in women age 40 to 45 originated at an age (> 40) when screening mammography would not have been recommended. Thus, given current screening recommendations, many breast cancers that develop in the fourth decade of life are not discovered by standard screening methods until the woman is older. The age-specific incidence for breast cancer among African American women is two times that of white women among those 30-39 years of age and the breast cancer mortality rate is nearly two times higher for African American women as compared to Caucasian women (12.8 versus 6.6 per 10,000 women, respectively). These differences have been attributed to significant ethnic and racial variations in the stage of disease at time of diagnosis, prevalence of adverse patient and tumor-related prognostic factors, and differences in provided cancer treatment. African American women are diagnosed with later stage disease and more aggressive tumors than white women accounting for the poorer overall survival among African American women with breast cancer. Reduction in ethnic and racial differences in breast cancer treatment outcome is a principal aim of modern day health care, and aggressive screening programs in young women aged 30-39 to detect disease at an early stage could assist in achieving this goal and improving cancer-related survival.

### BC-TRCOE Military Relevance:

Breast cancer is the most common non-skin cancer in women. It is the single greatest cause of cancer deaths among women under 40, and is a significant cause of mortality for women in the United States Armed Forces. Breast cancer mortality among women <50 years accounts for >40% of years of life lost due to this disease. The economic, social and emotional costs to families are far greater when a young woman dies than when an older woman dies of breast cancer. The more aggressive nature of the disease in young patients along with the attendant costs underscores the importance of early detection of breast cancer in young women. Breast cancer is a curable disease if it is detected early; as such early detection is related to survivorship, cost of treatment and quality of life for the affected woman.

The active duty military force is approximately 20% female. Most of these service members are in the age range (30-40 years) where routine screening for breast cancer consists only of clinical breast examination. Both mammography and clinical breast examination have a very poor accuracy in the young active duty force in determining which breast abnormalities require treatment, and which are benign and can be left alone. The immense scale and impact of this problem for the military can be assessed by the fact that there were over 2,000 cases of breast cancer diagnosed in active duty service members over the last ten years (source: ACTURS DoD Tumor Registry data). Furthermore, there were over 8,000 unnecessary breast biopsies done on active duty women during this time because it takes 4 breast biopsies of normal non-cancerous lesions to find each individual breast cancer. Hence, women often need to take lengthy amounts of time off from duty in order to undergo multiple tests leading up to the biopsy as well as time off from duty because of the biopsy itself. This translates into approximately 10,000 weeks, or 30 person-years, of time lost in the evaluation of normal, benign breast lesions in active duty service members. This would be unacceptable for any other healthcare issue, and should be so for this one. Unfortunately, at the present time there is no screening tool available to diagnose breast cancer in the early, curable stages for women under the age of 40, who make up the vast majority of women in military uniform.

#### BC-TRCOE Mission:

Utilize our unique biorepository of well characterized biospecimens from a broad subset of patients with breast cancer and other breast diseases to broaden our knowledge of the etiology and pathology of breast disease. Leverage the technological advances in genomic and proteomic research to further our understanding of breast cancer through discoveries in molecular biology, pathway analysis and systems biology that can be readily translated in to the clinic.

#### BC-TRCOE Specific Aims:

This project is structured around three major themes: clinically relevant molecular profiles, evaluation of genetic risk and tumor biology. These themes inform research across the five BCTRC pillars: (1) Breast Cancer Risk Reduction; (2) Biorepository; (3) Focused Research (Genomics, Proteomics); (4) Biomedical Informatics; and (5) Clinical Care.

#### Study Design:

The project utilizes a multidisciplinary approach for researching breast diseases and breast cancer. This multidisciplinary model integrates prevention, screening, diagnosis, treatment and continuing care, but the project is further unique in the incorporation of advances in risk reduction, biomedical informatics, tissue banking and translational research. The project is based on a Discovery Science paradigm, leveraging high-throughput molecular biology technology and our unique clinically well-characterized tissue repository with advances in biomedical informatics leading to hypothesis-generating discoveries that are then tested in hypothesis-driven experiments.

**Relevance:**

The BC-TRCOE is the continuation of the Clinical Breast Care Project (CBCP) that has been ongoing for more than 12 years. Its uniqueness and relevance has been attested to by numerous world-class cancer experts, from the innumerable public and private presentations made by BC-TRCOE investigators over the years, as well as by the extensive publication record of BC-TRCOE researchers. BC-TRCOE has developed the world's largest biorepository of human breast tissues and biospecimens. BCTRC has one of the few fully integrated genomic and proteomic molecular biology research programs in the nation devoted exclusively to research in breast diseases that is linked directly to the clinic and the patients (translational research).

**Hypothesis/Rationale/Purpose:**

The Walter Reed National Military Medical Center and Windber Research Institute have partnered in the BC-TRCOE since its inception. This program has become a leader in the fight against breast disorders and cancer. Recognition of this leadership has prompted Congress to establish the BC-TRCOE as the Breast Cancer Translational Research Center of Excellence (BC-TRCOE), one of five centers so designated by Congress in 2008. The project utilizes a multidisciplinary bed-to-bench-to-bedside approach as the standard for treating and studying breast diseases and breast cancer. This multidisciplinary model integrates advances in risk reduction, prevention, screening, diagnosis, treatment and continuing care with cutting edge research incorporating advanced methods from biomedical informatics, tissue banking, high throughput biology and translational research. These efforts focus on decreasing the morbidity and mortality of breast cancer among American women.

The BC-TRCOE currently utilizes the facilities, resources and expertise of the Walter Reed National Military Medical Center (WRNMMC), the Windber Research Institute (WRI), the Windber Medical Center (WMC) and the Anne Arundel Medical Center (AAMC). The BCTRC fosters a collaborative and collegial working relationship with its partners, other government agencies, academic institutions, commercial/industry leaders, and other non-profit organizations. The BC-TRCOE has a five pronged approach to the study and treatment of breast disease based on five interlocking pillars: (1) Breast Cancer Risk Reduction; (2) Biorepository; (3) Focused Research (including: Genomics, and Proteomics Research); (4) Biomedical Informatics; and (5) Clinical Care.

The overall purpose of the BC-TRCOE is to provide a balanced environment between the two competing and yet complementary research paradigms of hypothesis-driven research and hypothesis-generating research, in a translational research organization that unites clinical capabilities (patients, nurses, clinicians) with research capabilities (genomics, proteomics, immunohistochemistry and whole genome DNA sequencing) to analyze molecular and developmental pathways that are central to the diagnosis and treatment of breast disease. The critical foundations to this approach are provided by the tissue biorepository and biomedical informatics platforms.

We believe that there are three broad areas where the BC-TRCOE stands poised to make major contributions to breast cancer research and its translation into clinical practice. These areas include the identification of molecular profiles of disease with high clinical relevance, deepening our understanding of the genetic risk of breast disease and the enhancement of our understanding of breast tumor biology. These three themes are supported by the five pillars of the BC-TRCOE. There is no doubt that our understanding of the biology of Breast Cancer in all of its various forms and manifestations remains incomplete. We believe that our high-value repository of biospecimens, our strong biomedical informatics infrastructure and our research base with strong internal and external collaborations puts us in an excellent position to make contributions to the understanding of breast disease that will have impact on the quality of life for breast cancer patients and their families.

#### Clinically Relevant Molecular Profiling:

This is cross-cutting theme with clinical, risk assessment and basic research components. The primary focus of this theme is to evaluate the utility of existing molecular profiles that have relevance to risk assessment, diagnosis, prognosis and therapy in a clinical setting and to discover new profiles that can be evaluated in the clinic. Projects within this theme have well defined translational goals. The development of comprehensive and highly informative molecular profiles will be a foundation for the development and delivery of personalized/individualized medicine. A variety of research modalities will be used to identify these profiles including immunohistochemistry, gene and protein expression analysis and genetic profiling including Next Generation DNA Sequencing. Two major new initiatives are outlined below one involving the development and testing of clinically relevant immunohistochemical profiles for disease stratification and therapeutic guidance and the other using complete genomics sequencing of tumor and matched normal DNA to develop clinically relevant profiles that could aid in disease diagnosis, prognosis and therapy selection.

#### Genetic Risk:

The rapid developments of high throughput genotyping and genomic sequencing of individuals has reminded the research community of the power of family studies in the assessment of genetic risk. Evaluating family risk and translating that into individual risk is the primary goal of this theme. There is both clear clinical relevance and a strong basic research component to this theme. Understanding the underlying biology of observed racial disparities in disease prevalence, presentation and outcome will also be a major part of this effort. The interaction of the theme with the Risk Reduction pillar of the BCTRC and the number of projects outlined below that deal with research into the basis of the observed racial disparities in breast cancer morbidity and mortality point out the relevance of this theme to the overall goals of the BCTRC.

#### Tumor Biology:

A unique combination of resources and expertise put the BCTRC in a strong position to further our understanding of the basic biology of breast disease including breast

cancer. Many of the projects outlined in the Focused Research pillar address basic problems associated with tumor heterogeneity. The tumor microenvironment and stromal interactions, metastasis and recurrence, as well as the role of cancer stem cells and tumor evolution affecting the efficacy of treatment are emphasized. We firmly believe that a robust understanding of breast tumor biology is a key to the successful translation of the research performed at the BCTRC to the clinic.

In summary the Clinical Breast Care Project, a collaborative effort between Walter Reed and Windber, has resulted in excellent working relationships and collaborations between the two sites on all five of the project's main pillars. The project continues to achieve its goals and looks forward to further continuance of this great vision and what will be a national resource, into the future.

## **VI. REFERENCES**

N/A

## **VII. APPENDICES**

- APPENDIX 1: Publications and Abstracts

## **VIII. ATTACHMENTS**

- ATTACHMENT 1: CBCP Personnel 2005 - 2012
- ATTACHMENT 2: WRI Personnel 2005 - 2012
- ATTACHMENT 3: Blood Library Protocol 22 May 2012
- ATTACHMENT 4: Blood ICD (Informed Consent Document)
- ATTACHMENT 5: Tissue Protocol 22 May 2012
- ATTACHMENT 6: Tissue ICD (Informed Consent Document)
- ATTACHMENT 7: Questionnaire
- ATTACHMENT 8: Caffeine Intake Scale
- ATTACHMENT 9: Fat Intake Scale
- ATTACHMENT 10: Job Stress Survey
- ATTACHMENT 11: Offsite invitation
- ATTACHMENT 12: Offsite Agenda
- ATTACHMENT 13: Publications 2005-2007
- ATTACHMENT 14: Publications 2008-2012 Part 1
- ATTACHMENT 15: Publications 2008-2012 Part 2

## APPENDIX 1 PUBLICATIONS, PRESENTATIONS AND ABSTRACTS

### 2005

Moroni M, Pedraza L, Storrer CE, Foley RJ, Shriver CD, Ponniah S, Peoples GE. Immunoprofiling the Cancer-Specific Humoral Immune Response in Sera from Node-Positive and Node-Negative Breast Cancer Patients. American Association for Cancer Research (AACR), Apr 2005.

Proctor C, Ponniah S, Peoples GE, Shriver CD, Mueller SC, Moroni M. Inverse Relationship Between Syk and HER2/neu in Breast Epithelial Cancer Cell Lines. American Association for Cancer Research (AACR), Apr 2005.

Dehqanzada ZA, Storrer CE, Hueman MT, Foley R, Harris KA, Jama Y, Shriver CD, Ponniah S, Peoples GE. Serum Cytokine Levels in Breast Cancer Patients Before and After Receiving a HER2/neu Vaccine Assessed Using Multiplex Technology. American Association for Cancer Research (AACR), Apr 2005.

Mittendorf EA, Storrer CE, Dehqanzada ZA, Hueman MT, Foley RJ, Jama Y, Harris KA, Shriver CD, Peoples GE, Ponniah S. Induction of CD8 T Lymphocytes Specific for the Recognition of a Breast Cancer-Associated Polymorphism in the HER2/neu Protein in a Clinical Trial with the HER2 E75 Peptide Vaccine. American Association for Cancer Research (AACR), Apr 2005.

Dehqanzada ZA, Storrer CE, Hueman MT, Foley R, Harris KA, Jama Y, Shriver CD, Ponniah S, Peoples GE. Correlation of Serum MCP-1 Levels with Clinical Prognostic Variables and Endogenous and Induced Anti-Tumor Immunity in Breast Cancer Patients. Gary Wratten Surgical Symposium, Apr 2005.

Hueman MT, Stojadinovic A, Storrer CE, Dehqanzada ZA, Foley RJ, Jama Y, Harris KA, Shriver CD, Peoples GE, Ponniah S. Reduction of Regulatory T Cell Levels and Enhanced Anti-Tumor Immune Responses in Breast Cancer Patients Receiving a HER2 Peptide (E75) Vaccine. American Association for Cancer Research (AACR), Apr 2005.

Hueman MT, Stojadinovic A, Gurney JM, Dehqanzada ZA, Storrer CE, Shriver CD, Ponniah S, Peoples GE. Circulating Regulatory CD4+CD25+ T Cells, Which Enable Malignant Tumor to Evade the Immune System, Are Present in Higher Levels in Breast Cancer Patients and Are Decreased to Lower, Healthy Patient Levels in Breast Cancer Patients After Vaccination with a HER2/neu Peptide (E75) Vaccine. Gary Wratten Surgical Symposium, Apr 2005.

Gurney JM, Hueman MT, Woll MM, Ryan GB, Storrer CE, Fisher C, Shriver CD, Ioannides CG, Ponniah S, Peoples GE. Clinical Trial Results of a HER2/neu (E75)

Vaccine to Prevent Recurrence in High Risk Breast Cancer Patients. American Society of Clinical Oncology (ASCO), May 2005.

Maskery S, Zhang Y, Jordan R, Hu H, Shriver CD, Hooke JA, Liebman M. A Novel Computational Analysis of Heterogeneity in Breast Tissue. 2005. Proceedings from the 18th IEEE Symposium on Computer Based Medical Systems, Dublin, Ireland 24-25 Jun 2005 pp. 395-400.

Mittendorf EA, Storrer CE, Shriver CD, Ponniah S, Peoples GE. Evaluation of the CD107 Cytotoxicity Assay for the Detection of Cytolytic CD8+ Cells Recognizing HER2/neu Vaccine Peptides. Breast Cancer Res Treat. 2005 Jul, 92 (1): 85-93. PMID: 15980996.

Peoples GE, Gurney JM, Hueman MT, Woll MM, Ryan GB, Storrer CE, Fisher C, Shriver CD, Ioannides CG, Ponniah S. Clinical Trial Results of a HER2/neu (E75) Vaccine to Prevent Recurrence in High-Risk Breast Cancer Patients. J Clin Oncol. 20 Oct 2005; 23(30):7536-45. ePub 12 Sep 2005. PMID: 16157840.

Ellsworth RE, Ellsworth DL, Deyarmin B, Hoffman LR, Love B, Hooke JA, Shriver CD. Timing of Critical Genetic Changes in Human Breast Disease. Annals of Surgical Oncology, Dec 2005; 12(12):1054-60. ePub 10 Oct 2005. PMID: 16228814.

Shriver CD, Gutchell V, Bronfman L, Field LA, Ellsworth RE. The Clinical Breast Cancer Project: An Important Resource for Studying Breast Cancer in African American Women. American Society of Human Genetics (ASHG), Oct 2005.

Ellsworth RE, Ellsworth DL, Deyarmin B, Hooke JA, Shriver CD. HER2 Gene Amplification is a Marker for Global Genomic Instability. American Society of Clinical Oncology (ASCO), Oct 2005.

Somiari SB, Somiari RI, Heckman CM, Olsen CH, Jordan RM, Russell SJ, Shriver CD. Quantitative Analysis of Active and Total MMP2 and MMP9 in Serum of Individuals with Breast Disease, Breast Cancer, and at Low or High Risk of Developing Breast Cancer. Cancer Research, Oct 2005.

Dehqanzada ZA, Storrer CE, Foley RJ, Harris KA, Jama YH, Shriver CD, Ponniah S, Peoples GE. Correlation of Serum MCP-1 Levels with Clinical Prognostic Variables and Endogenous and Induced Anti-Tumor Immunity in Breast Cancer Patients. American College of Surgeons, Oct 2005.

Mittendorf EA, Storrer CE, Shriver CD, Ponniah S, Peoples GE. Evaluation of the HER2/neu-Derived Peptide GP2 for Use in a Peptide-Based Breast Cancer Vaccine Trial. American College of Surgeons, Oct 2005.

Maskery S, Tsavachidou D, Liebman M. Modeling and Simulation of the Perimenopause Transition. North American Menopause Society 16th Annual Meeting, 28 Sep-1 Oct 2005, San Diego, CA.

Maskery S. Co-Occurrence Analysis for Discovery of Novel Breast Cancer Pathology Patterns. Oral presentation. 6th Annual CBCP offsite meeting, 13-16 Nov 2005, Cumberland, MD.

Field L. Identification of Molecular Changes Associated with Breast Tumor Phenotype in African American Women. Oral presentation, 6th Annual CBCP offsite agenda, 13-16 Nov 2005, Cumberland, MD.

Ellsworth DL. Molecular Alterations in Tumor-Adjacent Normal Tissues: Implications for Defining Tumor-Stromal Interactions. Oral Presentation, 6th Annual CBCP offsite meeting, 13-16 Nov 2005, Cumberland, MD.

Dehqanzada ZA, Storrer CE, Foley RJ, Harris KA, Jama YH, Shriver CD, Ponniah S, Peoples GE. Elevated Serum MCP-1 Level is Associated with Favorable Clinical Prognostic Factors and Immune Response and Can be Induced by a HER2/*neu* (E75) Vaccine in Breast Cancer Patients. San Antonio Breast Cancer Symposium (SABCS), Dec 2005, San Antonio, TX

Field L, Bradley L, Kane J, Deyarmin B, Hooke JA, Ellsworth RE, Shriver CD. Gene Expression Differences in Disease-Free Breast Tissue in African American Women Versus Caucasian Women. 5 A.D.a. Poster Presentation. San Antonio Breast Cancer Symposium (SABCS), 8-11 December 2005.

Ellsworth RE, Ellsworth DL, Love B, Hoffman LR, Deyarmin B, Hooke JA, Shriver CD. Genomic Alterations as a Pathological Aid to Classifying Ductal Carcinoma In-Situ. Abstract. San Antonio Breast Cancer Symposium (SABCS), Dec 2005, San Antonio,

Mittendorf EA, Storrer CE, Shriver CD, Ponniah S, Peoples GE. The Impact of a Polymorphism in the HER2/Protein on a Peptide Vaccination Strategy. BC Research & TX.

Guo X, Shriver CD, Hu H, Liebman M. Semantic Similarity-Based Validation of Human Protein-Protein Interactions. Poster Presentation. Stanford, CA.

Field L. Breast Cancer in African American and Caucasian Women. Penn Highlands Community College.

Bukowski V, Maity T, Kolli K, Kirchner D, Brzeski H, Liebman M, Shriver CD, Keeney J. Optimization of Gel Filtration Chromatography for High Molecular Weight Protein Separation in Human Serum. Poster Presentation. Showcase for BioTechnology, 2005, Johnstown, PA.

Ellsworth R, Hoffman R, Zhang Y, Karinsky Y, Snyder C, Lynch H. Identification of Genetic Modifiers in Large Families with BRCA1 and BRCA2 Mutations. 5 A.D.c, Presentation. Komen Mission Conference, Washington, DC.

Hu H, Liebman M. Data Warehouse Development for Integrative Biomedical Informatics Research. 2005b. Annual Oracle Life Science Users Group Meeting.  
Liebman, M. Translational Medicine: An Engineering Approach. American Scientist. 2005.

Ru Z A, Zhang Y, Shriver CD, Jacobs N, Liebman M. Clinical Proteomics of Breast Cancer with the Application of Multi-Dimensional Protein Identification Technology. 96th AACR Annual Meeting, Anaheim, CA.

Ru Q, Silberman J, Yang S, Brzeski H, Shriver CD, Liebman M. Human Serum Proteome of Breast Cancer. 5 A.D.c, Gordon Research Conferences: New Frontiers in Cancer Detection and Diagnosis, Santa Ynez, CA.

Ru Q, Zhu A, Zhang Y, Shriver CD, Jacobs N, Liebman M. Clinical Proteomics of Breast Cancer with the Application of Multi-dimensional Protein Identification Technology. 96th AACR Annual Meeting, 16 -20, 2005, Anaheim, CA .

Ru Q, Zhu A, Zhang Y, Silberman J, Brzeski H, Shriver CD, Liebman M. Human Serum Proteome of Breast Cancer. 2005e. 1st US HUPO Symposium, Washington, DC.

**American Association for Cancer Research (AACR), 96th Annual Meeting, 16-20 Apr 2005, Anaheim, CA. 9 Poster presentations and Abstracts.....**

**AACR # 390 “Serum Proteomic Analysis of Breast Cancer by 2D DIGE/MS Analysis”** Kolli K, Kirchner D, Russell S, Maity T, Preuss A, Harshberger T, Brzeski H, Shriver CD, Liebman M

**394 “Clinical Proteomics of Breast Cancer with the Application Multi-Dimensional Protein Identification Technology”** Ru Q, Zhu L, Shriver CD, Jacobs N, Liebman M.

**1514 “Inverse Relationship Between Syk and HER2/neu in Breast Epithelial Cancer Cell Lines”** Proctor C, Ponniah S, Peoples GE, Jr, Shriver CD, Mueller SC, Moroni M.

**1541 “Immunoprofiling the Cancer-Specific Humoral Immune Response in Sera from Node Positive and Node Negative Breast Cancer Patients”** Moroni M, Pedraza L, Storrer CE, Foley RJ, Shriver CD, Ponniah S, Peoples GE. Inverse relationship between Syk and HER2/neu in breast epithelial cancer cell lines. Proctor C, Ponniah S, Peoples GE, Jr, Shriver CD, Mueller SC, Moroni M.

**1605 “Body Mass Index and Mammography Screening in Caucasian and African-American Women”** Zhu K, Wu H, Jatoi I, Potter JF, Shriver CD

**3449 “Induction of CD8 T Lymphocytes Specific for the Recognition of a Breast Cancer-Associated Polymorphism in the HER2/neu Protein in a Clinical Trial with**

**the HER2 E75-Peptide Vaccine”** Mittendorf EA, Storrer CE, Dehqanzada Z, Hueman MT, Foley RJ, Jama Y, Harris KA, Shriver CD, Peoples GE Jr, Ponniah S.

3451 **“Reduction of Regulatory T Cell Levels and Enhanced Antitumor Immune Responses in Breast Cancer Patients Receiving a HER2 Peptide (E75) Vaccine”** Hueman MT, Stojadinovic A, Storrer CE, Dehqanzada Z, Foley RJ, Jama Y, Harris KA, Shriver CD, Peoples GE Jr, Ponniah S.

3452 **“Serum Cytokine Levels in Breast Cancer Patients Before and After Receiving a HER2/neu Vaccine Assessed Using Multiplex Technology”** Dehqanzada Z, Storrer CE, Hueman MT, Foley RJ, Harris KA, Jama Y, Shriver CD, Ponniah S, Peoples GE, Jr

4818 **“Development of a Method to Isolate and Analyze Low Abundance/Low Molecular Weight Proteins/Peptides in the Serum of Breast Cancer Patients”** Maity T, Kirchner D, Russell S, Kolli K, Brzeski H, Shriver CD, Liebman M Ellsworth RE, Leach R, Hoffman L, Snyder C, Lynch H. Identification of Genetic Modifiers in a Large BRCA1 Family. Poster Presentation. American Society of Human Genetics 55th Annual Meeting, Oct 2005.

Hu H. Biomedical Informatics at Windber Research Institute. Oral Presentation. 6th Annual CBCP offsite meeting, 13-16 Nov 2005, Cumberland, MD.

Hu H. Biomedical Informatics Research and Development at Windber Research and Development at Windber Research Institute. Presentation. Showcase for BioTechnology 2005, Johnstown, PA.

Kaufman M, Brzeski H, Maity T, Liebman M. Per Amplification of a CpG Island in the Prdm15 Gene Prior to Developing a Hypermethylation Assay for Genes Deregulated in Breast Cancer. Poster Presentation. Showcase for BioTechnology 2005, Johnstown, PA.

Hu H, Somiari SB, Copper J, Everly RD II, Heckman C, Jordan R, Somiari RI, Hooke JA, Shriver CD, Liebman MN. In Wang L, Chen K, Ong YS (Eds): Can Circulating Matrix Metalloproteinases be Predictors of breast cancer? - A Neural Network Modeling Study. Lecture Notes in Computer Science (LNCS) 2005; 3610:1039-42. Springer-Verlag. DOI: 10.1007/11539087-138. pp 1039-1042.

Brzeski H, Russell S, Sullivan A, Somiari R, Shriver CD. In Walker J. Large-format Two Dimensional Polyacrylamide Gel Electrophoresis. (Editor) The Proteomics Protocols Handbook. Humana Press Inc., NJ. 2005; pp. 119-132.

Ellsworth DL, Russell S, Deyarmin B, Sullivan A, Brzeski H, Somiari RI, Shriver CD. Laser-Assisted Microdissection in Proteomic Analyses. In Walker J. (Editor) The Proteomics Protocols Handbook. Humana Press Inc., NJ, 2005; pp. 59-66.

Maskery S, Shinbrot T. Deterministic and Stochastic Elements of Axonal Guidance. Annual Review of Biomedical Engineering, vol. 7, pp. 187-221. 2005.

Sullivan A, Brzeski H, Russell S, Somiari R, Shriver C. In Walker J. Serum or Plasma Sample Preparation for Two-Dimensional Gel Electrophoresis. (Editor) The Proteomics Protocols Handbook. Humana Press Inc, NJ. 2005; pp. 49-54

Dehqanzada ZA, Storrer CE, Hueman MT, Foley R, Harris K, Jama Y, Shriver CD, Ponniah S, Peoples GE. Assessing Serum Cytokine Profiles in Breast Cancer Patients Receiving a HER2/*neu* Vaccine Using Luminex® Technology. *Ann Surg Oncol* 2005;12:S47 (#P20).

Hu H, Brzeski H, Hutchins J, Ramaraj M, Qu L, Xiong R, Meadow S, McMichael J, Hooke JA, Somiari SB, Shriver CD, Somiari RI, Liebman MN. Development of a Comprehensive Data Warehouse for Clinical and Genomic Breast Cancer Research. *Archives of Pathology and Laboratory Medicine*: 2005 Vol. 129, No 6, pp 811-34.

## **2006**

Khoo S, Storrer CE, Harris KA, Jama YH, Dehqanzada ZA, Shriver CD, von Hofe E, Ponniah S, Peoples GE. Immunological Monitoring of In Vitro and In Vivo Immune Responses to the Ii-Key/HER2/*neu* MHC class II Peptide Vaccine in a Phase I Clinical Trial in Breast Cancer Patients. *Association for Academic Surgery/Society of University Surgeons*, Feb 2006.7

Mittendorf EA, Khoo S, Dehqanzada ZA, Hueman MT, Storrer CE, Harris KA, Shriver CD, Ponniah S, and Peoples GE. Phase Ib Trial of the HER2/*neu* Peptide (E75) Vaccine for Optimal Dosing in Node-Negative Breast Cancer Patients. *Association for Academic Surgery/Society of University Surgeons*, Feb 2006.

Sotiriou C, Wirapati P, Loi S, Harris A, Fox S, Smeds J, Nordgren H, Farmer P, Praz V, Haibe-Kains B, Desmedt C, Larsimont D, Cardoso F, Peterse H, Nuyten D, Buyse M, Van de Vijver MJ, Bergh J, Piccart M, Delorenzi M. Gene Expression Profiling in Breast Cancer: Understanding the Molecular Basis of Histologic Grade to Improve Prognosis. *Jour of the NCI*, Vol 98, No 4, 15 Feb 2006.

Somiari SB, Shriver CD, Heckman C, Olsen C, Hu H, Jordan R, Arciero C, Russell S, Garguillo G, Hooke JA, Somiari RI. Plasma Concentrations and Activity of Matrix Metalloproteinase 2 and 9 in Patients with Breast Disease, Breast Cancer and At Risk of Developing Breast Cancer. *Cancer Letters*, EL Sevier Ireland Ltd, 2006 Feb 20; 233(1):98-107. PMID:16473671.

Seebach J, Ellsworth R, Shriver CD. A Molecular Profiling of Breast Cancer. *Seminars in Breast Disease*, Mar 2006, Vol 9, No. 1, pp. 19-24. DOI: 10.1053/j.sembd.2006.10.005.

Khoo S, Kao TC, Dehqanzada ZA, Storrer CE, Harris KA, Jama YH, Shriver CD, Ponniah S, Peoples GE. Developing a Predictive Model for Disease Status in Breast Cancer Patients Using Th1 and Th2 Serum Cytokine Profiles. *Society of Surgical Oncology (SSO)*, Mar 2006.

Dehqanzada ZA, Storrer CE, Foley RJ, Harris KA, Jama YH, Kao TC, Shriver CD, Ponniah S, Peoples GE. Luminex Analysis of Serum IL-6, Eotaxin and MCP-1 Levels can Predict Stage of Disease in Breast Cancer Patients. Society of Surgical Oncology (SSO), Mar 2006.

Hamsher L, Smith A, Dehqanzada ZA, Khoo S, Shriver CD, Ponniah S, Peoples GE, Moroni M. Pattern of Serum Immunoreactivity Against Breast Cancer Cell Lysates may Predict Severity of Disease in Breast Cancer Patients. Society of Surgical Oncology (SSO), Mar 2006.

Mittendorf EA, Storrer CE, Flora MN, Shriver CD, Peoples GE, Ponniah S. The Development of a Novel Method to Identify Immunogenic Peptides for Use in Cancer Vaccines. Society of Surgical Oncology (SSO), Mar 2006.

Dehqanzada ZA, Storrer CE, Hueman MT, Foley RJ, Harris KA, Jama YH, Shriver CD, Ponniah S, Peoples GE. Assessing Serum Cytokine Profiles in Breast Cancer Patients Receiving a HER2/*neu* Vaccine Using Luminex® Technology. *Oncol Reports* Mar 2007; 17(3):687-94.

Ellsworth RE, Janssen M, Meyers S, Patney H, Hadix H, Seth B, Ellsworth DL, Shriver CD. Identification of cSNPs in Environmental Response Genes Contributing to Breast Cancer Etiology. Poster Presentation. American Association for Cancer Research (AACR), 1–5 Apr 2006, Washington, DC.

Maskery S, Zhang Y, Hu H, Hooke J, Shriver CD, Liebman M. Breast Pathology Co-Occurrence in Stratified Populations, Implications for Breast Cancer Development in Different Populations. Abstract. American Association for Cancer Research (AACR), 1–5 Apr 2006, Washington, DC.

Qu L, Hu H, Xiong R, Brzeski H, Raab M, and Liebman MN. A Web-Based Tissue Gene Expression Visualization Tool Developed on the caCORE Cancer Research Informatics Platform. caBIG 2006 Annual Meeting, 9-11 Apr 2006, Arlington, VA.

Hu H, Correll M, Sheldon J, Liebman M. Development of an Oracle-based Integrative Biomedical Informatics Data Warehouse. Platform Presentation. Oracle Life Science Users Group Meeting, 3 Apr 2006, Boston, MA.

Jordan R, Brzeski H, London W, Mural R, Liebman M, Hooke JA, Shriver CD, Hu H. Observations of Different Laterality Imbalance of Breast Biopsies and Breast Disease Between Caucasian and African American Women. Proceedings of the American Society of Clinical Oncology (ASCO), 2-6 Jun 2006, Atlanta, GA.

Maskery S, Zhang Y, Hu H, Shriver CD, Hooke JA, Liebman M. Caffeine Intake, Race, and Risk of Invasive Breast Cancer, Lessons Learned from Data Mining a Clinical Database. Podium Presentation and Paper, Proceedings. 19th IEEE International Symposium on Computer-Based Medical Systems, 22-23 Jun 2006, Salt Lake City, UT.

Maskery S, Hu H, Mural R, Hooke JA, Shriver CD, Liebman M. Separate DCIS Progression Pathways Derived from Breast Disease Heterogeneity Data. Poster Abstract Accepted American Society of Clinical Oncology Annual Meeting, Atlanta, GA, 2-6 Jun 2006.

Ellsworth RL, Becker T, Ellsworth DL, Love B, Hooke JA, Shriver CD. Patterns of Genomic Instability Clinical Outcomes in Patients with Invasive Breast Cancer. Presentation. American Society of Clinical Oncology (ASCO), 2 Jun 2006, Atlanta, GA.

Maskery S, Zhang Y, Jordan R, Hu H, Hooke JA, Shriver CD, Liebman M. Co-Occurrence Analysis for Discovery of Novel Breast Pathology Patterns. IEEE Transactions on Information Technology in Biomedicine 2006 Jul; 10(3):497-503. PMID: 16871717.

Mittendorf EA, Storrer CE, Shriver CD, Ponniah S, Peoples GE. Investigating the combination of trastuzumab and HER2/neu peptide vaccines for the treatment of breast cancer. Ann Surg Oncol. 2006 Aug;13(8):1085-98. Epub 2006 Jul 24.

Ellsworth RE, Janssen M, Meyers S, Patney H, Hadix H, Seth B, Ellsworth DL, Shriver CD. Identification of cSNPs in Environmental Response Genes Contributing to Breast Cancer Etiology. Poster Presentation. American Association for Cancer Research (AACR), 1-5 Apr 2006, Washington, DC.

Maskery S, Zhang Y, Hu H, Hooke J, Shriver CD, Liebman M. Breast Pathology Co-Occurrence in Stratified Populations, Implications for Breast Cancer Development in Different Populations. Abstract. American Association for Cancer Research (AACR), 1-5 Apr 2006, Washington, DC.

Qu L, Hu H, Xiong R, Brzeski H, Raab M, and Liebman MN. A Web-Based Tissue Gene Expression Visualization Tool Developed on the caCORE Cancer Research Informatics Platform. caBIG 2006 Annual Meeting, 9-11 Apr 2006, Arlington, VA.

Hu H, Correll M, Sheldon J, Liebman M. Development of an Oracle-based Integrative Biomedical Informatics Data Warehouse. Platform Presentation. Oracle Life Science Users Group Meeting, 3 Apr 2006, Boston, MA.

Jordan R, Brzeski H, London W, Mural R, Liebman M, Hooke JA, Shriver CD, Hu H. Observations of Different Laterality Imbalance of Breast Biopsies and Breast Disease Between Caucasian and African American Women. Proceedings of the American Society of Clinical Oncology (ASCO), 2-6 Jun 2006, Atlanta, GA.

Maskery S, Zhang Y, Hu H, Shriver CD, Hooke JA, Liebman M. Caffeine Intake, Race, and Risk of Invasive Breast Cancer, Lessons Learned from Data Mining a Clinical Database. Podium Presentation and Paper, Proceedings. 19th IEEE International Symposium on Computer-Based Medical Systems, 22-23 Jun 2006, Salt Lake City, UT.

Maskery S, Hu H, Mural R, Hooke JA, Shriver CD, Liebman M. Separate DCIS Progression Pathways Derived from Breast Disease Heterogeneity Data. Poster Abstract Accepted American Society of Clinical Oncology Annual Meeting, Atlanta, GA, 2-6 Jun 2006.

Ellsworth RL, Becker T, Ellsworth DL, Love B, Hooke JA, Shriver CD. Patterns of Genomic Instability Clinical Outcomes in Patients with Invasive Breast Cancer. Presentation. American Society of Clinical Oncology (ASCO), 2 Jun 2006, Atlanta, GA.

Maskery S, Zhang Y, Jordan R, Hu H, Hooke JA, Shriver CD, Liebman M. Co-Occurrence Analysis for Discovery of Novel Breast Pathology Patterns. IEEE Transactions on Information Technology in Biomedicine 2006 Jul; 10(3):497-503. PMID: 16871717.

Mittendorf EA, Storrer CE, Shriver CD, Ponniah S, Peoples GE. Investigating the combination of trastuzumab and HER2/neu peptide vaccines for the treatment of breast cancer. Ann Surg Oncol. 2006 Aug;13(8):1085-98. Epub 2006 Jul 24.

Ellsworth DL, Seeley EH, Ellsworth RE, Deyarmin B, Sanders ME, Cornett DS, Hooke JA, Caprioli RM, Shriver CD. Direct Tissue Characterization of Protein Expression in Metastatic Breast Cancer. Podium Presentation. San Antonio Breast Cancer Symposium (SABCS), Dec 2006, San Antonio, TX.

Ellsworth RE, Weyandt J, Patney H, Shriver CD. Role of the Q356R BRCA1 Variant in Sporadic Breast Cancer. Poster Presentation. San Antonio Breast Cancer Symposium (SABCS), 14-17 Dec 2006, San Antonio, TX.

Ellsworth RE, Weyandt J, Patney H, Shriver CD. Utilizing Genes Identified from Peripheral Blood Leukocyte Microarray Data to Identify Clinical Subsets of Breast Histopathologies. Poster Presentation. San Antonio Breast Cancer Symposium (SABCS), 14-17 Dec 2006, San Antonio, TX.

Seebach J, Field L, Love B, Hooke JA, Ellsworth RE, Shriver CD. Identification of Gene Expression Differences in Primary Breast and Metastatic Lymph Node Tumors. Poster Presentation. San Antonio Breast Cancer Symposium (SABCS), 14-17 Dec 2006, San Antonio, TX.

Jordan R, Brzeski H, Hu H, Shriver CD, Mural R, Liebman MN, Yang Y. A Small Set of Genes Identified from Peripheral Blood Leukocyte Microarray Data can Classify Breast Cancer Patients on the Basis of Clinical Parameters. San Antonio Breast Cancer Symposium (SABCS), 14-17 Dec 2006, San Antonio, TX.

Somiari SB, Kovatich AJ, Baran PN, Jordan R, Maskery S, Zhang Y, Deyarmin B, Hooke JA, Shriver CD. The Expression of Cell Cycle Regulating Proteins and Steroid Hormone Receptor Status in Invasive Breast Cancer: Association with Tumor Grade and

Stage. San Antonio Breast Cancer Symposium (SABCS), 14-17 Dec 2006, San Antonio, TX

Ellsworth RE, Hooke JA, Kane J, Baran P, Kovatich A, Shriver CD. Discordance of Pathological Characteristic Between Primary Breast Tumors and Metastatic Lymph Nodes. Poster Presentation. San Antonio Breast Cancer Symposium (SABCS), 14-17 Dec 2006, San Antonio, TX.

Ellsworth RE, Love B, Kane J, Ellsworth DL, Hooke JA, Shriver CD. Levels and Patterns of Genomic Instability are Correlated with Histological Grading of Invasive Breast Tumors. San Antonio Breast Cancer Symposium (SABCS), 14-17 Dec 2006, San Antonio, TX.

Dunn M, Vernalis M. Translational Cardiovascular Health. Podium Presentation at the Translational Health: The Next Generation of Medicine Conference, Pasquerilla Conference Center, Johnstown, PA.

Ellsworth DL. Physiological, Psychometric, and Molecular Responses to a Coronary Heart Disease Reversal Program. Podium Presentation. Translational Health: The Next Generation of Medicine Conference, Pasquerilla Conference Center, Johnstown, PA.

Ellsworth RE. The Clinical Breast Care Project: An Important Resource in Investigating Environmental and Genetic Contributions to Breast Cancer in African American Women. Podium Presentation. Translational Health: The Next Generation of Medicine Conference, Pasquerilla Conference Center. Johnstown, PA.

Field L. Identification of Blood-based Biomarkers to Detect Breast Cancer. Podium Presentation. Translational Health: The Next Generation of Medicine Conference, Pasquerilla Conference Center, Johnstown, PA.

Hu H. A Data Warehouse for Translational Research. Podium Presentation. Translational Health: The Next Generation of Medicine Conference, Pasquerilla Conference Center, Johnstown, PA.

Liebman M, Mural R. Translational Medicine: Defining It's Role and Future. Podium Presentation. Translational Health: The Next Generation of Medicine Conference, Pasquerilla Conference Center, Johnstown, PA.

Mural R. The Impact of Genomic on Translational Research. Podium Presentation at the Translational Health: The Next Generation of Medicine Conference, Pasquerilla Conference Center, Johnstown, PA.

Yang Y, Heckman C., Shriver Cd, Becker T, Liebman M, Brzeski H. Gene Expression Profiles in White Blood Cells of Patients with Breast Cancer – Correlation with Detectability of Tumor by Mammographic Screening. Breast Cancer Res Treat, vol. 94, no. Suppl 1, p. S59-S60. 2006.

Yang S, Guo X, Yang YC, Papcunik D, Heckman C, Hooke JA, Shriver CD, Liebman MN, Hu H. Detecting Outlier Microarray Slides by Correlation and Percentage of Outliers Spots. *Cancer Informatics* 2006;2:351-360.

Hu H, Mural RJ, Liebman MN (Editors). *Biomedical Informatics: A Translations Approach*. Artech House Publishers.

Deng HW, Shen H, Liu Y, Hu H (Editors). *Current Topics in Human Genetics: Studies of Complex Diseases*. World Scientific Publishing Co.

Hu H, Kvecher L. Data Tracking Systems. In: Hu H, Mural RJ, Liebman MN (Editors). *Biomedical Informatics: A Translations Approach*. Artech House Publishers. Chapter 7 (submitted; to be published in 2008).

Ellsworth DL. Physiological, Psychometric, and Molecular Responses to a Coronary Heart Disease Reversal Program. Podium Presentation. *Translational Health: The Next Generation of Medicine Conference*, Pasquerilla Conference Center, Johnstown, PA.

Ellsworth RE. The Clinical Breast Care Project: An Important Resource in Investigating Environmental and Genetic Contributions to Breast Cancer in African American Women. Podium Presentation. *Translational Health: The Next Generation of Medicine Conference*, Pasquerilla Conference Center. Johnstown, PA.

Field L. Identification of Blood-based Biomarkers to Detect Breast Cancer. Podium Presentation. *Translational Health: The Next Generation of Medicine Conference*, Pasquerilla Conference Center, Johnstown, PA.

Hu H. A Data Warehouse for Translational Research. Podium Presentation. *Translational Health: The Next Generation of Medicine Conference*, Pasquerilla Conference Center, Johnstown, PA.

Liebman M, Mural R. *Translational Medicine: Defining It's Role and Future*. Podium Presentation. *Translational Health: The Next Generation of Medicine Conference*, Pasquerilla Conference Center, Johnstown, PA.

Mural R. The Impact of Genomic on Translational Research. Podium Presentation at the *Translational Health: The Next Generation of Medicine Conference*, Pasquerilla Conference Center, Johnstown, PA.

Yang Y, Heckman C., Shriver Cd, Becker T, Liebman M, Brzeski H. Gene Expression Profiles in White Blood Cells of Patients with Breast Cancer – Correlation with Detectability of Tumor by Mammographic Screening. *Breast Cancer Res Treat*, vol. 94, no. Suppl 1, p. S59-S60. 2006.

Yang S, Guo X, Yang YC, Papcunik D, Heckman C, Hooke JA, Shriver CD, Liebman MN, Hu H. Detecting Outlier Microarray Slides by Correlation and Percentage of Outliers Spots. *Cancer Informatics* 2006;2:351-360.

Hu H, Mural RJ, Liebman MN (Editors). *Biomedical Informatics: A Translations Approach*. Artech House Publishers.

Deng HW, Shen H, Liu Y, Hu H (Editors). *Current Topics in Human Genetics: Studies of Complex Diseases*. World Scientific Publishing Co.

Hu H, Kvecher L. Data Tracking Systems. In: Hu H, Mural RJ, Liebman MN (Editors). *Biomedical Informatics: A Translations Approach*. Artech House Publishers. Chapter 7 (submitted; to be published in 2008).

Hu H. Data Centralization. In: Hu H, Mural RJ, Liebman MN (Editors). *Biomedical Informatics: A Translations Approach*. Artech House Publishers. Chapter 8 (submitted; due to be published in 2008).

Hu H, Liebman MN. In: Deng HW, Shen H, Liu Y, and Hu H (Editors). *An introduction to biomedical informatics*. *Current Topics in Human Genetics: Studies of Complex Diseases*. World Scientific Publishing Co. Chapter 18.

## **2007**

No author. Intergration of Effective Patient Partnership. *U.S. Medicine & Navy Medicine*. February 2007.

Callaghan K, Ellsworth DL, Ellsworth RE, Becker T, Hooke J, Shriver CD. Genomic Instability and the Development of Metastatic Lymph Node Tumors. *Society of Surgical Oncology (SSO)*, 16 Apr 2007.

Ellsworth D, Ellsworth R, Becker T, Deyarmin B, Patney H, Jordan R, Hooke J, Shriver CD. Genomic Heritage of Sentinel Lymph Node Metastases: Implications for Clinical Management of Breast Cancer Patients. *American Society for Clinical Oncology (ASCO)*, 1-5 Jun 2007, Chicago, IL.

Ellsworth RE, Deyarmin B, Shriver CD. Molecular Relationship Between HER2 and BRCA 1 in Invasive Breast Carcinoma. *Am Soc of Hum Genetics*, 7 Jun 2007.  
Ellsworth RE, Shriver CD. HER2\_BRCA 1a. *American Society of Clinical Oncology (ASCO)*, 1-4 Jun 2007.

Ellsworth RE, Seeley EH, Ellsworth DL, Sanders ME, Hooke JA, Caprioli RM, Shriver CD. Identification of Proteins Promoting Development of Metastatic Breast Tumors. *American Society of Clinical Oncology (ASCO)*, 1-4 Jun 2007.

Collins PM. Hidden in Plain View. *Navy Medicine*, p 24-25, Jun 2007.

Liebman MN, Deyarmin B, Shriver CD, Stegmaier P, Karas H, Kel A, Wingender E. Analysis of HER2/neu as a Diagnostic/Biomarker using ExPlain to Examine Signaling Pathways and Transcription Factors. 15th Annual Intern Conf on Intell Systems for Molecular Biology & 6th European Conf on Computational Biology, Vienna, Austria, 21-25 July 2007.

Bronfman L, Shriver CD, Gutchell E. The Clinical Perspective. Chapter 2, Biomedical Informatics: A Translational Approach, Aug 2007.

Ellsworth RE, Hooke JA, Shriver CD. Pathological Characteristics of Breast Tumors in African American Women Treated Within An Equal-Access Health-Care System: Biological and Molecular Contributions to the Aggressive Phenotype and Poor Clinical Outcomes. Society of Surgical Oncology (SSO), 5 Sep 2007.

Ellsworth RE, Ellsworth DL, Patney HL, Deyarmin B, Love B, Hooke J, Shriver CD. Amplification of HER2 is a Marker for Global Genomic Instability. BioMed Central Ltd. BMC Cancer 2008 Oct 14; 8:297. DOI: 10.1186/1471-2407-8-297 PMID: 18854030.

Ellsworth RE, Weyandt JD, Patney HL, Anthony K, Shriver CD. Identification of cSNPs in Environmental Response Genes Contributing to Breast Cancer. San Antonio Breast Cancer Symposium (SABCS), 23-27 Oct 2007, San Antonio, TX.

Ellsworth RE, Bronfman L, Gutchell V, Love B, Field LA, Weyandt JD, Shriver CD. The Clinical Breast Care Project: An Important Resource in Investigating Environmental and Genetic Contributions to Breast Cancer in African American Women. Cell and Tissue Banking, Jun 2008; 9(2): 109-20. DOI: 10.1007/s10561-007-9054-z. Springer Science & Business Media B.V. e-pub, 2007 Oct. PMID: 17929197.

Ellsworth RE, Ellsworth DL, Love B, Patney H, Hoffman L, Kane J, Hooke JA, Shriver CD. Correlation of Levels and Patterns of Genomic Instability with Histological Grading of DCIS. Annals of Surg Onc. 2007 Nov; 14(11):3070-7. DOI: 10.1245/s10434-007-9459-8. PMID: 17549568.

Callaghan KA, Becker TE, Ellsworth DL, Hooke JA, Ellsworth RE, Shriver CD. Genomic Instability and the Development of Metastatic Lymph Node Tumors. Ann Surg Oncol Nov 2007;14(11):3125-3132. DOI: 10.1245/s10434-007-9504-7. PMID: 17653592.

Ellsworth RE, Hooke J, Ellsworth DL, Shriver CD. Contribution of Chromosomal Alterations to the Development of Poorly-Differentiated Invasive Breast Carcinomas. San Antonio Breast Cancer Symposium (SABCS), 13-16 Dec, 2007, San Antonio, TX.

Ellsworth RE, Deyarmin B, Patney H, Hooke J, Shriver CD. HER2 Gene Amplification is a Marker for Global Genomic Instability. San Antonio Breast Cancer Symposium (SABCS), 13-16 Dec, 2007, San Antonio, TX.

Ellsworth RE, Love B, Hooke JA, Shriver CD. Identification of Chromosomal Changes Associated with the Transition from *in situ* to Invasive Breast Cancer. San Antonio Breast Cancer Symposium (SABCS), 13-16 Dec 2007, San Antonio, TX.

Ellsworth RE, Patney H, Ellsworth DL, Love B, Hooke J, Shriver CD. Genomic Discrimination of Metastatic From Non-Metastatic Primary Breast Tumors. San Antonio Breast Cancer Symposium (SABCS), 13-16 Dec 2007, San Antonio, TX.

Ellsworth RE, Seeley EH, Ellsworth DL, Hooke JA, Caprioli RM, Shriver CD. Identification of Protein Expression Differences in Invasive Breast Tumors from African American Compared to Caucasian Women. Annals of Surg Onc. and San Antonio Breast Cancer Symposium (SABCS), 13-16 Dec 2007 San Antonio, TX.

Ellsworth DL, Seeley EH, Ellsworth RE, Deyarmin B, Sanders ME, Hooke JA, Caprioli RM, Shriver CD. Tumor Microenvironment in Breast Cancer Metastasis: Direct Tissue Protein Profiling of Tumor-Associated Stroma from Invasive Breast Cancer Patients with versus without Axillary Lymph Node Metastasis. San Antonio Breast Cancer Symposium (SABCS), 13-16 Dec 2007, San Antonio, TX.

Ellsworth DL, Ellsworth RE, Patney HL, Becker TE, Deyarmin B, Jordan RM, Hooke JA, Shriver CD. Primary Tumor Heterogeneity and Sentinel Lymph Node Metastasis: Understanding Molecular Processes of Breast Cancer Metastasis. San Antonio Breast Cancer Symposium (SABCS), 13-16 Dec 2007, San Antonio, TX.

Field LA, Seebach JF, Love BJ, Deyarmin B, Kane J, Hooke JA, Ellsworth RE, Shriver CD. Identification of Gene Expression Differences in Primary Breast Tumors from Node-Negative and Node-Positive Women. San Antonio Breast Cancer Symposium (SABCS), 13-16 Dec 2007, San Antonio, TX.

Field A, Ellsworth RE, Gutchell V, Hooke JA, Shriver CD. The Clinical Breast Care Project: An Important Resource for Studying Breast Cancer. American Association for Cancer Research (AACR), 27 Nov-1 Dec 2007.

Stojadinovic A, Hooke JA, Shriver CD, Nissan A, Kovatich AJ, Tzu-Cheg K, Ponniah S, Peoples GE, Moroni M. HYOU1/Orp150 Expression in Breast Cancer. MedSciMon, Intern Medical Jour of Experimental & Clin Research, 2007 Nov; 13(11):231-39.

Bekash A, Maskery S, Kvecher L, Hooke J, Liegman MN, Shriver CD, Mural RJ, Hu H. A Pilot Study of Controversial Breast Cancer Risk Factors Using the Clinical Breast Care Project Database as a Research Environment. San Antonio Breast Cancer Symposium (SABCS), 13-16 Dec 2007, San Antonio, TX.

Field LA, Love BJ, Hadix JA, Ellsworth RE, Shriver CD. Identification of Blood-Based Biomarkers for the Detection of Breast Cancer and Lymph Node Status. San Antonio Breast Cancer Symposium (SABCS), 13-16 Dec 2007, San Antonio, TX.

Maskery S, Hu H, Liebman M, Hooke J, Shriver CD, Taioli E. Traditional Breast Cancer Risk Factors and Common Breast Pathologies in Post-Menopausal Women. San Antonio Breast Cancer Symposium (SABCS), 13-16 Dec 2007, San Antonio, TX.

Callaghan KA, Weyandt JD, Ellsworth RE, Shriver CD. Identification of Genetic Predisposition to Development of Lymph Node Metastasis in Breast Cancer Patients. San Antonio Breast Cancer Symposium (SABCS), 13-16 Dec 2007, San Antonio, TX.

Hu H, Field L, Stegmaier P, Love B, Ellsworth RE, Shriver CD, Liebman MN, Mural R. A Transcription Factor-Centric Computational Analysis of Genes Differentially Expressed in Healthy Breast Tissues from African American and Caucasian Women. San Antonio Breast Cancer Symposium (SABCS), 13-16 Dec 2007, San Antonio, TX.

Maskery S, Hu H, Liebman M, Hooke J, Shriver CD, Verbanac K, Tafra L, Rosman M. Bayesian Analysis of Recurrence in Lymph Node Positive and Lymph Node Negative Breast Cancer Patients. San Antonio Breast Cancer Symposium (SABCS), 13-16 Dec 2007, San Antonio, TX.

Tkaczuk K. A Retrospective Clinical Correlation Study of the Breast Cancer Patient's Responses to Anticancer Drugs with the Expression Level of Drug Response Indicators (DRI) Measured in Patients Archival Tumor Tissue. San Antonio Breast Cancer Symposium (SABCS), 13-16 Dec 2007, San Antonio, TX.

Ellsworth RE, Hooke JA, Love B, Ellsworth DL, Shriver CD. Contribution of Chromosomal Alterations to the Development of Poorly-Differentiated Invasive Breast Carcinomas. Am Jour of Surg Path, Dec 2007.

Jordan R. A Small Set of Genes Identified From Peripheral Blood Leukocyte Microarray Data Can Classify Breast Cancer Patients on the Basis of Clinical Parameters. San Antonio Breast Cancer Symposium (SABCS), 13-16 Dec 2007 San Antonio, TX.

Jordan RM, Hu H, Heckman CM, Kvecher L, Shriver CD, Mural R, Yang YC. Peripheral Blood Microarray Data May Aid in Predicting Lymph Node Status of Breast Cancer Patients. San Antonio Breast Cancer Symposium (SABCS), 13-16 Dec 2007, San Antonio, TX.

Seebach J, Field L, Love B, Hooke J, Ellsworth R, Shriver CD. Identification of a Gene Expression Breast Cancer Metastasis Profile. San Antonio Breast Cancer Symposium (SABCS), 13-16 Dec 2007, San Antonio, TX.

## **2008**

Ellsworth RE, Hooke JA, Love B, Kane JL, Patney HL, Ellsworth DL, Shriver CD. Correlation of Levels and Patterns of Genomic Instability with Histological Grading of Invasive Breast Tumors. Breast Cancer Res Treat, 2008 Jan; 107(2):259-65.

DOI:10.1007/s10549-007-9547-2. e-pub, 2007. Springer Science & Business Media B.V.  
PMID: 17351743.

Sann NJ, Ellsworth RE, Kvecher L, Shriver CD, Ellsworth DL. Is a Diagnosis of Invasive Breast Cancer an Effective Motivational Factor for Lifestyle Change?. Society of Surgical Oncology (SSO), 5-9 Mar 2009, Phoenix, AZ.

Ellsworth RE, Hooke JA, Love B, Kane JL, Patney HL, Ellsworth DL, Shriver CD. Correlation of Levels and Patterns of Genomic Instability with Histological Grading of Invasive Breast Tumors. *Breast Cancer Res Treat*, 2008 Jan; 107(2):259-65.  
DOI:10.1007/s10549-007-9547-2. e-pub, 2007. Springer Science & Business Media B.V.  
PMID: 17351743.

Ellsworth RE, Becker TE, Progar C, Deyarmin B, Shriver CD. Characterization of Screen Versus Non-Screen Detected Breast Tumors. Society of Surgical Oncology (SSO), 5-9 Mar 2009, Phoenix, AZ.

Weyandt J, Ellsworth RE, Hooke JA, Shriver CD, Ellsworth DL. Molecular Characteristics of Early Breast Disease. American Association for Cancer Research (AACR), 18-22 Apr 2009, Denver, CO.

Ellsworth RE, Hooke JA, Shriver CD. Effect of Equal-Access to Breast Care on Clinicopathological Phenotypes of Invasive Breast Tumors in African American Women. American Association for Cancer Research (AACR), 12-16 Apr 08, San Diego, CA.

Fields L. Gene Expression Differences in Primary Breast Tumors from African American and Caucasian Women. American Association for Cancer Research (AACR) 12-16 Apr 2008, San Diego, CA.

Becker TE, Ellsworth RE, Deyarmin B, Patney HL, Jordan RM, Hooke JA, Shriver CD, Ellsworth DL. Genomic Heritage of Lymph Node Metastases: Implications for Clinical Management of Patients with Breast Cancer. *Annals Surg Oncol* Apr 2008; 15(4):1056-63.  
DOI: 10.1245/s10434-008-98153 PMID:1846400.

Shoemaker A, Heckman C, Zhang Y, Colledge A, Deyarmin B, Kane J, Hooke J, Shriver C, Mural R, Somiari S. The Prognostic Role of Cell Cycle Regulators in Invasive Breast Cancer. American Association for Cancer Research (AACR), 12-16 Apr 2008 San Diego, CA.

Shoemaker A, Heckman C, Zhang Y, Colledge A, Deyarmin B, Kane J, Hooke J, Shriver C, Mural R, Somiari S. RNA Quality of Human Breast Tissue after Surgical Removal-Effect of Processing Method, Temperature and Storage. American Association for Cancer Research (AACR), 12-16 Apr 2008, San Diego, CA.

Fields L. Gene Expression Differences in Primary Breast Tumors from African American and Caucasian Women. American Association for Cancer Research (AACR) 12-16 Apr 2008, San Diego, CA.

Becker TE, Ellsworth RE, Deyarmin B, Patney HL, Jordan RM, Hooke JA, Shriver CD, Ellsworth DL. Genomic Heritage of Axillary Lymph Node Metastases in Breast Cancer Patients. *Ann Surg Oncol*. 2008 Apr;15(4):1056-63. Epub 2008 Feb 2.

Shriver CD. Star Wars Defense Battles Breast Cancer. Conjoint Royal Australasian College of Surgeons and the College of Surgeons of Hong Kong, 12-14 May 2008.

Shriver CD. Well-documented Breast Cancer Tissue Collection Meets High Throughput Proteomics to Find Aggressive-tumor Markers in Blood. Pacific Northwest National Laboratory (PNNL), May 2008.

Ellsworth RE, Patney HL, Ellsworth DL, Love B, Deyarmin B, Hooke JA, Shriver CD. Genomic Alterations Associated with Early Stages of Breast Tumor Metastasis. Society of Surgical Oncology (SSO), May 2008, Chicago, IL.

Shoemaker A, Heckman C, Zhang Y, Colledge A, Deyarmin B, Kane J, Hooke J, Shriver C, Mural R, Somiari S. RNA Quality of Human Breast Tissue after Surgical Removal-Effect of Processing Method, Temperature and Storage. International Society for Biological and Environmental Repositories (ISBER), 18-21 May 2008, Bethesda, MD.

Ellsworth RA, Ellsworth DL, Patney HL, Love B, Deyarmin B, Hooke JA, Shriver CD. Genomic Alterations Associated with Early Stages of Breast Tumor Metastasis. *AnnSurgOncol*, 2008 Jul; 15(7):1989-95. DOI:10.1245/s10434-008-9902-5. PMID: 18401664.

Ellsworth DL, Ellsworth RE, Patney HL, Becker TE, Deyarmin B, Hooke JA, Shriver CD. Molecular Heterogeneity Among Sentinel Lymph Node Metastases: Implications for the Sentinel Node Hypothesis. Breast Cancer Symposium, 5-7 Sep 08, Washington, DC.

Bronfman L, Shriver CD. A New Paradigm for Integrating Translational Research into Comprehensive Breast Care: The Walter Reed Army Medical Center / Windber Research Institute Experience. *Seminars in Breast Disease Journal, Exploring the Economics of Breast Care: New Perspectives, New Directions*, Sep 2008; 11(3):160-63. OI:10.1053/j.sembd.2009.04.001.

Ellsworth RE, Vertrees A, Love B, Hooke JA, Ellsworth DL, Shriver CD. Chromosomal Alterations Associated with the Transition from In Situ to Invasive Breast Cancer. *AnnalsSurgOncol*. Sep 2008; 15(9):2519-25. DOI: 10.1245/s10434-008-0051-7. PMID:18622645.

Patney HL, Deyarmin B, Hooke JA, Ellsworth RE, Love B, Shriver CD. Relationship Between Alterations of Chromosome 8q24 and Poorly-Differentiated Breast Carcinoma. San Antonio Breast Cancer Symposium (SABCS) 10-14 Dec 2008, San Antonio, TX.

Croft D, Mao X, Hooke JA, Shriver CD, Shriver M, Ellsworth RE. Admixture Mapping to Identify Breast Cancer Susceptibility Loci in African American Women. San Antonio Breast Cancer Symposium (SABCS) 10-14 Dec 2008, San Antonio, TX.

Ellsworth DL, Ellsworth RL, Patney HL, Oviedo A, George A, Croft, DT, Jr., Love B, Jordan RM, Deyarmin B, Becker TE, Hooke JA, Shriver CD. Fingerprinting Genomic Heterogeneity in Primary Breast Carcinomas and Among Sentinel Lymph Node Metastases: Implications for Clinical Management of Breast Cancer Patients. San Antonio Breast Cancer Symposium (SABCS) 10-14 Dec 2008, San Antonio, TX.

Ellsworth DL, Gillard D, Love B, Ellsworth RE, Deyarmin B, Hooke JA, Kostyniak PJ, Shriver CD. Abundance and Distribution of Polychlorinated Biphenyls (PCBs) in Breast Tissue. San Antonio Breast Cancer Symposium (SABCS) 10-14 Dec 2008, San Antonio, TX.

Field LA, Love BJ, Kane J, Deyarmin B, Hooke JA, Ellsworth RE, Shriver CD. Role of Phosphoserine Phosphatase-like in Breast Disease in African American Women. San Antonio Breast Cancer Symposium (SABCS) 10-14 Dec 2008, San Antonio, TX.

Ellsworth RE, Heckman C, Love B, Shriver CD. Identification of Breast Cancer Metastasis Initiation and Virulence Genes. San Antonio Breast Cancer Symposium (SABCS) 10-14 Dec 2008, San Antonio, TX.

Callaghan KA, Ellsworth RE, Ellsworth DL, Shriver CD. HER2 in the Era of Molecular Medicine: A Review. *Curr Cancer Ther Rev*, 2008, 4, 00-00. Bentham Science Publishers Ltd. 1573-3947/08.

Maskery SM, Hu H, Hooke J, Shriver CD, Liebman MN. A Bayesian Derived Network of Breast Pathology Co-Occurrence. *Journal of Biomedical Informatics*. 2008;41:242-50.

Ellsworth RL, Seebach J, Field LA, Heckman C, Kane J, Hooke JA, Love B, Shriver CD. Breast Metastasis Gene Signature Identification of a Gene Expression Breast Cancer Nodal Metastasis Profile. *Clinical and Environmental Metastasis Journal*, Springer, NY.

## **2009**

Field LA, Love B, Kane J, Deyarmin B, Hooke JA, Ellsworth RE, Shriver CD. Gene Expression Differences in Normal Breast Tissue from Disease-Free African American and Caucasian Women. *Breast Cancer Research Journal*, BioMed Central Ltd, Middlesex House, London, UK, Jan 2009.

Ellsworth RE, Weyandt JM, Fantacone-Campbell JL, Deyarmin B, Ellsworth DL, Hooke JA, Shriver CD. Molecular Characterization of Breast Cancer Progression: Early Lesions are not Genetically Advanced. American Association for Cancer Research (AACR) 18-22 Apr 2009 Denver, CO.

Iida J, Nesbella M, Lehman J, Mural R, Shriver CD. Circulating Tumor Cells: Emerging Technologies for Detection, Diagnostics and Treatment –Role for CD44 in Enhancing Invasion, Migration, and Growth of Triple negative (TN) Breast Cancer Cells. NIH Circulating Tumor Cells Conference, 10-11 Sep 2009, Bethesda, MD.

Stojadinovic A, Peoples GE, Jurgens JS, Howard RS, Schuyledr B, Kwon KH, Henry LR, Shriver CD, Buckenmaier CC. Standard versus pH-adjusted and Lidocaine supplemented radiocolloid for patients undergoing sentinel-lymph-node mapping and biopsy for early breast cancer (PASSION-P trial): a double-blind, randomized controlled trial. *Lancet Oncology* 2009 Sep 10(9):849-54. PMID:19664956.

Holliday C, Rummel S, Hooke JA, Shriver CD, Ellsworth DL, Ellsworth RE. Genomic Instability in the Breast Microenvironment? A Critical Evaluation of the Evidence. *Expert Reviews in Molecular Diagnosis Journal*. Oct 2009; 9(7):667-78. PMID: 19817552 (PubMed).

Ellsworth RE, Patney HL, Kane JL, Weyandt J, Hooke JA, Shriver CD. Copy Number Variation in Low- and High-Grade Breast Tumors. *Amer Society of Human Genetics (ASHG)*, 20-24 Oct 2009, Bethesda, MD.

Ellsworth RE, Seeley EH, Ellsworth DL, Deyarmin B, Hooke JA, Sanders M, Caprioli R, Shriver CD. Proteomic Discrimination of Well- from poorly-Differentiated Breast Carcinomas. *SABCS* 9-13 Dec 2009, San Antonio, TX.

Ellsworth RE, Deyarmin B, Hooke JA, Shriver CD. Clinical Implications of The Molecular Relationship Between HER2 and BRCA1. *SABCS*, 9-13 Dec 2009, San Antonio, TX .

Ellsworth RE, Field LA, Hooke JA, Kane JA, Love BL, Shriver CD. Identification of a Molecular Signature of Breast Cancer Metastasis. *SABCS*, 9-13 Dec 2009, San Antonio, TX.

Ellsworth RE, Weyandt JD, Fantacone-Campbell JL, Deyarmin B, Ellsworth DL, Hooke JL, Shriver CD. Genetic Characterization of Columnar Cellular Lesions and Atypical Ductal Hyperplasia of the Breast. *SABCS* 9-13 Dec 2009, San Antonio, TX.

Bekhash A, Saini J, Yu Fan X, Hooke JA, Mural RJ, Shriver CD, Hu H. Differential Benign Breast Disease Co-occurrence with Cancer in Caucasian and African American Women. *SABCS* 9-13 Dec 2009, San Antonio, TX.

Saini J, Kovatich A, Li X, Bekash A, Hooke JA, Mural R, Shriver CD, Hu H. Association of Clinicopathologic Characteristics with IHC-Based Breast Cancer Subtypes. SABCS 9-13 Dec 2009, San Antonio, TX.

Kvecher L, Wu V, Hooke JA, Shriver CD, Mural RJ, Hu H. An Approach to Correlate the Temporal Information to Facilitate Specimen Selection in the Breast Cancer Research Project. SABCS 9-13 Dec 2009, San Antonio, TX.

Shay M, Duman J, Hu H, Shriver CD. Converting Paper Medical Records to Electronic Version to Support Breast Cancer Translational Research and Clinical Practice. SABCS 9-13 Dec 2009, San Antonio, TX.

Iida J, Nesbella M, Lehman JM, Mural R, Shriver CD. Role for CD44 in Enhancing Invasion, Migration, and Growth of Triple negative (TN) Breast Cancer Cells. San Antonio Breast Cancer Symposium (SABCS), 9-13 Dec 2009, San Antonio, TX.

Ellsworth RE, Hooke JA, Shriver CD, Ellsworth DL. Genomic Heterogeneity of Breast Tumor Pathogenesis. ClinMedOnc 2009; 3:77-85. Libertas Academica Ltd. Available on web at <http://www.la-press.com>.

Kvecher L, Wu V, Hooke JA, Shriver CD, Mural RJ, Hu H. An Approach to Correlate the Temporal Information to Facilitate Specimen Selection in the Breast Cancer Research Project. Cancer Research. 2009;69; 763s.

## **2010**

Ellsworth RE, Ellsworth DL, Weyandt JD, Fantacone-Campbell JL, Deyarmin B, Hooke JA, Shriver CD. Chromosomal Alterations in Pure Nonneoplastic Breast Lesions: Implications for Breast Cancer Progression. AnnSurgOncol, DOI: 10.1245/s10434-010-0910-x, 27 Jan 2010. PMID: 20107913 [PubMed].

Ellsworth RE, Ellsworth DL, Weyandt JD, Fantacone-Campbell JL, Deyarmin B, Hooke JA, Shriver CD. Chromosomal Alterations in Pure Breast Lesions: Implications for Breast Cancer Progression. Society of Surgical Oncology (SSO), 3-7 Mar 2010, St Louis, MO.

Ellsworth RE, Sann NJ, Kvecher L, Ellsworth DL, Shriver CD. Is a Diagnosis of Invasive Breast Cancer an Effective Motivational Factor for Lifestyle Change?. National Consortium of Breast Centers (NCBC), 20-24 Mar 2010, Las Vegas, NV.

Ellsworth RE, Oviedo AL, Kane JL, Shriver CD. Breastfeeding, Baby Boomers and Breast Cancer. Society of Surgical Oncology (SSO) Mar 2011.

Stojadinovic A, Eberhardt C, Henry L, Eberhardt J, Elster E, Peoples G, Nissan A, Shriver C. Development of a Bayesian Classifier for Breast Cancer Risk Stratification: A

Feasibility Study. *Eplasty*, vol 10, pp 203-216, 29 Mar 10, ISSN: 1937-5719. Open Science Co. LLC, Springfield, IL.

Ellsworth RE, Field LA, Love B, Deyarmin B, Hooke JA, Shriver CD. Identification of Differentially Expressed Genes in Breast Tumors from African American Compared to Caucasian Women. AACR 2-6 Apr 2011, Orlando, FL.

Valente AL, Ellsworth RE, Lubert S, Holliday C, Shriver CD. Identification of CDH1 mutations in patients with lobular carcinoma with and without a family history of gastric cancer. AACR 2-6 Apr 2011, Orlando, FL.

Peck AR, Witkiewicz AK, Liu C, Stringer GA, Klimowicz AC, Pequignot E, Freydin B, Tran TH, Yang N, Rosenberg AL, Hooke JA, Kovatich AJ, Nevalainen MT, Shriver CD, Hyslop T, Sauter G, Rimm DL, Magliocco AM, Rui H. Loss of Nuclear Localized and Tyrosine Phosphorylated Stat5: A Predictor of Poor Clinical Outcome and Increased Risk of Antiestrogen Therapy Failure in Breast Cancer. AACR 2-6 Apr 2011, Orlando, FL.

Ellsworth RE, Patney HL, Kane JL, Hooke JA, Shriver CD. Copy Number Variation in Low-and High-Grade Breast Tumor. AACR 17-21 Apr 2010, Washington, DC.

Rummel S, Ellsworth RE, Kane J, Deyarmin B, Hooke JA, Ellsworth DL, Shriver CD. Genomic Instability of Breast Stroma. American Association for Cancer Research (AACR) 17-21 Apr 2010, Washington, DC.

Ellsworth RE, Field LA, Love B, Kane JL, Hooke JA, Shriver CD. Breast Cancer Metastasis Gene Signature: The Search for the Holy Grail. AACR, 17-21 Apr 2010, Washington, DC.

Milburn M, Rosman M, Mylander C, Liang, Wen, Hooke JA, Kovatich AJ, Campbell L, Shriver C, Taft L. Does Tumor Immunohistochemistry Discordance Between Core Needle Biopsy and Mastectomy Raise Concerns for Treatment Recommendations? American Society of Breast Surgery (ASBS), 27 Apr-1 May 2011.

Ryder A, Witkiewicz AK, Liu C, Stringer GA, Klimowicz AC, Tran TH, Torhorst J, Yang N, Rosenberg AL, Hooke JA, Kovatich AJ, Nevalainen MT, Shriver CD, Hyslop T, Sauter G, Rimm DL, Magliocco AM, Rui H. Loss of Nuclear Localized and Tyrosine Phosphorylated Stat5 in Breast Cancer Predicts Unfavorable Clinical Outcome and Increased Risk of Anti-Estrogen Therapy Failure. *Clinical Cancer Research* Apr-May 2010.

Ellsworth RE, Field LA, Love B, Kane JL, Hooke JA, Shriver CD. Breast Cancer Metastatic Gene Signature: The Search for the Holy Grail. *J Clin Oncol* (Meeting Abstracts) May 2010 vol. 28 no. 15\_suppl 1075.

Ellsworth RE, Field LA, Love B, Kane JL, Hooke JA, Shriver CD. Breast Cancer Metastasis Gene Signature: The Search for the Holy Grail. American Society of Clinical Oncology (ASCO), Jun 2010, Chicago, IL.

Ellsworth RE, Sann NJ, Kvecher L, Ellsworth DL, Shriver CD. Is a Diagnosis of Invasive Breast Cancer an Effective Motivational Factor for Lifestyle Change? American Journal of Clinical Oncology (AJCO), Oct 2010.

Ellsworth RE, Field LA, Love B, Kane JL, Hooke JA, Shriver CD. Breast Cancer Metastasis Gene Signature: The Search for the Holy Grail. Personalized Cancer Therapy & Prevention, 8-9 Oct 2010, Houston, TX.

Ellsworth RE, Patney HL, Kane JA, Hooke JA, Shriver CD. Role of Chromosome 16q in Breast Tumorigenesis. SABCS 8-12 Dec 2010, San Antonio, TX.

Ellsworth RE, Valente AL, Field LA, Kane JL, Love B, Tafra L, Shriver CD. Genetic Signature Discriminating Metastatic from Non-Metastatic Small Tumors. SABCS 8-12 Dec 2010, San Antonio, TX.

Ellsworth R, Deyarmin B, Patney H, Shriver C, Ellison K, Thornton D, Tafra L, Cheng Z, Rosman M. Genetic Discrimination of Aggressive from Indolent DCIS. SABCS 8-12 Dec 2010, San Antonio, TX.

Ellsworth RE, Fantacone-Campbell JL, Valente A, Shriver C. Chromosomal Alterations Associated with Poor Outcome in Breast Cancer Patients. SABCS 8-12 Dec 2010, San Antonio, TX.

Bekhash A, Saini J, Li X, Rapuri P, Hooke JA, Kovatich A, Mural RJ, Shriver CD, Hu H. Ethnicity Difference of Benign Breast Diseases in Breast Cancer and Non-Cancer Patients. SABCS 8-12 Dec 2010, San Antonio, TX.

Rapuri PB, Li X, Brillhart G, Deyarmin B, Kvecher L, Hu H, Hooke J, Shriver C, Mural R. Comparison of Gene Expression Profiles of Lymph Node Metastatic and Non-Metastatic ER+ve Breast Tumors in Pre and Post Menopausal Women. SABCS 8-12 Dec 2010.

Zhou J, Somiari S, Lubert S, Saini J, Hooke J, Mural R, Shriver C, Brinckerhoff C. The Impact of Matrix Metalloproteinase-1 Promoter 1G/2G Polymorphism on Breast Diseases. SABCS 8-12 Dec 2010, San Antonio, TX.

Kopco R, Iida J, Lehman J, Somiari S, Somiari R, Mural R, Shriver C. Development of DNA aptamers that inhibit invasion, migration, and growth of breast cancer cells. SABCS 8-12 Dec 2010, San Antonio, TX.

Saini J, Li X, Kvecher L, Larson C, Croft D, Yang Y, Hooke JA, Shriver CD, Mural RJ, Hu H. Differential Gene Expression Analysis among Post-menopausal Caucasian

Invasive Breast Cancer, Benign and Normal Subjects. SABCS 8-12 Dec 2010, San Antonio, TX.

Field L, Rummel S, Shriver CD, Ellsworth RE. The Role of *PSPHL* in Breast Cancer in African American Women. SABCS 8-12 Dec 2010, San Antonio, TX.

Ellsworth RE, Decewicz DJ, Shriver CD, Ellsworth DL. Breast Cancer in the Personal Genomics Era. *Current Genomics*, 2010; 11:146-161.

Buckenmaier III CC, Kwon KH, Howard RS, McKnight GM, Shriver CD, Fritz WT, Garguilo GA, Joltes KH, Stojadinovic A. Double-Blinded, Placebo-Controlled Prospective Randomized Trial Evaluating the Efficacy of Paravertebral Block With and Without Continuous Paravertebral Block Analgesia in Outpatient Breast Cancer Surgery. *Pain Medicine* 2010; 11:790-799. Wiley Periodicals, Inc.

Field LA, Deyarmin B, Shriver CD, Ellsworth DL, Ellsworth RE. Laser Microdissection for Gene Expression Profiling. *Methods in Molecular Biology*, Vol 218.

Zhou J, Mural R, Shriver CD, Somiari SB. The Potential Role of MMPs as Prognostic Markers in Breast Cancer. *Biochemistry Research International*.

## **2011**

Bronfman L, Friend M. Are We Army or Navy? Don't Ask, Don't Tell - The Challenge of Merging Two Breast Centers into One. NCBC 13-16 Mar 11, Las Vegas, NV.

Hu H, Correll M, Kvecher L, Osmond M, Clark J, Bekash A, Schwab G, Gao D, Gao J, Kubatin V, Shriver CD, Hooke JA, Maxwell LG, Sheldon JG, Liebman MN, Mural RJ. A Data Warehouse for Translational Research. 10th Annual BIO-IT World Conference and Expo, 12-14 Apr 2011, Boston, MA.

Peck AR, Witkiewicz AK, Liu C, Klimowicz AC, Stringer GA, Pequignot E, Freydin B, Yang N, Tran TH, Rosenberg AL, Hooke JA, Kovatich AJ, Shriver CD, Rimm DL, Magliocco AM, Hyslop T, Rui H. Loss of nuclear localized Stat5a in breast cancer is associated with tumor progression and unfavorable clinical outcomes. *J Clin Oncol*. 2011 Jun 20;29(18):2448-58. Epub 2011 May 16.

Ellsworth RE, Field L, Shriver CD. Molecular signature discriminating triple negative breast tumors between African American and Caucasian women. AACR Conference, 18-21 Sep 2011, Washington, DC.

Field LA, Deyarmin B, Ellsworth RE, Shriver CD. Identification of blood-based biomarkers for breast cancer detection. AACR Advances in Breast Cancer Research Meeting, 12-15 Oct 2011, San Francisco, CA.

Kvecher L, Wu W, Kohr J, Shriver CD, Mural RJ, Hu H. DCU: A Data Correction Utility to Correct Clinicopathologic Data in a Data Warehouse. AIMA 2011 Annual Symposium, 22-26 Oct 2011, Washington, DC.

Zhou J, Brinckerhoff C, Lubert S, Yang K, Saini J, Hooke JA, Mural RJ, Shriver CD, Somiari S. Analysis of Matrix Metalloproteinase-1 Gene Polymorphisms and Expression in Benign and Malignant Breast Tumors. *Cancer Invest.* 2011 Nov;29(9):599-607.

Luo C, Hu H, Kvecher L, Kovatich AJ, Hooke JA, Chen Y, Mural RJ, Shriver CD. A comparative study of the triple-negative and basal-like breast cancer subtypes. 1st Annual TCGA Scientific Symposium, 17-18 Nov 2011, National Harbor, MD.

Bekhash A, Hooke JA, Chen Y, Kovatich AJ, Kvecher L, Mural RJ, Shriver CD, Hu H. Fibroadenomatoid changes have a higher occurrence rate in middle-aged benign breast disease patients with the trend retained in cancer patients. San Antonio Breast Center Symposium (SABCS), 6-10 Dec 2011, San Antonio TX.

Kovatich AJ, Kvecher L, Chen Y, Bekhash A, Hooke JA, Shriver CD, Mural RJ, Hu H. Subtype-specific co-occurrence of atypical hyperplasia and in situ carcinoma with invasive breast cancers. San Antonio Breast Center Symposium (SABCS), 6-10 Dec 2011, San Antonio TX.

Bekhash A, Kovatich AJ, Chen Y, Hooke JA, Kvecher L, Mural RJ, Shriver CD, Hu H. Fibrocystic changes have different age-dependent patterns in benign, in situ, and invasive breast cancer patients. San Antonio Breast Center Symposium (SABCS), 6-10 Dec 2011, San Antonio TX.

Ellsworth RE, Valente AL, Kane JK, Shriver CD. The effect of breastfeeding on molecular characteristics of invasive breast cancer. San Antonio Breast Center Symposium (SABCS), 6-10 Dec 2011, San Antonio TX.

Ellsworth RE, Valente AL, Kane JL, Ellsworth DL, Shriver CD. Gene expression alterations in the lymph node microenvironment in response to successful metastatic colonization. San Antonio Breast Center Symposium (SABCS), 6-10 Dec 2011, San Antonio TX.

Ellsworth R, Croft D, Ellsworth D, Shriver C. Effect of obesity on gene expression in invasive breast tumors. San Antonio Breast Center Symposium (SABCS), 6-10 Dec 2011, San Antonio TX.

Ellsworth DL, Croft DT Jr, Field LA, Deyarmin B, Kane J, Ellsworth RE, Hooke JA, Shriver CD. Congruence between patterns of micro RNA expression and Histologic grading of invasive breast carcinomas. San Antonio Breast Center Symposium (SABCS), 6-10 Dec 2011, San Antonio TX.

Zhou J, Brinckerhoff C, Lubert S, Yang K, Saini J, Hooke JA, Mural R, Shriver CD, Somiari SB. Matrix Metalloproteinase-1 2G Insertion Polymorphism is a Prognostic Marker for Predicting Breast Cancer Severity. *Breast Cancer Research and Treatment, Cancer Investigation* DOI: 10.3109/073579072011.621915.

Field LA, Love B, Deyarmin B, Hooke JA, Shriver CD, Ellsworth RE. Identification of differentially expressed genes in breast tumors from African American compared to Caucasian women. *Cancer Journal*, 2012 Mar 1;118(5):1334-44. doi: 10.1002/cncr.26405. Epub 2011 Jul 28. DOI: 10.1002/cncr.26405.

Hu H. Long Term Collaboration between Windber Research Institute and IDBS Yields Major Impacts on Breast Cancer Care and Understanding. Press Release for Research on Data Management for Translational Research Published in *The Journal of Biomedical Informatics*, Dec 2011.

Voegtly LM, Campbell JL, Shriver CD, Ellsworth RE. Molecular alterations associated with early and late breast cancer mortality. *Ann Surg Oncol Journal*.

Ellsworth RE, Valente AL, Shriver CD, Bittman B, Ellsworth DL. Impact of Lifestyle Factors on Quality of Life and Prognosis in Breast Cancer Survivors in the United States. *Expert Reviews Pharmacoeconomics and Outcomes Research Journal* – submitted for publication.

Iida J, Lehman JR, Clancy R, Somiar S, Ellsworth RE, Mural RJ, Shriver CD. Role for  $\beta$ -catenin/TCF signaling pathways in regulation of migration and growth of triple negative breast cancer cells. *PLoS ONE Journal, Public Library of Science* – submitted for publication.

## **2011**

Hu H, Correll M, Kvecher L, Osmond M, Clark J, Bekash A, Schwab G, Gao D, Gao J, Kubatin V, Shriver CD, Hooke JA, Maxwell LG, Sheldon JG, Liebman MN, Mural RJ. A Data Warehouse for Translational Research. 10th Annual BIO-IT World Conference and Expo, 12-14 Apr 2011, Boston, MA.

Peck AR, Witkiewicz AK, Liu C, Klimowicz AC, Stringer GA, Pequignot E, Freydin B, Yang N, Tran TH, Rosenberg AL, Hooke JA, Kovatich AJ, Shriver CD, Rimm DL, Magliocco AM, Hyslop T, Rui H. Loss of nuclear localized Stat5a in breast cancer is associated with tumor progression and unfavorable clinical outcomes. *J Clin Oncol*. 2011 Jun 20;29(18):2448-58. Epub 2011 May 16.

Ellsworth RE, Field L, Shriver CD. Molecular signature discriminating triple negative breast tumors between African American and Caucasian women. AACR Conference, 18-21 Sep 2011, Washington, DC.

Field LA, Deyarmin B, Ellsworth RE, Shriver CD. Identification of blood-based biomarkers for breast cancer detection. AACR Advances in Breast Cancer Research Meeting, 12-15 Oct 2011, San Francisco, CA.

Zhou J, Brinckerhoff C, Lubert S, Yang K, Saini J, Hooke JA, Mural RJ, Shriver CD, Somiari S. Analysis of Matrix Metalloproteinase-1 Gene Polymorphisms and Expression in Benign and Malignant Breast Tumors. *Cancer Invest.* 2011 Nov;29(9):599-607.

Luo C, Hu H, Kvecher L, Kovatich AJ, Hooke JA, Chen Y, Mural RJ, Shriver CD. A comparative study of the triple-negative and basal-like breast cancer subtypes. 1st Annual TCGA Scientific Symposium, 17-18 Nov 2011, National Harbor, MD.

Bekhash A, Hooke JA, Chen Y, Kovatich AJ, Kvecher L, Mural RJ, Shriver CD, Hu H. Fibroadenomatoid changes have a higher occurrence rate in middle-aged benign breast disease patients with the trend retained in cancer patients. San Antonio Breast Center Symposium (SABCS), 6-10 Dec 2011, San Antonio, TX.

Kovatich AJ, Kvecher L, Chen Y, Bekhash A, Hooke JA, Shriver CD, Mural RJ, Hu H. Subtype-specific co-occurrence of atypical hyperplasia and in situ carcinoma with invasive breast cancers. San Antonio Breast Center Symposium (SABCS), 6-10 Dec 2011, San Antonio, TX.

Bekhash A, Kovatich AJ, Chen Y, Hooke JA, Kvecher L, Mural RJ, Shriver CD, Hu H. Fibrocystic changes have different age-dependent patterns in benign, in situ, and invasive breast cancer patients. San Antonio Breast Center Symposium (SABCS), 6-10 Dec 2011, San Antonio, TX.

Ellsworth RE, Valente AL, Kane JK, Shriver CD. The effect of breastfeeding on molecular characteristics of invasive breast cancer. San Antonio Breast Center Symposium (SABCS), 6-10 Dec 2011, San Antonio, TX.

Ellsworth RE, Valente AL, Kane JL, Ellsworth DL, Shriver CD. Gene expression alterations in the lymph node microenvironment in response to successful metastatic colonization. San Antonio Breast Center Symposium (SABCS), 6-10 Dec 2011, San Antonio, TX.

Ellsworth R, Croft D, Ellsworth D, Shriver C. Effect of obesity on gene expression in invasive breast tumors. San Antonio Breast Center Symposium (SABCS), 6-10 Dec 2011, San Antonio, TX.

Ellsworth DL, Croft DT Jr, Field LA, Deyarmin B, Kane J, Ellsworth RE, Hooke JA, Shriver CD. Congruence between patterns of micro RNA expression and Histologic grading of invasive breast carcinomas. San Antonio Breast Center Symposium (SABCS), 6-10 Dec 2011, San Antonio, TX.

Zhou J, Brinckerhoff C, Lubert S, Yang K, Saini J, Hooke JA, Mural R, Shriver CD, Somiari SB. Matrix Metalloproteinase-1 2G Insertion Polymorphism is a Prognostic Marker for Predicting Breast Cancer Severity. *Breast Cancer Research and Treatment, Cancer Investigation* DOI: 10.3109/073579072011.621915

Field LA, Love B, Deyarmin B, Hooke JA, Shriver CD, Ellsworth RE. Identification of differentially expressed genes in breast tumors from African American compared to Caucasian women. *Cancer Journal*, 2012 Mar 1;118(5):1334-44. doi:10.1002/cncr.26405. Epub 2011 Jul 28. DOI: 10.1002/cncr.26405.

Hu H. Long Term Collaboration between Windber Research Institute and IDBS Yields Major Impacts on Breast Cancer Care and Understanding. Press Release for Research on Data Management for Translational Research Published in *The Journal of Biomedical Informatics*, Dec 2011.

Voegtly LM, Campbell JL, Shriver CD, Ellsworth RE. Molecular alterations associated with early and late breast cancer mortality. *Ann Surg Oncol Journal*.

Ellsworth RE, Valente AL, Shriver CD, Bittman B, Ellsworth DL. Impact of Lifestyle Factors on Quality of Life and Prognosis in Breast Cancer Survivors in the United States. *Expert Reviews Pharmacoeconomics and Outcomes Research Journal* – submitted for publication.

Iida J, Lehman JR, Clancy R, Somiar S, Ellsworth RE, Mural RJ, Shriver CD. Role for  $\beta$ -catenin/TCF signaling pathways in regulation of migration and growth of triple negative breast cancer cells. *PLoS ONE Journal, Public Library of Science* – submitted for publication.

## **2012**

Shriver CD. AACI Press Release. American Assoc of Cancer Institutes, related to USMCI position. Feb 2012

Ellsworth RE, van Laar R, Deyarmin B, Shriver CD. Effect of ASCO/CAP guidelines for determining ER status on molecular subtype. SSO, 21-24 Mar 2012, Orlando, FL.

Patney HL, Ellsworth RE, Shriver CD. Molecular drivers of breast tumor differentiation. SSO, 21-24 Mar 2012, Orlando, FL.

Ellsworth RE, Campbell JL, Voegtly L, Shriver CD. Molecular alterations associated with breast cancer mortalit. SSO, 21-24 Mar 2012, Orlando, FL.

Ellsworth RE, Gallagher C, Deyarmin B, Hooke JA, Shriver CD. The Effect of HER2 Expression on Luminal A Breast Tumors. SSO, 21-24 Mar 2012, Orlando, FL.

Greer L, Rosman M, Mylander C, Wareham J, Campbell L, Kovatich A, Hooke J, Liang W, Buras R, Shriver CD, Tafra L. Does Breast Tumor Heterogeneity Necessitate Further

Immunohistochemical Staining on Surgical Specimens? SSO, 21-24 Mar 2012, Orlando, FL.

Ellsworth R. Effect of ASCO/CAP guidelines for determining ER status on molecular subtype. *General Surgery News*, May 2012.

Greer LT, Rosman M, Mylander WC, Kovatich A, Hooke, J, Woodward J, Wareham J, Campbell JL, Liang Wen, Buras R, Shriver CD, Tafra L. Should Immunohistochemical Markers be Performed on the Surgical Specimens Rather than the Core Needle Biopsy in Breast Cancer Patients? ASBS, 2-6 May 2012, Phoenix, AZ.

PNNL, Walter Reed & Windber. PNNL, Walter Reed & Windber to scour triple negative breast cancer for leads on diagnostics and drugs. PNNL to EurekAlert! and Newswise, June 2012.

Hu, Hai. Challenges in data warehousing for translational research. 2012 IDBS Translational Medicine Symposium, Jun 2012, Boston, MA.

Iida J, Dorchak J, Clancy R, Hu H, Mural R, Shriver C. Role for wnt/beta-catenin pathways for facilitating triple-negative breast cancer migration and growth. Gordon Research Conference, 7-13 Jul 2012, Andover, NH.

The Cancer Genome Atlas Network. Comprehensive Molecular Portraits of Human Breast Tumors. *Nature*, 23 Sep 2012;490(7418):61-70. doi:10.1038/nature11412. Epub 2012 23 Sep.

Iida J, Dorchak J, Lehman JR, Clancy R, Luo C, Chen Y, Somiari S, Ellsworth RE, Hu H, Mural R, Shriver CD. FH535 inhibited migration and growth of breast cancer cells. *PLOS ONE*, Sept 2012; 7(9) e44418.

Hu H. A biomedical informatics infrastructure to support translational research. Invited to speak at NCI Center for Biomedical Informatics and Information Technology Speaker Series, 5 Sep 2012.

Zhang Y, Sun W, Gutchell EM, Kvecher L, Kohr J, Bekhash A, Shriver CD, Liebman MN, Mural RJ, Hu H. QAIT: A Quality Assurance Issue Tracking Tool to Facilitate the Improvement of Clinical Data Quality. *Comput Methods Programs Biomed*. 2012 Sep 14. pii: S0169-2607(12)00192-7. doi: 10.1016/j.cmpb.2012.08.010. [Epub ahead of print]

Ellsworth RE, Shriver CD. Demographic differences in African American compared to Caucasian women with luminal A breast cancer. *AACR Science of Cancer Health Disparities*, 27-30 Oct 2012, San Diego, CA.

Mural R. Prestigious Journal to Publish Collaborative Breast Cancer Research. Article for HJF SCOOP Oct 2012

**CBCP Personnel 2012**

<b>Last Name</b>	<b>First Name</b>	<b>Role</b>
Shriver	Craig	Principal Investigator
Basham	Janice	Licensed Practical Nurse
Boone	Jaime	Program Manager/Budget Analyst
Bronfman	Eileen	Administrative Director
Chestang	Allan	Data Manager
Cronin	Kerri	Administrative Assistant
Eckhauser	Peggy Lee	Research Nurse
Ellsworth	Rachael	Director of Translational Genomics
Enwold	Lindsey	Epidemiologist
Campbell	Jamie Leigh	Pathology Assistant
Hilton	Karrie	Research Nurse
Hooke	Jeffrey	Head of Pathology
Kelly	Kay	Research Protocol Coordinator
Means	Marilyn	Lab Tech
Neilson	Deborah	Research Nurse
Patterson	Carol	Medical Assistant
Tracey	Dianne	Administrative Assistant/Office Mngr.
Simmons	Stephanie	Research Nurse
Vilakazi	Patricia	Research Nurse
Williamson	Eric	Clinic Administrator
Yambaka	Baredu	Receptionist/Research Assistant
Zingmark	Rebecca	Histology Tech
Zhu	Kangmin	Epidemiologist
Weiss	Raymond	Physician
Wareham	Janet	Pathologist Assistant
Pangaro	Katherine	Clinical Oncology Research Nurse
Miskovsky	Vicki Jones	Admin Reviewer CCC Protocols

### CBCP Personnel 2011

<b>Last Name</b>	<b>First Name</b>	<b>Role</b>
Shriver	Craig	Principal Investigator
Basham	Janice	Licensed Practical Nurse
Boone	Jaine	Program Manager/Budget Analyst
Bronfman	Bileen	Administrative Director
Chestang	Allan	Data Manager
Cronin	Kerri	Administrative Assistant
Eckhauser	Peggy Lee	Research Nurse
Del	Ismail	Data manager
Ellsworth	Rachael	Director of Translational Genomics
Enwold	Lindsey	Epidemiologist
Campbell	Jamie Leigh	Pathology Assistant
Hilton	Karrie	Research Nurse
Hooke	Jeffrey	Head of Pathology
Kelly	Kay	Research Protocol Coordinator
Means	Marilyn	Lab Tech
Miller	Donald	Data Manager
Patterson	Carol	Medical Assistant
Tracey	Dianne	Administrative Assistant/Office Mngr.
Simmons	Stephanie	Research Nurse
Vilakazi	Patricia	Research Nurse
Williamson	Erio	Clinic Administrator
Yambaka	Baredu	Receptionist/Research Assistant
Zingmark	Rebecca	Histology Tech
Zhu	Kangmin	Epidemiologist
Weiss	Raymond	Physican
Wareham	Janet	Pathologist Assistant
Pangaro	Katherine	Clinical Oncology ResearchNurse

**CBCP PERSONNEL RECEIVING PAY FROM  
THE RESEARCH EFFORT  
January 1, 2010 -- December 31, 2010**

<b>Last Name</b>	<b>First Name</b>	<b>Role</b>
Shriver	Craig	Principal Investigator
Basham	Janice	Licensed Practical Nurse
Boone	Jaime	Admin Mgr, CBCP Physician Staff
Bronfinan	Eileen	Administrative Director
Chestang	Allan	Data Manager
Chou	Carolyn	Budget Analyst
Hamsher	Walter	Budget Analyst
Courville	Faith	Research Nurse
Cronin	Kerri	Administrative Assistant/ Office Manager
Del	Ismail	Data Manager
Bokhauser	Peggy	Research Nurse
Ellsworth	Rachael	Genetist
Enowold	Lindsey	Biostatitician
Fantacone	Jamie Leigh	Pathology Assistant
Gutshell	Veronica	Head Nurse, CBCP/ Nurse Practitioner
Hamsher	Walter	Budget Analyst
Hilton	Karrle	Research Nurse
Hodgson	Carol	Nurse Practitioner
Hooko	Jeffrey	Head of Pathology
Kelley	Kay	Research Protocol Coordinator
Leuth	Sunita	Administrative Assistant/ Office Manager
Means	Marilyn	Lab Technician

Miller	Donald	Data Manager
Patterson	Carol	Medical Assistant
Progar	Tina	Biomedical Informatics Coordinator
Reeco	Helke	Data Manager
Rojas	Winifred	Lab Tech/Phlebotomist
Rosenquist	Monica	Budget Analyst
Smith	Stephante	Research Nurse
Stojadinovic	Alexander	Breast Surgeon
Tracey	Dianne	Administrative Assistant/Office Manager
Vilakasi	Patricia	Research Nurse
Wareham	Janet	Pathologist's Assistant
Williamson	Eric	Clinic Administrator
Weiss	Raymond	Surgeon, Staff Researcher
Yambaka	Baredu	Receptionist
Zhao	Xinyan	Histology Technician
Zhu	Kangmin	Epidemiologist
Zhuo	Jing	Biostatistician
Zingman	Rebecca	Histology Technician

**CBCP PERSONNEL RECEIVING PAY FROM  
THE RESEARCH EFFORT  
October 1, 2008 – September 30, 2009**

<b>Last Name</b>	<b>First Name</b>	<b>Role</b>
Shriver	Craig	Principal Investigator
Basham	Janice	Licensed Practical Nurse
Boone	Jaine	Admin Mgr, CBCP Physician Staff
Bronfman	Eileen	Administrative Director
Campbell	Jamie Leigh	Pathology Assistant
Chestang	Allan	Data Manager
Chou	Carolyn	Budget Analyst
Courville	Faith	Research Nurse
Cronin	Kerri	Administrative Assistant/ Office Manager
Del	Ismail	Data Manager
Enowold	Lindsey	Biostatistician
Gutchell	Veronica	Head Nurse, CBCP/ Nurse Practitioner
Hilton	Karrie	Research Nurse
Hodgson	Carol	Nurse Practitioner
Hooke	Jeffrey	Head of Pathology
Kelley	Kay	Research Protocol Coordinator
Leuth	Sunita	Administrative Assistant/ Office Manager
Meaus	Marilyn	Lead Medical Clerk/Receptionist
Miller	Donald	Data Manager
Park	Kathleen	Clinical Nurse Specialist
Patterson	Carol	Medical Assistant
Progar	Christina	Biomedical Informatics Coordinator
Reece	Heike	Data Manager
Rojas	Winifred	Lab Tech/Phlebotomist
Rosenquist	Monica	Budget Analyst
Stojadinovic	Alexander	Breast Surgeon
Vilakasi	Patricia	Research Nurse
Williamson	Eric	Clinic Administrator
Zhao	Xinyan	Histology Technician
Zhu	Kangmin	Epidemiologist

**CBCP PERSONNEL RECEIVING PAY FROM  
THE RESEARCH EFFORT  
July 1, 2007 – June 30, 2008**

<b>Last Name</b>	<b>First Name</b>	<b>Role</b>
Shriver	Craig	Principal Investigator
Basham	Janice	Licensed Practical Nurse
Boone	Jaime	Admin Mgr, CBCP Physician Staff
Bronfman	Eileen	Administrative Director
Campbell	Jamie Leigh	Pathology Assistant
Chestang	Allan	Data Manager
Courville	Faith	Research Nurse
Cronin	Kerri	Receptionist/Administrative Assistant
Del	Ismail	Data Manager
Enowold	Lindsey	Biostatistician
Gutshell	Veronica	Head Nurse, CBCP/ Nurse Practitioner
Hilton	Karrle	Research Nurse
Hodgson	Carol	Nurse Practitioner
Hooke	Jeffrey	Head of Pathology
Kelley	Kay	Research Protocol Coordinator
Means	Marilyn	Lead Medical Clerk/Receptionist
Miller	Donald	Data Manager
Park	Kathleen	Clinical Nurse Specialist
Patterson	Carol	Medical Assistant
Progar	Christina	Biomedical Informatics Coordinator
Reece	Heike	Data Manager
Rojas	Winifred	Lab Tech/Phlebotomist
Rosenquist	Monica	Budget Analyst
Stojadinovic	Alexander	Breast Surgeon
Vilakasi	Patricia	Research Nurse
Williamson	Eric	Clinic Administrator
Zhao	Xinyan	Histology Technician
Zhu	Kangmin	Epidemiologist

**CBCP PERSONNEL RECEIVING PAY FROM  
THE RESEARCH EFFORT  
July 1, 2006 through June 30, 2007**

Last Name	First Name	Role
Shriver	Craig	Principal Investigator
Awunor	Angela	Research Nurse
Basham	Janice	Licensed Practical Nurse
Bronfman	Bileen	Administrative Director
Chestang	Allan	Data Manager
Cooper	Leslie	Psychologist
Courville	Faith	Research Nurse
Cronin	Kerri	Receptionist
Del	Ismail	Data Manager
Fantacomo	Jamie Leigh	Pathology Assistant
Glasco	Tiffany	Executive Office Coordinator
Gutchell	Veronica	Head Nurse, CBCP/ Nurse Practitioner
Hall	Muhammad	Postdoctoral Fellow Epidemiology
Hansher	Carlyle	Research Assistant
Harris	Katie	Research Assistant
Hilton	Karlo	Research Nurse
Hooke	Jeffrey	Head of Pathology
Jama	Yusuf	Research Assistant
Jones	Anita	Research Nurse
Kolley	Kay	Research Protocol Coordinator
Mnlady	Lori	Staff Research Nurse
Means	Marilyn	Lead Medical Clerk/Receptionist
Miller	Donald	Data Manager
Minochar	Nallini	Research Assistant
Moroni	Maria	Scientist
O'Neill	Stacy	Research Nurse
Pappas	Jennifer	Research Associate
Park	Kathleen	Clinical Nurse Specialist
Patterson	Carol	Medical Assistant
Ponniah	Sathibalan	Scientist-Immunologist
Progar	Christina	Program Administrator

W81XWH-05-2-0053

Principal Investigator: Craig D. Shriver, COL MC

Roece	Holko	Data Manager
Rojas	Whitfred	Lab Tech/Phlebotomist
Rosquist	Monica	Budget Analyst
Russo	Jamie	Staff Research Nurse
Sessoms	Lisa	Protocol Coordinator
Smith	Anna	Research Assistant
Smith	Rosson	Research Nurse
Stojadinovic	Alexander	Breast Surgeon
Storrer	Catherine	Senior Research Associate
Vilakasi	Patricia	Research Nurse
Yardlav	Margaret	Nurse Practitioner
Williamson	Eric	Clinic Administrator
Zhao	Xinyan	Histology Technician
Zhu	Kangmin	Epidemiologist

**CBCP PERSONNEL RECEIVING PAY FROM  
THE RESEARCH EFFORT IN 2005**

<b>Last Name</b>	<b>First Name</b>	<b>Role</b>
Shriver	Craig	Principal Investigator
Allen	Peter	Breast Surgeon
Baker	Benjamin	Lab Assistant
Basham	Janice	Licensed Practical Nurse
Bauchiero	Samantha	Research Assistant
Bronfman	Bileen	Protocol Coordinator
Chestang	Allan	Data Manager
Cooper	Leslie	Psychologist
Del	Ismail	Data Manager
Fantacone	Jamie Leigh	Pathology Assistant
Glasco	Tiffani	Administrative Clerk
Gutchell	Veronica	Head Nurse, CBCP/ Nurse Pract.
Hansher	Carlyle	Research Assistant
Hapner	Arthur	Executive Director & COO
Harris	Katie	Research Assistant
Hearrell	Elizabeth	Research Nurse
Hooke	Jeffrey	Head of Pathology
Jama	Yusuf	Research Assistant
Jones	Anita	Research Nurse
Lim	Sophie	Histology Technician
Lisenby	Susan	Research Nurse
Malady	Lori	Staff Research Nurse
Means	Marilyn	Lead Medical Clerk/Receptionist
Miller	Donald	Data Manager
Minohar	Nallini	Research Assistant
Moroni	Maria	Scientist
O'Neill	Stacy	Research Nurse
Overby	Julie	Receptionist
Papay	Diane	Staff Research Nurse
Pappas	Jennifer	Research Associate
Park	Kathleen	Clinical Nurse Specialist
Patterson	Carol	Medical Assistant
Peoples	George	Breast Surgeon
Ponniah	Sathibalan	Scientist-Immunologist
Progar	Christina	Program Administrator

Reece	Heike	Data Manager
Reynolds	Paula	Research Nurse
Rojas	Winifred	Lab Tech/Phlebotomist
Rosenquist	Monica	Budget Analyst
Russo	Jamie	Staff Research Nurse
Salpeas	Veronica	Staff Research Nurse
Sessoms	Lisa	Administrative Assistant
Smith	Anna	Research Assistant
Stojadinovic	Alexander	Breast Surgeon
Storrer	Catherine	Senior Research Associate
Thomas	Lisa	Health Tech / Procedure Room
Williamson	Eric	Clinic Administrator

WRI Personnel - 2012

Full Name	Department Name	Title	FTE
Chen, Yaqin	BCTR	Statistical Analyst	
Clancy, Rebecca	BCTR	Research Associate I	
Croft, Daniel T.	CADRE	Research Associate III	
Cruse, Steven M.	Administration	Housekeeper	
Deyarmin, Brenda	BCTR	Research Associate III	
Deyarmin, Dennis R.	Administration	Housekeeper	
Ditton, Dana L.	CADRE	Research Assistant	
Dorchak, Jesse A.	BCTR	Research Associate I	
Ellsworth, Darrell	CADRE	Senior Director	
Elston, Edward J.	Administration	It Manager	
Ferrau, Annalisa	halt	Administrative Assista	
Full Name	Department Name	Title	FTE
Furmanchik, Lydia	WRI Administration	Finance Assistant	
Greenawalt, Amber M.	BCTR	Research Associate I	
Holliday, Charvonne	Bullying Institute	Public Health Associat	
Hu, Hai	BCTR	Deputy Chief Scientific	
Ilda, George	BCTR	Director	
Kane, Jennifer	BCTR	Research Associate II	
Kohr, Joni	BCTR	Administrative Assista	
Kurtz, Thomas M.	WRI Administration	President	
Kvecher, Leonid	BCTR	Data Manager	
Larson, Caroline M.	BCTR	Resource Manager	
Lubert, Susan M.	BCTR	Lab Manager	
Mamula, Kimberly A.	CADRE	Sr. Statistical Analyst	
Masiello, Matthew	Dr. Masiello	Chief Medical Officer	
Melley, Jennifer	BCTR	Research Associate II	
Meyer, Jeffrey B.	BCTR	Research Associate II	
Mural, Richard J.	WRI Administration	Chief Scientific Officer	
O'Donnell, Amy	BCTR	Research Associate II	
Patney, Heather	CADRE	Research Associate II	
Penatzer, Cayla E.	CBCP	Intern/student/paid	
Rigby, Sean	BCTR	Research Assistant	
Rummel, Seth	BCTR	Research Associate II	
Seltz, Brianna J.	CADRE	Intern/student/paid	
Shaffer, Apryl	BCTR	Data Entry Technician	
Shawley, Cynthia M.	WRI Administration	Executive Assistant	
Slavik, Julianna E.	CADRE	Research Associate II	
Somlari, Stella	BCTR	Senior Director	
Sturtz, Lori	BCTR	Scientist	
Trostle, Lynn	WRI Administration	Grant Program Coordi	
Valente, Allyson	BCTR	Research Associate III	
Voegtly, Eric	Administration	Manager	
Wang, Hua	BCTR	It Developer	
Welse, Jonathan P.	Administration	It Support Specialist	
Wu, Weiqiang	BCTR	Research Associate III	

## WRI Personnel - 2011

Title	Last Name	First Name	FTE
Administrative Assistant Count Administrative Assistant	Ferrau	Annalisa	
Administrative Assistant II Count Administrative Assistant II	Kohr	Joni	
Chief Medical Officer Count Chief Medical Officer	Masiello	Matthew	
Chief Scientific Officer Count Chief Scientific Officer	Mural	Richard	
Data Entry Technician Count Data Entry Technician	Shaffer	Apryl	
Data Manager Count Data Manager	Kvecher	Leonid	
Deputy Chief Scientific Office Count Deputy Chief Scientific Office	Hu	Hai	
Director Count Director	Iida	George	
Executive Assistant Count Executive Assistant	Shawley	Cynthia	
Finance Assistant Count Finance Assistant	Furmanchik	Lydia	
Grant Program Coordinator Count Grant Program Coordinator	Trostle	Lynn	
Housekeeper Housekeeper Count Housekeeper	Cruse Deyarmin	Steven Dennis	
Intern/student/paid Intern/student/paid Count Intern/student/paid	Penatzer Seitz	Cayla Brienne	
It Developer Count It Developer	Wang	Hua	
It Manager Count It Manager	Elston	Edward	

WRI Personnel - 2011

It Support Specialist Count It Support Specialist	Weise	Jonathan
Lab Manager Count Lab Manager	Lubert	Susan
Manager Count Manager	Voegtly	Eric
President Count President	Kurtz	Thomas
Public Health Associate Count Public Health Associate	Holliday	Charvonne
Research Assistant Research Assistant Count Research Assistant	Rigby Ditton	Sean Dana
Research Associate I Research Associate I Research Associate I Count Research Associate I	Dorchak Greenawalt Clancy	Jesse Amber Rebecca
Research Associate II Research Associate II Research Associate II Research Associate II Research Associate II Research Associate II Research Associate II Count Research Associate II	Melley Kane Rummel Patney O'Donnell Meyer Slavik	Jennifer Jennifer Seth Heather Amy Jeffrey Julianna
Research Associate III Research Associate III Research Associate III Research Associate III Count Research Associate III	Valente Croft Wu Deyarmin	Allyson Daniel Weiqiang Brenda
Resource Manager Count Resource Manager	Larson	Caroline
Scientist Count Scientist	Sturtz	Lori
Senior Director	Somiari	Stella

WRI Personnel - 2011

Senior Director  
Count Senior Director

Ellsworth

Darrell

Sr. Statistical Analyst  
Count Sr. Statistical Analyst

Mamula

Kimberly

Statistical Analyst  
Count Statistical Analyst

Chen

Yaqin

Grand Count

44

WRI Personnel - 2010

Title	Last Name	First Name	FTE
Administrative Assistant Count Administrative Assistant	Ferrau	Annalisa	
Administrative Assistant II Count Administrative Assistant II	Kohr	Joni	
Chief Medical Officer Count Chief Medical Officer	Masiello	Matthew	
Chief Scientific Officer Count Chief Scientific Officer	Mural	Richard	
Data Entry Technician Count Data Entry Technician	Raymore	Susan	
Data Manager Count Data Manager	Kvecher	Leonid	
Deputy Chief Scientific Office Count Deputy Chief Scientific Office	Hu	Hai	
Director Director Count Director	Slavich Ilda	Lacey George	

WRI Personnel - 2010

Executive Assistant  
Count Executive Assistant

Shawley

Cynthia

Finance Assistant  
Count Finance Assistant

Furmanchik

Lydia

Grant Program Coordinator  
Count Grant Program Coordinator

Trostle

Lynn

Housekeeper  
Count Housekeeper

Lovette

Merry

Intern/student/paid  
Intern/student/paid  
Count Intern/student/paid

Hicks

Marisa

Fee

Brian

It Developer  
Count It Developer

Wang

Hua

It Manager  
Count It Manager

Elston

Edward

It Support Specialist

Weise

Jonathan

WRI Personnel - 2010

Count It Support Specialist

Lab Manager  
Count Lab Manager

Lubert

Susan

Manager  
Count Manager

Voegtly

Eric

Post Doctoral Fellow  
Post Doctoral Fellow  
Count Post Doctoral Fellow

Voegtly  
Zhou

Laura  
Jing

President  
Count President

Kurtz

Thomas

Public Health Associate  
Count Public Health Associate

Holliday

Charvonne

Receptionist  
Count Receptionist

Pavlosky

Cathy

Research Assistant  
Research Assistant  
Count Research Assistant

Leslie  
Rigby

Matthew  
Sean

WRI Personnel - 2010

Research Associate I	Aberts	Gerlanne
Research Associate I	Lehman	John
Research Associate I	Greenawalt	Amber
Research Associate I	Clancy	Rebecca
Research Associate I	Dorchak	Jesse
Count Research Associate I		

Research Associate II	Decewicz	Alisha
Research Associate II	Melley	Jennifer
Research Associate II	Kane	Jennifer
Research Associate II	Patney	Heather
Research Associate II	Meyer	Jeffrey
Research Associate II	O'Donnell	Amy
Research Associate II	Rummel	Seth
Count Research Associate II		

Research Associate III	Wu	Weiqlang
Research Associate III	Deyarmin	Brenda
Research Associate III	Croft	Daniel
Research Associate III	Valente	Allyson
Research Associate III	Kirchner	David
Research Associate III	Katenhusen	Richard
Research Associate III	Jordan	Rick
Count Research Associate III		

Research Physician	Decewicz	David
Research Physician	Bekhash	Anthony
Count Research Physician		

Resource Manager	Larson	Caroline
Count Resource Manager		

WRI Personnel - 2010

Scientist  
Scientist  
Scientist  
Count Scientist

Li  
Rapuri  
Sturtz

Xing  
Prema  
Lori

Senior Director  
Senior Director  
Senior Director  
Count Senior Director

Kolli  
Ellsworth  
Somlari

Kumar  
Darrell  
Stella

Grand Count

WRI Personnel - 2009

Title	Last Name	First Name	FTE
Administrative Assistant II Count Administrative Assistant II	Kohr	Joni	
Administrative Fellow Count Administrative Fellow	Xie	Yuanli	
Chief Administrative Officer Count Chief Administrative Officer	Bates	James	
Chief Medical Officer Count Chief Medical Officer	Masiello	Matthew	
Chief Scientific Officer Count Chief Scientific Officer	Mural	Richard	
Data Entry Technician Count Data Entry Technician	Raymore	Susan	
Data Manager Count Data Manager	Kvecher	Leonid	
Deputy Chief Scientific Office Count Deputy Chief Scientific Office	Hu	Hai	

WRI Personnel - 2009

Director	Petrick	Lisa
Director	Slavich	Lacey
Director	Iida	George
Director	Brownlee	Robert
Count Director		

Executive Assistant	Shawley	Cynthia
Count Executive Assistant		

Finance Assistant	Furmanchik	Lydia
Count Finance Assistant		

Grant Program Coordinator	Trostle	Lynn
Count Grant Program Coordinator		

Housekeeper	Lovette	Merry
Count Housekeeper		

Intern/student/paid	Fee	Brian
Intern/student/paid	Piskuric	Michael
Count Intern/student/paid		

It Developer	Napotnik	David
Count It Developer		

WRI Personnel - 2009

It Manager  
Count It Manager

Elston Edward

Lab Manager  
Count Lab Manager

Lubert Susan

President  
Count President

Kurtz Thomas

Public Health Associate  
Count Public Health Associate

Holliday Charvonne

Receptionist  
Count Receptionist

Pavlosky Cathy

Research Assistant  
Research Assistant  
Count Research Assistant

Leslie Matthew  
Rigby Sean

Research Associate I  
Count Research Associate I

Greenawalt Amber  
Lehman John  
Swearinger Chad  
Clancy Rebecca  
Aberts Gerianne  
Nesbella Matthew

WRI Personnel - 2009

Research Associate II	Decewicz	Alisha
Research Associate II	Melley	Jennifer
Research Associate II	Kane	Jennifer
Research Associate II	O'Donnell	Amy
Research Associate II	Patney	Heather
Research Associate II	Rummel	Seth
Research Associate II	Seth	Bernard
Research Associate II	Weaver	Luke
Count Research Associate II		

Research Associate III	Deyarmin	Brenda
Research Associate III	Valente	Allyson
Research Associate III	Jordan	Rick
Research Associate III	Katenhusen	Richard
Research Associate III	Kirchner	David
Research Associate III	Wu	Weiqliang
Research Associate III	Saini	Jasmine
Research Associate III	Croft	Daniel
Count Research Associate III		

Research Physician	Bekhash	Anthony
Research Physician	Decewicz	David
Count Research Physician		

Resource Manager	Bombatch	James
Resource Manager	Larson	Caroline
Count Resource Manager		

Scientist	Rapuri	Prema
Scientist	Sturtz	Lori
Scientist	Li	Xing
Count Scientist		

WRI Personnel - 2009

Senior Director  
Senior Director  
Senior Director  
Count Senior Director

Somiari	Stella
Kolli	Kumar
Ellsworth	Darrell

Grand Count

58

WRI Personnel - 2008

Title	Last Name	First Name	FTE
Administrative Assistant II Count Administrative Assistant II	Kohr	Joni	
Administrative Fellow Count Administrative Fellow	Xie	Yuanli	
Chief Administrative Officer Count Chief Administrative Officer	Bates	James	
Chief Medical Officer Count Chief Medical Officer	Masiello	Matthew	
Chief Scientific Officer Count Chief Scientific Officer	Mural	Richard	
Courier Count Courier	Kush	Anthony	
Data Manager Count Data Manager	Kvecher	Leonid	
Deputy Chief Scientific Office Count Deputy Chief Scientific Office	Hu	Hai	

WRI Personnel - 2008

Director	Brownlee	Robert
Director	Petrick	Lisa
Director	Iida	George
Director	Slavich	Lacey
Count Director		

Executive Assistant	Shawley	Cynthia
Count Executive Assistant		

Finance Assistant	Furmanchik	Lydia
Count Finance Assistant		

Grant Program Coordinator	Trostle	Lynn
Count Grant Program Coordinator		

It Manager	Elston	Edward
Count It Manager		

Lab Manager	Lubert	Susan
Count Lab Manager		

Public Health Associate	Holliday	Charvonne
Count Public Health Associate		

WRI Personnel - 2008

Receptionist Count Receptionist	Pavlosky	Cathy
Research Assistant Research Assistant Count Research Assistant	Leslie Rigby	Matthew Sean
Research Associate I Research Associate I Research Associate I Count Research Associate I	Lehman Greenawalt Nesbella	John Amber Matthew
Research Associate II Research Associate II Research Associate II Research Associate II Research Associate II Research Associate II Research Associate II Count Research Associate II	Decewicz Patney Kane O'Donnell Seth Rummel Weaver	Alisha Heather Jennifer Amy Bernard Seth Luke
Research Associate III Research Associate III Research Associate III Research Associate III Research Associate III Research Associate III Research Associate III Count Research Associate III	Deyarmin Jordan Katenhusen Valente Croft Kirchner Saini Wu	Brenda Rick Richard Allyson Daniel David Jasmine Weiqlang
Research Physiclan	Bekhash	Anthony

WRI Personnel - 2008

Research Physician  
Research Physician  
Count Research Physician

Decewicz      David  
Mihaly        Csaba

Resource Manager  
Resource Manager  
Count Resource Manager

Bombatch      James  
Larson         Caroline

Scientist  
Scientist  
Count Scientist

Sturtz         Lori  
Rapuri         Prema

Senior Director  
Senior Director  
Senior Director  
Count Senior Director

Kolli           Kumar  
Somlari        Stella  
Ellsworth      Darrell

Grand Count

WRI Personnel - 2007

Title	Last Name	First Name	FTE
Administrative Assistant II Count Administrative Assistant II	Kohr	Joni	
Administrative Fellow Count Administrative Fellow	Xie	Yuanli	
Chief Administrative Officer Count Chief Administrative Officer	Bates	James	
Chief Scientific Officer Count Chief Scientific Officer	Mural	Richard	
Courier Count Courier	Kush	Anthony	
Data Manager Count Data Manager	Kvecher	Leonid	
Deputy Chief Scientific Office Count Deputy Chief Scientific Office	Hu	Hai	
Director	Brownlee	Robert	
Director	Petrick	Lisa	
Director	Yang	Yaw-Ching	
Count Director			

WRI Personnel - 2007

Executive Assistant  
Count Executive Assistant

Shawley      Cynthia

Finance Assistant  
Count Finance Assistant

Furmanchik      Lydia

Grant Program Coordinator  
Count Grant Program Coordinator

Trostle      Lynn

It Manager  
Count It Manager

Elston      Edward

Lab Manager  
Count Lab Manager

Lubert      Susan

President  
Count President

Jacobs      F. Nicholas

Receptionist  
Count Receptionist

Pavlosky      Cathy

Research Assistant

Rigby      Sean

WRI Personnel - 2007

Count Research Assistant

Research Associate I	Greenawalt	Amber
Count Research Associate I		

Research Associate II	Seth	Bernard
Research Associate II	Weaver	Luke
Research Associate II	O'Donnell	Amy
Research Associate II	Weyandt	Jamie
Research Associate II	Decewicz	Allisha
Research Associate II	Kane	Jennifer
Research Associate II	Patney	Heather
Count Research Associate II		

Research Associate III	Croft	Daniel
Research Associate III	Kirchner	David
Research Associate III	Deyarmin	Brenda
Research Associate III	Jordan	Rick
Research Associate III	Valente	Allyson
Research Associate III	Katenhusen	Richard
Count Research Associate III		

Research Physician	Sun	Weihong
Research Physician	Mihaly	Csaba
Research Physician	Bekhash	Anthony
Count Research Physician		

Resource Manager	Bombatch	James
Resource Manager	Larson	Caroline
Count Resource Manager		

WRI Personnel - 2007

Scientist  
Scientist  
Count Scientist

Sturtz  
Maskery

Lori  
Susan

Senior Director  
Senior Director  
Senior Director  
Count Senior Director

Somiari  
Ellsworth  
Koll

Stella  
Darrell  
Kumar

Senior Scientist  
Count Senior Scientist

Zhang

Younghong

Sr Research Fellow  
Count Sr Research Fellow

Liebman

Michael

Grand Count

WRI Personnel - 2006

Title	Last Name	First Name	FTE
Administrative Assistant II Count Administrative Assistant II	Kohr	Joni	
Administrative Fellow Count Administrative Fellow	Xie	Yuanli	
Chief Administrative Officer Count Chief Administrative Officer	Bates	James	
Chief Medical Officer Count Chief Medical Officer	Dunn	Michael	
Chief Scientific Officer Count Chief Scientific Officer	Mural	Richard	
Coord Biomedical Informatics Count Coord Biomedical Informatics	Gutchell	Emily	
Courier Count Courier	Kush	Anthony	
Data Entry Technician Count Data Entry Technician	Thomas	Davene	

WRI Personnel - 2006

Data Manager  
Count Data Manager

Kvecher      Leonld

Deputy Chief Scientific Office  
Count Deputy Chief Scientific Office

Hu              Hai

Director  
Director  
Director  
Director  
Count Director

Yang            Yaw-Ching  
Petrick        Lisa  
Brownlee      Robert  
Ellsworth      Rachel

Executive Assistant  
Count Executive Assistant

Shawley        Cynthia

Finance Assistant  
Count Finance Assistant

Furmanchik    Lydia

Grant Program Coordinator  
Count Grant Program Coordinator

Trostle        Lynn

Intern/student/paid  
Count Intern/student/paid

Silberman      Jordan

WRI Personnel - 2006

It Manager Count It Manager	Elston	Edward
Lab Manager Count Lab Manager	Lubert	Susan
President Count President	Jacobs	F. Nicholas
Receptionist Count Receptionist	Pavlosky	Cathy
Research Assistant Research Assistant Count Research Assistant	Rigby Brink	Sean Melissa
Research Associate I Research Associate I Count Research Associate I	Grof-Tisza Greenawalt	Patrick Amber
Research Associate II Research Associate II Research Associate II Research Associate II Research Associate II Count Research Associate II	Decewicz Weyandt Kane Seth Patney	Alisha Jamie Jennifer Bernard Heather

WRI Personnel - 2006

Research Associate III  
Count Research Associate III

Croft Daniel  
Deyarmin Brenda  
Jordan Rick  
Katenhusen Richard  
Kirchner David

Research Physician  
Research Physician  
Count Research Physician

Bekhash Anthony  
Sun Weihong

Resource Manager  
Resource Manager  
Count Resource Manager

Bombatch James  
Larson Caroline

Scientist  
Scientist  
Count Scientist

Maskery Susan  
Sturtz Lori

Senior Director  
Senior Director  
Senior Director  
Count Senior Director

Kolli Kumar  
Somari Stella  
Ellsworth Darrell

Senior Scientist  
Senior Scientist  
Count Senior Scientist

Zhang Younghong  
Maity Tapan

WRI Personnel - 2006

Sr Research Fellow  
Count Sr Research Fellow

Liebman Michael

Grand Count

48

WRI Personnel - 2005

Title	Last Name	First Name
Administrative Assistant II	Neiderer	Cecilia
Administrative Assistant II	Kohr	Jonl
Count Administrative Assistant II		
Chief Administrative Officer	Bates	James
Count Chief Administrative Officer		
Chief Medical Officer	Dunn	Michael
Count Chief Medical Officer		
Chief Scientific Officer	Mural	Richard
Count Chief Scientific Officer		
Data Entry Technician	Thomas	Davene
Count Data Entry Technician		
Data Manager	Kvecher	Leonid
Count Data Manager		
Deputy Chief Scientific Office	Hu	Hal
Count Deputy Chief Scientific Office		
Director	Petrick	Lisa
Director	Yang	Yaw-Ching
Director	Brownlee	Robert

WRI Personnel - 2005

Director Count Director	Ellsworth	Rachel
Executive Assistant Count Executive Assistant	Shawley	Cynthia
Finance Assistant Count Finance Assistant	Furmanchik	Lydia
Grant Program Coordinator Count Grant Program Coordinator	Trostle	Lynn
Intern/student/paid Intern/student/paid Intern/student/paid Count Intern/student/paid	Silberman Chung Roney	Jordan Sey-Kyo Elizabeth
It Manager Count It Manager	Elston	Edward
Lab Manager Count Lab Manager	Lubert	Susan
Pre-Doctoral Fellow Count Pre-Doctoral Fellow	Leech	Bernard

WRI Personnel - 2005

President Count President	Jacobs	F. Nicholas
Receptionist Count Receptionist	Pavlosky	Cathy
Research Assistant Research Assistant Count Research Assistant	Rigby Brink	Sean Melissa
Research Associate I Research Associate I Research Associate I Count Research Associate I	Greenawalt Grof-Tisza Meyers	Amber Patrick Sarah
Research Associate II Research Associate II Research Associate II Research Associate II Count Research Associate II	Patney Hadix Kane Decewicz	Heather Jennifer Jennifer Alisha
Research Associate III Research Associate III Research Associate III Research Associate III Research Associate III Research Associate III Count Research Associate III	Kirchner Deyarmin Jordan Croft Katenhusen Zhu	David Brenda Rick Daniel Richard Andy

WRI Personnel - 2005

Research Physician Count Research Physician	Sun	Welhong
--	-----	---------

Resource Manager Resource Manager Count Resource Manager	Bombatch Larson	James Carollne
--	--------------------	-------------------

Scientist Scientist Scientist Scientist Count Scientist	Maskery Sturtz Wingerd Ru	Susan Lori Mark Qin
---	------------------------------------	------------------------------

Senior Director Senior Director Senior Director Senior Director Count Senior Director	Kolli Somlari Brzeski Ellsworth	Kumar Stella Henry Darrell
---	--	-------------------------------------

Senior Scientist Senior Scientist Senior Scientist Senior Scientist Count Senior Scientist	Guo Zhang Yang Maity	Xiang Younghong Song Tapan
--	-------------------------------	-------------------------------------

Sr Research Fellow Count Sr Research Fellow	Liebman	Michael
--	---------	---------

Grand Count

54

MCHL-SG

DATE: 3 May 2012

[Amendments: 6 Nov 01; 10 Jul 02; 25 Jun 03; 19 Feb 04; 30 Jul 04; 21 Mar 05; 6 Jun 05; 24 Aug 05; 15 May 06; 21 Mar 07; 20 Oct 2010]

MEMORANDUM FOR CHIEF, DEPARTMENT OF RESEARCH PROGRAMS,  
WALTER REED NATIONAL MILITARY MEDICAL CENTER

SUBJECT: Application and Request for Approval of Clinical Investigation Study Proposal

**1. PROTOCOL TITLE: Creation of a Blood Library for the Analysis of Blood for Molecular Changes Associated with Breast Disease and Breast Cancer Development**

**2. PRINCIPAL INVESTIGATOR:** COL Craig D. Shriver, MC, USA  
Chief, General Surgery Service  
Program Director, NCC General  
Surgery  
PI, Clinical Breast Care Project  
Professor of Surgery, USUHS  
Interim Director, Cancer Center, WRNMMC/JTF

Phone: 301-400-2889  
Email: craig.shriver@med.navy.mil  
Research Course Training Date: Sep 2010

**CO-PRINCIPAL INVESTIGATOR:** CAPT Ralph C. Jones, MC  
General Surgeon  
Chair of Cancer Committee  
Phone: 301-295-6335  
Email: ralph.jones@med.navy.mil  
Research Course Training Date: Jan 2010

**3. ASSOCIATE INVESTIGATORS:**  
Jeffrey A. Hooke, MD  
Head of Pathology, CBCP  
Phone 301-400-1483  
Email: jeffrey.hooke@med.navy.mil  
Research Course Training Date: Mar 2010

CAPT David M. Larson, MC  
Laboratory Director  
Phone: 301-319-4350  
Email: david.larson@med.navy.mil  
Research Course Training Date: Mar 2010  
Claudia E. Galbo, MD

Head, Clinical Mammography Services  
Women's Imaging Center  
Phone: 301-319-8548  
Email: claudia.galbo@med.navy.mil  
Research Course Training Date: Jan 2010

#### **4. COLLABORATING PERSONNEL:**

Richard J. Mural, Ph.D.  
Chief Scientific Officer  
Clinical Breast Care Project  
Windber Research Institute (WRI)  
620 Seventh Street  
Windber, PA 15963-1300  
Phone: 814-361-6950  
Email: r.mural@wriwindber.org  
Research Course Training Date: Dec 2010

Stella Somiari, PhD  
Senior Director Tissue Bank  
Windber Research Institute  
620 Seventh Street  
Windber, PA 15963-1300  
Phone: 814-467-9844  
Email: s.somiari@wriwindber.org  
Research Course Training Date: Mar 2011

Responsibility of collaborating personnel: Carrying out successful receipt of anonymized donor specimens transferred to the off-site functional genomics, proteomics and tissue and blood repository facility, and to manage and oversee the scientific experiments and handling of results of all studies involving the said specimens at said facility.

Diana Craig, MD  
Breast Surgeon, Windber Medical Center  
600 Somerset Avenue  
Windber, PA 15963  
Phone 814-467-0044  
Email: dcraig@windbercare.org  
Research Course Training Date: Jun 2009

Responsibility: Acting as on-site PI at the collaborating center in Windber, PA, for the Tissue and Blood repository and high-throughput research initiatives as described in this protocol.

**5. MEDICAL MONITOR:** Minimal risk study, medical monitor not required.

**6. OBJECTIVES:** Although there have been remarkable improvements in breast cancer diagnosis and management, most of the complex molecular mechanisms associated with the onset, progression and/or severity of breast cancer are still not well understood. The Clinical Breast Care Project (CBCP) is a congressionally mandated and funded military-civilian collaboration between Windber Medical Center (WRI) Windber, PA, Walter Reed National Military Medical Center (WRNMMC) Bethesda, MD, and the Telemedicine and Advanced Technology Research Center (TATRC) Fort Detrick, MD. As part of the CBCP we propose to carry out molecular, biochemical and histologic analysis of breast tissue and/or blood to provide more insights on the molecular mechanisms that may be relevant in breast cancer development and breast diseases. To achieve this aim, a large supply and a wide variety of good quality tissue sample and blood specimens are needed. Unfortunately, good quality donor breast tissue samples and blood specimens are scarce and, when available, are often not backed by a comprehensive medical history and/or not a good representation of the target population or study area. The non-availability of a steady and consistent supply of good quality tissue samples and blood specimens limits the systematic analysis of tissues and negatively impacts the generation of biologically useful information at the laboratories and by extension clinical practice. The objectives of this protocol are therefore:

1. Acquisition and banking of blood and serum from informed and consenting donors.
2. To characterize gene and protein expression profiles and single nucleotide polymorphisms associated with breast disease and breast cancer development.
3. Identify factors within patient serum and/or blood-derived cellular components that correlate with patient risk factors or clinical status as defined in the corresponding clinical patient database.

**7. MEDICAL APPLICATION:** One of the major challenges facing researchers and clinicians today is to understand the mechanisms associated with the evolution of benign breast disease and/or transition of breast disorders to breast cancer. The creation of a comprehensive tissue/blood bank is essential to the application of modern molecular and genetic analysis of breast diseases. Amongst other goals, this entire (overall CBCP) project will:

A. Establish a repository of high quality breast tissue and related (lymph nodal, blood) specimens for research on breast cancer and associated breast diseases, and  
Permit the establishment of a single relational database with accurate and comprehensive biologically and clinically relevant information on breast diseases.

Since the standard of care for treating breast diseases and breast cancer is based on a multidisciplinary model that integrates prevention, screening, diagnosis, treatment and management, this CBCP project will provide the necessary framework for such an integrated approach, which will positively impact the future management of breast cancer.

This proposed study will identify differences in gene markers and protein expression patterns observed amongst a variety of patients attending the CBCP breast clinic at WRNMMC, and WMC,. The patients will be grouped either as (1) high-risk for breast cancer development; (2) those diagnosed with breast cancer; or (3) other breast diseases (e.g. benign, or non/pre-invasive). Information obtained from the blood analysis will be linked with other clinical data to

establish possible relationships between gene expression patterns and specific genetic alteration with disease onset and progression. The medical implication of this study is the identification of molecular markers that may enable early breast cancer detection, monitoring and prognosis/prediction through a less subjective and less intrusive procedure.

**8. BACKGROUND AND SIGNIFICANCE:** Breast cancer is one of the most common cancers among women. Like many other cancers, it is complex and involves multiple biological changes, which could be somatic or heritable. These changes characterize the transition of a cell from normal to neoplastic and its final progression into an invasive cancer<sup>1,2</sup>. Broad generalizations can be made about risk, natural history and clinical pattern, but it is not at present possible to accurately predict the future of an individual classified as being “at high-risk” for the development of breast cancer. This label of “high-risk” is applied generally to women who, upon undergoing screening via the computerized Gail Model (which is a validated NCI tool using a series of patient questions / answers resulting in an individualized computerized risk assessment), have any risk outcome that results in a >1.67% risk of breast cancer development over the next 5 years. Current diagnostic procedures are invasive involving biopsies with a high rate of negative results. To make significant future improvements on early breast cancer diagnosis and management, a more effective and minimally invasive method will be required that will be sensitive, predictive and allow accurate diagnosis at the earliest stage of disease onset. Neoplastic cells can be detected in bodily fluids that drain or bathe affected organs<sup>3</sup>. The use of peripheral blood as a sample source for clinical research purposes is very attractive since it requires a less invasive procedure for sample collection (i.e. a simple blood stick). Clearly the use of circulating soluble markers that can be easily quantified through sensitive procedures such as the polymerase chain reaction (PCR) would be considerably more accurate than subjective approaches such as immunohistochemical staining which also require more invasive procedures for sample (tissue) collection. The spectrum of genetic and physiological events associated with primary tumors can be potential targets for molecular detection in the blood of patients and circulating soluble molecules in blood have been used as biological markers in cancer providing diagnostic and prognostic value<sup>4,5</sup>. High levels of circulating DNA have been observed in patients with metastatic disease and even patients with early stage cancer may harbor these free circulating DNA<sup>3</sup>.

In order to diagnose breast cancer at the earliest possible stage and provide appropriate therapy after such detection, we need to understand the global profile of the genes and proteins in the blood. So far approximately 68 genes have been associated with breast cancer. They fall into several groups including DNA repair, cell cycle regulation, tumor suppression, internal/external signaling and transcription/growth factor regulation. However their exact expression pattern, which may impact upon disease onset and progression, are not yet fully understood. The techniques of Microarray Analysis and Single nucleotide polymorphism (SNPs) screening are high throughput gene expression analysis strategies that have been used successfully to characterize gene expression in many medical conditions. Microarray analysis allows the identification of gene expression differences through the simultaneous monitoring of thousands of genes in parallel using complementary DNA (cDNA) which are preloaded onto a small (size of a dime) chip. On the other hand, using the SNPs technique, single base variations in DNA can be identified. These new high-throughput technologies provide the tools for the identification of potential markers for disease detection and monitoring, and insights can be gained into individual

differences observed in relation to disease susceptibility, onset of disease and/or response to therapy.

Significant improvement in breast cancer mortality rates will be achieved when we can either prevent the disease, or correctly diagnose it at an early stage and provide treatment that will halt metastasis. Until we find a means of preventing cancer, we can reduce mortality rates through early and accurate diagnosis followed by proper management. Developing a diagnosis/prognostic strategy that will be sensitive and which will require only samples collected through a minimally invasive procedure such as a blood stick, would significantly impact breast cancer management.

The current standard of care for treating breast diseases and breast cancer is based on a multidisciplinary model that integrates prevention, screening, diagnosis, treatment, and continuing care. The CBCP is modeled after this multidisciplinary approach. However, the CBCP is unique in its field. It incorporates advances in breast cancer risk reduction, informatics, tissue banking, and research. The present protocol represents one aspect of this integrated approach. It will interdigitate and synergize with the clinical, informatics, and research components of the CBCP, to achieve the overall objective of improving breast care and reducing the morbidity and mortality associated with breast disease.

In 2002, a panel of microsatellite markers was developed to assess allelic imbalance (AI) at the 26 most commonly altered regions in breast cancer. This panel was used to investigate how tumors develop and progress. For example, this panel was applied to 42 columnar cell lesions (CCL) and 31 atypical ductal hyperplasias (ADH), both non-neoplastic diseases of the breast. Of note, we assessed genetic changes only in pure CCL and ADH without concomitant DCIS or invasive breast cancer (IBC). This approach was quite different than the then current body of literature that had evaluated synchronous non-neoplastic lesions and, based on shared genetic changes between the neoplastic and non-neoplastic diseases, suggested that CCL and ADH were non-obligatory precursors to DCIS/IBC. Our study was one of the first to look at pure CCL and ADH, and unlike the other studies, we found very few genetic changes, suggesting that these lesions are genetically naïve and that earlier models of progression based on synchronous lesions may not accurately reflect tumor progression<sup>6</sup>.

In contrast, the 100 pure DCIS lesions we assayed shared many of the same alterations found in IBC, suggesting that DCIS lesions are genetically advanced<sup>7</sup>. We also found that low-grade DCIS and IBC are genetically distinct from high-grade DCIS and IBC, based largely on high rates of loss of chromosome 16q in low-grade disease and high rates of alterations at 9p21, 11q23, 13q14, 17p13.1 and 17p12<sup>8,9</sup>. These data suggest that high-grade disease does not progress from low-grade disease but rather that low- and high-grade disease are distinct diseases.

CBCP has been investigating molecular changes in the tumor microenvironment. Tumor epithelial and corresponding adjacent stroma <850  $\mu$ m from the tumor were laser microdissected from 30 formalin-fixed, paraffin-embedded tumor specimens and evaluated using the allelic

imbalance panel. Discordant patterns of LOH/AI between epithelial and stromal components were detected such that hierarchical clustering samples from the same patient rarely clustered together<sup>10</sup>. In a second study evaluating LOH/AI data generated from archival breast quadrants from 21 patients who underwent mastectomy, levels of LOH/AI were significantly higher ( $P < 0.05$ ) in the tumor-adjacent (15.4%) compared to distant (3.7%) tissues, suggesting that chromosomal alterations in stroma close to the tumor may promote tumorigenesis<sup>11</sup>. These data lead to the invitation to write review papers for *Lancet Oncology* and *Expert Reviews in Molecular Diagnosis*<sup>12, 13</sup>.

We have also had an interest in health care disparities between Caucasian (CW) and African American women (AAW) with breast cancer. Enrollment of patients within the DOD health-care system affords us the opportunity to evaluate clinical, epidemiological and molecular differences in patients treated within an equal-access health-care system, thus minimizing the impact of disparate access to quality-health care seen in the general population. Recruitment of African Americans into the Clinical Breast Care Project (CBCP) has been effective: in 2007, 25% of the patients enrolled were self-described AAW<sup>14</sup>. No differences in education levels, breast health practices (e.g., self-breast exam, routine screening mammograms) or reproductive histories were seen between AAW and CW with invasive breast cancer, however, AAW were more likely to use oral contraceptives, less likely to use HRT or to breastfeed and were more likely to be obese and have extremely high fat-intake. Breast tumors from AAW were also more likely to be diagnosed at a younger age, and are larger, more frequently ER negative and/or triple negative, and of higher grade. Evaluation of gene expression patterns between AAW and CW matched by age, grade and ER status revealed differentially expression of 23 genes, suggesting that tumors from AAW and CW are molecularly different and these differences may contribute to the less favorable outcome in AAW<sup>15</sup>.

We have also performed a number of studies to determine the molecular mechanisms of metastasis. Using our AI panel, we evaluated patterns of chromosomal changes in primary breast tumors and matched metastatic lymph node tumors and found disparate patterns of alterations between the two tumor types, suggesting that metastatic cells diverge early in tumorigenesis<sup>16</sup>. A follow-up study evaluated genetic changes between primary tumors (n=26) and multiple lymph node tumors (n=146) to determine how metastasis spreads. Divergent patterns of chromosomal alterations were seen between lymph nodes and the primary tumor, but also between metastatic lymph node tumors from the same patients. This genetic heterogeneity may impact response to adjuvant therapy, recurrence, and survival, and thus may be important to improving clinical management of breast cancer patients<sup>17</sup>. Gene expression analysis was also performed on primary breast tumors and matched metastatic lymph node tumors to improve our understanding of the metastatic process. Fifty-one genes were differentially expression ( $P < 1 \times 10^{-5}$ , >2-fold differences). Genes expressed at higher levels in the primary tumors were involved in degradation of the extracellular matrix, enabling cells with metastatic potential to disseminate, while genes expressed at higher levels in metastases were involved in transcription,

signal transduction and immune response, providing cells with proliferation and survival advantages. These data improve our understanding of the biological processes involved in successful metastasis and provide new targets to arrest tumor cell dissemination and metastatic colonization<sup>13</sup>. Evaluation of primary tumors from patients with metastases and those without has also been performed; however, in this study we were unable to identify a molecular signature that could effectively classify the two tumor types<sup>18</sup>. The inability to derive molecular profiles of metastasis in primary tumors may reflect tumor heterogeneity, paucity of cells within the primary tumor with metastatic potential, influence of the microenvironment, or inherited host susceptibility to metastasis<sup>19</sup>.

Biomedical Informatics (BMIX) has been focused on: 1) development and implementation of an infrastructure system that supports the acquisition, storage, and maintenance of the clinical and molecular data generated in the study, and 2) application of existing algorithms and methods, and development of new ones as needed, for biomedical informatics research as well as for supporting other scientific research in the study. Our work has resulted in an advanced and recognized infrastructure system that supports translational research<sup>20, 21</sup>, a number of peer-reviewed publications and conference presentations, and competitive extramural funding. The following is a high-level summary of the activities and achievements in BMIX.

Quality Control (QC) and Quality Assurance (QA). To obtain high-quality data we have developed a set of QA programs, and many of the QC/QA procedures have been embedded into our SOPs. For example, the clinical QA program consists of: 1) a visual inspection of missing values and obvious inconsistencies, 2) double data-entry to reduce data entry errors, 3) a computer program named QAMetrics that deploys established QA rules to check for data integrity across the whole questionnaire, and 4) a web-based QA Issue Tracking system to enable efficient communications between WRI and the clinical sites for resolving QA problems of the clinical data<sup>22-25</sup>. In addition, we have developed a microarray QA program to identify outlier arrays from a population of similar arrays<sup>26, 27</sup>. Additional QA programs have been developed or are in development to ensure the quality of the data generated at WRI.

The Laboratory Information Management System (LIMS). A LIMS is needed to track and manage the sheer volume of data involved in translational research. For genomic and proteomic experimental data tracking, we have deployed the Synapseus system from GenoLogics deployed in 2008. Our clinical and tissue banking activities are tracked by the Clinical Laboratory Workflow System that we co-developed with Cimarron Software that was deployed in 2004<sup>28</sup>. To support the evolving needs of breast cancer studies, we are now in the process of developing a replacement system of the CLWS using up-to-date IT technologies.

The Data Warehouse for Translational Research (DW4TR). We have developed the DW4TR to integrate internally collected and generated clinicopathological data and molecular data from all the platforms. The development is done with our IT development partner, InforSense Ltd (currently part of IDBS). We have developed an innovative patient-centric, modularly-structured

data model for clinicopathologic data, which contains disease-independent and disease-specific submodules to enable easy expansion to support the study of new diseases. A specimen-centric data model has been developed for molecular data. We have also developed a temporal data model to resolve the problem of the temporal information alignment between the data and specimens that were collected at multiple points of the time. The DW4TR has two major intuitive interfaces, one is the Aggregated Biomedical-information Browser and the other is the Individual Subject Information Viewer. The DW4TR has successfully supported the Clinical Breast Care Project with a total of 799 data elements, and has been expanded to support additional translational research programs<sup>29-32</sup>.

Other Applications: We have developed a Biomedical Informatics Portal, using the InforSense analytical workflow platform, to host additional applications deployed or developed by us to facilitate integrative biomedical research. For example, we have developed a specimen selection tool for scientists to use to select specimens in our tissue bank using the temporal and other information in the DW4TR<sup>33</sup>. The user can specify the inclusion/exclusion criteria. We have also developed a Data Correction Utility to enable correction of data errors in the DW4TR.

Research: Our research focuses on clinicopathologic risk factors of breast cancers, biomarker identification, and mechanistic understanding of primary and metastatic breast cancer, and we do so mostly by applying existing data mining methods<sup>34, 35</sup>. We also have developed new algorithms and visualization tools for clinicopathologic risk factor analysis, and reported interesting new breast disease co-occurrence patterns among other findings<sup>36-42</sup>. We found ethnicity-specific benign breast disease co-occurrence with breast cancer<sup>43, 44</sup>, reported different characteristics of invasive breast cancer between Caucasian and African American women<sup>45</sup>, and identified subtype-specific clinicopathologic properties of breast cancers<sup>46, 47</sup>. We worked closely with laboratory scientists in identifying biomarkers of breast cancer and breast cancer metastasis, and used bioinformatics tools to seek mechanistic understanding of breast cancer development and metastasis<sup>48-50</sup>, and we collaborated with external scientists as well in study human diseases<sup>51, 52</sup>. In performing a peripheral blood gene expression analysis, we have identified differential enrichment of signaling pathways between breast cancer patients and normal subjects<sup>53</sup>. In another project using breast cancer tissues we have identified significant difference between lymph node positive and lymph node negative cases in a carefully stratified population. We also performed a transcription factor-centric computational analysis of genes differentially expressed in healthy breast tissues from African American and Caucasian women<sup>54</sup>. We used the semantic similarity measures to characterize human metabolic and regulatory pathways which will help us to understand human disease development mechanisms<sup>55, 56</sup>.

## 9. PLAN

**A. Subjects:** All participants will be patients attending the CBCP Breast Centers at WRNMMC, andWMC who consent to join in the study. They will include:

- i) patients seen in the Risk Reduction clinics;
- ii) patients diagnosed with breast cancer;
- iii) patients diagnosed with any other type of breast disorder.

The far majority of patients will be female, however, 1% of all breast cancers occur in males and these male patients with breast cancer (ii above) would be eligible as well.

**B. Inclusion and exclusion criteria:**

Inclusion criteria:

- 1) adult over the age of 18 years;
- 2) mentally competent and willing to provide informed consent;
- 3) military beneficiaries;
- 4) presenting to the Breast Center or the Women's Imaging Center of the WRNMMC with evidence of possible breast disease;
- 5) high risk beneficiaries

Exclusion criteria:

- 1) known history of Human Immunodeficiency Virus (HIV), or any type of hepatitis (e.g. HCV, HBV);
- 2) prion-mediated disease;
- 3) drug resistant tuberculosis or other infectious disease presenting a significant risk to personnel handling tissue or blood-derived products;
- 4) with pre-existing coagulopathies;
- 5) all other conditions, for whom invasive biopsy or surgery is medically contraindicated, shall be excluded

During their clinic visit patients will be offered the opportunity to participate in our studies. If they choose to participate, before signing an informed consent, they will have the study explained to them by a CBCP research team member (e.g. purpose of study, where samples will be stored, the GINA law, how to withdraw from the study) and provided time to discuss any concerns and/or ask questions regarding the study and their participation. They will be offered the opportunity to consent for future research to be performed on their samples. A copy of the ICD/HIPAA will be provided to the subject prior to departing.

**C. Study Design:** This will be a prospective study aimed at characterizing the gene and protein expression in blood. A blood sample will be collected from all participants at enrollment and later as follows for the different categories:

- 1) patients attending the risk reduction clinic will have blood collected every 6 months for five years
- 2) patients diagnosed with breast cancer will have blood collected every 3 months for 2 years then every 6 months for years 3-5 and thereafter yearly

3) patients diagnosed with all other breast disorders or presenting for regular breast screening (i.e. the “control” group) will have blood drawn at any clinically-directed subsequent follow-up visits (as recommended by their physician).

All of the follow-up visits after the initial visit are standard, clinically recommended follow-up times that are routinely recommended in these patient diagnostic categories. In other words, no patient will need to come in at future times merely for protocol-required blood drawing. They will be coming to the breast center only for physician directed and clinically indicated visits based on their clinical diagnosis. Then at those times, we will take the opportunity to obtain the blood samples for protocol patients. Additionally, at the time of the initial visit, patients will fill out a simple questionnaire, and some general diagnostic information will be obtained, assisted by one of the breast clinic nurses. The questionnaire is attached to this protocol as Appendix .

Specimen repository storage – All specimens will be held in CBCP freezers for an indefinite period of time into the future. Exceptions to this will occur when specimens are exhausted (consumed) due to their use in the approved research within CBCP labs and elsewhere, or when previously consented patients withdraw their consent and use of their specimens, at which time that patient’s specimens will be withdrawn from the repository and destroyed. The CBCP PI is responsible for maintaining the integrity of the Tissue Bank, Serum/Blood Bank Repository, and all databases. In the event that the named PI leaves WRNMMC, another WRNMMC investigator will be named through in-place processes), and that new CBCP PI will take full responsibility for the banked samples and data. If under any circumstance a new CBCP PI is not named, then the WRNMMC Institutional Review Board (IRB) will be notified, and the IRB will determine the disposition of this tissue and blood library.

#### **D. Methodology:**

**1. Blood collection and processing:** Patients will have up to 20 cc of venous blood collected by a phlebotomist or nurse. Before blood is drawn, a nurse will explain the study, consent the patient and also assist the patient in completing a questionnaire, along with a caffeine, fat intake, and stress level questionnaires. All consenting patients will be assigned a unique 9-digit identifier (CBCP number) and all associated tissue, serum, and/or blood samples to be archived will be assigned an 8-digit barcode. Both the CBCP numbers and the corresponding sample barcodes will be registered into the clinical database. The clinical database which describes the characteristics of the blood donors will be maintained. Linkage between the patient and research samples will be kept in a password protected-limited access database and in individual patient records held in gang-locked cabinets.

A portion of the blood will be placed in PAX-gene blood tubes (Quigen, Inc), which stabilizes the RNA for up to seven days. Plasma and serum will be processed within 2 hours by spinning blood down and aliquoting into separate tubes, after which they will be stored at -86 degrees in the CBCP freezer. The CBCP specimens will be transferred to the CBCP Genomics / Proteomics Center at the Windber Research Institute on a weekly scheduled basis, where the following procedures will be carried out:

**a. RNA, DNA and Protein from blood isolation from blood:** RNA will be purified from blood collected by PAXgene RNA tubes using the RNA Test Kit (PAXgene™, Quiagen Inc., CA). RNA will then be employed for other downstream analysis after determining concentration with a DNA/RNA calculator. DNA will be isolated from blood using the QIAamp DNA blood kit. This involves lysing the blood in appropriate buffer and loading this on a spin column. DNA gets bound to the silica gel-based membrane and pure DNA eluted in water or low-salt buffer. Protein will be isolated from plasma or serum samples using an extraction buffer (50mM Tris HCL, pH 7.5 and 0.1% Nonidet P-40) as described below. Approximately two volumes of buffer per volume of the sample are centrifuged at 14,000rpm/15minutes at room temperature. The supernatant is transferred to a fresh tube and about twice its volume of isopropanol added. The mixture is allowed to stand at room temperature for 15-20 minutes before centrifugation as described above. The supernatant is discarded and precipitate reconstituted in the Tris buffer. Total protein concentration will be determined by the Bradford protein assay procedure.

**b. Primary uses of the blood and serum specimens.**

The known primary uses under this protocol for the acquired serum and blood fall into seven major subsections (and the laboratory workflows for each are included in diagrammatic form in the Appendix to this protocol). The five major subsections are:

**i. Serum/Blood repository Banking** – this includes sample definition and receiving, flash freezing/labeling/storage, OCT embedding (placing the blood cells in a special preservative that protects the RNA/DNA during prolonged freezing)/labeling (putting identifier codes on each tissue sample for subsequent tracking)/storage, and Inventory/tracking. The inventory and tracking of all samples will be done electronically with barcodes and sample tracking software – initially the software will be Freezerworks™ to be followed by our own developed software module by Cimmaron Inc. that will be integrated with our laboratory analysis software GenoMax™.

**ii. Gene Expression Profiling** – RNA will be extracted from tissues using various kits appropriate to the storage conditions of the tissues (flash frozen, OCT, FFPE, etc). RNA will be used for Northern Analysis, RT-PCR, and mRNA expression analysis using Affymetrix arrays and in future RNA Seq analysis.

**iii. Sequencing** – Capillary electrophoresis, and Sequence Analysis is performed on plasmids, usually to confirm the structure of various constructions or by direct sequencing of genomic DNA

**iv. Genotyping** – Blood will undergo DNA quantification followed by PCR set-up, thermal cycling, SNP (single nucleotide polymorphism) reaction clean-up, capillary electrophoresis set-up, genotype calling, and genotype QC. Genotyping is also performed using Affymetrix arrays.

**v. Protein Expression Profiling** – After sample clean-up, serums/blood/tissues will undergo 1D- and 2D-Electrophoresis, Gel staining, Image acquisition and processing and analysis, isolation of proteins from the gel (“Spot picking”), enzymatic digestion of the selected proteins (“Spot digestion”), Mass spectral analysis, Spectrum processing, and protein identification

At the end of each of the above seven laboratory workflows, the data will be QA'd, analyzed using powerful genomics/proteomics software tools, and placed into the CBCP database / data

warehouse. QA of the data involves using software tools that interrogate the fields of the data that come out of the workflow stations, to ensure the data has consistency and is within expected or known ranges; any data found to be outside of expected ranges is not necessarily flawed, but is then identified for closer analysis by researchers.

Data will be stored in the CBCP server(s) and/or data warehouse (being developed with NCR Inc.), on CBCP sites at WRNMMC or WRI, and eventually at the CBCP Data Warehouse in Fort Detrick, MD (as part of our MANVT initiative, funded separately). The MANVT initiative is a Medical Area Network Virtual Technologies program. Its ultimate goal is to allow for the near-real time interaction between all of the CBCP sites (WRNMMC Windber clinical and WRI) with regards to movement of high volumes of research data across a virtual fiberoptic network linking the sites, as well as allow for creation of a data warehouse (situated at least in part at Fort Detrick, MD or to have a significant redundancy backup there).

**2. Data collection:** All individuals signed up for the study will be categorized into groups and further clinical classification and staging of patients with tumors will be based on the tumor-nodes-metastasis (TNM) classification and the Union Internationale Contre Cancer (UICC) staging system for breast cancer incorporating the TNM classification. Blood sample will be obtained from all participants at sign-up and at all other recommended visits as indicated in section “c”. There will be no patient identifiers sent to the laboratory with the blood samples but all participants will be coded for easy tracking of results and other clinical data. All results will be saved on the CBCP database.

At the time of the initial visit, patients will fill out a questionnaire, a caffeine, fat intake, and stress level. Some general diagnostic information will be obtained, assisted by breast clinic nurses. The questionnaire is attached to this protocol .

Patients will have an update to their basic information collected annually as follow-up visits permit.

**3. Sample size/Data Analysis:** Sample size will grow according to the frequency of patients attending the breast care center. Primary analysis will be comparison of gene expression patterns, protein expression profiles, and single base variations in DNA across the different patient categories. Laboratory analysis data will be correlated with clinical data on the individual patients to determine relationship of gene expression patterns and variation within DNA sequences with disease onset and progression. Specific secondary testing and analysis of samples will be submitted as a separate protocol.

Other institutions are intended to join in this protocol. These other sites will be both military and civilian. As with multi-center protocols, a Principal Investigator will be responsible at each separate institution for overseeing the protocol, conforming the protocol and consent form to their own institutional IRB requirements, and for achieving their own institution’s IRB approval prior to proceeding.

**4. Addendum:** The oversight of the CBCP, by charter, consists of the following two groups:

1. Steering Committee – this committee provides internal review and oversight, meets quarterly on average and consists of members from WRNMMC, Windber Research Institute, and the

Henry M. Jackson Foundation. The Steering Committee is chartered with CBCP budgetary oversight, review, and approval.

2. Scientific Advisory Board – provides external peer review and scientific direction. Comprised of seven noted leaders in their respective fields of Biomedical Informatics, Pathology, Cancer Research, Breast Cancer Pathology, Tissue Banking, Breast Nursing/Education, and Breast Center Management. Board serves to provide scientific oversight of protocols, research studies and future vision of project. SAB meets three times annually.

The Steering Committee first met in February 2000, and has continued to meet and provide Project oversight and direction continuously. The SAB was formed and, helps determine the future direction of research.

## **10. BIBLIOGRAPHY:**

1. Deng G et al., (1996). Loss of heterogeneity in normal tissue adjacent to breast carcinomas. *Science* 274:2057-2059.
2. Liotta LA et al., (1974). Quantitative relationship of intravascular tumor cells, tumor vessels and pulmonary metastasis following tumor implantation. *Cancer Res.* 34: 997-1004.
3. Sidransky, D. (2000). Circulating DNA. What we know and what we need to learn. In circulating nucleic acids in plasma or serum. Editors: Anker, P and Stroun M. *Annals of the New York Academy of Science* pg 1-4, Vol 906.
4. Postumus MD et al., (2000). Serum metalloproteinase 3 in early rheumatoid arthritis is correlated with disease activity and radiological progression. *J Rheumatol* 27(12):2761-2768.
5. Adams J et al., (2000). Vascular endothelial growth factor (VEGF) in breast cancer: comparison of plasma, serum and tissue VEGF and microdensity and effect of tamozifen. *Cancer Res.* 60(11):2898-905.
6. Ellsworth RE et al (2010) Chromosomal alterations in pure nonneoplastic breast lesions: implications for breast cancer progression. *Ann.Surg.Oncol.* epub ahead of print.
7. Ellsworth RE et al (2007) Correlation of levels and patterns of genomic instability with histological grading of DCIS. *Annals Surg Onc* 14 (11):3070-3077.
8. Ellsworth RE et al (2008a) Correlation of levels and patterns of genomic instability with histological grading of invasive breast tumors. *Breast Cancer Res.Treat.* 107 (2):259-265.
9. Ellsworth RE et al (2008b) Chromosomal alterations associated with the transition from in situ to invasive breast cancer. *Ann.Surg.Oncol.* 15 (9):2519-2525.
10. Ellsworth DL et al (2004c) Genomic patterns of allelic imbalance in disease free tissue adjacent to primary breast carcinomas. *Breast Cancer Res.Treat.* 88 (2):131-139.

11. Ellsworth DL et al (2004b) Outer breast quadrants demonstrate increased levels of genomic instability. *Annals Surg.Oncol.* 11 (9):861-868.
12. Ellsworth DL et al (2004a) Genomic instability in histologically normal breast tissues: implications for carcinogenesis. *Lancet Oncol.* 5 (12):753-758.
13. Ellsworth RE et al (2009) A gene expression signature that defines breast cancer metastases. *Clin.Exp.Metastasis* 26 (3):205-213.
14. Ellsworth RE et al (2008c) The Clinical Breast Care Project: an important resource in investigating environmental and genetic contributions to breast cancer in African American women. *Cell Tissue Bank* 9 (2):109-120.
15. Field LA et al (2011) Identification of differentially expressed genes in breast tumors from African American compared to Caucasian women. *Cancer under revision.*
16. Ellsworth RE et al (2005) Allelic imbalance in primary breast carcinomas and metastatic tumors of the axillary lymph nodes. *Mol.Cancer Res.* 3:71-77.
17. Becker TE et al (2008) The genomic heritage of lymph node metastases: implications for clinical management of patients with breast cancer. *Ann.Surg.Oncol.* 15 (4):1056-1063.
18. Ellsworth RE et al (2011) Differential gene expression in primary breast tumors associated with lymph node metastasis. *Int.J.Breast Cancer* epub.
19. Holliday C et al (2009) Genomic instability in the breast microenvironment? A critical evaluation of the evidence. *Expert Rev.Mol.Diagn.* 9 (7):667-678.
20. Hu H, Liebman M. Chapter 18. An introduction to biomedical informatics. In: Deng H, Shen H, Liu Y, Hu H, editors. *Current Topics in Human Genetics: Studies of Complex Diseases*: World Scientific Publishing Co.; 2007.
21. Hu H, Mural RJ, Liebman MN, editors. *Biomedical Informatics in Translational Research*. Boston, London: Artech House; 2008.
22. Zhang Y, Kvecher L, Sun W, et al. A quality assurance issue tracking tool to facilitate the enhancement of clinical data quality (submitted).
23. Sun W, Gutchell EM, Zhang Y, et al. Clinical data quality assurance and quality control measures in the Clinical Breast Care Project (in preparation).
24. Zhang Y, Sun W, Gutchell EM, et al. An issue tracking system to facilitate the enhancement of clinical data quality in the clinical breast care project. *AMIA Annu Symp Proc.* 2007:1173.

25. Sun W, Zhang Y, Brzeski H, et al. Quality assurance of a clinical breast disease database and preliminary data mining. *Advancing Practice, Instruction and Innovation through Informatics*. Lake Tahoe, CA; 2005.
26. Yang S, Guo X, Yang YC, et al. Detecting outlier microarray arrays by correlation and percentage of outliers spots. *Cancer Inform*. 2007;2:351-60.
27. Yang S, Hu H, Shriver C, et al. Identification of genes differentially expressed in blood of patients with breast abnormalities. *The 27th Annual San Antonio Breast Cancer Symposium*. San Antonio, TX; 2004.
28. Hu H, Kvecher L. Chapter 7. Data Tracking Systems. In: Hu H, Mural RJ, Liebman MN, editors. *Biomedical Informatics in Translational Research*; 2008. p. 111-36.
29. Hu H. Chapter 8. Data Centralization. In: Hu H, Mural RJ, Liebman MN, editors. *Biomedical Informatics in Translational Research*; 2008. p. 141-71.
30. Hu H, Brzeski H, Hutchins J, et al. Biomedical informatics: development of a comprehensive data warehouse for clinical and genomic breast cancer research. *Pharmacogenomics*. 2004 Oct;5(7):933-41.
31. Hu H, Correll M, Kvecher L, et al. DW4TR: a data warehouse for translational research (revision submitted).
32. Hu H, Correll M, Osmond M, et al. A clinical data warehouse to support translational research. *15th Annual International Conference on Intelligent Systems for Molecular Biology (ISMB)*. Vienna, Austria; 2007.
33. Kvecher L, Wu W, Hooke J, Shriver C, Mural R, Hai H. An Approach to Correlate the Temporal Information to Facilitate Specimen Selection in the Breast Cancer Research Project. *Cancer Research*. 2009;69:763s.
34. Zhang Y, Correll M, Hu H. Chapter 9. Data Analysis. In: Hu H, Mural RJ, Liebman MN, editors. *Biomedical Informatics in Translational Research*; 2008. p. 175-202.
35. Correll M, Beulah S, Munro R, Sheldon J, Guo Y, Hu H. Data Mining Tools and Techniques for Identification of Biomarkers for Cancer. In: Chen J, Lonardi S, editors. *Biological Data Mining: Chapman and Hall/CRC Press*; 2009.
36. Maskery S, Hu H, Liebman MN, Hooke J, Shriver CD, Taioli E. Traditional breast cancer risk factors and common breast pathologies in post-menopausal women. *The 30th San Antonio Breast Cancer Symposium*. San Antonio, TX; 2007.
37. Maskery SM, Hu H, Hooke J, Shriver CD, Liebman MN. A Bayesian derived network of breast pathology co-occurrence. *J Biomed Inform*. 2008 Apr;41(2):242-50.

38. Maskery SM, Hu H, Mural RJ, Hooke J, Shriver CD, Liebman MN. Separate DCIS progression pathways derived from breast disease heterogeneity data. American Society of Clinical Oncology Annual Meeting. Atlanta, GA; 2006.
39. Maskery SM, Zhang Y, Jordan RM, et al. Co-occurrence analysis for discovery of novel breast cancer pathology patterns. *IEEE Trans Inf Technol Biomed.* 2006 Jul;10(3):497-503.
40. Maskery SM, Zhang YH, Hu H, Shriver CD, Hooke J, Liebman MN. Caffeine intake, race, and risk of invasive breast cancer, lessons learned from data mining a clinical database Proceedings of the 19th IEEE International Symposium on Computer-Based Medical Systems. Salt Lake City, UT; 2006.
41. Sun W, Zhang Y, Hooke J, et al. Relationship between educational level and breast cancer The 28th San Antonio Breast Cancer Symposium. San Antonio, TX; 2006.
42. Bekhash A, Maskery SM, Kvecher L, et al. Clinical Breast Care Project data warehouse as a research environment for breast cancer risk factor studies (submitted).
43. Bekhash A, Saini J, Fan X, et al. Differential benign Breast Disease Co-occurrence with Cancer in Caucasian and African American Women *Cancer Research.* 2009;69:667s.
44. Bekhash A, Saini J, Li X, et al. Ethnicity Difference of Benign Breast Diseases in Breast Cancer and Non-Cancer Patients. *Cancer Research.* 2010 Dec 8-12:288s-9s.
45. Hu H, Zhang Y, Kvecher L, et al. Different characteristics of invasive breast cancers between Caucasian and African American women. The 29th San Antonio Breast Cancer Symposium. San Antonio, TX; 2006.
46. Saini J, Kovatich A, Bekhash A, et al. Association of Clinicopathologic Characteristics with IHC-Based Breast Cancer Subtypes. *Cancer Research.* 2009;69:635s.
47. Saini J, Kovatich A, Li X, et al. Association of Clinicopathologic Characteristics with IHC-based Breast Cancer Subtypes (in preparation).
48. Hu H, Somiari SB, Copper J, et al. Can circulating matrix metalloproteinases be predictors of breast cancer? – A neural network modeling study. In: Wang L, Chen K, Ong YS, editors. *Lecture Notes in Computer Science (LNCS)* Springer-Verlag; 2005.
49. Li X, Rapuri P, Melley J, et al. Comparative analysis of gene expression profiles in human breast cancer from microarray data using breast tissues and peripheral blood samples. International Conference on Intelligent Systems for Molecular Biology (ISMB). Boston, MA; 2010.

50. Saini J, Li X, Kvecher L, et al. Differential Gene Expression Analysis among Postmenopausal Caucasian Invasive Breast Cancer, Benign and Normal Subjects. *Cancer Research*. 2010 Dec 8-12:245s.
51. Serretti A, Olgiati P, Liebman MN, et al. Clinical prediction of antidepressant response in mood disorders: linear multivariate vs neural network models. *Psychiatry Research* 2007(152):223-31.
52. Deng H-W, Shen H, Liu Y, Hu H. *Current Topics in Human Genetics: Studies of Complex Diseases*: World Scientific Publishing Co; 2007.
53. Li X, Hu H, Shriver C, Mural R. Microarray Data Analysis Using Peripheral Blood Samples Suggests Differential Enrichment of Signaling Pathways between Breast Cancer Patients and Normal Subjects. *Cancer Research*. 2009;69:654s.
54. Hu H, Stegmaier P, Field L, et al. A transcription factor-centric computational analysis of genes differentially expressed in healthy breast tissues from African American and Caucasian women. *The 30th San Antonio Breast Cancer Symposium*. San Antonio, TX; 2007.
55. Guo X, Liu R, Shriver CD, Hu H, Liebman MN. Assessing semantic similarity measures for the characterization of human regulatory pathways. *Bioinformatics*. 2006 Apr 15;22(8):967-73.
56. Guo X, Shriver CD, Hu H, Liebman MN. Analysis of metabolic and regulatory pathways through Gene Ontology-derived semantic similarity measures. *AMIA Annu Symp Proc*. 2005:972.

**11. FACILITIES TO BE USED:** The facilities to be used in this research study include the following:

1. Walter Reed National Military Medical Center:

- a. The CBCP bldg 19, 3<sup>rd</sup> floor, which contains an Ambulatory Surgery Suite with a breast biopsy procedure room / recovery area; a fully-functional pathology laboratory, the tissue bank area with freezers, the separate patient blood-drawing room, and the clinical and counseling areas.
- b. The Operating Rooms on the 3<sup>rd</sup> floor in bldg 9 where cancer-directed surgeries take place.
- c. The Women's Imaging Center bldg 19, 3<sup>rd</sup> floor, where non-surgical breast biopsies will take place.

2. Windber Research Institute / Windber Medical Center:

The WRI Functional Genomics and Proteomics Facility and Tissue Banks, located in Windber, PA. The facility is a 6000 sq ft highly integrated laboratory that is adequately equipped for tissue processing and archiving, functional genomics, proteomics, high-throughput gene and protein expression profiling, microarray construction, DNA sequencing, single nucleotide

polymorphiam analysis, micro satellite analysis, mass spectral analysis, laser capture microdissection, bioimaging and microscopy. The CBCP tissue / blood/serum bank storage capability on-site at WRI is in isothermic liquid nitrogen freezers with a storage capacity of up to 80,000 specimens

**12. TIME REQUIRED TO COMPLETE:**

Anticipated start – November 2011  
 Expected completion – Indefinite

**13. FUNDING IMPLICATIONS:**

**Funding or other resources will be from outside sources: Yes X No \_\_\_\_\_**

This grant is administratively managed by Lee Bronfman of the Henry M. Jackson Foundation, 301-295-3890. No funds will be transferred into WRNMMC. All aspects of the funding of this protocol, to include personnel and disposable supplies, will come from the above grant.

**Budget request from DCI:**

	<u>FY 2001</u>	<u>FY 2002</u>	<u>TOTAL</u>
<b>Equipment:</b>			None
<b>Other -</b>			None
<b>Travel:</b>			None
<b>Reprints:</b>			
	<u>TOTAL</u>	\$ _____	\$ _____
			<u>\$ NONE</u>

**14. NUMBER OF PATIENTS/CHARTS/SPECIMENS TO BE STUDIED:** We have an expected (estimated) number of patients per year that will be enrolled at WRNMMC in the tissue protocol. We estimate that the number of patients newly diagnosed with breast cancer and treated at WRNMMC will average 2-300 per annum. of a breast biopsy for any reason at our institution may average 1100 per annum.

The total number to be studied therefore will be about 1500 per year for five years, up to 7500 total patients.

**15. TYPE OF SUBJECT POPULATION TO BGE STUDIED:** Military health care beneficiaries over the age of 18 years presenting with the diagnosis of breast-specific complaint

or diagnosis, breast cancer, or any radiologic or clinical breast lesion requiring assessment, follow-up, biopsy or tissue diagnosis.

**16. ENVIRONMENTAL IMPACT STATEMENT: for Questions contact:  
Laura Broussard, NNMC Environmental Program Office at 301-295-6436.**

a. Does any part of this protocol generate hazardous chemical waste as defined by Title 40 of the Code of Federal Regulation YES ( ) NO (X) ? If yes, at what stages and how much?

b. Have you considered any alternative procedures YES ( ) NO ( ) N/A (X) ? If yes, why were they rejected?

**17.**

**INVESTIGATOR AGREEMENT**

By submitting this protocol and providing an electronic signature in IRBNet, or an ink signature below, I agree to the following statements:

**General Assurance:** I agree to conduct the study as outlined herein. I certify that all procedures involving human subjects have been described in full.

**Starting the Study:** I understand that I cannot begin the study until I have received an approval letter documenting approval by the WRNMMC IRB.

**Consent:** I am responsible for assuring the quality of each subject's consent in accordance with current federal regulations. This includes ensuring that any "designee" that obtains consent on my behalf is completely familiar with the protocol and is qualified to perform this responsibility.

**Adverse Events:** I understand that I must report serious adverse events within two days to the IRB and provide this information to the Medical / Research Monitor. I will report unexpected (but not serious) adverse events that may possibly be related to participation in the protocol within 10 working days to the IRB using the same procedure.

**Training:** I verify that the personnel performing these procedures described in this protocol are technically competent, have been properly trained, and are appropriately qualified.

**Compensation:** I am aware that members of the research team are not authorized to accept any form of personal compensation for our efforts in conducting this research.

**Modifications:** I am aware that all changes to the protocol must be approved by the IRB before implementation. Examples of changes to protocols that require IRB approval include change of on-site PI, addition of personnel on study, increased sample size, addition of other data points, sources of outside funding, and addition of data collection sites.

**Deviations to the Protocol:** I am aware that any protocol deviations discovered by either the PI or auditing official will be immediately reported to the IRB. All corrective actions will be documented and become a part of the master study file, along with the report.

**Duplication of Effort:** I have made a reasonable good faith effort to ensure that this protocol is not an unnecessary duplication of previous experiments.

**Reports:** I agree to provide a Continuing Review Progress Report 30 days prior to the anniversary of the protocol's initial approval or as stipulated by the IRB. I agree to submit a final report within 30 days following completion or termination of the study.

**Maintain Study Files:** I agree to maintain a Study File that must be kept for three years from the date the study is closed (32 CFR 219.115(b)). If IND medication or IDE appliances are used, the file must be kept for 2 years after Food and Drug Administration (FDA) approval and can then be destroyed; or if no application is filed or approved, until 2 years after the study is discontinued and FDA notified (21CFR 312.62(c)). I acknowledge that research data is the property of the Command and will not be removed without prior approval. When I am scheduled to PCS or ETS, study records will be given to a new PI or the Department Chief, or turned over to the Department of Research Programs.

This file may be inspected at any time by Department of Research Programs, the FDA, and/or other applicable regulatory agencies responsible for the oversight of research. This file will include:

- A. The approved protocol and applicable amendments.
- B. The IRB minutes granting approval to initiate the study.
- C. IRB approval letter.
- D. Each Consent Form/HIPAA Authorization signed by the subject or surrogate.
- E. Continuing Review Progress Reports.
- F. Reports of adverse effects.
- G. Reports of any significant new findings found during the course of the study.
- H. All study documents generated from study date.
- I. Publications, abstracts, reprints resulting from study data.
- J. All information pertaining to an investigational drug or device.

**Publications:** I am aware that advertisements, abstracts, presentations or publications resulting from research protocols must have their products cleared by the Public Affairs Office, undergo Operation Security (OPSEC) review, undergo review for release of actionable medical information, and Publication Clearance.

**Applicable Regulations:** I am familiar with applicable regulations governing research, and will adhere to all of the requirements outlined in the DoD Multiple Project Assurance for the North Atlantic Regional Medical Command.

I understand that if I fail to comply with any of these responsibilities, all projects for which I am an investigator may be suspended.

[Provide electronic signature in IRBNet; if not available, provide ink signature with signature block.](#)

PRINCIPAL INVESTIGATOR

WRNMMC LEAD INVESTIGATOR

Rank, Name, Corps

Rank, Name, Corps

Title

Title

Department

Department

## **19. APPENDICES**

**APPENDIX A** – Study Workflows uploaded in IRBNet

**APPENDIX B** - Questionnaires uploaded in IRBNet

**APPENDIX C** – General Impact Statement included with original submission

**APPENDIX D** - Pathology Impact Statement uploaded in IRBNet

**APPENDIX E** – Signed Conflict of Interest Statement uploaded in IRBNet

**APPENDIX F** – MTA uploaded in IRBNet

**APPENDIX G** – NA (Advertisement Brochure/Flyer)

**APPENDIX H** - Consent Form uploaded in IRBNet

**APPENDIX I** - HIPAA uploaded in IRBNet



Walter Reed National Military Medical Center, Bethesda

IRBNet# 20705

IRB Approval Date: 17 May 2012

IRB Expiration Date: 16 May 2013

*Do Not Sign Past the Expiration Date*

Last Name \_\_\_\_\_

## WALTER REED NATIONAL MILITARY MEDICAL CENTER (WRNMMC)

**This Specimen banking consent form is valid only if it contains the IRB stamped date.**

### **Consent for Voluntary Participation in a Specimen Banking Study Entitled:**

“Creation of a Blood Library for the Analysis of Blood for Molecular Changes Associated with Breast Disease and Breast Cancer Development”

**Principal Investigator:** Craig D. Shriver, COL, MC, General Surgery Service, Department of Surgery

**Study site:**  WRNMMC

### **1. INTRODUCTION OF THE STUDY**

You are being asked to be in this research study because you are a patient in our Breast Center. All patients can participate, those with breast diseases and those without breast cancer or any breast disease. Your participation is entirely voluntary. Refusal to participate will not result in any penalty or loss of benefits to which you are otherwise entitled.

### **2. PURPOSE OF THE STUDY**

The purpose of the study is to store blood samples from patients seen in the Breast Center to be used in laboratory studies to identify gene marker and protein expression changes that may be associated with the onset and progression of breast disease and possible breast cancer. By carefully examining blood samples from many different breast conditions and from non-breast disease patients, we hope to improve our understanding of the causes and best treatments for breast disease. Another purpose is to correlate patient clinical information with the results of these laboratory studies.



### **3. PROCEDURES TO BE FOLLOWED**

We propose to carry out this study using blood collected from consenting donors. You would most likely be giving blood for other routine procedures during your visit and we would like to use some of your blood for our study. However, we may ask for blood even if you are not having blood drawn for other purposes.

If you agree to be in this study, a nurse or another person trained in drawing blood will draw up to 20 mls (about 4 teaspoons) of blood from a vein in your arm. This will take approximately 5 minutes. At any future visits to the Breast Center, or as recommended by your treating doctor for medical care purposes, another blood sample up to 20 mls (about 4 teaspoons) may be taken from a vein in your arm.

In this study, you will be asked to fill out a questionnaire. A nurse will ask you a series of questions about your medical history. The questionnaire will also involve information on caffeine and dietary fat intake, along with lifestyle stress. These questions will take about 45 minutes to complete. You may refuse to answer any question if you so choose. There will be some basic information collected about your breast and health history.

At all future breast center visits, as recommended by your treating physician, an additional 5-15 minutes will be needed to have your blood drawn. A follow up questionnaire will be administered annually with the assistance of one of our clinic nurses. The information collected on the questionnaire will be used to correlate the laboratory research results with your medical history, which may be important in correctly interpreting research results.

### **4. IDENTIFICATION OF YOUR BLOOD AND/OR TISSUE SAMPLES, HOW AND WHERE THEY WILL BE STORED AND WHO WILL HAVE ACCESS TO YOUR SAMPLES**

You will be assigned a code number referred to as the Clinical Breast Care Program (CBCP) number [not your name or social security number (SSN)], that will be used for the blood samples, the questionnaire and any clinical information. The only connection between your "CBCP number" and your name or SSN will be kept in - double-locked secure files, and via a password protected secure database.



Walter Reed National Military Medical Center, Bethesda

IRBNet# 20705

IRB Approval Date: 17 May 2012

IRB Expiration Date: 16 May 2013

*Do Not Sign Past the Expiration Date*

The blood samples will be frozen and stored indefinitely in either the Breast Care Center laboratory sites at Walter Reed, or our off-site facility, Windber Research Institute (WRI), Windber, PA. All samples will be kept in a secured (locked) freezer and identified only by codes for the CBCP. Only researchers within the CBCP or associated with the CBCP may have access to the samples and related clinical information. The blood samples will undergo the primary research studies as described in the next paragraphs.

The primary research uses of the blood samples are to study the genetic makeup, to examine protein changes, and to look for other markers that may be associated with breast disease. This information will be analyzed and stored in a computer database using your assigned code number.

Any remaining blood samples left over after these studies are done will remain frozen indefinitely until needed for approved research projects, if you consent for future unknown breast disease research. You will be provided the opportunity at the end of this consent form to indicate whether you will allow your blood samples to be used for future breast research. It is not possible at this time to predict all the potential future research uses of the samples.

If other research is to be conducted using these samples, the research will only be conducted after the proposed study protocol has received approval from the local Institutional Research Board (IRB).

Because these research projects are experimental, the results of any research done with your sample will not be given to you or your doctor. These reports will not be put in your health records. The research will not be used in decisions regarding your medical care. Third parties, such as relatives, physicians, and insurance companies, will not have access to your information from these studies.

In addition to this study, there are other research studies ongoing in the breast center. You may also be asked to participate in these other studies. If you agree to participate in these other studies, any data collected may be linked to the database for this research study. Your name or your SSN will not be in the data.



## **5. AMOUNT OF TIME FOR YOU TO COMPLETE THIS STUDY**

Your participation in this study will consist of the time necessary to read and understand this consent form, and to complete the questionnaires, about 45 minutes. Follow up visits, if any, have already been discussed in Item 3 above. The collection of your blood samples to be used in this study will be done after you have had the study explained to you and you provide written consent. The blood will be collected by one of our research nurses by venipuncture and take about 5 minutes.

Your samples will be stored until they are no longer needed for research, or until you request removal of your blood samples from our repository where the samples are stored. You may withdraw from this study at any time, and may request that any remaining samples be destroyed. The PI or the Research Protocol Coordinator will be the point of contact, if you choose to withdraw or if destruction of the samples becomes necessary.

Removal of your samples from the bank must be requested in writing to the researcher at the address at the end of this consent form.

## **6. NUMBER OF PEOPLE THAT WILL TAKE PART IN THIS STUDY**

This study is a multi-site study with participants from several hospitals participating in the study. There will be up to around 1200 people taking part in this study each year.

## **7. POSSIBLE RISKS OR DISCOMFORTS FROM BEING IN THIS STUDY**

There may be some discomfort from drawing blood, and you may experience some pain, swelling and a bruise at the site of the needle stick. Some people feel dizzy or light-headed for a few minutes after blood is drawn.

These include employment problems, difficulty obtaining life or health insurance, mental stress for you or your family members, and other social discomforts and inconveniences. These risks can impact the mental state of both you and your family members, which could require counseling.



## **8. POSSIBLE BENEFITS FROM BEING IN THIS STUDY**

You will not benefit from being in this study, but the possible benefits of future research from your blood would include learning and understanding more about what causes breast cancer, and how to prevent and treat it.

## **9. CONFIDENTIALITY/PRIVACY OF YOUR IDENTITY AND YOUR RESEARCH RECORDS**

The principal investigator will keep your research records in a secure and locked place. These records may be looked at by staff from the WRNMMC Department of Research Programs, the Institutional Review Board (IRB), and other government agencies as part of their duties. These duties include making sure that the research participants are protected. Confidentiality of your records will be protected to the extent possible under existing regulations and laws but cannot be guaranteed. Complete confidentiality cannot be promised, particularly for military personnel, because information bearing on your health may be required to be reported to appropriate medical or command authorities. Your name will not appear in any published paper or presentation related to this study.

This research study meets the confidentiality requirements of the Health Insurance Portability and Accountability Act (HIPAA).

## **10. CONDITIONS UNDER WHICH YOUR PARTICIPATION IN THIS STUDY MAY BE STOPPED WITHOUT YOUR CONSENT**

Your taking part in this study may be stopped without your consent if remaining in the study might be dangerous or harmful to you. Your taking part in this study may also be stopped without your consent if the military mission requires it, or if you lose your right to receive medical care at a military hospital.

## **11. ELIGIBILITY AND PAYMENT FOR BEING IN THIS STUDY**

You will not receive any payment for being in this study.

## **12. COMPENSATION TO YOU IF INJURED & LIMITS TO YOUR MEDICAL CARE**

Should you be injured as a direct result of your participation in this study, you will be provided medical care for that injury at no cost to you. You will not



receive any compensation (payment) for injury. You should also understand that this is not a waiver or release of your legal rights.

Medical care is limited to the care normally allowed for Department of Defense (DOD) health care beneficiaries (patients eligible for care at military hospitals and clinics). Necessary medical care does not include in-home care or nursing home care. You should discuss this issue thoroughly with the PI before you enroll in this study.

If at any time you believe you have a study-related injury or illness as a result of participating in this research project, you should contact the PI. For questions about your rights as a research participant, call the Walter Reed National Military Medical Center Institutional Review Board (a group of people who review the research to protect your rights) at 301-295-2275, or the Staff Judge Advocate (SJA) Office at 301-295-2215.

### **13. COSTS THAT MAY RESULT FROM TAKING PART IN THIS STUDY**

There is no charge to you for taking part in this study.

### **14. IF YOU DECIDE TO STOP TAKING PART IN THIS STUDY AND THE INSTRUCTIONS FOR STOPPING EARLY**

You have the right to withdraw from this study at any time. This can be done by contacting the PI, or the research protocol coordinator, at the phone number in section 19 of this consent form. You may request that your sample be destroyed or that your identification be removed without destroying the sample. This request needs to be in writing to the address at the end of this consent form; by leaving the study at any time, you in no way risk losing your right to medical care.

### **15. RESEARCH RESULTS**

The results from tests that may be done on the blood that you donate for future research will not be given to you or to your doctor, even if you ask that this be done and the results will not become part of your hospital medical record because these tests will be done for research purposes only.

Your individual testing results that may be done on the blood that you donate for future research will not be released to any third party, including family members,



personal physicians, insurers or employers, under any circumstance unless required by law.

You should understand that there is a chance that the samples you are providing in this study may be used in other research studies related to breast disease. You may not be given any notice of future use of your sample. The confidentiality of your specimens will be protected in the same fashion as stated previously in the Description of This Study and/or Confidentiality section. You will not be personally identified in any published paper of these other research studies. Any other researcher using these samples will be under the direction of the CBCP, no direct identifiers such as your name, address, social security number (SSN), etc. will be provided to these researchers, but they could be provided with some identifying clinical information about you, such as date of surgery, diagnosis, treatment.

## **16. YOUR RIGHTS IF YOU TAKE PART IN THIS STUDY**

Participating in this study is your choice. You may choose either to participate or not to participate in the study. If you choose to participate in this study, you have the right to withdraw from this study at any time. This can be done by contacting the PI, or the research protocol coordinator, at the phone number in section 19 of this consent form. You may request that your sample be destroyed or that your identification be removed without destroying the sample. This request needs to be in writing to the address provided at the end of this consent form.

No matter what decision you make, there will be no penalty to you and you will not lose any of your regular benefits or risk losing your right to medical care.

## **17. THE NEW FEDERAL LAW – “GENETIC INFORMATION NON-DISCRIMINATION ACT” (GINA)**

This new Federal law generally makes it illegal for health insurance companies, group health plans, and most employers to discriminate against you based on your genetic information. This law generally will protect you in the following ways:

Health insurance companies and group health plans may not request your genetic information that we get from this research.



Health insurance companies and group health plans may not use your genetic information when making decisions regarding your eligibility or premiums.

Employers with 15 or more employees may not use your genetic information that we get from this research when making a decision to hire, promote, or fire you or when setting the terms of your employment.

All health insurance companies and group health plans must follow this law, dated 21 May 2010.

Be aware that this new Federal law does not protect you against genetic discrimination by companies that sell life insurance, disability insurance, or long-term care insurance."

## **18. AUTHORIZATION FOR RESEARCH USE OF PROTECTED HEALTH INFORMATION**

The Federal Health Insurance Portability and Accountability Act (**HIPAA**) includes a Privacy Rule that gives special safeguards to Protected Health Information (**PHI**) that is identifiable, in other words, can be directly linked to you (for example, by your name, social security number, etc.). We are required to advise you on how your PHI will be used.

### **(1) What information will be collected?**

For this research study, you will be asked some basic information about your breast, health, and medical history. The questionnaire will also involve information on caffeine intake, dietary fat intake, and lifestyle stress.

### **(2) Who may use your PHI within the Military Healthcare System?**

The members of the research team will have access to your health information in order to find out if you qualify to participate in this study, and for analyzing research data. Additionally, your PHI may be made available to health oversight groups such as the WRNMMC Department of Research Programs and Institutional Review Board, and other government agencies as part of their duties. These duties include making sure that research subjects are protected.

### **(3) What persons outside of the Military Healthcare System who are under the HIPAA requirements will receive your PHI?**

Information outside the Military Healthcare System such as dates of surgery, diagnosis, treatment, etc. related to the CBCP code assigned to your research file



may be shared with the research team for the purpose of performing an IRB approved research study. Your name, address and contact information will not be shared.

**(4) What is the purpose for using or disclosing your PHI?**

The members of the research team need to use your PHI in order to analyze the information related to the specific kind research study being performed. Researchers associated with the CBCP will use your blood samples to study the genetic makeup, protein changes, and other research markers that may be associated with breast disease, and unknown future research. The clinical information is important for performing analysis.

**(5) How long will the researchers keep your PHI?**

There is no expiration date for this study to maintain your PHI, since this study collects blood samples for placement within a tissue/blood bank.

**(6) Can you review your own research information?**

You will not be able to look at your research information.

**(7) Can you cancel this Authorization?**

Yes. If you cancel this Authorization, however, you will no longer be included in the research study. The information and samples we collected from you may be destroyed at your request. The research and samples that may have been used prior to your request will not be able to be destroyed or withdrawn because they will already be among the statistics of the study. If you wish to cancel your Authorization, please contact the Principal Investigator or the Research Protocol Coordinator in writing.

**(8) What will happen if you decide not to grant this Authorization?**

If you decide not to grant this Authorization, you will not be included in this research study. Refusal to grant this Authorization will not result in any loss of medical benefits to which you are otherwise entitled.

**(9) Can your PHI be disclosed to parties not included in this Authorization who are not under the HIPAA requirements?**

There is a potential that your research information will be shared with another party not listed in this Authorization in order to meet legal or regulatory requirements. Examples of persons who may access your PHI include representatives of the Food and Drug Administration, the Department of Health and Human Services (DHHS) Office for Human Research Protections (OHRP), and



Walter Reed National Military Medical Center, Bethesda

IRBNet# 20705

IRB Approval Date: 17 May 2012

IRB Expiration Date: 16 May 2013

Do Not Sign Past the Expiration Date

the DHHS Office for Civil Rights. This disclosure is unlikely to occur, but in that case, your health information would no longer be protected by the HIPAA Privacy Rule.

**(10) Who should you contact if you have any complaints?**

If you believe your privacy rights have been violated, you may send a written complaint to the WRNMMC Privacy Officer, located at 8901 Wisconsin Avenue, Bethesda, MD 20889-5600, telephone: 301-319-4775. Your signature at the end of this document acknowledges that you authorize WRNMMC personnel to use and disclose your Protected Health Information (PHI) collected about you for research purposes as described above.

**19. CONTACTS FOR QUESTIONS ABOUT THE STUDY**

If you have questions about the study, you should contact either the PI or the Research Protocol Coordinator at WRNMMC 301-319-2385.

**20. OTHER FUTURE RESEARCH FOR WHICH YOUR BLOOD AND/OR TISSUE SAMPLES TAKEN DURING THIS STUDY COULD BE USED**

Please read carefully each sentence below and think about your choices. After reading each sentence, **circle** "yes" or "no", include the date and your initials. If you have any questions please talk to your doctor.

**BY SIGNING THIS FORM, YOU ARE AGREEING THAT:**

a. My serum (blood) samples may be kept for use in future research to learn about, prevent, detect, or treat breast cancer, and to learn about whether certain genes relate to breast cancer.

YES NO Participant's Initials \_\_\_\_\_ Date \_\_\_\_\_

b. The principal investigator (or someone he or she chooses) may contact me in the future to ask me to take part in future research.

YES NO Participant's Initials \_\_\_\_\_ Date \_\_\_\_\_



Walter Reed National Military Medical Center, Bethesda

IRBNet# 20705

IRB Approval Date: 17 May 2012

IRB Expiration Date: 16 May 2013

*Do Not Sign Past the Expiration Date*

**SIGNATURE OF RESEARCH SUBJECT**

I have read (or someone has read to me) the information in this consent form. I have been given a chance to ask questions and all of my questions have been answered to my satisfaction.

**BY SIGNING THIS CONSENT FORM, YOU FREELY AGREE TO TAKE PART IN THE RESEARCH IT DESCRIBES.**

\_\_\_\_\_  
Participant's Signature

\_\_\_\_\_  
Date

\_\_\_\_\_  
[First name, MI, Last Name]  
Participant's **Printed** Name

\_\_\_\_\_  
SSN

\_\_\_\_\_  
Sponsor's Last 4

\_\_\_\_\_  
Street or mailing address

\_\_\_\_\_  
City

\_\_\_\_\_  
State and Zip Code

\_\_\_\_\_  
Date of Birth

**A copy of this consent form will be provided to you.**

**SIGNATURE OF INVESTIGATOR**

I have explained the research to the volunteer and answered all of his/her questions. I believe that the volunteer subject understands the information described in this document and freely consents to participate.

\_\_\_\_\_  
Investigator's Signature

\_\_\_\_\_  
Date (must be the same as the participant's)

\_\_\_\_\_  
Investigator's **Printed** Name

MCHL-SG

DATE: 3 May 2012

(rev 24 Apr 01; rev 3 May 2001; rev 7 July 01; rev 20 July 01; rev 14 Sep 01; rev 4 Oct 01; (USU rev 22 Oct 01) rev 11 Jun 02, 16 July 03, Nov 03, June 2004, 1 Aug 05, 20 Oct 2010)

MEMORANDUM FOR CHIEF, DEPARTMENT OF RESEARCH PROGRAMS, WALTER REED NATIONAL MILITARY MEDICAL CENTER

SUBJECT: Application and Request for Approval of Clinical Investigation Study Proposal

**1. PROTOCOL TITLE: Tissue and Blood Library establishment for Molecular, Biochemical and Histologic Study of Breast Disease**

**2. PRINCIPAL INVESTIGATOR:**

COL Craig D. Shriver, MC, USA  
Chief, General Surgery Service  
Program Director, NCC General Surgery

PI, Clinical Breast Care Project  
Professor of Surgery, USUHS  
Interim Director, Cancer Center, WRNMMC/JTF  
Phone: 301-400-2889  
Email: craig.shriver@med.navy.mil  
Research Course Training Date: Sep 2010

**CO-PRINCIPAL INVESTIGATOR:**

CAPT Ralph C. Jones, MC, USN  
General Surgeon  
Chair of Cancer Committee  
Phone: 301-295-6335  
Email: ralph.jones@med.navy.mil  
Research Course Training Date: Jan 2010

**3. ASSOCIATE INVESTIGATORS:**

CAPT David M. Larson, MC, USN  
Laboratory Director  
Phone: 301-319-4350  
Email: david.larson@med.navy.mil  
Research Course Training Date: Mar 2010

Claudia E. Galbo, MD  
Head Clinical Mammography Services  
Women's Imaging Center  
Phone: 301-319-8548  
Email: claudia.galbo@med.navy.mil  
Research Course Training Date: Jan 2010

Jeffrey A. Hooke, MD  
Head of Pathology, CBCP  
Phone 301-400-1483  
Email: jeffrey.hooke@med.navy.mil  
Research Course Training Date: Mar 2010

#### **4. COLLABORATING PERSONNEL:**

Richard J. Mural, Ph.D.  
Chief Scientific Officer  
Windber Research Institute (WRI)  
620 Seventh Street  
Windber, PA 15963-1300  
Phone 814- 361-6950  
Email: r.mural@wriwindber.org  
Research Course Training Date: Dec 2010

Responsibility: Carrying out successful receipt of anonymized donor specimens transferred to the off-site functional genomics, proteomics and tissue banking facility, and to manage and oversee the scientific experiments and handling of results of all studies involving the said specimens at said facility.

Diana Craig, MD  
Breast Surgeon, Windber Medical Center  
600 Somerset Avenue  
Windber, PA 15963  
Phone 814-467-4134  
Email: dcraig@windbercare.org  
Research Course Training Date: Jun 2009

Responsibility: Acting as on-site PI at the collaborating center in Windber, PA, for the Tissue and Blood repository and high-throughput research initiatives as described in this protocol.

**5. MEDICAL MONITOR:** Minimal risk study, medical monitor not required.

## 6. OBJECTIVES:

Although there have been remarkable improvements in breast cancer diagnosis and management, most of the complex molecular mechanisms associated with the onset, progression and/or severity of breast cancer are still not well understood. The Clinical Breast Care Project (CBCP) is a congressionally mandated and funded military-civilian collaboration between Windber Medical Center (WMC) in Windber, PA, Walter Reed National Military Medical Center (WRNMMC) Bethesda, MD, and the Telemedicine and Advanced Technology Research Center (TATRC) Fort Detrick, MD. As part of the CBCP we propose to carry out molecular, biochemical and histologic analysis of breast tissue and/or blood to provide more insights on the molecular mechanisms that may be relevant in breast cancer development and breast diseases. To achieve this aim, a large supply and a wide variety of good quality tissue samples is needed. Unfortunately, good quality donor breast tissue is extremely scarce and when available is often not backed by a comprehensive medical history and/or not a good representation of the target population or study area. The non-availability of a steady and consistent supply of good quality tissue limits the systematic analysis of tissues and negatively impacts the generation of biologically useful information at the laboratories and by extension clinical practice. The objectives of this project are therefore:

- A. Acquisition and banking of breast tissue, lymph nodes, and/or blood from informed and consenting donors,
- B. Experimental analysis of DNA, RNA and/or proteins isolated from donor tissues for molecular, biochemical, and/or histopathological analysis,
- C. Establishment of an integrated and relational database for tissue/serum and patient clinical characteristics that will provide the resources necessary to achieve the following future goals:
  - 1. Identify single nucleotide polymorphisms (SNPs) present in DNA from diseased breast tissue (as defined by histologic criteria) as compared to breast tissue without disease, lymph nodes with and without metastatic deposits, and/or DNA derived from patient leucocytes
  - 2. Identify differences in mRNA and protein expression associated with breast disease (as defined by histologic criteria) as compared to normal breast tissue and lymph nodes with and without metastatic deposits
  - 3. Correlate SNPs and differences in mRNA and protein expression associated with diseased breast and nodal tissue (as defined by histologic criteria) with the corresponding clinical patient database
  - 4. Identify factors within patient serum and/or blood-derived cellular components that correlate with patient risk factors or clinical status as defined in the corresponding clinical patient database

- 7. MEDICAL APPLICATION:** One of the major challenges facing researchers and clinicians today is to understand the mechanisms associated with the evolution of benign breast disease and/or transition of breast disorders to breast cancer. The creation of a comprehensive tissue/blood bank is essential to the application of modern molecular and genetic analysis of breast diseases. Amongst other goals, this entire (overall CBCP) project will:
- A. Establish a repository of high quality breast tissue and related (lymph nodal, blood) specimens for research on breast cancer and associated breast diseases, and
  - B. Permit the establishment of a single relational database with accurate and comprehensive biologically and clinically relevant information on breast diseases. Since the standard of care for treating breast diseases and breast cancer is based on a multidisciplinary model that integrates prevention, screening, diagnosis, treatment and management, this CBCP project will provide the necessary framework for such an integrated approach, which will positively impact the future management of breast cancer.
- 8. BACKGROUND AND SIGNIFICANCE:** Breast cancer is the second leading cause of cancer-related deaths in American women and the most common cause of cancer-related deaths between the ages 15 and 54<sup>1</sup>. The National Cancer Institute's Surveillance, Epidemiology and End Results (SEER) program reported that 32% of all malignancies among women are breast cancer (1). Over the past forty years the incidence rate of breast cancer has been steadily increasing worldwide with a rate of one to two percent per year in the United States. The American Cancer Society estimated that 192,370 new invasive cases of breast cancer were diagnosed in the USA in the year 2009, as well as an estimated 62,280 additional cases of in situ breast cancer. With this high incidence rate, even small improvements in treatment could represent tens of thousands of lives saved every year. For a systematic study of breast disorders, a large supply of good quality breast tissue specimens is needed.

The concept of preserving tissue dates back to the ancient Egyptians. As early as 5000 BC the Egyptians preserved bodies by burying them in shallow graves in the desert where dry sand and the sun's heat combined to desiccate and preserve the body. By 2600 B.C. the Egyptians had successfully developed the process of intentional mummification. They first eviscerated the body and then used natron as a dehydrating agent on the body tissues and viscera. Thousands of years later researchers have identified schistosomiasis in many of these preserved bodies, now called "mummies"<sup>2</sup>. It is now the work of the early Egyptians that has made possible the study of the epidemiology of schistosomiasis over a 5,000-year period<sup>3</sup>. This serves as the earliest example of how tissue preservation provides a wealth of information that can be used to study disease.

The lessons learned from history demonstrate to researchers that preserved human tissue provides an invaluable tool to study the basic science of disease. This realization has led to the development of serum and tissue banks. Knowledge gained from cells at their most

primitive level may translate into information relevant to clinical treatment and prevention of disease. As an example, serum banks as described by Evans and Mueller were instrumental in discovering an association between EBV and nasopharyngeal carcinoma in 1966<sup>4</sup>. As a logical follow up to serum banks, tissue banks such as the National Psoriasis Tissue Bank (NPTB) were founded<sup>5</sup>. Tissue provided by this bank led Dr. Bowcock and colleagues to map the first psoriasis gene to chromosome 17q<sup>6</sup>.

The desire to find a cure for breast cancer has spurred increased interest in breast tumor tissue banks across the United States and Canada. One example is the National Institute of Canada-Manitoba Breast Tumor Bank Project that was established in 1993. This tumor bank provides a national resource consisting of a pre-assembled data set of matched samples of paraffin-embedded and frozen tissue with corresponding pathological and clinical data<sup>7</sup>. In the first three years of operation, the bank accrued over 1,000 cases and has provided support for 20 research projects in 16 laboratories in Canada. Unfortunately, a repository of good quality breast tissue, lymph nodes, and blood / serum is not available in the USA. Although a small number of tissues can be obtained from different tissue banks for research purposes, they are often not backed by a comprehensive clinical and/or histopathologic information database, and they often are not good representations of the study population and/or study area.

High throughput gene and protein expression profiling by microarray is on the forefront in the battle to understand the mechanisms associated with breast cancer development, progression and severity. Microarrays consist of thousands of different complementary DNA (cDNA) clones spotted onto known locations on glass slides which are screened with cDNA populations from 2-3 individuals that have been differentially labeled with fluorescent dyes. Complimentary strands in the labeled probe hybridize to the immobilized target sequences and the signal generated is captured by laser confocal microscopy. The data are fluorescent points that may be characterized using the variables of color and intensity to generate ratios between the cDNA of the control (normal breast tissue) and the experimental tissue (diseased breast tissue). In order to discover differentially expressed gene profiles of diseased breast tissue, it must be compared with normal breast tissue. Thus, the current protocol proposes to acquire and archive tissue and/or blood samples from both normal and abnormal breast tissues and lymph nodes. The gene-expression profile of one tissue sample generally include both disease-specific patterns as well as idiosyncratic, patient or sample-specific patterns. Therefore, to draw any useful conclusions concerning the association between gene expression patterns and disease, it is necessary to analyze a large number of good quality tissue samples from many patients with similar clinical characteristics and well-documented case history. This protocol aims to establish a tissue repository that will provide good quality specimens for studies aimed at identifying molecular, biochemical, and histologic mechanisms associated with breast cancer and its development.

It is, however, also necessary to examine the protein expression profile in the same tissues,

because proteins are functional units of genes and are therefore usually involved in, or associated with, the complex biochemical processes that lead to carcinogenesis. Current high throughput proteomic technology is somewhat analogous to microarray technology, except that it relies on the electrophoretic separation of protein species with subsequent analysis of oligopeptides by mass spectrometry and/or microsequencing of protein-derived peptides. For comparison of the protein-expression profiles, control and experimental tissues are differentially labeled with a fluorescent dyes and simultaneously separated by 2D electrophoresis. The fluorescent signals generated are captured by multiple wavelength scanning, digitized, signals superimposed and differentially expressed proteins identified on the basis of resultant fluorescence of each spot. “Spots” representing separated proteins that are differentially expressed between the control and experimental tissues are then excised from the gel, digested and subsequently analyzed by mass spectroscopy. Again, accurate identification of correlations between disease and alterations in proteins requires the use of large numbers of both “normal” control tissue samples as well as samples with various types and/or stages of diseases.

Microarray and proteomics technologies offer great promise for the study, classification, and treatment of breast cancer. The study of gene expression in primary breast tumors is complicated. The difficulty lies in the fact that breast tumors consist of many different cell types, such as carcinoma cells (which are morphologically and genetically diverse), epithelial cells, stromal cells, adipose cells, endothelial cells, and infiltrating lymphocytes. However, Pollack and Perou have used cDNA microarrays to identify patterns of gene expression in mammary epithelial cells in culture and in primary breast cancers<sup>8</sup> and quite recently, a similar strategy has been used to study gene expression profiles in hereditary breast cancer<sup>9</sup>. These studies demonstrated two clusters of co-expressed genes with patterns that correlated with variations in cell proliferation rates and with activation of the specific signal transduction pathway. Their results support the feasibility of microarray technology applied to breast tumors.

Microarray technology is a proven novel scientific approach to the study of tumor biology at the global scale. In the recent past it has been applied to the study of other diseases, such as lymphoma, and has led to ground breaking discoveries. For example, Alizadeh et al. identified two molecularly distinct forms of diffuse large B-cell lymphoma. One type expressed genes characteristic of germinal center B-cells and the other expressed genes normally induced during in vitro activation of peripheral blood B-cells. He further observed that patients with germinal center B-like lymphoma had a significantly better overall survival<sup>10</sup>. Microarrays have similarly been used in studies of patients with prostate cancer. For example, Bubendorf et al. have found that amplifications of the androgen receptor gene were seen in 22% of hormone refractory disease<sup>11</sup>. This finding implicates a role for this gene in hormone therapy failure in prostate cancer treatment.

The ability to analyze thousands of genes and proteins simultaneously has many potential

impacts. It allows researchers to study the gene's relationship to breast cancer and to understand the pathways of disease development and progression. These discoveries will enable researches to develop novel drugs targeting disease pathways and disease-specific targets. Additionally, research application of these technologies may yield more specific and clinically relevant classification and subtyping of breast disease. Expression profiling of tumor tissue therefore is a powerful method to identify novel genes involved in oncogenesis. However, the variability in the quality and quantity of fresh tissue is an obstacle to the application of this powerful technology<sup>12</sup>.

We aim to create a large library of fresh and formalin-fixed paraffin-embedded (FFPE) breast tissue, and analyze it with high-throughput technologies in our CBCP laboratories for expression profiling. This tissue will be collected as part of that obtained during the course of clinically required treatment for breast disorders or other clinically indicated breast surgeries. This would provide us with the resources to study the evolution of breast disease from normal epithelium to cancerous tumor at the genetic level. A concomitant blood library will be used by our functional genomics and proteomic laboratory to investigate gene and protein marker expressions in patients with various forms of breast disease, and in those patients at above average risk for developing breast cancer (Risk Reduction Clinic patients). The biologically relevant information generated from the laboratory will then be linked with our clinical database to identify and/or understand associations between biological properties of the tissue and clinical presentations. Such a multidisciplinary, linked database will enable researchers and clinicians to follow patients longitudinally, and ultimately will facilitate the correlation between discoveries made in the lab and patient care.

Now that the Human Genome has been completely mapped<sup>13</sup>, understanding the molecular basis of cancer, and specifically breast cancer, is a challenge for this century. Availability of a large breast-derived tissue bank backed by a comprehensive clinical database is a necessary prerequisite to the application of modern molecular analysis technologies to this endeavor. Employing functional genomic and proteomics technologies to understand the genetic information contained within such a repository is the hope for the future of preventing, arresting, and eradicating breast cancer.

The current standard of care for treating breast diseases and breast cancer is based on a multidisciplinary model that integrates prevention, screening, diagnosis, treatment, and continuing care. The CBCP is modeled after this multidisciplinary approach. However, the CBCP is unique in its field. It incorporates advances in breast cancer risk reduction, informatics, tissue banking, and research. The present protocol represents one aspect of this integrated approach. It will interdigitate and synergize with the clinical, informatics, and research components of the CBCP, to achieve the overall objective of improving breast care and reducing the morbidity and mortality associated with breast disease.

In 2002, a panel of microsatellite markers was developed to assess allelic imbalance (AI) at the 26 most commonly altered regions in breast cancer. This panel was used to investigate how tumors develop and progress. For example, this panel was applied to 42 columnar cell lesions (CCL) and 31 atypical ductal hyperplasias (ADH), both non-neoplastic diseases of the breast. Of note, we assessed genetic changes only in pure CCL and ADH without concomitant DCIS or invasive breast cancer (IBC). This approach was quite different than the then current body of literature that had evaluated synchronous non-neoplastic lesions and, based on shared genetic changes between the neoplastic and non-neoplastic diseases, suggested that CCL and ADH were non-obligatory precursors to DCIS/IBC. Our study was one of the first to look at pure CCL and ADH, and unlike the other studies, we found very few genetic changes, suggesting that these lesions are genetically naïve and that earlier models of progression based on synchronous lesions may not accurately reflect tumor progression<sup>14</sup>.

In contrast, the 100 pure DCIS lesions we assayed shared many of the same alterations found in IBC, suggesting that DCIS lesions are genetically advanced<sup>15</sup>. We also found that low-grade DCIS and IBC are genetically distinct from high-grade DCIS and IBC, based largely on high rates of loss of chromosome 16q in low-grade disease and high rates of alterations at 9p21, 11q23, 13q14, 17p13.1 and 17p12<sup>16, 17</sup>. These data suggest that high-grade disease does not progress from low-grade disease but rather that low- and high-grade disease are distinct diseases.

CBCP has been investigating molecular changes in the tumor microenvironment. Tumor epithelial and corresponding adjacent stroma <850  $\mu$ m from the tumor were laser microdissected from 30 formalin-fixed, paraffin-embedded tumor specimens and evaluated using the allelic imbalance panel. Discordant patterns of LOH/AI between epithelial and stromal components were detected such that hierarchical clustering samples from the same patient rarely clustered together<sup>18</sup>. In a second study evaluating LOH/AI data generated from archival breast quadrants from 21 patients who underwent mastectomy, levels of LOH/AI were significantly higher ( $P < 0.05$ ) in the tumor-adjacent (15.4%) compared to distant (3.7%) tissues, suggesting that chromosomal alterations in stroma close to the tumor may promote tumorigenesis<sup>19</sup>. These data lead to the invitation to write review papers for *Lancet Oncology* and *Expert Reviews in Molecular Diagnosis*<sup>20, 21</sup>.

We have also had an interest in health care disparities between Caucasian (CW) and African American women (AAW) with breast cancer. Enrollment of patients within the DOD health-care system affords us the opportunity to evaluate clinical, epidemiological and molecular differences in patients treated within an equal-access health-care system, thus minimizing the impact of disparate access to quality-health care seen in the general population. Recruitment of African Americans into the Clinical Breast Care Project (CBCP) has been effective: in 2007, 25% of the patients enrolled were self-described AAW<sup>22</sup>. No differences in education levels, breast health practices (e.g., self-breast exam, routine screening mammograms) or reproductive histories were seen between AAW and CW with invasive breast cancer,

however, AAW were more likely to use oral contraceptives, less likely to use HRT or to breastfeed and were more likely to be obese and have extremely high fat-intake. Breast tumors from AAW were also more likely to be diagnosed at a younger age, and are larger, more frequently ER negative and/or triple negative, and of higher grade. Evaluation of gene expression patterns between AAW and CW matched by age, grade and ER status revealed differentially expression of 23 genes, suggesting that tumors from AAW and CW are molecularly different and these differences may contribute to the less favorable outcome in AAW<sup>23</sup>.

We have also performed a number of studies to determine the molecular mechanisms of metastasis. Using our AI panel, we evaluated patterns of chromosomal changes in primary breast tumors and matched metastatic lymph node tumors and found disparate patterns of alterations between the two tumor types, suggesting that metastatic cells diverge early in tumorigenesis<sup>24</sup>. A follow-up study evaluated genetic changes between primary tumors (n=26) and multiple lymph node tumors (n=146) to determine how metastasis spreads. Divergent patterns of chromosomal alterations were seen between lymph nodes and the primary tumor, but also between metastatic lymph node tumors from the same patients. This genetic heterogeneity may impact response to adjuvant therapy, recurrence, and survival, and thus may be important to improving clinical management of breast cancer patients<sup>25</sup>. Gene expression analysis was also performed on primary breast tumors and matched metastatic lymph node tumors to improve our understanding of the metastatic process. Fifty-one genes were differentially expression ( $P < 1 \times 10^{-5}$ , >2-fold differences). Genes expressed at higher levels in the primary tumors were involved in degradation of the extracellular matrix, enabling cells with metastatic potential to disseminate, while genes expressed at higher levels in metastases were involved in transcription, signal transduction and immune response, providing cells with proliferation and survival advantages. These data improve our understanding of the biological processes involved in successful metastasis and provide new targets to arrest tumor cell dissemination and metastatic colonization<sup>21</sup>. Evaluation of primary tumors from patients with metastases and those without has also been performed; however, in this study we were unable to identify a molecular signature that could effectively classify the two tumor types<sup>26</sup>. The inability to derive molecular profiles of metastasis in primary tumors may reflect tumor heterogeneity, paucity of cells within the primary tumor with metastatic potential, influence of the microenvironment, or inherited host susceptibility to metastasis<sup>27</sup>.

Biomedical Informatics (BMIX) has been focused on: 1) development and implementation of an infrastructure system that supports the acquisition, storage, and maintenance of the clinical and molecular data generated in the study, and 2) application of existing algorithms and methods, and development of new ones as needed, for biomedical informatics research as well as for supporting other scientific research in the study. Our work has resulted in an advanced and recognized infrastructure system that supports translational research<sup>28, 29</sup>, a number of peer-reviewed publications and conference presentations, and competitive extramural funding. The following is a high-level summary of the activities and achievements in BMIX.

Quality Control (QC) and Quality Assurance (QA). To obtain high-quality data we have developed a set of QA programs, and many of the QC/QA procedures have been embedded into our SOPs. For example, the clinical QA program consists of: 1) a visual inspection of missing values and obvious inconsistencies, 2) double data-entry to reduce data entry errors, 3) a computer program named QAMetrics that deploys established QA rules to check for data integrity across the whole questionnaire, and 4) a web-based QA Issue Tracking system to enable efficient communications between WRI and the clinical sites for resolving QA problems of the clinical data<sup>30-33</sup>. In addition, we have developed a microarray QA program to identify outlier arrays from a population of similar arrays<sup>34, 35</sup>. Additional QA programs have been developed or are in development to ensure the quality of the data generated at WRI.

The Laboratory Information Management System (LIMS). A LIMS is needed to track and manage the sheer volume of data involved in translational research. For genomic and proteomic experimental data tracking, we have deployed the Synapseus system from GenoLogics deployed in 2008. Our clinical and tissue banking activities are tracked by the Clinical Laboratory Workflow System that we co-developed with Cimarron Software that was deployed in 2004<sup>36</sup>. To support the evolving needs of breast cancer studies, we are now in the process of developing a replacement system of the CLWS using up-to-date IT technologies.

The Data Warehouse for Translational Research (DW4TR). We have developed the DW4TR to integrate internally collected and generated clinicopathological data and molecular data from all the platforms. The development is done with our IT development partner, InforSense Ltd (currently part of IDBS). We have developed an innovative patient-centric, modularly-structured data model for clinicopathologic data, which contains disease-independent and disease-specific submodules to enable easy expansion to support the study of new diseases. A specimen-centric data model has been developed for molecular data. We have also developed a temporal data model to resolve the problem of the temporal information alignment between the data and specimens that were collected at multiple points of the time. The DW4TR has two major intuitive interfaces, one is the Aggregated Biomedical-information Browser and the other is the Individual Subject Information Viewer. The DW4TR has successfully supported the Clinical Breast Care Project with a total of 799 data elements, and has been expanded to support additional translational research programs<sup>37-40</sup>.

Other Applications: We have developed a Biomedical Informatics Portal, using the InforSense analytical workflow platform, to host additional applications deployed or developed by us to facilitate integrative biomedical research. For example, we have developed a specimen selection tool for scientists to use to select specimens in our tissue bank using the temporal and other information in the DW4TR<sup>41</sup>. The user can specify the inclusion/exclusion criteria. We have also developed a Data Correction Utility to enable correction of data errors in the DW4TR.

**Research:** Our research focuses on clinicopathologic risk factors of breast cancers, biomarker identification, and mechanistic understanding of primary and metastatic breast cancer, and we do so mostly by applying existing data mining methods<sup>42, 43</sup>. We also have developed new algorithms and visualization tools for clinicopathologic risk factor analysis, and reported interesting new breast disease co-occurrence patterns among other findings<sup>44-50</sup>. We found ethnicity-specific benign breast disease co-occurrence with breast cancer<sup>51, 52</sup>, reported different characteristics of invasive breast cancer between Caucasian and African American women<sup>53</sup>, and identified subtype-specific clinicopathologic properties of breast cancers<sup>54, 55</sup>. We worked closely with laboratory scientists in identifying biomarkers of breast cancer and breast cancer metastasis, and used bioinformatics tools to seek mechanistic understanding of breast cancer development and metastasis<sup>56-58</sup>, and we collaborated with external scientists as well in study human diseases<sup>59, 60</sup>. In performing a peripheral blood gene expression analysis, we have identified differential enrichment of signaling pathways between breast cancer patients and normal subjects<sup>61</sup>. In another project using breast cancer tissues we have identified significant difference between lymph node positive and lymph node negative cases in a carefully stratified population. We also performed a transcription factor-centric computational analysis of genes differentially expressed in healthy breast tissues from African American and Caucasian women<sup>62</sup>. We used the semantic similarity measures to characterize human metabolic and regulatory pathways which will help us to understand human disease development mechanisms<sup>63, 64</sup>.

## **9. PLAN:**

The rapid advances in basic research and data mining technologies fuels a growing demand for tissue banks. The development of new tools of molecular biology such as polymerase chain reaction (PCR), microarray and proteomic technologies, and laser capture microdissection (LCM) has made it possible to examine gene expression in very small tumor samples and at high throughput. Tissue banks allow these researchers to test their hypotheses rapidly and in a cost effective manner. By linking such molecular information to clinical data, we propose to translate knowledge from the laboratory to the clinic. These new technologies coupled with the information that can be generated with resources available in a tissue bank, will inspire researchers in the field of breast cancer to ask questions and develop hypotheses that 10 years ago were inconceivable.

**A. Subjects:** Tissue will be obtained from two general subject groups:

- 1) Consenting adult patients presenting to the Breast Center or the Women's Imaging Center of the WRNMMC with known breast cancer,
- 2) Consenting adult patients presenting to the Breast Center or the Women's Imaging Center of the WRNMMC with evidence of breast disease requiring clinical need for some form of tissue biopsy, and

Evidence of breast disease shall include potentially malignant breast lesions detected by mammography or ultrasound, palpable breast mass(es), abnormal breast discharge, abnormal physical breast morphology consistent with possible breast cancer, or axillary adenopathy without a known pre-existing condition. The far majority of patients will be female, however, 1% of all breast cancers occur in males and these male patients with breast cancer (“1” above), as well as males undergoing breast tissue biopsy (“2” above).

**Inclusion and Exclusion Criteria:**

For consideration of inclusion in this study:

**Inclusion:**

- 1) adult over the age of 18 years,
- 2) mentally competent and willing to provide informed consent,
- 3) military beneficiaries,
- 4) presenting to the Breast Center or the Women’s Imaging Center of the WRNMMC with evidence of possible breast disease.

**Exclusion:**

- 1) known history of Human Immunodeficiency Virus (HIV), or any type of hepatitis,
- 2) prion-mediated disease,
- 3) drug resistant tuberculosis or other infectious disease presenting a significant risk to personnel handling tissue or blood-derived products,
- 4) with pre-existing coagulopathies,
- 5) all other conditions, for whom invasive biopsy or surgery is medically contraindicated, shall be excluded.

During their clinic visit patients will be offered the opportunity to participate in our studies. If they choose to participate, before signing an informed consent, they will have the study explained to them by a CBCP research team member (e.g. purpose of study, where samples will be stored, the GINA law, how to withdraw from the study) and provided time to discuss any concerns and/or ask questions regarding the study and their participation. They will be offered the opportunity to consent for future research to be performed on their samples. A copy of the ICD/HIPAA will be provided to the subject prior to departing.

**B.Study Design:** The proposed study is designed as minimal risk, prospective acquisition and archiving of patient serum, leucocytes, and breast tissue removed in the course of diagnostic, therapeutic or elective treatment, but not otherwise required for diagnostic or therapeutic purposes. Portions of these tissues will then be analyzed for global expression

pattern screening analyses utilizing high-throughput genome and proteome research technologies, results of which will be analyzed using standard bioinformatics tools. Informed patient consent for archiving and later experimental use will be obtained prior to procedures involving tissue removal. The tissue will be handled and processed for archiving under the supervision of a licensed pathologist, and the remaining tissue will be coded and released for archiving only after all diagnostic tissue needs have been satisfied as determined by the supervising pathologist. Blood samples will be obtained (as described in section d). Each patient will be assigned a unique, 9-digit coded identifier (CBCP number) and all associated tissue, serum, and/or blood samples to be archived will be assigned an 8-digit barcode. Both the CBCP numbers and the corresponding sample barcodes will be registered into the clinical database. Under separate approved protocols (WRAMC DCI WU# 00-2006 and WU# 01-2001), a clinical database continues to be maintained which describes the characteristics of the blood and tissue donors. Linkage between the patient and research samples will be kept in a password protected-limited access database and in individual patient records held in gang-locked cabinets..

Tissues will be employed for molecular, biochemical and histological investigations that will be generally described to the prospective patients at the time of counseling for tissue bank consent, and as described in paragraph 8 above. Portions of these tissues will be analyzed for global expression pattern screening analyses utilizing high-throughput genome and proteome research technologies, results of which will be analyzed using standard bioinformatics tools (eg. GenoMax™). These molecular biologic investigations will be carried out at the CBCP's research facility, The Functional Genomics / Proteomics / Gene Sequencing Center located in Windber, PA. Along with the global genome and proteome expression profiling studies, other studies on tissue and blood/serum specimens will be defined as per the scientific (clinical and basic science) panel of personnel of the CBCP, which consists of personnel from all sites and which meets via Video-Teleconferencing not fewer than every two weeks. In addition, the CBCP Scientific Advisory Board, made up of outside (non-CBCP) physicians, scientists, and other professionals, will meet semi-annually, and will provide input as to scientific direction.

All specimens will be held in CBCP freezers for up to an indefinite period of time into the future. Exceptions to this will occur when specimens are exhausted (consumed) due to their use in the approved research within CBCP labs and elsewhere, or when previously consented patients withdraw their consent, at which time that patient's specimens will be withdrawn from the repository and destroyed. Specimens are carefully monitored for the amount of each sample remaining, and place a "No Further Use" block on the specimen when it reaches a minimum threshold remaining. The PI would make the final decision as to whether the last remaining sample would be utilized further for research. The CBCP PI is responsible for maintaining the integrity of the Tissue Bank, Serum Bank, and all databases. In the event that the named PI leaves WRNMMC, another WRNMMC investigator will be named (through in-place processes), and that new CBCP PI will take full responsibility for the banked samples and data. If under any circumstance a new CBCP PI is not named, then the WRNMMC

Institutional Review Board (IRB) will be notified, and the IRB will determine the disposition of this tissue and blood library.

## **1. Methodology:**

a. The approach to collection of the samples to be achieved is as follows, based on the grouping of the patients presenting (corresponding to the subjects groups in section 9a. above).

- 1) Patients already diagnosed with breast cancer: This group will consist of patients who already have undergone a breast biopsy of any type, at WRNMMC or another institution (after confirmation of the pathology diagnosis at that other institution), which has confirmed a cancer diagnosis that requires further surgical therapy as per the consensus recommendation of the multidisciplinary breast conference.
- 2) All consenting adult patients presenting to the Breast Center or the Women's Imaging Center at WRNMMC with evidence of breast disease for which a breast tissue biopsy (to include ductal lavage, open breast biopsy, tru-cut biopsy, image-directed biopsy) is clinically indicated. After consent is obtained, patients will be taken to procedure or surgery.
- 3) Either at the time of initial visit and consent to the protocol, during recovery from surgery or at follow up visit, a nurse will assist the patient in completing a questionnaire, and caffeine, fat intake, and stress level instruments.

## **2. Sample Collection and Processing:**

a. Up to 20cc of blood will be obtained via a peripheral venous access, prior to the administration of any anesthetic or fluids. Once the breast tissue is surgically removed, as clinically indicated, the specimen(s) will be taken to the pathology laboratory where a licensed pathologist will ensure that the tissue is adequate for routine pathology analyses (diagnosis, margin status assessment, and other indicated purposes).

b. Then and only then, if any actual excess tissue (cancerous or benign) remains, samples of that tissue will be harvested for archiving in the tissue bank. This archival tissue will be divided and preserved using several methods: the tissue will be OCT embedded in a cryomold and frozen on dry ice, the tissue will be placed in a vial and flash-frozen in liquid nitrogen, and/or the tissue will be placed in a tissue cassette for paraffin embedding.

c. The OCT-embedded and flash-frozen tissue samples will be labeled with an 8-digit coded identifier and placed into the CBCP Tissue Bank freezer at temperatures -80C pending shipment to WRI.

d. Blood samples will be fractionated into their components using a centrifuge. The blood components will be aliquotted into labeled cryovials and placed into the -190C freezer.

e. This tissue will remain in the freezer at the breast care center laboratory site for at least a two-week period of time, or longer if needed to fulfill the requirement set in section (ii) below. During this time, no analyses will be performed on the specimen; this period of time will be known as the “Fail-Safe” time period. The Fail-Safe time period is intended to allow the diagnostic pathologists the opportunity to withdraw banked tissue for any additional diagnostic testing they determine is necessary to patient care.

f. After the pathologist determines with final certainty, by the publishing of the official final pathologic report with no outstanding addenda, that there is no diagnostic pathologic requirement for the frozen specimen(s), then the archived specimen(s) on that patient (identified only by a unique 8-digit barcode) will be released to the CBCP for research analyses. This will include transfer of the tissues to offsite locations for specialized studies (to include functional genomics, proteomics analyses). Similarly research-collected FFPE tissue will be appropriately labeled and transferred to offsite facilities for immunohistochemical analysis and tissue microarray construction.

g. The serum and non-serum contents of the blood will, after appropriate labeling and removal of all patient identifiers, linked only to the patient via a unique coded identifier be transferred to WRI CBCP offsite research facility for proteomic analysis.

### 3. Uses of the tissue and serum specimens.

The known primary uses under this protocol for the acquired tissues and serums/blood fall into seven major subsections:

a. Tissue Banking – this includes sample definition and receiving, freezing/labeling/storage, OCT embedding (placing the tissues in a special preservative that protects the RNA/DNA during prolonged freezing)/labeling (putting identifier codes on each tissue sample for subsequent tracking)/storage, and inventory/tracking. The inventory and tracking of all samples will be done electronically with unique identifiers (8-digit barcodes)

using Clinical Laboratory Workflow System (CLWS), which tracks each specimen throughout its lifetime in the repository. .

b. Imaging/Microscopy – after sample definition, receiving, and fixation and fluorescence in situ hybridization (FISH) will be performed on tumor samples. Laser capture microdissection (LCM) will be performed by our LCM-trained CBCP pathologist (trained at NCI in laser capture microdissection; CBCP equipment by Arcturus Inc.) for RNA isolation from specimens and subsequent cDNA synthesis, array analysis, and in-situ RT-PCR. Image acquisition will be performed on digital microscopes and images archived in the CBCP server(s) and/or data warehouse (being developed with NCR Inc.), initially on CBCP sites at WRNMMC or WRI, and eventually at the CBCP Data Warehouse in Fort Detrick, MD (as part of our MANVT initiative, funded separately). The MANVT initiative is a Medical Area Network Virtual Technologies program, which provides near-real time interaction between all of the main CBCP sites (WRNMMC, Windber clinical and WRI) with regards to movement of high volumes of research data across a virtual fiberoptic network linking the main sites, as well as allow for creation of a data warehouse (situated at least in part at Fort Detrick, MD or to have a significant redundancy backup there).

c. Gene Expression Profiling – RNA will be extracted from tissues using various kits appropriate to the storage conditions of the tissues (flash frozen, OCT, FFPE, etc). RNA will be used for Northern Analysis, RT-PCR, and mRNA expression analysis using Affymetrix arrays and in future RNA Seq analysis.

d. Sequencing –Capillary electrophoresis and Sequence Analysis is performed on plasmids, usually to confirm the structure of various constructions or by direct sequencing of genomic DNA.

e. Genotyping – Tissues/blood will undergo DNA quantification followed by PCR set-up, thermal cycling, SNP (single nucleotide polymorphism) reaction clean-up, capillary electrophoresis set-up, genotype calling, and genotype QC. Genotyping is also performed using Affymetrix arrays.

f. Protein Expression Profiling – After sample clean-up, serums/blood/tissues will undergo 1D- and 2D-Electrophoresis, Gel staining, Image acquisition and processing and analysis, isolation of proteins from the gel (“Spot picking”), enzymatic digestion of the selected proteins (“Spot digestion”), Mass spectral analysis, Spectrum processing, and protein identification.

At the end of each of the above seven laboratory workflows, the data will be QA’d,

analyzed using powerful genomics/proteomics software tools, and placed into the CBCP database / data warehouse. QA of the data involves using software tools that interrogate the fields of the data that come out of the workflow stations, to ensure the data has consistency and is within expected or known ranges; any data found to be outside of expected ranges is not necessarily flawed, but is then identified for closer analysis by researchers.

4. Data Collection: The main objective of this tissue banking and analysis initiative is to collect tissue, along with biological and clinical data, prospectively from all informed and consenting patients enrolled in the tissue repository protocols of the CBCP. The patient data collected will include the following categories: demographics, socioeconomic, clinical, pathologic, radiologic, treatment, follow-up, and risk factors. It is emphasized that all tissue samples will be anonymized and all direct patient identifiers of the tissue and/or serum, as well as of the database itself, will be removed from them. All tissue and data will be identified in the research arenas by a “CBCP #”, which will be a unique, 9-digit, individual patient-specific number. This CBCP number will be linked to all patient items via barcoding of samples and questionnaires. Connection between the “CBCP #” and the patient identifiers are located in two places: in patient files, maintained by CBCP laboratory personnel, which will be kept in a gang-locked file cabinet, and in a password-protected, limited-access, secure database to which the Research Protocol Coordinator, pathology, and data team have access. There will be no way for researchers anywhere along the chain of tissue, serum, or data collection or analysis, to identify the actual identity of the patient via the barcodes or CBCP number. Other researchers within the CBCP or associated with the CBCP who have permission to use the blood or tissue samples may also have access to the clinical information (indeed, this linkage is what makes the tissue bank so powerful as a research tool); however, this linkage will not involve the patient’s name, so no researchers will ever know the identity of the specimen’s origin. Other identifiers, i.e., date of birth, date of death, date of surgery, etc. may be identified to researchers pending a separate IRB approved protocol.

The data collection instruments and protocol that will accompany the tissue of the diagnosed breast cancer patients (Section 9.d.1. above) and patients not diagnosed with breast cancer, but requiring clinically-directed tissue biopsies of any type (Section 9.d.2. above), or elective reduction mammoplasty (Section 9.d.3. above) have been previously submitted to the WRAMC DCI, and was IRB-approved in December 2000 (WU# 00-2006). Enclosed with this protocol are the data-collection sheets approved under that protocol, and which will be used for the associated tissue, should the patient independently consent for both IRB-approved protocols.

5. Sample Size / Data Analysis: No sample size justification required as this is an open-ended tissue repository protocol, and will not undergo specific data analysis per se. However, we do have an expected (estimated) number of patients per year that will be

enrolled at WRNMMC in the tissue protocols. We estimate that the number of patients newly diagnosed with breast cancer and treated at WRNMMC will average 1-300 per annum. The number of patients undergoing some or any form of a breast biopsy for any reason at our institution may average 1100 per annum.

Other institutions are intended to join in this protocol. These other sites will be both military and civilian. As with multi-center protocols, a Principal Investigator will be responsible at each separate institution for overseeing the protocol, conforming the protocol and consent form to their own institutional IRB requirements, and for achieving their own institution's IRB approval prior to proceeding.

6. Addendum: The oversight of the CBCP, by charter, consists of the following two groups:

1. Steering Committee – this committee provides internal review and oversight, meets quarterly on average and consists of members from WRNMMC, Windber Research Institute, and the Henry M. Jackson Foundation. . The Steering Committee is chartered with CBCP budgetary oversight, review, and approval.

2. Scientific Advisory Board – provides external peer review and scientific direction. Comprised of seven noted leaders in their respective fields of Biomedical Informatics, Pathology, Cancer Research, Breast Cancer Pathology, Tissue Banking, Breast Nursing/Education, and Breast Center Management. Board serves to provide scientific oversight of protocols, research studies and future vision of project. SAB meets three times annually.

The Steering Committee first met in February 2000, and has continued to meet and provide Project oversight and direction continuously. The SAB was formed and the initial stated goals of the Steering Committee regarding the five pillars have been implemented, the SAB helps determine the future direction of research that was not initially included as part of what the Steering Committee developed.

## **10.BIBLIOGRAPHY:**

1. SEER data; American Cancer Society Annual Report, 1998.
2. Zimmerman MR, Brier B, Wade RS. Brief communication: twentieth-century replication of an Egyptian mummy: implications for paleopathology. *Am J Phys Anthropol.* 1998 Dec; 107(4):417-20.
3. Reyman TA, Zimmerman MR, Lewin PK. Autopsy of an Egyptian mummy: Histopathologic investigation. *Can Med Assoc J.* 1977 Sep 3; 117(5):470-2. Pollack JR, Perou CM, Alizadeh AA, et al. Genome-wide analysis of DNA copy-number changes using cDNA

microarrays. *Nat Genet.* 1999 Sep;23(1):41-6.

4. Evans AS, Mueller NE. The past is prologue: use of serum banks in cancer research. *Cancer Res.* 1992 Oct 1;52(19 Suppl):5557s-5560s.
5. Kuykendall LH, Lory DS. Teamwork and leadership: the development of the National Psoriasis Tissue Bank. *Dermatol Nurs.* 1995 Oct; 7(5):298-301.
6. Bowcock AM. Genetic locus for psoriasis identified. *Ann Med.* 1995 Apr; 27(2):183-6.
7. Watson PH, Snell L, Parisien M. The NCIC-Manitoba Breast Tumor Bank: a resource for applied cancer research. *CMAJ.* 1996 Aug 1; 155(3): 281-3.
8. Pollack JR, Perou CM, Alizadeh AA, et al. Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nat Genet.* 1999 Sep;23(1):41-6.
9. Hendenfalk, I, Duggan, Chen, Y., D et al (2001). Gene expression profiles in hereditary breast cancer. *New England Journal of Medicine*, 344 (2001): 8; 539-548.
10. Alizadeh AA, Eizen MB, Davis RE, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature.* 2000 Feb 3; 403(6769): 503-11.
11. Bubendorf L, Kolmer M, Kononen J, et al. Hormone therapy failure in human prostate cancer: analysis by complementary DNA and tissue microarrays. *J Natl Cancer Inst.* 1999 Oct 20;91(20):1758-64.
12. Goncalves J, Brenton JD, Gaasterland T, Zanke B, Woodgett JR, Quin J, Schmitt K, Sorge J: Critical issues in microarray data processing and analysis of tumor tissues. Abstracts of the 2nd Annual Advances in Genome Biology & Technology Meeting, Marco Island, FL; 4 Feb 2001.
13. International Human Genome Sequencing Consortium: Initial sequencing and analysis of the Human Genome. *Nature* 409: 15 Feb 2001; 860 – 922.
14. Ellsworth RE et al (2010) Chromosomal alterations in pure nonneoplastic breast lesions: implications for breast cancer progression. *Ann.Surg.Oncol.* epub ahead of print.
15. Ellsworth RE et al (2007) Correlation of levels and patterns of genomic instability with histological grading of DCIS. *Annals Surg Onc* 14 (11):3070-3077.
16. Ellsworth RE et al (2008a) Correlation of levels and patterns of genomic instability with histological grading of invasive breast tumors. *Breast Cancer Res.Treat.* 107 (2):259-

265.

17. Ellsworth RE et al (2008b) Chromosomal alterations associated with the transition from in situ to invasive breast cancer. *Ann.Surg.Oncol.* 15 (9):2519-2525.
18. Ellsworth DL et al (2004c) Genomic patterns of allelic imbalance in disease free tissue adjacent to primary breast carcinomas. *Breast Cancer Res.Treat.* 88 (2):131-139.
19. Ellsworth DL et al (2004b) Outer breast quadrants demonstrate increased levels of genomic instability. *Annals Surg.Oncol.* 11 (9):861-868.
20. Ellsworth DL et al (2004a) Genomic instability in histologically normal breast tissues: implications for carcinogenesis. *Lancet Oncol.* 5 (12):753-758.
21. Ellsworth RE et al (2009) A gene expression signature that defines breast cancer metastases. *Clin.Exp.Metastasis* 26 (3):205-213.
22. Ellsworth RE et al (2008c) The Clinical Breast Care Project: an important resource in investigating environmental and genetic contributions to breast cancer in African American women. *Cell Tissue Bank* 9 (2):109-120.
23. Field LA et al (2011) Identification of differentially expressed genes in breast tumors from African American compared to Caucasian women. *Cancer under revision.*
24. Ellsworth RE et al (2005) Allelic imbalance in primary breast carcinomas and metastatic tumors of the axillary lymph nodes. *Mol.Cancer Res.* 3:71-77.
25. Becker TE et al (2008) The genomic heritage of lymph node metastases: implications for clinical management of patients with breast cancer. *Ann.Surg.Oncol.* 15 (4):1056-1063.
26. Ellsworth RE et al (2011) Differential gene expression in primary breast tumors associated with lymph node metastasis. *Int.J.Breast Cancer* epub.
27. Holliday C et al (2009) Genomic instability in the breast microenvironment? A critical evaluation of the evidence. *Expert Rev.Mol.Diagn.* 9 (7):667-678.
28. Hu H, Liebman M. Chapter 18. An introduction to biomedical informatics. In: Deng H, Shen H, Liu Y, Hu H, editors. *Current Topics in Human Genetics: Studies of Complex Diseases: World Scientific Publishing Co.; 2007.*
29. Hu H, Mural RJ, Liebman MN, editors. *Biomedical Informatics in Translational Research.* Boston, London: Artech House; 2008.

30. Zhang Y, Kvecher L, Sun W, et al. A quality assurance issue tracking tool to facilitate the enhancement of clinical data quality (submitted).
31. Sun W, Gutchell EM, Zhang Y, et al. Clinical data quality assurance and quality control measures in the Clinical Breast Care Project (in preparation).
32. Zhang Y, Sun W, Gutchell EM, et al. An issue tracking system to facilitate the enhancement of clinical data quality in the clinical breast care project. *AMIA Annu Symp Proc.* 2007:1173.
33. Sun W, Zhang Y, Brzeski H, et al. Quality assurance of a clinical breast disease database and preliminary data mining. *Advancing Practice, Instruction and Innovation through Informatics.* Lake Tahoe, CA; 2005.
34. Yang S, Guo X, Yang YC, et al. Detecting outlier microarray arrays by correlation and percentage of outliers spots. *Cancer Inform.* 2007;2:351-60.
35. Yang S, Hu H, Shriver C, et al. Identification of genes differentially expressed in blood of patients with breast abnormalities. *The 27th Annual San Antonio Breast Cancer Symposium.* San Antonio, TX; 2004.
36. Hu H, Kvecher L. Chapter 7. Data Tracking Systems. In: Hu H, Mural RJ, Liebman MN, editors. *Biomedical Informatics in Translational Research;* 2008. p. 111-36.
37. Hu H. Chapter 8. Data Centralization. In: Hu H, Mural RJ, Liebman MN, editors. *Biomedical Informatics in Translational Research;* 2008. p. 141-71.
38. Hu H, Brzeski H, Hutchins J, et al. Biomedical informatics: development of a comprehensive data warehouse for clinical and genomic breast cancer research. *Pharmacogenomics.* 2004 Oct;5(7):933-41.
39. Hu H, Correll M, Kvecher L, et al. DW4TR: a data warehouse for translational research (revision submitted).
40. Hu H, Correll M, Osmond M, et al. A clinical data warehouse to support translational research. *15th Annual International Conference on Intelligent Systems for Molecular Biology (ISMB).* Vienna, Austria; 2007.
41. Kvecher L, Wu W, Hooke J, Shriver C, Mural R, Hai H. An Approach to Correlate the Temporal Information to Facilitate Specimen Selection in the Breast Cancer Research Project. *Cancer Research.* 2009;69:763s.

42. Zhang Y, Correll M, Hu H. Chapter 9. Data Analysis. In: Hu H, Mural RJ, Liebman MN, editors. *Biomedical Informatics in Translational Research*; 2008. p. 175-202.
43. Correll M, Beulah S, Munro R, Sheldon J, Guo Y, Hu H. *Data Mining Tools and Techniques for Identification of Biomarkers for Cancer*. In: Chen J, Lonardi S, editors. *Biological Data Mining*: Chapman and Hall/CRC Press; 2009.
44. Maskery S, Hu H, Liebman MN, Hooke J, Shriver CD, Taioli E. Traditional breast cancer risk factors and common breast pathologies in post-menopausal women. The 30th San Antonio Breast Cancer Symposium. San Antonio, TX; 2007.
45. Maskery SM, Hu H, Hooke J, Shriver CD, Liebman MN. A Bayesian derived network of breast pathology co-occurrence. *J Biomed Inform*. 2008 Apr;41(2):242-50.
46. Maskery SM, Hu H, Mural RJ, Hooke J, Shriver CD, Liebman MN. Separate DCIS progression pathways derived from breast disease heterogeneity data. American Society of Clinical Oncology Annual Meeting. Atlanta, GA; 2006.
47. Maskery SM, Zhang Y, Jordan RM, et al. Co-occurrence analysis for discovery of novel breast cancer pathology patterns. *IEEE Trans Inf Technol Biomed*. 2006 Jul;10(3):497-503.
48. Maskery SM, Zhang YH, Hu H, Shriver CD, Hooke J, Liebman MN. Caffeine intake, race, and risk of invasive breast cancer, lessons learned from data mining a clinical database Proceedings of the 19th IEEE International Symposium on Computer-Based Medical Systems. Salt Lake City, UT; 2006.
49. Sun W, Zhang Y, Hooke J, et al. Relationship between educational level and breast cancer The 28th San Antonio Breast Cancer Symposium. San Antonio, TX; 2006.
50. Bekhash A, Maskery SM, Kvecher L, et al. Clinical Breast Care Project data warehouse as a research environment for breast cancer risk factor studies (submitted).
51. Bekhash A, Saini J, Fan X, et al. Differential benign Breast Disease Co-occurrence with Cancer in Caucasian and African American Women *Cancer Research*. 2009;69:667s.
52. Bekhash A, Saini J, Li X, et al. Ethnicity Difference of Benign Breast Diseases in Breast Cancer and Non-Cancer Patients. *Cancer Research*. 2010 Dec 8-12:288s-9s.
53. Hu H, Zhang Y, Kvecher L, et al. Different characteristics of invasive breast cancers between Caucasian and African American women. The 29th San Antonio Breast Cancer

Symposium. San Antonio, TX; 2006.

54. Saini J, Kovatich A, Bekhash A, et al. Association of Clinicopathologic Characteristics with IHC-Based Breast Cancer Subtypes. *Cancer Research*. 2009;69:635s.
55. Saini J, Kovatich A, Li X, et al. Association of Clinicopathologic Characteristics with IHC-based Breast Cancer Subtypes (in preparation).
56. Hu H, Somiari SB, Copper J, et al. Can circulating matrix metalloproteinases be predictors of breast cancer? – A neural network modeling study. In: Wang L, Chen K, Ong YS, editors. *Lecture Notes in Computer Science (LNCS)* Springer-Verlag; 2005.
57. Li X, Rapuri P, Melley J, et al. Comparative analysis of gene expression profiles in human breast cancer from microarray data using breast tissues and peripheral blood samples. *International Conference on Intelligent Systems for Molecular Biology (ISMB)*. Boston, MA; 2010.
58. Saini J, Li X, Kvecher L, et al. Differential Gene Expression Analysis among Post-menopausal Caucasian Invasive Breast Cancer, Benign and Normal Subjects. *Cancer Research*. 2010 Dec 8-12:245s.
59. Serretti A, Olgiati P, Liebman MN, et al. Clinical prediction of antidepressant response in mood disorders: linear multivariate vs neural network models. *Psychiatry Research* 2007(152):223-31.
60. Deng H-W, Shen H, Liu Y, Hu H. *Current Topics in Human Genetics: Studies of Complex Diseases*: World Scientific Publishing Co; 2007.
61. Li X, Hu H, Shriver C, Mural R. Microarray Data Analysis Using Peripheral Blood Samples Suggests Differential Enrichment of Signaling Pathways between Breast Cancer Patients and Normal Subjects. *Cancer Research*. 2009;69:654s.
62. Hu H, Stegmaier P, Field L, et al. A transcription factor-centric computational analysis of genes differentially expressed in healthy breast tissues from African American and Caucasian women. *The 30th San Antonio Breast Cancer Symposium*. San Antonio, TX; 2007.
63. Guo X, Liu R, Shriver CD, Hu H, Liebman MN. Assessing semantic similarity measures for the characterization of human regulatory pathways. *Bioinformatics*. 2006 Apr 15;22(8):967-73.
64. Guo X, Shriver CD, Hu H, Liebman MN. Analysis of metabolic and regulatory

pathways through Gene Ontology-derived semantic similarity measures. AMIA Annu Symp Proc. 2005:972.

**11. FACILITIES TO BE USED:** The facilities to be used in this research study include the following:

1. Walter Reed National Military Medical Center (WRNMMC):
  - a. The Comprehensive Breast Center (CBCP) located in Bldg 19, third floor, which contains a breast biopsy procedure room / recovery area, a fully-functional pathology laboratory, the tissue bank area with freezer, the separate patient blood-drawing room, and the clinical and counseling areas.
  - b. The Operating Rooms on the third floor located in Bldg 9 where cancer-directed surgeries are performed.
  - c. The Women's Imaging Center located in Bldg 19, third floor, where non-surgical breast biopsies will take place.
  - d. The Pathology Laboratory located in Bldg 9, basement, where some breast specimens are accessioned.
  
2. At Windber Research Institute (CBCP) / Windber Medical Center: The WRI Functional Genomics and Proteomics Facility and Tissue Banks, located in Windber, PA. The facility is a 6000sq ft highly integrated laboratory that is adequately equipped for tissue processing and archiving, functional genomics, proteomics, high-throughput gene and protein expression profiling, microarray construction, DNA sequencing, single nucleotide polymorphism analysis, micro satellite analysis, mass spectral analysis, laser capture microdissection, bioimaging and microscopy.

The address for the CBCP is:

Clinical Breast Care Project (CBCP)  
WRNMMC, Bldg 19, third floor  
8901 Wisconsin Avenue  
Bethesda, MD 20889-5600  
Phone: 301-400-1477

**12. TIME REQUIRED TO COMPLETE:**

Anticipated start – November 2011  
Expected completion - Indefinite

**13. FUNDING IMPLICATIONS:**

Funding or other resources will be from outside sources: Yes X No

This grant is administratively managed by Lee Bronfman of the Henry M. Jackson Foundation, phone 301-295-3890. No funds will be transferred into WRNMMC.

**Budget request from DCI:**

**FY 2001   FY 2002   TOTAL**

**Equipment:**

Consumable Supplies (Itemize):

**Other -**

**Travel -**

**Reprints -**

**TOTAL   \$ \_\_\_\_\_   \$ \_\_\_\_\_   \$ NONE**

**14. NUMBER OF PATIENTS/CHARTS/SPECIMENS TO BE STUDIED:** We have an expected (estimated) number of patients per year that will be enrolled at WRNMMC in the tissue protocol. We estimate that the number of patients newly diagnosed with breast cancer and treated at WRNMMC will average 1-300 per annum. The number of patients undergoing some or any form of a breast biopsy for any reason at our institution may average 1100 per annum.

**15. TYPE OF SUBJECT POPULATION TO BE STUDIED:** Military health care beneficiaries over the age of 18 years presenting with the diagnosis of breast cancer, or any radiologic or clinical breast lesion requiring biopsy or tissue diagnosis.

**16. ENVIRONMENTAL IMPACT STATEMENT: For Questions contact:  
Laura Broussard, Environmental Program Office at 301-295-6436.**

a. Does any part of this protocol generate hazardous chemical waste as defined by Title 40 of the Code of Federal Regulation **YES ( ) NO (X)**? If yes, at what stages and how much?

b. Have you considered any alternative procedures **YES ( ) NO ( ) N/A (X)**? If yes, why were they rejected?

## **17. INVESTIGATOR AGREEMENT**

By submitting this protocol and providing an electronic signature in IRBNet, or an ink signature below, I agree to the following statements:

**General Assurance:** I agree to conduct the study as outlined herein. I certify that all procedures involving human subjects have been described in full.

**Starting the Study:** I understand that I cannot begin the study until I have received an approval letter documenting approval by the WRNMMC IRB.

**Consent:** I am responsible for assuring the quality of each subject's consent in accordance with current federal regulations. This includes ensuring that any "designee" that obtains consent on my behalf is completely familiar with the protocol and is qualified to perform this responsibility.

**Adverse Events:** I understand that I must report serious adverse events within two days to the IRB and provide this information to the Medical / Research Monitor. I will report unexpected (but not serious) adverse events that may possibly be related to participation in the protocol within 10 working days to the IRB using the same procedure.

**Training:** I verify that the personnel performing these procedures described in this protocol are technically competent, have been properly trained, and are appropriately qualified.

**Compensation:** I am aware that members of the research team are not authorized to accept any form of personal compensation for our efforts in conducting this research.

**Modifications:** I am aware that all changes to the protocol must be approved by the IRB before implementation. Examples of changes to protocols that require IRB approval include change of on-site PI, addition of personnel on study, increased sample size, addition of other data points, sources of outside funding, and addition of data collection sites.

**Deviations to the Protocol:** I am aware that any protocol deviations discovered by either the PI or auditing official will be immediately reported to the IRB. All corrective actions will be documented and become a part of the master study file, along with the report.

**Duplication of Effort:** I have made a reasonable good faith effort to ensure that this protocol is not an unnecessary duplication of previous experiments.

**Reports:** I agree to provide a Continuing Review Progress Report 30 days prior to the anniversary of the protocol's initial approval or as stipulated by the IRB. I agree to submit a final report within 30 days following completion or termination of the study.

**Maintain Study Files:** I agree to maintain a Study File that must be kept for three years from the date the study is closed (32 CFR 219.115(b). If IND medication or IDE appliances are used, the file must be kept for 2 years after Food and Drug Administration (FDA) approval and can then be destroyed; or if no application is filed or approved, until 2 years after the study is discontinued and FDA notified (21CFR 312.62(c). I acknowledge that research data is the property of the Command and will not be removed without prior approval. When I am scheduled to PCS or ETS, study records will be given to a new PI or the Department Chief, or turned over to the Department of Research Programs.

This file may be inspected at any time by Department of Research Programs, the FDA, and/or other applicable regulatory agencies responsible for the oversight of research. This file will include:

- A. The approved protocol and applicable amendments.
- B. The IRB minutes granting approval to initiate the study.
- C. IRB approval letter.
- D. Each Consent Form/HIPAA Authorization signed by the subject or surrogate.
- E. Continuing Review Progress Reports.
- F. Reports of adverse effects.
- G. Reports of any significant new findings found during the course of the study.
- H. All study documents generated from study date.
- I. Publications, abstracts, reprints resulting from study data.
- J. All information pertaining to an investigational drug or device.

**Publications:** I am aware that advertisements, abstracts, presentations or publications resulting from research protocols must have their products cleared by the Public Affairs Office, undergo Operation Security (OPSEC) review, undergo review for release of actionable medical information, and Publication Clearance.

**Applicable Regulations:** I am familiar with applicable regulations governing research, and will adhere to all of the requirements outlined in the DoD Multiple Project Assurance for the North Atlantic Regional Medical Command.

I understand that if I fail to comply with any of these responsibilities, all projects for which I am an investigator may be suspended.

[Provide electronic signature in IRBNet; if not available, provide ink signature with signature block.](#)

PRINCIPAL INVESTIGATOR  
Rank, Name, Corps  
Title  
Department

WRNMMC LEAD INVESTIGATOR  
Rank, Name, Corps  
Title  
Department

Requirements of U.S. Army Medical Research and Materiel Command's Human Subjects Research Review Board for secondary review:

1. Unanticipated problems involving risk to volunteers or others should be promptly reported by phone (301-619-2165), by email (hsrrb@amedd.army.mil), or by facsimile (301-619-7803) to the U.S. Army Medical Research and Materiel Command's Human Subjects Research Review Board. A complete written report should follow the initial notification. In addition to the methods above, the complete report can be sent to the U.S. Army Medical Research and Materiel Command, ATTN: MCMR-ZB-P, 504 Scott Street, Fort Detrick, MD 21702-5012.
2. The HSRRB must be notified of any deviation to the protocol that may have an affect on the safety of the subject and the integrity of the study. This notification should occur as soon as the deviation is identified. In addition, all deviations must be reported in the continuing review report (Annual Progress Report) and final study report.
3. Amendments to this protocol will be submitted to the WRNMMC Department of Research Programs. Upon approval, the amendment will be submitted to the HSRRB. This amendment will receive either full HSRRB review or expedited review by the Chair, as appropriate. The amendment may only be implemented after HSRRB approval has been granted.
4. A copy of the approved continuing review report (Annual Progress Report) and final study report will be submitted to the HSRRB for acceptance.

## **19. APPENDICES**

**APPENDIX A** – Study Workflows uploaded in IRBNet

**APPENDIX B** - Questionnaires uploaded in IRBNet

**APPENDIX C** – Included with original submission

**APPENDIX D** - Pathology Impact Statement uploaded in IRBNet

**APPENDIX E** – Signed Conflict of Interest Statement uploaded in IRBNet

**APPENDIX F** – MTA uploaded in IRBNet

**APPENDIX G** – NA (Advertisement Brochure/Flyer)

**APPENDIX H** - Consent Form uploaded in IRBNet

**APPENDIX I** - HIPAA uploaded in IRBNet



Walter Reed National Military Medical Center, Bethesda

IRBNet# 20704

IRB Approval Date: 17 May 2012

IRB Expiration Date: 16 May 2013

Do Not Sign Past the Expiration Date

Last Name \_\_\_\_\_

## WALTER REED NATIONAL MILITARY MEDICAL CENTER (WRNMMC)

**This Specimen banking consent form is valid only if it contains the IRB stamped date.**

### **Consent for Voluntary Participation in a Specimen Banking Study Entitled:**

“Tissue and Blood Library Establishment for Molecular, Biochemical and Histologic Study of Breast Disease”

**Principal Investigator:** Craig D. Shriver, COL, MC, General Surgery Service, Department of Surgery

**Study Sites:**  WRAMC

### **1. INTRODUCTION OF THE STUDY**

You are being asked to be in this research study because you are scheduled to have part or all of your breast tissue surgically removed as part of your clinical care. Your participation is voluntary. Refusal to participate will not result in any penalty or loss of benefits to which you are otherwise entitled.

Please read the information below, and ask questions about anything you do not understand, before deciding whether to take part in the study.

### **2. PURPOSE OF THE STUDY**

The purpose of the study is to gather samples of breast tissue, lymph node, and blood for use in future research studies. As part of the normal procedure, small amounts of breast tissue and lymph node are usually left over after all the clinical testing has been done. We would like to keep some of this left over tissue, and take some blood from you. By carefully examining breast cells from many different breast conditions, we hope to improve our understanding of the causes and best treatments of these medical problems. All normal hospital regulations will be followed in sending the tissue removed from your body, to pathology for diagnosis.



### **3. PROCEDURES TO BE FOLLOWED**

All necessary diagnostic tests will be performed by the pathologist on your surgically removed breast tissue prior to the release of any excess tissue by the pathologist. Then, if you agree to be in this study, a small amount of excess breast tissue will be obtained from the pathologist, only tissue that would otherwise be thrown away will be used for this study. If your surgical procedure also involves the removal of any lymph node(s), then a small amount of lymph node will also be obtained from the pathologist. We will also take a small amount of blood from you. Some tube(s) of blood (up to 20cc or about 4 teaspoons) will be drawn either before surgery or drawn separately at the time of your needle breast biopsy.

Participating in this study will not delay the processing of the pathology report and will not add to the time of your surgery. No additional surgery will be done on you to obtain these samples and only tissue that would otherwise be thrown away will be used for this study. In other words, all necessary diagnostic tests will be done first, prior to release of any excess tissue by the pathologist. If your breast tissue is being removed for therapeutic reasons and you consent to participate in this study, it is important for us to study the normal breast tissue because it allows our researchers to learn what non-diseased, normal breast tissue looks like at the genetic and protein levels, which we are analyzing in this study.

You will be asked to fill out a questionnaire at the time of the first visit. There will be some basic information collected about your breast, health, and medical history.

### **4. IDENTIFICATION OF YOUR BLOOD AND/OR TISSUE SAMPLES, HOW AND WHERE THEY WILL BE STORED AND WHO WILL HAVE ACCESS TO YOUR SAMPLES**

You will be assigned a code number referred to as the Clinical Breast Care Program (CBCP) number [not your name or social security number (SSN)], that will be used for the blood and tissue samples, the questionnaire and any clinical information. The only connection between your "CBCP number" and your name or SSN will be kept in double-locked secure files and via a password protected secure database.



Walter Reed National Military Medical Center, Bethesda

IRBNet# 20704

IRB Approval Date: 17 May 2012

IRB Expiration Date: 16 May 2013

*Do Not Sign Past the Expiration Date*

The blood samples will be frozen and stored indefinitely in either the Breast Care Center laboratory sites at Walter Reed or our off-site facility, Windber Research Institute (WRI), Windber, PA. All samples will be kept in a secured (locked) freezer and identified only by codes for the CBCP. Only researchers within the CBCP or associated with the CBCP may have access to the samples and related clinical information. The tissue and blood samples will undergo the primary research studies as described in the next paragraphs.

This breast tissue can be used for many types of laboratory research looking at cell changes during breast cancer development, identification of risk factors that lead to breast cancer, learning why and how breast cancer spreads to lymph nodes, and breast tissue biology.

The primary research uses of the blood and tissue samples are to study the genetic makeup, to examine protein changes, and to look for other markers that may be associated with breast disease. This information will be analyzed and stored in a computer database using your assigned code number.

Any remaining blood and tissue samples left over after these studies are done will remain frozen indefinitely until needed for approved research projects, if you consent for future unknown breast disease research. You will be provided the opportunity at the end of this consent form to indicate whether you will allow your blood and tissue samples to be used for future breast research. It is not possible at this time to predict all the potential future research uses of the samples.

If other research is to be conducted using these samples, the research will only be conducted after the proposed study protocol has received approval from the local Institutional Research Board (IRB).

Because these research projects are experimental, the results of any research done with your tissue will not be given to you or your doctor. These reports will not be put in your health records. The research will not be used in decisions regarding your medical care. Third parties, such as relatives, physicians, and insurance companies, will not have access to your information from these studies.

## **5. AMOUNT OF TIME FOR YOU TO COMPLETE THIS STUDY**



Your participation in this study will consist of the time necessary to read and understand this consent form, and to complete the questionnaires. The collection of all of the samples to be used in this study will be done at the time of your surgery. The blood draw will occur prior to your IV placement for the procedure, or if no IV is to be placed, then the collection will be done separately and take about 5 minutes.

Your samples will be stored until they are no longer needed for research, or until you request removal of your samples from our repository where the samples are stored. You may withdraw from this study at any time, and may request that any remaining samples be destroyed. The PI or the Research Protocol Coordinator will be the point of contact, if you choose to withdraw or if destruction of the samples becomes necessary. Removal of your samples from the bank must be requested in writing to the researcher at the address at the end of this consent form.

## **6. NUMBER OF PEOPLE THAT WILL TAKE PART IN THIS STUDY**

This study is called a multi-site study because participants from several hospitals will be in the study. There will be up to 1200 people taking part in this study each year.

## **7. POSSIBLE RISKS OR DISCOMFORTS FROM BEING IN THIS STUDY**

There may be some discomfort from drawing blood, and you may experience some pain, swelling and a bruise at the site of the needle stick. Some people feel dizzy or light-headed for a few minutes after blood is drawn.

These include employment problems, difficulty obtaining life or health insurance, mental stress for you or your family members, and other social discomforts and inconveniences. These risks can impact the mental state of both you and your family members, which could require counseling.

## **8. POSSIBLE BENEFITS FROM BEING IN THIS STUDY**



Walter Reed National Military Medical Center, Bethesda

IRBNet# 20704

IRB Approval Date: 17 May 2012

IRB Expiration Date: 16 May 2013

*Do Not Sign Past the Expiration Date*

You will not benefit from being in this study, but the possible benefits of future research from your blood/and or tissue include learning and understanding more about what causes breast cancer, and how to prevent and treat it.

## **9. CONFIDENTIALITY/PRIVACY OF YOUR IDENTITY AND YOUR RESEARCH RECORDS**

The principal investigator will keep your research records in a secure and locked place. These records may be looked at by staff from the WRNMMC Department of Research Programs, the Institutional Review Board (IRB), the representatives of the US Army Medical Research and Materiel Command, and other government agencies as part of their duties. These duties include making sure that the research participants are protected. Confidentiality of your records will be protected to the extent possible under existing regulations and laws but cannot be guaranteed. Complete confidentiality cannot be promised, particularly for military personnel, because information bearing on your health may be required to be reported to appropriate medical or command authorities. Your name will not appear in any published paper or presentation related to this study.

This research study meets the confidentiality requirements of the Health Insurance Portability and Accountability Act (HIPAA).

## **10. CONDITIONS UNDER WHICH YOUR PARTICIPATION IN THIS STUDY MAY BE STOPPED WITHOUT YOUR CONSENT**

Your taking part in this study may be stopped without your consent if remaining in the study might be dangerous or harmful to you. Your taking part in this study may also be stopped without your consent if the military mission requires it, or if you lose your right to receive medical care at a military hospital.

## **11. ELIGIBILITY AND PAYMENT FOR BEING IN THIS STUDY**

You will not receive any payment for being in this study.

## **12. COMPENSATION TO YOU IF INJURED AND LIMITS TO YOUR MEDICAL CARE**

Should you be injured as a direct result of your participation in this study, you will be provided medical care for that injury at no cost to you. You will not



Walter Reed National Military Medical Center, Bethesda

IRBNet# 20704

IRB Approval Date: 17 May 2012

IRB Expiration Date: 16 May 2013

*Do Not Sign Past the Expiration Date*

receive any compensation (payment) for injury. You should also understand that this is not a waiver or release of your legal rights.

Medical care is limited to the care normally allowed for Department of Defense (DOD) health care beneficiaries (patients eligible for care at military hospitals and clinics). Necessary medical care does not include in-home care or nursing home care. You should discuss this issue thoroughly with the PI before you enroll in this study.

If at any time you believe you have a study-related injury or illness as a result of participating in this research project, you should contact the PI. For questions about your rights as a research participant, call the Walter Reed National Military Medical Center Institutional Review Board (a group of people who review the research to protect your rights) at 301-295-2275, or the Staff Judge Advocate (SIA) Office at 301-295-2215.

### **13. COSTS THAT MAY RESULT FROM TAKING PART IN THIS STUDY**

There is no charge to you for taking part in this study.

### **14. IF YOU DECIDE TO STOP TAKING PART IN THIS STUDY AND THE INSTRUCTIONS FOR STOPPING EARLY**

You have the right to withdraw from this study at any time. This can be done by contacting the PI, or the research protocol coordinator, at the phone number in section 19 of this consent form. You may request that your sample be destroyed or that your identification be removed without destroying the sample. This request needs to be in writing to the address at the end of this consent form; by leaving the study at any time, you in no way risk losing your right to medical care.

### **15. RESEARCH RESULTS**

The results from tests that may be done on the (blood and/or tissue) that you donate for future research will not be given to you or to your doctor, even if you ask that this be done and the results will not become part of your hospital medical record because these tests will be done for research purposes only.



Your individual testing results that may be done on the blood and/or tissue that you donate for future research will not be released to any third party, including family members, personal physicians, insurers or employers, under any circumstance unless required by law.

You should understand that there is a chance that the samples you are providing in this study may be used in other research studies related to breast disease. You may not be given any notice of future use of your sample. The confidentiality of your specimens will be protected in the same fashion as stated previously in the Description of This Study and/or Confidentiality section. You will not be personally identified in any published paper of these other research studies. Any other researcher using these samples will be under the direction of the CBCP, no direct identifiers such as your name, address, social security number (SSN), etc. will be provided to these researchers, but they could be provided with some identifying clinical information about you, such as date of surgery, diagnosis, treatment.

## **16. YOUR RIGHTS IF YOU TAKE PART IN THIS STUDY**

Participating in this study is your choice. You may choose either to participate or not to participate in the study. If you choose to participate in this study, you have the right to withdraw from this study at any time. This can be done by contacting the PI, or the research protocol coordinator, at the phone number in section 19 of this consent form. You may request that your sample be destroyed or that your identification be removed without destroying the sample. This request needs to be in writing to the address provided at the end of this consent form.

No matter what decision you make, there will be no penalty to you and you will not lose any of your regular benefits or risk losing your right to medical care.

## **17. THE NEW FEDERAL LAW – “GENETIC INFORMATION NON-DISCRIMINATION ACT” (GINA)**

This new Federal law generally makes it illegal for health insurance companies, group health plans, and most employers to discriminate against you based on your genetic information. This law generally will protect you in the following ways:



Health insurance companies and group health plans may not request your genetic information that we get from this research.

Health insurance companies and group health plans may not use your genetic information when making decisions regarding your eligibility or premiums.

Employers with 15 or more employees may not use your genetic information that we get from this research when making a decision to hire, promote, or fire you or when setting the terms of your employment.

All health insurance companies and group health plans must follow this law, dated 21 May 2010.

Be aware that this new Federal law does not protect you against genetic discrimination by companies that sell life insurance, disability insurance, or long-term care insurance."

## **18. AUTHORIZATION FOR RESEARCH USE OF PROTECTED HEALTH INFORMATION**

The Federal Health Insurance Portability and Accountability Act (**HIPAA**) includes a Privacy Rule that gives special safeguards to Protected Health Information (**PHI**) that is identifiable, in other words, can be directly linked to you (for example, by your name, social security number, etc.). We are required to advise you on how your PHI will be used.

### **(1) What information will be collected?**

For this research study, you will be asked some basic information about your breast, health, and medical history. The questionnaire will also involve information on caffeine intake, dietary fat intake, and lifestyle stress.

### **(2) Who may use your PHI within the Military Healthcare System?**

The members of the research team will have access to your health information in order to find out if you qualify to participate in this study, and for analyzing research data. Additionally, your PHI may be made available to health oversight groups such as the WRNMMC Department of Research Programs and Institutional Review Board, and other government agencies as part of their duties. These duties include making sure that research subjects are protected.



**(3) What persons outside of the Military Healthcare System who are under the HIPAA requirements will receive your PHI?**

Information outside the Military Healthcare System such as dates of surgery, diagnosis, treatment, etc. related to the CBCP code assigned to your research file may be shared with the research team for the purpose of performing the IRB approved research study. Your name, address and contact information will not be shared.

**(4) What is the purpose for using or disclosing your PHI?**

The members of the research team need to use your PHI in order to analyze the information related to the specific kind research study being performed. Researchers associated with the CBCP will use your blood and/or tissue samples to study the genetic makeup, protein changes, and other research markers that may be associated with breast disease, and unknown future research. The clinical information is important for performing analysis.

**(5) How long will the researchers keep your PHI?**

There is no expiration date for this study to maintain your PHI, since this study collects samples for placement within a tissue bank.

**(6) Can you review your own research information?**

You will not be able to look at your research information.

**(7) Can you cancel this Authorization?**

Yes. If you cancel this Authorization, however, you will no longer be included in the research study. The information and samples we collected from you may be destroyed at your request. The research and samples that may have been used prior to your request will not be able to be destroyed or withdrawn because they will already be among the statistics of the study. If you wish to cancel your Authorization, please contact the Principal Investigator or the Research Protocol Coordinator in writing.

**(8) What will happen if you decide not to grant this Authorization?**

If you decide not to grant this Authorization, you will not be included in this research study. Refusal to grant this Authorization will not result in any loss of medical benefits to which you are otherwise entitled.

**(9) Can your PHI be disclosed to parties not included in this Authorization who are not under the HIPAA requirements?**



There is a potential that your research information will be shared with another party not listed in this Authorization in order to meet legal or regulatory requirements. Examples of persons who may access your PHI include representatives of the Food and Drug Administration, the Department of Health and Human Services (DHHS) Office for Human Research Protections (OHRP), and the DHHS Office for Civil Rights. This disclosure is unlikely to occur, but in that case, your health information would no longer be protected by the HIPAA Privacy Rule.

**(10) Who should you contact if you have any complaints?**

If you believe your privacy rights have been violated, you may send a written complaint to the WRNMMC Privacy Officer, located at 8901 Wisconsin Avenue, Bethesda, MD 20889-5600, telephone: 301-319-4775.

Your signature at the end of this document acknowledges that you authorize WRAMC personnel to use and disclose your Protected Health Information (PHI) collected about you for research purposes as described above.

**19. CONTACTS FOR QUESTIONS ABOUT THE STUDY**

If you have questions about the study, you should contact either the PI or the Research Protocol Coordinator at WRNMMC 301-319-2385.

**20. OTHER FUTURE RESEARCH FOR WHICH YOUR BLOOD AND/OR TISSUE SAMPLES TAKEN DURING THIS STUDY COULD BE USED**

Please read carefully each sentence below and think about your choices. After reading each sentence, **circle** "yes" or "no", include the date and your initials. If you have any questions please talk to your doctor.



Walter Reed National Military Medical Center, Bethesda

IRBNet# 20704

IRB Approval Date: 17 May 2012

IRB Expiration Date: 16 May 2013

*Do Not Sign Past the Expiration Date*

**BY SIGNING THIS FORM, YOU ARE AGREEING THAT:**

a. My tissue and serum (blood) samples may be kept for use in future research to learn about, prevent, detect, or treat breast cancer, and to learn about whether certain genes relate to breast cancer:

YES NO Participant's Initials \_\_\_\_\_ Date \_\_\_\_\_

b. My clinical data/images, mammogram, breast ultrasound may be used or reviewed for future research:

YES NO Participant's Initials \_\_\_\_\_ Date \_\_\_\_\_

c. The principal investigator (or someone he or she chooses) may contact me in the future to ask me to take part in future research:

YES NO Participant's Initials \_\_\_\_\_ Date \_\_\_\_\_



Walter Reed National Military Medical Center, Bethesda

IRBNet# 20704

IRB Approval Date: 17 May 2012

IRB Expiration Date: 16 May 2013

Do Not Sign Past the Expiration Date

**SIGNATURE OF RESEARCH SUBJECT**

I have read (or someone has read to me) the information in this consent form. I have been given a chance to ask questions and all of my questions have been answered to my satisfaction.

**BY SIGNING THIS CONSENT FORM, YOU FREELY AGREE TO TAKE PART IN THE RESEARCH IT DESCRIBES.**

\_\_\_\_\_  
Participant's Signature

\_\_\_\_\_  
Date

\_\_\_\_\_  
**Printed** First name

\_\_\_\_\_  
MI

\_\_\_\_\_  
**Printed** Last Name]

\_\_\_\_\_  
SSN

\_\_\_\_\_  
Sponsor's Last 4

\_\_\_\_\_  
Date of Birth

\_\_\_\_\_  
Street or mailing address

\_\_\_\_\_  
City

\_\_\_\_\_  
State

\_\_\_\_\_  
Zip Code

**A copy of this consent form will be provided to you.**

**SIGNATURE OF INVESTIGATOR**

I have explained the research to the volunteer and answered all of his/her questions. I believe that the volunteer subject understands the information described in this document and freely consents to participate.

\_\_\_\_\_  
Investigator's Signature

\_\_\_\_\_  
Date (must be the same as the participant's)

\_\_\_\_\_  
Investigator's Printed Name

Clinical Breast Care Project  
Core Questionnaire

This cover page will be removed and destroyed.  
**Do not attach a bar code to this page.**

Patient Name: \_\_\_\_\_

# Clinical Breast Care Project

## Core Questionnaire

This questionnaire is to be filled out by the nurse or other trained healthcare professional in the presence of the protocol-consented patient. Take as much time as the patient needs to understand the questions presented to them. All questions are asked on the initial questionnaire.

1. CBCP#: \_\_\_\_\_
2. Date: \_\_\_\_\_ / \_\_\_\_\_ / \_\_\_\_\_ (mm/dd/yyyy)
3. Date of birth: \_\_\_\_\_ / \_\_\_\_\_ / \_\_\_\_\_ (mm/dd/yyyy)
4. Age \_\_\_\_\_
5. Sex:
  - Female
  - Male
6. Ethnic background:
  - African American
  - African
  - Caribbean
  - Native American
  - Asian
  - Asian American
  - Hispanic (white)
  - Hispanic (other)
  - White
  - Other \_\_\_\_\_
7. Marital status:
  - Single
  - Married
  - Divorced
  - Widowed
  - Separated
8. Education:
  - No high school
  - Some high school
  - High school graduate
  - Some College
  - Associate Degree
  - Bachelor Degree
  - Advanced Degree
9. Occupation: \_\_\_\_\_
10. Size of family:
  - # of brothers \_\_\_\_\_
  - # of half brothers \_\_\_\_\_
  - # of sons \_\_\_\_\_
  - # of sisters \_\_\_\_\_
  - # of half sisters \_\_\_\_\_
  - # of daughters \_\_\_\_\_
11. Do you have any **primary** blood family members (mother, sister, half sister, daughter, father, brother, half brother, son) with a history of breast cancer?
  - Yes
  - No (go to #15)
  - Unknown (go to #15)
12. If yes, # of known **female** primary family members with a history of breast cancer? \_\_\_\_\_
13. If yes, # of known **male** primary family members with a history of breast cancer? \_\_\_\_\_
- 14.1. Does your **mother** have a history of breast cancer?
  - a. History
    - Yes
    - No
    - Unknown
  - b. Menopausal status at diagnosis
    - Pre-menopausal
    - Post-menopausal
    - Unknown
- 14.2. Does your **sister 1** have a history of breast cancer?
  - a. History
    - Yes
    - No
    - Unknown
  - b. Menopausal status at diagnosis
    - Pre-menopausal
    - Post-menopausal
    - Unknown

CBCP#

Today's Date:

- 14.3. Does your **sister 2** have a history of breast cancer?  
a. History  
 Yes  
 No  
 Unknown  
b. Menopausal status at diagnosis  
 Pre-menopausal  
 Post-menopausal  
 Unknown
- 14.4. Does your **sister 3** have a history of breast cancer?  
a. History  
 Yes  
 No  
 Unknown  
b. Menopausal status at diagnosis  
 Pre-menopausal  
 Post-menopausal  
 Unknown
- 14.5. Does your **half sister 1** have a history of breast cancer?  
a. History  
 Yes  
 No  
 Unknown  
b. Menopausal status at diagnosis  
 Pre-menopausal  
 Post-menopausal  
 Unknown  
c. Relationship  
 Maternal  
 Paternal
- 14.6. Does your **half sister 2** have a history of breast cancer?  
a. History  
 Yes  
 No  
 Unknown  
b. Menopausal status at diagnosis  
 Pre-menopausal  
 Post-menopausal  
 Unknown  
c. Relationship  
 Maternal  
 Paternal
- 14.7. Does your **daughter 1** have a history of breast cancer?  
a. History  
 Yes  
 No  
 Unknown  
b. Menopausal status at diagnosis  
 Pre-menopausal  
 Post-menopausal  
 Unknown
- 14.8. Does your **daughter 2** have a history of breast cancer?  
a. History  
 Yes  
 No  
 Unknown  
b. Menopausal status at diagnosis  
 Pre-menopausal  
 Post-menopausal  
 Unknown
- 14.9. Does your **father** have a history of breast cancer?  
a. History  
 Yes  
 No  
 Unknown
- 14.10. Does your **brother** have a history of breast cancer?  
a. History  
 Yes  
 No  
 Unknown
- 14.11. Does your **half brother** have a history of breast cancer?  
a. History  
 Yes  
 No  
 Unknown  
b. Relationship  
 Maternal  
 Paternal
- 14.12. Does your **son** have a history of breast cancer?  
a. History  
 Yes  
 No  
 Unknown

CBCP#

Today's Date:

15. Do you have any **secondary** blood family members (aunt, cousin, niece, grandmother, great grandmother, granddaughter, uncle, nephew, grandfather, great grandfather, grandson) with a history of breast cancer?  
 Yes  
 No (go to #19)  
 Unknown (go to #19)
16. If yes, # of known **female** secondary family members with a history of breast cancer? \_\_\_\_\_
17. If yes, # of known **male** secondary family members with a history of breast cancer? \_\_\_\_\_
- 18.1. Does your **aunt 1** have a history of breast cancer?  
a. History  
 Yes  
 No  
 Unknown  
b. Menopausal status at diagnosis  
 Pre-menopausal  
 Post-menopausal  
 Unknown  
c. Relationship  
 Maternal  
 Paternal
- 18.2. Does your **aunt 2** have a history of breast cancer?  
a. History  
 Yes  
 No  
 Unknown  
b. Menopausal status at diagnosis  
 Pre-menopausal  
 Post-menopausal  
 Unknown  
c. Relationship  
 Maternal  
 Paternal
- 18.3. Does your **aunt 3** have a history of breast cancer?  
a. History  
 Yes  
 No  
 Unknown  
b. Menopausal status at diagnosis  
 Pre-menopausal  
 Post-menopausal  
 Unknown  
c. Relationship  
 Maternal  
 Paternal
- 18.4. Does your **cousin 1** have a history of breast cancer?  
a. History  
 Yes  
 No  
 Unknown  
b. Menopausal status at diagnosis  
 Pre-menopausal  
 Post-menopausal  
 Unknown  
c. Relationship  
 Maternal  
 Paternal
- 18.5. Does your **cousin 2** have a history of breast cancer?  
a. History  
 Yes  
 No  
 Unknown  
b. Menopausal status at diagnosis  
 Pre-menopausal  
 Post-menopausal  
 Unknown  
c. Relationship  
 Maternal  
 Paternal
- 18.6. Does your **cousin 3** have a history of breast cancer?  
a. History  
 Yes  
 No  
 Unknown  
b. Menopausal status at diagnosis  
 Pre-menopausal  
 Post-menopausal  
 Unknown  
c. Relationship  
 Maternal  
 Paternal
- 18.7. Does your **niece 1** have a history of breast cancer?  
a. History  
 Yes  
 No  
 Unknown  
b. Menopausal status at diagnosis  
 Pre-menopausal  
 Post-menopausal  
 Unknown  
c. Relationship  
 Maternal  
 Paternal

CBCP#

Today's Date:

- 18.8. Does your **niece 2** have a history of breast cancer?  
a. History  
 Yes  
 No  
 Unknown  
b. Menopausal status at diagnosis  
 Pre-menopausal  
 Post-menopausal  
 Unknown  
c. Relationship  
 Maternal  
 Paternal
- 18.9. Does your **niece 3** have a history of breast cancer?  
a. History  
 Yes  
 No  
 Unknown  
b. Menopausal status at diagnosis  
 Pre-menopausal  
 Post-menopausal  
 Unknown  
c. Relationship  
 Maternal  
 Paternal
- 18.10. Does your **grandmother 1** have a history of breast cancer?  
a. History  
 Yes  
 No  
 Unknown  
b. Menopausal status at diagnosis  
 Pre-menopausal  
 Post-menopausal  
 Unknown  
c. Relationship  
 Maternal  
 Paternal
- 18.11. Does your **grandmother 2** have a history of breast cancer?  
a. History  
 Yes  
 No  
 Unknown  
b. Menopausal status at diagnosis  
 Pre-menopausal  
 Post-menopausal  
 Unknown  
c. Relationship  
 Maternal  
 Paternal
- 18.12. Does your **great grandmother 1** have a history of breast cancer?  
a. History  
 Yes  
 No  
 Unknown  
b. Menopausal status at diagnosis  
 Pre-menopausal  
 Post-menopausal  
 Unknown  
c. Relationship  
 Maternal  
 Paternal
- 18.13. Does your **great grandmother 2** have a history of breast cancer?  
a. History  
 Yes  
 No  
 Unknown  
b. Menopausal status at diagnosis  
 Pre-menopausal  
 Post-menopausal  
 Unknown  
c. Relationship  
 Maternal  
 Paternal
- 18.14. Does your **granddaughter 1** have a history of breast cancer?  
a. History  
 Yes  
 No  
 Unknown  
b. Menopausal status at diagnosis  
 Pre-menopausal  
 Post-menopausal  
 Unknown
- 18.15. Does your **granddaughter 2** have a history of breast cancer?  
a. History  
 Yes  
 No  
 Unknown  
b. Menopausal status at diagnosis  
 Pre-menopausal  
 Post-menopausal  
 Unknown

CBCP#

Today's Date:

18.16. Does your **uncle** have a history of breast cancer?

- a. History  
 Yes  
 No  
 Unknown  
b. Relationship  
 Maternal  
 Paternal

18.17. Does your **nephew** have a history of breast cancer?

- a. History  
 Yes  
 No  
 Unknown

18.18. Does your **grandfather** have a history of breast cancer?

- a. History  
 Yes  
 No  
 Unknown  
b. Relationship  
 Maternal  
 Paternal

18.19. Does your **great grandfather** have a history of breast cancer?

- a. History  
 Yes  
 No  
 Unknown  
b. Relationship  
 Maternal  
 Paternal

18.20. Does your **grandson** have a history of breast cancer?

- a. History  
 Yes  
 No  
 Unknown

19. Do you have any personal or family history of colon cancer?

- Yes  
 No (go to #25)  
 Unknown (go to #25)

20. If yes, describe which relative(s).

1. Self  
 2. Maternal Grandmother  
 3. Maternal Grandfather  
 4. Paternal Grandmother  
 5. Paternal Grandfather  
 6. Mother  
 7. Father  
 8. Sister(s) (answer question #21)  
 9. Brother(s) (answer question #22)  
 10. Daughter(s) (answer question #23)  
 11. Son(s) (answer question #24)

21. Please enter # of sister(s) with colon cancer \_\_\_\_\_

22. Please enter # of brother(s) with colon cancer \_\_\_\_\_

23. Please enter # of daughter(s) with colon cancer \_\_\_\_\_

24. Please enter # of son(s) with colon cancer \_\_\_\_\_

25. Do you have any personal or family history of ovarian cancer?

- Yes  
 No (go to #29)  
 Unknown (go to #29)

26. If yes, describe which relative(s)

1. Self  
 2. Maternal Grandmother  
 3. Paternal Grandmother  
 4. Mother  
 5. Sister(s) (answer question #27)  
 6. Daughter(s) (answer question #28)

27. Please enter # of sister(s) with ovarian cancer \_\_\_\_\_

28. Please enter # of daughter(s) with ovarian cancer \_\_\_\_\_

29. Do you have any personal or family history of thyroid cancer?

- Yes  
 No (go to #35)  
 Unknown (go to #35)

CBCP#

Today's Date:

- 30. If yes, describe which relative(s).
  - 1. Self
  - 2. Maternal Grandmother
  - 3. Maternal Grandfather
  - 4. Paternal Grandmother
  - 5. Paternal Grandfather
  - 6. Mother
  - 7. Father
  - 8. Sister(s) (answer question #31)
  - 9. Brother(s) (answer question #32)
  - 10. Daughter(s) (answer question #33)
  - 11. Son(s) (answer question #34)
- 31. Please enter # of sister(s) with thyroid cancer \_\_\_\_\_
- 32. Please enter # of brother(s) with thyroid cancer \_\_\_\_\_
- 33. Please enter # of daughter(s) with thyroid cancer \_\_\_\_\_
- 34. Please enter # of son(s) with thyroid cancer \_\_\_\_\_
- 35. Do you have any personal or family history of any other type of cancer?
  - Yes
  - No (go to #38)
  - Unknown (go to #38)

- 36. If yes, describe which relative(s)
  - 1. Self
  - 2. Maternal Grandmother
  - 3. Maternal Grandfather
  - 4. Paternal Grandmother
  - 5. Paternal Grandfather
  - 6. Mother
  - 7. Father
  - 8. Sister1
  - 9. Sister2
  - 10. Sister3
  - 11. Sister4
  - 12. Brother1
  - 13. Brother2
  - 14. Brother3
  - 15. Brother4
  - 16. Daughter1
  - 17. Daughter2
  - 18. Daughter3
  - 19. Daughter4
  - 20. Son1
  - 21. Son2
  - 22. Son3
  - 23. Son4
- 37. Describe which type of cancer
  - 1. Self \_\_\_\_\_
  - 2. Maternal Grandmother \_\_\_\_\_
  - 3. Maternal Grandfather \_\_\_\_\_
  - 4. Paternal Grandmother \_\_\_\_\_
  - 5. Paternal Grandfather \_\_\_\_\_
  - 6. Mother \_\_\_\_\_
  - 7. Father \_\_\_\_\_
  - 8. Sister1 \_\_\_\_\_
  - 9. Sister2 \_\_\_\_\_
  - 10. Sister3 \_\_\_\_\_
  - 11. Sister4 \_\_\_\_\_
  - 12. Brother1 \_\_\_\_\_
  - 13. Brother2 \_\_\_\_\_
  - 14. Brother3 \_\_\_\_\_
  - 15. Brother4 \_\_\_\_\_
  - 16. Daughter1 \_\_\_\_\_
  - 17. Daughter2 \_\_\_\_\_
  - 18. Daughter3 \_\_\_\_\_
  - 19. Daughter4 \_\_\_\_\_
  - 20. Son1 \_\_\_\_\_
  - 21. Son2 \_\_\_\_\_
  - 22. Son3 \_\_\_\_\_
  - 23. Son4 \_\_\_\_\_

CBCP#

Today's Date:

38. Do you have a history of leg clots (deep vein thrombosis) or embolism?  
 Yes  
 No  
 Unknown
39. Have you ever received any hormonal therapy for prevention/treatment of breast cancer?  
 Yes  
 No (go to #47)  
 Unknown (go to #47)
40. Which type of treatment did you receive?  
 Tamoxifen (go to #42)  
 Raloxifene (go to #42)  
 Other (go to #41)  
 Unknown (go to #42)
41. If other, please specify \_\_\_\_\_
42. Is the length of time on hormone treatment known?  
 Yes  
 No (go to #44)
43. How many months of hormone treatment? \_\_\_\_\_
44. Are you currently receiving hormone treatment?  
 Yes (go to #47)  
 No (go to #45)
45. Do you know when hormone treatment was discontinued?  
 Yes  
 No (go to #47)
46. If yes, month and year treatment was discontinued?  
\_\_\_\_ / \_\_\_\_\_
47. History of severe breast pain lasting over 3 months (Mastalgia)?  
 Yes  
 No (If male, go to #108. If female, go to #51)
48. If yes, did you ever seek treatment for mastalgia?  
 Yes  
 No (If male, go to #108. If female, go to #51)
49. Do you know which treatment you received?  
 Yes  
 No (If male, go to #108. If female, go to #51)
50. If yes, please describe your treatment  
\_\_\_\_\_  
\_\_\_\_\_  
If male, go to #108. If female, go to #51
51. Do you know at what age you began your menstrual period?  
 Yes  
 No (go to #53)
52. Age began menstrual period? \_\_\_\_\_
53. Are you?  
 Pre-menopausal (go to #60)  
 Post-menopausal (go to #58)  
 Surgically menopausal: Both ovaries removed (go to #54)  
 Status post hysterectomy: Ovaries not removed (go to #56)
54. Do you know at what age your ovaries were removed?  
 Yes  
 No (go to #56)
55. At what age were your ovaries removed? \_\_\_\_\_
56. Do you know at what age you had a hysterectomy?  
 Yes  
 No (go to #58)  
 N/A (go to #58)
57. At what age did you have a hysterectomy? \_\_\_\_\_
58. Do you know at what age you experienced menopause?  
 Yes  
 No (go to #60)
59. At what age did you become menopausal? \_\_\_\_\_
60. Have you ever been pregnant?  
 Yes  
 No (go to #63)  
 Unknown (go to #63)
61. Number of times pregnant? \_\_\_\_\_
62. Number of live births? \_\_\_\_\_
63. Have you had any miscarriages?  
 Yes  
 No (go to #65)  
 Unknown (go to #65)

CBCP#

Today's Date:

64. Number of miscarriages? \_\_\_\_\_
65. Have you ever terminated a pregnancy surgically?  
 Yes  
 No (go to #67)
66. Number of surgical terminations? \_\_\_\_\_
67. Have you ever terminated a pregnancy medically?  
 Yes  
 No (go to #69)
68. Number of medical terminations? \_\_\_\_\_
69. Age at your first pregnancy? \_\_\_\_\_
70. Age at your last pregnancy? \_\_\_\_\_
71. Did you ever breastfeed?  
 Yes  
 No (go to #74)
72. Do you know how long you breast fed for?  
 Yes  
 No (go to #74)
73. If yes, how long (total - in months)? \_\_\_\_\_
74. Have you used birth control pills?  
 Yes  
 No (go to #83)  
 Unknown (go to #83)
75. Do you know how long you used birth control pills?  
 Yes  
 No (go to #77)
76. Length of birth control pill usage? (total - in months)  
\_\_\_\_\_
77. Do you know how old you were when you first used birth control pills?  
 Yes  
 No (go to #79)
78. Age at first birth control pill usage? \_\_\_\_\_
79. Do you know how old you were when you last used birth control pills?  
 Yes  
 No (go to #81)
80. Age at last birth control pill usage? \_\_\_\_\_
81. Do you know the name of the birth control pill?  
 Yes  
 No (go to #83)
82. Name of birth control pill(s) used?  
\_\_\_\_\_
83. Have you used hormonal birth control other than the pill?  
 Yes  
 No (go to #92)  
 Unknown (go to #92)
84. If yes, what kind of hormonal birth control?  
 Depo-Provera (go to #86)  
 Norplant (go to #86)  
 IUD with a progestin (go to #86)  
 Other
85. Please describe the kind of hormonal birth control?  
\_\_\_\_\_
86. Do you know how long you used this hormonal birth control?  
 Yes  
 No (go to #88)
87. Length of hormonal birth control? (total - in months)  
\_\_\_\_\_
88. Do you know how old you were when you first used this hormonal birth control?  
 Yes  
 No (go to #90)
89. Age of first hormonal birth control usage? \_\_\_\_\_
90. Do you know how old you were when you last used this hormonal birth control?  
 Yes  
 No (go to #92)
91. Age of last hormonal birth control usage? \_\_\_\_\_
92. History of hormone replacement therapy (HRT)?  
 Yes  
 No (go to #102)  
 Unknown (go to #102)
93. Do you know how long you have been on hormone replacement therapy?  
 Yes  
 No (go to #95)

CBCP#

Today's Date:

94. How long (in years) have you been on hormone replacement therapy? \_\_\_\_\_
95. Do you know at what age you began hormone replacement therapy?  
 Yes  
 No (go to #97)
96. Age at first usage of hormone replacement therapy?  
\_\_\_\_\_
97. Are you still using hormone replacement therapy?  
 Yes (go to #100)  
 No
98. Do you know at what age you stopped hormone replacement therapy?  
 Yes  
 No (go to #100)
99. Age at last usage of hormone replacement therapy?  
\_\_\_\_\_
100. Type of hormone replacement therapy?  
 Estrogen alone (go to #102)  
 Estrogen and Progesterone combo (go to #102)  
 Other  
 Unknown (go to #102)
101. Describe your other hormone replacement therapy?  
\_\_\_\_\_
102. Do you routinely wear a bra?  
 Yes  
 No
103. Do you know at what age you started wearing a bra?  
 Yes  
 No (go to #105)
104. At what age did you start wearing a bra? \_\_\_\_\_
105. Do you routinely wear an underwire bra?  
 Yes  
 No
106. Do you routinely sleep in a bra?  
 Yes  
 No

107. Do you shave your underarms?  
 Yes  
 No
108. Do you use:  
 A deodorant  
 An antiperspirant  
 Both  
 Neither (go to #111)
109. Do you know how long you have used a deodorant or an antiperspirant?  
 Yes  
 No (go to #111)
110. Please enter how long you have used a deodorant or an antiperspirant (in years). \_\_\_\_\_

A drink is defined by the following chart:

Beer	Hard Liquor	Wine
One 12oz can or bottle	1.5oz of liquor (shot)	4oz glass

Using the above chart, please answer the following question through #124

111. Over the past year, how often did you drink?  
 Never (go to #113)  
 1 time per month or less  
 2-3 times per month  
 1-2 times per week  
 3-4 times per week  
 5-6 times per week  
 Daily
112. Over the past year, each time you drank, how much did you consume?  
 1 drink or less  
 2-3 drinks  
 4-5 drinks  
 6 drinks or more
113. Between the ages of 18 and 25, how often did you drink alcohol?  
 Never (go to #115)  
 1 time per month or less  
 2-3 times per month  
 1-2 times per week  
 3-4 times per week  
 5-6 times per week  
 Daily

CBCP#

Today's Date:

114. During this time period, each time you drank, how much did you consume?  
 1 drink or less  
 2-3 drinks  
 4-5 drinks  
 6 drinks or more
115. Between the ages of 26 and 35, how often did you drink alcohol?  
 I am younger than this age range. (go to #125)  
 Never (go to #117)  
 1 time per month or less  
 2-3 times per month  
 1-2 times per week  
 3-4 times per week  
 5-6 times per week  
 Daily
116. During this time period, each time you drank, how much did you consume?  
 1 drink or less  
 2-3 drinks  
 4-5 drinks  
 6 drinks or more
117. Between the ages of 36 and 45, how often did you drink alcohol?  
 I am younger than this age range. (go to #125)  
 Never (go to #119)  
 1 time per month or less  
 2-3 times per month  
 1-2 times per week  
 3-4 times per week  
 5-6 times per week  
 Daily
118. During this time period, each time you drank, how much did you consume?  
 1 drink or less  
 2-3 drinks  
 4-5 drinks  
 6 drinks or more
119. Between the ages of 46 and 55, how often did you drink alcohol?  
 I am younger than this age range. (go to #125)  
 Never (go to #121)  
 1 time per month or less  
 2-3 times per month  
 1-2 times per week  
 3-4 times per week  
 5-6 times per week  
 Daily
120. During this time period, each time you drank, how much did you consume?  
 1 drink or less  
 2-3 drinks  
 4-5 drinks  
 6 drinks or more
121. Between the ages of 56 and 65, how often did you drink alcohol?  
 I am younger than this age range. (go to #125)  
 Never (go to #123)  
 1 time per month or less  
 2-3 times per month  
 1-2 times per week  
 3-4 times per week  
 5-6 times per week  
 Daily
122. During this time period, each time you drank, how much did you consume?  
 1 drink or less  
 2-3 drinks  
 4-5 drinks  
 6 drinks or more
123. Between the ages of 66 and 75, how often did you drink alcohol?  
 I am younger than this age range. (go to #125)  
 Never (go to #125)  
 1 time per month or less  
 2-3 times per month  
 1-2 times per week  
 3-4 times per week  
 5-6 times per week  
 Daily
124. During this time period, each time you drank, how much did you consume?  
 1 drink or less  
 2-3 drinks  
 4-5 drinks  
 6 drinks or more
125. Did you smoke cigarettes in the past year?  
 Yes  
 No (go to #127)
126. Average number of packs a day in the last year  
 < 1 pack per day  
 1 pack per day (approximately)  
 > 1 pack per day

CBCP#

Today's Date:

127. Smoking history prior to the past year?  
# of years smoked (if 0, go to #131) \_\_\_\_\_

128. Prior to the past year, average number of packs a day while you were smoking?  
 < 1 pack per day  
 1 pack per day (approximately)  
 > 1 pack per day

129. If you quit smoking prior to the past year, do you know how long it is since you quit?  
 Yes  
 No (go to #131)

130. # years smoke-free since you quit? \_\_\_\_\_

131. Frequency of routine exercise?  
 Never (go to #133)  
 1 time per week or less  
 1 to 3 times a week  
 3 times a week or more  
 Unknown (go to #133)

132. When you exercise routinely, how long do you exercise for each session?  
 15 min or less  
 15 to 20 minutes  
 20 to 30 minutes  
 30 minutes or more

133. Weight \_\_\_\_\_ lbs

134. Height: \_\_\_\_\_ ft. \_\_\_\_\_ in.

135. Body Mass Index \_\_\_\_\_ kg ( $BMI = kg/m^2$ ) (Data Manager will calculate)

136. Do you have any significant medical disorders? (Please check all that apply)  
 1. Diabetes  
 2. Hypertension  
 3. CAD (Coronary Artery Disease)  
 4. Stroke  
 5. Other  
 6. None (go to #138)  
 7. Unknown (go to #138)

137. Please enter other significant medical disorders present in the patient.

\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_

138. Do you use any medications?  
 Yes  
 No (go to #140)  
 Unknown (go to #140)

139. List all present medications:

\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_

140. Frequency of self-breast examination?  
 Never  
 Every 2-3 months  
 Every month  
 2-3 times a month  
 Every week  
 Every day

141. Have you had any breast biopsies? (Not counting cyst aspirations, but including any current breast biopsy.)  
 Yes  
 No (go to #152)  
 Unknown (go to #152)

CBCP#

Today's Date:

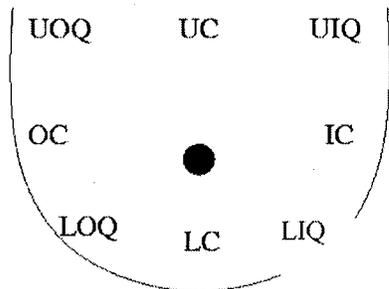
142. Have you had any biopsies on the **RIGHT** breast?  
 Yes  
 No (go to #147)  
 Unknown (go to #147)

143. Number of total biopsies on the **RIGHT** breast?  
 \_\_\_\_\_

144. Number of results of **PREVIOUS** biopsies on the **RIGHT** breast (Previous results only)
1. Atypical \_\_\_\_\_
  2. Benign \_\_\_\_\_
  3. Malignant \_\_\_\_\_
  4. Indeterminate \_\_\_\_\_

145. Is the current **RIGHT** breast abnormally palpable?  
 Yes  
 No  
 Unknown (go to #147)

146. **RIGHT** breast abnormality location
- UOQ
  - UIQ
  - LOQ
  - LIQ
  - Central
  - UC
  - LC
  - OC
  - IC
  - Unknown



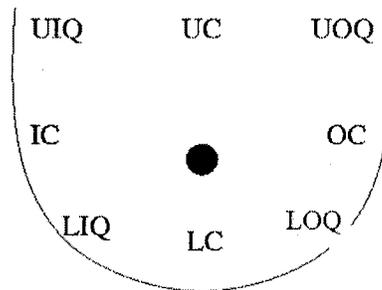
147. Have you had any biopsies on the **LEFT** breast?  
 Yes  
 No (go to #152)  
 Unknown (go to #152)

148. Number of total biopsies on the **LEFT** breast?  
 \_\_\_\_\_

149. Number of results of **PREVIOUS** biopsies on the **LEFT** breast (Previous results only)
1. Atypical \_\_\_\_\_
  2. Benign \_\_\_\_\_
  3. Malignant \_\_\_\_\_
  4. Indeterminate \_\_\_\_\_

150. Is the current **LEFT** breast abnormally palpable?  
 Yes  
 No  
 Unknown (go to #152)

151. **LEFT** breast abnormality location
- UOQ
  - UIQ
  - LOQ
  - LIQ
  - Central
  - UC
  - LC
  - OC
  - IC
  - Unknown



The following questions might require the assistance of a nurse.

152. Have you had a mammogram?  
 Yes  
 No (go to #156)  
 N/A (go to #156)

153. Do you know the date of your last mammogram?  
 Yes  
 No (go to #156)

CBCP#

Today's Date:

154. Date of your last mammogram?  
\_\_\_ / \_\_\_ / \_\_\_\_\_ (mm/dd/yyyy)

155. Mammogram results  
\_\_\_ BI-RADS 0  
\_\_\_ BI-RADS 1  
\_\_\_ BI-RADS 2  
\_\_\_ BI-RADS 3  
\_\_\_ BI-RADS 4  
\_\_\_ BI-RADS 5  
\_\_\_ Unknown

156. Have you had an ultrasound examination?  
\_\_\_ Yes  
\_\_\_ No (go to #160)

157. Do you know the date of your last ultrasound examination?  
\_\_\_ Yes  
\_\_\_ No (go to #159)

158. Ultrasound date  
\_\_\_ / \_\_\_ / \_\_\_\_\_ (mm/dd/yyyy)

159. Ultrasound results  
\_\_\_ Normal  
\_\_\_ Suspicious  
\_\_\_ Indeterminate  
\_\_\_ Cyst  
\_\_\_ Highly suspicious  
\_\_\_ Unknown

160. Has there been an Alkaline Phosphatase blood test drawn within the last three months?  
\_\_\_ Yes  
\_\_\_ No (go to #163)

161. Date of lab result:  
\_\_\_ / \_\_\_ / \_\_\_\_\_ (mm/dd/yyyy)

162. Lab result \_\_\_\_\_

163. Has the patient consented for future use on 01-20006 (tissue)?  
\_\_\_ Yes  
\_\_\_ No

164. Has the patient consented for future use on 01-20007 (blood)?  
\_\_\_ Yes  
\_\_\_ No

165. Gail Model 5-year patient risk \_\_\_\_\_ %

166. Gail Model 5-year patient average risk \_\_\_\_\_ %

167. 5-year risk ratio \_\_\_\_\_ %  
Patients 5-year risk number/average 5-year risk

168. Gail Model lifetime patient risk (to age 90)  
\_\_\_\_\_ %

169. Gail Model average lifetime risk \_\_\_\_\_ %

170. Gail Model lifetime risk ratio \_\_\_\_\_ %  
Patients lifetime risk number/average lifetime risk

Be sure to have the job stress, caffeine, and fat intake scales completed.

170.1. Caffeine Score \_\_\_\_\_

170.2. Fat Intake Score \_\_\_\_\_

170.3. Job Stress Index \_\_\_\_\_

170.4. Job Stress Severity \_\_\_\_\_

170.5. Job Stress Frequency \_\_\_\_\_

170.6. Job Pressure Index \_\_\_\_\_

170.7. Job Pressure Severity \_\_\_\_\_

170.8. Job Pressure Frequency \_\_\_\_\_

170.9. Support Index \_\_\_\_\_

170.10. Support Severity \_\_\_\_\_

170.11. Support Frequency \_\_\_\_\_

**If this is a breast cancer patient, please answer the following questions:**

171. Prior history of breast cancer?  
\_\_\_ Yes  
\_\_\_ No (go to #180)

CBCP#

Today's Date:

172. If yes, which breast?  
 Right  
 Left  
 Both
173. Do you know the year you were previously diagnosed with breast cancer?  
 Yes  
 No (go to #175)
174. Month and year prior breast cancer diagnosed?  
\_\_\_ / \_\_\_ (mm/yyyy)
175. Type of treatment received? (Check all that apply)  
 1. Surgery  
 2. Radiation Treatment  
 3. Chemotherapy  
 4. High Dose Chemotherapy  
 5. Hormonal Therapy  
 6. Immunotherapy  
 7. Complementary Treatment(s)  
 8. Unknown (go to #177)
176. List complementary treatments  
\_\_\_\_\_
177. Did prior cancer metastasize to distant organs?  
 Yes  
 No (go to #180)  
 Unknown (go to #180)
178. If yes, which organs?  
 1. Lung  
 2. Liver  
 3. Bones  
 4. Other
179. Please list which other organs.  
\_\_\_\_\_
180. Has the patient had a chest x-ray?  
 Yes  
 No (go to #183)  
 Unknown (go to #183)
181. What is the date of the chest x-ray?  
\_\_\_ / \_\_\_ / \_\_\_ (mm/dd/yyyy)
182. What is the result of the chest x-ray  
 Evidence of disease  
 No evidence of disease  
 Inconclusive
183. Has the patient had an MRI?  
 Yes  
 No (go to #186)  
 Unknown (go to #186)
184. What is the date of the MRI?  
\_\_\_ / \_\_\_ / \_\_\_ (mm/dd/yyyy)
185. What is the result of the MRI?  
 Evidence of disease  
 No evidence of disease  
 Inconclusive
186. Has the patient had a bone scan?  
 Yes  
 No (go to #189)  
 Unknown (go to #189)
187. What is the date of the bone scan?  
\_\_\_ / \_\_\_ / \_\_\_ (mm/dd/yyyy)
188. What is the result of the bone scan?  
 Evidence of disease  
 No evidence of disease  
 Inconclusive
189. Has the patient had an CT scan?  
 Yes  
 No (go to #192)  
 Unknown (go to #192)
190. What is the date of the CT scan?  
\_\_\_ / \_\_\_ / \_\_\_ (mm/dd/yyyy)
191. What is the result of the CT scan?  
 Evidence of disease  
 No evidence of disease  
 Inconclusive
192. Has the patient had an PET scan?  
 Yes  
 No (go to #195)  
 Unknown (go to #195)
193. What is the date of the PET scan?  
\_\_\_ / \_\_\_ / \_\_\_ (mm/dd/yyyy)
194. What is the result of the PET scan?  
 Evidence of disease  
 No evidence of disease  
 Inconclusive

CBCP#

Today's Date:

195. Is the patient on tamoxifen?  
 Yes  
 No  
 Unknown

196. Have there been any medical complications since last visit?  
 Yes  
 No (go to #199)

197. If yes, what complications?  
 2. Deep Vein Thrombosis (DVT)  
 3. Pulmonary Embolism (PE)  
 4. Endometrial changes  
 5. Cataracts  
 6. Other

198. Please list other complications.  
\_\_\_\_\_

199. Is the patient on aromatase inhibitor?  
 Yes  
 No

200. Has the breast cancer recurred in the same breast?  
 Yes  
 No

201. Has the breast cancer recurred in the opposite breast?  
 Yes  
 No

202. Has breast cancer metastasized?  
 Yes  
 No (go to #205)

203. If yes, to which organs?  
 1. Lung  
 2. Liver  
 3. Bones  
 4. Other

204. Please list other organs.  
\_\_\_\_\_

205. Has any new primary cancer developed?  
 Yes  
 No

206. If yes, which organs?  
 1. Lung  
 2. Liver  
 3. Bones  
 4. Other

207. Please list other organs.  
\_\_\_\_\_

**If this is a high-risk patient, please answer the following questions:**

208. Was the patient's 5-year Gail Model risk > 1.67%?  
 Yes  
 No (go to #215)

209. If yes, was the patient offered chemoprevention?  
 Yes (go to #212)  
 No

210. If no, why not?  
 Medical contraindication  
 Other

211. Please describe other reason  
\_\_\_\_\_

212. If offered chemoprevention, did the patient accept?  
 Yes (go to #215)  
 No

213. If patient did not accept, why?  
 1. Worry of complications  
 2. Blood Clots  
 3. Endometrial cancer  
 4. Wants to think about it and will decide Later  
 5. Unknown  
 6. Other

214. Please list other reasons why patient did not accept.  
\_\_\_\_\_

215. New Gail Model 5-year patient risk \_\_\_\_\_ %

216. New Gail Model 5-year average risk \_\_\_\_\_ %

217. New 5-year Risk Ratio \_\_\_\_\_ %  
Patients 5-year risk number/average 5-year risk

218. New Gail Model lifetime patient risk (to age 90)  
\_\_\_\_\_ %

CBCP#

Today's Date:

219. New Gail Model lifetime average risk (to age 90) \_\_\_\_\_ %

220. New Lifetime Risk Ratio: \_\_\_\_\_ %  
Patients lifetime risk number/average lifetime risk

221. Date of last GYN visit.  
\_\_\_\_ / \_\_\_\_ / \_\_\_\_\_ (mm/dd/yyyy)

222. Is the patient on tamoxifen?  
 Yes  
 No (end)  
 Unknown

223. Have there been any medical complications since your last visit?  
 Yes  
 No (end)

224. If yes, what complications?  
 1. None  
 2. Deep Vein Thrombosis (DVT)  
 3. Pulmonary Embolism (PE)  
 4. Endometrial changes  
 5. Cataracts  
 6. Other

225. Please list other medical complications.  
\_\_\_\_\_

Thank you for helping us in research!

CBCP # \_\_\_\_\_  
DATE: \_\_\_\_\_

### CAFFEINE INTAKE SCALE

Write in the amount that describes your drinking habits within the last week. Eight (8) ounces of coffee, tea, cocoa, or cola = 1 cup, 1 normal coffee mug = 1 cup, 1 can of carbonated soda = 1.5 cups. Be as accurate as possible.

1. How many cups of caffeinated coffee did you drink last week?  
(1 normal size mug = 1 cup, 8 ounces of coffee = 1 cup) \_\_\_\_\_
2. How many cups of tea did you drink last week?  
(1 normal size mug = 1 cup, 8 ounces of tea = 1 cup) \_\_\_\_\_
3. How many cups of cocoa did you drink last week?  
(1 normal size mug = 1 cup, 8 ounces of cocoa = 1 cup) \_\_\_\_\_
4. How many cups of caffeinated cola or carbonated drink did you drink last week?  
(1 can of carbonated drink = 1.5 cups) \_\_\_\_\_

CS \* 92: \_\_\_\_\_

TS \* 70: \_\_\_\_\_

CoS + 5: \_\_\_\_\_

CDS \* 40: \_\_\_\_\_

SS: \_\_\_\_\_

CBCP # \_\_\_\_\_  
DATE: \_\_\_\_\_

### NWLRC FAT INTAKE SCALE

Check the answer which best describes the way you have been eating over the month.

1. How many ounces of meat, fish or poultry do you usually eat?\*

- \_\_\_\_\_ 1. I do not eat meat, fish or poultry
- \_\_\_\_\_ 2. I eat 3 ounces or less per day
- \_\_\_\_\_ 3. I eat 4-6 ounces per day
- \_\_\_\_\_ 4. I eat 7 or more ounces per day

\*3 ounces of meat, fish or chicken is any ONE of the following: 1 regular hamburger, 1 chicken breast, 1 chicken leg (thigh & drumstick), 1 pork chop or 3 slices of pre-sliced lunch meat.

2. How much cheese do you eat per week?

- \_\_\_\_\_ 1. I do not eat cheese
- \_\_\_\_\_ 2. I eat whole milk cheese less than once a week and/or use only low fat cheese such as diet cheese, low fat cottage cheese, or ricotta
- \_\_\_\_\_ 3. I eat whole milk cheese once or twice per week (such as Cheddar, Swiss, and Monterey-Jack)
- \_\_\_\_\_ 4. I eat whole milk cheese three or more times per week

3. What type of milk do you use?

- \_\_\_\_\_ 1. I use only skim or 1% milk or don't use milk
- \_\_\_\_\_ 2. I usually use skim milk or 1% milk but use others occasionally
- \_\_\_\_\_ 3. I usually use 2% or whole milk

4. How many visible egg yolks do you use per week?

- \_\_\_\_\_ 1. I avoid all egg yolks or use less than one per week and / or use only egg substitute
- \_\_\_\_\_ 2. I eat 1-2 egg yolks per week
- \_\_\_\_\_ 3. I eat 3 or more egg yolks per week

5. How often do you eat these meats: regular hamburger, bologna, salami, hot dogs, corned beef, spareribs, sausage, bacon, braunsweiger, or liver? Do not count others.

- \_\_\_\_\_ 1. I do not eat any of these meats
- \_\_\_\_\_ 2. I eat them about once per week or less
- \_\_\_\_\_ 3. I eat them about 2 to 4 times per week
- \_\_\_\_\_ 4. I eat more than servings per week

CBCP # \_\_\_\_\_

DATE: \_\_\_\_\_

6. How many commercial baked goods and how much regular ice cream do you usually eat? (Examples: cake, cookies, coffee cake, sweet rolls, donuts, etc. Do not count low fat versions)

- \_\_\_\_\_ 1. I do not eat commercial baked goods and ice cream
- \_\_\_\_\_ 2. I eat commercial baked goods or ice cream once per week or less
- \_\_\_\_\_ 3. I eat commercial baked goods or ice cream 2-4 times per week
- \_\_\_\_\_ 4. I eat commercial baked goods or ice cream more than 4 times per week

7. What is the main type of fat you cook with?

- \_\_\_\_\_ 1. I use nonstick spray or I do not use fat in cooking
- \_\_\_\_\_ 2. I use a liquid oil (examples: Safflower, sunflower, corn, soybean, and olive oil)
- \_\_\_\_\_ 3. I use margarine
- \_\_\_\_\_ 4. I use butter, shortening, bacon drippings or lard

8. How often do you eat snack foods such as chips, fries or party crackers?

- \_\_\_\_\_ 1. I do not eat these foods
- \_\_\_\_\_ 2. I eat one serving of these snacks per week
- \_\_\_\_\_ 3. I eat these snacks 2-4 times per week
- \_\_\_\_\_ 4. I eat these snack foods more than 4 times per week

9. What spread do you usually use on bread, vegetables, etc?

- \_\_\_\_\_ 1. I do not use any spread
- \_\_\_\_\_ 2. I use diet or light margarine
- \_\_\_\_\_ 3. I use margarine
- \_\_\_\_\_ 4. I use butter

10. How often do you eat candy bars, chocolate, or nuts?

- \_\_\_\_\_ 1. Less than once per week
- \_\_\_\_\_ 2. One to three times per week
- \_\_\_\_\_ 3. More than 3 times per week

11. When you use recipes or convenience foods, how often are they low fat?

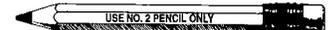
- \_\_\_\_\_ 1. Almost always
- \_\_\_\_\_ 2. Usually
- \_\_\_\_\_ 3. Sometimes
- \_\_\_\_\_ 4. Seldom or never

12. When you eat away from home, how often do you choose low fat foods?

- \_\_\_\_\_ 1. Almost always
- \_\_\_\_\_ 2. Usually
- \_\_\_\_\_ 3. Sometimes
- \_\_\_\_\_ 4. Seldom or never

Sum Score: \_\_\_\_\_

**JOB STRESS SURVEY**  
by C.D. Spielberger, PhD, and P.R. Vagg, PhD



NAME \_\_\_\_\_

<p><b>DIRECTIONS FOR MARKING ANSWER SHEET</b></p> <ul style="list-style-type: none"> <li>• Use a No. 2 black lead pencil. Do NOT use ink or ball point pen.</li> <li>• Make each mark heavy and black. Mark should fill circle completely.</li> <li>• Erase clearly any answer you wish to change. Make no stray marks.</li> </ul>	<b>IDENTIFICATION NUMBER</b>									<b>AGE</b>	<b>TODAY'S DATE</b>			<b>GENDER</b>	<b>YEARS OF EDUCATION</b>		<b>MARITAL STATUS</b>	
												MONTH	DAY		YR.	<input type="radio"/> M <input type="radio"/> F MALE FEMALE	<input type="radio"/> 8th grade or less <input type="radio"/> 9 <input type="radio"/> 13 <input type="radio"/> 10 <input type="radio"/> 14 <input type="radio"/> 11 <input type="radio"/> 15 <input type="radio"/> 12 <input type="radio"/> 16+	<input type="radio"/> Single <input type="radio"/> Married <input type="radio"/> Widowed <input type="radio"/> Separated <input type="radio"/> Divorced
	<input type="radio"/> 0	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5	<input type="radio"/> 6	<input type="radio"/> 7	<input type="radio"/> 8	<input type="radio"/> 9	<input type="radio"/> 0	Jan	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>			
	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5	<input type="radio"/> 6	<input type="radio"/> 7	<input type="radio"/> 8	<input type="radio"/> 9	<input type="radio"/> 0	Feb	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>				
	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5	<input type="radio"/> 6	<input type="radio"/> 7	<input type="radio"/> 8	<input type="radio"/> 9	<input type="radio"/> 0	<input type="radio"/> 1	Mar	<input type="radio"/>	<input type="radio"/> 0	<input type="radio"/> 0	<input type="radio"/> 0			
	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5	<input type="radio"/> 6	<input type="radio"/> 7	<input type="radio"/> 8	<input type="radio"/> 9	<input type="radio"/> 0	<input type="radio"/> 1	<input type="radio"/> 2	Apr	<input type="radio"/>	<input type="radio"/> 1	<input type="radio"/> 1	<input type="radio"/> 1			
	<input type="radio"/> 4	<input type="radio"/> 5	<input type="radio"/> 6	<input type="radio"/> 7	<input type="radio"/> 8	<input type="radio"/> 9	<input type="radio"/> 0	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	May	<input type="radio"/>	<input type="radio"/> 2	<input type="radio"/> 2	<input type="radio"/> 2			
	<input type="radio"/> 5	<input type="radio"/> 6	<input type="radio"/> 7	<input type="radio"/> 8	<input type="radio"/> 9	<input type="radio"/> 0	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	Jun	<input type="radio"/>	<input type="radio"/> 3	<input type="radio"/> 3	<input type="radio"/> 3			
	<input type="radio"/> 6	<input type="radio"/> 7	<input type="radio"/> 8	<input type="radio"/> 9	<input type="radio"/> 0	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5	Jul	<input type="radio"/>	<input type="radio"/> 4	<input type="radio"/> 4	<input type="radio"/> 4			
	<input type="radio"/> 7	<input type="radio"/> 8	<input type="radio"/> 9	<input type="radio"/> 0	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5	<input type="radio"/> 6	Aug	<input type="radio"/>	<input type="radio"/> 5	<input type="radio"/> 5	<input type="radio"/> 5			
<input type="radio"/> 8	<input type="radio"/> 9	<input type="radio"/> 0	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5	<input type="radio"/> 6	<input type="radio"/> 7	Sep	<input type="radio"/>	<input type="radio"/> 6	<input type="radio"/> 6	<input type="radio"/> 6				
<input type="radio"/> 9	<input type="radio"/> 0	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5	<input type="radio"/> 6	<input type="radio"/> 7	<input type="radio"/> 8	Oct	<input type="radio"/>	<input type="radio"/> 7	<input type="radio"/> 7	<input type="radio"/> 7				
										Nov	<input type="radio"/>	<input type="radio"/> 8	<input type="radio"/> 8	<input type="radio"/> 8				
										Dec	<input type="radio"/>	<input type="radio"/> 9	<input type="radio"/> 9	<input type="radio"/> 9				

**CORRECT MARK** ● ● ●

**INCORRECT MARKS** ✓ ✗ ○

Job stress can have serious effects on the lives of employees and their families. The impact of stressful job events is influenced by both the amount of stress associated with a particular event and the frequency of its occurrence. This survey will determine your perception of important sources of stress in your work. The survey lists 30 job-related events that many employees find stressful. First, you will be asked to rate the amount of stress associated with each event. Then, indicate the number of times within the last 6 months that you have experienced each event.

In making your ratings of the amount of stress for each stressor event, use all of your knowledge and experience. Consider the amount of time and energy that you would need to cope with or adjust to the event. Base your ratings on your personal experience as well as what you have seen to be the case for others. Rate the average amount of stress that you feel is associated with each event, rather than the extreme.

The first event, **ASSIGNMENT OF DISAGREEABLE DUTIES**, was rated by persons in a variety of occupations as producing an average amount of stress. This event has been given a rating of "5" and will be used as the standard for evaluating the other events. Compare each event with this standard. Then assign a number from "1" to "9" to indicate whether you judge the event to be less or more stressful than being assigned disagreeable duties.

**Part A. Instructions:** For job-related events judged to produce approximately the same amount of stress as the **ASSIGNMENT OF DISAGREEABLE DUTIES**, fill in the ⑤. For those events that you feel are more stressful than the standard, fill in a number proportionately larger than ⑤. If you feel an event is less stressful than the standard, fill in a number proportionately lower than ⑤.

STRESSFUL JOB-RELATED EVENTS	Amount of Stress								
	Low	Moderate				High			
1A. ASSIGNMENT OF DISAGREEABLE DUTIES	①	②	③	④	●	⑥	⑦	⑧	⑨
2A. Working overtime	①	②	③	④	⑤	⑥	⑦	⑧	⑨
3A. Lack of opportunity for advancement	①	②	③	④	⑤	⑥	⑦	⑧	⑨
4A. Assignment of new or unfamiliar duties	①	②	③	④	⑤	⑥	⑦	⑧	⑨
5A. Fellow workers not doing their job	①	②	③	④	⑤	⑥	⑦	⑧	⑨
6A. Inadequate support by supervisor	①	②	③	④	⑤	⑥	⑦	⑧	⑨
7A. Dealing with crisis situations	①	②	③	④	⑤	⑥	⑦	⑧	⑨
8A. Lack of recognition for good work	①	②	③	④	⑤	⑥	⑦	⑧	⑨
9A. Performing tasks not in job description	①	②	③	④	⑤	⑥	⑦	⑧	⑨
10A. Inadequate or poor quality equipment	①	②	③	④	⑤	⑥	⑦	⑧	⑨
11A. Assignment of increased responsibility	①	②	③	④	⑤	⑥	⑦	⑧	⑨
12A. Periods of inactivity	①	②	③	④	⑤	⑥	⑦	⑧	⑨
13A. Difficulty getting along with supervisor	①	②	③	④	⑤	⑥	⑦	⑧	⑨
14A. Experiencing negative attitudes toward the organization	①	②	③	④	⑤	⑥	⑦	⑧	⑨
15A. Insufficient personnel to handle an assignment	①	②	③	④	⑤	⑥	⑦	⑧	⑨
16A. Making critical on-the-spot decisions	①	②	③	④	⑤	⑥	⑦	⑧	⑨
17A. Personal insult from customer/consumer/colleague	①	②	③	④	⑤	⑥	⑦	⑧	⑨
18A. Lack of participation in policy-making decisions	①	②	③	④	⑤	⑥	⑦	⑧	⑨

**SIDE B**

**STRESSFUL JOB-RELATED EVENTS (Continued)**

	Amount of Stress								
	Low	Moderate	High						
19A. Inadequate salary	1	2	3	4	5	6	7	8	9
20A. Competition for advancement	1	2	3	4	5	6	7	8	9
21A. Poor or inadequate supervision	1	2	3	4	5	6	7	8	9
22A. Noisy work area	1	2	3	4	5	6	7	8	9
23A. Frequent interruptions	1	2	3	4	5	6	7	8	9
24A. Frequent changes from boring to demanding activities	1	2	3	4	5	6	7	8	9
25A. Excessive paperwork	1	2	3	4	5	6	7	8	9
26A. Meeting deadlines	1	2	3	4	5	6	7	8	9
27A. Insufficient personal time (e.g., coffee breaks, lunch)	1	2	3	4	5	6	7	8	9
28A. Covering work for another employee	1	2	3	4	5	6	7	8	9
29A. Poorly motivated coworkers	1	2	3	4	5	6	7	8	9
30A. Conflicts with other departments	1	2	3	4	5	6	7	8	9

**Part B. Instructions:** For each of the job-related events listed, please indicate the approximate number of days during the past 6 months on which you have personally experienced this event. Fill in the ① if the event did not occur; fill in the number ② for each event that you experienced personally on 9 or more days during the past 6 months.

**STRESSFUL JOB-RELATED EVENTS**

**Number of Days on Which the Event Occurred During the Past 6 Months**

1B. Assignment of disagreeable duties	①	1	2	3	4	5	6	7	8	9+
2B. Working overtime	①	1	2	3	4	5	6	7	8	9+
3B. Lack of opportunity for advancement	①	1	2	3	4	5	6	7	8	9+
4B. Assignment of new or unfamiliar duties	①	1	2	3	4	5	6	7	8	9+
5B. Fellow workers not doing their job	①	1	2	3	4	5	6	7	8	9+
6B. Inadequate support by supervisor	①	1	2	3	4	5	6	7	8	9+
7B. Dealing with crisis situations	①	1	2	3	4	5	6	7	8	9+
8B. Lack of recognition for good work	①	1	2	3	4	5	6	7	8	9+
9B. Performing tasks not in job description	①	1	2	3	4	5	6	7	8	9+
10B. Inadequate or poor quality equipment	①	1	2	3	4	5	6	7	8	9+
11B. Assignment of increased responsibility	①	1	2	3	4	5	6	7	8	9+
12B. Periods of inactivity	①	1	2	3	4	5	6	7	8	9+
13B. Difficulty getting along with supervisor	①	1	2	3	4	5	6	7	8	9+
14B. Experiencing negative attitudes toward the organization	①	1	2	3	4	5	6	7	8	9+
15B. Insufficient personnel to handle an assignment	①	1	2	3	4	5	6	7	8	9+
16B. Making critical on-the-spot decisions	①	1	2	3	4	5	6	7	8	9+
17B. Personal insult from customer/consumer/colleague	①	1	2	3	4	5	6	7	8	9+
18B. Lack of participation in policy-making decisions	①	1	2	3	4	5	6	7	8	9+
19B. Inadequate salary	①	1	2	3	4	5	6	7	8	9+
20B. Competition for advancement	①	1	2	3	4	5	6	7	8	9+
21B. Poor or inadequate supervision	①	1	2	3	4	5	6	7	8	9+
22B. Noisy work area	①	1	2	3	4	5	6	7	8	9+
23B. Frequent interruptions	①	1	2	3	4	5	6	7	8	9+
24B. Frequent changes from boring to demanding activities	①	1	2	3	4	5	6	7	8	9+
25B. Excessive paperwork	①	1	2	3	4	5	6	7	8	9+
26B. Meeting deadlines	①	1	2	3	4	5	6	7	8	9+
27B. Insufficient personal time (e.g., coffee breaks, lunch)	①	1	2	3	4	5	6	7	8	9+
28B. Covering work for another employee	①	1	2	3	4	5	6	7	8	9+
29B. Poorly motivated coworkers	①	1	2	3	4	5	6	7	8	9+
30B. Conflicts with other departments	①	1	2	3	4	5	6	7	8	9+



Clinical Breast  
Care Project

The Director of the Clinical Breast Care Project would like to invite you to the 11<sup>th</sup> Annual Clinical Breast Care Project Offsite Meeting

**April 3<sup>rd</sup>-6<sup>th</sup>, 2011**

**Nemacolin Woodlands Resort  
Farmington, Pennsylvania**

This meeting will provide a forum for discussion of current and future opportunities for the Clinical Breast Care Project. Presentations will focus on clinical care, science and research.

Individuals will need to coordinate their travel which will be reimbursed by contacting Kerri Cronin or Dianne Tracey at the number listed below. Lodging and meals will be contracted and arranged for attendees who RSVP. Please RSVP by **March 11<sup>th</sup>, 2011** by completing the attached registration and meal forms. The forms may be returned via email to Kerri Cronin or Dianne Tracey at the email addresses listed below or by faxing your completed forms to 202-782-4260.

For questions or concerns, please contact the meeting coordinators:

Kerri Cronin

Ph: 202-782-0002

Fax: 202-782-4260

[Kerri.cronin@amedd.army.mil](mailto:Kerri.cronin@amedd.army.mil)

Dianne Tracey

Ph: 202-782-0002

Fax: 202-782-4260

[Dianne.tracey.ctr@amedd.army.mil](mailto:Dianne.tracey.ctr@amedd.army.mil)

Sincerely,

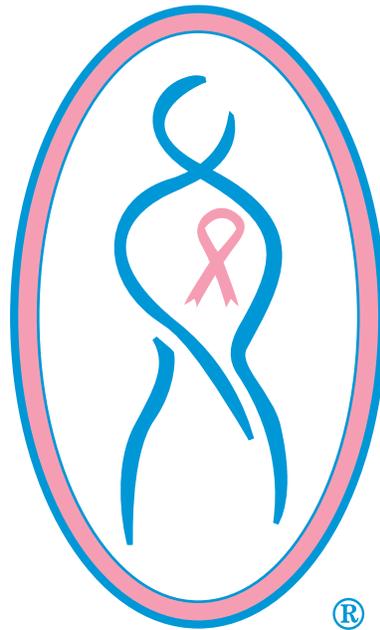
Craig D. Shriver, MD FACS COL MC  
Director and Principle Investigator,  
Clinical Breast Care Project



**Clinical Breast  
Care Project**

Walter Reed Army Medical Center  
Comprehensive Breast Center, Ward 55  
6900 Georgia Avenue, NW  
Washington, DC 20307-5001  
(202) 782-0002 [www.cbcp.info](http://www.cbcp.info)

ELEVEN YEARS OF SUCCESS  
THROUGH SOUND VISION, SUSTAINED GROWTH, & COLLEGIAL TEAMWORK



# Clinical Breast Care Project

11TH ANNUAL CBCP OFFSITE  
AGENDA

*"Crusading against breast disorders through  
prevention, diagnosis, treatment and research"*

## Sunday, April 3, 2011

Time	Meeting	Facilitator	Location
3:00 PM - 6:00 PM	Registration		Marquis Foyer
6:00 PM - 8:30 PM	<i>Cocktail Reception &amp; Dinner</i>		<i>Salons 1 &amp; 2</i>

## Monday, April 4, 2011

Time	Meeting	Facilitator	Location
7:00 AM - 4:00 PM	Registration		Marquis Foyer
7:00 AM - 08:00 AM	<i>Continental Breakfast</i>		<i>Marquis Foyer 3, 4</i>
7:55 AM - 8:00 AM	Administrative Announcements	Lee Bronfman, RN, MA, CCRP	Salons 3 & 4
8:00 AM - 8:30 AM	The Patient's Perspective	Jodi Petit	Salons 3 & 4
8:30 AM - 9:30 AM	Director's remarks	Craig D. Shriver, MD, FACS, COL, MC	Salons 3 & 4
9:30 AM - 10:00AM	<i>Break/Group Photo</i>		<i>Marquis Foyer 3, 4</i>
10:00 AM - 11:00 AM	Keynote Address: A Global Perspective on Breast Healthcare: The Breast Health Global Initiative	Benjamin O. Anderson, MD	Salons 3 & 4
11:00 AM - 11:30 AM	President and CEO Windber Research Inst.	Tom Kurtz	Salons 3 & 4
11:30 AM - 12:00 PM	WRI Chief Scientific Officer	Richard Mural, PhD	Salons 3 & 4
12:00 PM - 1:00 PM	<i>Lunch</i>		<i>Salons 1 &amp; 2</i>
1:00 PM - 1:30 PM	The NINA Study	Barbara Urban, MD & Erin Seeley, PhD	Salons 3 & 4
1:30 PM - 2:00 PM	Whither goest thou, Axillary Dissection?	Philip Perdue, MD	Salons 3 & 4
2:00 PM - 2:45 PM	New And Interesting Points Regarding Genetic Testing For Hereditary Cancers	Raymond Weiss, MD	Salons 3 & 4
2:45 PM - 3:15 PM	Axillary lymph nodes: friend or foe to metastatic colonization?	Rachel Ellsworth, PhD	Salons 3 & 4
3:15 PM - 3:30 PM	<i>Break</i>		<i>Marquis Foyer 3, 4</i>
3:30 PM - 4:00 PM	Tissue Microarray	Jeffrey Hooke, MD	Salons 3 & 4
4:00 PM - 5:30 PM	The Komen Promise Grant - Panel	Albert Kovatich, MS, Hai Hu, PhD, John Eberhardt, Hallgeir Rui, MD, PhD	Salons 3 & 4
6:30 PM - 9:00 PM	<i>Cocktail Reception, Dinner</i>		<i>Grand Ballroom</i>
7:30 PM - 9:30 PM	Steering Committee Meeting	Craig D. Shriver, MD, FACS, COL, MC	Nemacolin 1, 6

## Tuesday, April 5, 2011

Time	Meeting	Facilitator	Location
7:00 AM - 12:00 PM	Registration		Marquis Foyer
7:00 AM - 08:00 AM	<i>Continental Breakfast</i>		<i>Marquis Foyer 3, 4</i>
7:55 AM - 8:00 AM	Administrative Announcements	Lee Bronfman, RN, MA, CCRP	Salons 3, 4
8:00 AM - 8:30 AM	Comparison of gene expression of triple-negative breast tumors from African American and Caucasian women	Lori Field, PhD	Salons 3, 4

8:30 AM - 9:00 AM	Epigenetic Changes in Breast Care	Karin Michaels, ScD, PhD	Salons 3, 4
9:00 AM - 9:30 AM	Treatment-related functional morbidity - The NNCM experience over 10 years	Nicole Stout	Salons 3, 4
9:30 AM - 10:00 AM	Stress Therapy Empowers Prevention (STEP): A Healthy-Lifestyle Program for Breast Cancer Patients	Darrell Ellsworth, PhD	Salons 3, 4
<i>10:00 AM - 10:30 AM</i>	<i>Break</i>		<i>Marquis Foyer 3, 4</i>
10:30 AM - 11:00 AM	Breast Cancer Update in Medical Oncology	Chris Gallagher, MD	Salons 3, 4
11:00 AM - 11:30 AM	The CBCPTissue Bank at WRI - The Year in Review	Stella Somiari, PhD	Salons 3, 4
11:30 AM - 11:50 PM	Tankyrase as a Target of Breast Cancer Therapy	George Iida, PhD	Salons 3, 4
11:50 AM - 12:10 PM	Matrix Metalloproteinase-1 2G Insertion Polymorphism: A Prognostic Marker for Predicting Breast Cancer Severity	Jing Zhou, PhD	Salons 3, 4
12:10 PM - 12:30 PM	Copy Number Variation in Low and High-grade Breast Tumors	Heather Patney	Salons 3, 4
<i>12:30 PM - 1:30 PM</i>	<i>Lunch</i>		<i>Salons 1, 2</i>
1:30 PM - 2:00 PM	The International Breast Care Program: An Exciting Concept	Nick Jacobs, FACHE	Salons 3, 4
2:00 PM - 2:20 PM	Development of CD44 Specific DNA Aptamers that Inhibit Growth, Migration, and Invasion of Breast Cancer Cells	Rebecca Clancy	Salons 3, 4
2:20 PM - 2:40 PM	Wnt Signaling Pathway: Regulation of Migration and Growth of Triple-negative Breast Cancer Cells	George Iida, PhD	Salons 3, 4
<i>2:40 PM - 3:10 PM</i>	<i>Break</i>		<i>Marquis Foyer 3, 4</i>
3:10 PM - 3:40 PM	Biomedical Informatics: the Past Year and the Next Year	Hai Hu, PhD	Salons 3, 4
3:40 PM - 4:00 PM	CBCP Clinical Data QA Measures and Utilities	Leonid Kvecher, MS	Salons 3, 4
4:00 PM - 4:20 PM	Ethnicity Difference of the Concurrence of Benign Breast Diseases with Breast Cancer	Tony Bekhash, MD	Salons 3, 4
4:20 PM - 4:45 PM	Summary and Discussion	Richard Mural, PhD	Salons 3, 4
<i>6:30 PM - 8:30 PM</i>	<i>Cocktail Reception and Dinner</i>		<i>Grand Ballroom</i>
7:30 PM - 9:30 PM	Scientific Advisory Board Meeting	Craig D. Shriver, MD, FACS, COL, MC	Nemacolin 1, 6

## Wednesday, April 6, 2011

Time	Meeting	Facilitator	Location
<i>7:15 AM - 8:15 AM</i>	<i>Continental Breakfast</i>		<i>Marquis Foyer 3, 4</i>
8:00 AM - 8:05 AM	Administrative Announcements	Lee Bronfman, RN, MA, CCRP	Salons 3, 4
8:00 AM - 10:00 AM	Working group: Windber	Richard Mural, PhD	Salon 1
8:00 AM - 10:00 AM	Working group: National Capital Area	Lee Bronfman, RN, MA, CCRP	Salon 2
10:00 AM - 10:30 AM	Closing Remarks	Craig D. Shriver, MD, FACS, COL, MC	Salons 3, 4

## **Congruence between patterns of microRNA expression and histologic grading of invasive breast carcinomas**

Ellsworth DL, Croft DT, Jr, Field LA, Deyarmin B, Kane J, Ellsworth RE, Hooke JA, Shriver CD. Windber Research Institute, Windber, PA; Henry M. Jackson Foundation for the Advancement of Military Medicine, Rockville, MD; Walter Reed Army Medical Center, Washington, DC.

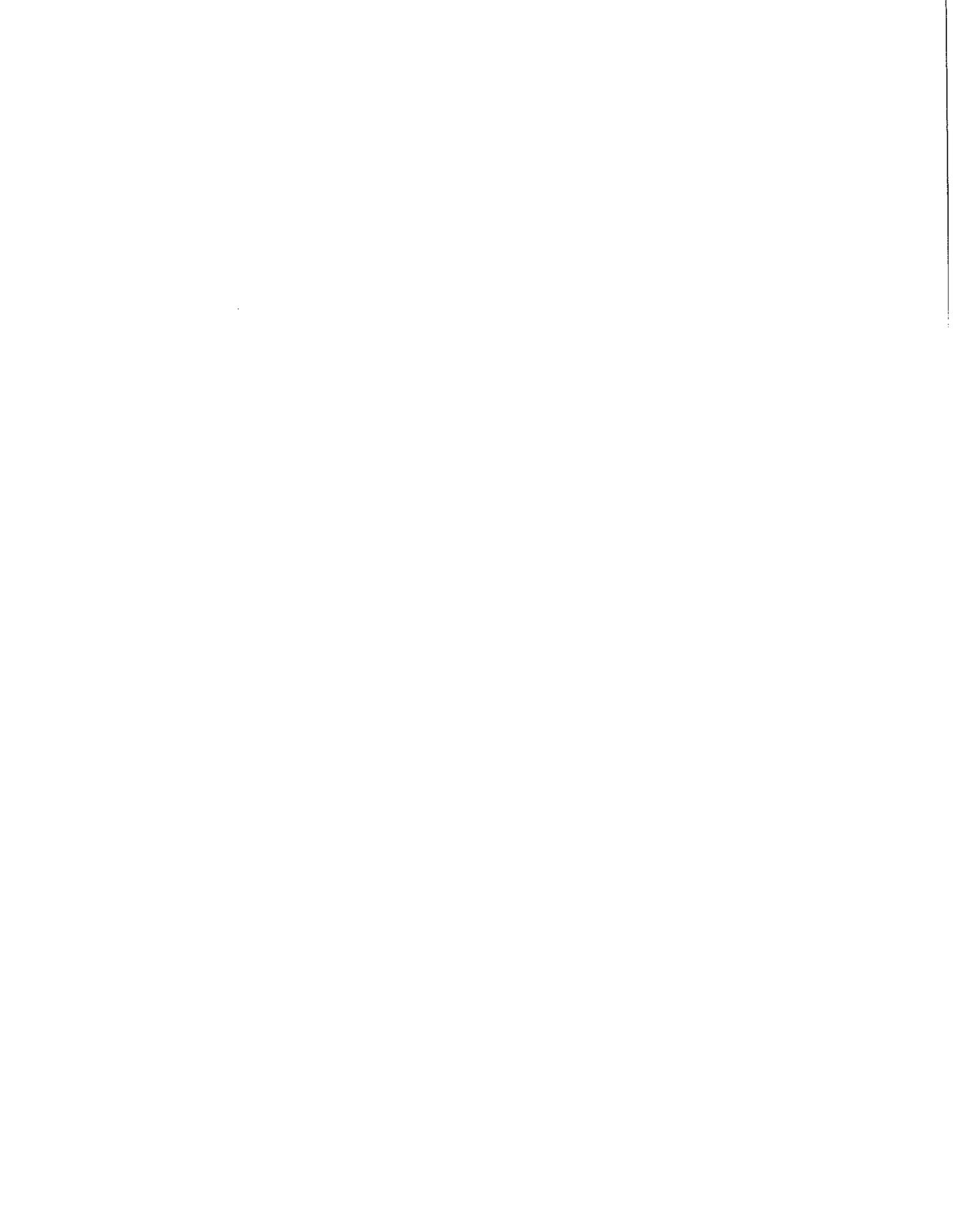
**Background:** Histologic grading may be used as an indicator of prognosis in breast cancer; patients with low-grade carcinomas have ~85% ten-year survival compared to just 45% survival in patients with high-grade disease. Although useful for risk stratification, assigning nuclear grade is subjective, and a large proportion of carcinomas are classified as intermediate-grade with uncertain prognosis, thus limiting clinical utility. MicroRNAs (miRNAs) regulate gene expression and serve an important role in breast cancer development. In this study we examined miRNA expression profiles in low-grade and high-grade breast carcinomas to determine if miRNA expression is associated with pathological classifications of tumor grade.

**Methods:** Breast tumors were obtained from 69 patients enrolled in the Clinical Breast Care Project. Samples were partitioned into low-grade (n=30) or high-grade (n=39) categories using the Nottingham Histologic Score. Following laser microdissection of frozen tissue sections, miRNA was isolated from pure populations of breast tumor cells and hybridized to Affymetrix GeneChip<sup>®</sup> miRNA arrays containing over 800 human miRNA probes. Expression profiles were analyzed with Partek Genomics Suite using the miRNA Expression Module.

**Results:** We identified 30 unique miRNAs that showed differential expression at a False Discovery Rate (FDR)  $p < 0.05$  between low-grade and high-grade breast carcinomas. Gene targets for these miRNAs function primarily in metabolism and cell communication. Expression of hsa-miR-18a and hsa-miR-572 was significantly different between histologic grades at an FDR  $p < 1 \times 10^{-8}$  and hierarchical clustering based on these miRNAs correctly classified 97% (29/30) of low-grade and 90% (35/39) of high-grade tumors. miR-18a has been shown to inhibit ER signaling and promote cellular differentiation, while the role of miR-572 in breast carcinogenesis is not well known.

**Conclusions:** Dysregulation of miRNAs may accompany changes in cellular morphology typically used in histologic classification of breast carcinomas. Patterns of miRNA expression may improve reproducibility and clinical utility of tumor grading and may prove useful for prediction of recurrence and survival for patients with intermediate-grade carcinomas.

<http://www.call4abstracts.com/sabcs/>



## **Genomic heritage of sentinel lymph node metastases: implications for clinical management of breast cancer patients**

Darrell L. Ellsworth, Rachel E. Ellsworth, Tyson E. Becker, Brenda Deyarmin, Heather L. Patney, Jeffrey A. Hooke, Craig D. Shriver

Clinical Breast Care Project, Windber Research Institute, Windber, PA; Clinical Breast Care Project, Walter Reed Army Medical Center, Washington, DC.

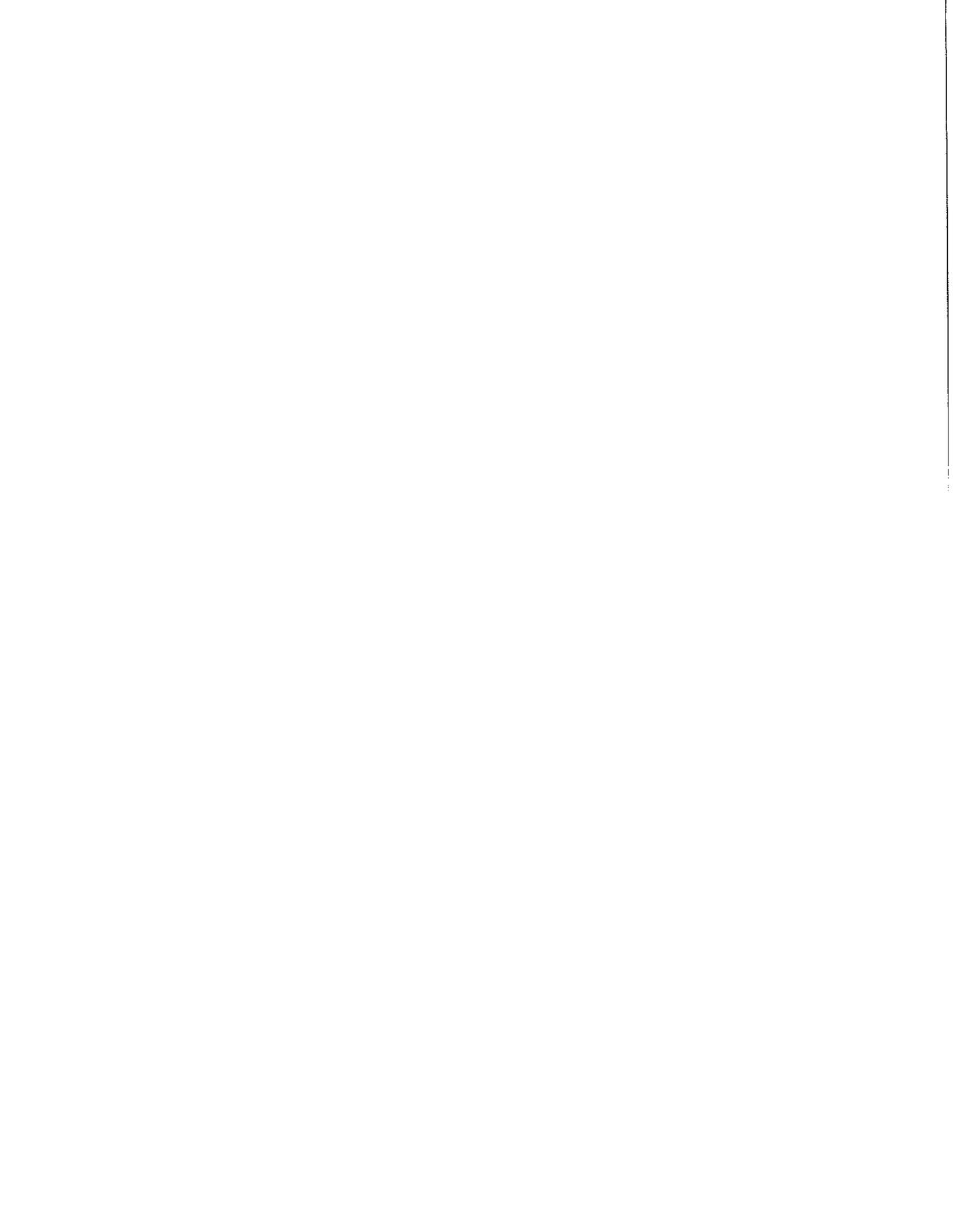
**Background:** Sentinel lymph node (SLN) biopsy status is a key prognostic factor for breast cancer patients. Sentinel nodes are believed to receive early disseminating cells from the primary tumor, but little is known about the origin of metastases colonizing the sentinel nodes. We used allelic imbalance (AI) to examine genomic relationships among metastases in the sentinel and non-sentinel axillary lymph nodes from complete axillary node dissections in 15 patients with lymph node positive breast cancer.

**Methods:** Sentinel nodes were localized by standard scintigraphic and gamma probe techniques using 1.0 mCi technetium-99m sulfur colloid (Tc-99). Pathologically positive nodes were identified by H&E histology and immunohistochemistry. Primary breast tumors and metastases in sentinel and axillary nodes were isolated by laser-assisted microdissection. AI was assessed at 26 chromosomal regions and used to examine the timing and molecular mechanisms of metastatic spread to the sentinel and axillary nodes.

### **Results:**

Overall AI frequencies were significantly higher ( $p < 0.05$ ) in primary breast tumors compared to lymph node metastases. A high level of discordance was observed in patterns and frequencies of AI events between metastases in the sentinel versus the non-sentinel axillary nodes. Phylogenetic analyses showed that (1) multiple genetically-divergent lineages of metastatic cells independently colonize the lymph nodes; (2) some lymph node metastases appeared to acquire metastatic potential early in tumorigenesis, while other metastases evolved later; and (3) importantly, lineages colonizing the sentinel nodes appeared to originate at different times and to progress by different molecular mechanisms.

**Conclusions:** Genomic diversity and timing of metastatic nodal spread may be important factors in determining outcomes of breast cancer patients. Metastases successfully colonizing the sentinel nodes appear to arise at different times during disease progression and may not be descendants of progenitor cells that colonize the lymph nodes early in tumorigenesis. Metastatic growth in the sentinel nodes thus may be a consequence of stimulating factors from the primary tumor that affect proliferation of previously disseminated cells rather than the timing of metastatic spread.



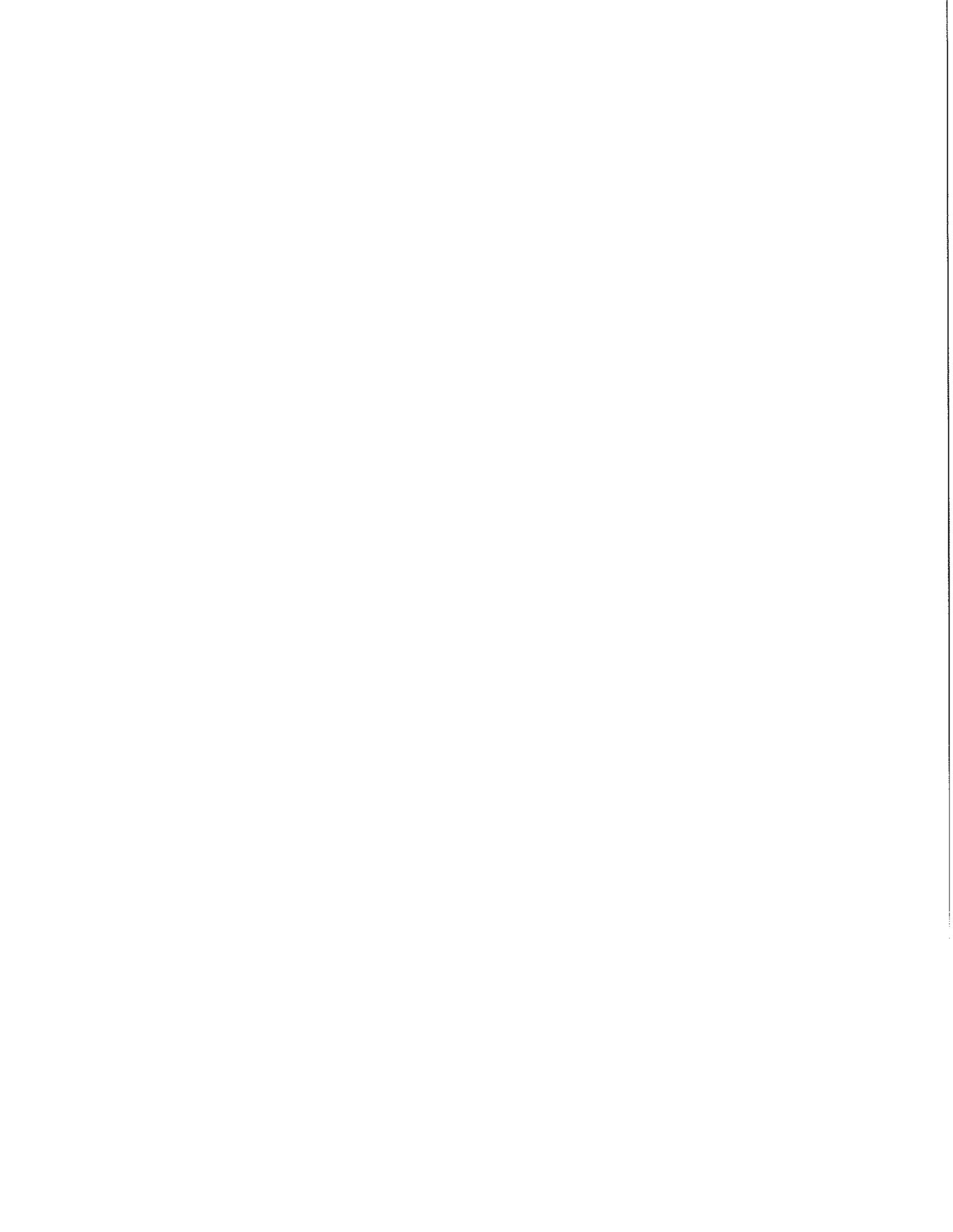
## HER2 BRCA 1a

**Background:** While HER2 status is used prognostically, the effects of HER2 can be influenced by other genes; co-amplification of TOP2A modulates the effectiveness of treatment with Herceptin. BRCA1, deleted in >50% of sporadic breast tumors, is located distal to HER2 on chromosome 17q and is marked by structural instability. To determine how alterations at the BRCA1 locus may modify amplification of HER2, FISH analysis of BRCA1 was performed in breast tumors with HER2 amplification and/or allelic imbalance on chromosome 17q12.

**Methods:** HER2 status (n=68) was determined using the PathVysion HER2 kit. For BRCA1 analysis, DNA probes were generated by nick translation from a BAC clone containing the entire BRCA1 gene in conjunction with a CEP 17 probe. Copy number gains and deletions were defined as BRCA1:CEP17 >1.3 and <0.8, respectively; aneusomy was defined when >10% of tumor cells had copy numbers > (polysomy) or < (monosomy) than two.

**Results:** Amplification of the HER2 gene was detected in 57% of tumors; an additional 16% of tumor were polysomic for chromosome 17q. The remaining 27% of tumors had previously detected allelic imbalance at the 17q12 region. The majority of samples (47%) were characterized by an amplification of HER2 with a downstream deletion of BRCA1, while HER2 and BRCA1 were amplified concomitantly in 21% of tumors, including those with polysomy of 17q. Fifteen percent of samples were monosomic for chromosome 17q; using current clinical scoring criteria, each of those samples were reported as having, despite the altered copy number, negative HER2 status. Levels of overall genomic instability, which were previously measured at 26 chromosomal regions, were significantly higher ( $P < 0.05$ ) in tumors with deletion of BRCA1 compared to those without loss of BRCA1 (38% and 24%, respectively).

**Conclusions:** Alterations, especially deletions, of BRCA1 are highly correlated with copy number changes in HER2. Because deletion of BRCA1 has been associated with overall genomic instability, patients with amplification of HER2 and deletion of BRCA1 may prove unresponsive treatment with Herceptin. Loss of BRCA1 should, therefore, be considered as a genetic modifier to HER2 amplification and the status of BRCA1 should be considered in selecting treatment options for patients with HER2 amplified tumors.



## **Identification of proteins promoting development of metastatic breast tumors**

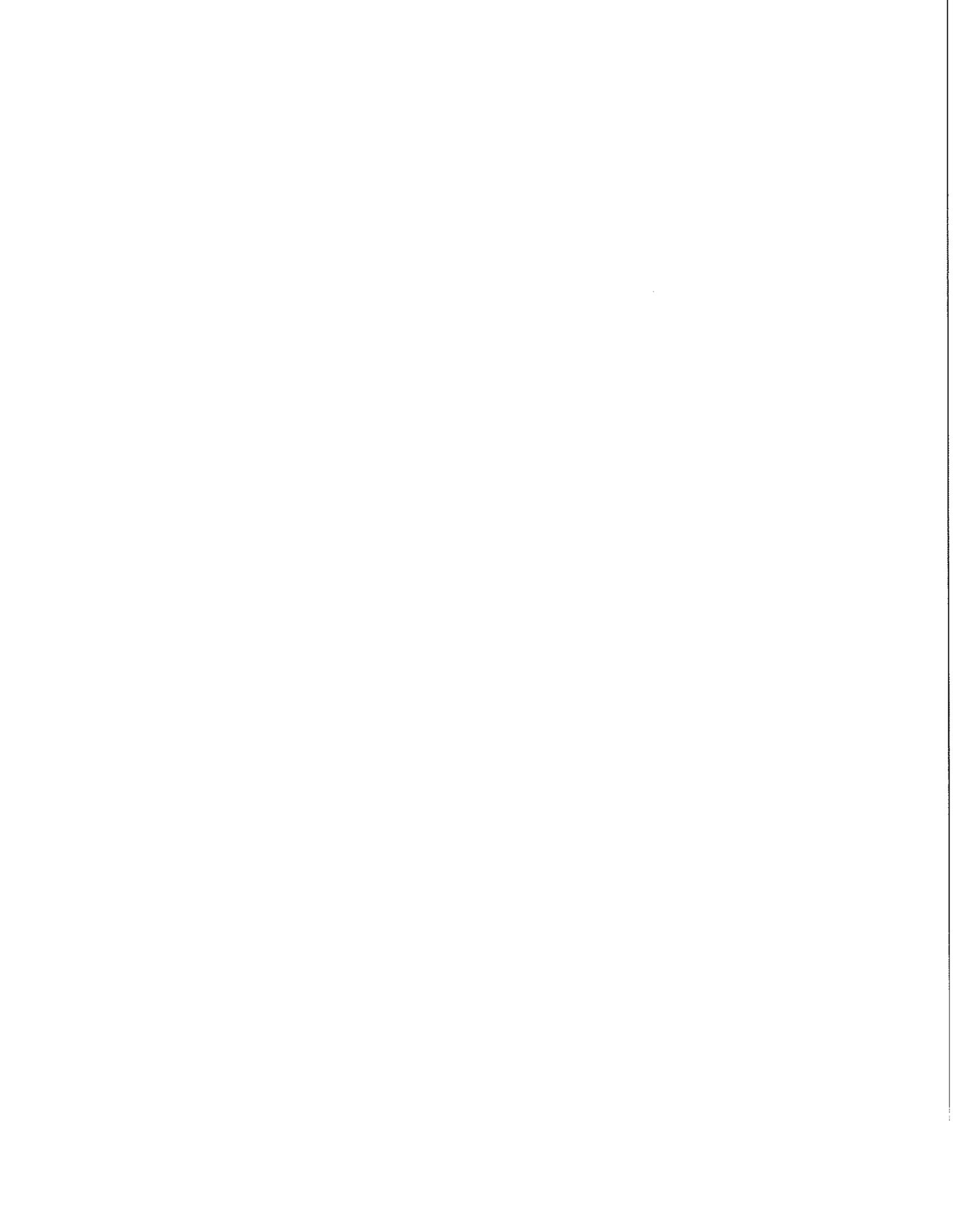
Rachel E. Ellsworth, Erin H. Seeley, Darrell L. Ellsworth, Melinda E. Sanders, Jeffrey A. Hooke, Richard M. Caprioli, Craig D. Shriver

**Background:** Primary breast tumors are constantly shedding tumor cells into the circulatory and lymphatic systems. Although detection of occult tumor cells is a risk factor for recurrence and progression, tumor cells remain detectable years after initial diagnosis in patients without clinical or histological detection of metastasis. Thus the question remains as to why some women with circulating tumor cells develop metastatic breast cancer while others do not.

**Methods:** Negative lymph nodes from women with node negative (n=22) and node positive disease (n=36) were obtained from patients enrolled in the Clinical Breast Care Project. Negative lymph node status was confirmed by IHC analysis. Frozen tissues were sectioned and mounted on gold coated MALDI target plates for protein expression profiling. Hematoxylin and eosin (H&E) stained slides were prepared from serial sections for histological characterization. MALDI matrix was deposited as individual spots on the tissue sections in a histology directed manner to assay specific areas and tissue types of interest. Mass spectral data were then acquired from multiple sites across each tissue section.

**Results:** 131 features were observed in negative nodes from patients without metastatic disease and 129 in negative nodes from patients with lymph node metastases. While the majority of features detected were similar between the two groups, 8.5% were differentially expressed. Two of the features which were expressed at significantly higher levels in nodes from patients with metastatic disease have been putatively identified as thymosin  $\beta$ 4 and thymosin  $\beta$ 10.

**Conclusions:** Thymosin  $\beta$ 4 and 10 have been associated with disease progression and metastatic capacity in a number of tumor types. The overexpression of these proteins in tumor-negative nodes from patients with metastatic disease in other regional nodes suggests lymph nodes do not play a passive role in metastasis, rather, expression of a specific subset of proteins creates an hospitable environment to facilitate colonization. These markers of metastasis may permit molecular discrimination of those patients with indolent disease from those at risk for metastasis and will thus allow for the design of customized treatment regimens to more effectively treat, or prevent, metastatic spread.



## **HER-2 Gene Amplification is a Marker for Global Genomic Instability**

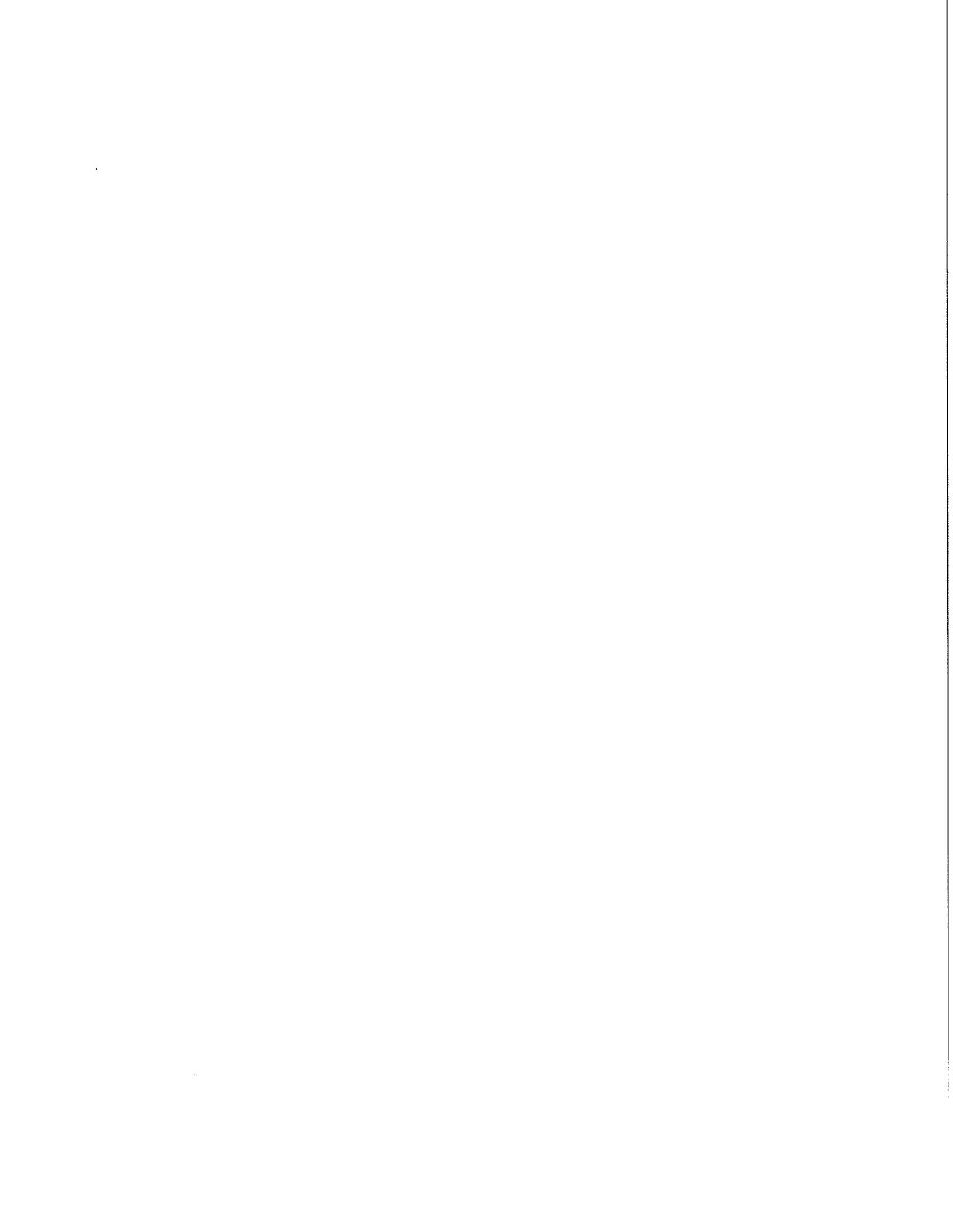
Rachel E. Ellsworth, Brenda Deyarmin, Heather Patney, Jeffrey A. Hooke, Craig D. Shriver

**BACKGROUND** Genomic alterations of the proto-oncogene c-erbB-2 (HER-2/neu) are associated with poor prognosis in patients with breast cancer. While HER-2 status has predictive value for the efficacy of different therapies, its long-term prognostic relevance remains controversial. Because some patients with HER-2 gene amplification do not exhibit corresponding over-expression of the HER-2 protein, it is unclear how genomic alterations predict clinical response. To assess the broader genomic implications of structural changes at the HER-2 locus, we investigated relationships between genomic instability and HER-2 status in patients with invasive breast cancer.

**METHODS** DNA was extracted after laser microdissection from 190 paraffin-embedded invasive breast tumor specimens, 22% (41/190) of which had HER-2 amplification determined by the PathVysion<sup>®</sup> fluorescence in situ hybridization (FISH) assay, with an additional 17 specimens having polysomy for chromosome 17q. Referent DNA was extracted from blood, disease-free skin, or negative lymph node samples. Allelic imbalance (AI) was assessed using a panel of microsatellite markers representing 26 chromosomal regions commonly altered in breast cancer. Exact unconditional and Student t-tests were used to investigate relationships between genomic instability and HER-2 status.

**RESULTS** Overall levels of AI did not differ significantly in patients stratified by stage, age or hormonal status. The frequency of AI was significantly higher ( $P < 0.0001$ ) in patients with HER-2 amplification/polysomy (28%) compared to those without chromosome 17 alterations (19%). When individual chromosomal regions were examined, samples with HER-2 amplification showed significantly higher levels of AI ( $P < 0.05$ ) at chromosomes 11q23, 17p13.3, 17q12-q21 and 18q21.

**CONCLUSIONS** HER-2 gene amplification may serve as a marker of global genomic instability. Concurrent genetic alterations in other regions of the genome may contribute to poor prognosis and influence associations between HER-2 status and outcomes of adjuvant chemotherapy and Herceptin<sup>®</sup> treatment. The addition of molecular profiling for AI may thus enhance treatment decisions for individuals with HER-2 amplified breast cancer.



Identification of chromosomal changes associated with the transition from *in situ* to invasive breast cancer

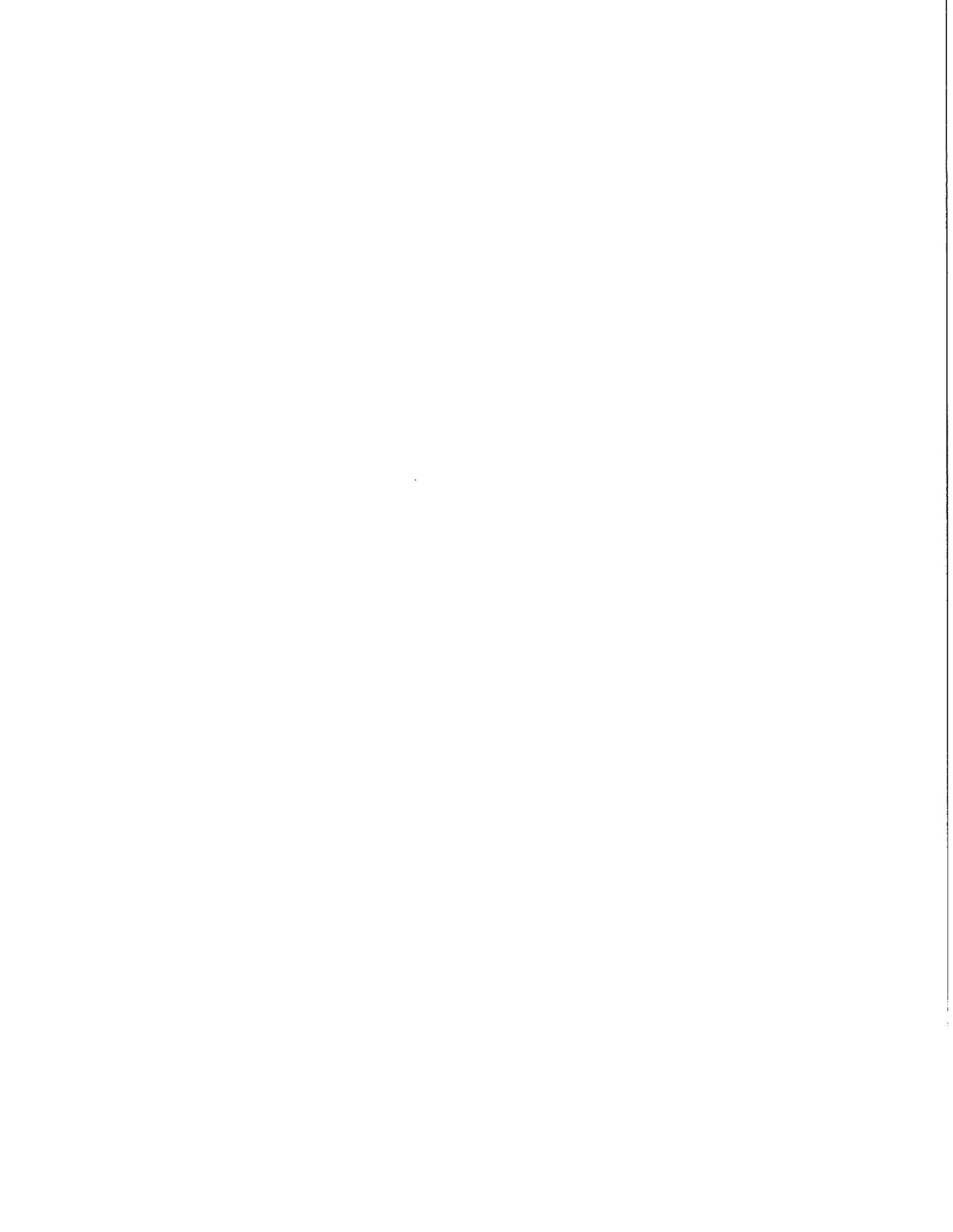
RE Ellsworth, B Love, JA Hooke, CD Shriver

**Background:** DCIS is a preinvasive lesions of the breast in which the duct is filled with malignant epithelial cells. Although these malignant cells are contained within the breast, DCIS is a precursor to invasive breast cancer. Molecular and biological changes responsible for the transition from *in situ* to invasive breast disease are unknown.

**Methods:** Allelic imbalance data representing 26 chromosomal regions frequently altered in breast cancer was generated from 65 DCIS lesions (18 grade 1, 47 grade 3) and 103 invasive breast tumors (55 well-differentiated, 58 poorly-differentiated) was generated. Based on molecular data suggesting breast disease develops along either a low- or high-grade genetic pathway, overall levels and patterns of AI were compared between grade 1 DCIS and well-differentiated breast cancers and between grade 3 DCIS and poorly-differentiated invasive breast cancers using exact unconditional and student's t-tests.

**Results:** Overall levels of AI were significantly higher ( $P < 0.01$ ) in grade 1 DCIS (12%) compared to well-differentiated breast cancer (19%); levels were not significantly different between grade 3 DCIS and poorly-differentiated breast cancer. No significant differences were detected at any of the 26 chromosomal regions between grade 1 DCIS and well-differentiated breast tumors, however, levels of AI were significantly higher in poorly-differentiated tumors compared to grade 3 DCIS at chromosomes 1p36 and 11q23.

**Conclusions:** Although low-grade DCIS and well-differentiated breast cancer did not differ in patterns of AI, two chromosomal regions were effective in discriminating high-grade *in situ* from invasive disease. Candidate genes include RUNX3 on chromosome 1p36 and HYOU1 on chromosome 11q23, both of which function in invasion. Further characterization of these critical changes may provide molecular assays to identify DCIS lesions with invasive potential as well as new targets for molecular therapeutics.



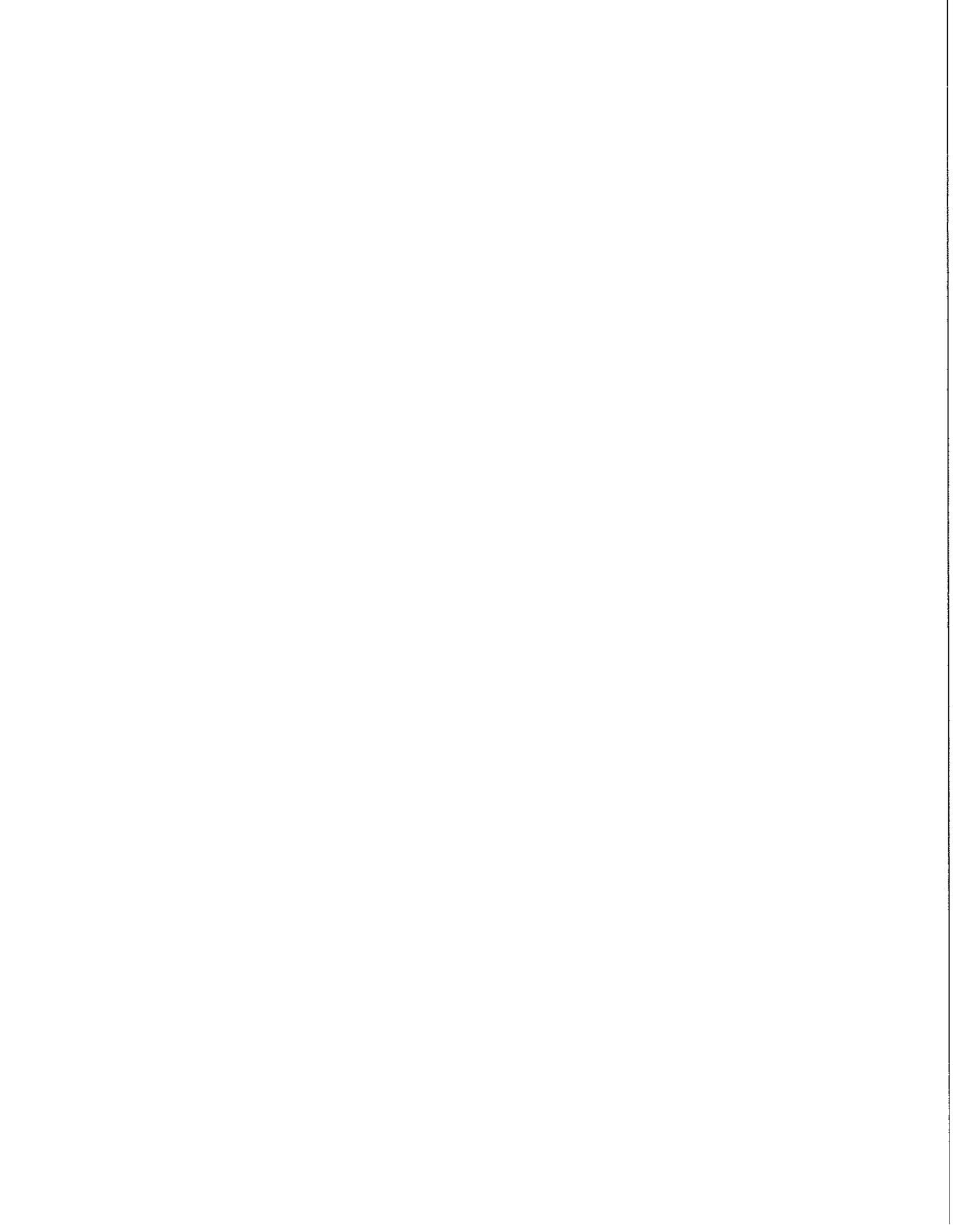
Molecular relationship between HER2 and BRCA1 in invasive breast carcinoma  
RE Ellsworth, B Deyarmin, CD Shriver

**Background:** Deletion of 17q including the BRCA1 gene region at 17q21 is one of the most common alterations found in sporadic breast cancers. Deletion of this region often results in inactivation of the BRCA1 tumor suppressor gene, which inhibits normal DNA repair, thus rendering cells susceptible to additional DNA damage. In contrast, the HER2 gene, located at 17q12 within 5 Mb of the BRCA1 gene, is found to be amplified in ~20% of breast tumors. Although efforts have sought to identify the smallest common region amplified in patients with HER2 amplification, the status of BRCA1 in these patients, whether co-amplified or deleted, is unknown, thus to determine the molecular relationships between BRCA1 and HER2, fluorescence in situ hybridization (FISH) analysis of BRCA1 was performed in breast tumors with documented HER2 amplification and/or allelic imbalance at 17q12-q21.

**Methods:** HER2 status was determined for 77 invasive breast tumors using the PathVysion HER2 kit. For BRCA1 analysis, DNA probes were generated by nick translation from BAC clone RP11-831F13 in conjunction with a CEP 17 probe. Chromosomal content was defined using copy number values of the CEP17 probe: monosomy =  $\leq 1.75$ , disomy = 1.76-2.25, and polysomy =  $> 2.25$ . Chromosomal gains and losses for HER2 and BRCA1 were defined as copy number ratios of  $\geq 3.0$  and  $< 1.75$ , respectively.

**Results:** The majority (73%) of breast tumors were aneusomic (42% polysomy, 31% monosomy) for CEP17. Frequency of HER2 gains and losses was 73% and 4%, respectively. For BRCA1, 56% of specimens had a loss of BRCA1, but no tumors showed a gain of BRCA1. HER2 (gain) and BRCA1 (deletion) status were discordant in 35% of specimens and were divided equally between polysomic, disomic, and monosomic centromere backgrounds.

**Conclusions:** The high frequency of BRCA1 deletion and absence of copy number gain in breast tumors selected for amplification of HER2 suggests that alterations on 17q12-q21 do not represent a single large amplicon, rather amplification and deletion occur independently within this 5 Mb region. Although the timing of these separate events is unknown, 20% of samples had copy number  $\leq 1.75$  at CEP17 and BRCA1 with normal or amplified chromosomal content at HER2 suggesting an early large deletion of 17q followed by targeted amplification of the HER2 region. This data, in conjunction with the high levels of repetitive elements within the BRCA1 gene region, suggest that genomic instability of BRCA1 may influence HER2 amplification.

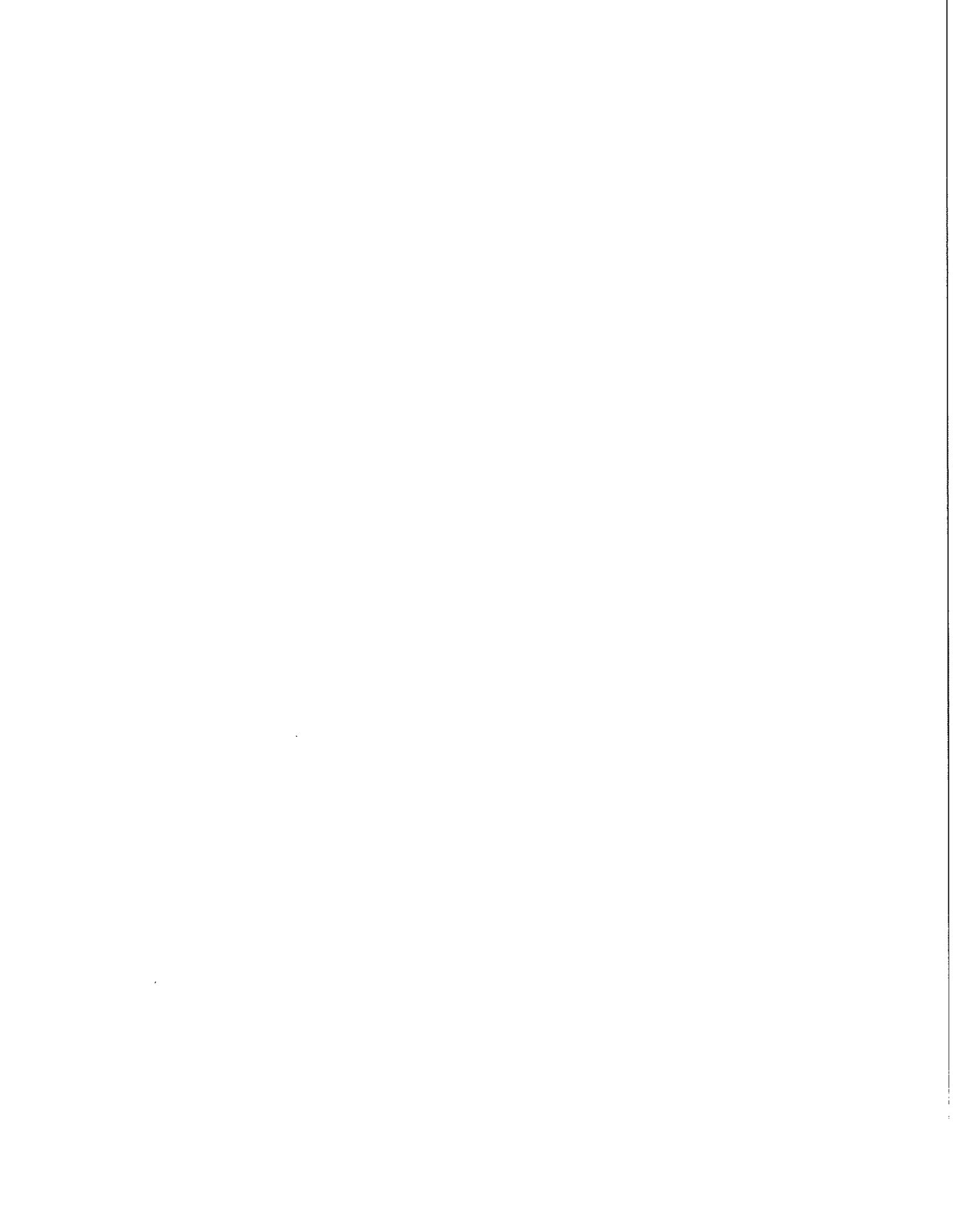


Identification of cSNPs in environmental response genes contributing to breast cancer etiology

RE Ellsworth<sup>1</sup>, JD Weyandt<sup>1</sup>, HL Patney<sup>1</sup>, K Anthony<sup>1</sup>, CD Shriver<sup>2</sup>

<sup>1</sup>Windber Research Institute, Windber, PA; <sup>2</sup>Walter Reed Army Medical Center, Washington, DC

Sporadic cancer is likely caused by a number of DNA variants, each making a small contribution to overall cancer risk. The Environmental Genome Project has identified 57 potentially deleterious cSNPs in 40 genes involved in metabolism of toxins, drug clearance, and DNA repair with population frequencies ranging from 0.01 – 0.38. To determine whether these cSNPs contribute to breast cancer etiology, we examined the prevalence of these variants in women with invasive breast cancer (n=212) and age- and ethnicity-matched (n=212) female controls enrolled in the Clinical Breast Care Project (CBCP). The patient population was comprised of 21% African American and 79% Caucasian women; other minority populations were excluded due to small sample size. Genotypes were determined by RFLP assays or by direct sequencing. Minor allele frequencies (MAFs) in controls were in agreement with published values; most cSNPs had MAFs  $\leq 0.01$ , while SNPs in GSTZ1 (rs3177427), GCKR (rs1260326), MTHFR (rs1801133 and rs1801131) and ERCC5 (rs17655) had relatively high minor allele frequencies ( $>0.20$ ). Overall allele frequencies differed significantly ( $P < 0.05$ ) between cases and controls for SNPs rs1799950 (BRCA1), rs1799853 (CYP2C9), and rs3218778 (POLI). Examination of the data by ethnic group revealed that a number of the cSNPs differed between cases and controls within a single population: the Q356R BRCA1, A232V RFC2 and R71G POLI variants in Caucasians, and the W452C ERBB2 variant in African Americans. Altered expression of ERBB2 has been associated with poor prognosis in women with breast cancer and African American women frequently have aggressive tumors with poor clinical outcomes. CYP2C9 is involved in estrogen metabolism, and BRCA1, POLI, and RFC2 are largely involved in proper DNA synthesis and thus deleterious SNPs may contribute to increased risk of DNA damage, genomic instability and disease pathogenesis. Functional variants in environmental response genes may, therefore, impair the ability to respond to exogenous exposures and increase the risk of developing breast cancer.



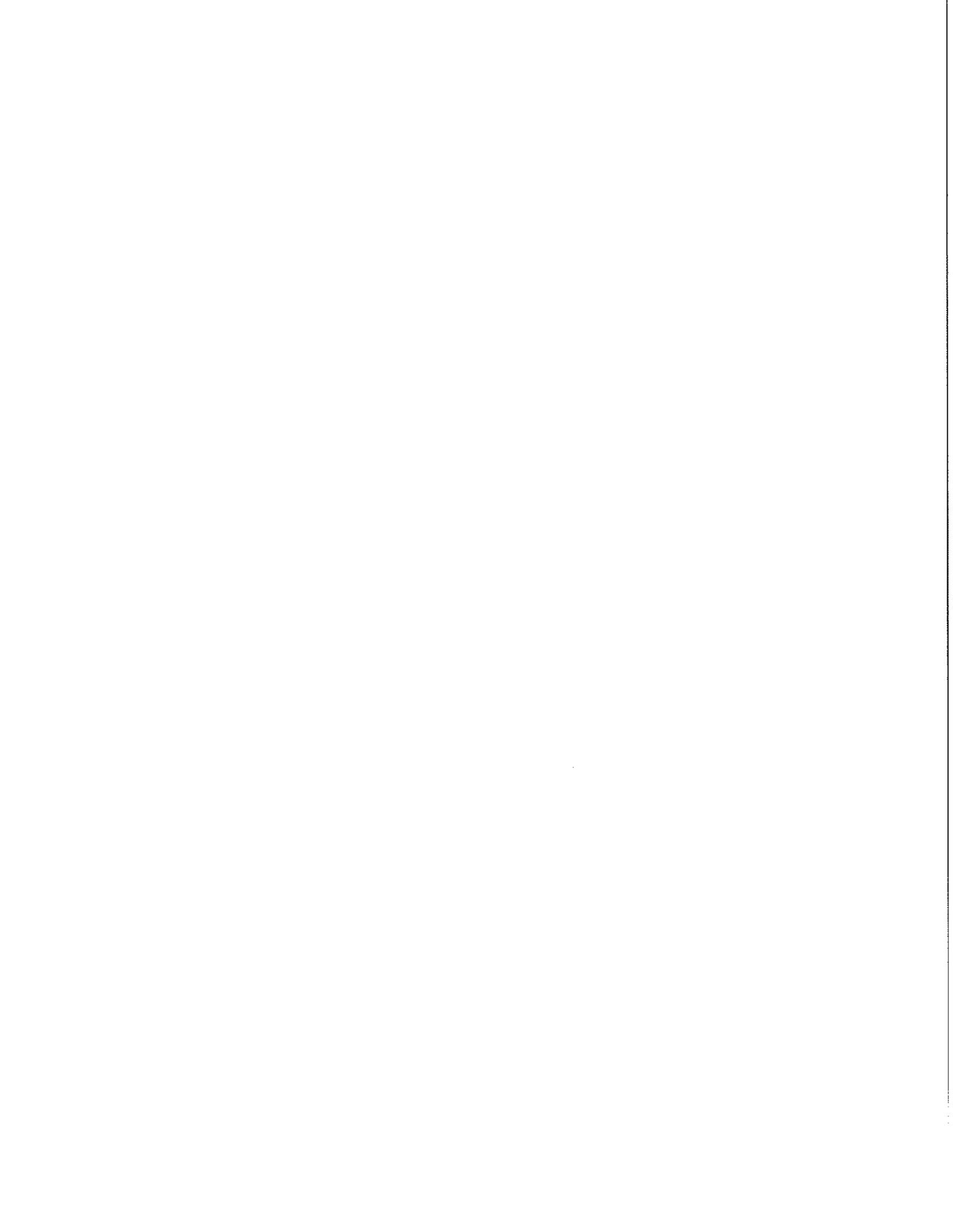
Genomic discrimination of metastatic from non-metastatic primary breast tumors  
Rachel E. Ellsworth, PhD, Heather L. Patney, BS, Darrell L. Ellsworth, PhD Brad Love,  
PhD, Jeffrey A. Hooke, MD, Craig D. Shriver, MD

**Background:** Axillary lymph node status is one of the most important prognostic factors in treatment selection for patients with breast cancer. Several genetic studies attempting to identify molecular profiles have produced conflicting results, thus it is unclear whether genetic changes in the primary breast tumor are predictive of lymph node involvement. This study examined allelic imbalance (AI) in primary breast tumors from node-negative and node-positive patients to determine whether patterns of genomic changes can be used to identify metastatic breast cancers.

**Methods:** Laser microdissected DNA samples were collected from primary breast tumors of 102 patients with positive lymph nodes and 98 patients with negative lymph nodes. Fifty-two microsatellite markers representing 26 chromosomal regions commonly deleted in breast cancer were used to assess patterns of AI.

**Results:** Overall AI frequencies between node negative (21.7%) and node positive (23.5%) breast tumors did not differ significantly ( $P=0.1986$ ). When the data were examined by chromosomal region, chromosome 8q24 had significantly higher ( $P<0.005$ ) levels of AI events in the node positive primary tumor samples (27%) compared to the node negative samples (9%). No node negative samples had significantly higher levels of AI compared to node positive samples at any of the 26 chromosomal regions.

**Conclusions:** Examination of the 8q24 region revealed two candidate genes - metastasis suppressor 1 (MTSS1) and MYC. MTSS1 is involved in actin binding, and loss MTSS1 is associated with disruption of the cytoskeleton while overexpression of MYC has been associated with changes in bone marrow microenvironment allowing successful colonization by metastatic breast cells, as well as with lymphangiogenesis, promoting the proliferation of metastatic cells. Thus, alterations of the 8q24 region may be a key determinant in development of metastatic lymph nodes.



Identification of protein expression differences in invasive breast tumors from African American compared to Caucasian women

Rachel E. Ellsworth, Erin H. Seeley, Darrell L. Ellsworth, Jeffrey A. Hooke, Richard M. Caprioli, Craig D. Shriver

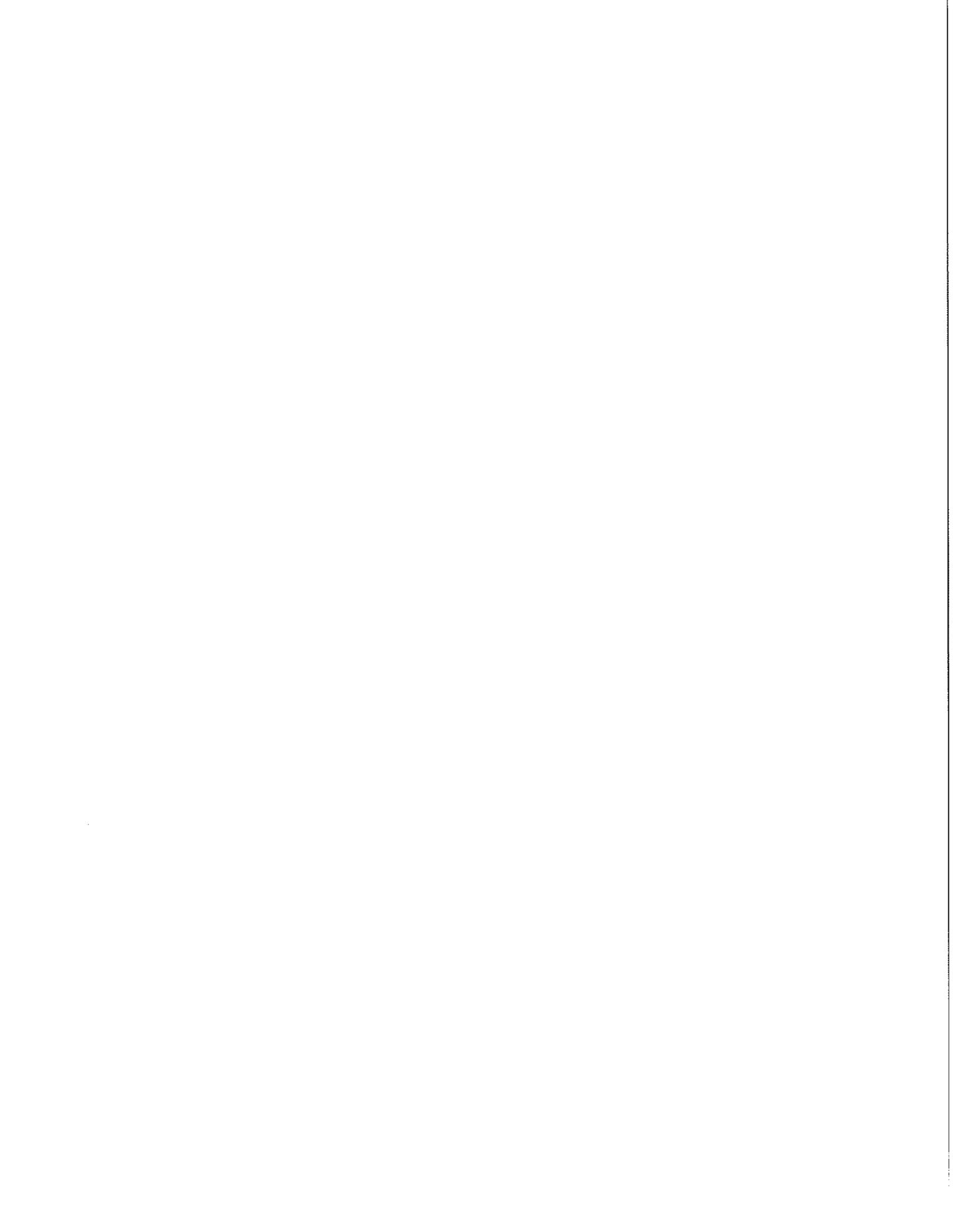
**Background:** Breast cancer in African American women (AAW) occurs at a younger age than in Caucasian women (CW) and has a significantly more aggressive phenotype including negative ER, positive lymph node and p53 status and more frequently have the triple negative (ER, PR, HER2 negative). The aggressive nature of these tumors, coupled with the decreased availability and efficacy of hormone and chemotherapeutics render breast cancer outcomes in AAW significantly less favorable than in CW. Underlying molecular causes of this more aggressive phenotype remain unknown.

**Methods:** OCT embedded frozen primary breast tumors from African American (n=12) and Caucasian women (n=16) were obtained from patients enrolled in the Clinical Breast Care Project. Frozen tissues were sectioned and mounted on gold coated MALDI target plates for protein expression profiling. Hematoxylin and eosin (H&E) stained slides were prepared from serial sections for histological characterization. MALDI matrix was deposited as individual spots on the tissue sections in a histology directed manner to assay specific areas and tissue types of interest. Mass spectral data were then acquired from multiple sites across each tissue section.

**Results:** 141 features were observed in tumors from AAW and 149 in those from CW. While the majority of features detected were similar between the two groups, 11 were differentially expressed: 8 expressed at higher levels in tumors from AAW and 3 from CW. Protein features with *m/z* charges of 10850, 3371, and 3443 were expressed at levels more than 2-fold higher in tumors from AAW compared to CW.

**Conclusions:** Differential expression of proteins in tumors from young AAW patients compared to those from CW supports a model in which aggressive tumor phenotypes and less favorable outcomes are driven by molecular differences within the tumors.

Identification of these differentially expressed protein features will allow for the design of improved diagnostics and customized treatment regimens for this underserved population.



**Tumor microenvironment in breast cancer metastasis: direct tissue protein profiling of tumor-associated stroma from invasive breast cancer patients with versus without axillary lymph node metastasis**

Ellsworth DL, Seeley EH, Ellsworth RE, Deyarmin B, Sanders ME, Hooke JA, Caprioli RM, Shriver CD. Windber Research Institute, Windber, PA; Vanderbilt University, Nashville, TN; Walter Reed Army Medical Center, Washington, DC.

**Background:** Interactions between metastatic breast tumors and their microenvironment likely play an important role in controlling local proliferation, tissue invasion, and metastatic spread. As regional metastasis to the axillary lymph nodes is an important step in the progression of malignant breast cancer, it is important to understand molecular events unique to the tumor microenvironment that may be associated with the metastatic process. This study used direct tissue profiling to examine protein expression in tumor-associated stroma from breast cancer patients with and without nodal metastasis to better understand the potential role of the tumor microenvironment in breast cancer metastasis.

**Methods:** All patients were enrolled in the Clinical Breast Care Project at Walter Reed Army Medical Center. Primary breast tumors were obtained from 30 patients with axillary lymph node metastasis and matched, by patient age and tumor size, to 23 patients free from nodal metastasis. Histological characterization of H&E slides served as a guide for protein expression profiling on serial tissue sections that were mounted on gold coated MALDI target plates. Matrix was deposited as individual spots (200 microns in diameter) on the tissue sections in a histology directed manner to assay specific areas of normal appearing tissue within ~1 mm of each primary carcinoma. Mass spectral data were then acquired from multiple sites of the microenvironment by laser ablation and mass spectrometry analysis. Differentially expressed proteins were defined by statistical analysis of microarray (SAM) methods using a fold-change threshold of 2X and a false discovery rate of <1%.

**Results:** High quality mass spectra were obtained from multiple locations of stroma in the immediate vicinity of the primary carcinoma. The spectra had on average ~130 identifiable protein features, from which we identified a subset of four proteins that were expressed at significantly higher levels ( $\geq 2.5$ -fold) in the microenvironment of primary tumors from node-positive patients compared to primary tumors from node-negative patients. No proteins were expressed at significantly higher levels in the microenvironment of non-metastatic primary tumors.

**Discussion:** Effective therapeutic targeting of metastatic breast cancer will require an in-depth understanding of interactions between primary tumors and surrounding stroma. Proteins highly expressed in the microenvironment of metastatic primary breast carcinomas compared to non-metastatic tumors may play an important role in the metastatic process, providing a rationale for targeting the tumor-associated microenvironment in managing metastatic breast cancer. Further elucidation of molecular mechanisms underlying tumor-stromal interactions may yield improved medical treatments for targeting breast cancer growth and metastasis.



## **Primary tumor heterogeneity and sentinel lymph node metastases: understanding molecular processes of breast cancer metastasis**

Darrell L. Ellsworth, Rachel E. Ellsworth, Heather L. Patney, Tyson E. Becker, Brenda Deyarmin, Rick M. Jordan, Jeffrey A. Hooke, Craig D. Shriver

Clinical Breast Care Project, Windber Research Institute, Windber, PA; Clinical Breast Care Project, Walter Reed Army Medical Center, Washington, DC.

**Background:** Sentinel lymph node (SLN) status is a key prognostic factor for breast cancer patients. Sentinel nodes are believed to receive early disseminating cells from the primary tumor, but it is unclear whether the timing of metastatic dissemination or heterogeneity among cell lineages in the primary tumor defines the genetic makeup and ultimately the clinical behavior of lymph node metastases. This study used allelic imbalance (AI) to examine genomic relationships among metastases in the sentinel and non-sentinel axillary lymph nodes and multiple microdissected areas of the primary carcinoma in 20 patients with lymph node positive breast cancer.

**Methods:** Heterogeneity was assessed in three dimensions by laser microdissection of 6-15 areas (each ~2.86 sq mm in size) from multiple sections of the primary tumor. Sentinel nodes were localized by standard scintigraphic and gamma probe techniques using 1.0 mCi technetium-99m sulfur colloid (Tc-99). Pathologically positive nodes were identified by H&E histology/IHC and metastases were isolated by microdissection. AI was assessed at 26 chromosomal regions and used to examine molecular mechanisms and infer timing of metastatic spread to the sentinel and axillary nodes.

### **Results:**

A high level of discordance was observed in patterns and frequencies of AI events between metastases in the sentinel versus the non-sentinel axillary nodes. Hierarchical clustering and phylogenetic analyses suggest that (1) multiple genetically-divergent metastatic cells independently colonize the lymph nodes with vastly different patterns of genomic alterations; (2) genomic heterogeneity in the primary tumor at the time of diagnosis may not fully explain patterns of molecular changes in the corresponding metastases; and (3) the genomic composition of sentinel node metastases varies considerably and may reflect timing of metastatic dissemination rather than primary tumor heterogeneity.

**Conclusions:** Genomic diversity and timing of metastatic nodal spread may be important factors in determining outcomes of breast cancer patients. Metastases successfully colonizing the sentinel nodes have different patterns of molecular alterations not completely explained by primary tumor heterogeneity, suggesting that these metastases arise at different times during disease progression. Early growth of metastases in the sentinel nodes thus may be a consequence of stimulating factors that affect proliferation of previously disseminated cells rather than the timing of metastatic spread.



**Identification of gene expression differences in primary breast tumors from node-negative and node-positive women.**

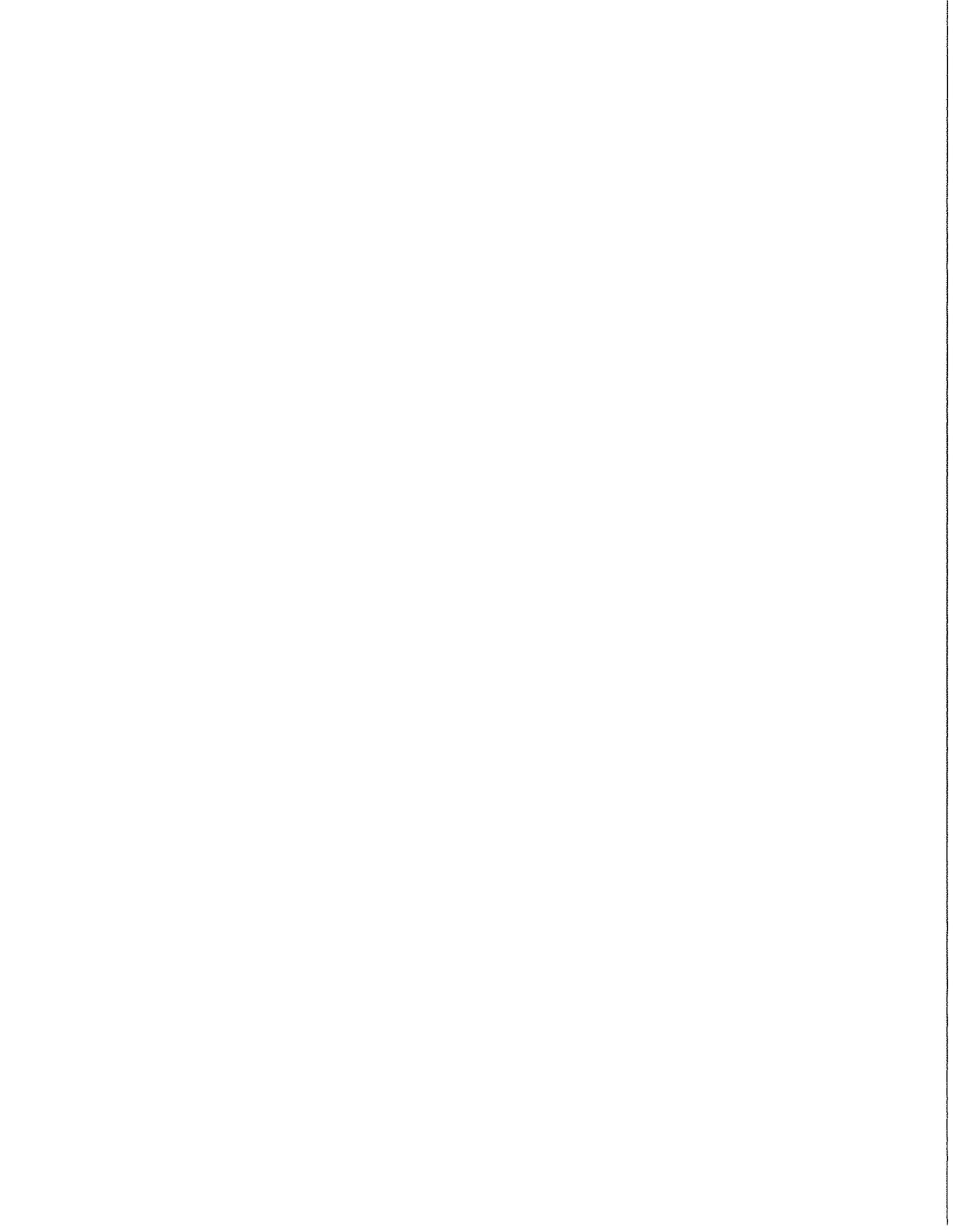
Field LA, Seebach JF, Love BJ, Deyarmin B, Kane J, Hooke JA, Ellsworth RE, Shriver CD. Windber Research Institute, Windber, PA; Invitrogen Informatics, Carlsbad, CA; Walter Reed Army Medical Center, Washington, D.C.

**Background:** One of the most useful prognostic indicators in breast cancer patients is lymph node status. Nodal status is currently determined through sentinel and axillary lymph node dissections; however, these procedures disrupt the lymphatic system and can lead to secondary complications such as lymphedema. Here, we have performed microarray analysis using RNA from node-negative and node-positive primary tumors to identify gene expression differences that may improve our understanding of the metastatic process and allow for the development of molecular tests to discriminate patients at risk for developing metastatic breast cancer.

**Methods:** Primary breast tumors from node-negative (n=20) and node-positive (n=20) women were laser microdissected to isolate pure tumor cells. Total RNA was isolated using the RNAqueous Micro kit (Ambion). RNA was amplified with two rounds of amplification using the MessageAmp II aRNA Amplification kit (Ambion) and labeled with biotin-11-UTP (PerkinElmer). Fragmented aRNA samples were hybridized to Affymetrix HG U133A 2.0 expression arrays. Differentially expressed genes were identified by Mann-Whitney testing and Bonferroni correction with p-values of 0.05 and 0.001.

**Results:** Sixty-five genes were differentially expressed in primary breast tumors from node-negative vs. node-positive women. Fifteen genes had higher expression in node-positive women. Two of these genes, GNGT1 and RASGRP1, were differentially expressed at  $p < 0.001$  and had  $\geq 4$ -fold increase in expression levels compared to node-negative women. Of the 50 genes with lower expression in primary tumors from node-positive women, 8 genes, including MUC1, had both a p-value  $< 0.001$  and  $> 4$ -fold difference in expression between the two groups.

**Discussion:** Gene expression differences identified here can discriminate those tumors from which cells have disseminated and metastasized to the lymph nodes from those without metastatic spread. The functions of the identified genes include cell proliferation, invasion, and migration, cell adhesion and cytoskeleton organization. These data suggest that primary breast tumors from patients with metastasis are genetically different from those without. These differences may prove useful in the development of molecular assays to improve risk stratification and treatment options.

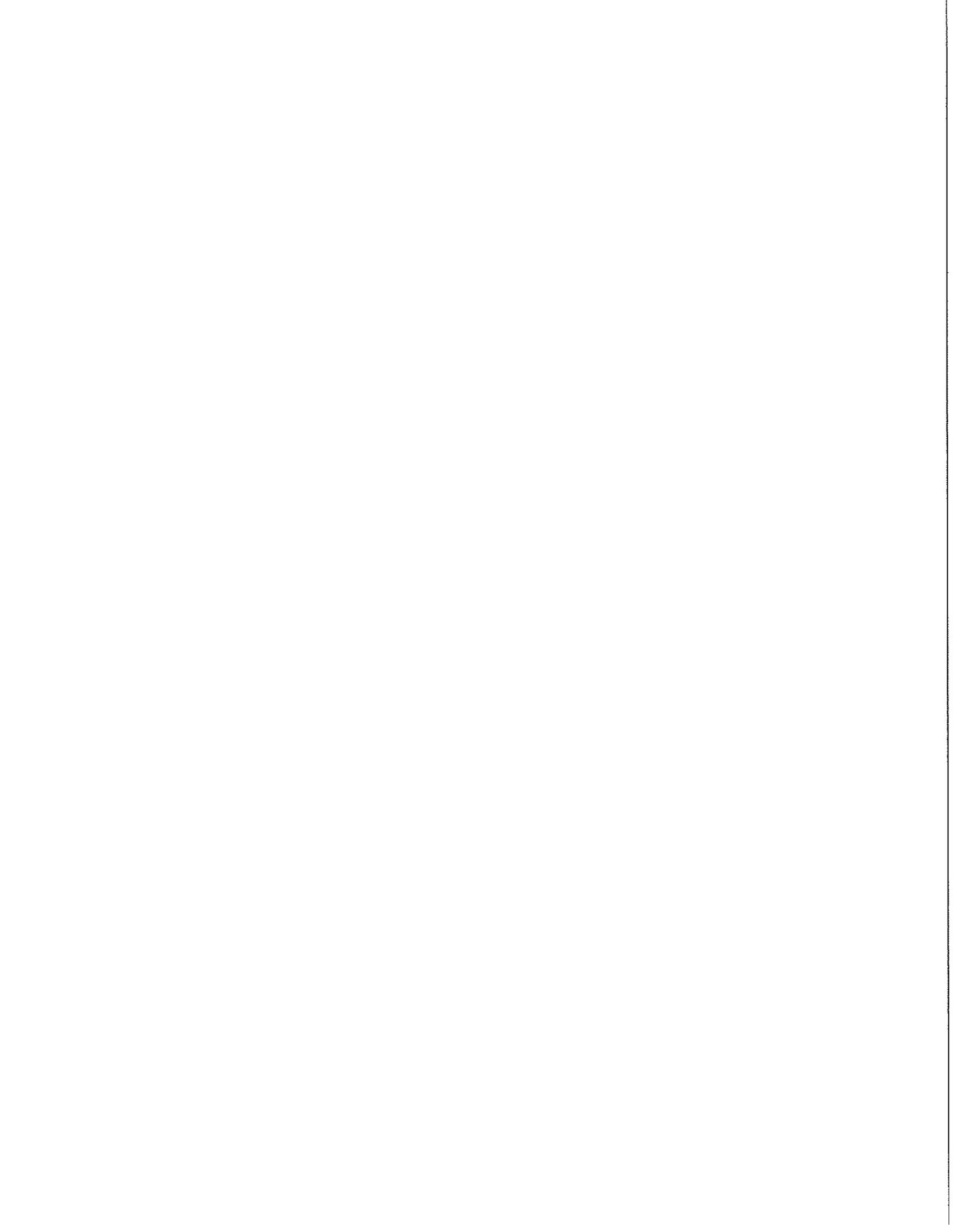


## Analysis of HER2/neu as a Diagnostic/Biomarker using ExPlain™ to examine signaling pathways and transcription factors.

Michael N. Liebman<sup>1</sup>, Brenda Deyarmin<sup>1</sup>, Craig D. Shriver<sup>3</sup>, Philip Stegmaier<sup>2</sup>, Holger Karas<sup>2</sup>, Alexander Kel<sup>2</sup> and Edgar Wingender<sup>2</sup>, (1) Windber Research Institute, 620 Seventh St. Windber, PA, USA 15963; (2) BIOBASE GmbH, Halchtersche Str. 33, D-38304 Wolfenbüttel, Germany; (3) Walter Reed Army Medical Center, Clinical Breast Care Project, Washington, DC, USA

HER2/neu expression levels are a biomarker that is used for qualifying a breast cancer patient for their potential response to herceptin. Over-expression only occurs in approximately 20% of all breast cancer patients, and in addition, only 40% of those qualified patients respond to herceptin. Clinically, there are two methods for measurement, one measuring gene copy number by fluorescent in situ hybridization (FISH), the other through the use of antibodies through immuno-histochemistry (IHC). Ideally, these independent assessments enable the accurate identification of individuals who will respond to treatment with herceptin, but this is not always the case. In a significant population of patients, there is a distinct difference between these two tests in evaluating the expression status of HER2/neu, thus leaving the oncologist with a difficult decision. In addition, although Herceptin has been established as a critical breast cancer treatment for women who over-express the HER2/neu gene, it is not a panacea without side-effects and high cost. The Clinical Breast Care Project, involving Walter Reed Army Medical Center and Windber Research Institute, utilizes both assays, and approximately 20% show lack of concurrence between FISH and IHC measurements. In a cohort of 76 patients, we observe that 16 exhibit differences in FISH and IHC, with over-expression of HER2/neu always being assessed by IHC. We have further analyzed these patients with gene expression analysis, using both Affymetrix and CodeLink platforms, in an effort to identify further correlated profiles to reconcile these differences and assist in improving prediction of response to herceptin. This study compares FISH (-)/IHC(-); FISH(+)/IHC(+); and FISH(-)/IHC(+) patients to identify additional biomarkers/diagnostics for use in improving patient treatment and outcome.

We have analyzed the gene expression data using the ExPlain™ tool (BIOBASE) in an effort to explore the relationship of transcription processes in the causal basis for HER2/neu expression. We found that FISH(+)/IHC(+) differ from FISH (-)/IHC(-) patients in regulation of expression of about 200 genes through transcription factors of PPAR, CREB, STAT, p53, Egr-1 and AP1 families in the first of group of patients, which seems are triggered by p38 and TNF-alpha signaling pathways, and through transcription factors of SMAD, ER-alpha, NF-AT and SRF families in the second group of patients, clearly associated with TGF-beta as well as calcium release pathways. We speculate that, Inosit-1,4,5-trisphosphate receptor subtype 2 (IP3R2) which is downregulated in the FISH (-)/IHC(-) group of patients and known to be functionally interacting with the anti-apoptotic protein Bcl-2 can play a key role in regulation of activity of the calcium pathway dependent transcription factors. This study can help identifying new potential biomarkers for better prediction of potential positive responders to treatment with herceptin.





**Correlation of levels and patterns of genomic instability with histological grading of DCIS**

Journal:	<i>Annals of Surgical Oncology</i>
Manuscript ID:	ASO-2007-03-0122
Manuscript Type:	Original Article
Date Submitted by the Author:	01-Mar-2007
Complete List of Authors:	Ellsworth, Rachel; Windber Research Institute, Clinical Breast Care Project Ellsworth, Darrell; Windber Research Institute, Clinical Breast Care Project Love, Brad; Invitrogen Informatics Patney, Heather; Windber Research Institute, Clinical Breast Care Project Hoffman, Laurel; Windber Research Institute, Clinical Breast Care Project Kane, Jennifer; Windber Research Institute, Clinical Breast Care Project Hooke, Jeffrey; Walter Reed Army Medical Center, Clinical Breast Care Project Shriver, Craig; Walter Reed Army Medical Center, Clinical Breast Care Project



1  
2  
3  
4  
5 Correlation of levels and patterns of genomic instability with histological grading of  
6 DCIS  
7

8 Rachel E. Ellsworth<sup>1,4</sup>, Darrell L. Ellsworth<sup>1</sup>, Brad Love<sup>2</sup>, Heather L. Patney<sup>1</sup>, Laurel R.  
9 Hoffman<sup>1</sup>, Jennifer Kane<sup>1</sup>, Jeffrey A. Hooke<sup>3</sup>, Craig D. Shriver<sup>3</sup>  
10

11  
12 <sup>1</sup>Clinical Breast Care Project, Windber Research Institute, Windber, PA, USA;  
13 <sup>2</sup>Invitrogen Informatics, Carlsbad, CA, USA; <sup>3</sup>Clinical Breast Care Project, Walter Reed  
14 Army Medical Center, Washington, DC, USA; <sup>4</sup>Henry M. Jackson Foundation for the  
15 Advancement of Military Medicine, Rockville, MD, USA  
16  
17

18 Running title:  
19 Allelic Imbalance in DCIS  
20  
21

22 Contact for page proofs and correspondence:

23 Rachel E. Ellsworth  
24 Windber Research Institute  
25 620 Seventh Street  
26 Windber, PA 15963  
27 Phone: (814) 467-9844  
28 FAX: (814) 467-6334  
29 Email: [r.ellsworth@wriwindber.org](mailto:r.ellsworth@wriwindber.org)  
30  
31  
32

33  
34 Contact for reprints:

35 Ms. Tiffani Glasco  
36 Clinical Breast Care Project  
37 Walter Reed Army Medical Center  
38 6900 Georgia Avenue, NW  
39 Washington, DC 20307  
40 Phone: (202) 782-0002  
41 [tiffani.glasco@na.amedd.army.mil](mailto:tiffani.glasco@na.amedd.army.mil)  
42  
43  
44  
45  
46  
47  
48

49 **Key words:** Allelic imbalance, ductal carcinoma *in situ*, genomic instability, histological  
50 grade, tumor classification  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Ellsworth R

**ABSTRACT**

**Background** Histological grading of ductal carcinoma *in situ* (DCIS) lesions separates DCIS into three subgroups (well-, moderate- or poorly-differentiated). It is unclear, however, whether breast disease progresses along a histological continuum or whether each grade represents a separate disease. In this study, levels and patterns of allelic imbalance (AI) were examined in DCIS lesions to develop molecular models that can distinguish pathological classifications of DCIS.

**Methods** Laser microdissected DNA samples were collected from DCIS lesions characterized by a single pathologist including well- (n=18), moderately- (n=35) and poorly-differentiated (n=47) lesions. A panel of 52 microsatellite markers representing 26 chromosomal regions commonly altered in breast cancer was used to assess patterns of AI.

**Results** The overall frequency of AI increased significantly ( $P<0.001$ ) with increasing grade (well-, 12%; moderately-, 17%; poorly-, 26%). Levels of AI were not significantly different between well- and moderate-grades of disease but were significantly higher ( $P<0.0001$ ) in poorly-differentiated compared to well-/moderately-differentiated disease. No significant differences in patterns of AI were detected between well- and moderately-differentiated disease; however, AI occurred significantly more frequently ( $P<0.05$ ) in high-grade lesions at chromosomes 6q25-q27, 8q24, 9p21, 13q14 and 17p13.1, and significantly more frequently in low-grade lesions at chromosome 16q22.3-q24.3.

**Conclusions** The inability to discriminate DCIS at the genetic level suggests that grades 1 and 2 DCIS may represent a single, non-high grade form of DCIS, while poorly-

1  
2  
3 differentiated DCIS appears to be a genetically more advanced disease that may represent  
4  
5 a discrete disease entity, characterized by a unique spectrum of genetic alterations.  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Ellsworth R

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Synopsis: Patterns of chromosomal alterations suggest that DCIS does not represent a histological continuum; rather two distinct genetic pathways are associated with either low- or high-grade disease. Grade 2 DCIS can largely be genetically characterized as either high- or low- grade.

## Introduction

Incidence of ductal carcinoma *in situ* (DCIS) has increased dramatically over the past 30 years from 1% to ~20% of all diagnosed breast cancers<sup>1</sup>. DCIS represents a heterogeneous group of preinvasive breast cancer lesions, with different histological patterns, growth rates, and cytological features. As a result, pathologists do not agree on a unified classification system for diagnosing DCIS<sup>2</sup>. Current histological grading based largely on the degree of nuclear atypia is the most widely used system for classifying *in situ* lesions<sup>3</sup>.

The inability to define universally accepted criteria for stratifying DCIS underscores the difficulties in developing standardized practices of care for treating DCIS. Options include breast-conserving surgery by wide margin excision with or without radiation and/or chemopreventive therapy, or mastectomy<sup>4</sup>. Pathological data alone may be insufficient to identify patients requiring aggressive treatment; for example, a small high-grade lesion with negative margins may not require additional intervention, while a large low-grade lesion with positive margins may require additional surgery, including mastectomy<sup>5</sup>. Thus, to avoid over-treating women with indolent disease while under-treating women with aggressive disease, new molecular tools are being evaluated to supplement pathological information to classify DCIS lesions, predict clinical outcome, and optimize treatment regimens.

Several genes have been examined as potential tools for differentiating aggressive from indolent DCIS at the molecular level. For example, amplification of the epidermal growth factor *HER2/neu*, which is associated with poor prognosis in invasive tumors, has been detected in 63% of high-grade but only 14% of low-grade DCIS<sup>6</sup>. Similarly,

Ellsworth R

1  
2  
3 amplification of the cyclin D1 gene and the frequency of p53 mutations occur at low  
4  
5 levels in low-grade DCIS and become more frequent in high-grade lesions<sup>7,8</sup>. In addition,  
6  
7 allelic imbalance (AI) studies have shown that regions on chromosomes 16q, 17p, and  
8  
9 17q are frequently altered in DCIS, suggesting that these regions may be important for  
10  
11 early tumor development<sup>9-11</sup>. Despite these observations, a molecular signature has not  
12  
13 been identified that can accurately differentiate aggressive from indolent disease.  
14  
15

16  
17 In this study, we assayed levels and patterns of AI at 26 chromosomal regions  
18  
19 frequently altered in breast cancer in 100 well-characterized DCIS samples. Our  
20  
21 objectives were to (1) identify genomic patterns of AI that correlate with important  
22  
23 clinical characteristics of DCIS lesions, and (2) develop molecular models that can  
24  
25 distinguish pathological classifications of DCIS.  
26  
27

## 28 29 30 31 **Materials and methods**

32  
33 Paraffin-embedded pure DCIS specimens (n=100) with no evidence of an  
34  
35 invasive component were obtained from the Windber Medical Center (WMC) and  
36  
37 Memorial Medical Center (MMC) Pathology Departments or from the Clinical Breast  
38  
39 Care Project (CBCP) Pathology Laboratory. Samples from WMC and MMC (n=32) were  
40  
41 archival in nature and anonymized with no links between the assigned research number  
42  
43 and patient identifiers. Clinical information including age at diagnosis, clinical outcome,  
44  
45 and pathological details were provided anonymously by the MMC Cancer Registry.  
46  
47 Tissue and blood samples from CBCP patients (n=68) were collected with approval from  
48  
49 the Walter Reed Army Medical Center Human Use Committee and Institutional Review  
50  
51 Board. All subjects enrolled in the CBCP voluntarily agreed to participate and gave  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 written informed consent. Clinical information was provided for all CBCP samples using  
4 questionnaires designed by and administered under the auspices of the CBCP.  
5  
6

7  
8 Diagnoses of all samples were made by the CBCP pathologist (JAH) from  
9  
10 hematoxylin and eosin (H&E) stained slides. DCIS was categorized according to nuclear  
11 grade as recommended by the 1997 Consensus Conference<sup>2</sup>. Specifically, low-grade  
12 (grade 1) nuclei had a monomorphic, rounded appearance, 1.5 to 2 times the size of a  
13 normal duct epithelial cell nucleus, finely dispersed chromatin, and only occasional  
14 nucleoli and mitoses. High-grade (grade 3) nuclei were markedly pleomorphic, more than  
15 1.5 to 2 times the size of a normal duct epithelial cell nucleus, with coarse, vesicular  
16 chromatin, prominent nucleoli, and conspicuous mitoses. Intermediate grade (grade 2)  
17 nuclei exhibited features between low-grade and high-grade. In addition to nuclear grade,  
18 the architectural pattern (cribriform, micropapillary, solid, comedo, papillary, clinging)  
19 and presence or absence of necrosis were recorded. Estrogen (ER) and progesterone  
20 receptor (PR) status was determined by immunohistochemistry analysis (MDR Global,  
21 Windber, PA) and HER2 status assessed by FISH using the PathVysion kit (Abbott  
22 Laboratories, Abbott Park, IL). Clinical information for all samples is provided in Table  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42 1.

43  
44 DNA was extracted from homogeneous populations of DCIS cells following  
45 laser-assisted microdissection on an *ASLMD* laser microdissection system (Leica  
46 Microsystems, Wetzlar, Germany) (Figure 1)<sup>12</sup> All microdissected sections were  
47 examined by the CBCP pathologist, who identified and marked regions of DCIS before  
48 microdissection; when both high grade and non-high grade lesions were present within  
49 the same section, a single type of lesion was selected for microdissection. The integrity of  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Ellsworth R

1  
2  
3 multiple serial sections was established by pathological verification of the first and last  
4  
5 sections stained with H&E. Referent DNA was obtained from blood clots using Clotspin  
6  
7 and Puregene DNA purification kits (Gentra Systems, Minneapolis, MN) or from  
8  
9 microdissected disease-free skin or negative lymph node tissues.  
10  
11

12  
13 A panel of 52 microsatellite markers was amplified, genotyped, and analyzed as  
14  
15 previously described<sup>13-15</sup>. Samples with a normalized ratio of  $\leq 0.35$  were considered to  
16  
17 show genomic instability, which indicates that a substantial proportion of cells in that  
18  
19 sample contained the same chromosomal aberration compared to normal somatic cells. A  
20  
21 normalized ratio of  $\leq 0.35$  can result from a deletion (loss of heterozygosity) or increase  
22  
23 in copy number (allelic amplification) at a chromosomal region (Figure 2). Because each  
24  
25 chromosomal region was represented by two polymorphic markers, genomic instability in  
26  
27 each region was classified as previously described<sup>16</sup>.  
28  
29  
30

31  
32 Analysis of variance (ANOVA) was used to test for homogeneity in AI  
33  
34 frequencies among histological grades. Mann-Whitney tests using median frequencies  
35  
36 were then used in pair-wise comparisons to identify significant differences in AI  
37  
38 frequencies between histological classifications. Fisher's Exact Test was used to detect  
39  
40 differences in AI frequency at each chromosomal region. A *P*-value  $< 0.05$  was  
41  
42 considered statistically significant in all analyses.  
43  
44  
45  
46  
47

## 48 Results

### 49 Clinical and Pathological Characteristics

50  
51 All samples were collected from female patients with an average age at diagnosis  
52  
53 of 59.7 years (SD = 12.7); 30% of patients were pre-menopausal ( $< 50$  years of age).  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4 Pathological diagnoses showed that 18% of samples were grade 1, 35% were grade 2, and  
5  
6 47% were grade 3. Lumpectomies were performed in 71% of patients while 29% of  
7  
8 patients underwent mastectomies. Three individuals had bilateral mastectomies after a  
9  
10 diagnosis of DCIS in one breast. The frequency of mastectomy was significantly higher  
11  
12 in the archival samples ( $P<0.01$ ) than in CBCP samples, likely reflecting changes in  
13  
14 surgical practices over the last 15 years. To date, all patients are alive; three (3%) have  
15  
16 had cancer recurrence.  
17

18  
19  
20 ER, PR and HER2/*neu* status were determined for all patients where sufficient *in*  
21  
22 *situ* lesions were available. Overall, 83% of specimens were ER positive, 71% were PR  
23  
24 positive, and 40% of specimens showed either HER2/*neu* gene amplification or  
25  
26 aneusomy for chromosome 17. When the data were stratified by grade, the frequency of  
27  
28 ER and PR negativity and HER2/*neu* amplification increased with increasing grade  
29  
30 (Table 2). The differences in PR ( $P<0.01$ ) and HER2 ( $P<0.001$ ) status were significantly  
31  
32 higher in grade 3 than grade 1 patients, with grade 2 lesions having histological  
33  
34 characteristics between grades 1 and 3.  
35  
36  
37  
38  
39  
40

#### 41 Overall AI Frequencies

42  
43 Within the 100 lesions studied, 485 chromosomal alterations were detected. The  
44  
45 frequency of AI at each locus varied widely, with the most frequent AI (45%) seen on  
46  
47 chromosome 16q11-q22 and the least frequent AI (1%) on chromosome 1p36. The  
48  
49 average AI frequency was 12% (range 0 – 25%) in grade 1, 17% (range 0-52%) in grade  
50  
51 2, and 26% (range 0 – 59%) in grade 3 lesions. Mann-Whitney tests for pair-wise  
52  
53 comparisons showed that the level of AI in grade 3 carcinomas was significantly higher  
54  
55  
56  
57  
58  
59  
60

Ellsworth R

1  
2  
3 than that in grade 1 ( $P<0.0001$ ) or grade 2 ( $P<0.01$ ) lesions. Stratification of samples into  
4 high- (grade 3) and non-high-grade (grades 1 and 2) revealed similar results, with  
5  
6 significantly higher levels of AI in high-grade lesions ( $P<0.0001$ ). Levels of AI did not  
7  
8 differ significantly between grades 1 and 2. Although overall levels of AI were higher in  
9  
10 samples with central necrosis (21%) compared to those without (16%), this difference  
11  
12  
13 was not significant.  
14  
15  
16  
17  
18

### 19 **Differential patterns of genomic instability between grades 1 and 3 DCIS**

20  
21 To determine whether low- and high-grade DCIS are characterized by different genetic  
22  
23 alterations, the frequency of AI at each chromosomal region was examined in grade 1 and  
24  
25 3 lesions (Table 2). The frequency of AI events increased from grade 1 to grade 3 DCIS  
26  
27 at 23 of 26 chromosomal regions, with AI occurring significantly more frequently at  
28  
29 chromosomes 6q25-q27, 8q24, 9p21, 13q14, and 17p13.1 in grade 3 lesions. AI was not  
30  
31 detected in any grade 1 lesions at chromosomes 1p36, 9p21, 13q14 or 14q31-q32. Levels  
32  
33 of AI were significantly higher at 16q22.3-q24.3 in grade 1 compared to grade 3 DCIS.  
34  
35  
36  
37  
38  
39  
40

### 41 **Comparison of genomic instability patterns in grade 2 compared to grades 1 and 3**

#### 42 **DCIS**

43  
44 The highest frequency of AI in grade 2 DCIS occurred at chromosome 16q11-q22 (41%).  
45  
46 There were no significant patterns of AI that could distinguish grade 2 DCIS as  
47  
48 genetically different from either grades 1 or 3. Patterns of AI did not differ significantly  
49  
50 between grades 1 and 2 for any of the chromosomal regions. As with grade 1 DCIS,  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 levels of AI were significantly higher in grade 3 compared to grade 2 DCIS at  
4  
5 chromosomes 6q25-q27, 8q24 and 9p21.  
6  
7

8 Because grades 1 and 2 did not have significantly different patterns of AI, data  
9  
10 from grades 1 and 2 were pooled into a non-high grade DCIS group and compared to  
11  
12 grade 3 DCIS. When pooled, AI occurred significantly more frequently at chromosomes  
13  
14 6q25-q27, 8q24, 9p21, and 13q14.  
15  
16  
17  
18

### 19 20 **Disease Prediction Modeling** 21

22 Chromosomal alterations at 6q25-q27, 8q24, 9p21, 13q14 and 17p13.1 were  
23  
24 highly informative in discriminating high- from non-high grade lesions. Over 80% of  
25  
26 grade 1 DCIS did not have AI at any of these chromosomal regions; the three samples  
27  
28 with AI showed instability at a single region and none showed AI at chromosomes 9p21  
29  
30 or 13q14. Grade 2 DCIS was split equally between specimens with no AI at any of these  
31  
32 critical regions, and those with AI at at least one region. In contrast, 86% of grade 3  
33  
34 DCIS had AI at at least one critical region; three (7%) grade 3 lesions had AI at  
35  
36 chromosome 16q22.3-q24.3 without concomitant AI at 6q25-q27, 8q24, 9p21, 13q14 or  
37  
38 17p13.1.  
39  
40  
41  
42  
43  
44  
45

### 46 **Discussion** 47

48 Despite advances in breast cancer screening and early diagnosis, which have led  
49  
50 to a significant increase in the detection of preinvasive lesions, the heterogeneous clinical  
51  
52 behavior of DCIS has made it difficult to develop a standard treatment regimen.  
53  
54  
55 Currently, clinical management of DCIS is largely case specific; patients may receive  
56  
57  
58  
59  
60

Ellsworth R

1  
2  
3 surgical interventions ranging from excisional biopsy to mastectomy. Risk of recurrence  
4 in DCIS also varies widely among patients, ranging from 10–40%. Thus, a critical need  
5 in translational medicine is to utilize molecular information to identify why certain  
6 preinvasive lesions behave more aggressively than others.  
7  
8  
9  
10  
11

12 Breast cancer development has been characterized as progression along a non-  
13 obligatory histological continuum from ADH to low-, moderate- and high-grade DCIS to  
14 invasive cancer. Previous work in our laboratory has shown that overall levels of AI in  
15 DCIS did not differ significantly from levels in invasive breast carcinomas, but were  
16 significantly higher than in ADH<sup>17</sup>. In this study, stratification of DCIS lesions by  
17 histological grade revealed significantly higher levels of genomic instability in high-  
18 versus non-high grade DCIS (Figure 3). Under a model of linear progression, these data  
19 suggest that critical periods in breast cancer development include the transitions from  
20 ADH to non-high grade DCIS and from non-high to high-grade DCIS.  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32

33 An alternative to the linear progression model is one in which each grade of DCIS  
34 represents a single disease entity, developing along discrete genetic pathways<sup>18</sup>. Models  
35 of disease progression have been developed that describe three distinct genetic pathways  
36 correlated with the three grades of DCIS disease<sup>11,19</sup>. In these models, loss of  
37 chromosome 16q was detected significantly more frequently in low- and intermediate-  
38 grade compared to high-grade DCIS. AI at chromosome 16q11-q22.1 did not differ  
39 significantly between high- (50%) and low-grade (56%) DCIS, representing the most  
40 frequently altered chromosomal region for both high- and low-grade disease. In contrast,  
41 AI at chromosome 16q22.3-q24.3 occurred significantly more frequently in low-(56%)  
42 compared to high-grade (29%) DCIS. Localization of the critical region that  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 differentiates low- from high-grade DCIS to 16q22-q24 supports earlier reports<sup>11,19</sup>, and  
4  
5 suggests that alterations of the more proximal region of chromosome 16, present in ~50%  
6  
7 of lesions studied here, may be a common and early alteration in the development of *in*  
8  
9 *situ* carcinoma.

10  
11  
12 In addition to loss of chromosome 16q, amplification of chromosomes 11q13 and  
13  
14 17q12 and deletion of 17p have been associated with intermediate- and high-grade  
15  
16 DCIS<sup>11,20,21 11,19-21</sup>. In this study, chromosome 17p13.1 was altered significantly more  
17  
18 frequently in high- (39%) than low-grade (7%) DCIS; mutations in the tumor protein, p53  
19  
20 gene on chromosome 17p13.1 have been associated with increased genomic instability in  
21  
22 breast tumors<sup>22</sup>, and increased from low-grade DCIS (0%) through intermediate-  
23  
24 (4.35%) to high-grade DCIS (40%)<sup>7</sup>. Thus, alterations of the p53 gene may contribute to  
25  
26 the differences in pathology and behavior of high-grade lesions.

27  
28  
29 In addition to AI at 17p13.1, alterations at 6q25-q27, 8q24, 9p21 and 13q14 were  
30  
31 effective at discriminating low- from high-grade DCIS (Figure 4). A number of  
32  
33 candidate genes that may be biologically involved in disease development have been  
34  
35 identified in these regions. For example, the insulin-like growth factor 2 receptor  
36  
37 (IGF2R) on chromosome 6q26 has shown deletions in ~75% of poorly-differentiated  
38  
39 DCIS<sup>23</sup>, and amplification of the V-MYC oncogene (MYC) on chromosome 8q24 has  
40  
41 been associated with more aggressive behavior in DCIS<sup>24</sup>. AI at 9p21 and 13q14 are  
42  
43 especially interesting in this study because imbalance at either region was detected in  
44  
45 >50% of high-grade DCIS while no AI events were detected at these loci in any low-  
46  
47 grade lesions. Within the 13q14 region is the potential candidate gene retinoblastoma  
48  
49 (RB1). Underexpression of RB1 has been correlated with high-grade invasive breast  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Ellsworth R

1  
2  
3 carcinomas and increased cellular proliferation<sup>25</sup>. Another tumor suppressor gene,  
4 cyclin-dependent kinase inhibitor 2A (TP16) on chromosome 9p21 is also involved in  
5  
6 cell cycle control and loss of the ability of p16 to induce cell cycle arrest has been  
7  
8 demonstrated in cells lacking functional RB1<sup>26</sup>.  
9  
10

11  
12 Differences patterns of chromosomal alterations at 6q25-q27, 8q24, 9p21, 13q14  
13 and 17p13.1 as well as 16q22-q24 in high- and low-grade lesions suggest that DCIS is  
14  
15 not a single disease, but represents different pathological subtypes that develop along  
16  
17 unique genetic pathways. At the clinicopathological level, intermediate-grade lesions  
18  
19 have characteristics between high- and low-grade DCIS, while at the genetic level, grade  
20  
21 2 DCIS usually resembled low-grade DCIS, with no significant differences in patterns of  
22  
23 AI compared to low-grade disease. Patterns of AI at the six critical regions categorized  
24  
25 intermediate-grade samples as genetically high-(35%) or low-grade (50%), with only  
26  
27 15% having imbalance patterns characteristic of both genetic pathways. The ability to  
28  
29 genetically discriminate grade 2 DCIS into high- or low-grade categories, coupled with  
30  
31 the finding that there were no patterns of AI unique to grade 2 disease, suggests that  
32  
33 intermediate-grade DCIS represents the phenotypic extremes of low- or high-grade  
34  
35 disease, rather than a third subtype of *in situ* carcinoma.  
36  
37  
38  
39  
40  
41  
42

43  
44 In conclusion, patterns of AI appear to support a model in which DCIS develops  
45  
46 along two discrete genetic pathways. Physical loss of chromosome 16q, especially in the  
47  
48 16q22-q11 region, appears critical to the development of well-differentiated DCIS, while  
49  
50 AI at 6q25-q27, 8q24, 9p21, 13q14, and 17p13.1 were associated with the development  
51  
52 of poorly-differentiated lesions. Grade 2 DCIS likely represent a blend of high- and low-  
53  
54 grade DCIS, because 85% of grade 2 carcinomas could be categorized as high- or low-  
55  
56  
57  
58  
59  
60

1  
2  
3 grade based on the underlying genetic profile. Identification of these molecular profiles  
4  
5 not only improves our biological understanding of DCIS but may allow pathologists to  
6  
7 more accurately characterize these lesions, resulting in more personalized and appropriate  
8  
9 treatments directed at the specific form of breast disease.  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Ellsworth R

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**Acknowledgements**

The authors thank Dr. John Yerger for generously granting access to archived samples, and Sue Lubert for assistance in genotyping. This research was supported by a grant from Susan G. Komen for the Cure.

## References

1. Greenlee RT, Murray T, Bolden S, Wingo PA. Cancer Statistics, 2000. *CA Cancer J Clin* 2000; 50:7-33.
2. The Consensus Conference Committee. Consensus conference on the classification of ductal carcinoma in situ. *Hum Pathol* 1997; 28:1221-1225.
3. Tavassoli FA, Hoeffler H, Rosai J, Holland R, Ellis IO, Schnitt SJ, Boecker W, Heywang-Kobrunner SH, Moinfar F, Lakhani SR. Intraductal proliferative lesions. In Tavassoli FA, Devilee P (eds). *WHO Classification: Tumors of the breast and female genital organs*. Lyon: IARC Press; 2003:63-80.
4. Schwartz GF. The current treatment of ductal carcinoma in situ. *The Breast Journal* 2001; 7:308-310.
5. Skinner KA, Silverstein MJ. The management of ductal carcinoma in situ of the breast. *Endocrine-Related Cancer* 2001; 8:33-45.
6. Ho GH, Calvano JE, Bisogna M, Borgen PI, Rosen PP, Tan LK, Van Zee KJ. In microdissected ductal carcinoma in situ, HER-2/neu amplification, but not p53 mutation, is associated with high nuclear grade and comedo histology. *Cancer* 2000; 89:2153-2160.
7. Done SJ, Eskandarian S, Bull S, Redston M, Andrulis IL. p53 missense mutations in microdissected high-grade ductal carcinoma in situ of the breast. *J Natl Cancer Inst* 2001; 93:700-704.
8. Simpson JL, Quan DE, O'Malley F, Odom-Maryon T, Clarke PE. Amplification of CCND1 and expression of its protein product, cyclin D1, in ductal carcinoma in situ of the breast. *Am J Pathol* 1997; 151:161-168.
9. Allred DC, Mohsin SK, Fuqua SAW. Histological and biological evolution of human premalignant breast disease. *Endocrine-Related Cancer* 2001; 8:47-61.
10. Fujii H, Szumel R, Marsh C, Zhou W, Gabrielson E. Genetic progression histologic grade and allelic loss in ductal carcinoma in situ of the breast. *Cancer Research* 1996; 56:5260-5265.
11. Vos CB, ter Haar NT, Rosenberg C, Peterse JL, Cleton-Jansen AM, Cornelisse CJ, van de Vijer MJ. Genetic alteration on chromosome 16 and 17 are important features of ductal carcinoma in situ of the breast and are associated with histologic type. *Br J Cancer* 1999; 81:1410-1418.
12. Ellsworth DL, Shriver CD, Ellsworth RE, Deyarmin B, Somiari RI. Laser capture microdissection of paraffin-embedded tissues. *Biotechniques* 2003; 34:42-46.

Ellsworth R

- 1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60
13. Ellsworth DL, Ellsworth RE, Love B, Deyarmin B, Lubert SM, Mittal V, Hooke JA, Shriver CD. Outer breast quadrants demonstrate increased levels of genomic instability. *Annals Surg Onc* 2004; 11:861-868.
14. Ellsworth RE, Ellsworth DL, Lubert SM, Hooke J, Somiari RI, Shriver CD. High-throughput loss of heterozygosity mapping in 26 commonly deleted regions in breast cancer. *Cancer Epidemiol Biomarkers Prev* 2003; 12:915-919.
15. Medintz IL, Lee C-CR, Wong WW, Pirkola K, Sidransky D, Mathies RA. Loss of heterozygosity assay for molecular detection of cancer using energy-transfer primers and capillary array electrophoresis. *Genome Res* 2000; 10:1211-1218.
16. Ellsworth RE, Ellsworth DL, Neatrour DM, Deyarmin B, Lubert SM, Sarachine MJ, Brown P, Hooke JA, Shriver CD. Allelic imbalance in primary breast carcinomas and metastatic tumors of the axillary lymph nodes. *Mol Cancer Res* 2005; 3:71-77.
17. Ellsworth RE, Ellsworth DL, Deyarmin B, Hoffman LR, Love B, Hooke JA, Shriver CD. Timing of critical genetic changes in human breast disease. *Annals Surg Onc* 2005; 12:1054-1060.
18. Hwang ES, DeVries S, Chew KL, Moore II DH, Kerlikowske K, Thor A, Ljung B-M, Waldman FM. Patterns of chromosomal alterations in breast ductal carcinoma in situ. *Clin Cancer Res* 2004; 10:5160-5167.
19. Buerger H, Otterbach F, Simon R, Poremba C, Diallo R, Decker T, Riethdorf L, Brinkschmidt C, Dockhorn-Dworniczak B, Boecker W. Comparative genomic hybridization of ductal carcinoma in situ of the breast - evidence of multiple genetic pathways. *J Pathol* 1999; 187:396-402.
20. O'Connell P, Pekkel V, Fuqua SAW, Osborne CK, Allred DC. Analysis of loss of heterozygosity in 399 premalignant breast lesions at 15 genetic loci. *J Natl Cancer Inst* 1998; 90:697-703.
21. Ando Y, Iwase H, Ichihara S, Toyoshima S, Nakamura T, Yamashita H, Toyama T, Omoto Y, Karamatsu S, Mitsuyama S, Fujii Y, Kobayashi S. Loss of heterozygosity and microsatellite instability in ductal carcinoma in situ of the breast. *Cancer Lett* 2000; 156:207-214.
22. Livingstone LR, White A, Sprouse J, Livanos E, Jacks T, Tlsty TD. Altered cell cycle arrest and gene amplification potential accompany loss of wild-type p53. *Cell* 1992; 70:923-935.
23. Chappell SA, Walsh T, Walker RA, Shaw JA. Loss of heterozygosity at the mannose 6-phosphate insulin-like growth factor 2 receptor gene correlates with poor differentiation in early breast carcinomas. *Br J Cancer* 1997; 76:1558-1561.

- 1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60
24. Aulmann S, Bentz M, Sinn HP. C-myc oncogene amplification in ductal carcinoma in situ of the breast. *Breast Cancer Res Treat* 2002; 74:25-31.
  25. Bieche I, Lidreau R. Loss of heterozygosity at 13q14 correlates with RB1 gene underexpression in human breast cancer. *Mol Carcinog* 2000; 29:151-158.
  26. Lukas J, Parry D, Aagaard L, Mann DJ, Bartkova J, Strauss M, Peters G, Bartek J. Retinoblastoma-protein-dependent cell-cycle inhibition by the tumour suppressor p16. *Nature* 1995; 375:503-506.

### Figure Legends

**Figure 1** Hemotoxylin and eosin stained sections showing specificity of laser microdissection in obtaining homogeneous populations of DCIS cells. (a-b) grade 1, (c-d) grade 2, (e-f) grade 3.

**Figure 2** Detection of allelic imbalance in grade 3 DCIS with fluorescence-based genotyping. Alleles for marker D17S1298 on chromosome 17q13.3 were detected as fluorescent peaks in (a) microdissected breast tumor DNA and (b) referent DNA. A normalized peak height ratio of 0.23 was calculated for the tumor sample using the following peak heights in relative fluorescence units (rfu): tumor DNA – 17,459 rfu and 4,002 rfu, referent DNA – 34,728 rfu and 34,097 rfu.

**Figure 3** Genomic instability (% informative markers showing AI) across a continuum of histological diagnoses in breast disease. Significant differences ( $P < 0.01$ ) in AI were observed between ADH and grade 1 DCIS and between grade 2 and grade 3 DCIS.

**Figure 4** Model of genetic pathways defining DCIS. Deletion of chromosome 16q and 13p14 have been established in previous studies as critical genetic changes discriminating grade 1 from grade 3 disease. For the other chromosomal regions, AI analysis cannot distinguish chromosomal amplifications from deletions and thus are depicted without reference to the nature of the genetic change.

Table 1. Clinical and pathological characteristics of 100 DCIS lesions

Characteristics	Grade 1	Grade 2	Grade 3	P-value <sup>a</sup>
Age (years)				
≤50	28%	26%	35%	0.7667
>50	72%	74%	65%	
Central Necrosis				
Present	39%	87%	93%	0.0000
Absent	61%	13%	7%	
Distribution				
Focal <sup>b</sup>	60% (0.63 cm)	39% (0.63 cm)	44% (1.54 cm)	0.4756
Multifocal	40%	61%	56%	
Hormonal Status				
ER+/PR+	79%	77%	39%	0.0126 <sup>c</sup>
ER+/PR-	8%	13%	23%	
ER-/PR+	7%	3%	2%	
ER-/PR-	6%	7%	36%	
HER2 Status <sup>d</sup>				
HER2+	7%	38%	74%	0.0000
HER2-	93%	62%	26%	
Surgical Treatment				
Lumpectomy <sup>e</sup>	77%	71%	64%	0.3748
Mastectomy	23%	29%	36%	

<sup>a</sup>The P-value listed here compares clinical characteristics in grade 1 compared to grade 3 DCIS.

<sup>b</sup>The average size of single lesions is listed in parenthesis for each grade

<sup>c</sup>P-value for hormone receptor positive (ER and/or PR) compared to hormone receptor negative

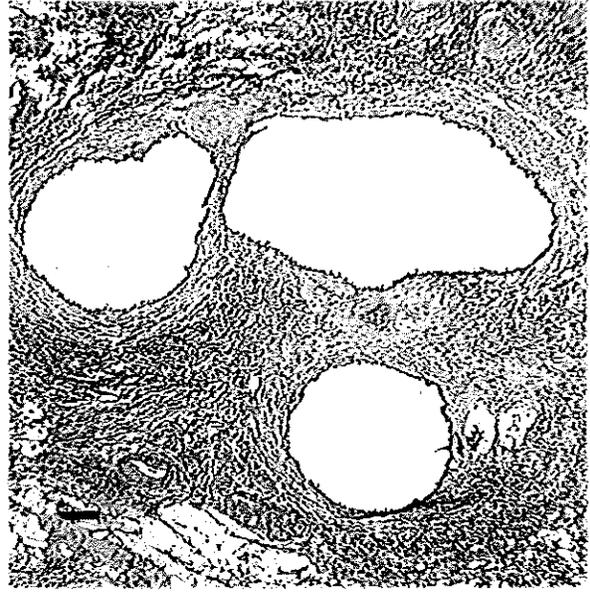
<sup>d</sup>HER2 status was determined by FISH

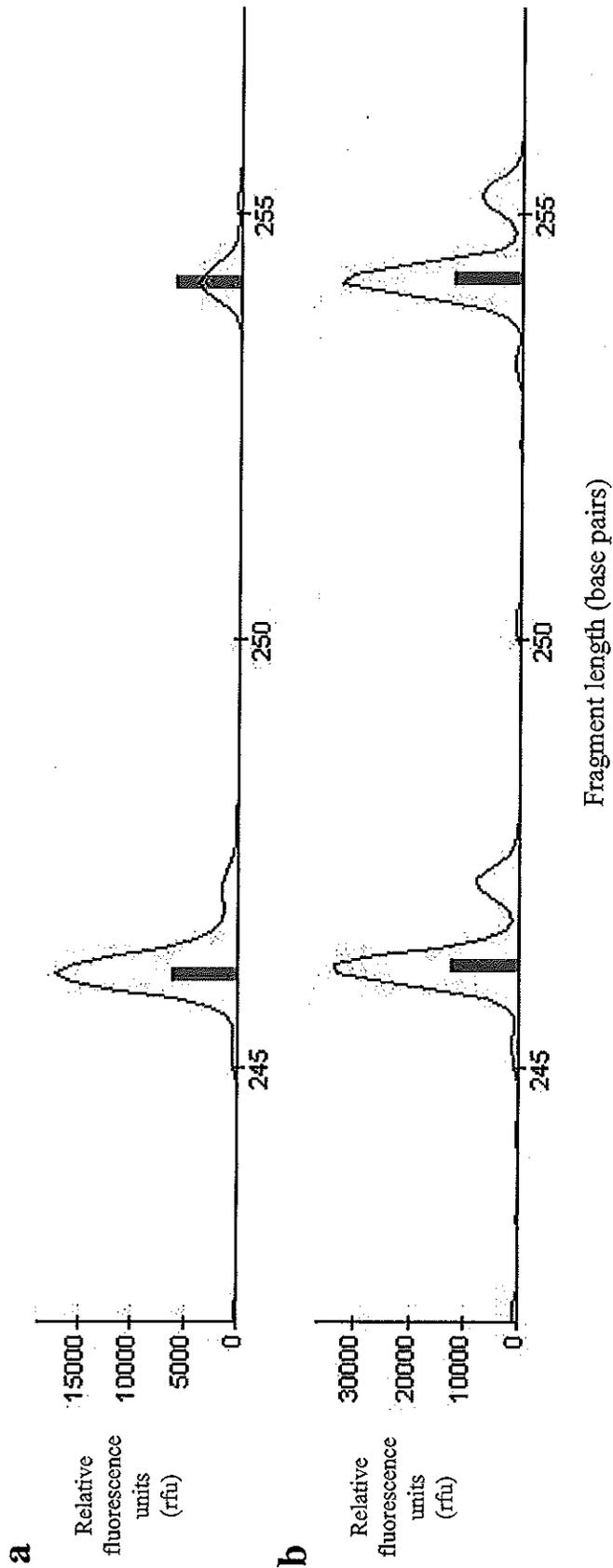
<sup>e</sup>Includes incisional biopsy, excisional biopsy, and re-excision

Ellsworth R

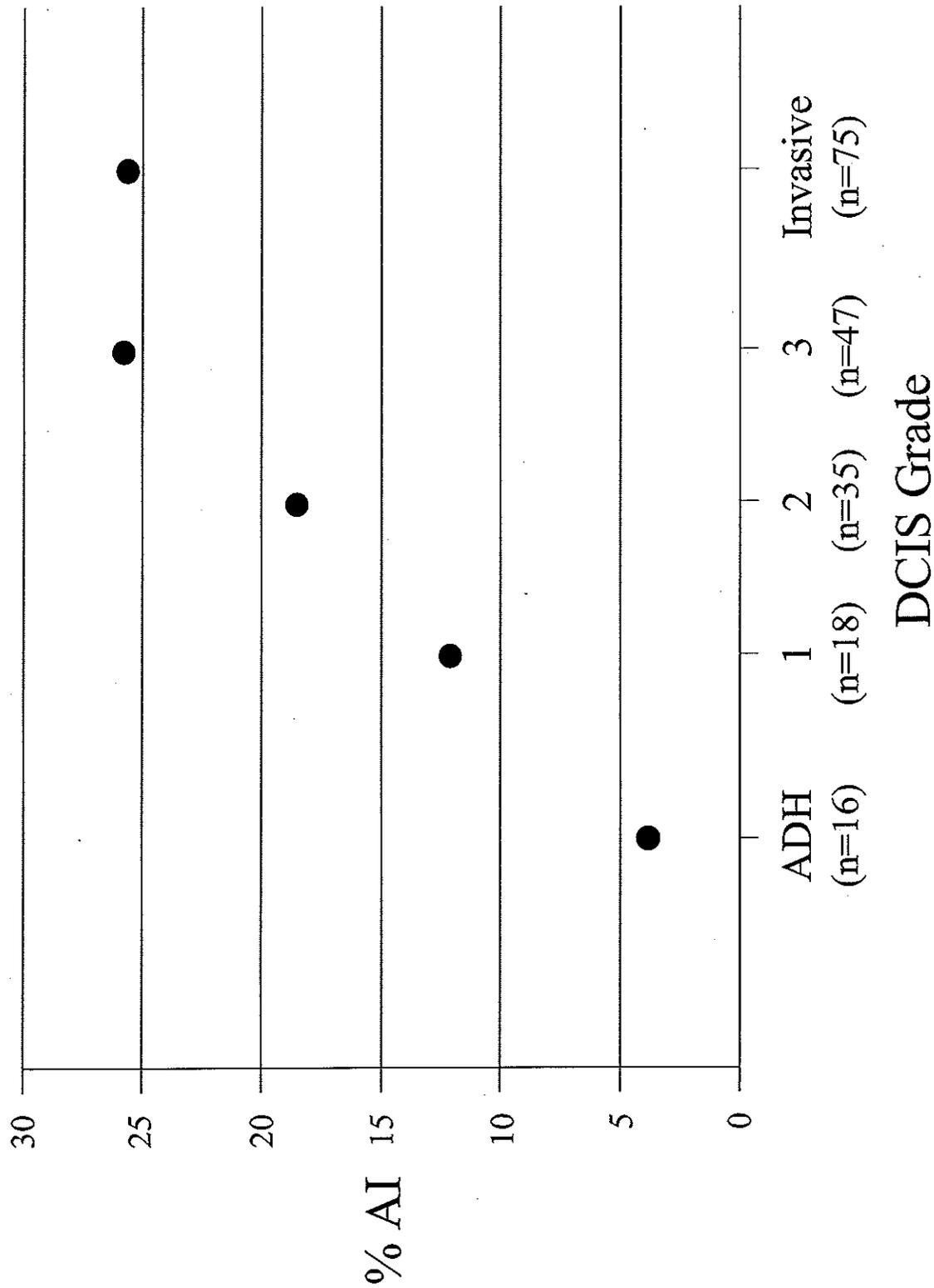
Table 2. Frequency of AI by grade at 26 chromosomal regions. Numbers in bold are statistically significant.

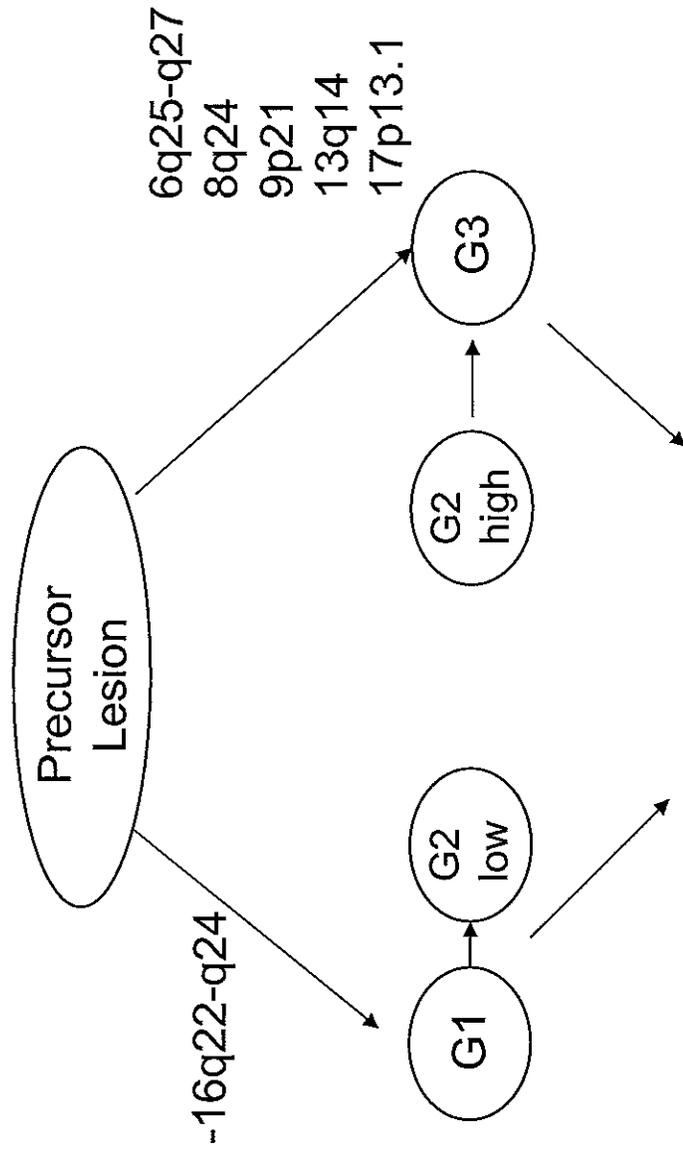
Chromosomal Region	Grade 1	Grade 2	Grade 3	P 1 vs.3	P 1 vs. 2	P 1/2 vs. 3
1p36.1-p36.2	0	0	0.04	1	1	0.2124
2q21.3-23.3	0.06	0.13	0.13	0.6655	0.6463	0.7593
3p14.1	0.13	0.27	0.20	0.7104	0.3002	1
5q21.1-q21.3	0.12	0.06	0.18	.7110	0.5495	0.1959
6q15	0.12	0.09	0.21	0.4836	1	0.1576
6q22.1-q23.1	0.12	0.18	0.28	0.3165	0.7039	0.2177
6q25.2-q27	0.06	0.12	0.29	<b>0.0489</b>	0.6477	<b>0.0285</b>
7q31.1-q31.31	0.17	0.06	0.10	0.5986	0.2578	1
8p22-p21.3	0.07	0.15	0.32	0.0847	0.6518	<b>0.0254</b>
8q24	0.06	0.16	0.38	<b>0.0133</b>	0.3934	<b>0.0048</b>
9p21	0	0.07	0.30	<b>0.0131</b>	0.5264	<b>0.0034</b>
10q23.31-q23.33	0.12	0.07	0.21	0.7123	0.6262	0.1394
11p15	0.07	0.13	0.14	0.6622	1	0.7499
11q13.1	0.06	0.13	0.22	0.2594	0.6484	0.1641
11q23	0.15	0.31	0.19	1	0.3247	0.3728
13q12.3	0.19	0.19	0.32	0.5181	1	0.2279
13q14.2-q14.3	0	0.23	0.35	<b>0.0033</b>	0.0798	<b>0.0308</b>
14q32.11-q31	0	0.15	0.15	0.1747	0.1555	0.5407
16q11.2-q22.1	0.56	0.29	0.50	0.7733	0.1171	0.2989
16q22.3-q24.3	0.59	0.40	0.32	<b>0.0416</b>	0.2407	0.2050
17p13.3	0.23	0.29	0.44	0.3237	1	0.1599
17p13.1	0.07	0.27	0.39	<b>0.0239</b>	0.2344	0.0647
17q12-q21	0.13	0.15	0.32	0.1924	1	<b>0.0486</b>
18q21.1-q21.3	0.13	0.15	0.33	0.1959	1	<b>0.0498</b>
22q12.3	0.12	0.16	0.15	1	1	1
22q13.1	0.06	0.23	0.25	0.1513	0.2300	0.4403



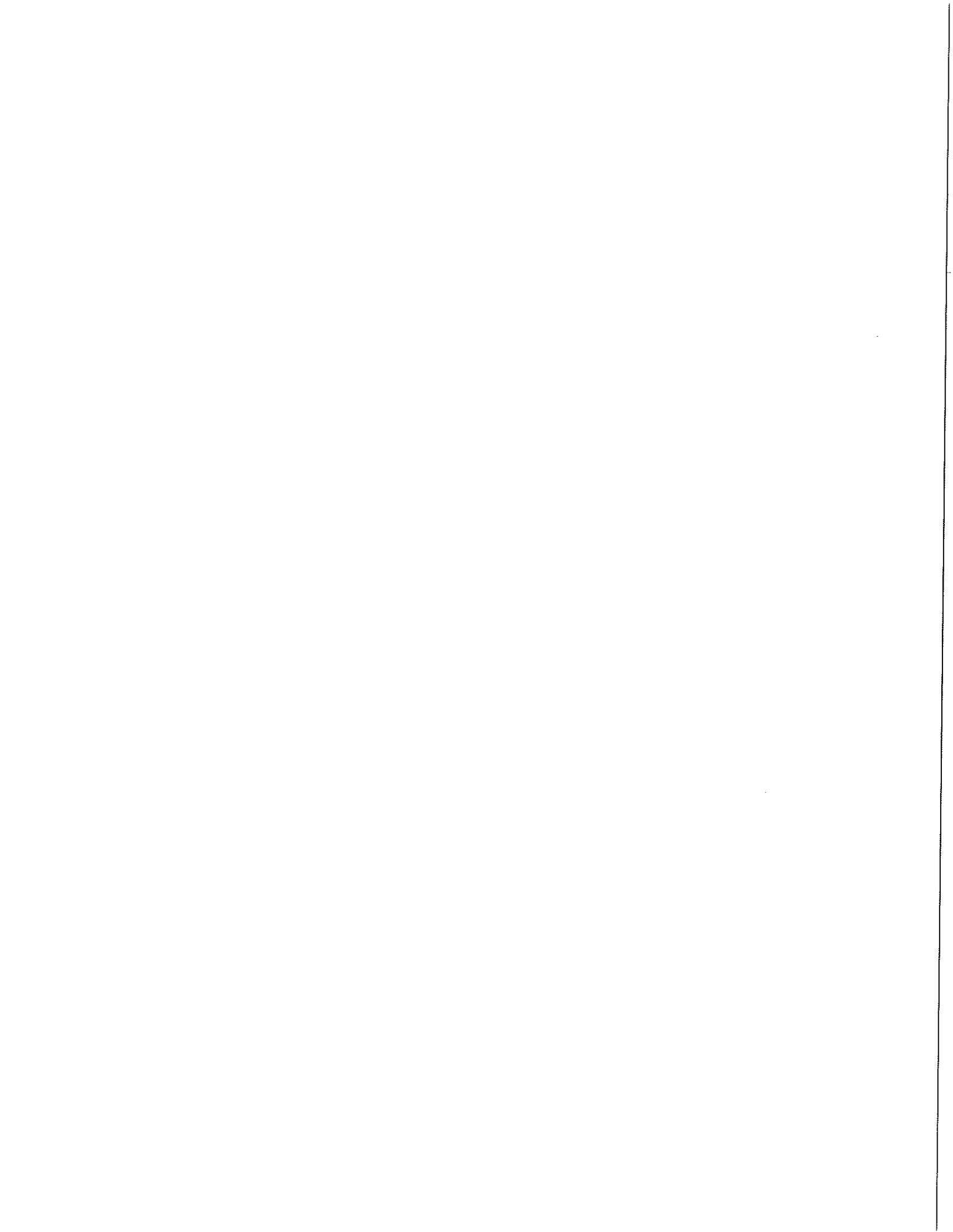


1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49





### Invasive Carcinoma



## Identification of blood-based biomarkers for the detection of breast cancer and lymph node status.

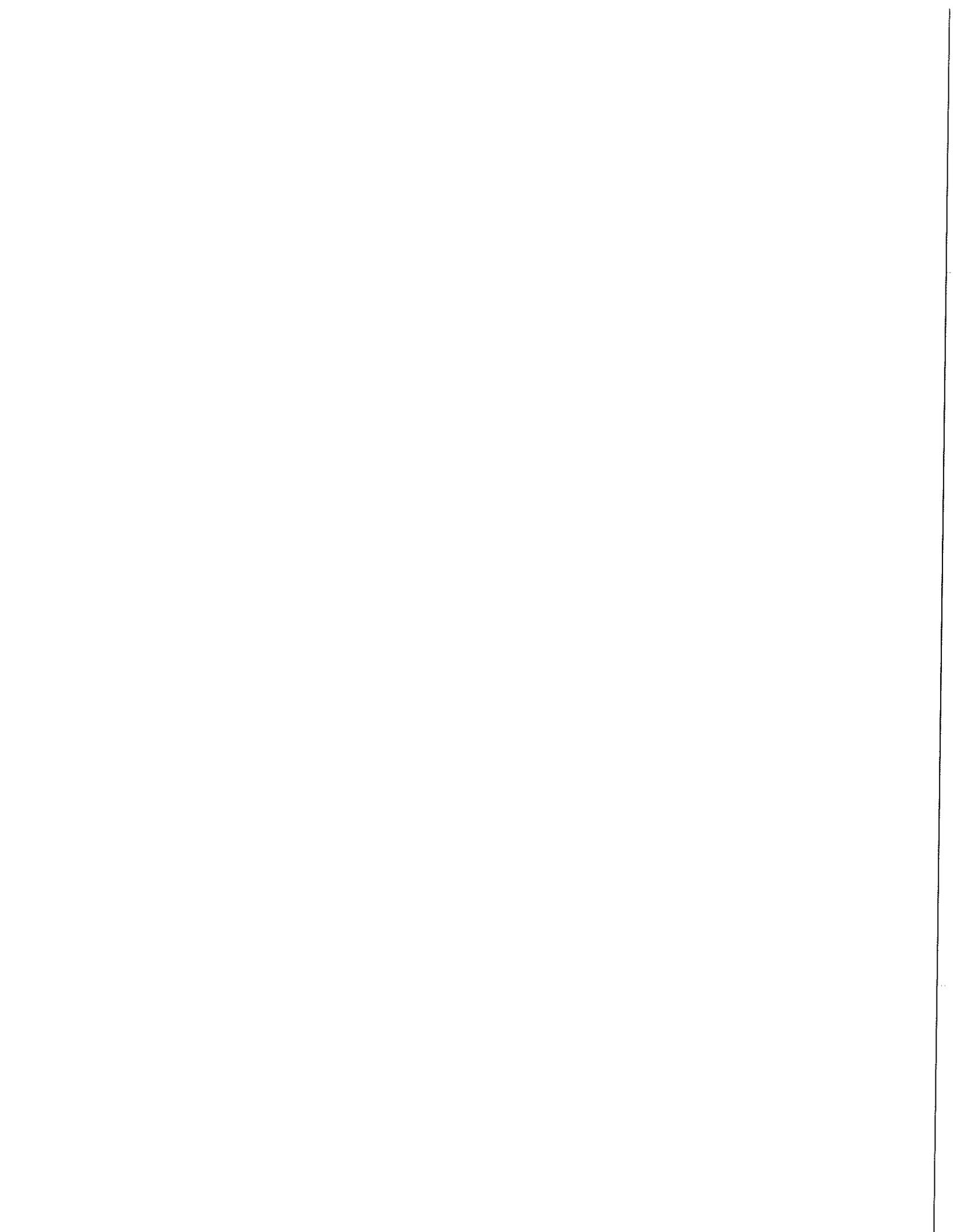
Field LA, Love BJ, Hadix JA, Ellsworth RE, Shriver CD. Windber Research Institute, Windber, PA; Invitrogen Informatics, Carlsbad, CA; Walter Reed Army Medical Center, Washington, D.C.

**Background:** Current methods for determining whether a patient has breast cancer and/or lymph node metastases involve invasive surgical procedures and pathological evaluation. In addition to being more cost-effective, the development of reliable blood tests that can discriminate benign versus malignant breast disease with or without lymph node involvement would ease the anxiety and stress associated with more invasive procedures. Through comparison of whole-blood gene expression profiles, we have identified differences in gene expression between 1) patients with and without breast cancer and 2) breast cancer patients with and without metastatic lymph node tumors.

**Methods:** Whole blood from 20 benign, 19 lymph node negative, and 15 lymph node positive patients was collected into PAXgene tubes. RNA was isolated using the PAXgene Blood RNA kit (PreAnalytiX) and globin mRNA levels reduced using the GLOBINclear kit (Ambion). Globin-reduced RNA was amplified with the MessageAmp II aRNA Amplification kit and labeled with biotin-11-UTP (Ambion). aRNA was fragmented and hybridized to the Affymetrix HG U133A 2.0 expression array. Differentially expressed genes were identified by Mann-Whitney testing.

**Results:** Comparison of blood-based gene expression patterns between patients with benign versus invasive disease identified 115 probe sets with differential expression at  $p < 0.0001$ . Among these genes, PPP1R12A, PPP2R5A, and ATRX had higher expression in cancer patients whereas EI24 and RNH1 had lower expression in cancer cases. Comparisons between lymph node negative and lymph node positive patients identified 5 differentially expressed genes at  $p \leq 0.001$ . VEGFA, CCNA1, and PDPN had higher expression levels in node-positive patients.

**Discussion:** These data suggest that gene expression profiles from whole blood can be used to differentiate between patients with and without breast cancer. Those genes with higher expression in cancer function in cell growth and division, chromatin remodeling, transcriptional regulation, and signal transduction. Genes with lower expression in cancer patients function in cytoskeleton organization and biogenesis, apoptosis, and regulation of angiogenesis. Several changes in gene expression were also detected in blood from node-negative and node-positive patients; these genes function in protein glycosylation, angiogenesis, cell growth and migration, cell cycle regulation, and inhibition of apoptosis. These results suggest that blood may represent a potential screening tool for the detection of breast cancer and may prove useful in the determination of lymph node status in breast cancer patients.



**Title:** Traditional Breast Cancer Risk Factors and Common Breast Pathologies in Post-Menopausal Women

S Maskery, PhD<sup>1</sup>, H Hu, PhD<sup>1</sup>, M Liebman, PhD<sup>1</sup>, J Hooke, MD<sup>2</sup>, C D Shriver, MD<sup>2</sup>, and E Taioli<sup>3</sup>, MD PhD, <sup>1</sup>Biomedical Informatics, Windber Research Institute, Windber, PA, 15963 and <sup>2</sup>Clinical Breast Care Project, Walter Reed Army Medical Center, Washington, DC, 20307, Department of Epidemiology<sup>3</sup>, University of Pittsburgh, Pittsburgh, PA, 15260.

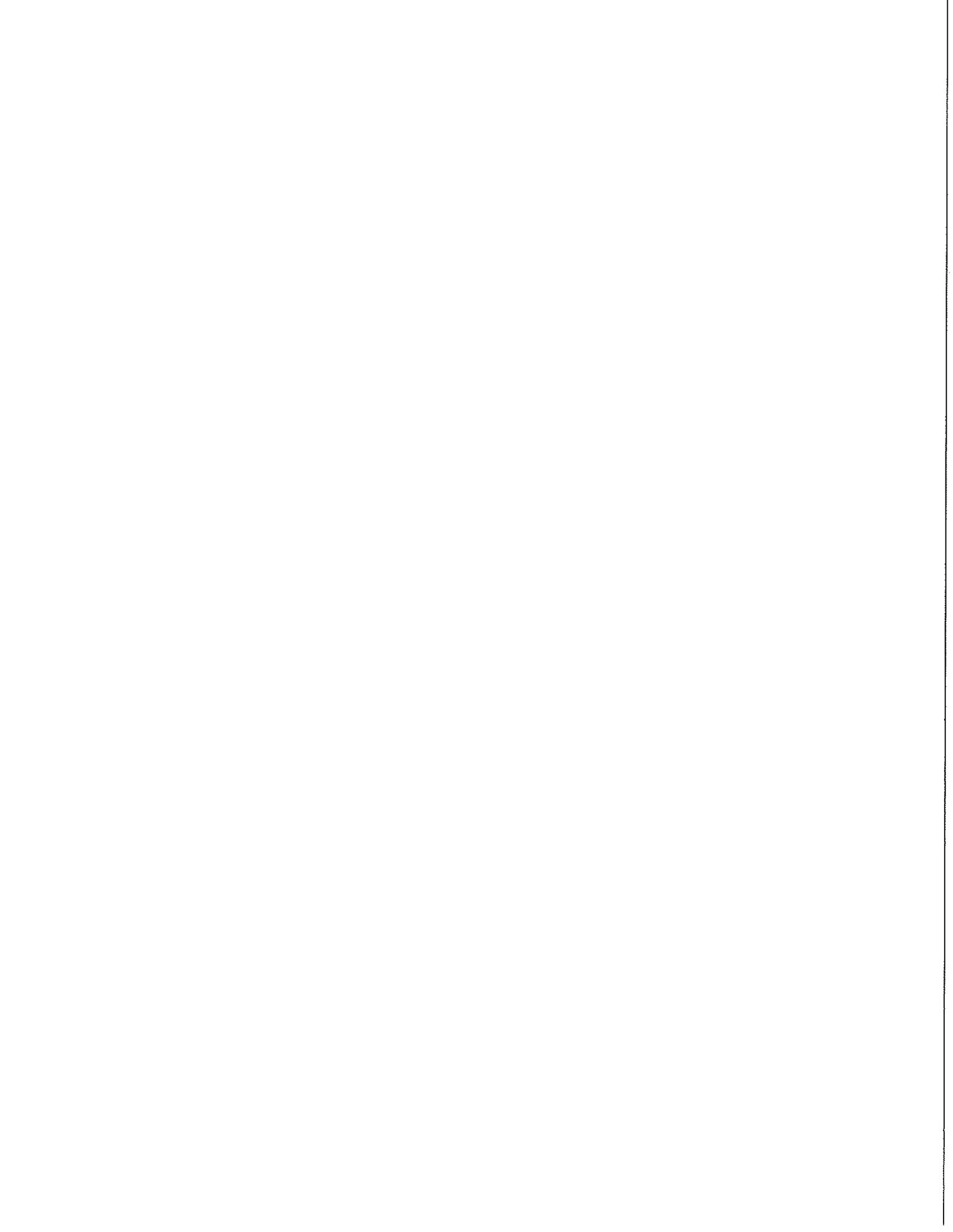
**OBJECTIVE:** Identification of environmental risk factors for non-cancerous breast pathologies may enable the elucidation of the mechanisms by which these factors perturb the normal breast microenvironment. The risk associated with breast cancer risk factors on the occurrence of common breast pathologies was determined in post-menopausal women in a clinical cohort recruited into the Clinical Breast Care Program (CBCP), a joint research project between Walter Reed Army Medical Center (WRAMC) and Windber Research Institute (WRI).

**DESIGN:** In this retrospective study, the contribution of multiple breast cancer risk factors to the occurrence of common breast pathologies was assessed by logistic regression. The inclusion criteria (female, post-menopausal, all diagnoses made by the same pathologist, no missing data) restricted the analysis to 444 records. Odds ratios for stromal fibrosis, apocrine metaplasia, sclerosing adenosis, moderate or florid hyperplasia of the usual type (HUT), and single or multiple papillomas were calculated. These ratios were adjusted for co-occurring breast pathologies (the pathologies listed above, ductal carcinoma in situ, and invasive ductal carcinoma) and breast cancer risk factors (age, hormone replacement therapy (HRT), body mass index (BMI), ethnicity, exercise frequency, alcohol intake frequency, smoking history, early age of menarche, breast feeding, and family history of breast cancer).

**RESULTS:** Age, exercise frequency, early age of menarche, and breast-feeding are not associated with the occurrence of any of the studied breast pathologies in this cohort. HRT use (reference: never used HRT) is significantly associated with stromal fibrosis occurrence (OR=2.5, HRT use > 7yrs), sclerosing adenosis occurrence (OR=2.1, estrogen/progesterone combo), and HUT occurrence (OR=0.52, estrogen/progesterone combo). BMI (reference: normal BMI) is significantly associated with stromal fibrosis occurrence (OR=0.33, obese BMI) and HUT occurrence (OR=1.7, overweight BMI). Ethnicity (reference: Caucasian) is borderline significant (p=0.051) in stromal fibrosis occurrence (OR=0.48, African American). Alcohol intake frequency (reference: < 1/week) is significantly associated with sclerosing adenosis (OR=2.1, 3/week or more) and HUT occurrence (OR=0.46 1-2/week, OR=0.47 3/week or more). Smoking (reference: never smoked) is significantly associated with sclerosing adenosis occurrence (OR=2.0, ever smoked). Family history (reference: no primary relative with breast cancer) is significantly associated with sclerosing adenosis occurrence (OR=1.9, primary relative with breast cancer).

**CONCLUSIONS:** Several documented breast cancer risk factors are also risk factors in the development of common breast pathologies in post-menopausal women in this cohort.

Understanding how breast cancer risk factors affect other breast pathologies may yield new insight into the mechanisms of these risk factors on post-menopausal breast cancer development.



“Identification of genetic predisposition to development of lymph node metastasis in breast cancer patients.”

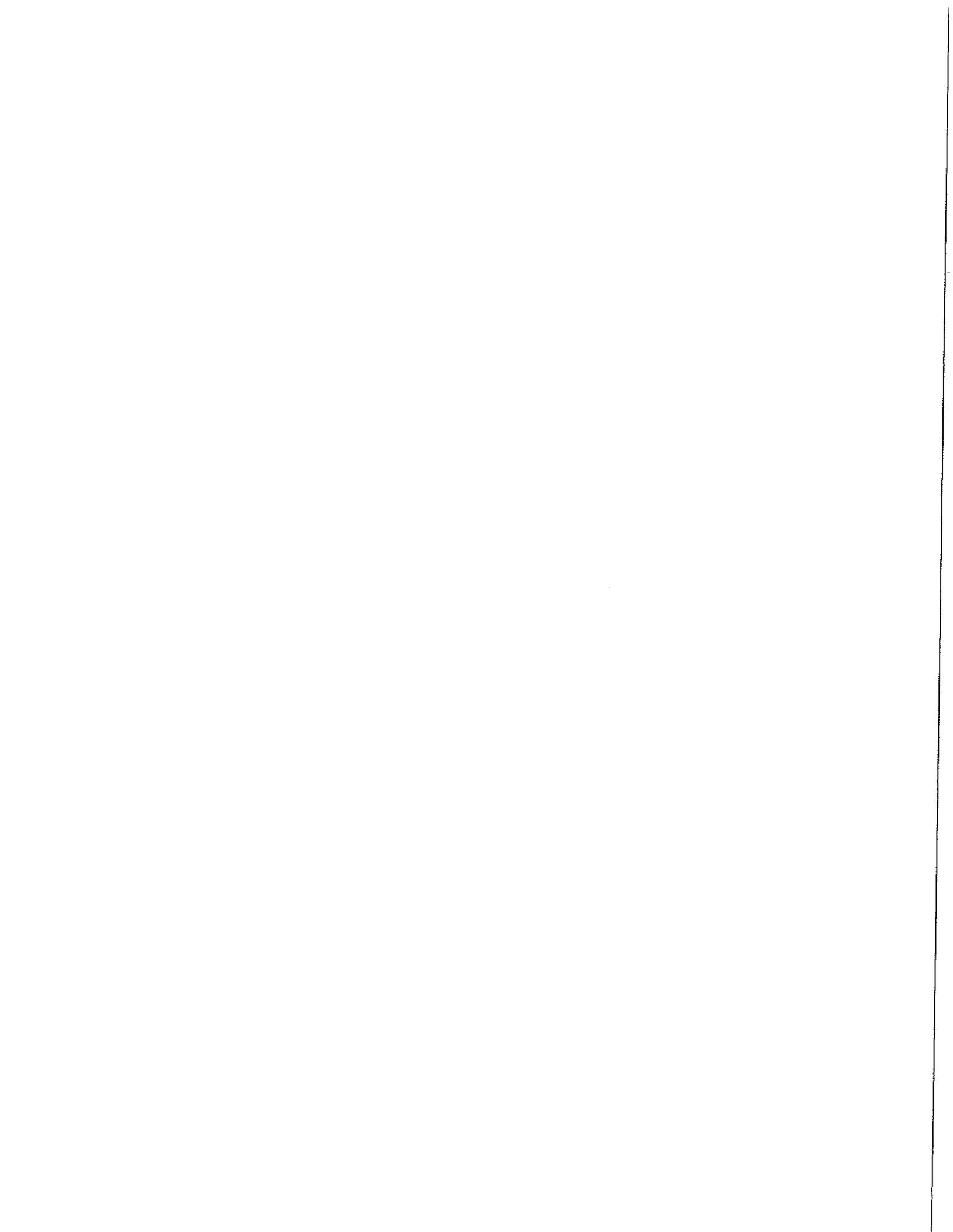
Karen A. Callaghan, Jamie D. Weyandt, Rachel E. Ellsworth, Craig D. Shriver

**Background:** Breast cancer metastasis to ipsilateral axillary nodes is one of the most important predictors of poor outcome in breast cancer cases, and significantly influences the recommendations for adjuvant treatment such as chemotherapy. There is currently no way to predict which patients will develop lymph node metastases despite multiple studies examining the potential impact of primary tumor characteristics on this phenomenon. Here we attempt to identify DNA variants that potentially predispose individuals with breast cancer to the development of lymph node metastases.

**Methods:** Genomic DNA was obtained from blood samples from node-positive (n=37) and node-negative (n=40) Caucasian female breast cancer patients under 50 years of age. Following digestion, ligation, and PCR, the biotin-labeled DNA product was fragmented and hybridized to GeneChip Human Mapping 100K single nucleotide polymorphism (SNP) arrays (Affymetrix). Data was analyzed using Exemplar software and significance was evaluated using a P value of <0.001. Results were compared between node-positive and node-negative samples to identify significant differences in polymorphisms between the two groups.

**Results:** Association studies using Fisher’s exact test revealed that allele frequencies from 39 SNP markers representing 31 chromosomal regions were significantly different between the two groups (P<0.001). Additional model building analysis supported 36 of these markers, with each of these regions demonstrating a difference in two or more statistical tests when comparing node-negative to node-positive samples. Of note, there were six chromosomal regions (5p13.2, 6q22.31, 8q13.3, 9p24.1, 14q32.2, and Xq23) in which multiple SNP markers exhibited significance in both association analysis and model building algorithms.

**Conclusions:** Identification of SNPs within six chromosomal regions with significant differences between node-negative and node-positive breast cancer patients suggests that the propensity to develop metastases may be influenced by genetic factors. Of the potential candidate genes that have been discovered in these regions, several have been found to function in immunomodulation, cell growth regulation, and lipid metabolism, thus polymorphisms within these genes may either confer a protective effect by enhancing the host’s immuno-surveillance or render individuals susceptible to metastatic spread by creating a host tissue environment favorable to metastatic colonization.



**A transcription factor-centric computational analysis of genes differentially expressed in healthy breast tissues from African American and Caucasian women**

Hai Hu, PhD<sup>1</sup>, Lori Field, PhD<sup>1</sup>, Philip Stegmaier<sup>2</sup>, Bradley Love, PhD<sup>3</sup>, Rachel Ellsworth, PhD<sup>1</sup>, Craig D. Shriver, MD<sup>4</sup>, Michael N. Liebman, PhD<sup>1</sup>, and Richard J. Mural, PhD<sup>1</sup>

<sup>1</sup>Windber Research Institute, Windber, PA, USA; <sup>2</sup>BIOBASE GmbH, Wolfenbüttel, Germany; <sup>3</sup>Invitrogen Informatics, Carlsbad, CA; <sup>4</sup>Clinical Breast Care Project, Walter Reed Army Medical Center, Washington, DC, USA

**Background:** Breast cancer in African American women (AA) generally differs from that in Caucasian American women (CA). To explore the molecular basis of such differences, a gene expression study has been conducted with healthy breast tissues from AA and CA, which identified differentially expressed genes. In an *in silico* study using Explain™ (BIOBASE), we examined whether those genes are under the control of common transcription factors (TF) and upstream signaling network. Explain is an integrative system for analysis based on knowledge libraries created by BIOBASE following manual curation of published literature.

**Method:** The RNA microarray experiments (Affy GeneChip U133A 2.0) with healthy breast tissues (reductive mammoplasty or disease-free biopsy) used 26 AA and 22 CA. Differentially expressed genes were identified with false-positive corrected Mann-Whitney analysis. Applying  $p < 0.001$ , 89 gene probes (dataset1, corresponding to 117 genes in Explain) were considered differentially expressed. This dataset was input to Explain for analysis using established statistical methods, together with two other datasets, i.e., a set of 631 probes with  $p < 0.01$  (dataset2), and a background set with  $p > 0.5$  and  $|\log_2(\text{fold})| < 0.01$  (dataset3).

**Result:** Only one TF was contained in dataset1. To collect additional TFs to facilitate the subsequent analysis, we combined TFs in dataset2 with TFs known to regulate genes in dataset1, resulting in a total of 36 TFs. Additional analysis based on site predictions indicated that binding sites of 18 of the 36 TFs were significantly enriched in promoters of genes in dataset1 when compared to those in dataset3. Adding 18 TFs with genes in dataset1 yielded 134 genes coding for 49 proteins contained in the knowledge library. These molecules were used for upstream signaling network analysis, applying Key Node analysis for identification of key molecules that can reach many (set at  $>19$  here) of the other molecules within set number of reaction steps (here 6), as well as Network Cluster analysis for identification of stable subnetworks. Seven genes were identified, i.e., STAT1 in dataset2, and ADRBK2, CD22, THPO, GRB2, IFNAR1, and ARF4 in dataset1.

**Discussion:** Using a bioinformatics knowledge library system we identified 7 genes that may help explain the differential gene expression between healthy breast tissues from CA and AA. These genes are all upregulated in CA, and are involved in nucleotide or protein binding, receptor binding, and cytokine and catalytic activities. Cancer development may

involve uncontrolled cell growth or compromised tumor suppression/apoptosis capability. Additional experiments, both computationally and experimentally, will need to be conducted to determine if the latter was the reason to the more aggressive growth and worse prognosis of breast cancers in AA women.

Pathological characteristics of breast tumors in African American women treated within an equal-access health-care system: biological and molecular contributions to the aggressive phenotype and poor clinical outcomes

Rachel Ellsworth, Jeff Hooke, Craig Shriver

**Introduction:** Incidence of breast cancer is higher in young African American (AAW) compared to Caucasian women (CW) and is associated with poor clinical outcome. These differences have often been attributed to access to quality-health care. >600 AAW have enrolled in the Clinical Breast Care Project (CBCP) through the equal-access health care system provided by the Department of Defense. Epidemiological and pathological factors were compared in AAW and CW with invasive breast cancer to identify environmental and biological factors associated with aggressive tumor phenotype and poor clinical outcome.

**Methods:** Information for over 500 demographic and pathological fields was collected. Data was analyzed patients from patients with invasive breast cancer who were self-described as CW (n=356) or AAW (n=84) using Student's t-test and exact unconditional tests; a significance value of  $P < 0.05$  was used for all analyses.

**Results:** Education, co-morbidity, age of menarche, age at first parity, exercise, BMI and use of screening mammogram did not differ significantly between the two groups. A significantly higher proportion of AAW performed monthly BSE, used oral birth controls pills and had high fat intake while a lower proportion breast fed, used HRT, alcohol, cigarettes or caffeine. The average age of diagnosis was significantly lower in AAW, with 16% diagnosed under age 40 compared to 4% of CW. Tumors from AAW more frequently had the triple negative phenotype (ER, PR, and HER-2/*neu* negative), and positive lymph nodes. The mortality rate was significantly higher in AAW (6%) compared to CW (1.5%).

**Conclusions:** Despite the provision of standardized health care, high rates of screening mammograms and BSE, tumors from AAW are associated with poor prognosis and higher mortality rates. Although environmental components may influence tumor growth, the young age of onset and aggressive pathological characteristics described here suggest that aggressive tumor phenotype and outcome are reflective of population-specific molecular and biological differences.



**Title:** Bayesian Analysis of Recurrence in Lymph Node Positive and Lymph Node Negative Breast Cancer Patients

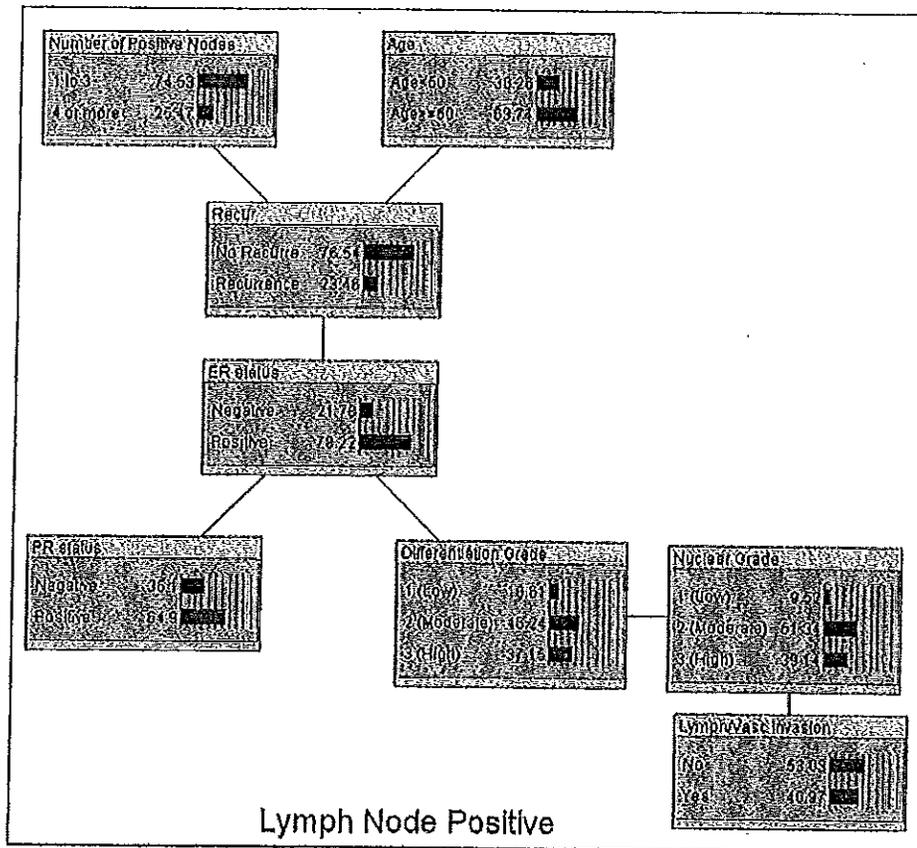
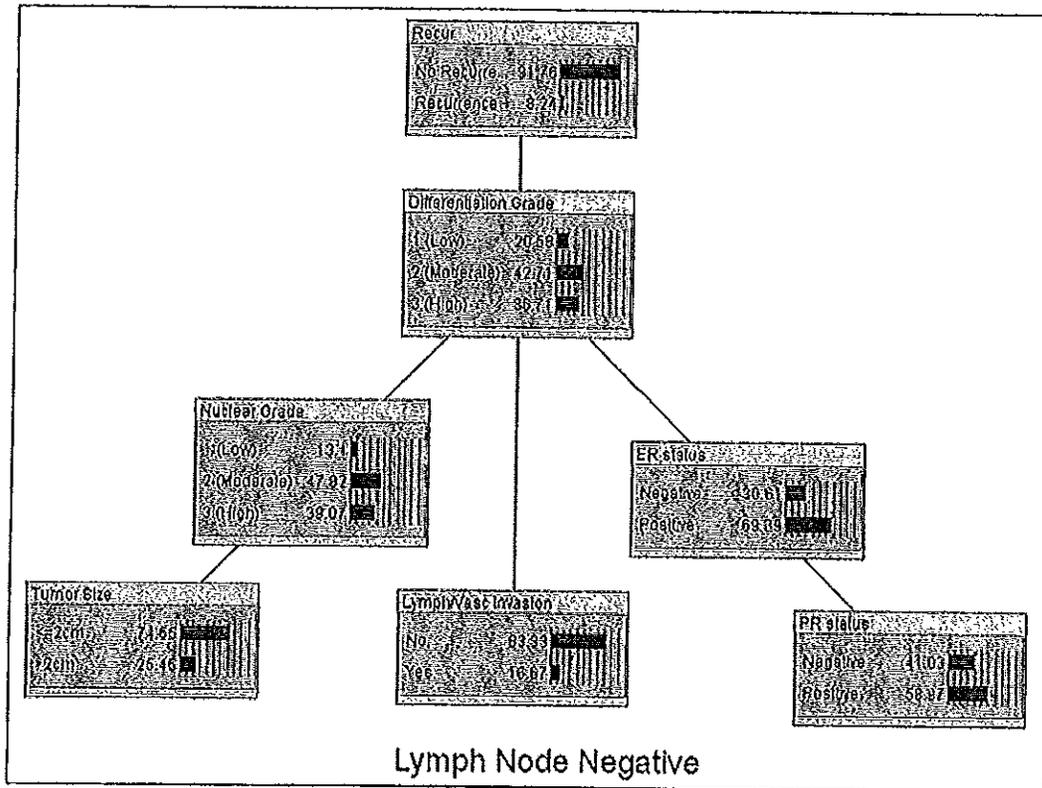
S Maskery, PhD<sup>1</sup>, H Hu, PhD<sup>1</sup>, M Liebman, PhD<sup>1</sup>, C D Shriver, MD<sup>2</sup>, K Verbanac<sup>3</sup>, L Taft<sup>4</sup>, and M Rosman<sup>4</sup>, MD <sup>1</sup>Biomedical Informatics, Windber Research Institute, Windber, PA, 15963 and <sup>2</sup>Clinical Breast Care Project, Walter Reed Army Medical Center, Washington, DC, 20307, <sup>3</sup>Department of Surgery, Eastern Carolina University, Greenville, NC, 27834, <sup>4</sup>Ann Arundel Medical Center, Annapolis, MD, 21401.

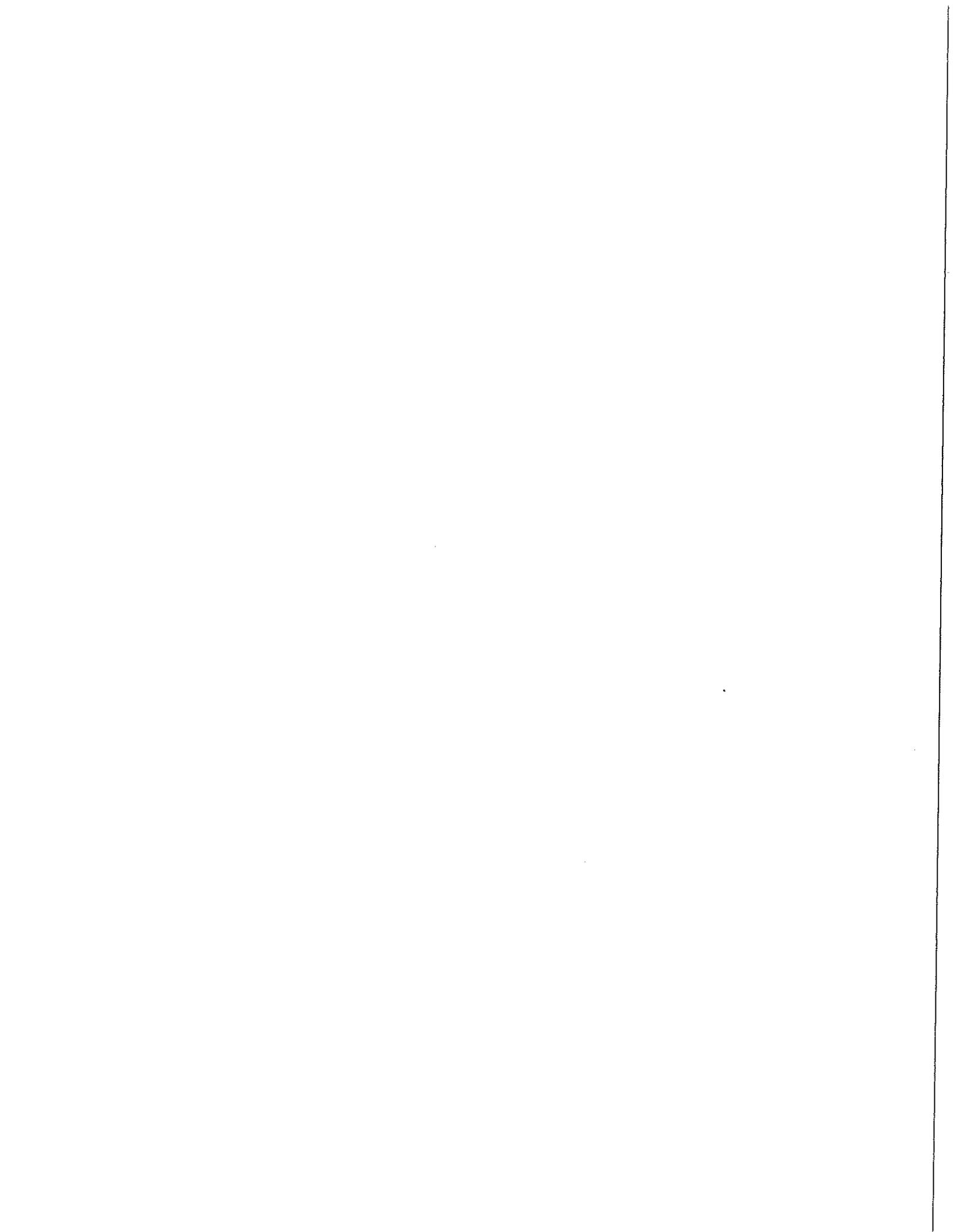
**BACKGROUND:** A Bayesian network is essentially a network of conditional probabilities and as such encodes the probabilistic relationships between model variables (e.g. prognostic factors and breast cancer recurrence). The graphical nature of these networks makes them an intuitive tool to explore the relative significance of prognostic variables on breast cancer recurrence. We applied Bayesian analysis to a multi-institutional cohort of breast cancer patients in which prognostic variables and recurrence were recorded.

**DESIGN:** A multi-institute IRB approved study followed a 1267 invasive breast cancer patient cohort over an average of 4.3 years. Estrogen receptor (ER) status, progesterone receptor (PR) status, tumor size, lymphatic or vascular invasion, ethnicity, age, number of positive lymph nodes, differentiation, and nuclear grade were recorded. The cohort was divided into lymph node positive (N=200, 46 recurrences) and lymph node negative groups (N=418, 21 recurrences). Inclusion criteria for this analysis were no missing data in any variables analyzed and follow-up of at least 2.5 years in the non-recurrence population. Bayesian network analysis was performed separately on each group using a high MDL bias = 1.0 to prevent over-fitting of the data. Recurrence predictors found by Bayesian analysis were compared to a multivariate survival analysis using the Cox Proportional Hazards (CPH) model.

**RESULTS:** The resultant Bayesian networks are displayed in Fig. 1. A line between nodes (variables) in the network indicates these two variables are conditionally dependent, one variable can predict the other and visa versa. In the lymph node negative group only differentiation is a direct predictor of recurrence. However, in the lymph node positive group age, number of positive nodes, and ER status are all direct predictors of recurrence. A parallel multi-variate survival analysis found the same predictor (differentiation) in the lymph node negative group, and mostly the same predictors (age, number of positive nodes, and differentiation) in the lymph node positive group.

**CONCLUSIONS:** Bayesian network analysis enables a graphically intuitive representation of the predictive interdependence of prognostic factors for breast cancer recurrence. The resultant networks identify essentially the same prognostic factors for recurrence as those identified by the traditional CPH model. Validation of these models in an independent cohort may lead to the adoption of these networks as an intuitive mechanism of predicting recurrence in breast cancer patients.





Research article

**Open Access****Amplification of HER2 is a marker for global genomic instability**Rachel E Ellsworth\*<sup>1</sup>, Darrell L Ellsworth<sup>2</sup>, Heather L Patney<sup>2</sup>,  
Brenda Deyarmin<sup>2</sup>, Brad Love<sup>3</sup>, Jeffrey A Hooke<sup>4</sup> and Craig D Shriver<sup>4</sup>

Address: <sup>1</sup>Clinical Breast Care Project, Henry M. Jackson Foundation for the Advancement of Military Medicine, Rockville, MD, USA, <sup>2</sup>Clinical Breast Care Project, Windber Research Institute, Windber, PA, USA, <sup>3</sup>Invitrogen, Carlsbad, CA, USA and <sup>4</sup>Clinical Breast Care Project, Walter Reed Army Medical Center, Washington, DC, USA

Email: Rachel E Ellsworth\* - r.ellsworth@wriwindber.org; Darrell L Ellsworth - d.ellsworth@wriwindber.org;  
Heather L Patney - h.patney@wriwindber.org; Brenda Deyarmin - b.deyarmin@wriwindber.org; Brad Love - drblove@yahoo.com;  
Jeffrey A Hooke - jeffrey.hooke1@us.army.mil; Craig D Shriver - craig.shriver@us.army.mil

\* Corresponding author

Published: 14 October 2008

Received: 1 April 2008

BMC Cancer 2008, 8:297 doi:10.1186/1471-2407-8-297

Accepted: 14 October 2008

This article is available from: <http://www.biomedcentral.com/1471-2407/8/297>

© 2008 Ellsworth et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.**Abstract**

**Background:** Genomic alterations of the proto-oncogene c-erbB-2 (HER-2/neu) are associated with aggressive behavior and poor prognosis in patients with breast cancer. The variable clinical outcomes seen in patients with similar HER2 status, given similar treatments, suggests that the effects of amplification of HER2 can be influenced by other genetic changes. To assess the broader genomic implications of structural changes at the HER2 locus, we investigated relationships between genomic instability and HER2 status in patients with invasive breast cancer.

**Methods:** HER2 status was determined using the PathVysion<sup>®</sup> assay. DNA was extracted after laser microdissection from the 181 paraffin-embedded HER2 amplified (n = 39) or HER2 negative (n = 142) tumor specimens with sufficient tumor available to perform molecular analysis. Allelic imbalance (AI) was assessed using a panel of microsatellite markers representing 26 chromosomal regions commonly altered in breast cancer. Student t-tests and partial correlations were used to investigate relationships between genomic instability and HER2 status.

**Results:** The frequency of AI was significantly higher ( $P < 0.005$ ) in HER2 amplified (27%) compared to HER2 negative tumors (19%). Samples with HER2 amplification showed significantly higher levels of AI ( $P < 0.05$ ) at chromosomes 11q23, 16q22-q24 and 18q21. Partial correlations including ER status and tumor grade supported associations between HER2 status and alterations at 11q13.1, 16q22-q24 and 18q21.

**Conclusion:** The poor prognosis associated with HER2 amplification may be attributed to global genomic instability as cells with high frequencies of chromosomal alterations have been associated with increased cellular proliferation and aggressive behavior. In addition, high levels of DNA damage may render tumor cells refractory to treatment. In addition, specific alterations at chromosomes 11q13, 16q22-q24, and 18q21, all of which have been associated with aggressive tumor behavior, may serve as genetic modifiers to HER2 amplification. These data not only improve our understanding of HER in breast pathogenesis but may allow more accurate risk profiles and better treatment options to be developed.

## Introduction

The HER2 (*c-erb-B2*, *HER-2/neu*) gene, located on chromosome 17q12, is a member of the epidermal growth factor receptor family with tyrosine kinase activity [1]. Amplification of the HER2 gene and/or over-expression of the corresponding protein have been detected in 15–25% of human breast cancers and is associated with poor prognosis [2,3]. Under current standards of clinical care, patients with HER2 amplified (HER2+) tumors receive trastuzumab in combination with standard chemotherapy [4], however, despite treatment with trastuzumab, many HER2+ patients develop distant and progressive metastatic disease [5,6].

To develop more effective treatments for patients with HER2+ breast tumors, efforts have focused on the identification of genes that modify clinical response to trastuzumab including cyclin-dependent kinase inhibitor 1B (p27), phosphatase and tensin homolog (PTEN), insulin-like growth factor 1 receptor (IGF1R) and topoisomerase II  $\alpha$  (TOP2A) [7–11]. In addition, efforts to characterize molecular changes associated with HER2 amplification revealed a cluster of genes from the 17q12–q21 region with similar patterns of amplification, and concordant changes in gene expression [12–14]. Furthermore, multiple gene expression analyses have defined molecular signatures for breast tumors with varying pathological characteristics [13,15,16], improving the ability to accurately characterize tumor sub-types.

Despite advances in molecular characterization of HER2+ breast tumors, mechanisms by which HER2 amplification contributes to breast cancer pathogenesis remain unknown. In this study, we examined levels and patterns of allelic imbalance (AI) in primary breast tumors with and without HER2 gene amplification to 1) examine associations between amplification of the HER2 gene and global genomic instability and 2) identify chromosomal changes commonly observed in HER2 amplified tumors.

## Methods

Paraffin-embedded primary breast tumors were obtained from the Windber Medical Center Pathology Department or the Clinical Breast Care Project (CBCP) Pathology Laboratory. Samples from patients with a previous history of breast cancer or who had received neoadjuvant therapy were excluded from this study. Samples from the Windber Medical Center were archival in nature, having been diagnosed between 1991 and 2003; clinical information was provided for these de-identified samples by the Memorial Medical Center Cancer Registry. Tissue and blood samples from CBCP patients were collected between 2001 and 2006 with approval from the Walter Reed Army Medical Center Human Use Committee and Institutional Review Board. All subjects enrolled in the CBCP voluntarily

agreed to participate and gave written informed consent. Clinical information was collected for all CBCP samples using questionnaires designed by and administered under the auspices of the CBCP.

To ensure consistency, diagnosis of all tumor samples were made by one pathologist from hematoxylin and eosin (H&E) stained slides; staging was performed using guidelines defined by the AJCC *Cancer Staging Manual* sixth edition [17,18]. HER2 status was assayed using the PathVysion® HER-2 DNA Probe kit (Abbott Laboratories, Downers Grove, IL). Amplification was defined as a HER2:CEP 17 signal ratio of  $\geq 2.2$  [3]. Patients with either equivocal HER2 status (1.8 – 2.2) or aneusomy were not evaluated in this study. Clinicopathological information for all samples is summarized in Table 1.

DNA was obtained from pure populations of primary breast tumor cells following laser-assisted microdissection on an AS LMD laser microdissection system (Leica Microsystems, Wetzlar, Germany) [19] (Figure 1). All microdissected sections were examined by the CBCP pathologist, who identified and marked regions of tumor before dissection. To avoid PCR artifacts,  $\geq 5,000$  cells were captured from each of six consecutive breast tumor sections, with the sixth section reserved for all confirmatory reruns. Referent DNA samples for the archival samples were extracted from disease-free skin or negative lymph node tissue from each patient using the QIAamp DNA Mini Kit (Qiagen, Valencia, CA). Referent DNA for the CBCP samples was obtained from blood clots using Clotspin and Puregene DNA purification kits (Gentra, Minneapolis, MN).

Microsatellite markers were amplified as previously described [20], purified using Sephadex G-50 resin and genotyped on a MegaBACE-1000 capillary electrophoresis apparatus (Amersham Biosciences, Piscataway, NJ) following standard protocols. Genotypes were determined using Genetic Profiler version 2.0 software. AI was detected as previously described [21] using a cutoff value of 0.35 (Figure 2), which provides >80% reproducibility when AI events are confirmed on a second aliquot of DNA [22].

To increase the ability to detect AI, two microsatellite markers from each chromosomal region were assayed. Marker information was then pooled and AI at each chromosomal region was defined as follows: 1) when at least one marker for a given region showed an allelic ratio  $\leq 0.35$ , the region was considered to show AI; 2) when neither marker had an allelic ratio  $\leq 0.35$  and at least one marker was informative, the region was considered normal; and 3) when both markers were homozygous, the region was considered uninformative.

**Table 1: Clinical and pathological characteristics of 181 invasive breast tumors at the time of diagnosis**

	HER2 amplified (n = 39)	HER2 negative (n = 142)	P-value <sup>a</sup>
<b>Menopausal Status</b>			
Pre (<50 years)	31%	31%	NS
Menopausal (≥ 50 years)	69%	69%	
<b>Tumor Type</b>			
Infiltrating ductal	84%	73%	NS
Infiltrating lobular	8%	16%	
Mixed ductal and lobular	5%	5%	
Other	3%	6%	
<b>Tumor Grade</b>			
Well (Grade 1)	11%	41%	P < 0.0001 <sup>b</sup>
Moderate (Grade 2)	26%	36%	
Poor (Grade 3)	63%	23%	
<b>Hormone Receptor Status</b>			
			P < 0.05 <sup>c</sup>
ER+/PR+	53%	63%	
ER+/PR-	11%	17%	
ER-/PR+	0%	3%	
ER-/PR-	36%	17%	
<b>Lymph Node Status</b>			
Negative	47%	53%	NS
Positive	53%	47%	
<b>TNM Stage</b>			
			NS
Stage I	36%	47%	
Stage II	40%	35%	
Stage III	21%	16%	
Stage IV	3%	2%	
<b>HER2 - IHC</b>			
			P < 0.0005 <sup>d</sup>
0+	0%	17%	
1+	6%	33%	
2+	17%	49%	
3+	77%	1%	

<sup>a</sup>P-values calculated for HER2 positive versus negative tumors.

<sup>b</sup>Comparison between well- and poorly-differentiated tumors.

<sup>c</sup>Comparison between hormone receptor negative (ER-/PR-) and hormone receptor positive tumors.

<sup>d</sup>Comparison between 0+/1+ and 2+/3+.

Comparison of the clinicopathological factors and levels and patterns of AI by HER2 status were performed using Mann-Whitney and Fisher's exact tests. Potential confounding factors were investigated by computing the significance of partial correlations between HER2 and AI while holding grade and ER constant as mitigating factors. Correlations were calculated non-parametrically from the 2 by 2 tables for ordinal scores (using Phi coefficient of association) and then a direct application of the partial correlation calculation was conducted. P-values were estimated from the non-parametric partial correlation using parametric assumptions to attempt to indicate if true correlation was observed. A significance value of  $P < 0.05$  was used for all analyses.

## Results

### Clinicopathologic features

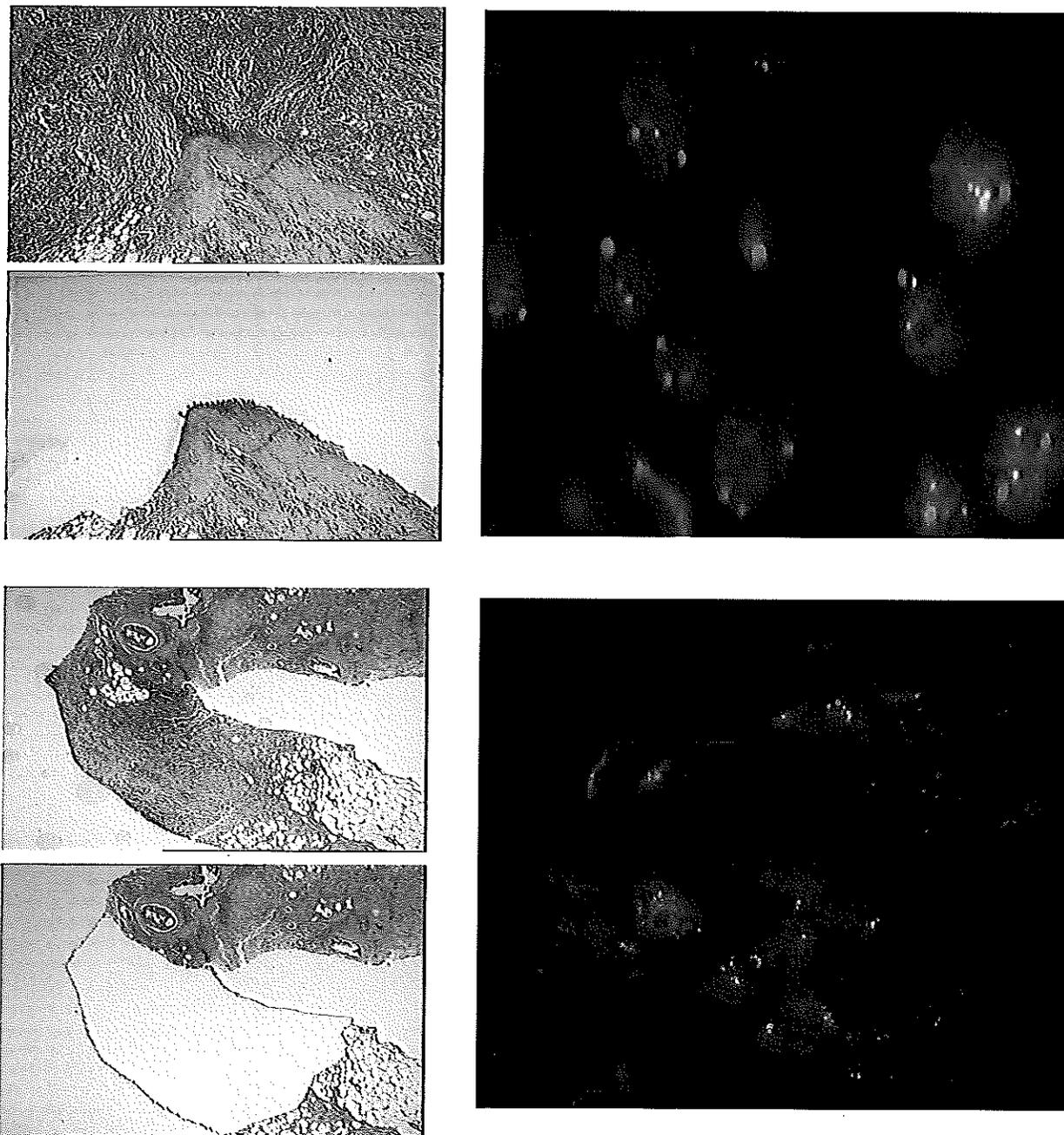
In total, 181 samples were included in this study. All samples were collected from female patients; 22% (n = 39) were HER2+. While age at diagnosis, lymph node status,

tumor histology and stage did not differ significantly between the two groups, poor differentiation and hormone receptor negative status were observed significantly more frequently in HER2+ compared to HER2- tumors. Ninety-four percent of HER2 amplified samples had IHC scores of 2+ or 3+.

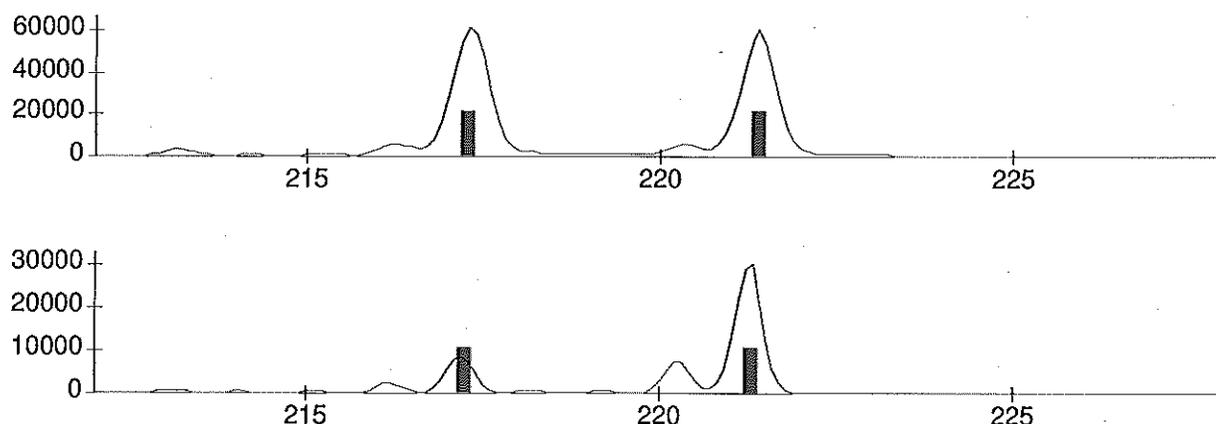
Of the 39 HER2+ tumors, six were from patients diagnosed with invasive breast cancer prior to 1998, thus treatment with trastuzumab was not available for these patients. Of the remaining 33 HER2+ patients, ten were treated with trastuzumab (Table 2). Four HER2+ patients have died of disease and three have developed distant metastasis. The large number of specimens collected within the last 5 years precluded the analysis of outcome data in this study.

### Genetic differences in HER2+ and HER2- tumors

Median AI levels were 25% (range 0–80%) and 13% (range 0–67%) in HER2+ and HER2- tumors, respectively.



**Figure 1**  
**Images of tumors before and after laser-assisted microdissection with corresponding FISH data.** The tumor specimen on the top was taken from a pre-menopausal woman with stage IIb IDCA, without amplification of the HER2 gene. The tumor specimen on the bottom was taken from a pre-menopausal woman stage IIIB IDCA and HER2 amplification of 3.3. Green signals = CEP17 probe, orange = HER2.



**Figure 2**

**Detection of allelic imbalance in a HER2+ breast tumor with fluorescence-based genotyping.** Alleles for marker D17S250 on chromosome 17q12 were detected as fluorescent peaks in reference DNA (top panel) and microdissected breast tumor DNA (bottom panel). A normalized peak height ratio of 0.28 was calculated for the tumor sample using the following peak heights in relative fluorescence units (rfu): tumor DNA – 8,791 rfu and 30,585 rfu, referent DNA – 61,195 rfu and 59,934 rfu.

Levels of AI were significantly higher ( $P < 0.005$ ) in HER2 amplified (mean = 27%) compared to HER2 negative (mean = 19%). When stratified by chromosomal region, AI events were detected significantly more frequently in HER2+ tumors at 11q23, 16q22-q24 and 18q21 (Table 3). No region was altered at significantly higher levels in HER2- tumors.

#### **Confounding factor analysis of AI and HER2 with ER status and tumor grade**

Because ER status and tumor grade have been associated with HER2 status and may confounding the relationship between HER2 amplification and AI, partial correlations were calculated at each chromosomal region between HER2+ and HER2- samples, holding ER status and grade constant. While the correlation between AI and HER2 status at chromosome 11q23 was influenced by grade, three chromosomal regions, 11q13.1, 16q22-q24 and 18q21 showed significant correlations with HER2 status that cannot be explained by either ER status or grade.

#### **Discussion**

The development of trastuzumab has been cited as a successful example of pharmacogenomics where treatment choices are personalized to individual patients based on specific tumor characteristics. Not all patients with HER2+ tumors, however, will benefit from trastuzumab, and given that the cost of trastuzumab ranges from \$20,000 – \$80,000/year with potential side effects including fever and chills, gastrointestinal toxicity, myelosuppression, and cardiotoxicity with heart failure [23], more precise

prediction of which HER2+ patients will derive benefit from trastuzumab and improved understanding of how amplification and/or overexpression of HER2 contribute to aggressive tumor biology are critical to improving patient treatment.

Breast cancer pathogenesis is associated with an accumulation of sequential genetic alterations. Early genetic changes that deregulate tumor suppressor and oncogenes may render these cells susceptible to additional genetic damage and lead to widespread instability in the tumor genome. Increasing levels of genetic changes have been associated with adverse characteristics such as poorly differentiated pathology, metastasis and decreased survival [24-26]. Isola et al. assessed global copy number changes using comparative genomic hybridization and found that tumor with HER2 amplification had significantly higher levels of aberrations compared to HER2- tumors, suggesting that these tumors were genetically more advanced [27]. In agreement with the results of Isola et al., we found that higher levels of chromosomal alterations were correlated with HER2 status, suggesting that HER2 amplification may serve as a surrogate marker for underlying genomic instability

In addition to the high levels of overall genomic instability associated with HER2 amplification, the positive association between chromosomal alterations at chromosomes 11q13.1, 16q22-q24 and 18q21 and HER2 amplification suggests that genes in these regions may contribute to pathogenesis of HER2+ tumors. Bertucci et

**Table 2: Clinical characteristics of 39 HER2+ patients**

Sample	Date Diagnosis	Stage Diagnosis	Trastuzumab	Status <sup>a</sup>
1	2003	IIA	No	NED
2	2003	IIB	No	DOD
3	2003	IIA	Yes	NED
4	2003	IIA	No	NED
5	2005	I	Yes	NED
6	2003	IIA	Yes	DOD
7	2004	IIA	No	NED
8	2004	I	No	NED
9	2004	IV	Yes	AWD
10	2004	IIIA	Yes	AWD
11	2004	IIIA	Yes	NED
12	2004	IIIC	Yes	AWD
13	2004	I	No	NED
14	2004	I	No	NED
15	2004	IIIC	Yes	NED
16	2004	IIIA	Yes	NED
17	2001	IIB	No	NED
18	2001	I	No	NED
19	2002	I	No	NED
20	2002	IIA	No	UNK
21	2002	IIIA	No	NED
22	2002	I	No	NED
23	2002	I	No	NED
24	2002	IIB	No	NED
25	2003	IIIC	No	NED
26	2003	IIA	No	NED
27	2005	I	No	NED
28	2004	IIA	No	NED
29	2005	IIIA	No	NED
30	2005	I	No	NED
31	1996	IIB	No	DOC
32	2002	I	Unknown	NED
33	2003	I	Yes	NED
34	1992	IIA	No	NED
35	1993	IIIA	No	DOD
36	1994	I	No	NED
37	1995	I	No	DOC
38	1995	I	No	DOD
39	1996	IIB	No	NED

<sup>a</sup>AWD = alive with disease, DOC = dead other causes, DOD = dead of disease, NED = no evidence of disease, UNK = status unknown

al. identified a 36-gene expression profile of HER2+ breast tumors which included altered expression of genes chromosomes 11q and 16q, including the fatty acid desaturase 2 (FADS2) gene [GenBank: AF084559] on chromosome 11q12.2 and the M-cadherin (CDH15) gene [GenBank: D83542] on chromosome 16q24 [28]. Deletions of chromosome 18q have been associated with poor survival [29]. Loss of chromosome 18q has been associated with amplification of HER2 as well tumor progression and poor prognosis [27,29,30]. Thus, chromosomal alterations at these regions may contribute to the aggressive pathology and poor prognosis associated with HER2+ tumors.

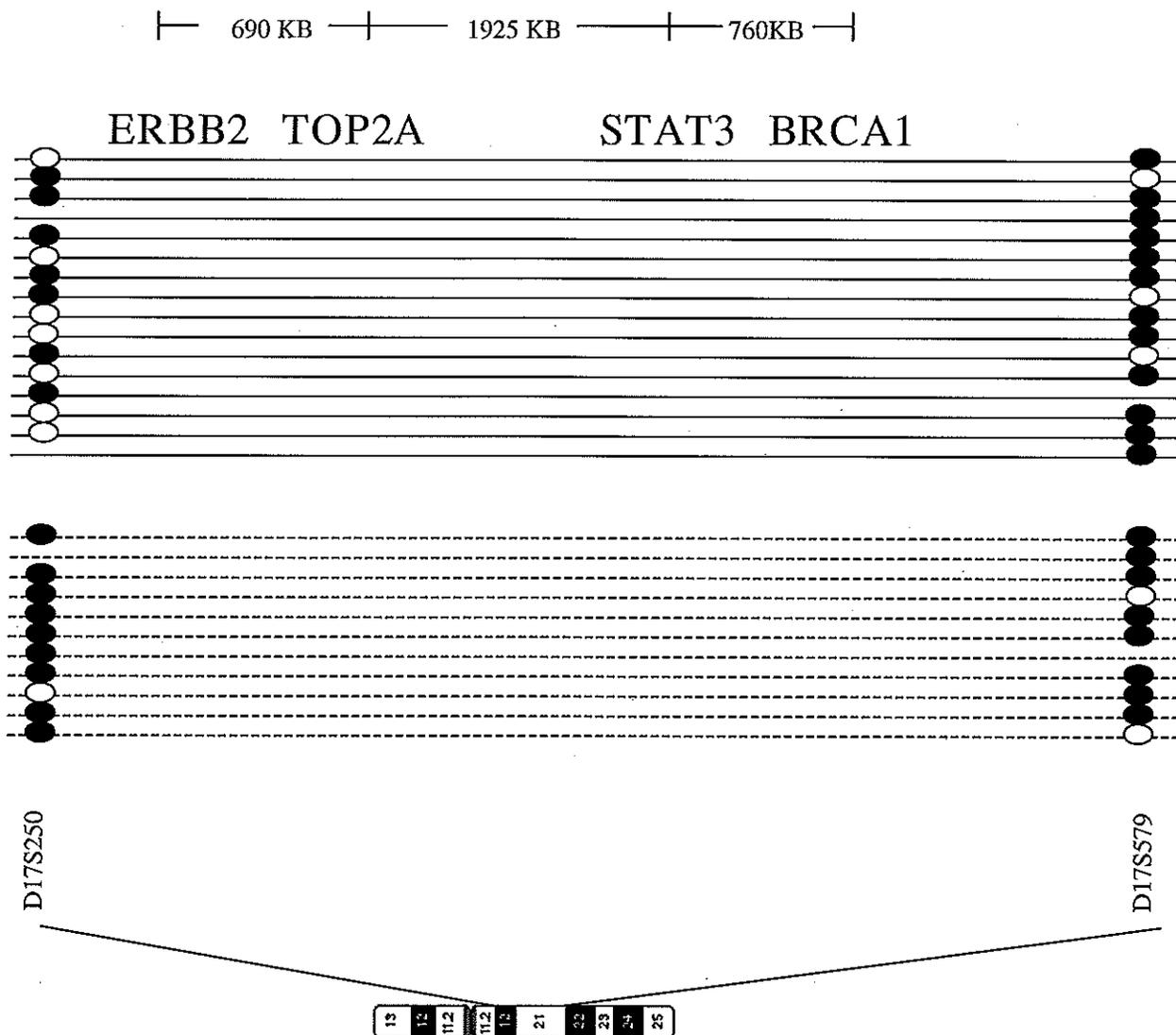
**Table 3: Frequency of AI by HER2 status at 26 chromosomal regions**

Chromosomal Region	HER2+	HER2-	P HER2+ v HER2-
1p36.1-p36.2	0.25 (36)	0.12 (140)	0.0658
2q21.3-23.3	0.14 (36)	0.15 (130)	1.0000
3p14.1	0.19 (36)	0.18 (129)	0.8100
5q21.1-q21.3	0.23 (35)	0.16 (133)	0.3233
6q15	0.18 (38)	0.20 (133)	1.0000
6q22.1-q23.1	0.25 (37)	0.16 (134)	0.2286
6q25.2-q27	0.32 (37)	0.20 (136)	0.1219
7q31.1-q31.31	0.20 (35)	0.08 (133)	0.0523
8p22-p21.3	0.26 (35)	0.18 (130)	0.3476
8q24	0.26 (39)	0.16 (129)	0.1577
9p21	0.20 (35)	0.12 (129)	0.2607
10q23.31-q23.33	0.14 (37)	0.15 (132)	1.0000
11p15	0.26 (35)	0.18 (133)	0.3411
11q13.1	0.33 (39)	0.19 (137)	0.0790
11q23	0.50 (36)	0.23 (128)	<b>0.0033</b>
13q12.3	0.27 (37)	0.30 (130)	0.6641
13q14.2-q14.3	0.29 (39)	0.18 (133)	0.1790
14q32.11-q31	0.18 (38)	0.20 (131)	1.0000
16q11.2-q22.1	0.30 (37)	0.30 (134)	1.0000
16q22.3-q24.3	0.50 (36)	0.26 (129)	<b>0.0076</b>
17p13.3	0.43 (30)	0.26 (117)	<b>0.0789</b>
17p13.1	0.33 (39)	0.28 (137)	0.5564
17q12-q21	0.26 (38)	0.17 (133)	0.2367
18q21.1-q21.3	0.27 (37)	0.13 (128)	<b>0.0416</b>
22q12.3	0.27 (38)	0.15 (132)	0.4565
22q13.1	0.33 (39)	0.26 (134)	0.6639

Numbers in bold are statistically significant

Because the HER2 gene is located within one of the 26 regions in our AI panel, patterns of AI at 17q12-q21 were examined. The two markers used to assess chromosomal content at 17q12-q21, D17S250 and D17S579, define a 5.7Mb region that includes not only the HER2 and TOP2A genes but also the signal transducer and activator of transcription 3 (STAT3) and breast cancer 1 (BRCA1) genes (Figure 3). In samples with AI at chromosome 17q12-q21, the majority of HER2- cases had AI at either D17S250 (24%) or D17S579 (59%), while in 50% of HER2+ tumors, AI was detected at both markers, suggesting that multiple genes from the 17q region may be altered in HER2+ tumors.

TOP2A alterations may contribute to pathogenesis in HER2+ breast carcinomas by altering sensitivity to anthracyclines [31]. Increased expression of STAT3, which may alter cellular proliferation, angiogenesis, and apoptosis, can be activated by HER2 [32,33], and has been associated with advanced disease and poor response to chemotherapy [34], thus genomic amplification of 17q12-q21 may increase STAT3 levels and contribute to aggressive, refractory breast disease. Although mutations in BRCA1 are uncommon in sporadic cancers, physical loss of BRCA1 has been detected in ~50% of sporadic tumors. BRCA1



**Figure 3**  
**Schematic diagram of the 17q12-q21 region.** Genes involved in breast cancer are listed across the top, with distance between genes noted. Patterns of Al are as follows: solid line = HER2- tumors, dotted line = HER2 amplified. Black circles represent an Al event, white circles, normal chromosomal content. Where there is no circle, the reference genotype was homozygous, and thus uninformative.

deletions may be attributed to the genomic structure of BRCA1, which is characterized by unusually high numbers of *Alu* and non-*Alu* repetitive sequences [35]. During replication, mis-pairing between repetitive elements can lead to large deletions in the BRCA1 gene region [36]. How the unstable genomic structure of BRCA1 affects HER2 is unknown, however, because BRCA1 is a DNA repair enzyme, deletions of BRCA1 may impair cellular DNA repair, leading to an accumulation of DNA damage across the genome.

**Conclusion**

In conclusion, breast tumors with copy number changes in the HER2 gene show higher levels of overall genomic instability. Alterations at chromosomes 11q13, 16q22-q24 and 18q21, all of which have been associated with aggressive tumor behavior, may serve as genetic modifiers to HER2 amplification. In addition, alterations within and across the 17q12-q21 region, including TOP2A, STAT3 and BRCA1 may modify the effects of HER2 amplification. Future studies to identify the gene alterations

associated with HER2 amplification as well as examination of a larger group of HER2+ patients treated with trastuzumab should improve our understanding of HER in breast pathogenesis and allow more accurate risk profiles and better treatment options to be developed.

### Competing interests

The authors declare that they have no competing interests.

### Authors' contributions

JAH performed all pathological characterizations, BD performed FISH analysis on all specimens, HP carried out the genotyping experiments, studies, BL oversaw all statistical analysis, DLE and CDS contributed to the preparation of the manuscript and REE conceived of the study, and participated in its design and coordination and helped to draft the manuscript. All authors read and approved the final manuscript.

### Acknowledgements

The authors thank Dr. Christopher M. Gallagher for critical evaluation of the manuscript. This work was performed under the auspices of the Clinical Breast Care Project, and supported by the United States Department of Defense (Military Molecular Medicine Initiative MDA W81XWH-05-2-0075). The opinion and assertions contained herein are the private views of the authors and are not to be construed as official or as representing the views of the Department of the Army or the Department of Defense.

### References

- Coussens L, Yang-Feng TL, Liao YC, Chen E, Gray A, McGrath J, Seeburg PH, Libermann TA, Schlessinger J, Francke U, Levinson A, Ullrich A: **Tyrosine kinase receptor with extensive homology to EGF receptor shares chromosomal location with NEU oncogene.** *Science* 1985, **230**:1132-1139.
- Slamon DJ, Godolphin W, Jones LA, Holt JA, Wong SG, Keith DE, Levin WJ, Stuart SG, Udove J, Ullrich A, Press MF: **Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer.** *Science* 1989, **244**:707-712.
- Wolff AC, Hammond ME, Schwartz JN, Hagerty KL, Allred DC, Cote RJ, Dowsett M, Fitzgibbons PL, Hanna WM, Langer A, McShane LM, Paik S, Pegram MD, Perez EA, Press MF, Rhodes A, Sturgeon C, Tauba SE, Tubbs R, Vance GH, Vijver M van de, Wheeler TM, Hayes DF, American Society of Clinical Oncology, and College of American Pathologists: **American Society of Clinical Oncology/College of American Pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer.** *J Clin Oncol* 2007, **25**:1118-1145.
- Slamon DJ, Leyland-Jones B, Shak S, Fuchs H, Paton V, Bajamonde A, Fleming T, Eiermann W, Wolter J, Pergman M, Baselga J, Norton L: **Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2.** *New Eng J Med* 2001, **344**:783-792.
- Cobleigh MA, Vogel CL, Tripathy D, Robert NJ, Scholl S, Fehrenbacher L, Wolter JM, Paton V, Shak S, Lieberman G, Slamon DJ: **Multinational study of the efficacy and safety of humanized anti-HER2 monoclonal antibody in women who have HER2-overexpressing metastatic breast cancer that has progressed after chemotherapy for metastatic disease.** *J Clin Oncol* 1999, **17**:2639-2648.
- Vogel CL, Cobleigh MA, Tripathy D, Guthrie JC, Harris LN, Fehrenbacher L, Slamon DJ, Murphy M, Novotny WF, Burchmore M, Shak S, Stewart SJ, Press M: **Efficacy and safety of trastuzumab as a single agent in first-line treatment of HER2-overexpressing metastatic breast cancer.** *J Clin Oncol* 2002, **20**:719-726.
- Nahta R, Takahashi T, Ueno NT, Hung MC, Esteva FJ: **P27(kip1) down-regulation is associated with trastuzumab resistance in breast cancer cells.** *Cancer Res* 2004, **64**:3981-3986.
- Cardoso F, Durbecq V, Larsimont D, Paesmans M, Leroy JY, Rouas G, Sotiriou C, Renard N, Richard V, Piccart MJ, Di Leo A: **Correlation between complete response to anthracycline-based chemotherapy and topoisomerase II-alpha gene amplification and protein overexpression in locally advanced/metastatic breast cancer.** *Int J Oncol* 2004, **24**:201-209.
- Fritz P, Cabrera CM, Dippon J, Gertels A, Simon W, Aulitzky WE, Kulp H van der: **c-erbB2 and topoisomerase IIalpha protein expression independently predict poor survival in primary human breast cancer: a retrospective study.** *Breast Cancer Res* 2005, **7**:R374-R384.
- Nagata Y, Lan KH, Zhou X, Tan M, Esteva FJ, Sahlin AA, Klos KS, Li P, Monia BP, Nguyen NT, Hortobagyi GN, Hung MC, Yu D: **PTEN activation contributes to tumor inhibition by trastuzumab, and loss of PTEN predicts trastuzumab resistance in patients.** *Cancer Cell* 2004, **6**:117-127.
- Lu Y, Li X, Zhao Y, Mascarenhas D, Pollack M: **Insulin-like growth factor-I receptor signaling and resistance to trastuzumab (Herceptin).** *J Natl Cancer Inst* 2001, **93**:1852-1857.
- Kauraniemi P, Barlund M, Monni O, Kallioniemi A: **New amplified and highly expressed genes discovered in the ERBB2 amplicon in breast cancer by cDNA microarrays.** *Cancer Res* 2001, **61**:8235-8240.
- Perou CM, Sorlie T, Eisen MB, Rijn M van de, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lonning PE, Borresen-Dale AL, Brown PO, Botstein D: **Molecular portraits of human breast tumours.** *Nature* 2000, **406**:747-752.
- Pollack JR, Perou CM, Alizadeh AA, Eisen MB, Pergamenschikov A, Williams CF, Jeffrey SS, Botstein D, Brown PO: **Genome-wide analysis of DNA copy-number changes using cDNA microarrays.** *Nat Genet* 1999, **23**:41-46.
- Sorlie T, Tibshirania R, Parker J, Hastie T, Marron JS, Nobel A, Deng S, Johnsen H, Pesich R, Geisler S, Demeter J, Perou CM, Lonning PE, Brown PO, Borresen-Dale AL, Botstein D: **Repeated observation of breast tumor subtypes in independent gene expression data sets.** *Proc Natl Acad Sci USA* 2003, **100**:8418-8423.
- Dressman MA, Baras A, Malinowski R, Alvis LB, Kwon I, Walz TM, Polymeropoulos MH: **Gene expression profiling detects gene amplification and differentiates tumor types in breast cancer.** *Cancer Res* 2003, **63**:2194-2199.
- American Joint Committee on Cancer: **AJCC Cancer Staging Manual** New York Springer; 2002.
- Elston CW, Ellis IO: **Pathological prognostic factors in breast cancer I. The value of histological grade in breast cancer: experience from a large study with long-term follow up.** *Histopathology* 1991, **19**:403-410.
- Ellsworth DL, Shriver CD, Ellsworth RE, Deyarmin B, Somiari RI: **Laser capture microdissection of paraffin-embedded tissues.** *BioTechniques* 2003, **34**:42-46.
- Ellsworth RE, Ellsworth DL, Lubert SM, Hooke J, Somiari RI, Shriver CD: **High-throughput loss of heterozygosity mapping in 26 commonly deleted regions in breast cancer.** *Cancer Epidemiol Biomarkers Prev* 2003, **12**:915-919.
- Medintz IL, Lee CCR, Wong WW, Pirkola K, Sidransky D, Mathies RA: **Loss of heterozygosity assay for molecular detection of cancer using energy-transfer primers and capillary array electrophoresis.** *Genome Res* 2000, **10**:1211-1218.
- Ellsworth RE, Ellsworth DL, Neatrou DM, Deyarmin B, Lubert SM, Sarachine MJ, Brown P, Hooke JA, Shriver CD: **Allelic imbalance in primary breast carcinomas and metastatic tumors of the axillary lymph nodes.** *Mol Cancer Res* 2005, **3**:71-77.
- Treish I, Schwartz R, Lindley C: **Pharmacology and therapeutic use of trastuzumab in breast cancer.** *Am J Health Syst Pharm* 2000, **57**:2063-2076.
- Ellsworth RE, Hooke JA, Love B, Kane JL, Patney HL, Ellsworth DL, Shriver CD: **Correlation of levels and patterns of genomic instability with histological grading of invasive breast tumors.** *Breast Cancer Res Treat* 2008, **107**:259-265.
- Nishizaki T, DeVries S, Chew K, Goodson WH, Ljung BM, Thor A, Waldman FM: **Genetic alterations in primary breast cancers and their metastases: direct comparison using modified comparative genomic hybridization.** *Genes Chromosomes Cancer* 1997, **19**:267-272.
- Al-Kuraya K, Schraml P, Torhorst J, Tapia C, Zaharleva B, Novotny H, Spichtin H, Mauer R, Mirlacher M, Kochli O, Zuber M, Dietrich H,

- Moss F, Wilber K, Simon R, Sauter G: **Prognostic relevance of gene amplifications and coamplifications in breast cancer.** *Cancer Res* 2004, **64**:8534-8540.
27. Isola J, Chu L, DeVries S, Matsumura K, Chew K, Ljung BM, Waldman FM: **Genetic alterations in ERBB2-amplified breast carcinomas.** *Clin Cancer Res* 1999, **5**:4140-4145.
  28. Bertucci F, Borle N, Ginestier C, Groulet A, Charafe-Jauffret E, Adelaide J, Geneix J, Bachelart L, Finetti P, Koki A, Hermitte F, Hasoun J, Debono S, Viens P, Fert V, Jacquemier J, Birnbaum D: **Identification and validation of an ERBB2 gene expression signature in breast cancers.** *Oncogene* 2004, **23**:2564-2575.
  29. Dellas A, Torhorst J, Schultheiss E, Mihatsch MJ, Moch H: **DNA sequence losses on chromosomes 11p and 18q are associated with clinical outcome in lymph node-negative ductal breast cancer.** *Clin Cancer Res* 2002, **8**:1210-1216.
  30. Frank CJ, McClatchey KD, Devaney KO, Carey TE: **Evidence that loss of chromosome 18q is associated with tumor progression.** *Cancer Res* 1997, **57**:824-827.
  31. Pritchard KI, Shepherd LE, O'Malley FP, Andrulis IL, Tu D, Bramwell VH, Levine MN: **HER2 and responsiveness of breast cancer to adjuvant chemotherapy.** *New Engl J Med* 2006, **354**:2103-2111.
  32. Hsieh FC, Cheng G, Lin J: **Evaluation of potential Stat3-regulated genes in human breast cancer.** *Biochem Biophys Res Commun* 2005, **335**:292-299.
  33. Fernandes A, Hamburger AW, Gerwin BI: **ErbB-2 kinase is required for constitutive stat 3 activation in malignant human lung epithelial cells.** *Int J Cancer* 1999, **83**:38486-38493.
  34. Diaz N, Minton S, Cox C, Bowman T, Gritsko T, Garcia R, Ewels I, Wloch M, Livingston S, Seljo E, Cantor A, Lee JH, Beam CA, Sullivan D, Jove R, Muro-Cacho CA: **Activation of stat3 in primary tumors from high-risk breast cancer patients is associated with elevated levels of activated SRC and survivin expression.** *Clin Cancer Res* 2006, **12**:20-28.
  35. Smith TM, Lee MK, Szabo CI, Jerome N, McEuen M, Taylor M, Hood L, King MC: **Complete genomic sequence and analysis of 117 kb of human DNA containing the gene BRCA1.** *Genome Res* 1996, **6**:1029-1049.
  36. Welsh PL, King MC: **BRCA1 and BRCA2 and the genetics of breast and ovarian cancer.** *Hum Mol Genet* 2001, **10**:705-713.

### Pre-publication history

The pre-publication history for this paper can be accessed here:

<http://www.biomedcentral.com/1471-2407/8/297/prepub>

Publish with **Bio Med Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

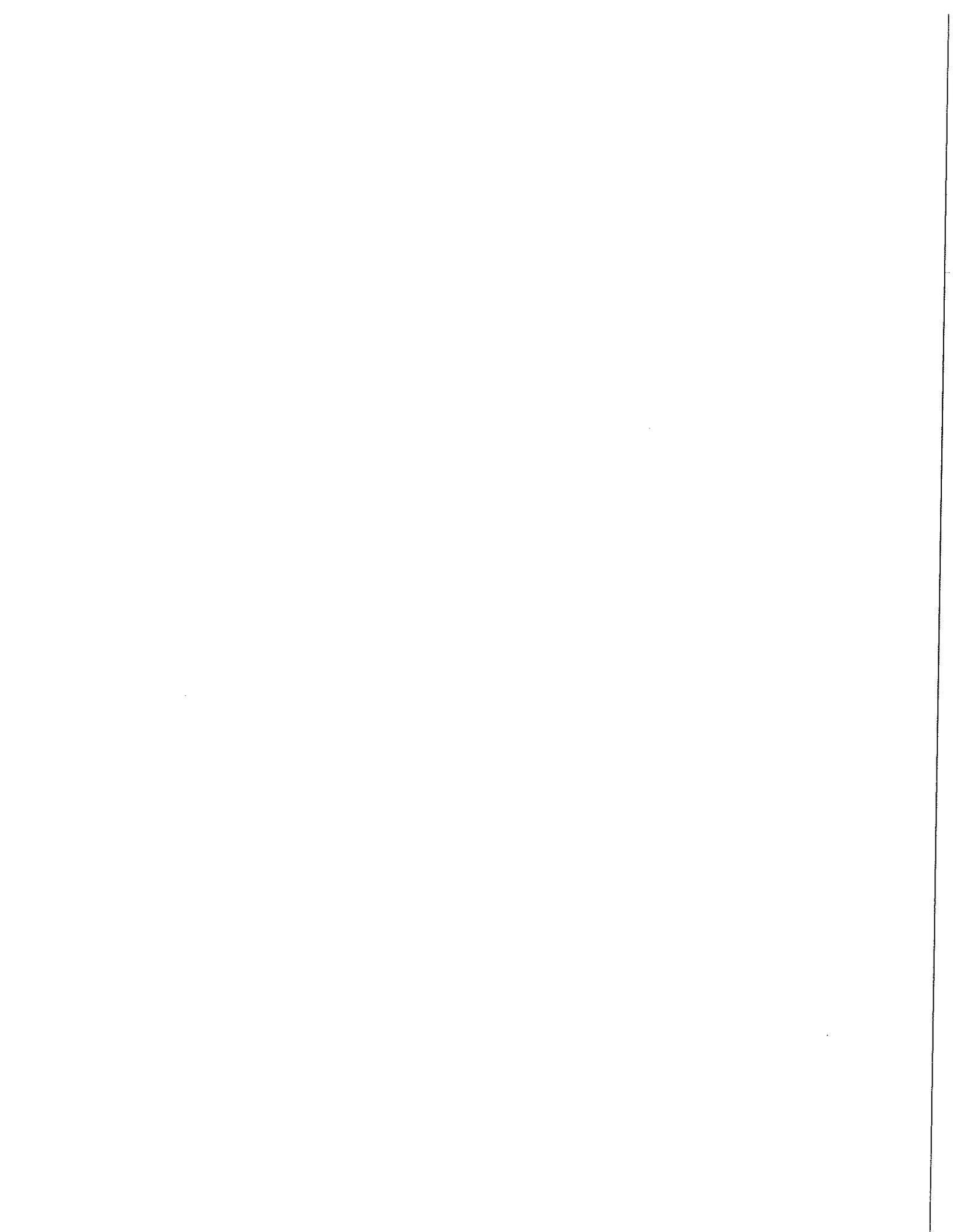
Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)





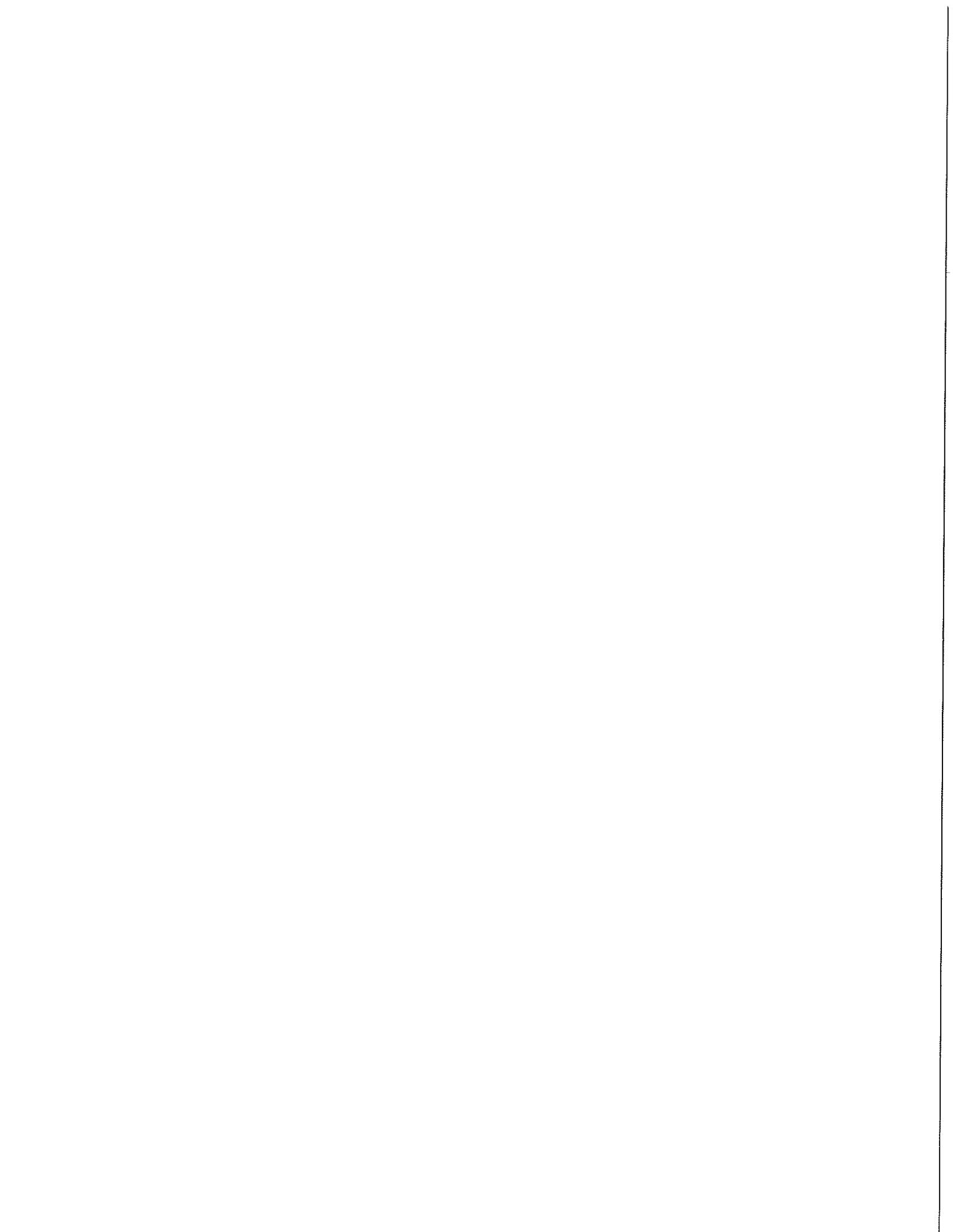
## ABSTRACT FOR 30<sup>th</sup> ANNUAL SAN ANTONIO BREAST CANCER SYMPOSIUM

**Title:** A Retrospective Clinical Correlation Study of the Breast Cancer Patients' Responses to Anticancer Drugs with the Expression Level of Drug Response Indicators (DRI) Measured in Patient's Archival Tumor Tissue

Katherine H Rak Tkaczuk, M.D.<sup>1</sup>, Nancy S Tait, R.N.<sup>1</sup>, William H Rogers, M.D.<sup>1</sup>, Ming Tan, PHD<sup>1</sup>, Olga Ioffe, M.D.<sup>1</sup>, Steve A Lesko, PHD<sup>2</sup>, Scott Deamond, PHD<sup>2</sup>, Zhao-Ping Lum, M.D.<sup>2</sup> and Paul O Ts'O, PHD<sup>2</sup>. <sup>1</sup>Medicine, University of Maryland Greenebaum Cancer Center, Baltimore, Maryland, United States, 21201 and <sup>2</sup>CCC Diagnostics LLC, Baltimore, Maryland, United States, 21227.

**Body:** We conducted a retrospective tumor tissue analysis study in patients with metastatic breast cancer at the University of Maryland Greenebaum Cancer Center. The purpose of our study was to correlate the clinical response to systemic therapy for metastatic breast cancer to the quantitative measurement of DRI/biomarkers with a drug response indicator test (DRIT) which was developed by CCC Diagnostics. The DRIT was first validated in breast cancer cell lines with known sensitivities to chemotherapy agents with activity in breast cancer. We then took formalin-fixed paraffin-embedded tumor tissue from breast cancer patients who received single agent therapy and analyzed it for the expression of the corresponding DRI. The DRI on the tumor tissue was classified as resistant or sensitive profile/group and then the corresponding patient clinical response to given therapy was obtained and correlated with the DRI. The non-responder/responder to the respective drug and vice versa for a resistant/sensitive tumor was then defined by the DRI. For example with the Topoisomerase II as the DRI for Anthracycline/Doxorubicin, the tumor will be more sensitive to the drug when the DRI expression is low and resistant to the drug when the DRI expression is high. A total of 39 cases of DRI studies on 5 classes of drugs (Trastuzumab, hormonal therapy, Taxane-based, Capecitabine, Anthracycline-based) given as monotherapy were evaluated in metastatic breast cancer patients. Among the 28 cases classified as being sensitive to an agent as predicted by the DRI, only one case was not correlated correctly. Among the 11 cases classified as resistant tumors as predicted by the DRI, only one case was not correlated correctly. In summary 39 BC cases were examined for the correlation between the predicted tumor tissue sensitivity/resistance by our DRIT and correlated with the clinical responsiveness to anticancer therapy. The DRIT prediction was not accurate in only two cases (one from predicted sensitive tumor, and one from predicted resistant tumor). The accuracy of DRIT in this small patient cohort is therefore 95% (37/39). A much larger number of patients' samples with rigorous statistical analyses of correlation are forthcoming, Our early results support the hypothesis that DRI expression in the tumor tissue may serve as a predictor of anticancer drug response of breast cancer patients.

*To be presented at the 30<sup>th</sup> Annual San Antonio Breast Cancer Symposium. The symposium will take place December 13-16, 2007 at the Henry B. Gonzalez Convention Center, San Antonio, Texas, USA.*



## **HER2 in the Era of Molecular Medicine: A Review**

Karen A. Callaghan, MD, Rachel E. Ellsworth, PhD, Darrell L. Ellsworth, PhD, Craig D. Shriver, MD

*From the Clinical Breast Care Project, Walter Reed Army Medical Center (KAC, CDS), Washington, DC and Clinical Breast Care Project, Windber Research Institute (REE, DLE), Windber, PA.*

**Grant support:** This work was performed under the auspices of the Clinical Breast Care Project with funding provided by the United States Department of Defense through the Army Medical Research and Materiel Command (MRMC)/Telemedicine and Advanced Technology Research Center (TATRC), Fort Detrick, MD and the Henry M. Jackson Foundation for the Advancement of Military Medicine, Rockville, MD (Military Molecular Medicine Initiative MDA W81XWH-05-2-0075).

The opinions expressed herein are solely those of the authors and do not represent the opinions of the Department of Defense.

Address correspondence to Karen Callaghan, MD and reprint requests to Tiffani Glasco, CBCP Breast Center, Ward 55, Walter Reed Army Medical Center, 6900 Georgia Ave, NW, Washington, DC 20307, Telephone: 202-782-0002, FAX: 202-782-0759, E-mail: karen.callaghan@na.amedd.army.mil

Key words: HER2, breast cancer, human epidermal growth factor receptor

Running head: HER2 in Breast Cancer Pathogenesis

ABSTRACT: Human epidermal growth factor receptor 2 (HER2) is overexpressed in 15%–30% of breast cancers. Women with HER2-positive breast cancer tend to have more aggressive cancer, increased risk of recurrence, and less favorable survival outcomes than women with HER2-negative breast cancer. This review focuses on HER2 and its role in breast cancer pathogenesis. We begin by providing background information on the biological function of HER2 and how this gene contributes to breast cancer development and progression. Next, we review the ongoing debate surrounding the accuracy of available modalities for detecting HER2, namely fluorescence *in-situ* hybridization (FISH) versus immunohistochemistry (IHC). We include current data examining the relationship between HER2 and possible genetic modifiers, such as topoisomerase II $\alpha$ , BRCA1, and genomic instability in breast cancer subjects, and how these relationships may influence response to current therapies directed against HER2. We then discuss trastuzumab, and, after providing an overview of the molecular processes involved in targeted therapy, we summarize the current literature regarding outcomes, as well as the potential impact on the overall health of patients, with special attention to cardiac risk involved with such therapy. Finally, we touch on future directions in this field, including newer targeted therapies in development.

## **BACKGROUND**

Breast cancer is the second leading cause of cancer-related death in women in the United States. One of the many factors influencing prognosis and treatment is overexpression of human epidermal growth factor receptor 2 (HER2, also known as *c-erbB2*, or *HER2/neu*), which occurs in 15%-30% of breast cancers and generally predicts a poor prognosis with decreased disease-free and overall survival [1]. A growing number of specific targeted therapies directed against HER2 have been and are being developed, although response to these therapies is often variable. Here we provide a comprehensive overview of HER2 and its associated clinical implications.

### **1. INTRODUCTION: WHAT IS HER2?**

The HER2 proto-oncogene was first identified by transfection studies in which cells were transformed with DNA from chemically induced rat neuroglioblastomas [2]. The gene is located at chromosomal region 17q12-q21 and encodes HER2, a 185-kDa transmembrane phosphoglycoprotein, which along with HER1, also known as epidermal growth factor receptor (EGFR), HER3, and HER4, is a member of the *erbB* Type I receptor tyrosine kinase family [3]. Although HER2 has no specific ligand, it plays a key role in receptor signaling as a preferred partner for heterodimer formation with all other *erbB* family members, and is involved in multiple signaling cascades including the phosphatidylinositol 3-kinase (PI3K) and mitogen-activated protein kinase (MAPK) pathways [3,4]. HER2 functions in receptor recycling in pathways involved in cell growth, proliferation, differentiation, and migration [4,5].

## 1.1 The role of HER2 in breast cancer development and progression

Based on its role in various molecular pathways, it is not surprising that HER2 has been implicated in breast cancer development and progression. In cell-lines and animal models, cells transfected with HER2 are more invasive and metastatic, while inhibition of HER2 gene function results in reversal of the malignant phenotype [4]. Cells that over-express HER2 have been found to exhibit down-regulation of integrins, up-regulation of matrix metalloproteinases (MMPs), altered cell-cell and cell-ECM (extracellular matrix) interactions, and reduced transcription of E-cadherin (a key cell-cell adhesion molecule), which in turn may lead to break down of the ECM and extravasation [4]. Additionally, enhanced migratory capability of HER2-overexpressing cells is thought to reflect the downstream effects of a second messenger, phosphatidylinositol (4,5)-bisphosphate (PIP<sub>2</sub>). Mobilization of gelsolin and changes in *F*-actin may lead to increased cell deformability [6], as well as the ability of cells over-expressing HER2 to induce endothelial cell retraction, allowing for transmigration of cancer cells [7]. HER2 also has been shown to influence angiogenesis via up-regulation of angiopoetin-2 [8].

Aside from its role in development of metastasis through its impact on cellular interactions, migration, and angiogenesis, HER2 plays an important role in disease progression by modulating cellular response to various therapies. Transfection studies and clinical data demonstrate increased sensitivity to anthracyclines of cells with overexpressed or amplified HER2 [9-12]. Conversely, other studies reveal that transfection of HER2 can decrease sensitivity to certain drugs, including paclitaxel and Tamoxifen [13]. Gene expression profiling supports this observation — overexpression

of the HER2 gene has been associated with low levels of estrogen receptor (ER) expression as well as many other genes normally associated with ER expression. These observations suggest that breast tumors over expressing HER2 may constitute a biologically distinct subtype with unique clinical attributes [14].

## **2. MODALITIES FOR DETECTING HER2 POSITIVITY**

As HER2-positive patients are eligible for targeted therapy that has the potential to improve survival, much time and effort has been devoted to determining the most accurate test for detecting HER2 overexpression and/or amplification in breast tumors. At present, immunohistochemistry and fluorescence *in-situ* hybridization (FISH) are the main methods of diagnosis, but other methods are currently under development.

### **2.1 Immunohistochemistry (IHC)**

IHC is a semi-quantitative method that identifies HER2 protein on the cell surface using one of several commercially-available antibodies. There are currently two IHC tests on the market that have been approved by the United States Food and Drug Administration (FDA): HercepTest™ (Dako Corporation, Glostrup, Denmark) and PATHWAY™ (Ventana Medical Systems, Tucson, AZ) [15,16]. Quantification is based on the amount of membrane staining and percentage (usually at least 10%) of cells stained. Assays are graded as 0, 1+, 2+, or 3+. Tumors with scores of 0 and 1+ are considered negative, while those with 3+ are positive for HER2 over-expression. Scores of 2+ represent weak to moderate over-expression, but are considered ambiguous and typically must be confirmed by FISH (see 2.2 below). IHC is the most widely used testing modality for

determining HER2 status due to its low cost, time-efficiency, and applicability to paraffin embedded tissue; however, IHC scoring can be subjective and may not be as accurate as FISH due to loss of antigen in the formalin fixation process [15].

## **2.2 Fluorescence *in-situ* hybridization (FISH)**

FISH is a quantitative assay that utilizes a DNA probe to determine the number of HER2 gene copies in the nucleus of each cell. Three FDA-approved commercially available kits have been developed: PathVysion<sup>®</sup> (Abbott Molecular, Des Plaines, IL), INFORM<sup>™</sup> (Ventana Medical Systems, Tucson, AZ), and HER2 FISH pharmDx<sup>™</sup> (Dako). All systems are based on counts of 30-60 cell nuclei. Both the PathVysion<sup>®</sup> and pharmDx<sup>™</sup> tests measure the number of HER2 gene copies and the number of chromosome 17 centromeres (CEP17) and consider a ratio of HER:CEP17 of  $\geq 2$  to be amplified. The Ventana INFORM<sup>™</sup> system defines amplification as  $\geq 5$  HER2 gene copies regardless of the number of copies of chromosome 17 [15]. Currently, FISH is considered the gold standard for HER2 assessment because the results tend to be more reproducible, especially in high volume laboratories; however, FISH is more expensive, time consuming, and requires special equipment. In addition, tissue sections used in the assays are difficult to preserve and cannot be used to assess tissue morphology [17,18]. Importantly, it remains unclear whether the absolute number of copies of HER2, or the ratio of HER2:copies of chromosome 17, should be used to define HER2 amplification.

## **2.3 FISH versus IHC**

In 1989, Slamon et al. showed near complete correlation between amplification of the HER2 gene and over-expression of the HER2 protein [19]. Despite the high concordance between IHC and FISH for 0, 1+, and 3+ cases, historically there has been low concordance between IHC and FISH for 2+ cases, with only 12-36% of 2+ cases showing true gene amplification by FISH [18]. More recent reports have shown discordant results between FISH and IHC on the same sample, and even sub-optimal concordance between different laboratories performing the same test on the same sample [15].

An initial concordance evaluation, conducted as part of the North Central Cancer Treatment Group (NCCTG) trial N9831, used a central laboratory to confirm the HER2 status of samples previously determined to be HER2 positive by a local laboratory. Central laboratory testing of 2,535 patients confirmed positive results obtained locally in 2,176 patients (85.8%). Concordance between laboratories was greater for FISH (88%) than the HercepTest™ (81.6%) — an improvement considering that earlier investigation by the same group found concordance rates of 67% and 74%, respectively. Discordant cases were sent to an external reference laboratory, which agreed with the findings of the central laboratory (i.e., confirmed HER2 negativity) in 90.7% of cases [15]. The discrepancy rate in HER2 testing between clinical laboratories is discouraging, as targeted therapy may be withheld from patients with false negative results who actually over-express HER2 and thus could benefit from targeted therapy, while patients with false positive results may be subjected to unnecessary treatment that is not without risk. Currently, the most widely accepted algorithm for HER2 testing (Figure 1) is based on

the National Comprehensive Cancer Network (NCCN) Taskforce recommendations for interpretation of results for both IHC and FISH [20] (Table 1).

#### **2.4 Alternative methods**

Chromogenic *in-situ* hybridization (CISH) determines gene amplification using a peroxidase-based chromogenic reaction. Like FISH, scoring is determined by counting the number of HER2 signals in each nucleus. When originally proposed by Tanner et al. [20], scores of 1-5 were considered to represent normal gene copy number, low-level amplification was defined by 6-10 signals in more than 50% of cancer cells, and high-level amplification was defined by more than 10 signals per nucleus in more than 50% of cancer cells. More recently,  $\geq 5$  signals per nucleus has been used to define HER2 amplification [17].

Concordance between CISH and FISH in validating IHC results has been shown to be  $>95\%$  [17,18]. The advantages of CISH over FISH and IHC are: (1) signals remain stable over time, (2) slides can be stored at room temperature for future reference, (3) evaluation is carried out using a standard microscope, and (4) histological assessment can be performed on the CISH slides [18,21]. Although currently in limited use, technical and clinical validation are required before CISH becomes a widely accepted method for validating equivocal IHC results.

Other methods for detecting HER2 gene amplification include differential polymerase chain reaction (D-PCR) and ELISA-based assays. The Cancer and Leukemia Group B

(CALGB) 8541 and 8869 trials compared IHC, FISH, and D-PCR and found relatively high concordance in determining HER2 status, but equivocal results in predicting outcome following adjuvant doxorubicin therapy [12]. In addition, serum levels of the extracellular domain (ECD) of HER2 have been quantified using ELISA, and sequential monitoring has been used to measure response to chemo- and targeted therapies [22]. Although this method utilizes easily-collectable peripheral blood, further validation will be needed before this test is widely used in clinical practice [22,23].

### **3. GENETIC CHANGES ASSOCIATED WITH HER2 ABNORMALITIES**

HER2 interacts with other genes to regulate many key cellular processes. Unfortunately, current therapies directed against HER2 are effective in <50% of patients, likely the result of potentially redundant pathways involving other genes, which may be under- or over-expressed as a result of genomic changes.

#### **3.1 HER2 and Topoisomerase II $\alpha$**

Another gene located in the 17q12-q21 region is Topoisomerase II $\alpha$  (TOP2A), a key enzyme involved in DNA replication, which catalyzes the unwinding of DNA by inducing single stranded breaks on both strands [24,25]. TOP2A is detectable only in actively dividing cells, and is therefore a useful marker of cellular proliferation [26]. Amplification of TOP2A occurs more frequently in HER2-amplified compared to non-amplified tumors (32%-90% versus 5%-10%) [25], prompting initial investigators to conclude that these two genes were on the same amplicon. The genes are now known to occupy separate amplicons [27,28]. Although HER2-overexpressing tumors are usually

more responsive to anthracyclines [29], which inhibit TOP2A and trap DNA-strand-break intermediates resulting in persistent DNA cleavage and apoptosis, more recent data suggest that increased responsiveness is attributable to TOP2A, rather than HER2, overexpression [24,26]. In a retrospective analysis evaluating response to anthracyclines, patients with TOP2A overexpression (and not gene amplification) seemed to have a more favorable responses [26]; other studies, however, have failed to show a relationship between TOP2A expression and response to anthracyclines *in vivo* [30]. Still others found that any copy number change in TOP2A (whether amplified or deleted) predicted benefit from anthracycline-based treatments, but HER2 status failed to provide any predictive information regarding response to such therapy [31].

In addition to predicting response to therapy, TOP2A over-expression has been correlated with several clinical parameters, including lower hormone receptor expression and increased histological grade. Several studies have correlated TOP2A levels with poor survival in patients with breast cancer. Depowski et al. [32] compared TOP2A expression in 184 primary tumors to several clinical variables, including presence of lymph node metastases, estrogen/progesterone receptor expression, HER2 status, disease recurrence, and survival. Univariate analyses showed that TOP2A over-expression correlated significantly with decreased survival, lymph node positivity, and HER2 amplification.

In contrast to the studies outlined above, other researchers have found no relationship between TOP2A and clinical outcomes in breast cancer patients. Arriola et al. evaluated 232 patients receiving neoadjuvant doxorubicin (an anthracycline) and found 43 tumors

positive for HER2 by standard IHC/FISH. Although TOP2A amplification correlated with high ki67 levels, only HER2 amplification correlated positively with negative estrogen receptor status and predicted clinical response (defined as decreased mammographic appearance of the tumor). TOP2A amplification alone did not show significant associations with clinical or pathological response [30].

### **3.2 HER2 and BRCA1**

The BRCA1 gene is located close to HER2 and TOP2A on chromosome 17q21. Chromosomal alterations that lead to amplification of HER2 thus may be expected to alter BRCA1 as well. In fact, loss of heterozygosity at BRCA1 has been found to correlate significantly with HER2 overexpression in older patients [33]. However, breast cancers that result from BRCA1 mutations infrequently over-express HER2 ( $\leq 3.7\%$ ), possibly because monosomy of chromosome 17 is one of the more frequent chromosomal alterations seen in BRCA1 carcinomas, and as such, could account for the low incidence of HER2 amplification and subsequent overexpression [34].

### **3.3 Overall genomic instability**

Research in our laboratory has shown that primary breast tumors with HER2 amplification and/or polysomy have significantly higher frequencies of genomic instability, assessed by allelic imbalance (AI), compared to primary tumors that do not exhibit chromosome 17 alterations (28% vs. 19%,  $P < 0.0001$ ). These findings suggest that HER2 gene amplification may serve as a marker of global genomic instability, which has been linked to more aggressive cancer phenotypes. When individual chromosomal

regions were examined, tumors with HER2 amplification showed significantly higher levels of AI ( $P < 0.05$ ) at chromosomal regions 6q25-q27, 11q23, 17p13.3, 17q12-q21, and 18q21. Concurrent genetic alterations in other regions of the genome may, therefore, contribute to poor prognosis and influence associations between HER2 status and efficacy of targeted therapies.

#### **4. TREATMENT OF HER2 POSITIVE (HER2+) PATIENTS**

##### **4.1 Trastuzumab**

###### **4.1.1 Mechanism of action**

Observations that a murine monoclonal antibody, which inhibits proliferation of HER2-overexpressing human cancer cells, was immunogenic in humans led to the development of Trastuzumab (Herceptin<sup>®</sup>, Genentech, South San Francisco, CA). To minimize immunogenicity, the antigen binding region was fused to the framework of human IgG, making trastuzumab a humanized monoclonal antibody that binds the extracellular domain (ECD) of HER2. Although the mechanism of action is not fully understood, trastuzumab is believed to induce an antibody-dependent cytotoxic response involving receptor internalization [35,36] and degradation [3]. Others propose that trastuzumab disrupts the relationship between HER2 and Src tyrosine kinase leading to inactivation of Src with subsequent activation of the PI3-kinase-inhibitor PTEN. This in turn leads to rapid Akt dephosphorylation and inhibition of cellular proliferation [37]. Conversely, *in vivo* studies suggest that trastuzumab may enable apoptosis, as levels of ki67 (a marker of proliferation) have not been found to decrease [23]. Trastuzumab also appears to negatively influence angiogenesis, as evidenced by decreased microvessel density *in vivo*

and reduced endothelial cell migration *in vitro*, effects that are enhanced in combination with paclitaxel [38].

Trastuzumab was approved by the FDA in 1998 as a first line treatment with paclitaxel for HER2+ metastatic breast cancer and has been proven to be effective as either a single agent [39] or in combination with chemotherapeutics. It is typically administered as a 4 mg/kg loading dose, with continued doses of 2 mg/kg on a weekly basis, but variations in usage may be prescribed on an individual patient basis [39-41]. Issues that remain to be addressed include dosing schedule, duration of therapy, degree of benefit if continued after disease progression, and whether other chemotherapeutic drugs should be given in combination with trastuzumab [3].

#### **4.1.2 Outcomes**

##### *4.1.2.1 Trastuzumab as a single agent*

Vogel et al. [39] investigated the use of trastuzumab as a single agent for treating metastatic disease in patients whose tumors had IHC scores of 2+ or 3+. The combined complete and partial response rate was 26%, and the clinical benefit rate (defined as response rate + cases of stable disease) was 38%, regardless of dosage. However, response was only seen in patients who had IHC results of 3+. When samples were re-evaluated with FISH, response was 34% in patients with HER2-amplified tumors compared with 7% in non-amplified cases — median time to progression was 4.9 months and 1.7 months, respectively. These observations suggest that FISH may be a superior method for selecting patients likely to benefit from trastuzumab.

#### *4.1.2.2 Trastuzumab in combination with adjuvant chemotherapy (summarized in Table2)*

Initial phase II clinical trials demonstrated that many women with HER2+ metastatic breast cancer, who relapsed after chemotherapy, responded to trastuzumab [42], and that response was greater when given in combination with chemotherapy [42,43]. Based on this data, Slamon et al. [44] conducted a phase III trial in which HER2+ women were randomized to receive six cycles of chemotherapy alone or in combination with weekly trastuzumab. Median time to progression, rate of overall response, duration of response, time to treatment failure, and overall survival were all significantly better in the group treated with trastuzumab, with risk of death decreased by 18%-20%. Patients with IHC scores of 3+ benefited more than 2+ patients, suggesting that women with highly amplified HER2+ tumors obtain the most benefit. Treatment was not without risk though, as congestive heart failure was a serious side effect that increased in incidence with concomitant doxorubicin administration and advancing age.

In an attempt to evaluate the efficacy of trastuzumab in combination with other adjuvant agents, while avoiding risk of cardiac toxicity associated with concomitant anthracycline and trastuzumab use, the Finland Herceptin (FinHer) study initiated a phase III trial to investigate administration of docetaxel or vinorelbine with or without trastuzumab, followed by fluorouracil, epirubicin, and cyclophosphamide [45]. Treatment with trastuzumab significantly decreased disease recurrence and death in patients with HER2+ cancer regardless of the companion chemotherapeutic agent used, with minimal cardiac adverse events.

The National Surgical Adjuvant Breast and Bowel Project (NSABP) B-31 trial compared the outcomes of HER2+, node positive, and high risk node negative patients randomized to receive four cycles of anthracycline and cyclophosphamide (AC) every three weeks followed by paclitaxel (either four cycles every 21 days or weekly for 12 weeks) with or without trastuzumab. When combined with results from the North Central Cancer Treatment Group (NCCTG) trial N9831, trastuzumab was found to reduce recurrence rates by 50% and mortality by 33% in HER2+ patients [46]. The Herceptin Adjuvant (HERA) study compared three groups of women (observation alone, trastuzumab for one year, trastuzumab for two years) who underwent at least four cycles of adjuvant chemotherapy after locoregional treatment for breast cancer [47]. Median follow-up was 12 months (maximum 36 months) and the primary endpoint was disease free survival, defined as recurrence of breast cancer, second non-breast malignancy, or death. Patients who received one year of treatment with trastuzumab showed reduced rates of recurrence compared to the observation only group, suggesting that trastuzumab can be given after initial chemotherapy and is efficacious in node positive as well as node-negative patients. Similar results were obtained in the Breast Cancer International Research Group (BCIRG) 006 trial, which found significantly longer disease free survival both when trastuzumab was added to docetaxel following AC, and when it was given in addition to docetaxel and carboplatin [48].

#### *4.1.2.3 Trastuzumab in the neoadjuvant setting*

Hurley and colleagues investigated the combination of docetaxel, cisplatin, and trastuzumab as primary systemic therapy for women with locally-advanced HER2+ breast cancer and found that all patients experienced an objective clinical response — 46% showed complete response and 54% showed a partial response. Patients were treated adjuvantly with cyclophosphamide and doxorubicin, radiotherapy, and tamoxifen when indicated. Four year progression free and overall survival rates were 81% and 86%, respectively [49]. Other investigations have shown similar results [50].

#### *4.1.2.4 Trastuzumab beyond disease progression*

Administration of cyclophosphamide and methotrexate, in combination with trastuzumab, has been examined in a relatively small group of patients who previously received trastuzumab [41]. Overall response rates (18%) and clinical benefit rates (46%) suggest possible benefits from continuing trastuzumab beyond disease progression. Bartsch and colleagues [51] reported that although response rates may decline with subsequent treatments, there can be benefit for some patients who did not respond favorably to earlier trastuzumab-containing regimens.

#### *4.1.3 Trastuzumab-related cardiotoxicity*

Several *in vitro* and animal studies suggest that HER2 plays a crucial role in cardiac embryogenesis, myocyte repair, and prevention of dilated cardiomyopathy [52]. These models are supported by human data, as a small subset of patients who receive trastuzumab develop cardiac dysfunction and congestive heart failure (CHF). Current standards of care mandate that adequate cardiac function be documented by

echocardiogram or multiple-gated acquisition (MUGA) scanning prior to initiating therapy, and then reassessed throughout therapy to rule out significant deterioration. Nonetheless, incidence of cardiac toxicity ranges from 4%-28% depending on the defining criteria and whether or not it is given in combination with other drugs such as anthracyclines and paclitaxel [44,53,54]. Results of a retrospective review examining cardiac toxicity in seven clinical trials [55] show that the incidence of cardiac dysfunction is greater when trastuzumab is given in combination with anthracyclines (27%) than when given with paclitaxel (13%) or alone (3%-7%).

Slamon et al. [44] performed a retrospective review of 63 patients with trastuzumab-related cardiotoxicity and discovered that although CHF occurs more frequently in patients who received anthracyclines (especially doxorubicin), cardiotoxicity is reversible in ~75% of patients. In patients with cardiac dysfunction who received an anthracycline prior to or concomitant with trastuzumab, average baseline left ventricular ejection fraction (LVEF; a measure of the heart's ability to pump blood) declined from 61% to 43% with trastuzumab therapy [53]. When trastuzumab was discontinued; however, LVEF improved to an average of 55% in less than two months with appropriate therapy (ACE inhibitors and beta-blockers). Trastuzumab was re-started in patients whose symptoms and LVEF stabilized, with only three patients experiencing recurrent cardiac dysfunction or symptoms, prompting permanent discontinuation of trastuzumab in these patients. Cardiac damage thus appears to be reversible, and patients can be safely re-treated after resolution of symptoms and return of acceptable cardiac function. Guarneri and colleagues [52] observed that cardiotoxicity induced by trastuzumab differs from that

related to anthracycline use on several levels. Anthracycline-induced cardiotoxicity is cumulative, dose dependent, irreversible, and associated with specific myocardial damage (vacuolization and loss of contractile elements) on microscopic evaluation. After examining 173 patients who received trastuzumab for one year or longer, overall incidence of cardiac dysfunction was 28%, with only ~11% exhibiting grade 3 cardiotoxicity. Three patients underwent endomyocardial biopsy, which showed focal vacuolar changes, pleomorphic mitochondria, myocardial cell hypertrophy, and mild interstitial fibrosis consistent with reversible cardiomyopathy on electron microscopy. As part of the NSABP B-31 trial, 1,664 patients were assessed for development of cardiac events. Five out of 814 (0.8%) control patients and 31 out of 850 (4.1%) trastuzumab-treated patients experienced a cardiac event [54]. In summary, there is wide agreement that given the improvement in disease-free survival and reduction in mortality afforded by trastuzumab administration, as well as reversible cardiotoxicity, trastuzumab is appropriate for most HER2-positive women with adequate cardiac function.

#### ***4.1.4 Mechanisms of resistance***

Many patients who initially respond to trastuzumab generally acquire resistance within one year [23,44], either through *de novo* or acquired mechanisms [56]. Based on observations in lung cancer patients, one potential mechanism of resistance is altered receptor-antibody interaction, possibly the result of mutations in HER2 that inhibit binding of trastuzumab. Other proposed mechanisms include compensatory increases in cell signaling via other members of the HER receptor family, constitutive Akt signaling that inhibits cell cycle arrest and apoptosis, or down-regulation of PTEN, which

ultimately blocks the action of trastuzumab and enables cell proliferation [37].

Trastuzumab has been found to increase the half-life of p27<sup>kip1</sup> by decreasing the cyclin E/cdk2-mediated phosphorylation of p27<sup>kip1</sup> and by blocking ubiquitin-dependent degradation. Increased levels of p27<sup>kip1</sup> lead to cell cycle arrest, but reduced expression of p27<sup>kip1</sup> can make trastuzumab less effective [16]. Because the insulin-like growth factor receptor (IGF-IR) functions through the same pathways (PI3K and MAPK) as the HER family of receptors, excessive IGF-IR signaling has been associated with resistance to trastuzumab [57].

#### **4.2 Other targeted therapies**

Many redundant molecular processes are involved in cell signaling, especially within the erbB receptor family. As a result, resistance to trastuzumab may develop through alternate pathways. For example, combined HER2 and EGFR overexpression occurs in 10%-36% of breast cancers and defines one of the most aggressive tumor phenotypes [5,23], prompting investigations of treatment strategies that simultaneously target multiple receptors.

Compounds currently being investigated include pertuzumab, a recombinant human monoclonal antibody that binds HER2, thereby blocking EGFR and HER3 dimerization with HER2 and inhibiting intracellular signaling [5,58]. Gefitinib is a tyrosine kinase inhibitor, which can be given in combination with pertuzumab and trastuzumab [23]. Lapatinib is a potent, reversible, selective tyrosine kinase inhibitor that has activity against both EGFR and HER2. Lapatinib can be given in combination with trastuzumab

[58] to inhibit tumor growth and restore tamoxifen sensitivity to resistant ER-positive breast cancer [5]. Additional therapies under development include introduction of an adenovirus gene to prevent synthesis of HER2 mRNA [23,57-59] and synthesis of a chimeric protein designed to deliver immune effector cells to sites of carcinogenesis [23].

DNA-based and peptide-based vaccines are designed to boost HER2 immunity without affecting cells with normal HER2 levels. E75 in particular, is a HER2-derived peptide capable of inducing an immune response mediated by cytotoxic T-lymphocytes. Given together with granulocyte-macrophage colony-stimulating factor (GM-CSF), this vaccine may reduce recurrence rates and improve overall survival in HER2+, node-positive breast cancer patients at high risk for recurrence. Phase III trials to assess the true clinical benefit of E75 are needed [60]. Recent trials involving GP2, administered together with E75, attempt to diversify the immune response and create a multi-epitope vaccine [60-62].

## **5. FUTURE DIRECTIONS**

Possibilities for diagnosis and treatment of patients with over-expression and/or amplification of HER2 may expand exponentially in the near future as our biological understanding of molecular mechanisms of HER2 action increases. Attempts to make IHC and FISH results more reproducible and concordant in clinical practice will continue, while other investigations are questioning the appropriateness of current scoring systems used to determine HER2 status for treatment purposes. Recent work indicates that polysomy of chromosome 17, rather than HER2 status determined by FISH,

predicts response to trastuzumab. Thus, some patients with multiple copies of the HER2 gene, but who are considered to be non-amplified because they have multiple copies of chromosome 17, may benefit from trastuzumab therapy [63]. Further studies are needed to identify whether chromosome 17 polysomy is superior to HER2 copy number ratio for predicting response to therapy. Another modality being developed to aid in predicting therapeutic response utilizes gene arrays to identify gene signatures predictive of response to chemotherapeutic drugs [25]. Such arrays may ultimately be useful for identifying patients likely to respond to HER2 targeted therapies such as trastuzumab, thereby eliminating unnecessary and potentially harmful administration of drugs to patients who would receive little if any benefit. Development of new therapies such as tyrosine kinase inhibitors, peptide vaccines, and modulators of other cellular pathways associated with HER2 hold much promise for breast cancer patients.

### HER2 Testing Algorithm

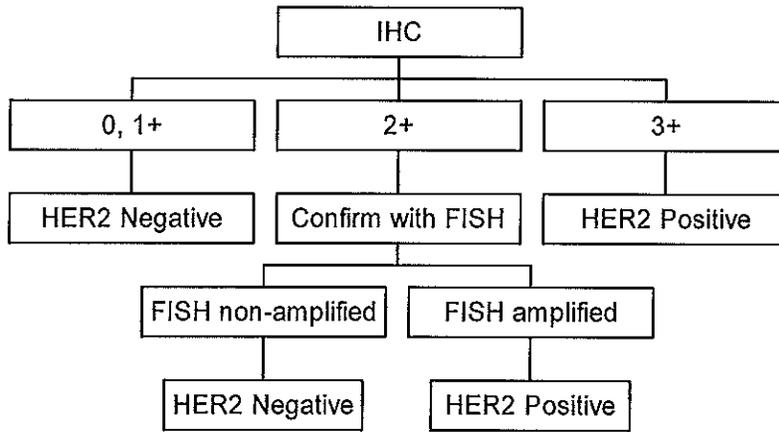


Figure 1. Generally accepted clinical algorithm for determining HER2 status based on recommendations of the National Comprehensive Cancer Network.

**TABLE 1. NCCN recommendations for interpreting IHC and FISH results to determine HER2 status [20].**

	Positive	Negative	Indeterminate*
IHC	3+	0 or 1+	2+
FISH HER2:CEP17 ratio or # of HER2 copies	>2.2 >6	<1.8 <4	1.8-2.2 4-6

\*In cases of indeterminate FISH results, the NCCN Task Force recommends counting more cells, repeating the FISH assay, or performing IHC if not already done. NCCN, National Comprehensive Cancer Network.

<b>Table 2. Summary of the major adjuvant trastuzumab trial designs and endpoints</b>			
<b>Trial</b>	<b>Treatment arms</b>	<b>Primary Endpoint (HR)</b>	<b>Secondary Endpoints</b>
HERA	Any CT +/- RT → observation Any CT +/- RT → T q3w x 12mo Any CT +/- RT → T q3w x 24mo	DFS (0.54)*	TTR, TTdR, OS
NSABP B-31	AC x 4 → P q3w x 4 or qw x 12 AC x 4 → P q3w x 4 or qw x 12 + T qw x 52	DFS** (0.48)	TTdR, OS**
NCCTG N9831	AC x 4 → P qw x 12 AC x 4 → P qw x 12 → T qw x 52 AC x 4 → P qw x 12 + T qw x 52		
BCIRG 006	AC x 4 → D q3w x 4 AC x 4 → D q3w x 4 + T qw x 12 → T q3w x 13 D + Cp q3w x 6 + T qw x 18 → T q3w x 11	DFS (0.49) (0.61)	OS, toxicity, evaluation of pathologic & molecular markers for predicting efficacy
FinHer	D q3w x 3 or V qw x 8 → CEF q3w x 3 D q3w x 3 or V qw x 8 + T qw x 9 → CEF q3w x 3	DFS (0.42)	AE, effect on LVEF, TTdR, OS
<p>Abbreviations: A, doxorubicin; AE, adverse events; BCIRG, Breast Cancer International Research Group; C, cyclophosphamide; Cp, carboplatin; CT, chemotherapy; D, docetaxel; DFS, disease free survival; E, epirubicin; F, 5-Fluorouracil; HERA, Herceptin® Adjuvant; HR, hazard ratio; LVEF, left ventricular ejection fraction; NCCTG, North Central Cancer Treatment Group; NSABP, National Surgical Adjuvant Breast and Bowel Project; OS, overall survival; P, paclitaxel; q3w, every 3 weeks; qw, weekly; RT, radiotherapy; T, trastuzumab; TTdR, time to distant recurrence; TTR, time to recurrence; V, vinorelbine.</p> <p>* = based on patients in the T x 12mo treatment arm ** = based on combined results of NSABP B-31 and NCCTG N9831</p>			

## References

1. Slamon, D. J., Clark, G. M., Wong, S. G., Levin, W. J., Ullrich, A., and McGuire, W. L. Human breast cancer: correlation of relapse and survival with amplification of the HER-2/*neu* oncogene. *Science* 1987; 235:177-182.
2. Shih, C., Padhy, L. C., Murray, M., and Weinberg, R. A. Transforming genes of carcinomas and neuroblastomas introduced into mouse fibroblasts. *Nature* 1981; 290:261-264.
3. Yeon, C. H. and Pegram, M. Anti-erbB-2 antibody trastuzumab in the treatment of HER2-amplified breast cancer. *Investigational New Drugs* 2005; 23:391-409.
4. Eccles, S. A. The role of *c-erbB-2/HER2/neu* in breast cancer progression and metastasis. *Journal of Mammary Gland Biology and Neoplasia* 2001; 6:393-406.
5. Reid, A., Vidal, L., Shaw, H., and deBono, J. Dual inhibition of ErbB1 (EGFR/HER1) and ErbB2 (HER2/*neu*). *European Journal of Cancer* 2007; 43:481-489.
6. Brandt, B. H., Roetger, A., Dittmar, T., Nikolai, G., Seeling, M., Merschjann, A., Nofer, J., Dehmer-Moller, G., Junker, R., Assmann, G., and Zaenker, K. S. *c-erbB-2/EGFR* as dominant heterodimerization partners determine a motogenic phenotype in human breast cancer cells. *The FASEB Journal* 1999; 13:1939-1949.
7. Carter, W. B., Hoying, J. B., Boswell, C., and Williams, S. K. HER2/*neu* over-expression induces endothelial cell retraction. *International Journal of Cancer* 2001; 91:295-299.
8. Carter, W. B. and Ward, M. D. HER2 regulatory control of angiopoietin-2 in breast cancer. *Surgery* 2000; 128:153-158.
9. Muss, H. B., Thor, A. D., Berry, D. A., Kute, T., Liu, E. T., Koerner, F., Cirincione, C. T., Budman, D. R., Wood, W. C., Barcos, M., and Henderson, I. C. *c-erbB-2* expression and response to adjuvant therapy in women with node-positive early breast cancer. *New England Journal of Medicine* 1994; 330:1260-1266.
10. Paik, S., Bryant, J., Park, C., Fisher, B., Tan-Chiu, E., Hyams, D., Fisher, E. R., Lippman, M. E., Wickerham, D. L., and Wolmark, N. *erbB-2* and response to doxorubicin in patients with axially lymph node-positive, hormone receptor-negative breast cancer. *Journal of the National Cancer Institute* 1998; 90:1361-1370.
11. Thor, A. D., Berry, D. A., Budman, D. R., Muss, H. B., Kute, T., Henderson, I. C., Barcos, M., Cirincione, C. T., Edgerton, S., Allred, C., Norton, L., and Liu, E. T. *erbB-2*, p53, and efficacy of adjuvant therapy in lymph node-positive breast cancer. *Journal of the National Cancer Institute* 1998; 90:1346-1360.
12. Dressler, L. G., Berry, D. A., Broadwater, G., Cowan, D., Cox, K., Griffin, S., Miller, A., Tse, J., Novotny, D., Persons, D. L., Barcos, M., Henderson, I. C., Liu, E. T., Thor, A. D., Budman, D. R., Muss, H. B., Norton, L., and Hayes, D. F. Comparison of HER2 status by fluorescence in situ hybridization and immunohistochemistry to predict benefit from dose escalation of adjuvant doxorubicin-based therapy in node-positive breast cancer patients. *Journal of Clinical Oncology* 2007; 23:4287-4297.
13. Yu, D., Jing, T., Liu, B., Yao, J., Tan, M., McDonnell, T. J., and Hung, M. Overexpression of ErbB2 blocks Taxol-induced apoptosis by upregulation of p21<sup>Cip1</sup>, which inhibits p34<sup>Cdc2</sup> kinase. *Molecular Cell* 1998; 2:581-591.

14. Perou, C. M., Sorlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., Pollack, J. R., Ross, D. T., Johnsen, H., Akslén, L. A., Fluge, O., Pergamenschikov, A., Williams, C., Zhu, S. X., Lonning, P. E., Borresen-Dale, A., Brown, P. O., and Botstein, D. Molecular portraits of human breast tumors. *Nature* 2000; 406:747-752.
15. Perez, E. A., Suman, V. J., Davidson, N. E., Martino, S., Kaufinan, P. A., Lingle, W. L., Flynn, P. J., Ingle, J. N., Visscher, D., and Jenkins, R. B. HER2 testing by local, central, and reference laboratories in specimens from the North Central Cancer Treatment Group N9831 Intergroup adjuvant trial. *Journal of Clinical Oncology* 2006; 24:3032-3038.
16. Tokunaga, E., Oki, E., Nishida, K., Koga, T., Egashira, A., Morita, M., Kakeji, Y., and Maehara, Y. Trastuzumab and breast cancer: developments and current status. *International Journal of Clinical Oncology* 2006; 11:199-208.
17. Hanna, W. M. and Kwok, K. Chromogenic *in-situ* hybridization: a viable alternative to fluorescence *in-situ* hybridization in the HER2 testing algorithm. *Modern Pathology* 2006; 19:481-487.
18. Li-Ning-T, E., Ronchetti, R., Torres-Cabala, C., and Merino, M. J. Role of chromogenic *in situ* hybridization(CISH) in the evaluation of HER2 status in breast carcinoma: comparison with immunohistochemistry and FISH. *International Journal of Surgical Pathology* 2005; 13:343-351.
19. Slamon, D. J., Godolphin, W., Jones, L. A., Holt, J. A., Wong, S. G., Keith, D. E., Levin, W. J., Stuart, S. G., Udove, J., Ullrich, A., and Press, M. F. Studies of the HER-2/*neu* proto-oncogene in human breast and ovarian cancer. *Science* 1989; 244:707-712.
20. Tanner, M., Gancberg, D., DiLeo, A., Larsimont, D., Rouas, G., Piccart, M., and Isola, J. Chromogenic *in Situ* hybridization: a practical alternative for fluorescence *in Situ* hybridization to detect HER-2/*neu* oncogene amplification in archival breast cancer samples. *American Journal of Pathology* 2000; 157:1467-1472.
21. Carlson, R. W., Moench, S. J., Hammond, M. E., Perez, E. A., Burstein, H. J., Allred, D. C., Vogel, C. L., Goldstein, L. J., Somlo, G., Gradishar, W. J., Hudis, C. A., Jahanzeb, M., Stark, A., Wolff, A. C., Press, M. F., Winer, E. P., Paik, S., and Ljung, B. M. HER2 testing in breast cancer: NCCN Task Force report and recommendations. *Journal of the National Comprehensive Cancer Network* 2006; 4:S1-S22.
22. Kostler, W. J., Schwab, B., Singer, C. F., Neumann, R., Rucklinger, E., Brodowicz, T., Tomek, S., Niedermayr, M., Hejna, M., Steger, G. G., Krainer, M., Wiltshcke, C., and Zielinski, C. C. Monitoring of serum Her-2/*neu* predicts response and progression-free survival to trastuzumab-based treatment in patients with metastatic breast cancer. *Clinical Cancer Research* 2004; 10:1618-1624.
23. Nahta, R. and Esteva, F. J. Herceptin: mechanisms of action and resistance. *Cancer Letters* 2006; 232:123-138.
24. Fritz, P., Cabrera, C. M., Dippon, J., Gerteis, A., Simon, W., Aulitzky, W. E., and van der Kuip, H. c-erbB2 and topoisomerase II $\alpha$  protein expression independently predict poor survival in primary human breast cancer: a retrospective study. *Breast Cancer Research* 2005; 7:R374-R384.
25. Mano, M. S., Rosa, D. D., DeAzambuja, E., Ismael, G. F. V., and Durbecq, V. The 17q12-q21 amplicon: Her2 and topoisomerase-II  $\alpha$  and their importance to the biology of solid tumors. *Cancer Treatment Reviews* 2007; 33:64-77.
26. Cardoso, F., Durbecq, V., Larsimont, D., Paesmans, M., Leroy, J. Y., Rouas, G., Sotiriou, C., Renard, N., Richard, V., Piccart, M. J., and DiLeo, A. Correlation between complete response to

anthracycline-based chemotherapy and topoisomerase II- $\alpha$  gene amplification and protein overexpression in locally advanced/metastatic breast cancer. *International Journal of Oncology* 2004; 24:201-209.

27. Jarvinen, T. A. H., Tanner, M., Barlund, M., Borg, A., and Isola, J. Characterization of topoisomerase II $\alpha$  gene amplification and deletion in breast cancer. *Genes, Chromosomes & Cancer* 1999; 26:142-150.
28. Mueller, R. E., Parkes, R. K., Andrulis, I. L., and O'Malley, F. P. Amplification of the TOP2A gene does not predict high levels of topoisomerase II alpha protein in human breast tumor samples. *Genes, Chromosomes & Cancer* 2004; 39:288-297.
29. Pritchard, K. I., Shepherd, L. E., O'Malley, F. P., Andrulis, I. L., Tu, D., Bramwell, V. H., and Levine, M. N. HER2 and responsiveness of breast cancer to adjuvant chemotherapy. *New England Journal of Medicine* 2006; 354:2103-2111.
30. Arriola, E., Moreno, A., Varela, M., Serra, J. M., Falo, C., Benito, E., and Escobedo, A. P. Predictive value of HER-2 and topoisomerase II $\alpha$  in response to primary doxorubicin in breast cancer. *European Journal of Cancer* 2006; 42:2954-2960.
31. Knoop, A. S., Knudsen, H., Balslev, E., Rasmussen, B. B., Overgaard, J., Nielsen, K. V., Schonau, A., Gunnarsdottir, K., Olsen, K. E., Mouridsen, H., and Ejlertsen, B. Retrospective analysis of topoisomerase II $\alpha$  amplifications and deletions as predictive markers in primary breast cancer patients randomly assigned to cyclophosphamide, methotrexate, and fluorouracil or cyclophosphamide, epirubicin, and fluorouracil: Danish Breast Cancer Cooperative Group. *Journal of Clinical Oncology* 2005; 23:7483-7490.
32. Depowski, P. L., Rosenthal, S. I., Brien, T. P., Stylos, S., Johnson, R. L., and Ross, J. S. Topoisomerase II $\alpha$  expression in breast cancer: correlation with outcome variables. *Modern Pathology* 2000; 13:542-547.
33. Querzoli, P., Albonico, G., Grazia di Iasio, M., Ferretti, S., Rinaldi, R., Cariello, A., Pedriali, M., Matteuzzi, M., and Nenci, I. Biophenotypes and survival of BRCA1 and TP53 deleted breast cancer in young women. *Breast Cancer Research and Treatment* 2001; 66:135-142.
34. Honrado, E., Benitez, J., and Palacios, J. Histopathology of BRCA1- and BRCA2-associated breast cancer. *Critical Reviews in Oncology/Hematology* 2006; 59:27-39.
35. Mittendorf, E. A., Storrer, C. E., Shriver, C. D., Ponniah, S., and Peoples, G. E. Investigating the combination of trastuzumab and HER2/neu peptide vaccines for the treatment of breast cancer. *Annals of Surgical Oncology* 2006; 13:1085-1098.
36. Harris, M. Monoclonal antibodies as therapeutic agents for cancer. *Lancet* 2004; 5:292-302.
37. Nagata, Y., Lan, K. H., Zhou, X., Tan, M., Esteva, F. J., and Sahin, A. A. PTEN activation contributes to tumor inhibition by trastuzumab, and loss of PTEN predicts trastuzumab resistance in patients. *Cancer Cell* 2004; 6:117-127.
38. Klos, K. S., Zhou, X., Lee, S., Zhang, L., Yang, W., and Nagata, Y. Combined trastuzumab and paclitaxel treatment better inhibits erbB-2-mediated angiogenesis in breast carcinoma through a more effective inhibition of Akt than either treatment alone. *Cancer* 2003; 98:1377-1385.
39. Vogel, C. L., Cobleigh, M. A., Tripathy, D., Gutheil, J. C., Harris, L. N., Fehrenbacher, L., Slamon, D. J., Murphy, M., Novotny, W. F., Burchmore, M., Shak, S., Stewart, S. J., and Press, M. Efficacy

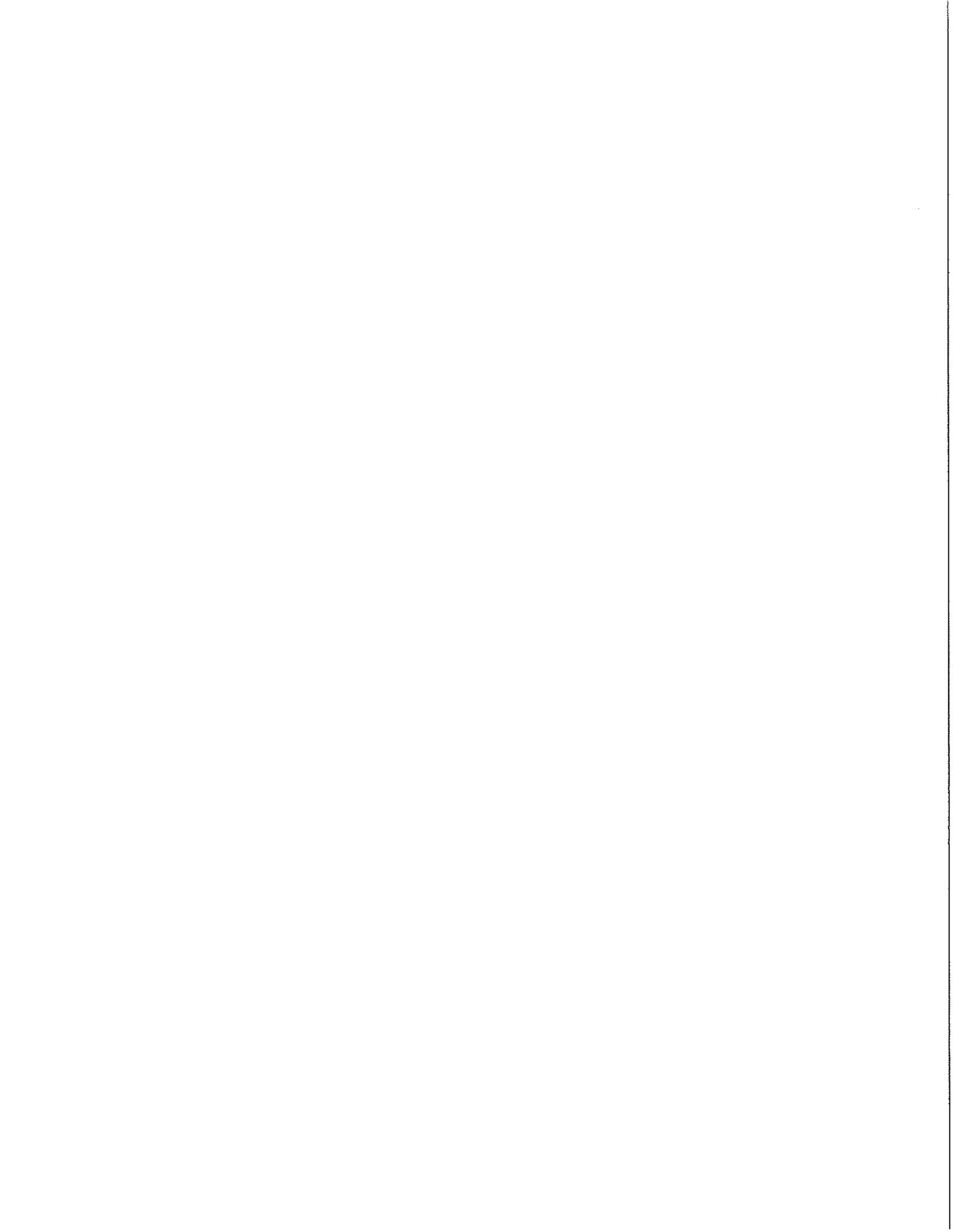
and safety of trastuzumab as a single agent in first-line treatment of HER2-overexpressing metastatic breast cancer. *Journal of Clinical Oncology* 2002; 20:719-726.

40. Baselga, J., Carbonell, X., Castaneda-Soto, N., Clemens, M., Green, M., Harvey, V., Morales, S., Barton, C., and Ghahramani, P. Phase II study of efficacy, safety, and pharmacokinetics of trastuzumab monotherapy administered on a 3-weekly schedule. *Journal of Clinical Oncology* 2005; 23:2162-2171.
41. Orlando, L., Cardillo, A., Ghisini, R., Rocca, A., Balduzzi, A., Torrissi, R., Peruzzotti, G., Goldhirsch, A., Pietri, E., and Colleoni, M. Trastuzumab in combination with metronomic cyclophosphamide and methotrexate in patients with HER-2 positive metastatic breast cancer. *BMC Cancer* 2006; 6:225.
42. Cobleigh, M. A., Vogel, C. L., Tripathy, D., Robert, N. J., Scholl, S., Fehrenbacher, L., Wolter, J. M., Paton, V., Shak, S., Lieberman, G., and Slamon, D. J. Multinational study of the efficacy and safety of humanized anti-HER2 monoclonal antibody in women who have HER2-overexpressing metastatic breast cancer that has progressed after chemotherapy for metastatic disease. *Journal of Clinical Oncology* 1999; 17:2639-2648.
43. Pegram, M., Konecny, G. E., O'Callaghan, C., Beryt, M., Pietras, R., and Slamon, D. J. Rational combinations of trastuzumab with chemotherapeutic drugs used in the treatment of breast cancer. *Journal of the National Cancer Institute* 2004; 96:739-749.
44. Slamon, D. J., Leyland-Jones, B., Shak, S., Fuchs, H., Paton, V., Bajamonde, A., Fleming, T., Eiermann, W., Wolter, J., Pegram, M., Baselga, J., and Norton, L. Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *New England Journal of Medicine* 2001; 344:783-792.
45. Joensuu, H., Kellokumpu-Lehtinen, P., Bono, P., Alanko, T., Kataja, V., Asola, R., Utriainen, T., Kokko, R., Hemminki, A., Tarkkanen, M., Turpeenniemi-Hujanen, T., Jyrkkio, S., Flander, M., Helle, L., Ingalsuo, S., Johansson, K., Jaaskelainen, A., Pajunen, M., Rauhala, M., Kaleva-Kerola, J., Salminen, T., Leinonen, M., Elomaa, I., and Isola, J. Adjuvant docetaxel or vinorelbine with or without trastuzumab for breast cancer. *New England Journal of Medicine* 2006; 354:809-820.
46. Romond, E. H., Perez, E. A., Bryant, J., Suman, V. J., Geyer, C. E., Davidson, N. E., Tan-Chiu, E., Martino, S., Paik, S., Kaufman, P. A., Swain, S. M., Pisansky, T. M., Fehrenbacher, L., Kutteh, L. A., Vogel, V. G., Visscher, D., Yothers, G., Jenkins, R. B., Brown, A. M., Dakhil, S. R., Mamounas, E. P., Lingle, W. L., Klein, P. M., Ingle, J. N., and Wolmark, N. Trastuzumab plus adjuvant chemotherapy for operable HER2-positive breast cancer. *New England Journal of Medicine* 2005; 353:1673-1684.
47. Piccart-Gebhart, M. J., Procter, M., Leyland-Jones, B., Goldhirsch, A., Untch, M., Smith, I., Gianni, L., Baselga, J., Bell, R., Jackisch, C., Cameron, D., Dowsett, M., Barrios, C. H., Steger, G. G., Huang, C., Andersson, M., Inbar, M., Lichinitser, M., Lang, I., Nitz, U., Iwata, H., Thomssen, C., Lohrisch, C., Suter, T. M., Ruschoff, J., Suto, T., Giatromanolaki, S., Ward, C., Strahle, C., McFadden, E., Dolci, M. S., and Gelber, R. D. Trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer. *New England Journal of Medicine* 2005; 353:1659-1672.
48. Slamon, D. J., Eiermann, W., Robert, N. J., Pienkowski, T., Martin, M., Pawlicki, M., Chan, A., Smylie, M., Liu, M., Falkson, C., Pinter, T., Fornander, T., Shiftan, T., Valero, V., Von Minckwitz, G., Mackey, J., Tabah-Fisch, I., Buyse, M., Lindsay, M. A., Riva, A., Bee, V., Pegram, M., Press, M., and Crown, J. Phase III randomized trial comparing doxorubicin and cyclophosphamide followed by docetaxel (AC -> T) with doxorubicin and cyclophosphamide followed by docetaxel and trastuzumab (AC -> TH) with docetaxel, carboplatin and trastuzumab (TCH) in HER2 positive early breast cancer patients: BCIRG 006 study. *Breast Cancer Research and Treatment* 2006; 94:S5.

49. Hurley, J., Doliny, P., Reis, I., Silva, O., Gomez-Fernandez, C., Velez, P., Pauletti, G., Powell, J. E., Pegram, M. D., and Slamon, D. J. Docetaxel, cisplatin, and trastuzumab as primary systemic therapy for human epidermal growth factor receptor 2-positive locally advanced breast cancer. *Journal of Clinical Oncology* 2006; 24:1831-1838.
50. Buzdar, A. U., Ibrahim, N. K., Francis, D., Booser, D. J., Thomas, E. S., Theriault, R. L., Pusztai, L., Green, M. C., Arun, B. K., Giordano, S. H., Cristofanilli, M., Frye, D. K., Smith, T. L., Hunt, K. K., Singletary, S. E., Sahin, A. A., Ewer, M. S., Buchholz, T. A., Berry, D., and Hortobagyi, G. N. Significantly higher pathologic complete remission rate after neoadjuvant therapy with trastuzumab, paclitaxel, and epirubicin chemotherapy: results of a randomized trial in human epidermal growth factor receptor 2-positive operable breast cancer. *Journal of Clinical Oncology* 2005; 23:3676-3685.
51. Bartsch, R., Wenzel, C., Hussian, D., Pluschnig, U., Sevelde, U., Koestler, W., Altorjai, G., Locker, G. J., Mader, R., Zielinski, C. C., and Steger, G. G. Analysis of trastuzumab and chemotherapy in advanced breast cancer after the failure of at least one earlier combination: an observational study. *BMC Cancer* 2006; 6:63.
52. Guarneri, V., Lenihan, D. J., Valero, V., Durand, J., Broglio, K., Hess, K. R., Michaud, L. B., Gonzalez-Angulo, A. M., Hortobagyi, G. N., and Esteva, F. J. Long-term cardiac tolerability of trastuzumab in metastatic breast cancer: the M.D. Anderson Cancer Center experience. *Journal of Clinical Oncology* 2006; 24:4107-4115.
53. Ewer, M., Vooletich, M. T., Durand, J., Woods, M. L., Davis, J. R., Valero, V., and Lenihan, D. J. Reversibility of trastuzumab-related cardiotoxicity: new insights based on clinical course and response to medical treatment. *Journal of Clinical Oncology* 2005; 23:7820-7826.
54. Tan-Chiu, E., Yothers, G., Romond, E., Geyer, C. E., Ewer, M., Keefe, D., Shannon, R. P., Swain, S. M., Brown, A., Fehrenbacher, L., Vogel, V. G., Seay, T. E., Rastogi, P., Mamounas, E. P., Wolmark, N., and Bryant, J. Assessment of cardiac dysfunction in a randomized trial comparing doxorubicin and cyclophosphamide followed by paclitaxel, with or without trastuzumab as adjuvant therapy in node-positive, human epidermal growth factor receptor 2-overexpressing breast cancer: NSABP B-31. *Journal of Clinical Oncology* 2005; 23:7811-7819.
55. Seidman, A. D., Hudis, C. A., Pierri, M. K., Shak, S., Paton, V., Ashby, M., Murphy, M., Stewart, S. J., and Keefe, D. Cardiac dysfunction in the trastuzumab clinical trials experience. *Journal of Clinical Oncology* 2002; 20:1215-1221.
56. Arteaga, C. L. Trastuzumab, an appropriate first-line single-agent therapy for HER2-overexpressing metastatic breast cancer. *Breast Cancer Research* 2003; 5:96-100.
57. Jerome, L., Alami, N., Belanger, S., Page, V., Yu, Q., Paterson, J., Shiry, L., Pegram, M., and Leyland-Jones, B. Recombinant human insulin-like growth factor binding protein 3 inhibits growth of human epidermal growth factor receptor-2-overexpressing breast tumors and potentiates Herceptin activity *in vivo*. *Cancer Research* 2006; 66:7245-7252.
58. Nahta, R. and Esteva, F. J. Trastuzumab: triumphs and tribulations. *Oncogene* 2007; 26:3637-3643.
59. Kiewe, P., Hasmmuller, S., Kahlert, S., Heinrigs, M., Rack, B., Marme, A., Korfel, A., Jager, M., Lindhofer, H., Sommer, H., Thiel, E., and Untch, M. Phase I trial of the trifunctional anti-HER2 x anti-CD3 antibody ertumaxomab in metastatic breast cancer. *Clinical Cancer Research* 2006; 12:3085-3091.
60. Peoples, G. E., Gurney, J. M., Hueman, M. T., Woll, M. M., Ryan, G. B., Storrer, C. E., Fisher, C., Shriver, C. D., Ioannides, C. G., and Ponniah, S. Clinical trial results of a HER2/*neu* (E75) vaccine

to prevent recurrence in high-risk breast cancer patients. *Journal of Clinical Oncology* 2005; 23:7536-7545.

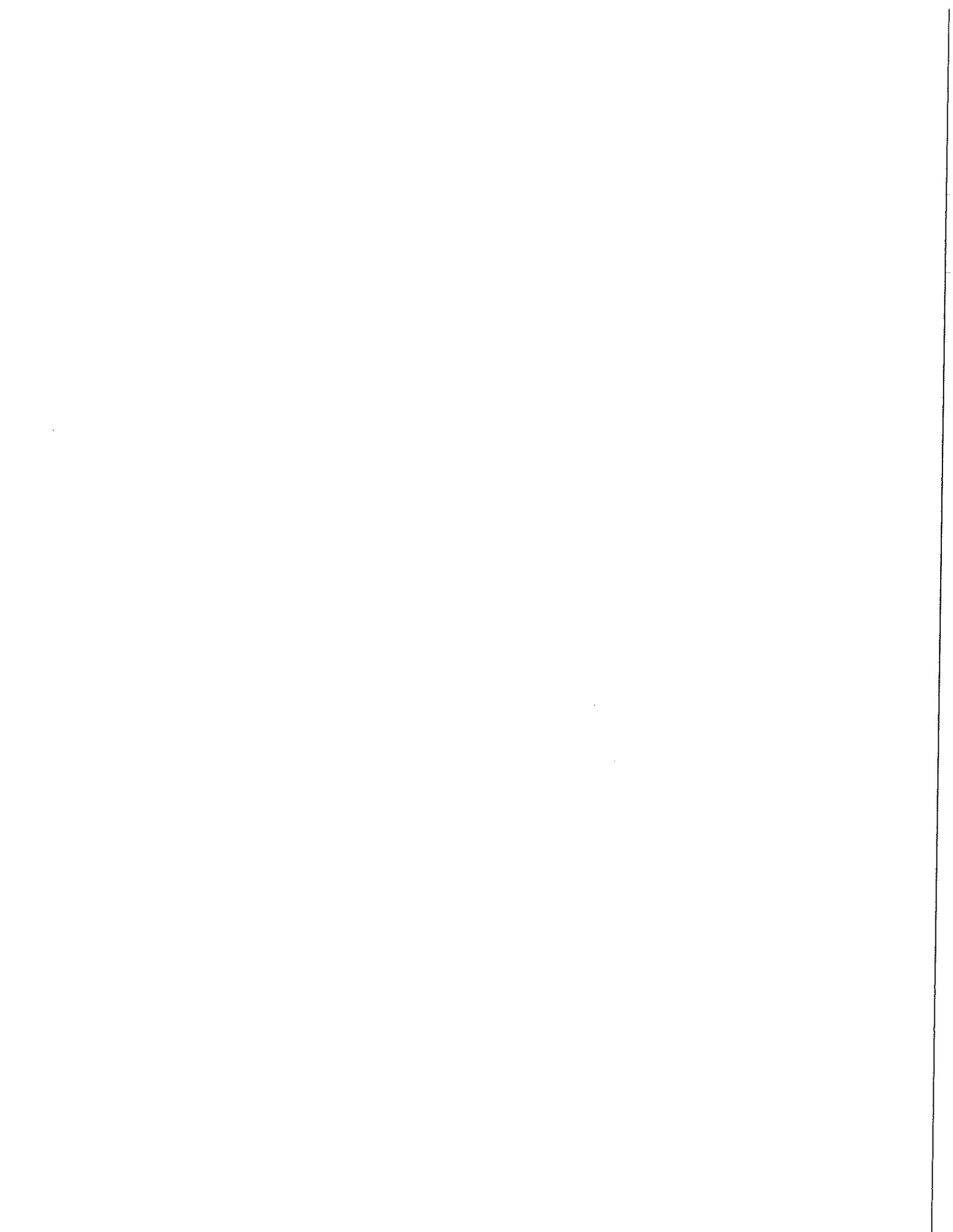
61. Mittendorf, E. A., Gurney, J. M., Storrer, C. E., Shriver, C. D., Ponniah, S., and Peoples, G. E. Vaccination with a HER2/*neu* peptide induces intra- and inter-antigenic epitope spreading in patients with early stage breast cancer. *Surgery* 2006; 139:407-418.
62. Mittendorf, E. A., Storrer, C. E., Foley, R. J., Harris, K., Jama, Y., Shriver, C. D., Ponniah, S., and Peoples, G. E. Evaluation of the HER2/*neu*-derived peptide GP2 for use in a peptide-based breast cancer vaccine trial. *Cancer* 2006; 106:2309-2317.
63. Kaufman, P. A., Broadwater, G., Lezon-Geyda, K., Dressler, L. G., Berry, D. A., Friedman, P., Winer, E. P., Hudis, C. A., Ellis, M. J., Seidman, A. D., and Harris, L. N. CALGB 150002: Correlation of HER2 and chromosome 17 (ch17) copy number with trastuzumab (T) efficacy in CALGB 9840, paclitaxel (P) with or without T in HER2+ and HER2- metastatic breast cancer (MBC). *Journal of Clinical Oncology* 25[18S]. 2007.



**The Clinical Breast Care Project: an important resource for studying breast cancer in African American women.**

Lori A. Field, Rachel E. Ellsworth, Veronica Gutchell, Jeffrey A. Hooke, Craig D. Shriver.

Age at diagnosis, pathological characteristics, and tumor behavior differ between African American (AAW) and Caucasian women (CW) with breast cancer, with AAW having more aggressive tumors and higher mortality rates. Although both societal and molecular contributions to these disparities have been suggested, the African American population has traditionally been under-represented in research and clinical protocols, limiting the power of epidemiologic and molecular studies to provide better understanding of disease pathogenesis in this minority population. The Clinical Breast Care Project (CBCP) has developed a large tissue and blood repository from patients undergoing treatment for breast cancer, with previous history of breast cancer, counseled in the Risk Reduction Clinic, screened by routine mammography, or undergoing elective reductive mastoplasty. Recruitment of AAW into the CBCP was successful; 25% of the 2,454 female patients were African American, including 35% disease-free, 3% high-risk, 40% benign, 8% pre-invasive and 14% with invasive breast disease. More than 500 data fields regarding lifestyle choices, socioeconomic status, health history and geography were collected from all participants, and all consenting individuals provided blood specimens for genomic and proteomic studies. Tissues were collected from all patients undergoing surgical procedures using protocols that preserve the macromolecules for downstream research applications. In addition, patients were followed after diagnosis, with 85% of patients providing ongoing and updated demographic and clinical information. Thus, recruitment efforts in the CBCP have resulted in collection of well-annotated information and research-quality specimens from a large number of AAW. This unique resource will allow for the identification of biological and environmental factors associated with the occurrence, detection, prognosis, and survival of breast cancer in AAW.



Maskery

## **A Bayesian Derived Network of Breast Pathology Diagnoses**

Susan M. Maskery, PhD, Windber Research Institute  
Hai Hu, PhD, Windber Research Institute  
Jeffrey Hooke, MD, Department of Pathology, Walter Reed Army Medical Center  
Craig D. Shriver, MD, Department of General Surgery, Walter Reed Army Medical Center  
Michael N. Liebman, PhD, Windber Research Institute

Corresponding Author: Susan Maskery  
Windber Research Institute  
620 7<sup>th</sup> Street  
Windber, PA 15963  
814 361-6949 (phone)  
814 467-6334 (fax)  
[s.maskery@wriwindber.org](mailto:s.maskery@wriwindber.org)

Maskery

## **Abstract**

Our objective is to use Bayesian network analysis as a tool for the discovery of novel patterns in breast pathology co-occurrence. The elucidation of such patterns will guide future molecular breast cancer research. In this paper we present the validation and verification of a machine learned Bayesian network of breast pathology co-occurrence patterns. To build the network, the present/not present occurrences of 69 breast disease and cancer diagnoses from 1409 pathology reports are used. All pathology reports are written by a single pathologist. The resulting network has 32 diagnosis nodes interconnected by 50 arcs. Each arc represents a co-occurrence prediction. Model validation is assessed through the original structure remaining after random exclusion of 25%, 50%, and 75% of the 1409 pathology report dataset. The structure of the network is stable- random removal 75% of the original data set (1057 pathology reports) leaves 80% of the original network intact. Model verification is assessed by review of breast pathology literature for each arc in the network. Almost all co-occurrences (96%), as represented by arcs in the network, are found in the breast pathology literature. In conclusion, the Bayesian network of breast pathology co-occurrence presented here is robust and verifiable. Molecular experiments are planned to explore the possible clinical meaning of the co-occurrence patterns found.

## **Introduction**

Breast cancer is a common disease in the Western world. In the United States 1 in 7 women will develop breast cancer at some point in their life [1]. Many laboratories actively research this disease from its molecular profile [reviewed in 2, 3] to its epidemiology [4]. Despite intensive research and the accumulation of vast amounts of molecular data, clinical care for invasive breast cancer remains based upon gross pathological features supplemented with immunohistochemistry test results.

## Maskery

Since the pathology of breast tumors remains the most reliable clinical predictor of risk for breast cancer progression, we focus on the pathology of breast cancer and breast disease. Breast disease in the context of breast cancer risk has been looked at epidemiologically. Early studies identified possible pre-cancerous lesions in the breast based upon non-cancerous pathologies that co-occur with breast cancer [5] and epidemiological studies have ascribed risk ratios for cancer development to these various pre-cancerous pathologies [6, 7]. In recent years most researchers have focused more on the molecular features of breast tumors such as RNA expression [8] and genomic instability [9]. However, although molecular markers for breast cancer prognosis may be more advanced than those of other cancers, the known markers do not always predict outcome correctly and new markers that might improve outcome prediction are very difficult to find under current discovery paradigms [10].

We hypothesize that patterns of pathology co-occurrence may refine the selection of tumor samples for molecular analysis, such that heterogeneity between tumors is decreased and thus, enable the linkage of genetic analysis results to pathology patterns. To derive pathology clusters from our dataset, first we develop a network of breast pathology interactions. For this task we create a Bayesian network of the co-occurrence of pathology diagnoses.

Bayesian networks can be used to model uncertain relationships between variables of interest. A Bayesian network is a directed acyclic graph (DAG) over a domain composed of variables of interest. Variables of interest, in this case pathology diagnoses, are represented as nodes in the network. Based upon scoring criterion, strongly associated variables become nodes in the network. Nodes are connected by arcs. Inherent in the arcs connecting network nodes are parent child relationships that define the joint probability distribution (JPD) of the entire network, such that the JPD is defined as the product of the conditional probabilities of all child

## Maskery

variables and the prior probabilities of all variables without parents (root variables). The conditional distribution of each “child” node can be defined using the node’s immediate parent nodes [11]. In this way, the probability of any scenario (variable  $v_1=x_1$ , variable  $v_2=x_2$  ... variable  $v_n=x_n$ .) can be calculated with knowledge of the parent child relationships between variables. For example, for a simple three node network where node C is the child of nodes A and B, and nodes A and B do not have a parent variable, the JPD is:

$$P(A,B,C)=P(C|A,B)*P(A)*P(B) \quad \text{Eqn. 1}$$

In this article we validate and verify a Bayesian network analysis based approach to deriving the topology of probabilistic patterns of breast disease and cancer diagnoses. Access to a large, well annotated, diverse set of pathology reports is necessary to the successful adaptation of Bayesian network technology to this research problem. We have access to 1409 de-identified pathology reports. Importantly, a single pathologist provides the pathology diagnoses for all biopsy patients in the dataset used to develop the network. In this way our data is not confounded by inter-pathologist variability or by the multiple ways the same lesion may be named in a pathology report. Importantly for this project, for each clinical case any and all pathologies present in *all* the biopsied tissue samples are included in the pathology report. This reporting method differs from a typical clinical pathology report where usually only the most severe diagnoses are recorded.

Using this high quality patient pathology data, we have successfully applied Bayesian network analysis to identify the topology of breast disease interactions. Future work will involve the use of pathology networks, with the addition of epidemiological data, treatment, and outcome data, to define pathology patterns of breast cancer presentation. These patterns will be used to design genomic and proteomic experiments to better understand the biological reason for why

## Maskery

different pathologies are observed to co-occur, and in the process gain a deeper understanding of the molecular origins of breast cancer.

## Materials and Methods

### *Dataset*

Data is drawn from 1409 CBCP pathology checklists. Each checklist has 131 possible diagnoses of which any and all that are found in a patient's biopsy or biopsies in the case of cancer are recorded. Diagnoses added after the study began or non-breast disease diagnoses (i.e. axillary lymph node status) are excluded. This left 114 diagnoses out of the original 131 in the dataset. Of these 114 diagnoses 69 are present in the study population. Disease classifications for these diagnoses are: inflammatory and reactive lesions, miscellaneous benign lesions, papillary lesions, biphasic tumors, benign and atypical hyperplasias, in situ carcinomas, infiltrating carcinomas common and special types, diagnoses of the nipple, benign mesenchymal lesions, and sarcomas. Pathology reports may contain information from multiple procedures. For example a woman diagnosed with cancer may have a core needle biopsy, an excisional biopsy, and a re-excision on the same breast over the course of days to weeks to months. Diagnoses in a single checklist may have been made off of tissue from any of these procedures.

### *Bayesian Analysis*

Bayesian networks are learned from the diagnosis data in binary (0-not present, 1-present) form. FasterAnalytics v5.1.7 (DecisionQ Corp., Washington, DC) software is used to build all models. The network structure is machine learned from the data; no expert opinion information is involved in learning the network structure. Details of the learning algorithm are in

## Maskery

reference 11. Briefly, FasterAnalytics uses an algorithm primarily based upon the Basic Greedy Search algorithm to which an additional optimization algorithm, “queue learning,” has been added[12]. MDL (minimum description length) scoring with bias = 1.0 is used to optimize the network. The number of “tries” the software can make to beat the last best score is set to 300. Each time the software is run a different number is used to seed the random number generator used in the learning algorithm. The structure with the highest MDL score is chosen as the “best” model. Through multiple tests we determined that running the software until the same best score is found at least 3 times reliably finds the highest scored model for each dataset.

## *Model Verification*

To estimate the stability of the derived Bayesian network model in relation to our dataset, we compare the original dataset derived network to networks derived from datasets with 25%, 50%, and 75% of the original data removed. To accomplish this comparison, we measure  $\mathfrak{S}_s$ - the fraction of the original network remaining in a network derived from a subset of the original dataset. To measure  $\mathfrak{S}_s$ , a metric, Network Arc Summation (NAS), is created. To calculate NAS, the “strength” of each arc in the complete dataset network is determined. The strength of each arc is defined as the change in MDL score between the full network and the network with just that arc removed. In this way it is possible to weight the importance of each arc in the construction of the complete dataset network. The complete dataset network’s NAS ( $NAS_o$ ) is the summation of all its arc strengths. To compare each non-complete dataset derived network to the complete dataset derived network, the non-complete dataset derived network’s NAS ( $NAS_s$ ) is calculated and divided by  $NAS_o$ . The arc strengths from the original dataset are used to calculate  $NAS_s$  such that, if an arc is found in both networks than that arc is weighted in  $NAS_s$  in

## Maskery

the same way as it is in  $NAS_o$ . However, if an arc is present in the non-complete dataset derived network and not present in the original dataset-derived network, that arc is weighted as 0. In this way the fraction of the original network remaining in a network derived from a subset of the data ( $\mathfrak{S}$ ) is calculated as

$$\mathfrak{S}_s = \frac{NAS_s}{NAS_o} \quad \text{Eqn. 2}$$

To summarize, the following steps are used to estimate the stability of the final Bayesian network as a function of variation in the dataset.

- 1) Find the highest MDL scored network for the original dataset.
- 2) Calculate the strength of each arc in the network
- 3) Calculate the NAS for the original dataset.
- 4) Create a new dataset where 25% of the records are randomly removed.
- 5) Find the highest MDL scored network for this dataset.
- 6) Calculate NAS for this network.
- 7) Calculate  $\mathfrak{S}$
- 8) Repeat steps 4-7 three times
- 9) Calculate the average and standard deviation for  $\mathfrak{S}_{25\%}$
- 10) Calculate a new dataset where 50% of the records are randomly removed.
- 11) Repeat steps 5-8.
- 12) Calculate the average and standard deviation for  $\mathfrak{S}_{50\%}$
- 13) Calculate a new dataset where 75% of the records are randomly removed.
- 14) Repeat steps 5-8.
- 15) Calculate the average and standard deviation for  $\mathfrak{S}_{75\%}$

Maskery

### *Model Validation*

There is a great deal of literature dedicated to breast pathology. This literature will be used to verify the model. Expert knowledge from a pathologist is also used to verify model results and to identify any incorrect findings. The model is considered fully verifiable if all connections in the network are found in the breast pathology literature or confirmed by expert knowledge.

### **Results**

#### *The Network*

The 69 diagnoses that occur in the patient population are entered into the analysis. Of the 69 entered, 32 diagnoses have associations strong enough for inclusion as nodes in the network.

Table 1 lists these diagnoses and their individual diagnosis code in the network.

Table 1. Abbreviation and Name for all Diagnoses in the Bayesian Analysis Derived Network

Abbreviation	Diagnosis Name	Abbreviation	Diagnosis Name
MicroCalcs	Microcalcifications	ModHUT	Moderate intraductal hyperplasia
FC	Fibrocystic changes	FloridHUT	Florid intraductal hyperplasia
SFibrosis	Stromal fibrosis	CCH	Columnar cell hyperplasia
Cysts	Cysts	ADH	Atypical ductal hyperplasia
ApoMeta	Apocrine metaplasia	AAH	Atypical apocrine hyperplasia
SAdeno	Sclerosing adenosis	Gynecomastia	Gynecomastia
FAdeno	Florid adenosis	ALH	Atypical lobular hyperplasia
AdenoNOS	Adenosis, not otherwise specified	DCISNonComedo	Ductal carcinoma in situ not comedo
RadScar	Radial Scar	DCISComedo	Ductal carcinoma in situ comedo
CollSph	Collagenous spherulosis	LCIS	Lobular carcinoma in situ
MultPap	Multiple (peripheral) papillomas	IDC	Infiltrating ductal carcinoma

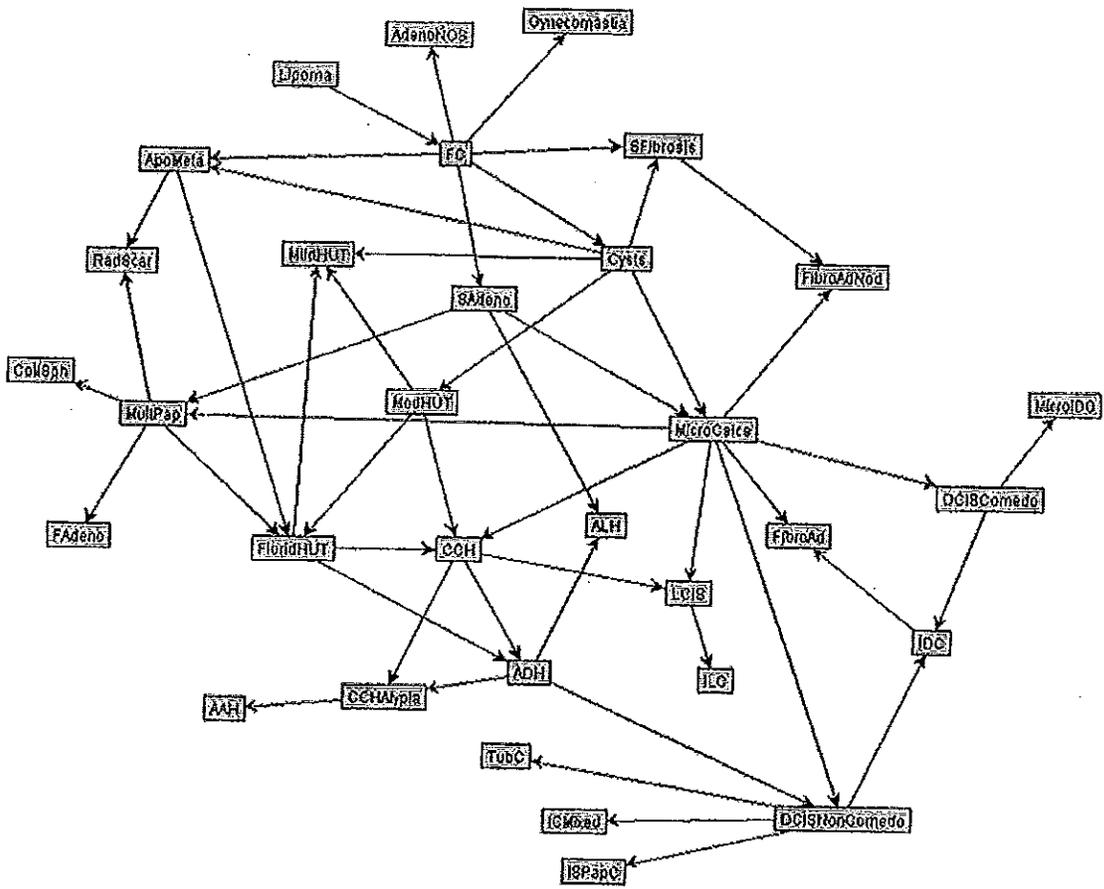
Maskery

ISPapC	In situ papillary carcinoma	MicroIDC	Microinvasive ductal carcinoma
FibroAd	Fibroadenoma	ILC	Infiltrating lobular carcinoma
FibroAdN	Fibroadenomatoid nodule	TubC	Tubular carcinoma
CCHAtypia	Columnar cell hyperplasia with atypia	ICMixed	Infiltrating carcinoma, mixed ductal and lobular
MildHUT	Mild intraductal hyperplasia	Lipoma	Lipoma

Figure 1 is a visual representation of the network.

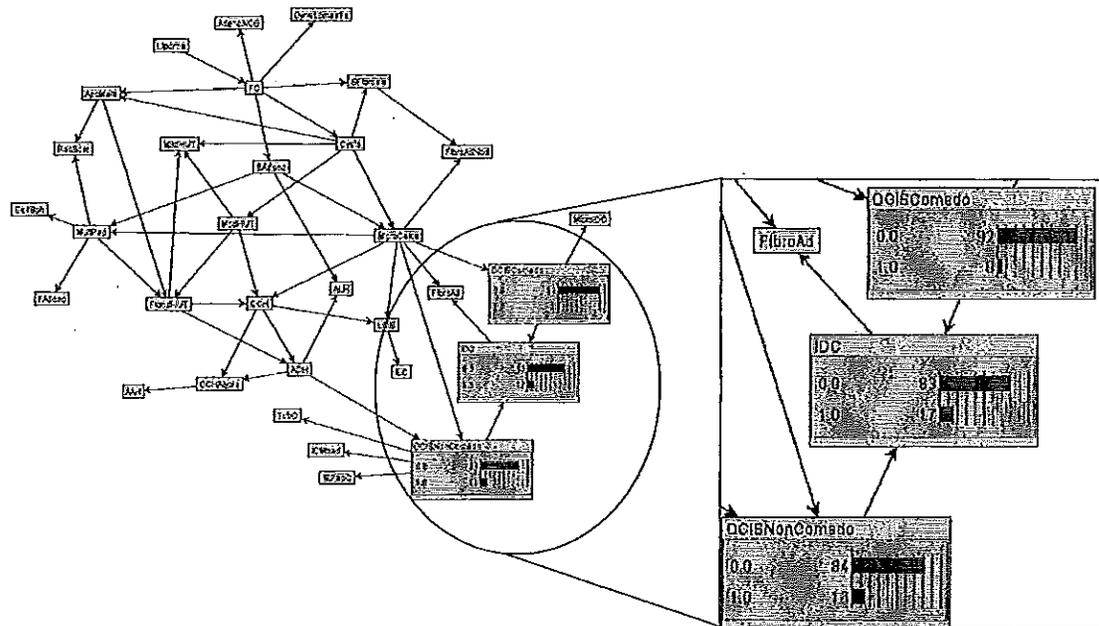
A.

Maskery



B.

## Maskery

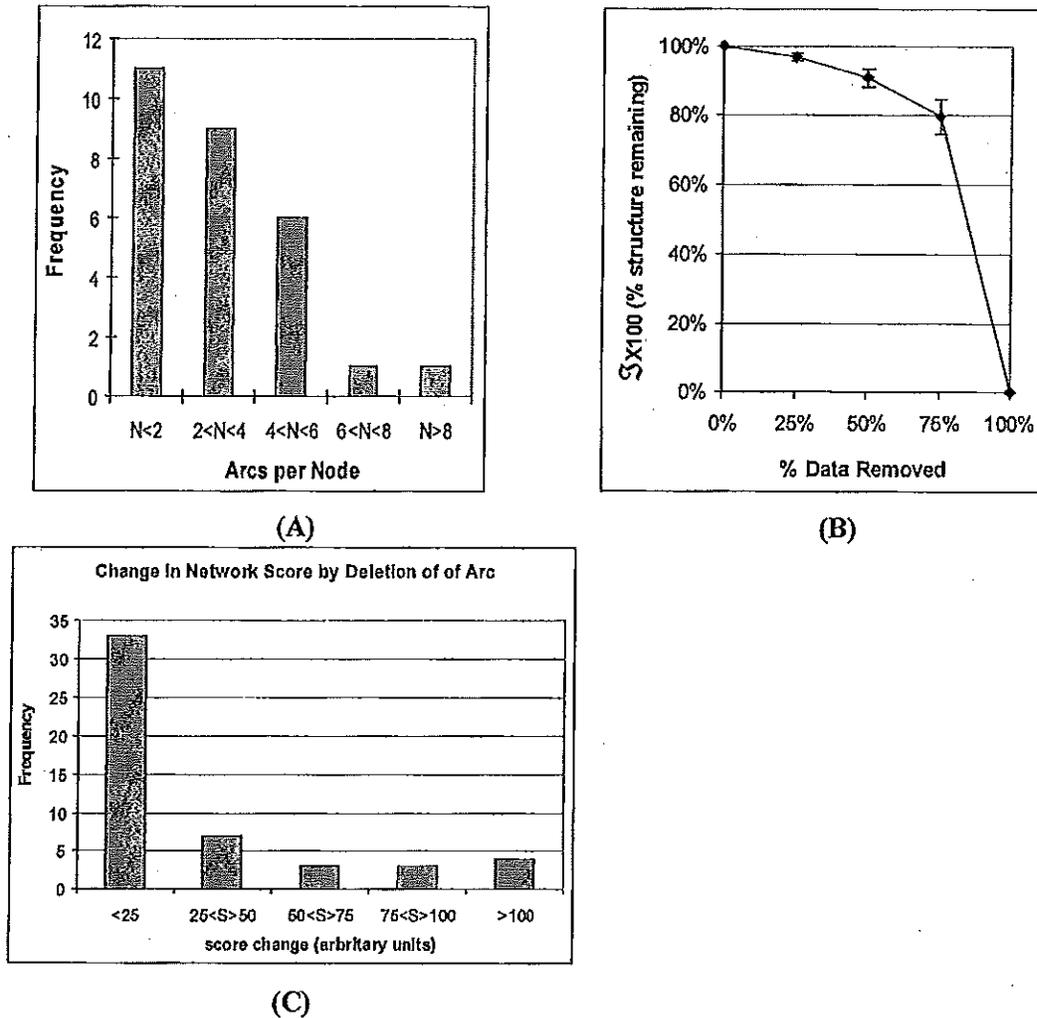


**Figure 1 Bayesian Network derived from Pathology Diagnosis Data**

A. Each node is a diagnosis whose complete name is listed in Table 1. Arrows do *not* imply causality, but instead are related to independence statements embedded in the structure of the network. B. It is possible to enlarge each node in the network to show the percent frequency of each diagnosis in the population. This is visualized in the enlarged cutout.

The average number of arcs per node is 3 and the maximum number of observed arcs per node is 9. Between the 32 network nodes are 50 arcs. Looking at the network structure the node with the largest number of arcs is MicroCalc (microcalcifications). The second most connected node, with 7 arcs, is FC (fibrocystic changes). Several nodes have 6 arcs each, including: MultiPap (multiple papillomas), FloridHUT (intraductal hyperplasia florid), CCH (columnar cell hyperplasia), and DCISNonComedo (DCIS non-Comedo). The frequency of arcs per node is summarized in a histogram in Fig. 2A.

## Maskery



**Figure 2 Network Diagnostics**

(A) Histogram of arcs per node in Figure 1. (B) Fraction of the original network remaining ( $\mathfrak{S}$ ) after the random removal of 25%, 50%, and 75% of the records in the original dataset. Error bars are the standard deviation in network scores scaled to  $\mathfrak{S}$ . (C) The strength of each arc in the original dataset derived network is calculated as the change in network score by its deletion. Plotted in C is a histogram of the number of arcs whose score change falls between the ranges on the x-axis.

To test the stability of the network in relation to our original dataset, we randomly removed 25%, 50%, and 75% of the cases in the initial data set, and measured what fraction of the original network remained in the networks constructed from subsets of the data. The result of this analysis is Fig. 2B. As seen from Fig. 2B removing 75% of the pathology reports can still

## Maskery

reconstitute a network that has ~80% of the original network's structure. Based on this result we conclude the network built from the original data set is stable.

The measurement of network structure is based upon arc strength, and arcs in the network can vary greatly in their measured strength (calculated from the change in network MDL score by their removal). The median arc strength is 17, and arc strength ranges from a low of <1 to a high of 934. Fig.2C is a histogram of the change in network score (i.e. arc strength) broken down in increments of 25.

## Model Validation

Any arc found in the network represents a reproducible co-occurrence for the two diagnoses it connects. All these co-occurrence patterns are verified with published evidence or expert opinion. Of the 49 arcs which strength greater than 1, published evidence and pathologist knowledge directly support 47 identified co-occurrences. The 2 arcs not directly verifiable with published evidence or expert opinion (columnar cell hyperplasia with atypia and atypical apocrine metaplasia; atypical ductal hyperplasia and atypical lobular hyperplasia) do not seem implausible to expert opinion. Each arc, the diagnoses whose co-occurrence it implies, and the evidence supporting this co-occurrence are listed in Table 2.

Table 2 Literature Based Evidence for Network Arcs

Pathology 1	Pathology 2	Arc Strength	Published Evidence
Fibrocystic changes (FC)	Cysts	934	In our pathologist's classification scheme, by definition, cyst is a sub-diagnosis of fibrocystic changes.
Ductal carcinoma in situ not Comedo	Invasive ductal carcinoma (IDC)	192	Retrospective and cytogenic studies point toward DCIS as a non-obligate precursor to

Maskery

(DCISNonComedo)			IDC [13].
Ductal carcinoma in situ Comedo (DCISComedo)	Invasive ductal carcinoma (IDC)	176	Retrospective and cytogenic studies point toward DCIS as a non-obligate precursor to IDC [13].
Fibrocystic changes (FC)	Sclerosing adenosis (SAdeno)	174	Sclerosing adenosis (adenosis more generally) commonly occurs as part of fibrocystic changes [14 pg. 139 ].
Lobular carcinoma in situ (LCIS)	Invasive lobular carcinoma (ILC)	91	Reported frequency of LCIS with ILC varies from 65%-98% cases of ILC [14 pg 635].
Moderate intraductal hyperplasia (ModHUT)	Cysts	84	Frequent coexistence is seen between fibrocystic changes (including cysts) and ductal hyperplasia [14 pg. 203-204].
Cysts	Microcalcifications (MicroCalcs)	78	Microcalcifications are commonly associated with microcysts, especially those with apocrine epithelium [14 pg. 946].
Fibrocystic changes (FC)	Stromal fibrosis (SFibrosis)	73	In our pathologist's classification scheme, by definition, stromal fibrosis is a sub-diagnosis of fibrocystic changes
Sclerosing adenosis (SAdeno)	Microcalcifications (MicroCalcs)	65	Microcalcifications are frequently formed in sclerosing adenosis [14 pg. 139].
Fibrocystic changes (FC)	Gynecomastia	55	This is an exclusionary finding - gynecomastia (male breast enlargement) excludes all other common female breast pathologies. It is probably linked to FC because this is the most

Maskery

			common female breast lesion.
Moderate intraductal hyperplasia (ModHUT)	Florid intraductal hyperplasia (FloridHUT)	43	This is an exclusionary finding based on how intraductal hyperplasias are recorded, in which only the most severe intraductal hyperplasia is recorded. Hence, a diagnosis of moderate intraductal hyperplasia excludes a diagnosis of florid intraductal hyperplasia in the CBCP classification scheme.
Atypical ductal hyperplasia (ADH)	Columnar cell hyperplasia (CCH)	42	Some authorities believe columnar cell lesions are non-obligate pre-cursors to ADH and well-differentiated DCIS [15].
Florid intraductal hyperplasia (FlroidHUT)	Apocrine metaplasia (ApoMeta)	38	Frequent coexistence is seen between fibrocystic changes (including apocrine metaplasia) and ductal hyperplasia [14 pg. 203-204].
Columnar cell hyperplasia with atypia (CCHAtypia)	Columnar cell hyperplasia (CCH)	33	CCH and CCH with atypia are part of the continuum of columnar cell lesions, and it is not uncommon for more than one columnar cell lesion type to exist in the same breast [16].
Columnar cell hyperplasia (CCH)	Florid intraductal hyperplasia (FlroidHUT)	25	CCH and Hyperplasia of the Usual Type (HUT) have been observed to co-occur[9]. Florid intraductal hyperplasia is a subset of HUT.
Atypical ductal hyperplasia (ADH)	Florid intraductal hyperplasia	25	Although the role of intraductal hyperplasia as a non-obligate precursor to ADH is

Maskery

			The differentiation between LCIS and atypical lobular hyperplasia is difficult [14 pg. 611]. LCIS can be found in sclerosing adenosis [14 pg. 601].
Ductal carcinoma in situ not Comedo (DCISNonComedo)	Tubular carcinoma (TubC)	1	DCIS is found in the majority of tubular carcinomas [14 pg. 373].

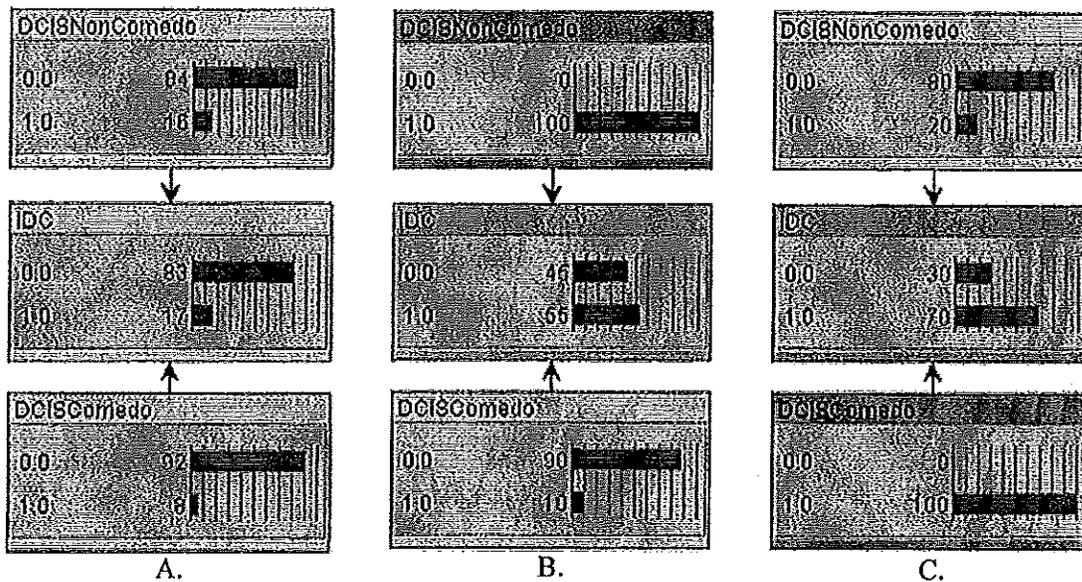
This verification analysis does not address lesions whose co-occurrence is expected from literature, but that do not appear in the network. For example, the lesion RadScar (radial scar) is thought to be associated with breast cancer by several investigators[21]. However, in our analysis we do not see this association included in the network. The reason for this disparity may be the relative rarity of the radial scar diagnosis in our population (population frequency – 3%), making all but the most solid associations with this lesion difficult to find. Another reason may be that radial scar and breast cancer truly do not co-occur more frequently than expected by chance given their relative frequencies in breast biopsies in not just our population, but in other populations as well.

In contrast, as seen in Fig. 1 the very well documented association between in situ carcinoma and invasive carcinoma is evident by the arcs between the two ductal in situ/invasive diagnosis pairs – DCIS non-Comedo & IDC and DCIS Comedo & IDC and one lobular in situ/invasive pair – LCIS & ILC. As shown in Table 2, these three arcs are very strong (as measured by the change in network score by their deletion).

Focusing on the ductal carcinoma diagnoses, if either DCIS non-Comedo or DCIS Comedo is present (value =1.0 in Fig 3 and Fig. 1B), then the likelihood of IDC greatly

## Maskery

increases. This scenario is illustrated in the 3 panels of Fig. 3. In Fig. 3A, the network nodes DCIS non-Comedo, DCIS Comedo, and IDC are enlarged to show the percent frequency of these diagnoses in the studied population - 0.0 indicates the percent of the population in which a diagnosis is not present, and 1.0 indicates the percent of the population in which the diagnosis is present. The baseline expected frequencies for these diagnoses in the study population are: DCIS non-Comedo - 16%; DCIS Comedo - 8%; and IDC - 17%. As demonstrated in Fig. 3B, when DCIS non-Comedo is set to present [1.0=100%], the expected incidence of IDC increases to 55% of the population from its baseline value of 17%. Likewise, when DCIS Comedo is set to present [1.0=100%] the expected incidence of IDC increases to 70% of the population (Fig. 3C).



**Figure 3. Effect of Setting Evidence for DCIS on the Expected Frequency of IDC**

In (A) the baseline expected frequencies of DCIS non-Comedo, DCIS Comedo, and IDC are displayed. In (B) evidence is set such that DCIS non-Comedo is present [1.0=100]. The expected frequency of IDC increases 38 percentage points from 17% to 55%. In (C) evidence is set such that DCIS Comedo is present [1.0=100]. The expected frequency of IDC increases 53 percentage points from 17% to 70%.

Expanding outward beyond the two arcs between DCIS non-Comedo & IDC and DCIS Comedo & IDC, and observing the arcs that link DCIS non-Comedo and DCIS Comedo to the rest of the

## Maskery

network (Fig. 1) one can see that MicroCalcs (microcalcifications) directly links to both DCIS non-Comedo and DCIS Comedo. This is expected since microcalcifications are frequently found in DCIS lesions[14 pg. 259]. In fact, mammogram detection of DCIS-associated microcalcifications is a major motivator for many breast biopsies. In contrast to DCIS Comedo, DCIS non-Comedo is also linked to ADH (Atypical Ductal Hyperplasia) which is in turn linked to CCH (Columnar Cell Hyperplasia) and FloridHUT (Florid Intraductal Hyperplasia). It is in this set of lesions that the traditional hyperplasia-atypical hyperplasia-in situ carcinoma-invasive carcinoma pattern of lesions is in evidence. This association pattern, specifically the lack of a co-occurring lesions with Comedo DCIS and co-occurrence of other lesions with non-Comedo DCIS, has been documented previously[15].

## Discussion

Bayesian network analysis provides a stable and verifiable method for the derivation of a topology of breast disease and cancer diagnosis interactions from the dataset analyzed. Validation has shown that the network described here can be re-generated to 80% of its original structure from only 25% of the original data. Additionally, there is evidence for 96% of the connections machine learned from this data set. From these two observations, high confidence is placed in both the stability and accuracy of the network in describing the probabilistic relations between breast disease diagnoses.

It is important to note that the algorithm generating this network is an optimization search algorithm finding a solution to a Non-deterministic Polynomial time hard (NP-hard) problem. Because of the way the network is learned, there is no guarantee that the final result is the global optimum. In fact the stability of the final network is a major concern given the type of algorithm

Maskery

(optimization) that derives the network and problem class (NP-hard). However, the reported network proves very robust to substantial deletion of the original learning dataset; even though no effort is made to make sure the deleted records leave a “balanced” distribution of diagnoses in the remaining pathology reports.

An additional concern in using an analysis technique that is machine learning based is the veracity of the results. However, the Bayesian network analysis algorithm used in this article provides a network for which almost every pathology interaction has been previously documented in the breast pathology literature or is well known to pathologists. This type of verification allows us to trust results that may be counterintuitive to the prevailing wisdom.

It is both a strength and a limitation of this study that a single pathologist writes all the pathology reports. On the one hand, consistency in the diagnosis of breast disease lesions is maintained at a level that would be impossible if done by multiple pathologists. On the other hand, the network of lesion co-occurrences described in this paper may be pathologist dependent and consequently, may change if a different pathologist had made the diagnoses on the 1409 cases used in this study. However, the comparison of the network to published literature does give confidence that the pathology diagnoses from which the network is constructed are in line with general pathology practice. Furthermore, a possible application of this work might be as an aid in standardizing pathology diagnosis. For example, if a pathologist diagnoses columnar cell hyperplasia the probability of all other possible breast lesions could be calculated and displayed as an ordered list of the most likely lesions to co-occur with columnar cell hyperplasia. Each additional lesion diagnosis the pathologist makes will update the network and the list of lesions likely to co-occur with the group of lesions found thus far. Consequently, the pathologist using

## Maskery

this network is alerted to look for lesions with a high likelihood of co-occurring in the specimen being diagnosed and alerted if a lesion diagnosis is unlikely given previous experience.

A strength of this analysis method is that the Bayesian network visualization leads to an intuitive understanding of the analysis results. When shown to non-computationally based researchers (i.e. pathologists, surgeons, cell biologists, etc.) the network is easy to understand. Strong caution must be given that the increased/decreased probabilities of the lesions are for co-occurrence not for causation. However once this caution is understood, the network itself is a fascinating window into the co-occurrence of breast lesions from 1400 plus pathology reports.

The original reason for generating this Bayesian network was to identify breast pathology co-occurrence patterns to aid the design of experiments into the molecular origins of breast cancer. Based upon this network we focus on two different pathology lesion profiles involving DCIS - comedo DCIS and invasive cancer; and non-comedo DCIS hyperplasia, atypical hyperplasia, and invasive cancer. These profiles may lead to the discovery of biomarkers for different grades of breast cancer. A poorly differentiated high grade IDC has a worse prognosis than a low grade well differentiated or medium grade moderately differentiated IDC. Furthermore, DCIS comedo is more likely to be associated with poorly differentiated high grade IDC. In contrast, DCIS not comedo is more likely to be associated with well differentiated low grade IDC or moderately differentiated medium grade IDC (data not shown). Thus, profiling of DCIS comedo lesions that do not co-occur with other pathologies may lead to the discovery of biomarkers for an environment where there is a high likelihood for high grade poorly differentiated IDC development. In contrast, profiling of lesions such as DCIS non-comedo and other lesions seen in its pathology co-occurrence profile with IDC may lead to biomarkers for an environment where low grade well differentiated or moderate grade moderately differentiated

## Maskery

IDC is more likely to occur. Note that we do not state there is a direct causal link between these lesions and the cancers they co-occur with, only that these lesions may be indicative of an environment where these cancers may develop.

Future work will look at how to further refine the pathology patterns and the design of experiments to uncover the biological reasons for why different pathology patterns are observed. One way to refine the observed pathology portraits is to stratify in situ and invasive diagnoses by grade. For example, instead of placing the present/not present profile IDC into the analysis, the present/not present profiles of IDC Grade 1 (well-differentiated), IDC Grade 2 (moderately differentiated), and IDC Grade 3 (poorly differentiated) might instead be analyzed. The same would be done for DCIS (i.e. DCIS well differentiated, DCIS moderately differentiated, and DCIS poorly differentiated). Further stratification of these diagnoses of greatest interest may provide finer detail into the pathology patterns observed. Incorporating epidemiological information, such as menopausal status, may also increase the resolution of the resultant pathology patterns.

In conclusion the Bayesian network of breast disease and cancer pathology interactions described here is robust and verifiable. From this network, intriguing patterns of pathology co-occurrence in our CBCP population are discovered.

## **Acknowledgements**

This work is supported by the Clinical Breast Care Project with funds from the US Department of Defense through the Henry Jackson Foundation for the Advancement of Military Medicine, Rockville, MD. The opinions and assertions contained herein are those of the authors and do not

## Maskery

necessarily reflect the official positions of the Department of the Army, the Department of Defense, or the United States government.

## References

- [1] Ries LAG, Eisner MP, Kosary CL, Hankey BF, Miller BA, Clegg L, et al. SEER Cancer Statistics Review, 1975-2001. Bethesda, MD: National Cancer Institute; 2004.
- [2] Reis-Filho JS, Simpson PT, Gale T, Lakhani SR. The molecular genetics of breast cancer: The contribution of comparative genomic hybridization. *Pathol Res Pract* 2005;201(11):713-725.
- [3] Sorlie T. Molecular portraits of breast cancer: tumour subtypes as distinct disease entities. *European Journal of Cancer* 2004;40(18):2667-2675.
- [4] Dumitrescu RG, Cotarla I. Understanding breast cancer risk -- where do we stand in 2005? *J Cell Mol Med* 2005;9(1):208-21.
- [5] Wellings SR, Jensen HM, Marcum RG. An Atlas of Subgross Pathology of the Human Breast with Special Reference to Possible Precancerous Lesions. *Journal of the National Cancer Institute* 1975;55(2):231-273.
- [6] Dupont WD, Page DL. Risk factors for breast cancer in women with proliferative breast disease. *N Engl J Med* 1985;312(3):146-51.
- [7] Page DL, Dupont WD, Rogers LW, Rados MS. Atypical hyperplastic lesions of the female breast. A long-term follow-up study. *Cancer* 1985;55(11):2698-708.
- [8] Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *PNAS* 2001;98(19):10869-10874.

## Maskery

- [9] Simpson PTP, Gale TB, Reis-Filho JSMD, Jones CP, Parry SM, Sloane JPF, et al. Columnar Cell Lesions of the Breast: The Missing Link in Breast Cancer Progression?: A Morphological and Molecular Analysis. *American Journal of Surgical Pathology* June 2005;29(6):734-746.
- [10] Natkunam Y, Mason DY. Prognostic immunohistologic markers in human tumors: why are so few used in clinical practice? *Lab Invest* 2006;86(8):742-7.
- [11] D'Ambrosio B. Inference in Bayesian Networks. *AI Magazine* 1999;20(2):21-36.
- [12] Moraleda J. New Algorithms, Data Structures, and User Interfaces for Machine Learning of Large Datasets with Applications [PhD]. Palo Alto: Stanford University; 2004.
- [13] Lakhani SR. The transition from hyperplasia to invasive carcinoma of the breast. *J Pathol* 1999;187(3):272-8.
- [14] Rosen P. *Rosen's Breast Pathology*. 2 ed. Philadelphia, PA: Lippincott Williams and Wilkins; 2001.
- [15] Simpson PT, Reis-Filho JS, Gale T, Lakhani SR. Molecular evolution of breast cancer. *J Pathol* 2005;205(2):248-54.
- [16] Schnitt SJ, Vincent-Salomon A. Columnar cell lesions of the breast. *Adv Anat Pathol* 2003;10(3):113-24.
- [17] Dupont WD, Page DL, Parl FF, Vnencak-Jones CL, Plummer WD, Jr., Rados MS, et al. Long-term risk of breast cancer in women with fibroadenoma. *N Engl J Med* 1994;331(1):10-5.
- [18] Valdes EK, Tartter PI, Genelus-Dominique E, Guilbaud DA, Rosenbaum-Smith S, Estabrook A. Significance of Papillary Lesions at Percutaneous Breast Biopsy. *Ann Surg Oncol* 2006;13(4):480-482.

## Maskery

- [19] Renshaw AA, Derhagopian RP, Tizol-Blanco DM, Gould EW. Papillomas and atypical papillomas in breast core needle biopsy specimens: risk of carcinoma in subsequent excision. *Am J Clin Pathol* 2004;122(2):217-21.
- [20] Ali-Fehmi R, Carolin K, Wallis T, Visscher DW. Clinicopathologic analysis of breast lesions associated with multiple papillomas. *Hum Pathol* 2003;34(3):234-9.
- [21] Kennedy M, Masterson AV, Kerin M, Flanagan F. Pathology and clinical relevance of radial scars: a review. *J Clin Pathol* 2003;56(10):721-724.

## Figure Captions

### Figure 1 Bayesian Network derived from Pathology Diagnosis Data

A. Each node is a diagnosis whose complete name is listed in Table 1. Arrows do *not* imply causality, but instead are related to independence statements embedded in the structure of the network. B. It is possible to enlarge each node in the network to show the percent frequency of each diagnosis in the population. This is visualized in the enlarged cutout.

### Figure 2 Network Diagnostics

(A) Histogram of arcs per node in Figure 1. (B) Fraction of the original network remaining ( $\mathfrak{S}$ ) after the random removal of 25%, 50%, and 75% of the records in the original dataset. Error bars are the standard deviation in network scores scaled to  $\mathfrak{S}$ . (C) The strength of each arc in the original dataset derived network is calculated as the change in network score by its deletion. Plotted in C is a histogram of the number of arcs whose score change falls between the ranges on the x-axis.

### Figure 3. Effect of Setting Evidence for DCIS on the Expected Frequency of IDC

## Maskery

In (A) the expected baseline frequencies of DCIS non-Comedo, DCIS Comedo, and IDC are displayed.

In (B) evidence is set such that DCIS non-Comedo is present [1.0=100]. The expected frequency of IDC increases 38 percentage points from 17% to 55%. In (C) evidence is set such that DCIS Comedo is present [1.0=100]. The expected frequency of IDC increases 53 percentage points from 17% to 70%.





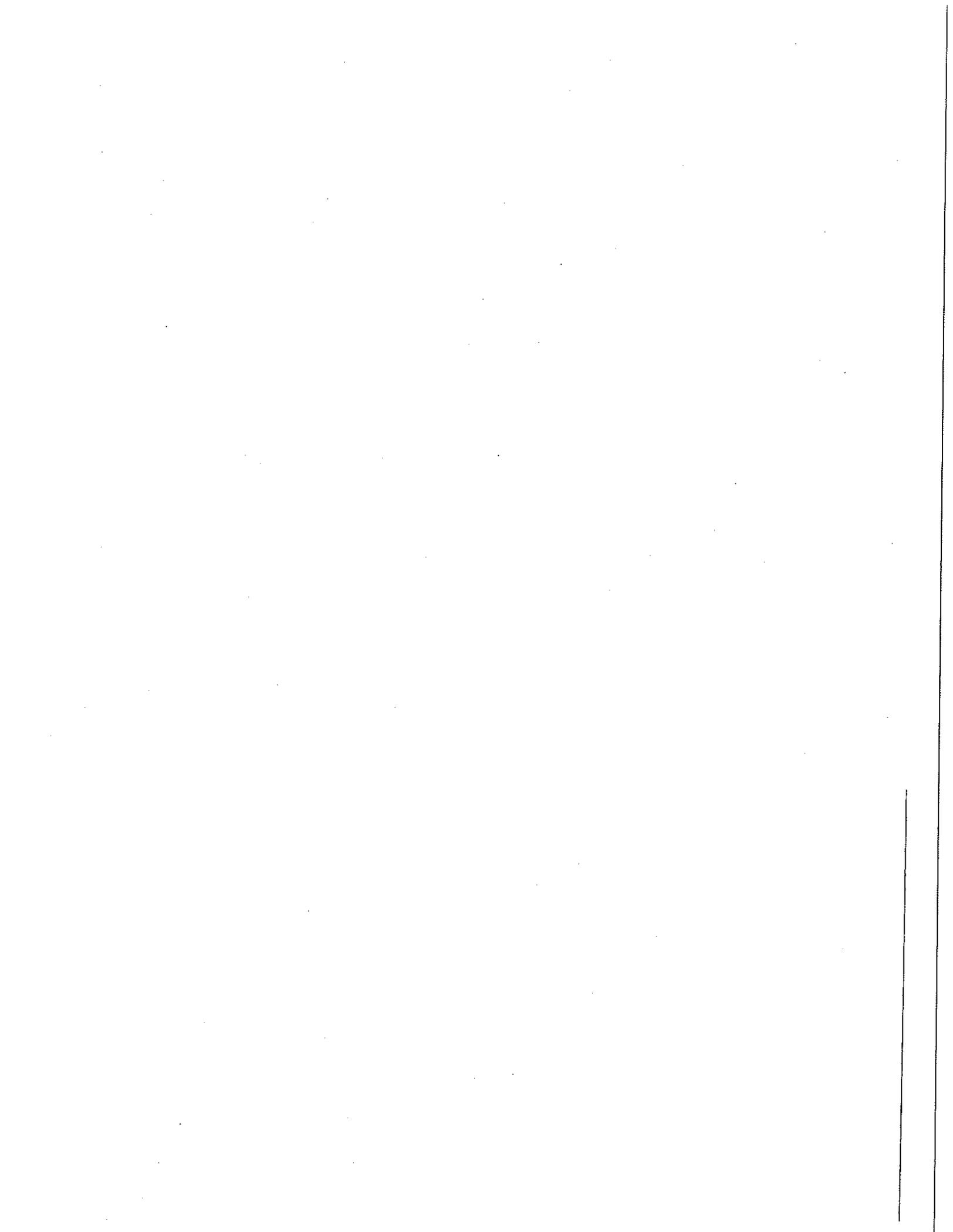


Figure2A

[Click here to download high resolution image](#)

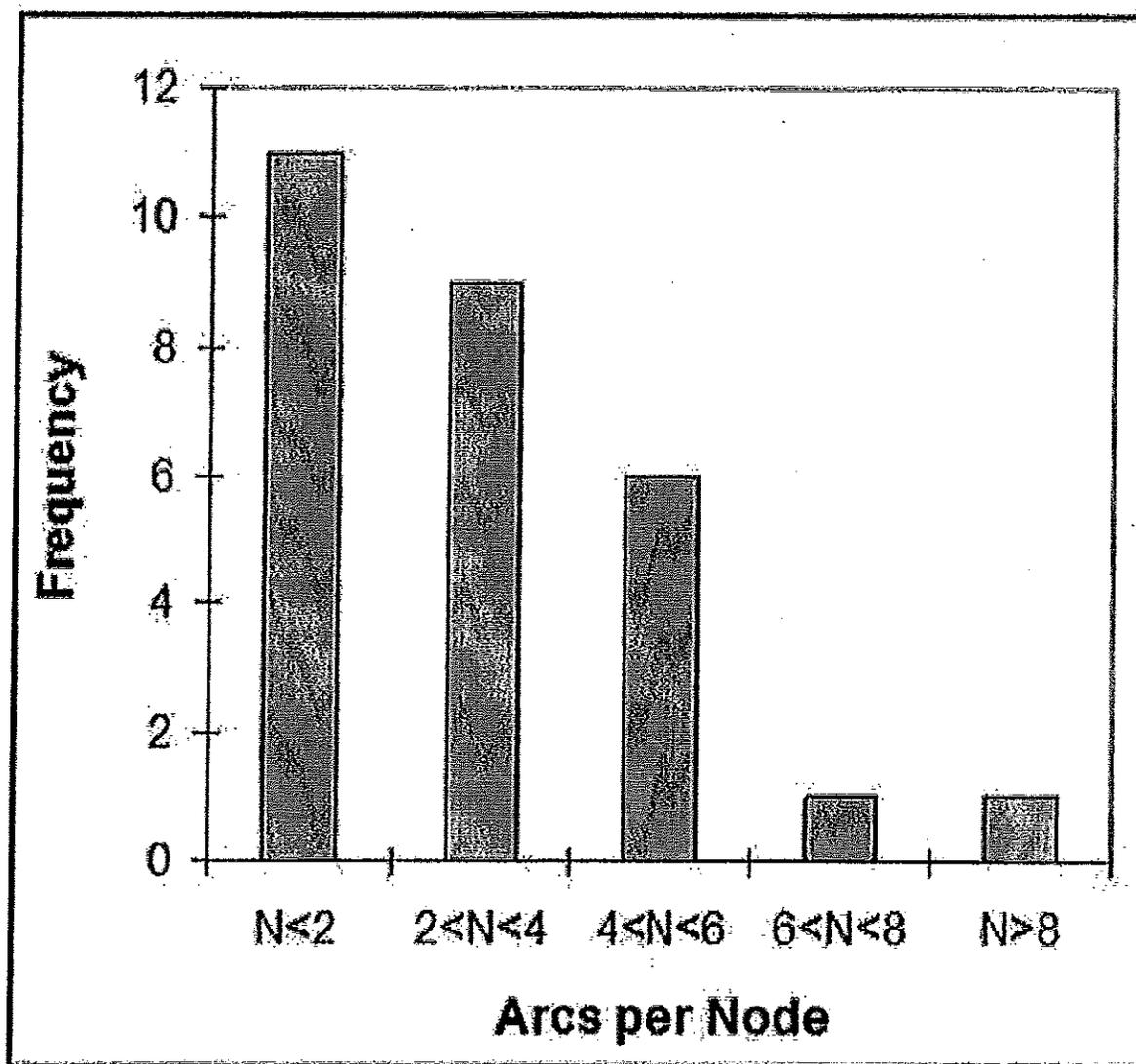


Figure2B

[Click here to download high resolution image](#)

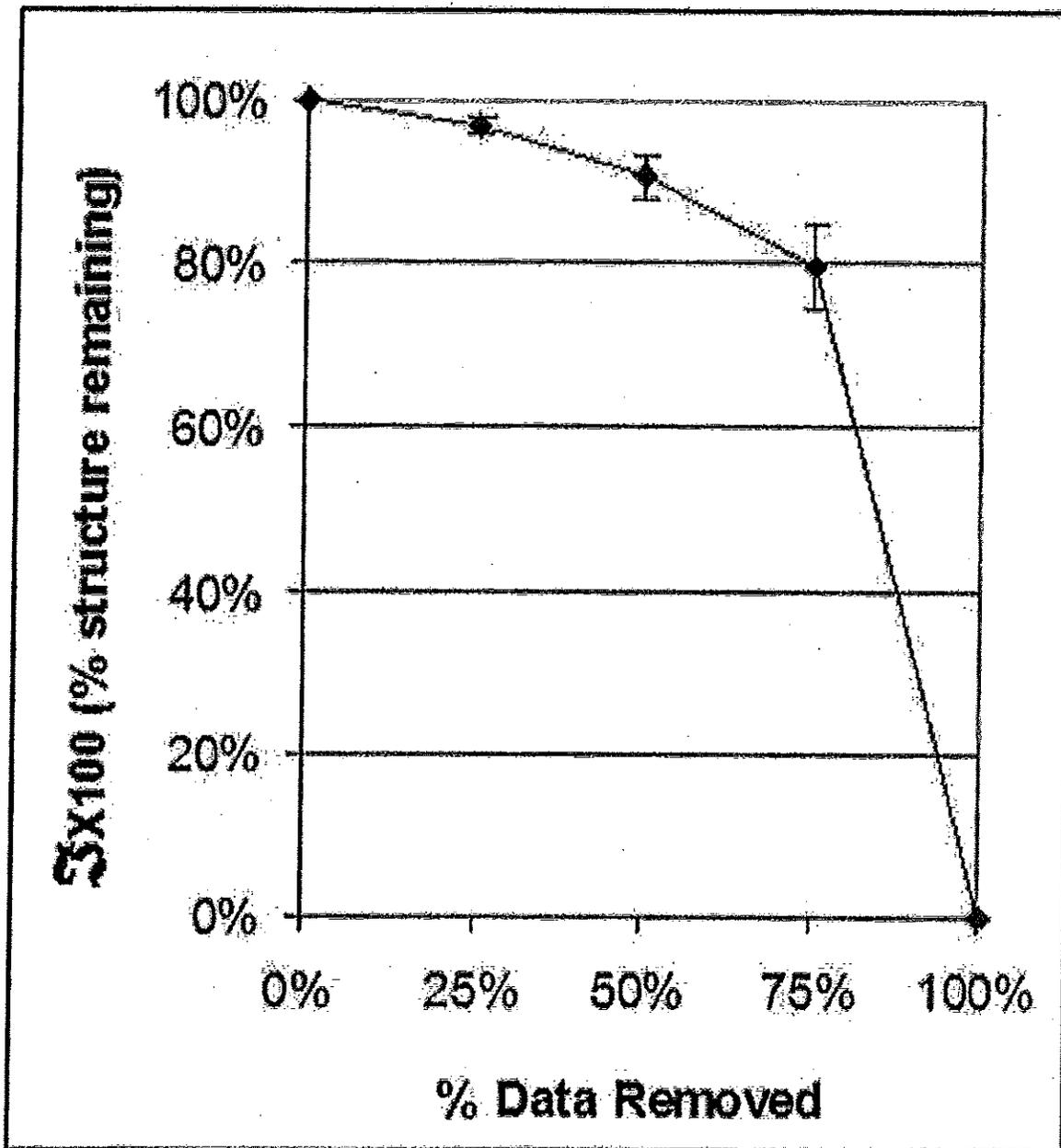
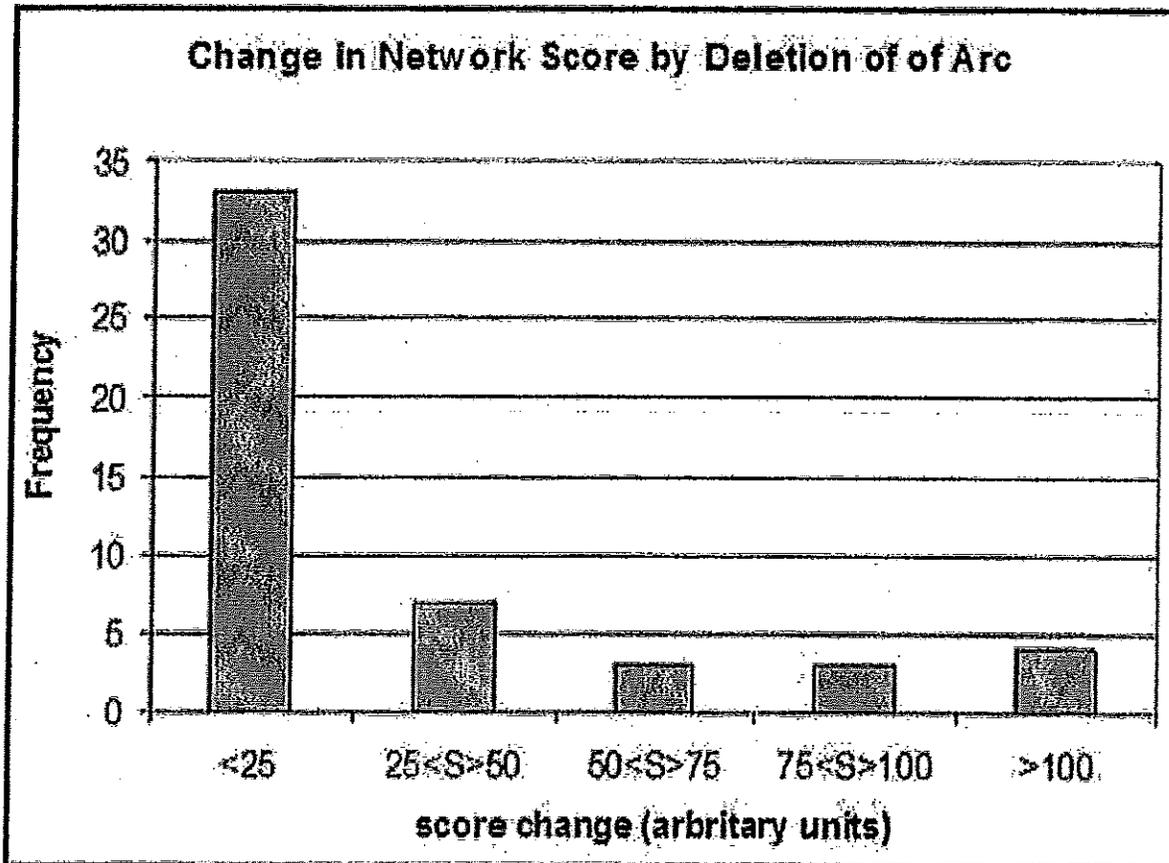


Figure2C

[Click here to download high resolution image](#)



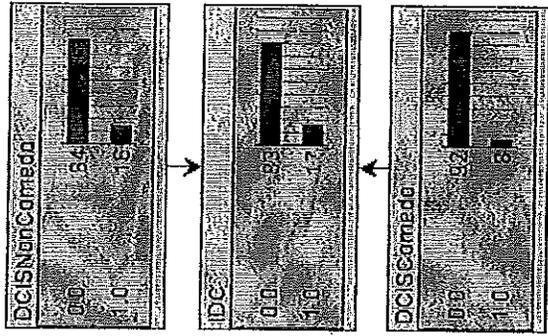


Figure3A

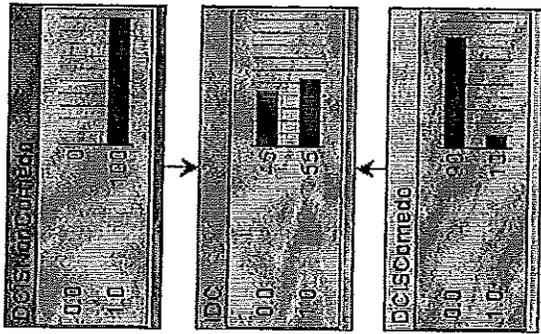


Figure3B

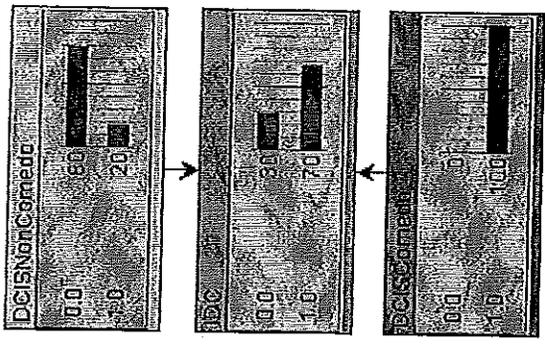


Figure3C

Table 1

Table 1. Abbreviation and Name for all Diagnoses in the Bayesian Analysis Derived Network

Abbreviation	Diagnosis Name	Abbreviation	Diagnosis Name
MicroCalcs	Microcalcifications	ModHUT	Moderate intraductal hyperplasia
FC	Fibrocystic changes	FloridHUT	Florid intraductal hyperplasia
SFibrosis	Stromal fibrosis	CCH	Columnar cell hyperplasia
Cysts	Cysts	ADH	Atypical ductal hyperplasia
ApoMeta	Apocrine metaplasia	AAH	Atypical apocrine hyperplasia
SAdeno	Sclerosing adenosis	Gynecomastia	Gynecomastia
FAdeno	Florid adenosis	ALH	Atypical lobular hyperplasia
AdenoNOS	Adenosis, not otherwise specified	DCISNonComedo	Ductal carcinoma in situ not comedo
RadScar	Radial Scar	DCISComedo	Ductal carcinoma in situ comedo
CollSph	Collagenous spherulosis	LCIS	Lobular carcinoma in situ
MultiPap	Multiple (peripheral) papillomas	IDC	Infiltrating ductal carcinoma
ISPapC	In situ papillary carcinoma	MicroIDC	Microinvasive ductal carcinoma
FibroAd	Fibroadenoma	ILC	Infiltrating lobular carcinoma
FibroAdN	Fibroadenomatoid nodule	TubC	Tubular carcinoma
CCHAtypia	Columnar cell hyperplasia with atypia	ICMixed	Infiltrating carcinoma, mixed ductal and lobular
MildHUT	Mild intraductal hyperplasia	Lipoma	Lipoma

Table2

Table 2 Literature Based Evidence for Network Arcs

Pathology 1	Pathology 2	Arc Strength	Published Evidence
Fibrocystic changes (FC)	Cysts	934	In our pathologist's classification scheme, by definition, cyst is a sub-diagnosis of fibrocystic changes.
Ductal carcinoma in situ not Comedo (DCISNonComedo)	Invasive ductal carcinoma (IDC)	192	Retrospective and cytogenic studies point toward DCIS as a non-obligate precursor to IDC [13].
Ductal carcinoma in situ Comedo (DCISComedo)	Invasive ductal carcinoma (IDC)	176	Retrospective and cytogenic studies point toward DCIS as a non-obligate precursor to IDC [13].
Fibrocystic changes (FC)	Sclerosing adenosis (SAdeno)	174	Sclerosing adenosis (adenosis more generally) commonly occurs as part of fibrocystic changes [14 pg. 139].
Lobular carcinoma in situ (LCIS)	Invasive lobular carcinoma (ILC)	91	Reported frequency of LCIS with ILC varies from 65%-98% cases of ILC [14 pg 635].
Moderate intraductal hyperplasia (ModHUT)	Cysts	84	Frequent coexistence is seen between fibrocystic changes (including cysts) and ductal hyperplasia [14 pg. 203-204].
Cysts	Microcalcifications (MicroCalcs)	78	Microcalcifications are commonly associated with microcysts, especially those with apocrine epithelium [14 pg. 946].
Fibrocystic changes	Stromal fibrosis	73	In our pathologist's classification scheme, by

(FC)	(SFibrosis)		definition, stromal fibrosis is a sub-diagnosis of fibrocystic changes
Sclerosing adenosis (SAdeno)	Microcalcifications (MicroCals)	65	Microcalcifications are frequently formed in sclerosing adenosis [14 pg. 139].
Fibrocystic changes (FC)	Gynecomastia	55	This is an exclusionary finding - gynecomastia (male breast enlargement) excludes all other common female breast pathologies. It is probably linked to FC because this is the most common female breast lesion.
Moderate intraductal hyperplasia (ModHUT)	Florid intraductal hyperplasia (FloridHUT)	43	This is an exclusionary finding based on how intraductal hyperplasias are recorded, in which only the most severe intraductal hyperplasia is recorded. Hence, a diagnosis of moderate intraductal hyperplasia excludes a diagnosis of florid intraductal hyperplasia in the CBCP classification scheme.
Atypical ductal hyperplasia (ADH)	Columnar cell hyperplasia (CCH)	42	Some authorities believe columnar cell lesions are non-obligate pre-cursors to ADH and well-differentiated DCIS [15].
Florid intraductal hyperplasia (FloridHUT)	Apocrine metaplasia (ApoMeta)	38	Frequent coexistence is seen between fibrocystic changes (including apocrine metaplasia) and ductal hyperplasia [14 pg. 203-204].
Columnar cell hyperplasia with	Columnar cell hyperplasia (CCH)	33	CCH and CCH with atypia are part of the continuum of columnar cell lesions, and it is

atypia (CCHAtypia)			not uncommon for more than one columnar cell lesion type to exist in the same breast [16].
Columnar cell hyperplasia (CCH)	Florid intraductal hyperplasia (FlroidHUT)	25	CCH and Hyperplasia of the Usual Type (HUT) have been observed to co-occur[9]. Florid intraductal hyperplasia is a subset of HUT.
Atypical ductal hyperplasia (ADH)	Florid intraductal hyperplasia (FlroidHUT)	25	Although the role of intraductal hyperplasia as a non-obligate precursor to ADH is hypothetical, these two lesions are known to often co-occur [9].
Ductal carcinoma in situ not Comedo (DCISNonComedo)	Atypical ductal hyperplasia (ADH)	25	ADH commonly co-occurs in the same breast that DCIS or invasive breast cancer is diagnosed [13].
Florid intraductal hyperplasia (FlroidHUT)	Multiple papillomas (MultPap)	24	Many benign papillomas have more complex structures caused by hyperplasia of the epithelium [14 pg. 80].
Mild intraductal hyperplasia (MildHUT)	Florid intraductal hyperplasia (FloridHUT)	24	This is an exclusionary finding based on how intraductal hyperplasias are recorded, in which only the most severe intraductal hyperplasia is recorded. Hence, a diagnosis of mild intraductal hyperplasia excludes a diagnosis of florid intraductal hyperplasia in the CBCP classification scheme.
Ductal carcinoma in situ Comedo	Microinvasive ductal carcinoma	24	Microinvasion is most commonly associated with comedo-type DCIS. -- expert opinion

(DCISComedo)	(MicroIDC)		
Mild intraductal hyperplasia (MildHUT)	Cysts	23	Frequent coexistence is seen between fibrocystic changes (including cysts) and ductal hyperplasia [14 pg 203-204].
Moderate intraductal hyperplasia (ModHUT)	Mild intraductal hyperplasia (MildHUT)	23	This is an exclusionary finding based on how intraductal hyperplasias are recorded, in which only the most severe intraductal hyperplasia is recorded. Hence, a diagnosis of mild intraductal hyperplasia excludes a diagnosis of moderate intraductal hyperplasia in the CBCP classification scheme.
Fibrocystic changes (FC)	Lipoma	19	This is an exclusionary finding – lipoma (fat deposit) is found in the subcutaneous fat of the breast, and does not involve the breast duct lobular units (where all other breast diagnoses listed so far are found). It is probably linked to FC because this is the most common female breast lesion.
Sclerosing adenosis (SAdeno)	Multiple papillomas (MultPap)	19	Sclerosing adenosis is sometimes incorporated into a papilloma [14 pg. 80].
Fibrocystic changes (FC)	Apocrine metaplasia (ApoMeta)	18	In our pathologist's classification scheme, by definition, apocrine metaplasia is a sub-diagnosis of fibrocystic changes.
Ductal carcinoma	Microcalcifications	16	There is a well-documented relationship

in situ not Comedo (DCISNonComedo)	(MicroCalcs)		between calcifications and DCIS [14 pg. 259].
Columnar cell hyperplasia with atypia (CCHAtypia)	Atypical ductal hyperplasia (ADH)	15	ADH can arise in CCH with atypia [16].
Ductal carcinoma in situ not Comedo (DCISNonComedo)	Microcalcifications (MicroCalcs)	14	There is a well-documented relationship between calcifications and DCIS [14 pg. 259].
Cysts	Stromal Fibrosis (SFibrosis)	13	Cysts and stromal fibrosis are part of the fibrocystic spectrum of lesions. Given cysts are the most common fibrocystic lesion seen, it is not unexpected to see co-occurrence between cysts and stromal fibrosis, another fibrocystic lesion.
Cysts	Apocrine metaplasia (ApoMeta)	13	"Apocrine metaplasia frequently is present in the epithelial lining of cysts ..." [14 pg. 96].
Columnar cell hyperplasia (CCH)	Lobular carcinoma in situ (LCIS)	13	LCIS is frequently associated with CCH [14 pg. 216].
Invasive ductal carcinoma (IDC)	Fibroadenoma (FibroAd)	13	Fibroadenoma is a negative finding that greatly decreases the likelihood of IDC.
Fibroadenomatoid nodule (FibroAdN)	Microcalcifications (MicroCalcs)	12	Fibroadenomatoid nodules are basically miniature fibroadenomas and can be thought of in that context. Microcalcifications can be seen

			in association with fibroadenomas [17].
Multiple papillomas (MultPap)	Microcalcifications (MicroCalcs)	11	Most lesions with multiple peripheral papillomas are found incidentally in breast biopsies with mammographic abnormalities such as calcifications, nodules, masses, and other various opacities [14 pg. 78].
Columnar cell hyperplasia (CCH)	Microcalcifications (MicroCalcs)	10	Calcifications are frequently formed in CCH lesions [14 pg. 216].
Columnar cell hyperplasia with atypia (CCHAtypia)	Atypical apocrine hyperplasia (AAH)	10	No direct published evidence.
Ductal carcinoma in situ not Comedo (DCISNonComedo)	In situ papillary carcinoma (ISPapC)	10	In situ papillary carcinoma is a special type of DCIS and occurs commonly with non-papillary DCIS [18, 19].
Lobular carcinoma in situ (LCIS)	Microcalcifications (MicroCalcs)	9	Although calcifications are infrequently found in LCIS lesions, they are frequently found in lesions that co-occur with LCIS. However, in cases of florid LCIS calcifications do occur [14 pg. 582].
Columnar cell hyperplasia (CCH)	Moderate intraductal hyperplasia (ModHUT)	9	CCH and HUT have been observed to co-occur [9]. Florid intraductal hyperplasia is a subset of HUT.
Radial scar	Multiple	7	Radial scars with a prominent papillary

(RadScar)	papillomas (MultPap)		component have been recorded [14 pg. 87].
Florid adenosis (FAdeno)	Multiple papillomas (MultPap)	6	Multiple papillomas typically presented as a mass of multiple ducts filled with papillomas intermingled with various fibrocystic lesions including, in particular, florid adenosis [20].
Ductal carcinoma in situ not Comedo (DCISNonComedo)	Infiltrating carcinoma, mixed ductal and lobular (ICMixed)	6	Approximately 2% of invasive ductal carcinomas are mixed ductal and lobular [14 pg. 325]. DCIS occurs commonly with invasive breast carcinomas.
Fibroadenomatoid nodule (FibroAdN)	Stromal fibrosis (SFibrosis)	5	Fibroadenomatoid nodules are basically miniature fibroadenomas and can be thought of in that context. Stromal fibrosis is part of the fibrocystic changes spectrum of diagnoses and the epithelial component of fibroadenomas are known to be prone to various alterations including fibrocystic changes [14 pg.170].
Atypical ductal hyperplasia (ADH)	Atypical lobular hyperplasia (ALH)	5	No direct published evidence.
Fibrocystic changes (FC)	Adenosis, not otherwise specified (AdenoNOS)	3	Adenosis occurs often as part of a spectrum lesions called fibrocystic changes [14 pg. 139].
Apocrine metaplasia (ApoMeta)	Radial scar (RadScar)	3	Apocrine metaplasia is frequently observed in the cystic component of radial scars or as a co-existing fibrocystic lesion [14 pg. 91].

Multiple papillomas (MultPap)	Collagenous spherulosis (CollSph)	1	"Collagenous spherulosis may be found in associations with other benign proliferative lesions including papilloma, papillary duct hyperplasia, atypical ductal hyperplasia, and sclerosing adenosis" [14 pg. 114].
Atypical lobular hyperplasia (ALH)	Sclerosing adenosis (SAdeno)	1	Atypical lobular hyperplasia is typically found incidentally and not associated with another lesion that prompted a biopsy [14 pg. 611]. The differentiation between LCIS and atypical lobular hyperplasia is difficult [14 pg. 611]. LCIS can be found in sclerosing adenosis [14 pg. 601].
Ductal carcinoma in situ not Comedo (DCISNonComedo)	Tubular carcinoma (TubC)	1	DCIS is found in the majority of tubular carcinomas [14 pg. 373].

**\* Reviewer Suggestions**

**Reviewer Suggestions**

I would like to recommend the following reviewers for this paper:

J. Robert Beck, M.D.  
Fox Chase Cancer Center  
[Robert.Beck@fccc.edu](mailto:Robert.Beck@fccc.edu)

Peter Watson, MD  
British Columbia Cancer Research Centre  
[pwatson@bccancer.bc.ca](mailto:pwatson@bccancer.bc.ca)

Mark Sherman, MD  
National Institute of Cancer  
[shermanm@mail.nih.gov](mailto:shermanm@mail.nih.gov)

Sincerely,  
Susan Maskery

**Title:** Traditional Breast Cancer Risk Factors and Common Breast Pathologies in Post-Menopausal Women

S Maskery, PhD<sup>1</sup>, H Hu, PhD<sup>1</sup>, M Liebman, PhD<sup>1</sup>, J Hooke, MD<sup>2</sup>, C D Shriver, MD<sup>2</sup>, and E Taioli<sup>3</sup>, MD PhD, <sup>1</sup>Biomedical Informatics, Windber Research Institute, Windber, PA, 15963 and <sup>2</sup>Clinical Breast Care Project, Walter Reed Army Medical Center, Washington, DC, 20307, Department of Epidemiology<sup>3</sup>, University of Pittsburgh, Pittsburgh, PA, 15260.

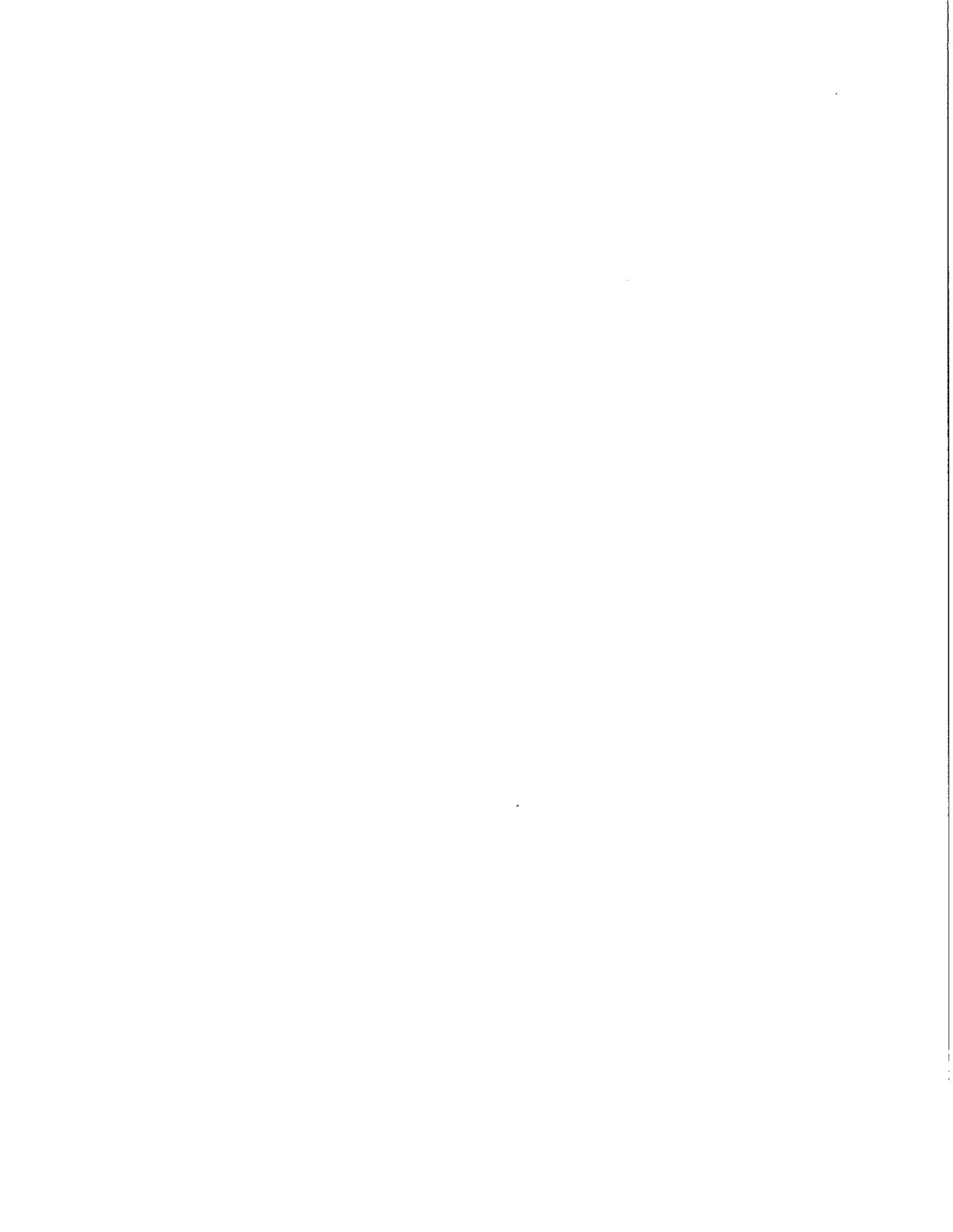
**OBJECTIVE:** Identification of environmental risk factors for non-cancerous breast pathologies may enable the elucidation of the mechanisms by which these factors perturb the normal breast microenvironment. The risk associated with breast cancer risk factors on the occurrence of common breast pathologies was determined in post-menopausal women in a clinical cohort recruited into the Clinical Breast Care Program (CBCP), a joint research project between Walter Reed Army Medical Center (WRAMC) and Windber Research Institute (WRI).

**DESIGN:** In this retrospective study, the contribution of multiple breast cancer risk factors to the occurrence of common breast pathologies was assessed by logistic regression. The inclusion criteria (female, post-menopausal, all diagnoses made by the same pathologist, no missing data) restricted the analysis to 444 records. Odds ratios for stromal fibrosis, apocrine metaplasia, sclerosing adenosis, moderate or florid hyperplasia of the usual type (HUT), and single or multiple papillomas were calculated. These ratios were adjusted for co-occurring breast pathologies (the pathologies listed above, ductal carcinoma in situ, and invasive ductal carcinoma) and breast cancer risk factors (age, hormone replacement therapy (HRT), body mass index (BMI), ethnicity, exercise frequency, alcohol intake frequency, smoking history, early age of menarche, breast feeding, and family history of breast cancer).

**RESULTS:** Age, exercise frequency, early age of menarche, and breast-feeding are not associated with the occurrence of any of the studied breast pathologies in this cohort. HRT use (reference: never used HRT) is significantly associated with stromal fibrosis occurrence (OR=2.5, HRT use > 7yrs), sclerosing adenosis occurrence (OR=2.1, estrogen/progesterone combo), and HUT occurrence (OR=0.52, estrogen/progesterone combo). BMI (reference: normal BMI) is significantly associated with stromal fibrosis occurrence (OR=0.33, obese BMI) and HUT occurrence (OR=1.7, overweight BMI). Ethnicity (reference: Caucasian) is borderline significant (p=0.051) in stromal fibrosis occurrence (OR=0.48, African American). Alcohol intake frequency (reference: < 1/week) is significantly associated with sclerosing adenosis (OR=2.1, 3/week or more) and HUT occurrence (OR=0.46 1-2/week, OR=0.47 3/week or more). Smoking (reference: never smoked) is significantly associated with sclerosing adenosis occurrence (OR=2.0, ever smoked). Family history (reference: no primary relative with breast cancer) is significantly associated with sclerosing adenosis occurrence (OR=1.9, primary relative with breast cancer).

**CONCLUSIONS:** Several documented breast cancer risk factors are also risk factors in the development of common breast pathologies in post-menopausal women in this cohort.

Understanding how breast cancer risk factors affect other breast pathologies may yield new insight into the mechanisms of these risk factors on post-menopausal breast cancer development.



Contribution of chromosomal alterations to the development of poorly-differentiated  
invasive breast carcinomas

Rachel E. Ellsworth, PhD<sup>1</sup>, Jeffrey A. Hooke, MD<sup>2</sup>, Brad Love, PhD<sup>3</sup>, Darrell L.  
Ellsworth, PhD<sup>1</sup>, Craig D. Shriver, MD<sup>2</sup>

<sup>1</sup>Clinical Breast Care Project, Windber Research Institute, Windber, PA; <sup>2</sup>Clinical Breast  
Care Project, Walter Reed Army Medical Center, Washington, DC; <sup>3</sup>Invitrogen  
Informatics, Carlsbad, CA

Contact for correspondence:

Rachel E. Ellsworth  
Windber Research Institute  
620 Seventh Street  
Windber, PA 15963  
Phone: (814) 467-9844  
FAX: (814) 467-6334  
Email: [r.ellsworth@wriwindber.org](mailto:r.ellsworth@wriwindber.org)

Contact for reprints:

Ms. Tiffani Glasco  
Clinical Breast Care Project  
Walter Reed Army Medical Center  
6900 Georgia Avenue, NW  
Washington, DC 20307  
Phone: (202) 782-0002  
[tiffani.glasco@na.amedd.army.mil](mailto:tiffani.glasco@na.amedd.army.mil)

The opinion and assertions contained herein are the private views of the authors and are  
not to be construed as official or as representing the views of the Department of the Army  
or the Department of Defense.

This work was performed under the auspices of the Clinical Breast Care Project with  
funding provided by federal appropriations from the United States Department of  
Defense and the Henry M. Jackson Foundation for the Advancement of Military  
Medicine [Military Molecular Medicine Initiative MDAW81XWH-05-2-0075 to C.D.S.].

## ABSTRACT

Background: Pathological grade is a useful prognostic factor for stratifying breast cancer patients into favorable (well-differentiated tumors) and less favorable (poorly-differentiated tumors) outcome groups. Chromosomal alterations at multiple regions have been useful in discriminating low- from high-grade breast tumors; however, molecular mechanisms by which these changes contribute to the underlying components of tumor grade – tubule formation, nuclear pleomorphism, and mitoses – remain unknown.

Methods: Allelic imbalance (AI) at 26 chromosomal regions was examined from DNA isolated after laser microdissection from 205 invasive breast cancers representing grades 1 (n=63), 2 (n=74), and 3 (n=68). Differences in overall levels and patterns of AI at each chromosomal region were compared for favorable (1) and unfavorable (3) scores within each component used for histological grading (tubule formation, nuclear pleomorphism and mitotic count) and analyzed using Student t-tests and Exact Unconditional tests.

Results: No significant differences in AI levels at any chromosomal region were observed between samples with low versus high scores for tubule formation. In contrast, overall levels of AI were significantly different between samples with high and low scores for both nuclear pleomorphism ( $P < 0.005$ ) and mitotic count ( $P < 0.05$ ). For individual chromosomal regions, 9p21, 13q14, 17p13, and 17q12 showed significantly higher frequencies of AI events in samples with high levels of nuclear atypia, while 16q11-q22 had significantly more AI events in samples with low nuclear atypia. For mitosis, AI was more frequent in samples with high mitotic counts at chromosomes 1p36, 8p21-p22, 11q23, and 13q14.

Conclusions: This study suggests that nuclear pleomorphism and mitotic counts may have a genetic basis. Specific chromosomal regions may harbor genes that influence nuclear morphology and cellular proliferation, and thus may be important in breast cancer progression. In contrast, the lack of association between rates of chromosomal alterations and tubule formation may indicate that tubule formation represents a phenotypic outcome, rather than a causative change, in poorly-differentiated breast tumors. Identification of genes involved in key biological processes that govern carcinogenesis may serve as attractive targets for improved therapeutics.

## BACKGROUND

The most widely used system for histological grading of breast tumors is based largely on the classification parameters developed by Bloom and Richardson (Bloom, Richardson, 1957) in which the cumulative scores from three components, tubule formation, nuclear pleomorphism and mitotic count, are used to define tumors as well- (grade 1), moderately- (grade 2) or poorly- (grade 3) differentiated. This system of grading has clinical utility in determine patient risk and outcome: patients with low-grade carcinomas had a 95% 5-year survival compared to just 50% of patients with high-grade disease (Elston, Ellis, 1991; Henson et al., 1991). Assignment of nuclear grade is, however, highly subjective (Robbins et al., 1995), limiting the ability to utilize grade as a reliable prognostic tool.

Breast cancer is a heterogeneous disease, involving multiple molecular pathways. Extensive gene expression analyses have identified patterns of expression that classify breast tumors into five major subtypes (Perou et al., 2000; Sorlie et al., 2001; Sorlie et al., 2003) as well as to predict recurrence and outcome (Paik et al., 2004; Ramaswamy et al., 2003; van't Veer et al., 2002). Recently, gene expression profiles have been identified that can successfully discriminate low- from high-grade invasive breast carcinomas (Sotiriou et al., 2006). In addition to these gene expression differences, genomic alterations at chromosomes 1p, 1q, 6q, 9p, 11q, 13q, 16q, 17p and 17q have been associated with either low- or high-grade breast tumors (Buerger et al., 2001; Ellsworth et al., 2007b; Roylance et al., 1999; Roylance et al., 2002), suggesting that invasive breast cancer represents two distinct diseases that develop along either high- or low-grade genetic pathways.

Although the development of molecular signatures associated with disease grade can be used to discriminate high- from low-grade carcinomas, how genetic changes at these critical regions contribute to disease pathogenesis remains unknown. To improve our understanding of how genetic alterations contribute to tumor grade, we used a panel of 52 microsatellite markers representing 26 chromosomal regions commonly altered in breast cancer to identify levels and patterns of allelic imbalance (AI) in 205 invasive breast carcinomas. Our objectives were to examine the individual contributions of tubule formation, nuclear pleomorphism and mitoses to overall grade and to identify chromosomal changes associated with each of these pathological components.

## **METHODS**

Paraffin-embedded primary breast tumors from 205 patients were obtained from the Windber Medical Center Pathology Department or the Clinical Breast Care Project (CBCP) Pathology Laboratory. Samples from the Windber Medical Center (n=49) were archival in nature and anonymized with no links between the assigned research number and patient identifiers. Clinical information was provided anonymously by the Memorial Medical Center Cancer Registry. Tissue and blood samples from CBCP patients (n=156) were collected with approval from the Walter Reed Army Medical Center (WRAMC) Human Use Committee and Institutional Review Board. All subjects enrolled in the CBCP voluntarily agreed to participate and gave written informed consent. Clinical information was collected for all CBCP samples using questionnaires designed by and administered under the auspices of the CBCP. To ensure consistency, diagnosis of all tumor samples were made by one pathologist from hemotoxylin and eosin (H&E) stained

slides; grade was assigned using the Bloom-Richardson system of classification (Bloom, Richardson, 1957; Elston, Ellis, 1991). Pathological diagnosis partitioned the samples into grade 1 (n=63), grade 2 (n=74), and grade 3 (n=68). Clinicopathological information for all samples is summarized in Table 1.

DNA was obtained from homogeneous populations of primary breast tumor cells following laser-assisted microdissection on an *ASLMD* laser microdissection system (Leica Microsystems, Wetzlar, Germany) as previously described (Ellsworth et al., 2003a). All microdissected sections were examined by the CBCP pathologist, who identified and marked regions of tumor before dissection. The integrity of multiple serial sections was established by pathological verification of the first and last sections stained with H&E. To avoid PCR artifacts,  $\geq 5,000$  cells were captured from each of six consecutive breast tumor sections, with the sixth section reserved for all confirmatory reruns. Referent DNA samples for the archived samples were extracted from disease-free skin or negative lymph node tissue from each patient using the QIAamp DNA Mini Kit (Qiagen, Valencia, CA). Referent DNA for the CBCP samples was obtained from blood clots using Clotspin and Puregene DNA purification kits (Gentra, Minneapolis, MN).

Microsatellite markers were amplified as previously described (Ellsworth et al., 2003b), purified using Sephadex G-50 resin and genotyped on a MegaBACE-1000 capillary electrophoresis apparatus (Amersham Biosciences, Piscataway, NJ) following standard protocols. Genotypes were determined using Genetic Profiler version 2.0 software. AI was detected using methods first described by Medintz et al. (Medintz et al., 2000) and defined by a threshold value of 0.35 which has been shown to have >80%

reproducibility when AI events were confirmed on a second aliquot of DNA (Ellsworth et al., 2005).

To increase our ability to detect AI, two microsatellite markers from each chromosomal region were assayed. Marker information was then pooled and the following criteria were used to define AI for each region: when at least one marker for a given region showed an allelic ratio  $\leq 0.35$ , the region was considered to show AI; when neither marker had an allelic ratio  $\leq 0.35$  and at least one marker was informative, the region was considered normal; and when both markers were homozygous, the region was considered uninformative.

Comparison of the clinicopathological factors and levels and patterns of AI by grade were performed using Student's t-tests and Fisher's exact tests. A significance value of  $P < 0.05$  was used for all analyses.

## RESULTS

### *Dissection of grade components*

The degree of nuclear pleomorphism was broadly distributed across samples from 24% with small, uniform cells (score = 1), to 36% of intermediate size, with moderate nuclear variation (score = 2), to 40% with large cells with marked variation (score = 3). In contrast, for tubule formation and mitotic count, reduced (<10%) tubule formation (score = 3) and low ( $\leq 5$  per 10 fields) mitotic count (score = 1) were the predominant scorings, observed in 80% and 62% of samples, respectively. When stratified by overall grade, reduced tubule formation, high (>10 per 10 fields) mitotic counts, and marked nuclear variation occurred significantly more frequently ( $P < 0.05$ ) in grade 3 compared to grade 1 tumors (Table 2).

*Association of genomic changes with tubule formation, nuclear pleomorphism and mitotic count*

Overall levels of genomic instability did not differ significantly between samples with high compared to low tubule formation. Likewise, none of the 26 chromosomal regions differed significantly in AI levels between samples with high and low tubule formation.

Overall levels of genomic instability were significantly higher ( $P < 0.01$ ) in tumors with marked nuclear variation (25.3%) compared to those without (17.8%). AI events were significantly more frequent in tumors with marked nuclear variation at chromosomes 9p21, 13q14, 17p13.1, 17p13.3 and 17q12-q21. In addition, AI at chromosome 16q11-q22 occurred at a significantly higher frequency in tumors without nuclear pleomorphism compared to those with marked variation.

Similar to nuclear pleomorphism, overall levels of AI were significantly higher ( $P < 0.05$ ) in tumors with high mitotic counts (26.8%) compared to those without (21.5%). Three chromosomal regions previously identified as discriminating high- from low-grade invasive tumors – 1p36, 11q23 and 13q14 – showed a significantly higher frequency of AI in tumors with high mitotic counts. AI at chromosome 8p21-p22 occurred at a significantly higher frequency in tumors with high mitotic counts.

*Evaluation of 2- versus 3- tiered grading system*

Tumor grade was recalculated using only two of the three components and compared to the original grade assignment. When grade was recalculated omitting scores for tubule formation, 97% of high- (66/68) and low-grade (61/63) samples retained their

original grade assignment. Two high-grade tumors were reclassified as intermediate grade, but had high-grade AI profiles. One low-grade discordant case had the low-grade genetic profile while the other sample demonstrated AI at three high-grade regions. Elimination of nuclear pleomorphism resulted in 43% of low-grade tumors being reclassified as intermediate-grade, ~40% of which had AI at chromosome 16q11-q22, a hallmark of low-grade disease. Elimination of mitotic count resulted in 57% of low-grade tumors being reclassified, including one specimen reclassified as high-grade. High-grade disease was not affected by omitting either nuclear pleomorphism or mitotic scores.

## DISCUSSION

Histological grading of primary breast carcinomas has become widely accepted as an important predictor of outcome (Fitzgibbons et al., 2000). Recent molecular studies have identified specific chromosomal alterations and gene expression profiles that correlate with tumor grade, and an array of candidate genes that may contribute to tumor behavior and phenotypic characteristics have been suggested (Ellsworth et al., 2007b; Sotiriou et al., 2006). Despite the importance of molecular changes in shaping tumor growth and differentiation in patients with breast cancer, genetic contributions to physiological and structural attributes that define histological grade have not been well-defined.

Previous research from our laboratory identified molecular signatures that were associated with low- (loss of chromosome 16q11-q22) and high-grade (AI at chromosome 1p36, 9p21, 11q23, 13q14, 17p13.1 and 17q12-q21) invasive breast tumors

(Ellsworth et al., 2007b). These molecular signatures persisted when components defining tumor grade were examined independently. AI at chromosomal regions 1p36, 11q23 and 13q14 was associated with high mitotic count; AI at 9p21, 13q14, 17p13.1 and 17q12-q21 was correlated with marked nuclear pleomorphism; AI at 16q11-q22 was associated with lack of nuclear atypia. Therefore, overall histologic grade may be determined by genetic changes that influence nuclear structure and rates of cellular proliferation.

Genomic alterations at chromosomes 1p36, 8p21-p22, 11q23 and 13q14 were associated with high mitotic count, a measure of proliferation. Based on current knowledge of functional attributes, several candidate genes in these regions may influence mitotic count and thus contribute to histologic grade. The chromodomain helicase DNA binding protein 5 (CHD5), located at chromosome 1p36.31, serves as a tumor suppressor by controlling cellular proliferation (Bagchi et al., 2007). On chromosome 8p22, the deleted in liver cancer 1 (DLC1) tumor suppressor gene has been shown to regulate cell growth and proliferation (Yuan et al., 2003). A recently identified gene SKCG-1 on chromosome 11q23 has been shown to have reduced expression in breast tumors; silencing of SKCG-1 results in increased cellular proliferation (Singh, Roy, 2006). Finally, genomic instability near the tumor suppressor gene RB1 on chromosome 13q14 has been associated with high-grade invasive and *in situ* carcinomas (Ellsworth et al., 2007a; Ellsworth et al., 2007b) while underexpression of RB1 has been correlated with increased cellular proliferation (Bieche, Lidreau, 2000). Thus, changes at each of these chromosomal regions may contribute to the increased mitotic counts associated with high-grade breast tumors.

Nuclear pleomorphism measures changes in cell size and uniformity. Previous studies have shown that AI at chromosome 16q11-q22 is associated with the development of low-grade invasive breast cancer (Buerger et al., 2001; Cleton-Jansen et al., 2004; Ellsworth et al., 2007b; Roylance et al., 2002). Because AI at 16q11-q22 is specifically associated with preservation of nuclear structure, alterations of this region may confer a protective advantage against nuclear damage. Although several genes on chromosome 16q have been investigated for a possible role in low-grade disease, (Cleton-Jansen et al., 2004; van Wezel et al., 2005), specific genetic changes have not been identified.

In contrast, a number of candidate genes have been identified that may contribute to the development of nuclear pleomorphism. Structural changes at chromosome 9p21, including the frequently co-altered genes CDKN2A, CDKN2B and MTAP, have been associated with poor prognosis in high grade breast tumors (Blegen et al., 2003). Similarly, the loss of the 17p13.1 region is commonly observed in large, poorly-differentiated tumors (Seitz et al., 2001; Winqvist et al., 1995), and, in particular, loss of the p53 gene strongly correlates with marked nuclear atypia (Tsuda, Hirohashi, 1994). The 17q12-q21 region, which includes both the HER2 and BRCA1 genes is frequently altered in tumors with nuclear pleomorphism (Hanby et al., 2007; Yang et al., 2001). Thus, changes at several chromosomal regions may contribute to nuclear instability and the resulting poor outcomes often associated with high-grade tumors.

In contrast to nuclear pleomorphism and mitotic count, specimens with high versus low levels of tubule formation were not genetically discernible. Omission of tubule formation scores in determining tumor grade did not significantly change the histological grade of high- and low-grade tumors. These data suggest that tubule

formation may reflect a resulting phenotype, rather than serve as a causative factor, in the process of tumor differentiation. At the clinical level, nuclear pleomorphism was superior to tubule formation and mitotic counts in predicting survival which may suggest the traditional tripartite histologic scoring system should be reconsidered (Rank et al., 1987). A two-tiered scoring system involving nuclear pleomorphism and mitotic counts, could discriminate high- from low-risk, node-negative breast cancer patients (Tsuda et al., 1998). Finally, Komaki et al. compared a two- and three-tiered scoring system for predicting disease recurrence and found that a two-tiered system comprised of mitotic counts and nuclear pleomorphism was significantly more accurate (Komaki et al., 2006).

In conclusion, patterns of AI that discriminate high- from low-grade invasive breast tumors are associated with levels of mitoses and nuclear pleomorphism. AI events at chromosomal regions associated with increased mitotic counts harbor candidate genes involved in increased cellular proliferation, thus alterations at these regions may promote cell cycle deregulation. Similarly, a number of genes located in genomic regions that are associated with high levels of nuclear variation may be critical in the development of high- compared to low-grade breast tumors. Molecular profiles could not discriminate samples with high- from low-levels of tubule formation, suggesting that tubule formation is a result of, not a cause, of tumor differentiation. Genetic data provide a molecular rationale for clinical models that suggest a two-tiered system of classification, based on nuclear pleomorphism and mitoses, may better reflect the underlying molecular differences between high- and low-grade breast carcinomas, thus increasing the ability to use tumor grading as a prognostic tool.

## **ACKNOWLEDGEMENTS**

The authors thank Jennifer Kane for technical assistance.

#### References

1. Bagchi A, Papazoglu C, Wu Y et al. CHD5 is a tumor suppressor at human 1p36. *Cell* 2007;128:459-75.
2. Bieche I, Lidreau R. Loss of heterozygosity at 13q14 correlates with RB1 gene underexpression in human breast cancer. *Mol Carcinog* 2000;29:151-8.
3. Blegen H, Will JS, Ghadimi BM et al. DNA amplifications and aneuploidy, high proliferative activity and impaired cell cycle control characterize breast carcinomas with poor prognosis. *Anal Cell Pathol* 2003;25:103-14.
4. Bloom HJ, Richardson WW. Histological grading and prognosis in breast cancer. *Br J Cancer* 1957;11:359-77.
5. Buerger H, Mommers EC, Littmann R et al. Ductal invasive G2 and G3 carcinomas of the breast are the end stages of at least two different lines of genetic evolution. *J Pathol* 2001;194:165-70.
6. Cleton-Jansen AM, Buerger H, ter Haar NT et al. Different mechanisms of chromosome 16 loss of heterozygosity in well- versus poorly differentiated ductal breast cancer. *Genes Chromosomes Cancer* 2004;41:109-16.
7. Ellsworth DL, Shriver CD, Ellsworth RE, Deyarmin B, Somiari RI. Laser capture microdissection of paraffin-embedded tissues. *Biotechniques* 2003a;34:42-6.
8. Ellsworth RE, Ellsworth DL, Love B et al. Correlation of levels and patterns of genomic instability with histological grading of DCIS. *Anns Surg Oncol* 2007a;e pub.
9. Ellsworth RE, Ellsworth DL, Lubert SM, Hooke J, Somiari RI, Shriver CD. High-throughput loss of heterozygosity mapping in 26 commonly deleted regions in breast cancer. *Cancer Epidemiol Biomarkers Prev* 2003b;12:915-9.

10. Ellsworth RE, Ellsworth DL, Neatrou DM et al. Allelic imbalance in primary breast carcinomas and metastatic tumors of the axillary lymph nodes. *Mol Cancer Res* 2005;3:71-7.
11. Ellsworth RE, Hooke JA, Love B et al. Correlation of levels and patterns of genomic instability with histological grading of invasive breast tumors. *Breast Cancer Res Treat* 2007b; Epub ahead of print.
12. Elston CW, Ellis IO. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow up. *Histopathology* 1991;19:403-10.
13. Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. *Cell* 1990;61:759-67.
14. Fitzgibbons PL, Page DL, Weaver D et al. Prognostic factors in breast cancer. College of American Pathologists Consensus Statement 1999. *Arch Pathol Lab Med* 2000;124:966-78.
15. Hanby AM, Kelsell DP, Potts HW et al. Association between loss of heterozygosity of BRCA1 and BRCA2 and morphological attributes of sporadic breast cancer. *Int J Cancer* 2007;88:204-8.
16. Henson DE, Riès L, Freedman LS, Carriaga M. Relationship among outcome, stage of disease, and histologic grade for 22,616 cases of breast cancer. *Cancer* 1991;68:2142-9.
17. Komaki K, Sano N, Tangoku A. Problems in histological grading of malignancy and its clinical significance in patients with operable breast cancer. *Breast Cancer* 2006;13:249-53.
18. Medintz IL, Lee C-CR, Wong WW, Pirkola K, Sidransky D, Mathies RA. Loss of heterozygosity assay for molecular detection of cancer using energy-transfer primers and capillary array electrophoresis. *Genome Res* 2000;10:1211-8.
19. Paik S, Shak S, Tang G et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med* 2004;351:2817-26.
20. Perou CM, Sorlie T, Eisen MB et al. Molecular portraits of human breast tumours. *Nature* 2000;406:747-52.

21. Ramaswamy S, Ross KN, Lander ES, Golub TR. A molecular signature of metastasis in primary solid tumors. *Nat Genet* 2003;33:49-54.
22. Rank F, Dombernowsky P, Jespersen NC, Pedersen BV, Keiding N. Histologic malignancy grading of invasive ductal breast carcinoma. A regression analysis of prognostic factors in low-risk carcinomas from a multicenter trial. *Cancer* 1987;60:1299-305.
23. Robbins P, Pinder S, de Klerk N et al. Histological grading of breast carcinomas: a study of interobserver agreement. *Hum Pathol* 1995;26:873-9.
24. Roylance R, Gorman P, Hanby A, Tomlinson I. Allelic imbalance analysis of chromosome 16q shows that grade I and grade III invasive ductal breast cancers follow different genetic pathways. *J Pathol* 2002;196:32-6.
25. Roylance R, Gorman P, Harris W et al. Comparative genomic hybridization of breast tumors stratified by histological grade reveals new insights into the biological progression of breast cancer. *Cancer Res* 1999;59:1433-6.
26. Seitz S, Poppe K, Fischer J et al. Detailed deletion mapping in sporadic breast cancer at chromosomal region 17p13 distal to the TP53 gene: association with clinicopathological parameters. *J. Pathol* 2001;194:318-26.
27. Singh KP, Roy D. SKCG-1: a new candidate growth regulatory gene at chromosome 11q23.2 in human sporadic Wilms tumours. *Br J Cancer* 2006;94:1524-32.
28. Sorlie T, Perou CM, Tibshirania R et al. Gene Expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl Acad Sci USA* 2001;98:10869-74.
29. Sorlie T, Tibshirania R, Parker J et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci USA* 2003;100:8418-23.

30. Sotiriou C, Wirapati P, Loi S et al. Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. *J Natl Cancer Inst* 2006;98:262-72.
31. Tsuda H, Akiyama F, Kurosumi M, Sakamoto G, Watanabe T. Establishment of histological criteria for high-risk node-negative breast carcinoma for a multi-institutional randomized clinical trial of adjuvant therapy. Japan National Surgical Adjuvant Study of Breast Cancer (NSAS-BC) Pathology Section. *Jpn J Clin Oncol* 1998;28:486-91.
32. Tsuda H, Hirohashi S. Association among p53 gene mutation, nuclear accumulation of the p53 protein and aggressive phenotypes in breast cancer. *Int J Cancer* 1994;57:498-503.
33. van't Veer LJ, Dai H, van de Vijver MJ et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002;415:530-6.
34. van Wezel T, Lombaerts M, van Roon EH et al. Expression analysis of candidate breast tumor suppressor genes on chromosome 16q. *Breast Cancer Res* 2005;7:R998-R1004.
35. Winqvist R, Hampton GM, Mannermaa A et al. Loss of heterozygosity for chromosome 11 in primary human breast tumors is associated with poor survival after metastasis. *Cancer Res* 1995;55:2660-4.
36. Yang Q, Mori I, Sakurai T et al. Correlation between nuclear grade and biological prognostic variables in invasive breast cancer. *Breast Cancer* 2001;8:105-10.
37. Yuan BZ, Zhou X, Durkin ME et al. DLC-1 gene inhibits human breast cancer cell growth and in vivo tumorigenicity. *Oncogene* 2003;22:445-50.

Table 1. Clinical and pathological features of 205 invasive breast tumor specimens

	Well-differentiated (n=63)	Moderately- differentiated (n=74)	Poorly-differentiated (n=68)	P Well- vs. Poorly- differentiated
<b>Menopausal Status</b>				
Pre (<50 years)	20%	28%	42%	$P < 0.05$
Menopausal (≥50 years)	80%	72%	58%	
<b>Histology</b>				
IDCA	56%	77%	93%	$P < 0.001^b$
ILCA	27%	15%	3%	
Mixed	8%	8%	0%	
Other <sup>a</sup>	9%	0%	4%	
<b>TNM Stage</b>				
I	52%	43%	28%	
II	33%	32%	40%	
III	13%	20%	22%	
IV	2%	5%	10%	
<b>Lymph Node Status</b>				
Negative	52%	53%	41%	
Positive	48%	47%	59%	
<b>Hormone Receptor Status</b>				
ER+/PR+	78%	69%	31%	
ER+/PR-	17%	14%	14%	
ER-/PR+	0%	6%	10%	
ER-/PR-	5%	11%	45%	$P < 0.001^c$
<b>HER2 Status</b>				
Positive	6%	23%	40%	$P < 0.001$
Negative	94%	77%	60%	

<sup>a</sup>Other histological types include tubular, medullary, apocrine, and mucinous carcinomas

<sup>b</sup>Comparison of frequencies in IDCA

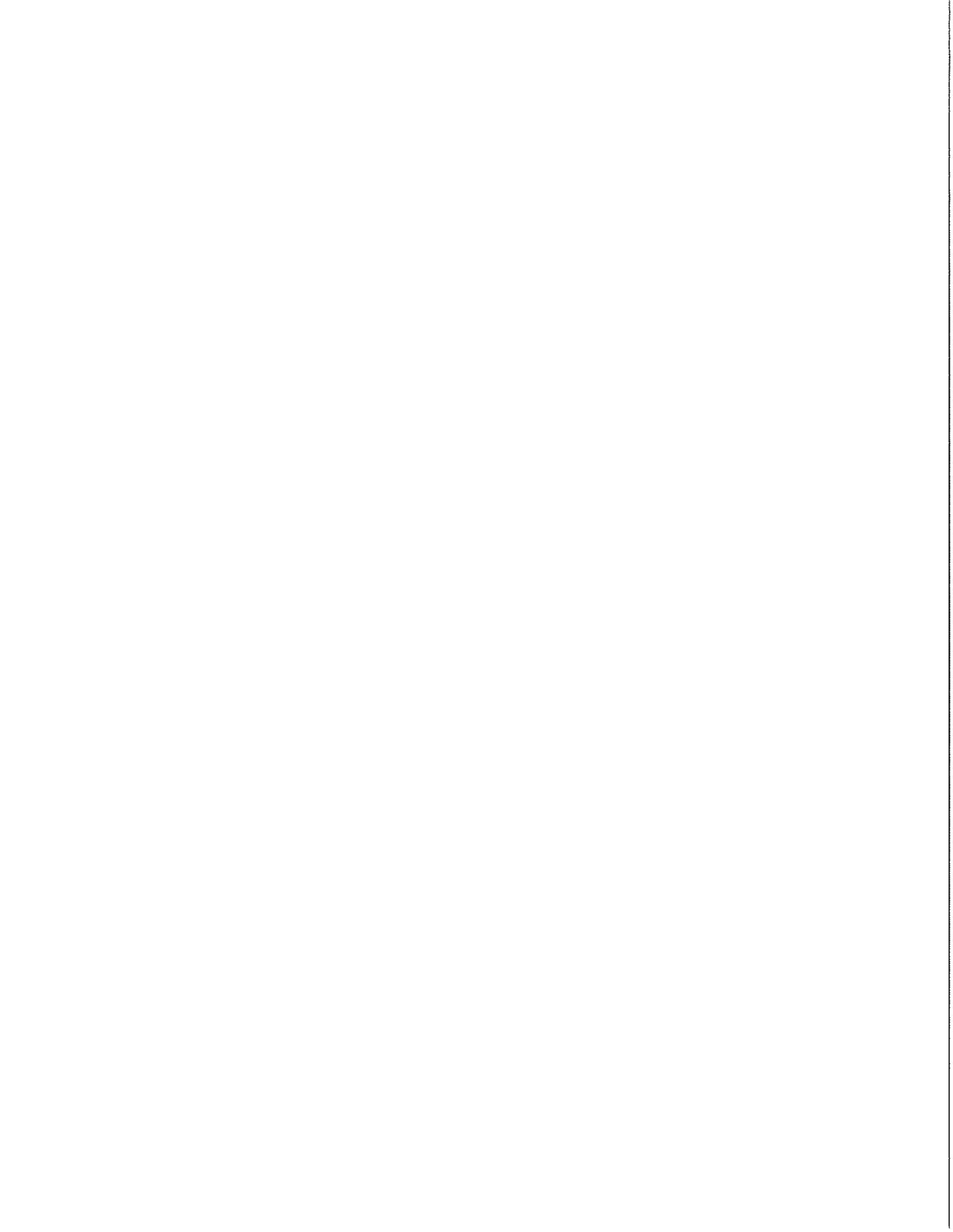
<sup>c</sup>Comparison of frequency of hormone receptor negative (ER-/PR-) compared to ER+ and/or PR+

Table 2. Tumor characteristics by histologic grade

	Well- differentiated (n=63)	Moderately- differentiated (n=74)	Poorly- differentiated (n=68)
<i>Tubule Formation</i>			
1	25%	0%	0%
2	32%	4%	1%
3	43%	96%	99%
<i>Nuclear Pleomorphism</i>			
1	77%	1%	0%
2	22%	76%	4%
3	1%	23%	96%
<i>Mitotic count</i>			
1	98%	85%	1%
2	2%	12%	34%
3	0%	3%	65%

Table 3. Frequency of AI stratified by grade component at 26 chromosomal regions. Numbers in bold are statistically significant

Chromosomal Region	Tubule Formation			Nuclear Pleomorphism			Mitosis					
	1+	2+	3+	P1 vs.3	1+	2+	3+	P1 vs.3	1+	2+	3+	P1 vs.3
1p36.1-p36.2	0	0.22	0.19	0.1337	0.17	0.16	0.23	0.4833	0.12	0.14	0.33	<b>0.0023</b>
2q21.3-23.3	0.2	0.18	0.17	0.7229	0.21	0.15	0.15	0.4623	0.18	0.12	0.17	1.0000
3p14.1	0.19	0.14	0.18	1.0000	0.11	0.23	0.17	0.5963	0.17	0.16	0.21	0.6444
5q21.1-q21.3	0.19	0.24	0.16	0.7314	0.18	0.13	0.21	1.0000	0.16	0.19	0.21	0.4739
6q15	0.11	0.2	0.24	0.3711	0.13	0.31	0.21	0.3373	0.25	0.15	0.21	0.6793
6q22.1-q23.1	0.13	0.04	0.22	0.5294	0.11	0.23	0.22	0.1474	0.21	0.17	0.19	0.8289
6q25.2-q27	0.21	0.22	0.28	0.6246	0.15	0.29	0.29	0.0873	0.27	0.14	0.25	0.8554
7q31.1-q31.31	0	0.1	0.11	0.3652	0.05	0.06	0.16	0.1342	0.07	0.1	0.17	0.0716
8p22-p21.3	0	0.1	0.24	0.0700	0.14	0.19	0.28	0.0737	0.18	0.11	0.38	<b>0.0109</b>
8q24	0.21	0.14	0.17	0.7135	0.12	0.15	0.21	0.2425	0.15	0.19	0.27	0.1079
9p21	0	0.17	0.17	0.1362	0.02	0.16	0.22	<b>0.0028</b>	0.14	0.13	0.26	0.0907
10q23.31-q23.33	0.23	0.24	0.15	0.4292	0.2	0.13	0.19	1.0000	0.15	0.14	0.24	0.2306
11p15	0.31	0.26	0.18	0.2009	0.2	0.18	0.19	1.0000	0.19	0.17	0.24	0.5171
11q13.1	0.25	0.18	0.24	1.0000	0.31	0.2	0.2	0.2003	0.25	0.23	0.18	0.4117
11q23	0.13	0.3	0.32	0.1519	0.2	0.3	0.37	0.0646	0.25	0.25	0.44	<b>0.0307</b>
13q12.3	0.14	0.37	0.26	0.5227	0.22	0.27	0.29	0.5229	0.27	0.32	0.28	0.8429
13q14.2-q14.3	0.07	0.17	0.27	0.1198	0.06	0.26	0.31	<b>0.0013</b>	0.19	0.23	0.4	<b>0.0077</b>
14q32.11-q31	0.43	0.1	0.2	0.0834	0.13	0.26	0.2	0.4642	0.19	0.17	0.25	0.5122
16q11.2-q22.1	0.25	0.43	0.28	1.0000	0.42	0.35	0.18	<b>0.0084</b>	0.35	0.19	0.23	0.1712
16q22.3-q24.3	0.15	0.33	0.31	0.3485	0.33	0.25	0.33	1.000	0.3	0.19	0.35	0.5660
17p13.3	0.21	0.37	0.35	0.3829	0.26	0.25	0.5	<b>0.0224</b>	0.32	0.48	0.38	0.5344
17p13.1	0.31	0.22	0.37	0.7894	0.28	0.24	0.49	<b>0.0237</b>	0.3	0.5	0.42	0.1425
17q12-q21	0.13	0.04	0.19	0.7398	0.07	0.13	0.28	<b>0.0047</b>	0.14	0.21	0.25	0.1080
18q21.1-q21.3	0.14	0.11	0.22	0.7369	0.1	0.28	0.22	0.1291	0.2	0.21	0.25	0.5202
22q12.3	0	0.25	0.2	0.0755	0.12	0.22	0.19	0.4420	0.18	0.23	0.18	1.0000
22q13.1	0.21	0.35	0.25	1.0000	0.26	0.28	0.23	0.6687	0.27	0.24	0.22	0.5567



**A small set of genes identified from peripheral blood leukocyte microarray data can classify breast cancer patients on the basis of clinical parameters**

Jordan R, Brzeski H, Hu H, Shriver CD, Mural R, Liebman MN, Yang Y.

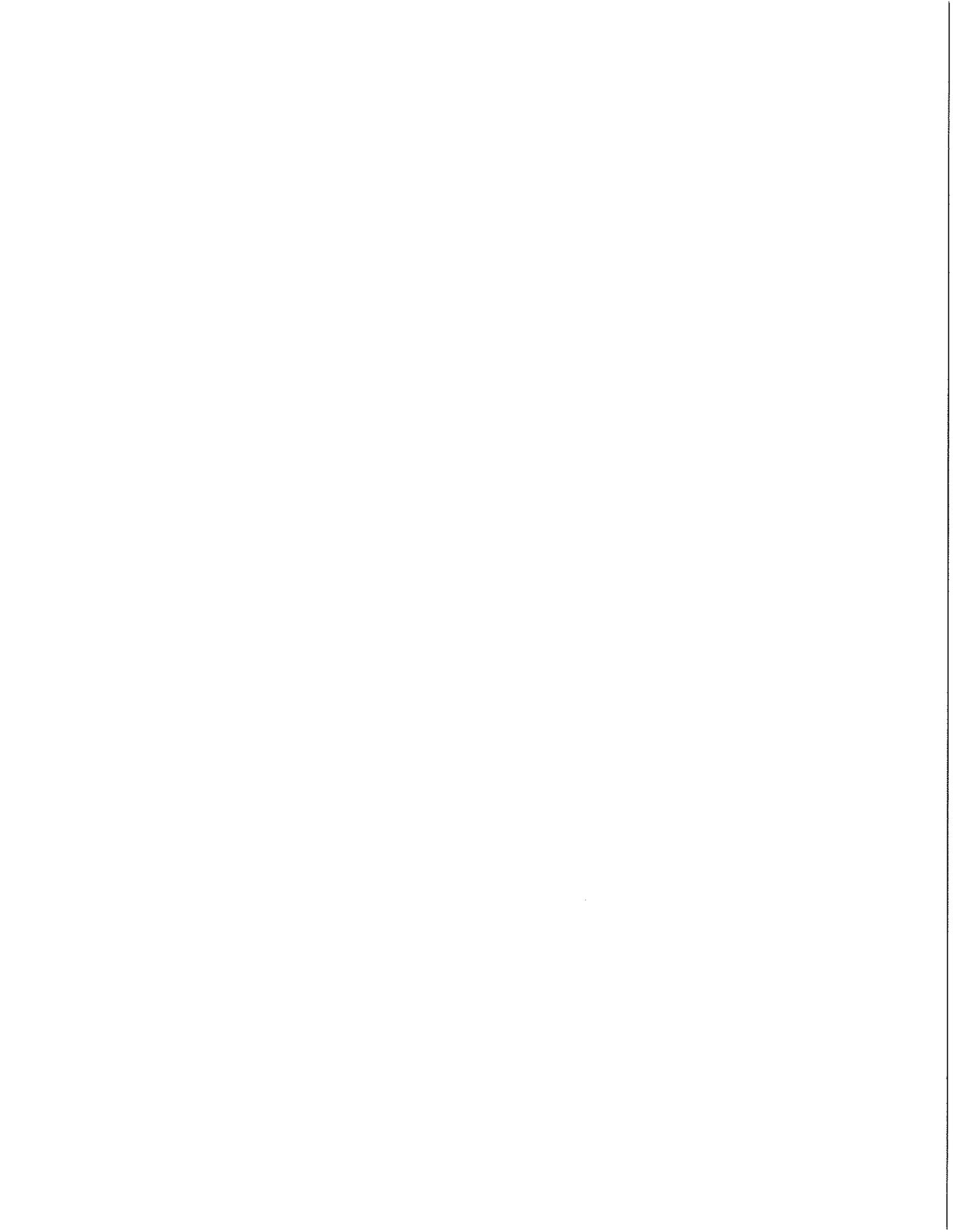
Windber Research Institute, Windber, PA, USA

**Background:** Biomarkers may be identified from microarray data, and may lead to early detection, drug or vaccine targets. There are several reports of biomarkers identified by microarray analysis of breast cancer tissue; however, these lists are not consistent between publications. Since tumors are heterogeneous we performed classifications of peripheral blood leukocyte microarray data from breast cancer patients, using a minimal number of genes to see whether there was a leukocyte gene expression pattern that characterized their response to the tumor.

**Material and Methods:** GE Codelink Human 20K Gene Expression Bioarrays were used to test blood samples, from patients enrolled in the Clinical Breast Care Project (CBCP), with proven breast pathology diagnosis categories of Invasive Cancer (57) and Benign pathologies (69). Patients were randomly divided into training (49%), test (21%), and validation (30%) datasets with respect to each diagnosis category. The Codelink normalized data was filtered to remove missing data, log transformed, and normalized by global mean adjustment. All procedures were carried out using caGBDA (<http://bioinformatics.upmc.edu>). Differentially expressed genes were found using the J5 test and Random Resampling Validation using a Naïve Bayes Classifier at 50 iterations was performed for predictions. Stratifications were based on three parameters: biomarker status (values determined by independent immunohistochemistry (IHC) assays, invasive only), menopausal status, and a benign vs. invasive comparison using ethnicity and menopausal stratifications. Whenever possible, balanced datasets were used. For the gene expression comparisons, the biomarker of interest was removed from the microarray data (e.g. for ER stratified samples, the ER microarray results were not used for the classifications). Small sets of genes were identified from the training and test datasets, and applied to the validation datasets for classification predictions.

**Results:** Analysis of independent validation datasets produced the following classification accuracies: post-menopausal invasive vs. post-menopausal benign, as well as pre-menopausal Caucasian invasive vs. post-menopausal Caucasian invasive both resulted in 100% (14 and 13 genes, respectively). In addition, Caucasian benign vs. Caucasian invasive returned 89% (16 genes), while pre-menopausal invasive vs. post-menopausal invasive resulted in 81% (8 genes). The biomarkers resulted in: ERBB2 status 88% (18 genes), PR status 82% (7 genes) and ER status 76% (5 genes).

**Discussion:** Our preliminary study suggests that a small set of genes identified from microarray experiments performed on peripheral blood leukocytes may be used for patient classifications. We are currently investigating the data for metabolite responsive genes (e.g. those that respond to estrogen). Gene ontology studies have shown a majority of the candidate genes to be involved in either immune response or molecular transport. These preliminary results provide us with an interesting blood signature that may lead to improvement of the diagnosis process.



**Title: Peripheral blood microarray data may aid in predicting lymph node status of breast cancer patients**

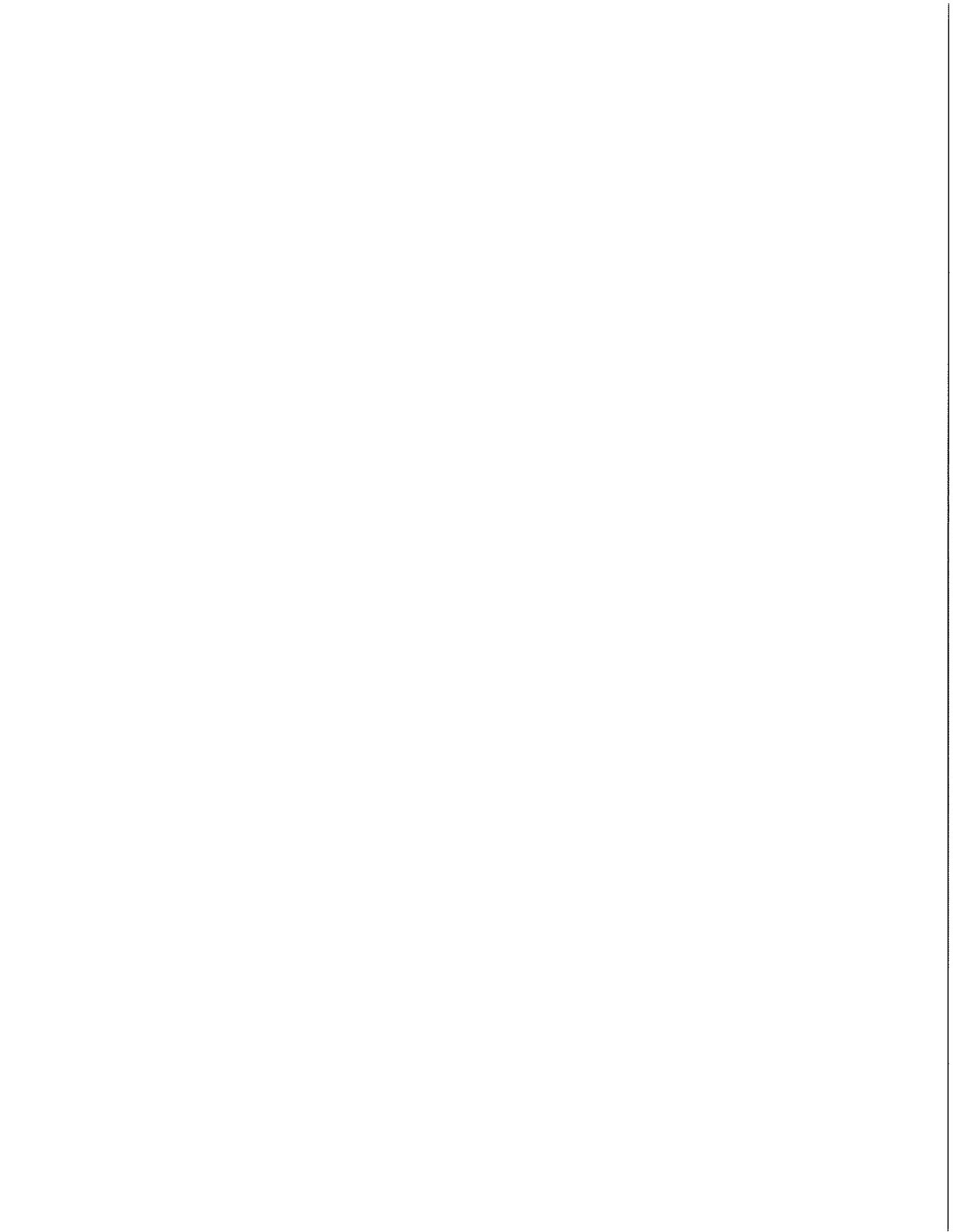
**Rick M Jordan, MS1, Hai Hu, PhD1, Caroline M Heckman, BS1, Leonid Kvecher, MS1, Craig D Shriver, MD2, Richard Mural, PhD1 and Yaw-Ching Yang, PhD1. 1Windber Research Institute, Windber, PA, United States, 15963 and 2Walter Reed Army Medical Center, Washington, DC, United States, 20307.**

**Background:** In order to identify metastatic breast cancer in its truly earliest stages, identification of spread to proximal lymph nodes is of clinical importance. Currently, the methods of diagnosis and detection can be described as invasive procedures. A less-invasive method such as a blood-based gene expression prediction model that could aid in lymph node status (LN) determination could be of immense importance. We performed classifications of peripheral leukocyte microarray data from breast cancer patients, using a minimal number of genes to see whether there was a leukocyte gene expression pattern that characterized lymph node status.

**Material and Methods:** Affymetrix HG U133 Plus 2.0 GeneChip Arrays were used to obtain blood microarray data from 100 patients. Patients enrolled in the Clinical Breast Care Project (CBCP), with proven breast pathology diagnosis categories of invasive pre-menopausal (n=26) and invasive post-menopausal (n=74), of which lymph node status was available for 19 pre-menopausal (10 LN- 9 LN+) and 58 post-menopausal (47 LN- 11 LN+) samples were chosen for this study. The dataset was randomly divided into training (49%), test (21%), and external validation (30%) subsets with respect to each diagnosis category. The data was median-normalized and filtered by removing genes meeting carefully chosen intensity criteria (leaving 17736 probes for analysis). All classification procedures were carried out using caGEDA (cancer Gene Expression Data Analysis tool; <http://bioinformatics.upmc.edu>). Differentially expressed genes were found using the D1 test and Random Resampling Validation with a Nave Bayes Classifier at 10 iterations for predictions. Stratifications were based on lymph node (IHC-verified) and menopausal status. The following comparisons were performed: pre- vs. post-menopausal, invasive LN+ vs. LN-, pre-menopausal LN+ vs. LN-, and post-menopausal LN+ vs. LN-. Small sets of genes (n 25) were identified from the training and test datasets, and applied to the validation datasets for classification predictions.

**Results:** Analysis of independent test validation datasets produced the following classification accuracies: menopausal status 90% (10 pre-menopausal and 20 post-menopausal samples), invasive LN-status in-general 70.8% (10 LN+ and 14 LN- samples), pre-menopausal LN-status 83.3% (3 LN+ and 3 LN-samples), and post-menopausal LN-status 94.4% (4 LN+ and 14 LN- samples).

**Discussion:** While our validation accuracies reached into the 90% range, the training/test accuracies averaged around 70%-80%. Our preliminary study suggests that genes identified from microarray experiments performed on peripheral blood leukocytes initially appear to be partially able to predict lymph node status in breast cancer patients. We are currently working to increase our sample size, as we expect larger sample sizes to provide more definitive predictions.



## Chapter 2 - The Clinical Perspective

Lee Bronfman, Craig D. Shriver, and Emily Gutchell

It is within the clinical setting that the need for biomedical research is revealed, and to this setting that the advancements of biomedical research return. No one truly needs to be convinced of the need for clinical research. There is intrinsic agreement that as human beings we have extraordinary needs for quality healthcare. The debates ensue regarding how research is conducted and move into a myriad of potential lively discussions from there. Yet it is rarely debated that research is essential enterprise, and that it should be conducted with the highest level of integrity. A research project that meets the highest standards will begin with a well-devised plan for launching the project from the clinical setting. This chapter will discuss how to implement those aspects of a biomedical informatics research project that are conducted in the clinical health care setting. Typically, these processes are consenting research subjects, collecting clinical information and biospecimens from research subjects, and transferring the many data and specimens to a research facility for storage and analysis. Easily done, one might think. And it is so, if one takes into account the complexity of getting there.

### 2.1 Developing a Research Protocol

When biomedical research incorporates clinical data to drive genomic and proteomic research, a specimen (tissue, blood, bone marrow, etc.) for analysis is required. In order to obtain that specimen and use it for research, a research

protocol must be developed. Reed and Jones-Wells suggest that a protocol can be thought of as a blueprint for what is to be done, how it is to be done, and who is to assume ultimate responsibility for it.[1] To effectively write a protocol, these three basic questions of what, how, and who must be fully addressed. The protocol will be submitted to the Institutional Review Board (IRB) affiliated with the organization where the research will be conducted, and must be approved prior to the beginning of any research. Institutional Review Boards exist to assure that research conducted within the organization is well designed and without undue risk to participants, and that human subjects are protected from unethical practices. Augmented by implications of the genomic era, as well as regulatory mandates to protect personal health information, protecting privacy has also become an important concern to IRBs. The specific requirements for writing and submitting a research protocol to an IRB are unique to any given organization, yet because the goals of an Institutional Review Board are the same everywhere, the required elements of a protocol are fairly consistent. Organizations such as the U.S. Food and Drug Administration (<http://www.fda.gov/oc/gcp/guidance.html#guidance>), the Office for Human Research Protections (<http://www.hhs.gov/ohrp/education/#activities>), the National Institutes of Health (<http://clinicalcenter.nih.gov/researchers/training/ippcr.shtml>), and the Society of Clinical Research Associates (<http://www.socra.org/>), offer courses and educational material on developing clinical research protocols. Let us consider each section of a standard clinical research protocol.

### 2.1.1 The Cover Page

The cover page of a protocol will include the title of the study, the name, title, and contact information of the Principal Investigator, any Associate Investigators and/or collaborators, and if required, a Medical Monitor. Figure 2.1 shows a sample cover page for a protocol.

Figure 2.1 goes here

The Principal Investigator (PI) is the individual primarily responsible for the actual execution of the clinical investigation. He/she is responsible for the conduct of the study, obtaining the consent of research subjects, providing necessary reports, and maintaining study documents. Some or all of these responsibilities can be, should be, and often are, delegated to other trained professionals with the requisite expertise, under the supervision of the PI. The PI is usually affiliated with the institution where the research is to be conducted and if not, an administrative point of contact will be identified.

Associate Investigators are other key personnel involved in the governance of the study who are affiliated with the sponsoring institution. For the example of a protocol to study breast disease, these may be a breast surgeon, a pathologist, a medical oncologist, a radiation oncologist, a radiologist, etc.. Collaborating personnel are all other key personnel involved in the governance of the study who come from outside the sponsoring institution. In our example, the collaborating personnel identified could be the vice president of the research facility where the biorepository is located, and the breast surgeon consenting patients to the protocol at the second site. A Medical Monitor is a physician who

monitors the study from a medical safety perspective and is not involved in any way with the research study. When a study is assessed to pose greater than minimal risk to its subjects, a Medical Monitor will likely be necessary. The responsibilities of a Medical Monitor include:

- 1 Periodic review of research files, medical records, etc.
- 2 Review of annual progress reports, adverse events, protocol deviations.
- 3 The ability to terminate the study or participation in the study if there is concern about the welfare of enrolled subjects.

Many research studies only pose minimal risk to participants and may not require a Medical Monitor. Minimal risk to human subject participants is defined within the Code of Federal Regulations. "Minimal risk means that the probability and magnitude of harm or discomfort anticipated in the research are not greater in and of themselves than those ordinarily encountered in daily life or during the performance of routine physical or psychological examinations or tests." [2]

### 2.1.2 The Body of the Research Protocol

The following sections, or variations thereof, comprise the body of a research protocol.

#### Objectives

The objectives of the protocol are stated, including the research hypothesis if any. Research objectives should be consistent with the Plan and Data Analysis sections. The medical application or medical importance and usefulness of the results of the study are often cited in the objectives.

#### Background and Significance

The Background and Significance section of a protocol is necessary to justify the need for the project and explain the rationale. It includes a summary of relevant literature. Making this section easy to understand—especially to those outside of the specialty of the author, is important.

### Plan

The protocol's plan is comprised of many subsections which together delineate the broad scope of processes and methodologies that will be implemented by the protocol. Each subsection must be addressed briefly. This assures the IRB that the plan for implementing the protocol has been thoroughly conceived. The subsections include, but may not be limited to:

- Study design
- Recruitment methods, with inclusion and exclusion criteria for subject participation
- An estimate of the sample size planned for the study
- Anticipated start date and expected completion date of the study
- Data collection practices, including types of data collected as well as practices for assuring patient confidentiality, including processes for handling protected health information as defined by the Health Insurance Portability and Accountability Act
- Data analysis practices , i.e., a general description of how the data collected will be analyzed and by whom

When a study includes the collection and storage of human biological specimens, additional information will be required. Such information will include but may not

be limited to:

- Obtaining informed consent for future use of specimens and clinical data
- Specimen collection procedures
- Specimen shipping procedures
- Specimen storage procedures
- Specimen confidentiality, and how it will be maintained
- Who will have access to specimens
- Length of time specimens will be stored
- Withdrawal and destruction of specimens

The research office associated with the IRB that will approve your protocol may have valuable guidelines and/or standard formats for submitting the plan, as well as other sections of the protocol.

### References

Include a section of references citing the publications used to develop the proposal.

### Budget

The budget for the research project is submitted with the protocol. Individual IRBs usually require specific budget formats.

### Attachments Relevant to the Research

The IRB will require copies of documents to be used in conducting the research such as consent forms, questionnaires/survey instruments, brochures describing the research, and marketing material. If there are additional, collaborative protocols or grant proposals affiliated with the project, these will be required by

the IRB.

Developing a protocol is a time consuming, labor intensive undertaking. There will be many drafts, revisions, and consultations with associate investigators before a final document is submitted to an IRB. Once the review process begins with the IRB, correspondence and communication will require additional clarification and revision before a final version is approved. Although all IRBs must adhere to the Common Rule, [3], individual IRBs may function quite differently from one another. It behooves the Principal Investigator and Study Coordinator to become familiar with the requirements of their IRB. Attention to detail in this phase of research will prove beneficial as the study is conducted. All research protocols involving human subjects must be reviewed at least once per year. After the IRB has granted initial approval of your protocol, a yearly, continuing review will be required. Your IRB will provide the format for what is called a Continuing Review application. This is likely to include a progress report of the number of enrollees over the past year, as well as the duration of the study, changes in such things as survey instruments, consent forms, or any changes to the protocol's activities that require approval from an IRB. If there have been any significant events requiring regulatory reporting, the original report to the IRB of the event is included in the progress report.

## 2.2 Ethics in Clinical Research

Institutional Review Boards exist to protect the rights and the welfare of human subjects participating in research. Various ethical codes and reports provide guidelines and mandates for the ethical conduct of clinical research. The

Nuremberg Code was developed during the trial of Nazi physicians after World War II, and was followed by the Declaration of Helsinki in 1964. A "living" document, the Declaration of Helsinki was revised in 1975, 1983 and 2000. The Council for International Organizations of Medical Sciences (CIOMS) published International Ethical Guidelines for Biomedical Research Involving Human Subjects in 1993, a collaborative effort with the World Health Organization, which targeted the unique concerns of human subject protections for research conducted in developing countries.

Henry Beecher, MD, published an expose in The New England Journal of Medicine in 1966 where he cited numerous examples of unethical research.[4] The Tuskegee Syphilis Study, begun in 1932 and finally stopped in 1972, was a United States Public Health Service Study where penicillin treatment was withheld from African American men in Tuskegee, Alabama, who had been diagnosed with syphilis. Dr. Beecher's exposure of the Tuskegee Study led to the development of the National Research Act in 1974. The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research was formed, and in 1979 that Commission produced the Belmont Report. The Belmont Report is a summary of ethical principles intended to provide national guidelines to all those involved in research with human subjects. The Belmont Report focuses upon three basic ethical concepts in research: Respect for Persons, Beneficence, and Justice.[5] The Belmont Report led to the Common Rule. The Common Rule is a federal policy which directs a research institution to assure the protection of human subjects participating in research. Title 45 of the

Code of Federal Regulations, Subpart A, Part 46 describes the Common Rule in its entirety.

#### 2.2.1 Regulatory Policies for Protecting a Research Subject's Privacy

Regulatory policies for safeguarding the privacy of personal information vary among countries and have impact upon the use of data in research. Privacy policies governing personal data originate from different angles in different countries, such as banking, e-commerce, business, and healthcare. Regardless, these laws can have regulatory reach into ensuring privacy protections for research subjects. Researchers in Canada are impacted by the Personal Information Protection and Electronic Document Act; researchers in the European Union, by the European Union Directive on Data Protection. In the United States, it is the Health Insurance Portability and Accountability Act, (HIPAA), that researchers need to understand. The HIPAA Privacy Rule is administered through the United States Department of Health and Human Services, Office for Civil Rights. Enacted to protect employee health and insurance when workers changed or lost their jobs, the HIPAA law also provided guidelines for agencies to protect an individual's health information. Increasing Internet usage and digital exchange of health information led to updated HIPAA guidelines which addressed protected health information in electronic privacy and confidentiality issues, as well as data collection in human use research protocols. Although HIPAA was enacted in 1996, nationwide compliance to the HIPAA Privacy Rule was implemented in April 2003.

The HIPAA Privacy Rule sets standards that limit the use and disclosure of

protected health information for research purposes. The Privacy Rule does not replace or modify the Common Rule or U.S. Food and Drug Administration (FDA) regulations. The Privacy Rule works with these other regulations to increase human subject protections. The intent of the HIPAA Privacy Rule is to protect the privacy of individually identifiable health information by establishing conditions for its use and disclosure. Covered entities are subject to the HIPAA Privacy Rule. A covered entity is a health care provider, health plan, or health care clearinghouse. Researchers working within or collaboratively with covered entities become bound by provisions of the Privacy Rule as well. Thus, many researchers must obtain written permission from study enrollees in the form of an HIPAA authorization, before the enrollee's "protected health information" (PHI) can be used in research. If the researcher elects to use only a limited data set as defined by the Privacy Rule, written authorization from the participant isn't required, however, a data use agreement between the covered entity and the researcher is required, assuring that specific safeguards will be taken with the PHI that remains within the limited data set. The IRB may require evidence of the data use agreement in the protocol submission or may require a reference to the data use agreement within the informed consent document. Informed consent is covered in the next section.

The other consideration for researchers is to use only de-identified health information, as defined by the Privacy Rule, for which there are no restrictions on use and disclose. There are two ways to de-identify health information:

1. Remove all of 18 unique identifiers specified by the Privacy Rule, thereby

assuring no information can identify the individual.

2. Statistically de-identify information whereby a statistician certifies that there is a "very small" risk that the information could be used to identify the individual.

The following are the 18 unique identifiers that must be removed from health information in order for that information to be considered de-identified: names, locations smaller than a state, dates that relate directly to an individual, telephone and fax numbers and e-mail addresses, social security numbers, medical record numbers, prescription numbers, health plan beneficiary numbers, other account numbers, certificate/license numbers, vehicle identification numbers, serial numbers, license plate numbers, device identifiers/serial numbers, WEB URLs and IP addresses, biometric identifiers such as fingerprints and voiceprints, photographic images, and any other unique identifying numbers or codes. Given this exhaustive list and the nature of clinical and demographic data collected for biomedical research projects, it is likely that the biomedical researcher will need to obtain written HIPAA authorization from research subjects.

HIPAA authorization is different from, but may be combined with, the informed consent document. It must be for a specific research study; one blanket HIPAA authorization for all research within an organization is not permitted. The National Institutes of Health provide user friendly guidance on the core elements and required statements that must be part of the HIPAA authorization. The website is <http://privacyruleandresearch.nih.gov/authorization.asp>. Additionally, the requirements are published in the Code of Federal Regulations, Title 45, Part

164.508 (c) and (d).

The IRB and privacy board of your institution don't need to review or approve the HIPAA authorization, unless you elect to combine it with your informed consent document, in which case the IRB will review it as part of informed consent. The authorization form need not expire, and the fact that it does not must be stated. All research participants must receive a signed copy of their HIPAA authorization.

### 2.3 Informed Consent: Writing the Consent Form

Informed consent must be obtained from every research subject. This is a federal requirement mandated by the Code of Federal Regulations, Title 21, Part 50, Section 50.25, and Title 45, Part 46, Section 116. The intent of the regulation is to provide research subjects with all the necessary information needed to make an informed decision whether or not to participate in the research. There are many required elements in a consent form. Each is listed within the code noted above. Among them, of key importance to the biomedical informatics researcher are a description of the study's research purposes and procedures, identification of any procedures in the protocol that are considered experimental, a disclosure of alternative treatments that may be of benefit to the patient, the expected duration of a subject's participation, a description of any reasonably foreseeable risks, a statement describing the extent to which confidentiality of records will be maintained, and, where appropriate, a statement of significant new findings that developed during the course of the research, which could impact the subject's willingness to continue participation.

Many Institutional Review Boards recommend that consent forms be written at an

eighth grade reading level. This becomes truly challenging when genomic and proteomic investigations are the goal of the researcher. Mark Hochhauser, Ph.D., is a readability consultant and advocates the use of plain English for consent forms. He recommends including a table of contents in a consent form and following a question and answer format.[6]

Microsoft Word offers functions for calculating readability scores for a document. These can be optionally displayed when using the Spelling and Grammar tools. A student's or children's dictionary can also be a useful tool. It is also important that the consent form be written in the language that the research subjects use. For this reason, it may be necessary to have more than one language version of the consent form. Be sure to include all of the required elements as outlined in the Code of Federal Regulations in your informed consent document. These are likely to be reviewed carefully by the IRB.

Below is an example of a portion of the informed consent document used by the Clinical Breast Care Project (CBCP). CBCP is a biomedical research collaborative between Walter Reed Army Medical Center in Washington DC, and Windber Research Institute in Windber, PA. The CBCP seeks to decrease morbidity and mortality from breast cancer and breast disease by providing patients with state of the art clinical care, including advances in risk reduction. The CBCP is also committed to a translational research goal of rapidly translating research findings into clinical practice. The CBCP has a state of the art biospecimen repository to which research subjects contribute blood and tissue samples. Multiple questionnaires containing clinical data are collected on each

patient and biospecimens are extensively annotated. Researchers use the repository to conduct proteomic and genomic research. Clinical data are analyzed in response to proteomic and genomic analysis and findings. Biomedical informatics scientists develop informatics structures and expand the informatics environment upon which the research relies for rapid translation into the clinical environment. Given the challenges of adequately describing research with many goals, in easy to understand language, the section of the informed consent document used by the CBCP that describes the study is two pages long.

The following is an excerpt from that section:

The tissue and blood samples will undergo the primary research studies as described in the next paragraphs.

This breast tissue can be used for many types of laboratory research looking at cell changes during breast cancer development, identification of risk factors that lead to breast cancer, learning why and how breast cancer spreads to lymph nodes, and breast tissue biology. Some of this type of research on tissue is genetic research. These research studies may include examination of the gene markers of the breast cells or lymph nodes, as well as other parts of the breast cells.

The primary research uses of the blood and tissue samples are to study the genetic makeup, protein changes, and to look for other research markers that may be associated with breast disease. This information will be analyzed and stored in a computer database using your assigned code number.

Any remaining blood and tissue samples left over after these studies are done, will remain frozen indefinitely until needed for approved research projects. There may be future uses of the remaining frozen samples of your blood that are not known to us at this time. If other research is to be conducted using these samples, this research also will only be conducted after the samples have been stripped of all identifiers. It is not possible at this time to predict all the potential future research uses of the samples. You will be given the opportunity at the end of this consent form to indicate if you will allow your blood and tissue samples to be used for future unknown breast disease research.

### 2.3.1 Consenting for Future Research

Although not required by the Code of Federal Regulations, an important item on a consent form for proteomic and genomic research, and biomedical research in general, is an option to consent for future research. If approved by the IRB, this will allow researchers to conduct proteomic and genomic research in the future without having to obtain a second consent from the subject. Many ethicists and, in fact, most patients prefer this approach to consenting in advance for related, but presently unknown, future research on specimens and data collected now. It avoids a "cold call" from the research team in what might be years after the patient has successfully completed a course of treatment for the disease.

Readdressing that difficult time in order to obtain an updated consent, can cause needless emotional turmoil. If the patient should die of his or her disease, and the research team then inadvertently contacts the family, this too can have significant emotional effects on all concerned. Enlightened IRBs, especially those with good patient or lay person representation, understand this important point about allowing for the possibility of patients to consent now, for future unknown research uses of their donated biospecimens and data. For those studies employing future use consenting, it is important that a reliable process exists that identifies to the appropriate study personnel in a timely manner which participants' data may not be used for future research. In the experience of the CBCP, only about 2% of the participants refuse to consent for future use research.

The means to obtain consent from patients for future research uses of their clinical data, including imaging data, as well as their biospecimens could prove

pivotal to the success of a biomedical research project in the genomic age.

#### 2.4 Collecting Clinical Data: Developing and Administering Survey Instruments

The array of clinical data collected by the biomedical research project will depend upon the goals of the project. Representatives from the clinical setting and the research setting will be involved in determining what information is necessary to collect. Reliable and valid instruments exist, and a review of them may provide an adequate tool for use. Also available are the services of consulting firms specializing in research survey instruments. These firms offer expertise in instrument development, optimal interviewing formats given the characteristics of the instrument, training of staff for standardized administration of the instrument, as well as data collection and data management services.

If you develop your own instrument, it may be necessary to convene a multidisciplinary task force to successfully compile all data elements necessary for the project. Staff with expertise along the continuum of collecting, processing, storing, and analyzing the data that will be collected should be consulted. These are not only clinical experts and scientists, but also business administrators at the organizations involved in the research, research specialists adept at consenting and administering questionnaires to research subjects, IT specialists in programming, web service and database management, and data quality specialists. Each of these knows something helpful about obstacles and solutions to successfully obtaining and securing data at their organization and/or their level of enterprise.

The method by which the survey instrument will be administered must also be decided upon. Although there are many ways to administer a research questionnaire (mail, Internet, telephone, computer, chart abstraction, focus groups, etc.), the personal, one-on-one interview conducted by qualified and trained staff may yield the highest quality data. Once the content of the instrument has been established, the reliability of the responses it generates depends upon the instrument being administered the same way to every respondent. This reliability is achieved by standardizing administration across interviewers. (American Institutes for Research, personal communication, January 2004).

To standardize administration across interviewers, training is required. Providing your staff with professional training on administering the survey will clarify any difficult or confusing questions, establish trust between interviewer and interviewee, empower research staff to confidently administer the survey, and result in better data collection. It may allow for a longer, more extensive survey instrument to be reliably completed by minimizing "questionnaire fatigue" that patients experience when filling out lengthy research tools.

## 2.5 Issues Important to Biomedical Informatics

Along with writing the protocol and developing the survey instrument(s), planning for the successful storage and analysis of all data generated by the processes involved in the study is important. Most of this will happen in the hands of the biomedical informatics experts. However, the clinical team is responsible for the initial collection of most of the data, and as already outlined, must inform the

research subjects of the types of data that will be collected, and how it will be privately and securely stored and analyzed.

#### 2.5.1 De-identifying Data

Questionnaires, medical images such as mammograms and ultrasound images, and biospecimens will need to be de-identified, to the extent stated on the Informed Consent and HIPAA Authorizations. Yet, these data must also be tracked in organized fashion for successful data management. A research project-specific ID number, not the subject's social security number or hospital number, etc., needs to identify all biospecimens and data collected for each enrolling individual. The project must plan for how a research subject's ID number will be assigned, where and how the connection between the assigned ID number and the subject's name will be stored, and who, besides the Principal Investigator, will have access to this information. Biospecimens and questionnaires will additionally require a tracking barcode or other identifying number that links to the research subject's assigned ID number and also the date these data are collected. There may be multiple surveys, images, and biospecimens from one research subject, collected throughout the lifetime of the project.

#### 2.5.2 Quality Assurance

The clinical team will need to communicate closely with the data management specialists who receive and review the questionnaires, biospecimens, and other clinical data such as mammograms and ultrasound images. It is vital to have open communication channels for resolving discrepancies, as well as electronic

data tracking tools for making changes to data collected. Neither medical charting, nor the processes of retrieving medical and social histories from patients are 100% reliable. The process of obtaining as accurate and complete a data record as possible from a patient can require much back and forth between data collectors and data reviewers, depending upon the volume of patients participating in the research study as well as the volume of data collected. At the Clinical Breast Care Project, the Windber Research Institute developed a user-friendly online data tracking tool that streamlined the process of identifying, communicating, and resolving discrepancies in clinical data collected. There are thousands of enrollees in the CBCP, and the project collects many hundreds of data elements from each enrollee, such as family cancer histories, dates of menarche and menopause, hormone replacement therapy use and duration, hormonal birth control use and duration, pregnancy and birth histories, intake of tobacco, alcohol, fat, caffeine, underarm cosmetic use and duration. Additionally, diagnosis and treatment data are collected, such as mammogram dates and findings, ultrasound dates and findings, CT scans, x-rays, bone scans, PET scans, MRI, chemotherapy, radiation therapy, hormonal and endocrine adjuvant therapies, pathologic diagnoses from biopsies, lumpectomies, mastectomies, etc. Without precise methods for correcting errors, the project would lose vital data strength and compromise its ability to conduct excellent research. The online tracking tool, called the Quality Assurance Issue Tracking System (QAITS) provides secure online communication between the data entry technicians and the quality specialists who review all the questionnaires for completeness and

accuracy before data are entered. The issue that requires clarification or correction is entered into the QAIT by a data entry technician. A quality specialist opens the issue and resolves it. This may require verifying information in a medical record, and/or verifying information through the nurse who initially collected it, and possibly will require a call back to the patient. When the quality specialist has successfully resolved the issue using the QAIT, the issue is then closed by a data entry technician. The steps, dates, and times along the way are recorded electronically by the tool, and the corrected data is then entered into the data tracking system. All this happens after the quality specialists have conducted initial reviews of questionnaires before they are transferred to the research facility for data entry. In addition to data problem resolution, the QAIT provides reports that enable management to better supervise, train, and track proficiency of staff involved in the data collection and quality assurance processes. Figure 2.2 depicts a list of issues that require clarification and/or correction by users of the QAIT.

Figure 2.2 goes here

### 2.5.3 Data transfer from the clinic setting to the research setting:

The biomedical information collected within the clinical setting must be transferred to the research specialists and scientists who will use and analyze the information. This may involve entry into the data tracking systems and repositories that are on site, and/or shipping to sites where the data and specimen management systems are located. Regardless, quality control practices that ensure the integrity of clinical data and biological specimens during

transfer are of paramount importance. Quality control methods for transferring biological specimens to the repository are covered in Chapter 3. Additionally, all clinical data should undergo quality review before being cleared for data entry. The data entry process should also have a quality oversight component. If you have a data tracking system whereby data entry occurs directly, with no opportunity for quality review before data entry, it is necessary for quality control procedures to begin at the point after which the data is entered, but before it is analyzed by the researchers. The challenge of the quality review process is to ensure the clinical data collected are as accurate and complete as possible. An organized data management system such as the QAIT tool described above, that tracks corrections and changes to data is essential for maintaining integrity in processing clinical data. Coordination of the quality review takes continual involvement and oversight. A Standard Operating Procedure that details a process for ensuring good quality control of clinical data collected on survey instruments may be essential for projects that collect many clinical data elements.

Once biospecimens and clinical data arrive at the research facility where they are stored and utilized, robust data tracking and data centralization systems will be necessary to meet the goals of a large biomedical research project. Such systems are discussed in Chapters 7 and 8.

#### 2.5.4 Standard Operating Procedures (SOPs)

Written policies and procedures can greatly assist in assuring standardized adherence to regulations, guidelines, local policies and institutional directives.

They provide staff with easy to locate guidance. SOPs should be living documents, meaning they should be developed, shared with appropriate staff, reviewed, and revised periodically. The format of SOPs should be standardized. Staff can then rely upon the existence of SOPs within a familiar format that is easy to find within the clinic and/or research setting. The SOP should be titled with the subject or standard it addresses. It should then state why the organization has this standard and who must comply with it. Lastly the steps of the procedure should be clearly delineated. A biomedical research project will require an SOP for the retrieval, storage and shipping of biospecimens. Also useful will be an SOP for consenting patients to the protocol. It will further be useful to have an SOP for receiving, reviewing, and entering questionnaire data. Figure 2.3 shows the table of contents for an SOP that delineates how the core questionnaire from the Clinical Breast Care Project is processed after it has been administered to a research participant.

Figure 2.3 goes here

SOPs define the scope of practice for an individual or a program. They provide guidelines for new members of the team. They document roles and responsibilities for the research team, and they set the standards of good practice for the operation of your project. It is worthwhile to keep the benefits in mind when faced with how time consuming it is to write useful SOPs, and the fact that they must be periodically reviewed and updated to reflect changes. Make every effort to write SOPs that can be, and are, enforced in practice—deviations from them may not be viewed kindly by an auditor.

## 2.6 Implementing the Research Protocol

Once the protocol, consent form, questionnaire/survey instrument(s), and HIPAA authorization have been approved by the IRB, and the biomedical informatics systems and repositories are functioning to receive data and specimens, a planned implementation of the research protocol can take place. Clinical staff responsible for specimen acquisition need training in exactly how specimens will be obtained, transported within the clinic, stored temporarily, and shipped to the repository. Team planning and "dry runs" for every aspect of the acquisition process, but with no actual specimen, should be considered. This will help ensure that the first batch of biological specimens sent to the repository location will not be compromised due to unforeseen problems. Similar "dry runs" should be done with subject enrollment and clinical data management.

Those staff responsible for obtaining informed consent and administering survey instruments to subjects need to be thoroughly knowledgeable about the protocol and proficient in the consenting and survey administration processes.

Completion of a live or online course in the protection of human subjects is recommended for all staff directly involved in the research, and is required by some IRBs. Courses may be offered through professional organizations such as the Society of Clinical Research Associates.[7] The Office for Human Research Protections and the National Cancer Institute offer free web-based resources for human subjects training.[8-9] Information about international codes and ethics for human subject protections is also covered by these sources. The Collaborative IRB Training Initiative (CITI), out of the University of Miami, provides a

comprehensive selection of educational modules for biomedical, and social and behavioral researchers. Continuing education modules are also offered as a refresher, to ensure that research staff remain competent in the area of Human Subject Protection.[10]

Staff responsible for consenting subjects to the protocol, as well as completing the questionnaire/survey instrument, should be encouraged to keep track of any issues or problems experienced with the consent form or the survey instrument. Staff meetings can be an ideal time to discuss "best practices" in administration of the instrument and consenting of subjects. Also discuss what the frequently asked questions by subjects are, which questions prove to need special attention and clarification. This will result in reliability and standardization of both the administration of the survey instrument and the obtaining of informed consent. A successful biomedical informatics research initiative will include a solid orientation and continuing education programs for staff.

## 2.9 Summary

Participating in a research protocol is a privilege. Many research participants and many staff members demonstrate this sentiment in the way they approach their roles and tasks within the research project. Conducting research in a clinical setting is an endeavor in which people take pride. It is beneficial to those doing the work now, as well as those who will be beneficiaries of that work in the future.

Preparing the documents and the processes for ensuring a well run protocol is a time consuming endeavor requiring collaboration from many experts. The

administration of the protocol must be planned, the formal protocol written, and the informed consent and HIPAA authorization documents written. A survey instrument must be developed. Approval from the IRB of the organization(s) conducting research must be obtained. All staff participating in research activities must be trained for their unique roles in the study as well as in research ethics, including the protection of human subjects. Data management systems for ensuring integrity and anonymity of biospecimens and clinical data must be developed and followed. Quality oversight and ongoing staff education must be conducted with regularity. Biospecimens and clinical data must be successfully transferred from the clinic setting to the research setting that will receive them and enable the work of the scientists to be carried forward in the activities of research.

## 2.10 References

- [1] Reed, E., and A. Jones-Wells, "Writing a Clinical Protocol: The Mechanics". In *Principles and Practice of Clinical Research*, p. 439, J.I. Gallin (ed), San Diego, CA: Academic Press, 2002.
- [2] Code of Federal Regulations, Title 45—Public Welfare; Subtitle A—Department of Health and Human Services, Part 46, subsection 46.102.
- [3] Code of Federal Regulations, Title 45, Subpart A, Part 46.
- [4] Beecher, H.K., "Ethics and Clinical Research," *The New England Journal of Medicine*, Vol. 274, No. 24, 1966, pp.1354-1360.
- [5] Website, "The Belmont Report",  
<http://www.hhs.gov/ohrp/humansubjects/guidance/belmont.htm>.
- [6] Hochhauser, M., "Informed Consent Reading, Understanding and Plain English," SoCRA Source, Issue 42, November 2004, pp. 24—26
- [7] Website, "SOCRA, Society of Clinical Research Associates",  
<http://www.socra.org/>.
- [8] Website, "United States Department of Health and Human Services, Office for Human Research Protections", <http://www.hhs.gov/ohrp/>.
- [9] Website, "National Cancer Institute, Human Participant Protections Education for Research Teams",  
<http://cme.cancer.gov/clinicaltrials/learning/humanparticipant-protections.asp>.
- [10] Website, "CITI Collaborative Institutional Training Initiative",  
<http://www.citiprogram.org/>.

## Figure Legends

### Figure 2.1

Sample cover page of a protocol to study breast disease by collecting biological specimens from research subjects at two different clinical sites, and storing and analyzing the specimens at a research facility

### Figure 2.2

The list of current issues being generated in the QAIT online tool used by the Clinical Breast Care Project. Notice the status column. Issues that are ready for a quality specialist to respond to are marked "Ready for QA". Issues that are still being formulated by a data entry technician are marked "Data Entry". Issues that have been resolved by a quality specialist are marked "QA: (no pending issue)" along with the quality specialist's name.

### Figure 2.3

Table of contents for a Standard Operating Procedure used by the Clinical Breast Care Project to standardize the way research questionnaires are processed.

1. **PROTOCOL TITLE:** Tissue and blood repository for molecular, biochemical and histologic study of breast disease

2. **PRINCIPAL INVESTIGATOR:**

[Name of Principal Investigator]  
[Title of Principal Investigator]  
Phone:  
Fax:  
Email:

3. **ASSOCIATE INVESTIGATORS:**

[Name of Associate Investigator 1]  
[Title of Associate Investigator 1]  
Phone:  
Fax:  
Email:

[Name of Associate Investigator 2]  
[Title of Associate Investigator 2]  
Phone:  
Fax:  
Email:

[Name of Associate Investigator 3]  
[Title of Associate Investigator 3]  
Phone:  
Fax:  
Email:

4. **COLLABORATING PERSONNEL:**

[Name of Collaborating Personnel 1]  
[Title of Collaborating Personnel 1]  
[Address of Collaborating Personnel 1]  
Phone:  
Email:

Responsibility: Carry out successful receipt of de-identified donor specimens transferred to the research and tissue banking facility, and to manage and oversee the scientific experiments and handling of results of all studies involving the said specimens at said facility.

[Name of Collaborating Personnel 2]  
[Title of Collaborating Personnel 2]  
[Address of Collaborating Personnel 2]  
Phone:  
Email:

Responsibility: Act as onsite Principal Investigator at the collaborating clinical center for the tissue and blood repository and high-throughput research initiatives as described in this protocol.

5. **MEDICAL MONITOR:** Minimal risk study, medical monitor not required.

# WRI Shared Resources

Jun 4, 2007 -- Monday  
 provided for Biomedical Informatics Group

Home | [COLWS](#) | [My Profile](#) | [Log out](#)

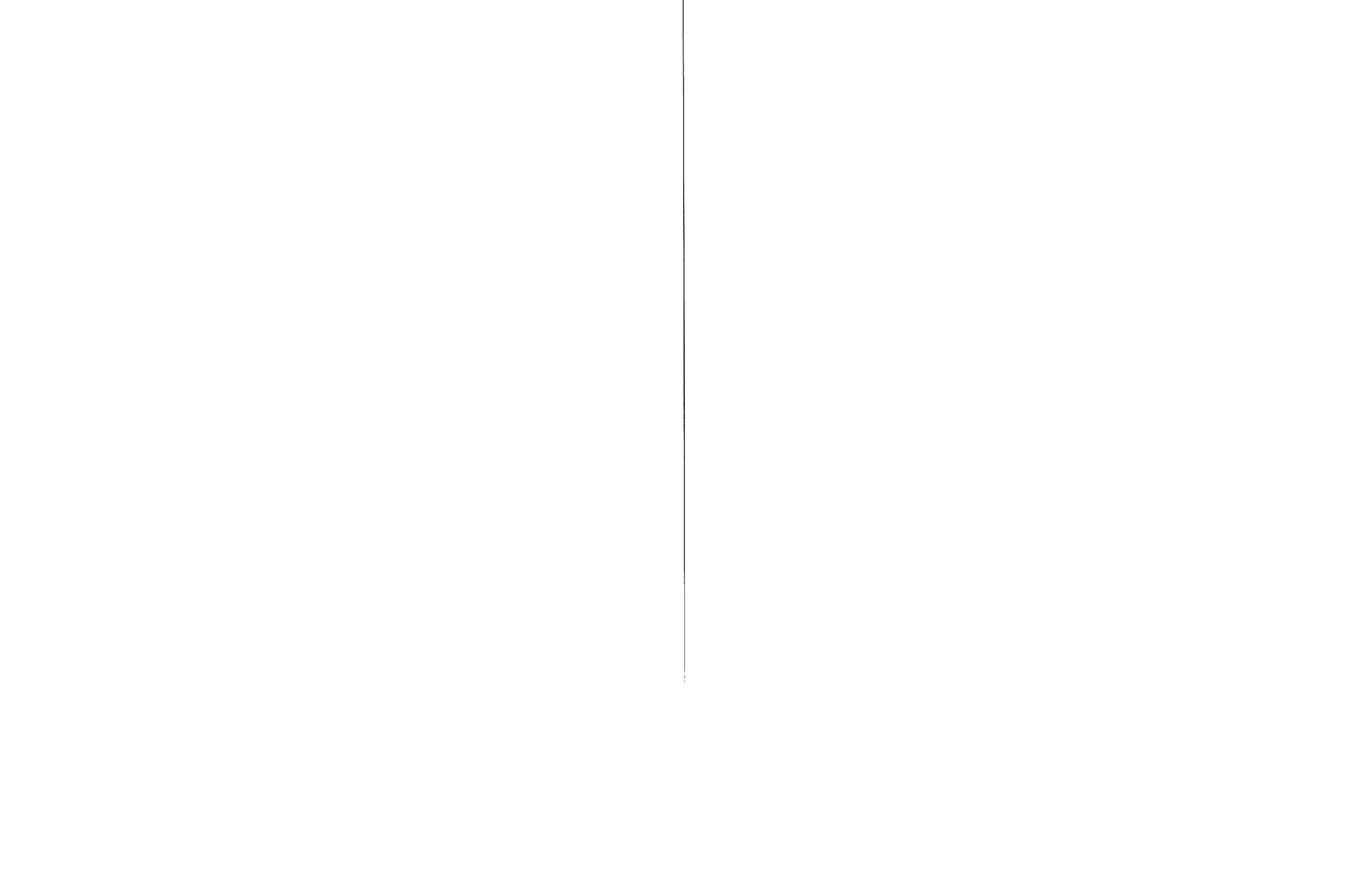
Emily Gurchell

- Core QA
- questionnaires report
- search
- QA summary
- recent issues/FU
- help

Display settings:

0, All  
 Active all

Number	Barcode	ISSN	Core Questionnaire Title	ISSN Date	Version	Status
1. [ + ]	000068BFP	100000146	04/12/2007	05/29/2007	Follow-up (04/14/2003)	Data Entry
2. [ + ]	000029R16	000200134	04/10/2007	05/29/2007	Follow-up (04/14/2003)	Data Entry
3. [ + ]	000068S1P	000000592	04/10/2007	05/25/2007	Follow-up (04/14/2003)	Data Entry
4. [ + ]	000066K9G	000000180	04/03/2007	05/25/2007	Follow-up (04/14/2003)	Data Entry
5. [ + ]	000068S1N	100000707	04/10/2007	06/04/2007	Follow-up (04/14/2003)	Ready for QA [Claim for QA]
6. [ + ]	000068BDY	000400014	02/20/2007	06/01/2007	Follow-up (04/14/2003)	Ready for QA [Claim for QA]
7. [ + ]	000068BC2	100001126	01/29/2007	06/01/2007	Core questionnaire (04/14/2003)	Ready for QA [Claim for QA]
8. [ + ]	000068BDY	100000357	01/22/2007	05/30/2007	Follow-up (04/14/2003)	Data Entry
9. [ + ]	000068BT	100001128	01/18/2007	05/31/2007	Core questionnaire (04/14/2003)	QA: Ismail Del (no pending issue)
10. [ + ]	000068BD6	000400006	12/08/2006	06/01/2007	Follow-up (04/14/2003)	Ready for QA [Claim for QA]
11. [ + ]	000068BD8	000400021	08/28/2006	06/01/2007	Follow-up (04/14/2003)	Ready for QA [Claim for QA]



Effect of equal-access to breast care on clinicopathological phenotypes of invasive breast tumors in African American

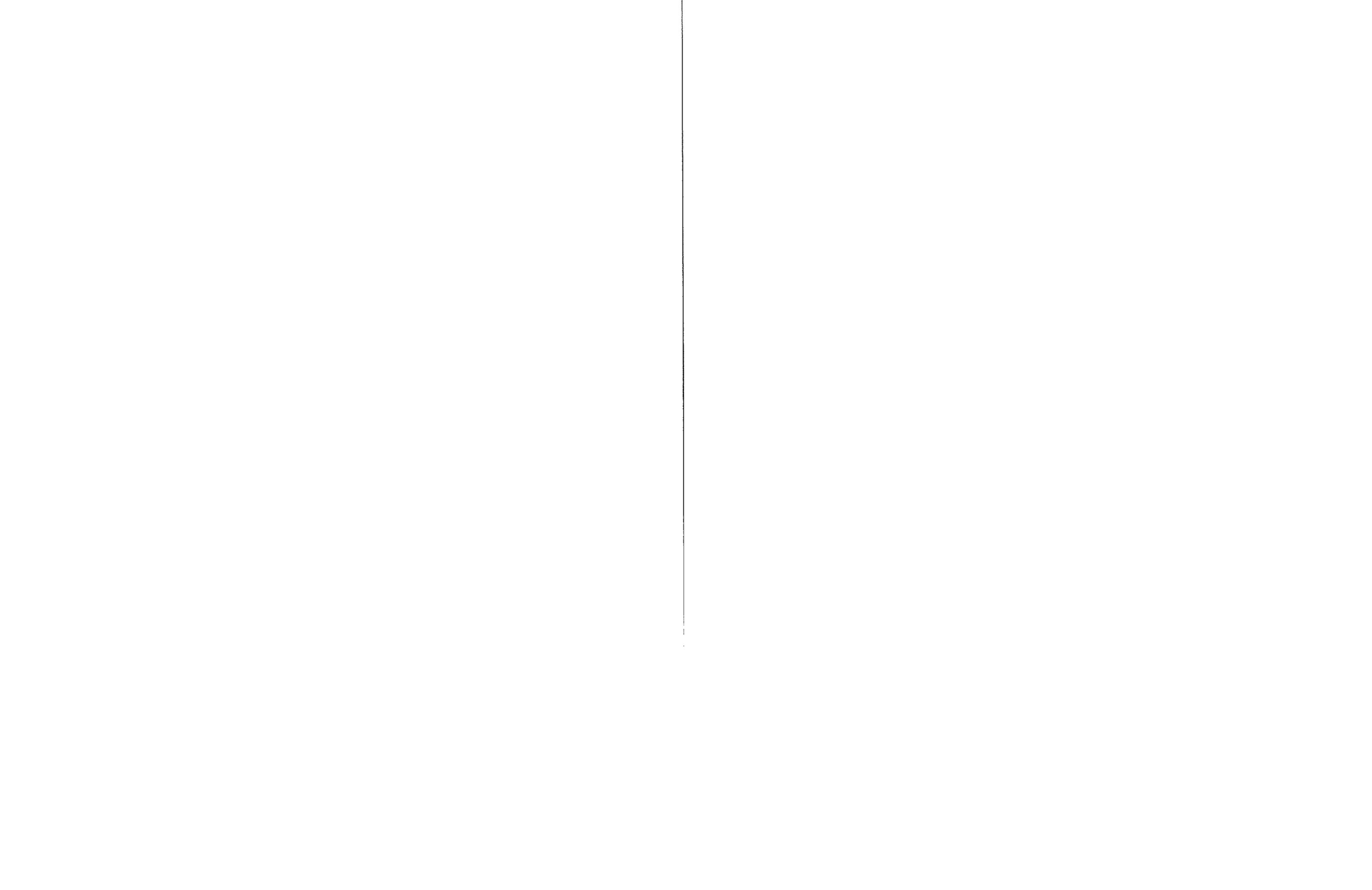
Rachel Ellsworth, Jeff Hooke, Craig Shriver

**Introduction:** Incidence of breast cancer is higher in young African American (AAW) compared to Caucasian women (CW) and is frequently associated with an aggressive phenotype and poor clinical outcome. These differences have often been attributed to differences in access to quality-health care. The Clinical Breast Care Center at Walter Reed Army Medical Center provides high-quality equal-access breast care to military personnel and their dependents. Clinical and pathological factors, including tumor stage, ER, PR and HER2 status and outcome, were examined in AAW and CW patients with invasive breast carcinomas.

**Methods:** AAW (n=76) and CW (n=253) with invasive breast cancer were enrolled in the Clinical Breast Care Project (CBCP). Clinical was collected using a large 500-field questionnaire and every specimen was reviewed by a single dedicated breast pathologist. ER, PR and HER2 protein levels were determined by IHC and/or FISH analysis. Data was analyzed patients from patients with invasive breast cancer using Student's t-test and exact unconditional tests.

**Results:** The average age of diagnosis was significantly lower ( $P<0.01$ ) in AAW (mean =51 yr) compared to CW (mean = 59 yr), with 14% of AAW diagnosed under age 40 compared to 5% of CW. Stage at diagnosis, lymph node and HER2 status did not differ between the two groups. Significantly more tumors from AAW were ER negative (33% v 21%;  $P<0.05$ ), PR negative (55% v 30%;  $P<0.0001$ ) and had the triple negative phenotype (ER, PR, and HER-2/*neu* negative; 18% v 6%;  $P<0.005$ ). The mortality rate was significantly higher in AAW (6% v 1.5%;  $P<0.05$ ) compared to CW.

**Conclusions:** A reported higher stage at diagnosis in AAW has been suggested to reflect delays in seeking medical treatment, suggesting that poor outcomes in AAW with breast cancer are associated with socioeconomic factors; however, in this population, provided with high-quality health care regardless of the ability to pay, while stage at diagnosis did not differ between AAW and CW, tumors from AAW were associated with poor pathological characteristics and higher mortality rates. Pathological and phenotypic differences between AAW and CW suggest that invasive breast tumors are different biological diseases, associated with population-specific molecular profiles.



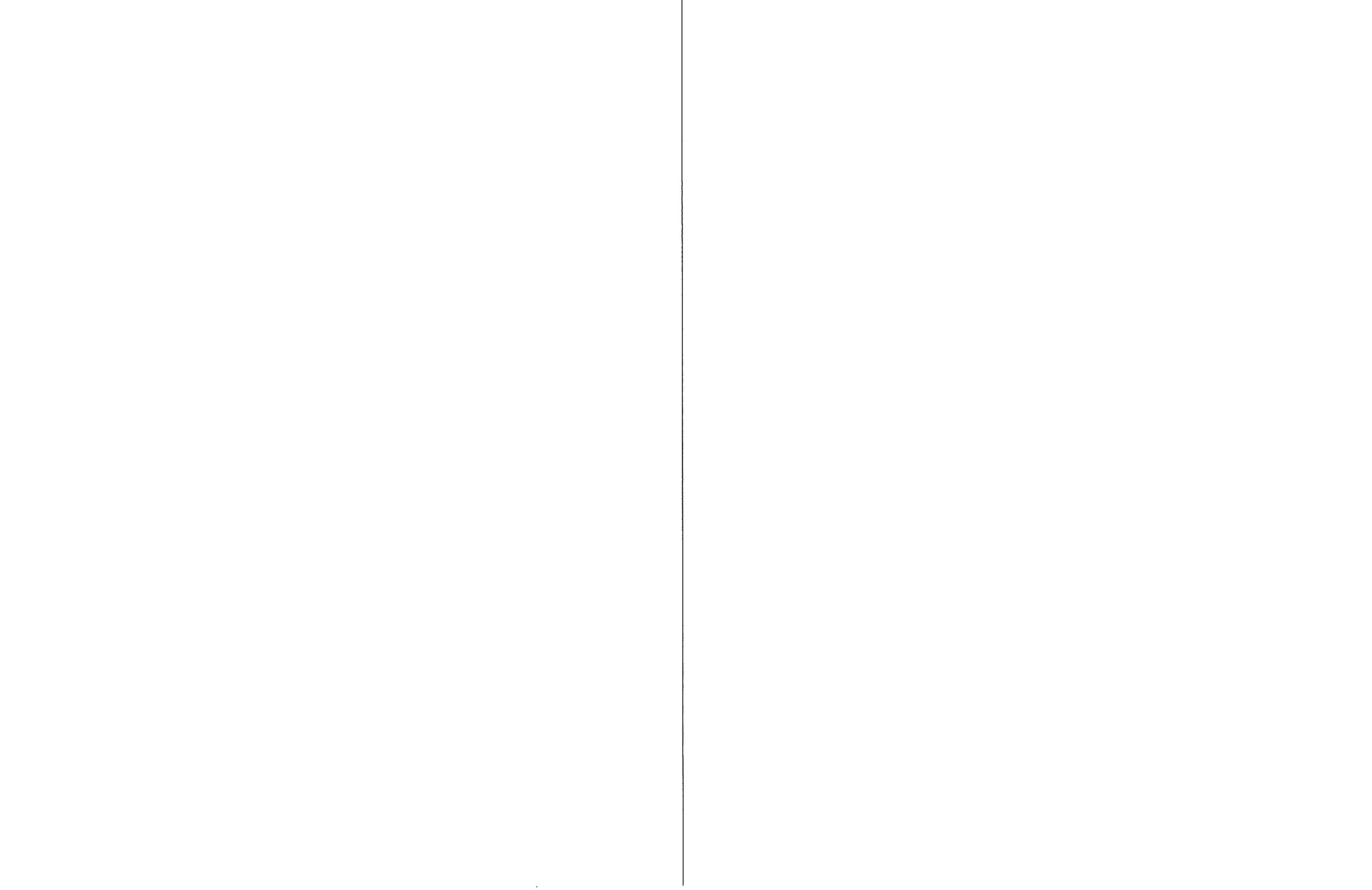
## **Gene expression differences in primary breast tumors from African American and Caucasian women.**

**Background:** Incidence of breast cancer varies across ethnic groups. Overall incidence is highest among Caucasian women (CW); however, incidence in women younger than 40 as well as mortality is highest among African American women (AAW). Although socioeconomic disparities are often blamed for these differences, AAW are more likely than CW to present with large, aggressive tumors with poor prognostic features such as estrogen receptor negativity, triple negative phenotype, and lymph node involvement, suggesting that molecular factors may also contribute to the younger age of onset and worse outcomes in AAW. In the present study, we have compared primary breast tumors from AAW and CW by gene expression profiling to identify differentially expressed genes that may explain why diagnosis at a young age, aggressive tumor biology and lower survival are more prevalent in AAW than CW with breast cancer.

**Materials and Methods:** Breast tumor specimens were obtained from 26 pairs of age and stage-matched AAW and CW enrolled in the Clinical Breast Care Project (CBCP). Tumor cells were isolated by laser microdissection (Leica) and RNA isolated using the RNAqueous Micro kit (Ambion). RNA was amplified and labeled with biotin-11-UTP using two rounds of amplification with the MessageAmp II aRNA Amplification kit (Ambion). Fragmented aRNA was hybridized to the HG U133A 2.0 array (Affymetrix). A Mann-Whitney U test was used to identify differentially expressed genes ( $p \leq 0.001$ ).

**Results:** Sixty-five genes were differentially expressed between AAW and CW. Twenty-eight genes were more highly expressed in AAW, including SOS1, TSPO, and PSPH, while 37 genes had lower expression in AAW, including ZNF228, SCAMP4, and PSD3.

**Discussion:** These results support a model in which molecular factors influence breast cancer onset, progression and prognosis in AAW. Genes with higher expression in tumors from AAW function in steroid, fatty acid, and serine biosynthesis, signal transduction, cell proliferation and adhesion, and cytoskeleton organization and biogenesis. Functions of genes with lower expression in AAW tumors include regulation of transcription, vesicular transport, iron metabolism, signal transduction, and extracellular matrix organization and biogenesis. These results suggest that breast tumors from AAW and CW differ at the gene expression level and that these differences may be responsible for the more aggressive clinical features of breast tumors in AAW compared to CW.



Identification of a gene expression breast cancer metastasis profile  
Jeff Seebach, Lori Field, Brad Love, Jeff Hooke, Rachel Ellsworth, Craig Shriver

**Background:** Lymph nodes status is currently the most important predictor of prognosis in breast cancer. Despite current knowledge about the various subtypes of breast carcinomas, little is known about genes responsible for invasion into the lymphatic system or the ability of the tumor cell to survive and multiply in the lymph node. Here, gene expression analysis was performed on primary breast and matched metastatic lymph nodes tumors to identify molecular profiles associated with metastasis.

**Methods:** OCT embedded primary breast and lymph node tumors were selected from 20 patients with positive lymph nodes. Pure tumor populations were isolated using laser capture microdissection. RNA was isolated using RNAqueous Microprotocol for LCM (Ambion) and followed by two rounds of amplification using the MessageAmp II aRNA Amplification Kit (Ambion). The cDNA product was labeled with biotin, fragmented, and hybridized to a GeneChip Human Genome U133A 2.0 array (Affymetrix). Differential expression was determined using Mann-Whitney testing; Bonferroni corrected P-values of 0.05 and 0.001 were calculated.

**Results:** 51 genes were differentially expressed between primary and metastatic lymph node tumors; of these, 20 had both a P-value  $<0.001$  and  $>4$ -fold differences in expression levels. Thirty-eight genes including C7, FOXF1 and EPHA3 had significantly lower expression in metastatic compared to primary breast tumors. Thirteen genes including TNN and TAC1 had lower expression in metastatic tumors compared to the primary tumors.

**Conclusions:** Identification of a set of genes differentially expressed between primary breast and metastatic lymph node tumors suggests that expression of a unique set of genes is required for metastatic breast tumor cells to successfully colonize foreign tissues. Genes identified here are involved in cell migration, adhesion, quiescence and signaling; these genetic changes may confer upon tumor cells the ability to disseminate, invade the lymph nodes and then enter an active growth phase. Chemotherapeutic agents targeted to these genes may be useful in retarding metastasis.



#### Genetic alterations associated with breast tumor metastasis

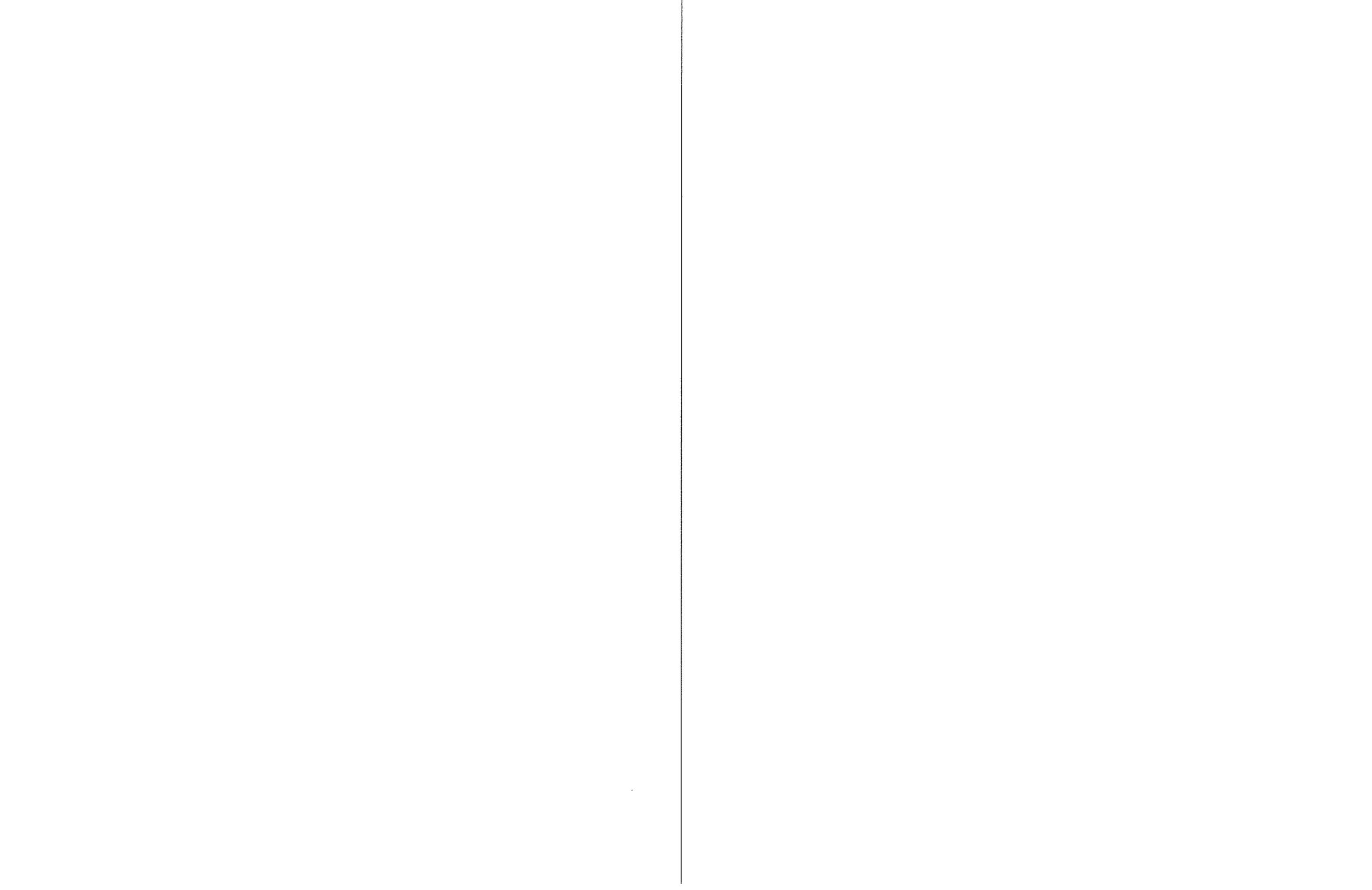
Rachel E. Ellsworth, PhD, Heather L. Patney, BS, Darrell L. Ellsworth, PhD Brad Love, PhD, Jeffrey A. Hooke, MD, Craig D. Shriver, MD

**Background:** Axillary lymph node status is one of the most important prognostic factors for patients with breast cancer. Genetic studies attempting to identify molecular profiles associated with metastasis have produced conflicting results. Here, we assayed chromosomal changes in primary breast tumors from node-negative (n=100) and node-positive (n=100) patients to determine whether specific genetic changes are associated with tumor metastasis.

**Methods:** Laser microdissected DNA samples were collected from node negative and node positive primary breast tumors. Fifty-two microsatellite markers representing 26 chromosomal regions commonly altered in breast cancer were used to assess levels and patterns of AI. Data was analyzed using exact unconditional and Student's t-tests; values of  $P < 0.05$  were considered significant.

**Results:** Overall AI frequencies between node negative (21.6%) and node positive (23.4%) breast tumors did not differ significantly ( $P = 0.1903$ ). When the data were examined by chromosomal region, chromosomes 6q25-q27 and 8q24 had significantly higher levels of AI events in the node positive primary tumor samples (34%, 27%) compared to the node negative samples (20%, 10%). No node negative samples had significantly higher levels of AI compared to node positive samples at any of the 26 chromosomal regions.

**Conclusions:** Chromosome 6q25-q27 is the site of the FRA6 fragile site, rendering this region susceptible to chromosomal breakage and loss. Alterations at FRA6 affect expression of the scaffold protein Afadin; loss of Afadin has been correlated with poor outcome in breast cancer patients. On chromosome 8q24, overexpression of the MYC has been associated with changes in bone marrow microenvironment allowing successful colonization by metastatic breast cells, as well as with lymphangiogenesis, promoting the proliferation of metastatic cells. Thus, alterations of both the 6q25-q27 and 8q24 regions may be key determinants in development of metastasis.



## The Genomic Heritage of Lymph Node Metastases: Implications for Clinical Management of Patients with Breast Cancer

Tyson E. Becker, MD,<sup>1</sup> Rachel E. Ellsworth, PhD,<sup>1,2</sup> Brenda Deyarmin, HT(ASCP),<sup>1</sup>  
Heather L. Patney, BS,<sup>1</sup> Rick M. Jordan, MS,<sup>1</sup> Jeffrey A. Hooke, MD,<sup>3</sup>  
Craig D. Shriver, MD,<sup>3</sup> and Darrell L. Ellsworth, PhD<sup>1</sup>

<sup>1</sup>Clinical Breast Care Project, Windber Research Institute, 620 Seventh Street, Windber, PA 15963, USA

<sup>2</sup>Henry M. Jackson Foundation for the Advancement of Military Medicine, 1401 Rockville Pike, Suite 600, Rockville, MD 20852, USA

<sup>3</sup>Clinical Breast Care Project, Walter Reed Army Medical Center, 6900 Georgia Avenue, NW, Washington, DC 20307, USA

**Background:** Metastatic breast cancer is an aggressive disease associated with recurrence and decreased survival. To improve outcomes and develop more effective treatment strategies for patients with breast cancer, it is important to understand the molecular mechanisms underlying metastasis.

**Methods:** We used allelic imbalance (AI) to determine the molecular heritage of primary breast tumors and corresponding metastases to the axillary lymph nodes. Paraffin-embedded samples from primary breast tumors and matched metastases ( $n = 146$ ) were collected from 26 patients with node-positive breast cancer involving multiple axillary nodes. Hierarchical clustering was used to assess overall differences in the patterns of AI, and phylogenetic analysis inferred the molecular heritage of axillary lymph node metastases.

**Results:** Overall frequencies of AI were significantly higher ( $P < 0.01$ ) in primary breast tumors (23%) than in lymph node metastases (15%), and there was a high degree of discordance in patterns of AI between primary breast carcinomas and the metastases. Metastatic tumors in the axillary nodes showed different patterns of chromosomal changes, suggesting that multiple molecular mechanisms may govern the process of metastasis in individual patients. Some metastases progressed with few genomic alterations, while others harbored many chromosomal alterations present in the primary tumor.

**Conclusions:** The extent of genomic heterogeneity in axillary lymph node metastases differs markedly among individual patients. Genomic diversity may be associated with response to adjuvant therapy, recurrence, and survival, and thus may be important in improving clinical management of breast cancer patients.

**Key Words:** Metastasis—Breast Cancer—Allelic Imbalance—Axillary Lymph Nodes.

Metastasis represents an important step in the progression of malignant breast cancer from a localized stage to an aggressive systemic disease. Although selection of adjuvant or hormonal therapy for patients

with metastatic breast cancer is guided by prognostic factors such as hormone receptor status, tumor size, pathological grade, lymphatic and/or vascular invasion, and age,<sup>1</sup> axillary lymph node status and the corresponding number of nodes involved are currently the most reliable predictors of survival. The 5-year survival rates differ dramatically between women with negative lymph nodes (>90%) and those with axillary lymph node metastasis (<70%).<sup>2</sup>

Address correspondence and reprint requests to: Darrell L. Ellsworth, PhD; E-mail: d.ellsworth@wriwindber.org

Published by Springer Science+Business Media, LLC © 2008 The Society of Surgical Oncology, Inc.

Models of the metastatic process differ with respect to the timing and extent of genomic alterations necessary for breast cancer metastasis.<sup>3</sup> At one end of the spectrum, the clonal-progression hypothesis suggests that epithelial cells sequentially accumulate genomic and epigenetic changes that eventually confer metastatic potential to a large proportion of cells in the primary tumor.<sup>4,5</sup> This model predicts that metastatic tumors would be genetically similar to the primary carcinomas from which they descended, because only a small proportion of molecular changes occurring late in disease progression would differentiate metastatic from non-metastatic cells. Recent studies have challenged this traditional hypothesis and suggest that important molecular changes confer metastatic potential to a small population of cells early in development of the primary tumor.<sup>6,7</sup> Accordingly, cells destined to form metastatic disease evolve independently, leading to greater genetic divergence between the primary tumor and the metastases.

These models of metastasis have important clinical implications for effective treatment of patients with breast cancer. As current risk allocations and treatment recommendations targeting breast cancer metastases are based largely on histological and molecular characteristics of the primary tumor,<sup>8</sup> metastases may be responsive to treatment if they are genetically similar to the primary carcinoma (clonal progression). Conversely, greater genetic differentiation between primary carcinomas and corresponding metastases may lead to metastatic disease that is refractory to current therapies. To provide new insights for defining molecular events associated with the acquisition of metastatic potential and to investigate mechanisms of metastatic dissemination, we analyzed levels and patterns of allelic imbalance (AI) in primary breast carcinomas and matched axillary lymph node metastases in women with multiple positive nodes. Our objectives were to (1) assess the degree of genomic similarity between primary breast tumors and metastatic tumors of the axillary lymph nodes, (2) determine the extent of genomic differentiation among metastatic tumors of individual patients, and (3) infer the genomic heritage of axillary lymph node metastases.

## MATERIALS AND METHODS

### Sample Collection and Processing

Paraffin-embedded samples of primary breast tumors and matched axillary lymph node metastases

( $n = 146$ ) from 26 patients with node-positive breast cancer involving multiple axillary nodes were obtained from the pathology archives of the Windber Medical Center (WMC) and from patients enrolled in the Clinical Breast Care Project (CBCP) at Walter Reed Army Medical Center (WRAMC). Approval for use of samples in this research was provided by the WMC Institutional Review Board (IRB) and the WRAMC Human Use Committee/IRB. All patients enrolled in the CBCP voluntarily agreed to participate and gave written informed consent. Clinical information regarding all CBCP samples was provided in questionnaires designed by and administered under the auspices of the CBCP. Diagnoses of all tumor samples were made after examination of hematoxylin and eosin (H&E) stained slides by the dedicated CBCP pathologist. Staging was carried out using guidelines defined in the 6<sup>th</sup> edition of the American Joint Committee on Cancer (AJCC) Cancer Staging Manual.<sup>9</sup> Clinical and pathological characteristics of all patients are shown in Table 1.

Laser microdissection of pure tumor cells was conducted on an *ASLMD* laser microdissection system (Leica Microsystems, Wetzlar, Germany) as previously described.<sup>10</sup> All slides used for microdissection were examined by the pathologist, who identified and delineated the primary tumors and metastases before dissection. The integrity of multiple serial sections was determined by pathological verification of the first and last sections stained with H&E. The number of 5- $\mu$ m-thick tissue sections required to achieve a minimum of 250 cells per genotype varied from 5 to 25 slides, depending on tumor size. A minimum cell count requirement was established to avoid PCR artifact associated with low concentrations of DNA obtained from formalin-fixed, paraffin-embedded tissues.<sup>11</sup> DNA was then extracted from microdissected tumor cells and genotyped as previously described.<sup>12</sup> Referent DNA was obtained from each patient from peripheral blood samples using Clotspin and Puregene DNA purification kits (Gentra, Minneapolis, MN) or from microdissected disease-free skin (nipple) or negative lymph node tissues.

### Allelic Imbalance

A custom panel of 52 microsatellite markers (Invitrogen, Carlsbad, CA) representing 26 chromosomal regions commonly altered in breast cancer was used to assess AI in each tumor sample. AI was defined by the formula  $(T1/T2)/(N1/N2)$ , where T1 and N1 are the smaller peak heights and T2 and N2 are

GENOMIC HERITAGE OF LYMPH NODE METASTASES

TABLE 1. Clinical and pathological characteristics of 26 patients with metastatic breast cancer

Patient no.	Age at Dx	Histology	Tumor size	No. of positive lymph nodes	No. of lymph nodes studied	TNM stage	Clinical outcome	FAI in primary tumor <sup>a</sup>	FAI in lymph node metastases <sup>b</sup>
1	65	IDCA	T2	4	4	IIB	AWD	0.56	0.20
2	77	ILCA	T3	4	4	IIIA	DOD	0.54	0.17
3	70	IDCA	T2	8	7	IIIA	DOD	0.68	0.15
4	60	IDCA	T2	2	2	IIB	NED	0.04	0.18
5	76	ILCA	T2	4	4	IIIA	NED	0.12	0.11
6	50	IDCA	T2	30	6	IV	DOD	0.28	0.18
7	30	IDCA	T3	6	5	IIIA	NED	0.32	0.16
8	48	IDCA	T3	10	5	IIIA	NED	0.04	0.03
9	73	ILCA	T2	10	2	IIIC	NED	0.12	0.02
10	48	IDCA	T2	19	3	IIIC	AWD	0.04	0.09
11	65	ILCA	T2	26	8	IIIC	AWD	0.24	0.17
12	60	ILCA	T3	4	3	IIIA	NED	0.17	0.08
13	80	Mixed	T2	4	4	IIIA	NED	0.48	0.30
14	46	IDCA	T1	7	7	IIIA	NED	0.44	0.16
15	72	ILCA	T2	3	2	IIB	AWD	0.20	0.22
16	36	IDCA	T2	18	6	IV	DOD	0.35	0.29
17	45	ILCA	T2	3	3	IIB	NED	0.04	0.05
18	47	IDCA	T2	11	6	IIIA	NED	0.12	0.21
19	38	IDCA	T2	6	4	IIIA	NED	0.04	0.17
20	84	IDCA	T2	8	6	IIIA	NED	0.20	0.07
21	47	ILCA	T2	2	2	IIB	NED	0.08	0.02
22	62	IDCA	T2	6	6	IIIA	DOD	0.13	0.14
23	56	IDCA	T3	15	8	IIIC	NED	0.12	0.04
24	63	IDCA	T1	9	6	IIIA	AWD	0.33	0.24
25	42	IDCA	T3	6	5	IIIA	AWD	0.28	0.22
26	63	IDCA	T2	3	2	IIA	NED	0.04	0.25

IDCA, invasive ductal carcinoma; ILCA, invasive lobular carcinoma; Mixed, invasive carcinoma with mixed ductal and lobular features; TNM, tumor-node-metastasis; AWD, alive with disease; DOD, dead of disease; NED, no evidence of disease; FAI, fractional allelic imbalance

<sup>a</sup> FAI calculated as the number of chromosomal regions showing AI divided by the number of informative genotypes.

<sup>b</sup> Lymph node metastases from each patient were pooled as a group.

the larger peak heights of the tumor and normal referent samples, respectively.<sup>13</sup> A threshold for determining an AI of 0.35 or less was used in this study because it has been shown to produce highly reproducible detection of chromosomal alterations in independent PCR reactions.<sup>14</sup>

Statistical Analysis

Fisher's Exact Test (2-tailed) and Student's *t*-tests were used to test for significant differences in the overall frequency of AI, and the frequency of AI at each chromosomal region, between tumor samples. Overall similarities in patterns of AI between primary tumors and corresponding metastases were assessed by hierarchical clustering using a Euclidean distance metric in Spotfire (Somerville, MA) and the Unweighted Pair Group Method with Arithmetic Mean (UPGMA) clustering algorithm.<sup>15</sup> A phylogenetic analysis to infer the molecular heritage of axillary lymph node tumors was conducted using the Phylogeny Inference Package (PHYLIP version 3.66).<sup>16</sup> These methods determined relationships among the tumor samples within each patient by defining the fewest number of molecular changes (AI events) be-

tween samples. AI was treated as a discrete character and coded as present or absent. Analysis in PHYLIP used Camin-Sokal parsimony<sup>17</sup> in the MIX algorithm—which assumes that AI events can originate independently but cannot revert to normal—and Wagner parsimony<sup>18</sup> in the PARS program. Input options included (1) equal weighting of all chromosomal regions in which AI was assessed, (2) random input order, and (3) outgroup rooting using a hypothetical (normal) tissue with no AI. When multiple equivalent (equally parsimonious) trees were identified, a consensus tree was produced by bootstrapping analysis, which generated multiple datasets (*n* = 100) that are re-sampled versions of the original dataset,<sup>19</sup> by running the SEQBOOT, MIX, and CONSENSE programs sequentially.

RESULTS

Frequencies of Allelic Imbalance

Overall frequencies of AI among all 26 patients were significantly higher (*P* < 0.01) in primary breast tumors (23%) than in lymph node metastases

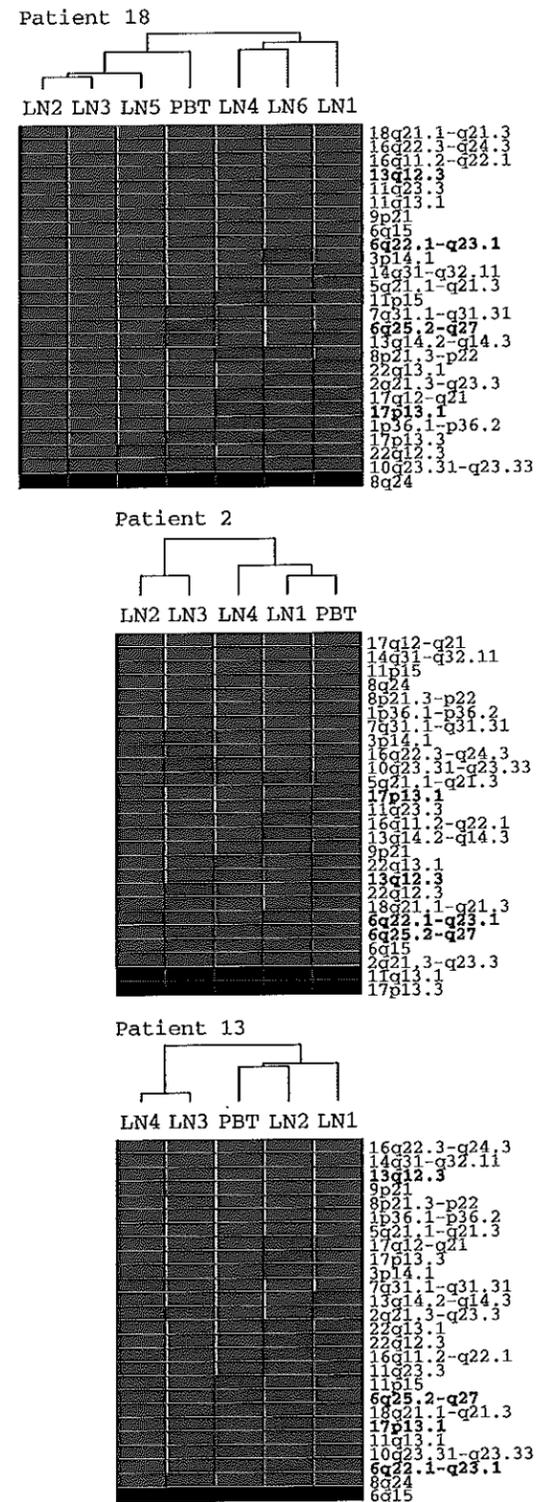


FIG. 1. Patterns of similarity in allelic imbalance (AI) determined by hierarchical cluster analysis among primary breast tumors (PBT) and axillary lymph node (LN) metastases from selected patients. Green normal, red AI, black uninformative. Chromosomal regions in bold show a significantly higher frequency of AI events in primary tumors than in corresponding metastases.

(15%). Chromosomal regions 6q22.1-q23.1 ( $P < 0.05$ ), 6q25.2-q27 ( $P < 0.05$ ), 13q12.3 ( $P < 0.01$ ), and 17p13.1 ( $P < 0.05$ ) showed a significantly higher frequency of AI events in primary tumors than in corresponding metastases; however, no region had significantly more AI events in lymph node metastases. When AI frequencies between primary tumors and lymph node metastases were examined in individual patients, 5 of the 26 had significantly higher levels in their primary tumors than in their pooled lymph node metastases ( $P < 0.01$ ), while 1 showed the opposite ( $P < 0.05$ ). Within a given patient, a high level of discordance was observed in the frequency of specific AI events between primary and metastatic carcinomas (data not shown).

**Hierarchical Clustering**

Hierarchical cluster analysis identified distinct groupings of tumors within patients based on overall similarity in patterns of genomic alterations. Usually one cluster contained the primary breast tumor and several metastases that were genetically similar to the primary carcinoma, while other clusters were composed exclusively of lymph node metastases, which were genetically different from the primary tumor and, in some patients, from each other. Importantly, some lymph node metastases possessed molecular changes not present in the primary tumor, while others had no detectable chromosomal alterations (Fig. 1).

**Phylogenetic Analysis**

The tree with the fewest AI events generated by the phylogenetic analysis for each patient grouped the tumors into multiple discrete lineages (Fig. 2). Trees generated by different algorithms were similar in overall structure. Several metastases shared a common ancestry with the primary tumor more recently than other metastases that appeared to diverge from the primary carcinoma early in disease progression. Within each patient, branch lengths were not equivalent for all tumor lineages, suggesting that rates of chromosomal changes varied among primary tumors and lymph node metastases (Fig. 3). In general,

GENOMIC HERITAGE OF LYMPH NODE METASTASES

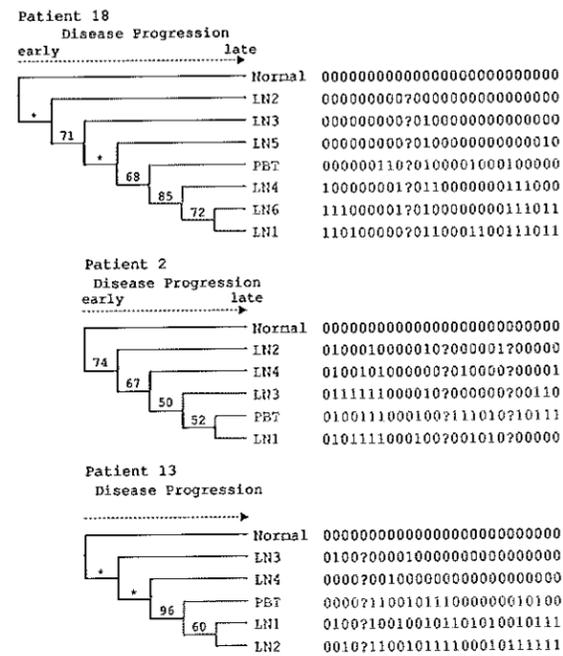


FIG. 2. Consensus trees based on allelic imbalance (AI) depicting relatedness among primary breast tumors (PBT) and axillary lymph node (LN) metastases from selected patients. The percentage of bootstrap replications supporting each lineage is indicated along the respective branches. Branches appearing in less than 50% of the replications are denoted by asterisks. Trees were rooted by a hypothetical tissue type—normal breast epithelium with no AI. The data matrix beside each tree contains data for that patient, coded as AI (1) or normal (0). Question marks denote regions where AI could not be determined because markers were uninformative (homozygous). Blue PBT, green metastases hypothesized to have diverged early in tumorigenesis, red metastases believed to have evolved late in tumorigenesis.

metastases that originated late in tumorigenesis appeared to exhibit higher rates of chromosomal alterations than those that had evolved early.

DISCUSSION

Overall patterns and frequencies of AI in this study were discordant between primary breast carcinomas and matched metastases in the axillary lymph nodes and, importantly, differed among separate axillary lymph node metastases of individual patients. Phylogenetic and hierarchical cluster analyses demonstrated that multiple genetically divergent lineages of metastatic cells independently colonize the axillary lymph nodes. Some lymph node metastases appeared to acquire metastatic potential early in tumorigenesis and successfully formed metastases with relatively few

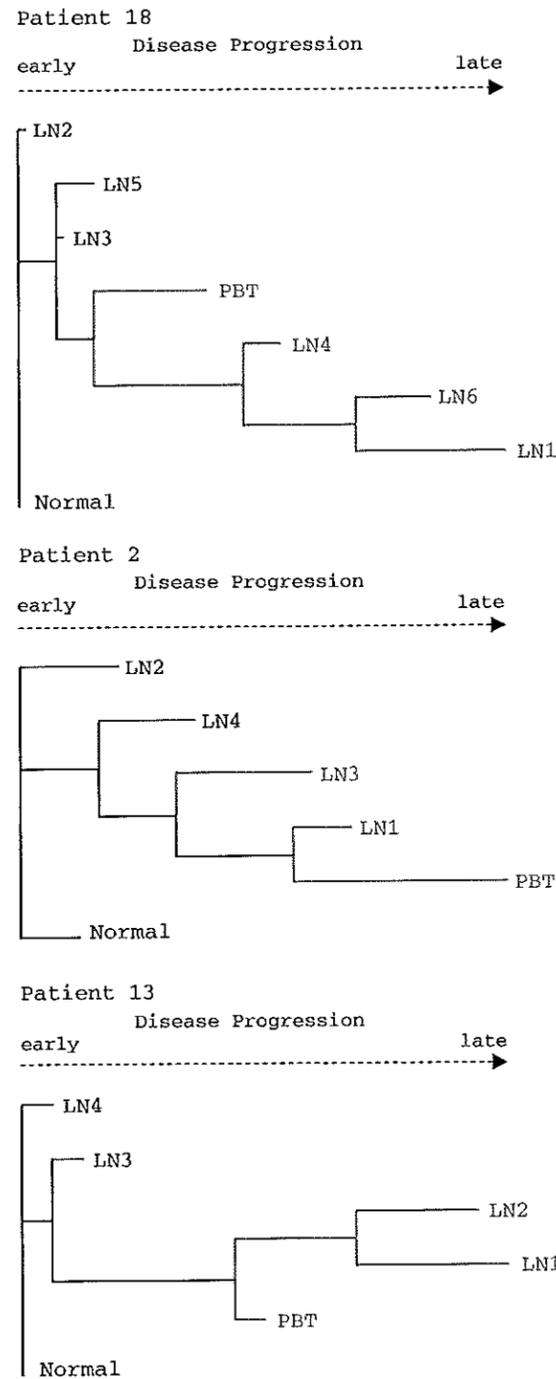


FIG. 3. Trees generated using the PARS algorithm depicting genomic heritage of primary breast tumors (PBT) and lymph node (LN) metastases. Branch lengths are proportional to the number of AI events in each tumor. Blue PBT, green LN metastases that were genetically divergent from the PBT, red LN metastases that were genetically similar to the PBT.

genomic alterations, while others seemed to have evolved later and thus harbored many chromosomal alterations that are also present in the primary tumor. These observations suggest that several molecular mechanisms may govern the process of breast cancer metastasis in any individual patient. Lymph node metastases that were genetically similar to the primary tumor may have been derived through clonal progression mechanisms, while those that were genetically divergent from the primary tumor may have evolved via parallel-and-independent evolution.<sup>20</sup>

The desire to improve clinical management of breast cancer coupled with observations of genomic variability between primary and metastatic carcinomas has led to numerous hypotheses attempting to model the metastatic process, which differ dramatically in their prediction of (1) timing of metastatic dissemination, (2) molecular mechanisms driving metastatic spread, and (3) number of progenitor cells with sufficient metastatic capacity to successfully colonize distant tissues.<sup>3</sup> Traditional models have suggested that somatic mutations accumulate over time and eventually confer metastatic capacity to small populations of tumor cells late in tumorigenesis.<sup>4,5,21,22</sup> Conversely, contemporary views of metastasis derived from genomic<sup>7,14,23</sup> and gene expression studies<sup>24,25</sup> propose that metastatic capacity is an inherent feature of breast tumors and that the propensity for metastatic development is acquired early in disease progression by subpopulations of cells, possibly cancer stem cells,<sup>26</sup> that evolve independently in the primary tumor and the metastases. These cells may disseminate in an undifferentiated state, but have the capacity to proliferate and efficiently form new tumors.

In our study, the presence of divergent lineages of axillary lymph node metastases in individual patients indicates the existence of several genomically distinct populations of cancer progenitor cells that may have originated or acquired metastatic potential at different times during primary tumorigenesis. These genetically diverse progenitor cells may give rise to breast cancer metastases that evolve with variable rates of genomic change and progress with vastly different chromosomal alterations. Because growth rates of metastases have been shown to be very homogeneous,<sup>27</sup> we believe patterns of genetic differentiation among metastatic tumors reflect variation in the timing of origin and are not simply a consequence of different rates of cellular proliferation among lymph node metastases.

The observation that tumor lineages originating at different times during disease progression were

growing concurrently in the axillary nodes of individual patients suggests that there may be a period of latency at some point in the dissemination process for progenitor cells that acquired metastatic potential early in tumorigenesis. In current models of disease recurrence and survival,<sup>28</sup> the biological behavior of breast cancer metastases is consistent with tumor latency in specific micrometastatic phases—as single cells and avascular foci.<sup>29</sup> Populations of latent tumor cells have been shown to remain clinically undetectable within host tissues for prolonged periods of time.<sup>30</sup> Metastatic growth of early disseminating tumor cells may therefore be arrested after initial colonization of the lymph nodes. In later stages of disease progression, metastatic growth may be induced by withdrawal of angiogenesis inhibitors,<sup>31</sup> release of growth-stimulating factors from the primary tumor,<sup>32</sup> or other mechanisms, which may permit the simultaneous proliferation of tumor progenitor cells that disseminated at different times during tumorigenesis.

Our findings suggest that there is extensive molecular heterogeneity among the axillary nodal metastases derived from a primary breast carcinoma; however, this study did not distinguish sentinel from non-sentinel axillary nodes. The sentinel lymph node (SLN) hypothesis postulates that cancer cells initially spread from the primary tumor to the first-draining node (the sentinel node) before reaching higher echelon nodes.<sup>33</sup> The sentinel node thus serves as a physical repository for the earliest tumor cells that are shed from the primary tumor and arrive in a nodal basin. These progenitor cells may have the capacity to establish a metastatic deposit or may remain viable but non-proliferative as isolated tumor cells. Because lymphatic drainage in the breast is based on an intricate system of branching vessels, genomically different cancer progenitor cells may colonize different lymph nodes, but development of metastases appears to occur in a sequential manner that maintains the known capability of SLN biopsy to accurately stage the axilla in breast cancer patients.<sup>34</sup> Further research is needed to assess the molecular heritage of sentinel node metastases, the influence of molecular heterogeneity in the primary tumor on genomic diversity among metastases, and mechanisms by which genomically distinct progenitor cells metastasize in a predictable manner to sentinel and then non-sentinel lymph nodes.

In conclusion, our results demonstrate the molecular variability of lymphatic breast cancer metastases and may help us understand why lymph node-positive breast cancer has variable response rates to

present therapies and often responds only transiently to available treatment options.<sup>35</sup> Although the risk of systemic metastasis has been shown to decrease significantly in patients receiving chest wall radiation and adjuvant tamoxifen after mastectomy,<sup>36,37</sup> rates of recurrence remain high (45–47%) because these treatments fail to effectively target all metastatic cells. As many patients with metastatic breast cancer do not respond to, or eventually escape, current treatments, both *de novo* and acquired mechanisms of therapeutic resistance may occur in individual patients. Genetic heterogeneity between metastatic carcinomas and the primary tumor from which they descended<sup>38,39</sup> may contribute to variability in outcomes among patients with metastatic disease.

Axillary lymph node involvement is a surrogate marker for distant metastases and may reflect the extent of genomic diversity in distant metastases. Accordingly, metastatic lineages that differ genomically from the primary tumor may be refractory to treatments based on characteristics of the primary tumor. Similarly, rapidly evolving metastatic lineages that exhibit high rates of chromosomal change may be more prone to escape treatment over time. Further studies are needed to determine whether genomic variability in lymph node or distant metastases is associated with different outcomes.

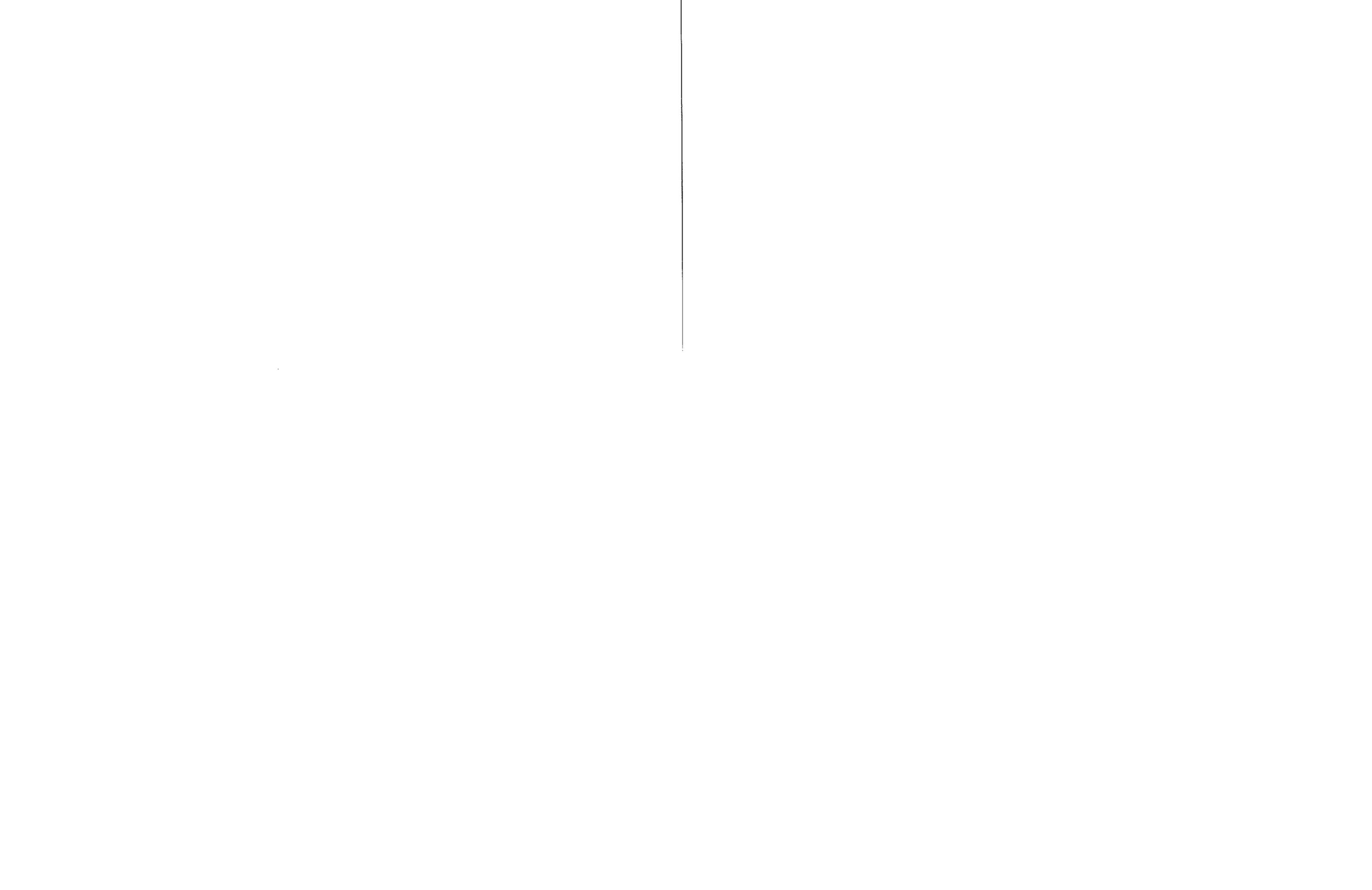
#### ACKNOWLEDGEMENTS

Presented in part at the Society of Surgical Oncology's 61<sup>st</sup> Annual Cancer Symposium, March 13–16, 2008, in Chicago, IL. This work was performed under the auspices of the Clinical Breast Care Project, a joint effort of many investigators and staff members whose contributions are gratefully acknowledged. We especially thank the program participants. Supported by the United States Department of Defense (Military Molecular Medicine Initiative MDA W81XWH-05-2-0075). The opinion and assertions contained herein are the private views of the authors and are not to be construed as official or as representing the views of the Department of the Army or the Department of Defense.

#### REFERENCES

- Eifel P, Axelson JA, Costa J, et al. National Institutes of Health consensus development conference statement: adjuvant therapy for breast cancer, November 1-3, 2000. *J Natl Cancer Inst* 2001; 93:979–89.
- Ries LAG, Melbert D, Krapcho M, et al. (eds). (2007) SEER Cancer Statistics Review, 1975–2004. National Cancer Institute, Bethesda, MD. Available: [http://www.seer.cancer.gov/csr/1975\\_2004](http://www.seer.cancer.gov/csr/1975_2004).
- Weigelt B, Peterse JL, van't Veer LJ. Breast cancer metastasis: markers and models. *Nat Rev Cancer* 2005; 5:591–602.
- Fidler IJ, Kripke ML. Metastasis results from preexisting variant cells within a malignant tumor. *Science* 1977; 197:893–5.
- Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. *Cell* 1990; 61:759–67.
- Bernards R, Weinberg RA. Metastasis genes: a progression puzzle. *Nature* 2002; 418:823.
- Schmidt-Kittler O, Ragg T, Daskalakis A, et al. From latent disseminated cells to overt metastasis: genetic analysis of systemic breast cancer progression. *Proc Natl Acad Sci U S A* 2003; 100:7737–42.
- Goldhirsch A, Glick JH, Gelber RD, et al. Meeting highlights: international expert consensus on the primary therapy of early breast cancer 2005. *Ann Oncol* 2005; 16:1569–83.
- Greene FL, Page DL, Fleming ID, et al. (eds). AJCC Cancer Staging Manual, Sixth Edition. Springer-Verlag, New York, 2002.
- Ellsworth DL, Shriver CD, Ellsworth RE, et al. Laser capture microdissection of paraffin-embedded tissues. *BioTechniques* 2003; 34:42–6.
- Bertheau P, Plassa LF, Lerebours F, et al. Allelic loss detection in inflammatory breast cancer: improvement with laser microdissection. *Lab Invest* 2001; 81:1397–402.
- Ellsworth RE, Ellsworth DL, Lubert SM, et al. High-throughput loss of heterozygosity mapping in 26 commonly deleted regions in breast cancer. *Cancer Epidemiol Biomark Prev* 2003; 12:915–9.
- Medintz IL, Lee C-CR, Wong WW, et al. Loss of heterozygosity assay for molecular detection of cancer using energy-transfer primers and capillary array electrophoresis. *Genome Res* 2000; 10:1211–8.
- Ellsworth RE, Ellsworth DL, Neatrour DM, et al. Allelic imbalance in primary breast carcinomas and metastatic tumors of the axillary lymph nodes. *Mol Cancer Res* 2005; 3:71–7.
- Sneath PHA, Sokal RR. Numerical taxonomy. W.H. Freeman, San Francisco, 1973.
- Felsenstein J. PHYLIP: phylogeny inference package (version 3.63). Cladistics 1989; 5:164–6. Available: <http://www.evolution.genetics.washington.edu/phylip.html>.
- Camin JH, Sokal RR. A method for deducing branching sequences in phylogeny. *Evolution* 1965; 19:311–26.
- Kluge AG, Farris JS. Quantitative phyletics and the evolution of Anurans. *Syst Zool* 1969; 18:1–32.
- Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 1985; 39:783–91.
- Gray JW. Evidence emerges for early metastasis and parallel evolution of primary and metastatic tumors. *Cancer Cell* 2003; 4:4–6.
- Nishizaki T, DeVries S, Chew K, et al. Genetic alterations in primary breast cancers and their metastases: direct comparison using modified comparative genomic hybridization. *Genes Chromosomes Cancer* 1997; 19:267–72.
- HAMPL M, Hampl JA, Reiss G, et al. Loss of heterozygosity accumulation in primary breast carcinomas and additionally in corresponding distant metastases is associated with poor outcome. *Clin Cancer Res* 1999; 5:1417–25.
- Kuukasjärvi T, Karhu R, Tanner M, et al. Genetic heterogeneity and clonal evolution underlying development of asynchronous metastasis in human breast cancer. *Cancer Res* 1997; 57:1597–604.
- van't Veer LJ, Dai H, van de Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002; 415:530–6.
- Lähdesmäki H, Hao X, Sun B, et al. Distinguishing key biological pathways between primary breast cancers and their lymph node metastases by gene function-based clustering analysis. *Int J Oncol* 2004; 24:1589–96.

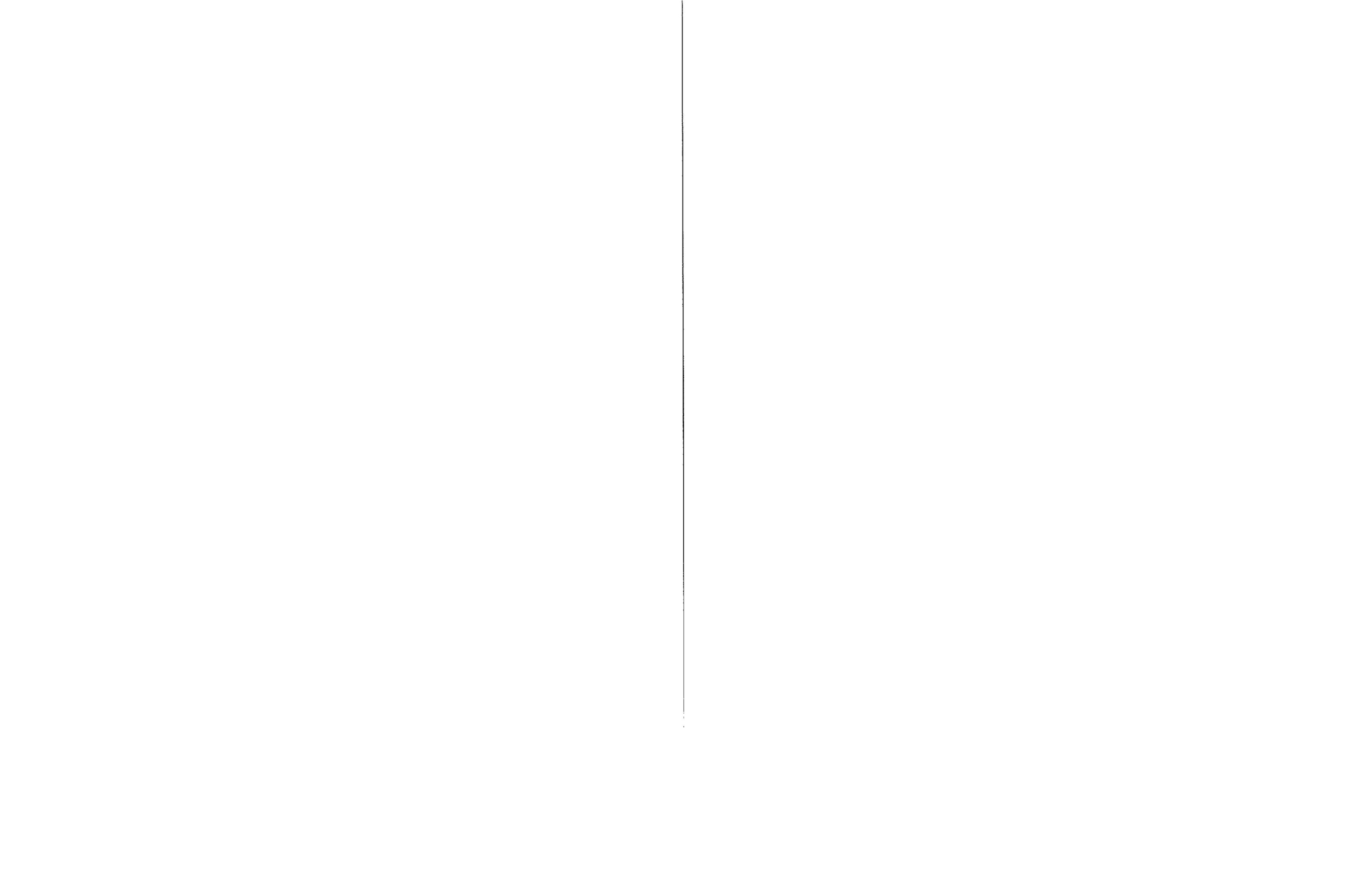
26. Al-Hajj M, Wicha MS, Benito-Hernandez A, et al. Prospective identification of tumorigenic breast cancer cells. *Proc Natl Acad Sci U S A* 2003; 100:3983-8.
27. Holmgren L, O'Reilly MS, Folkman J. Dormancy of micrometastases: balanced proliferation and apoptosis in the presence of angiogenesis suppression. *Nat Med* 1995; 1:149-53.
28. Klein CA, Hölzel D. Systemic cancer progression and tumor dormancy: mathematical models meet single cell genomics. *Cell Cycle* 2006; 5:1788-98.
29. Demicheli R, Miceli R, Moliterni A, et al. Breast cancer recurrence dynamics following adjuvant CMF is consistent with tumor dormancy and mastectomy-driven acceleration of the metastatic process. *Ann Oncol* 2005; 16:1449-57.
30. Luzzi KJ, MacDonald IC, Schmidt EE, et al. Multistep nature of metastatic inefficiency: dormancy of solitary cells after successful extravasation and limited survival of early micrometastases. *Am J Pathol* 1998; 153:865-73.
31. Guba M, Cernaianu G, Koehl G, et al. A primary tumor promotes dormancy of solitary tumor cells before inhibiting angiogenesis. *Cancer Res* 2001; 61:5575-9.
32. Fisher B, Gunduz N, Coyle J, et al. Presence of a growth-stimulating factor in serum following primary tumor removal in mice. *Cancer Res* 1989; 49:1996-2001.
33. Benson JR, della Rovere GQ Axilla Management Consensus Group. Management of the axilla in women with breast cancer. *Lancet Oncol* 2007; 8:331-48.
34. Giuliano AE, Dale PS, Turner RR, et al. Improved axillary staging of breast cancer with sentinel lymphadenectomy. *Ann Surg* 1995; 222:394-401.
35. Stockler M, Wilcken NRC, Gherzi D, et al. Systematic reviews of chemotherapy and endocrine therapy in metastatic breast cancer. *Cancer Treat Rev* 2000; 26:151-68.
36. Ragaz J, Jackson SM, Le N, et al. Adjuvant radiotherapy and chemotherapy in node-positive premenopausal women with breast cancer. *N Engl J Med* 1997; 337:956-62.
37. Overgaard M, Jensen M-B, Overgaard J, et al. Postoperative radiotherapy in high-risk postmenopausal breast-cancer patients given adjuvant tamoxifen: Danish Breast Cancer Cooperative Group DBCG 82c randomised trial. *Lancet* 1999; 353:1641-8.
38. Cavalli LR, Urban CA, Dai D, et al. Genetic and epigenetic alterations in sentinel lymph nodes metastatic lesions compared to their corresponding primary breast tumors. *Cancer Genet Cytogenet* 2003; 146:33-40.
39. Goodison S, Viars C, Urquidí V. Molecular cytogenetic analysis of a human breast metastasis model: identification of phenotype-specific chromosomal rearrangements. *Cancer Genet Cytogenet* 2005; 156:37-48.



## The prognostic role of cell cycle regulators in invasive breast cancer

Somiari, SB<sup>1</sup>, Kovatich, AJ<sup>2</sup>, Baran PN<sup>2</sup>, Zhang Y<sup>1</sup>, Deyarmin B<sup>1</sup>, Hooke J<sup>3</sup>, Shriver CD<sup>3</sup>.  
<sup>1</sup>CBCP/WRI, Windber, PA, <sup>2</sup>MDR Global Systems, Windber, PA, <sup>3</sup>CBCP/WRAMC, Washington DC.

The regulation of the cell cycle depends on a complex association between different molecules such as hormone receptors, proliferation/differentiation markers, oncogenes and tumor suppressors. Since these biological markers are often co-regulated, combined evaluation of their protein expression patterns may provide more accurate information about disease classification and prognosis than when investigated individually. The aim of our study was to determine the combination of markers that would more accurately classify and therefore provide correct prognosis for breast cancer. We studied the expression pattern of the steroid hormone receptors (ER/PR), the tyrosine kinase receptor family members, ERBB1 (EGFR) and ERBB2, together with cyclin D, c-MYC, TopoII alpha and p27. We utilized immunohistochemistry (IHC) for the protein expression and quantitative scoring of IHC was done by the Automated Cellular Imaging System (ACIS). A total of thirty six formalin fixed paraffin embedded breast cancer tumors were analyzed. This consisted of stages I (19%), II (47%), III (19%) and IV (15%) tumors which were either well (28%), moderately (25%) or poorly (47%) differentiated (tumor grade). At least 2-3 relevant tissue sections of each tumor were scored and the mean score obtained. Data analysis of the quantitative ACIS scores showed that the expression of EGFR ( $p = 0.037$ ), ER ( $p = 0.022$ ), PR ( $p = 0.029$ ) and p27 ( $p = 0.025$ ) were significantly associated with tumor stage. Also EGFR ( $p < 0.001$ ), TopoII alpha ( $p = 0.014$ ), ER ( $< 0.001$ ), PR ( $p = 0.001$ ), p27 ( $p = 0.021$ ) and cyclin D ( $p = 0.008$ ) expression were associated with tumor grade. We followed this analysis with Dunnett's T3 test for post-hoc comparison due to unequal variance and observed that only c-MYC was associated with tumor stage and stage I tumors had higher expression of c-MYC than stage IV ( $p = 0.039$ ). Further analysis based on the Dunnett's T3 test indicated that EGFR expression was significantly higher in the well differentiated than the moderately ( $p < 0.001$ ) and poorly differentiated tumors ( $p = 0.001$ ). Well differentiated tumors also had higher expression of TopoII alpha compared to moderately differentiated tumors ( $p = 0.003$ ). However, the expression of ER and p27 was lower in well differentiated tumors compared to moderate tumors ( $p = 0.013$  and  $0.53$  respectively) and ER expression was equally low in well differentiated tumors compared to poorly differentiated tumors ( $p = 0.031$ ). Our current data which relates to protein expression association with tumor stage was 100% concordant with manually scored IHC. We did not observe any significant associations of protein expression of TopoII alpha and cyclin D with tumor grade for manually scored IHC. Our studies provide evidence on the potential role of the combined analysis of some cell regulating molecules for the accurate classification of breast cancer, the concordance of the ACIS and manual scoring of IHC and the added advantage of the ACIS as a powerful tool capable of revealing additional information from IHC. Interestingly, ERBB2 did not show any association with either tumor stage or grade.



## Abstract

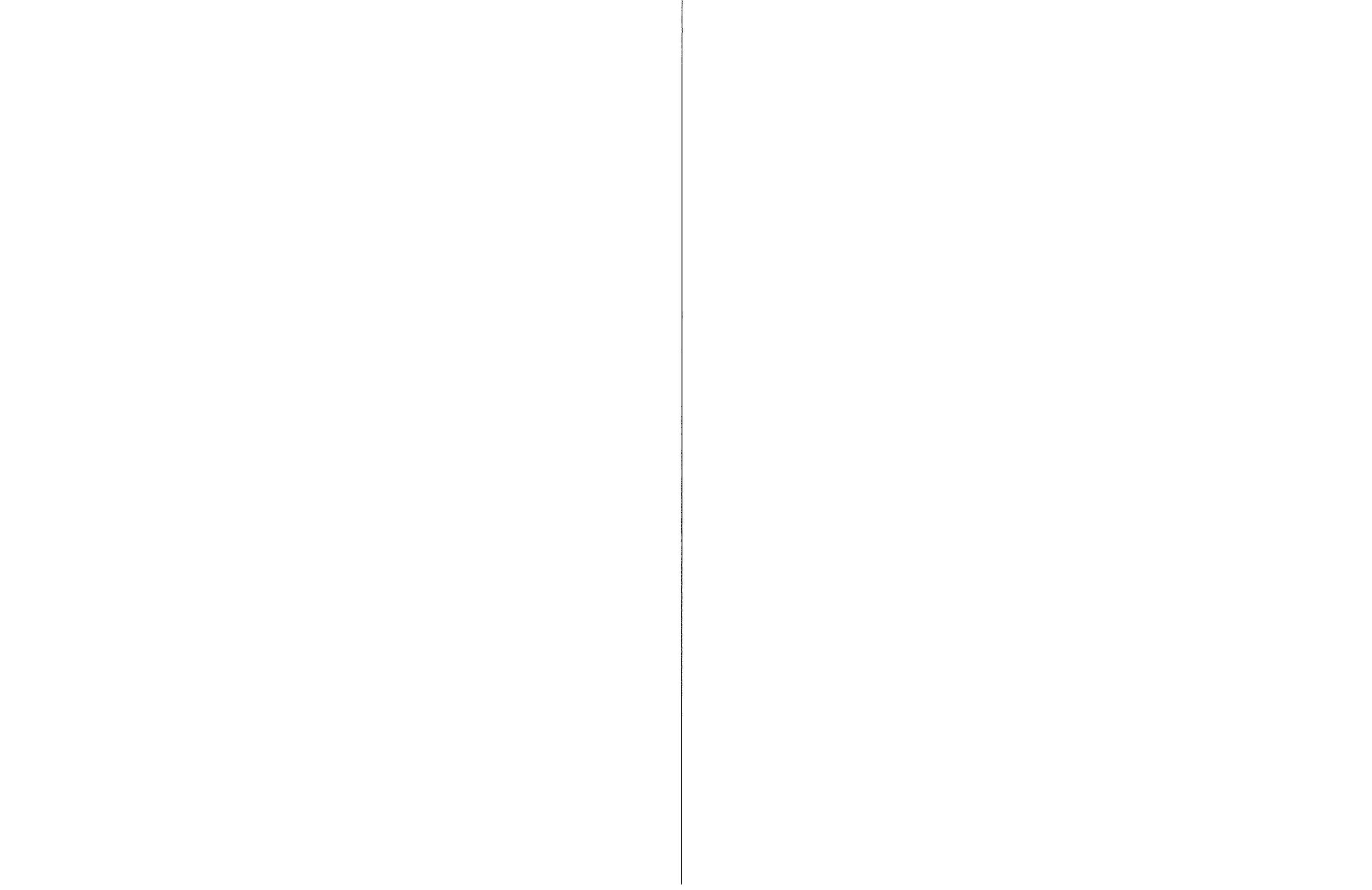
### **RNA quality of human breast tissue after surgical removal-effect of processing method, temperature and storage.**

Shoemaker A<sup>1</sup>, Heckman C<sup>1</sup>, Zhang Y<sup>1</sup>, Colledge A<sup>1</sup>, Deyarmin B<sup>1</sup>, Kane J<sup>1</sup>, Hooke J<sup>2</sup>, Shriver C<sup>2</sup>, Mural R<sup>2</sup>, Somiari S<sup>1</sup>.

Clinical Breast Care Project, <sup>1</sup>Windber Research Institute, 620 Seventh Street, Windber, PA 15963, USA.

<sup>2</sup>Walter Reed Army Medical Center, 6900 Georgia Ave., NW Washington DC 20307, USA.

The usefulness of human surgical tissue is directly related to the quality of the nucleic acid obtained from the tissue. Temperature conditions and length of storage can affect RNA quality. In this study we determined effect of processing/method of preservation, temperature and storage time on RNA integrity. Sixty-five breast specimens were kept for 1-6 hours at room temperature before embedding in OCT or flash freezing (FF) in liquid nitrogen. These were then stored for 20-170 days at -80°C before RNA processing. Shavings were obtained from OCT samples and RNA Integrity Number (RIN) determined for both OCT and FF samples using the Agilent Bioanalyzer. Average RIN was 8.25 for the OCT and 8.16 for the FF samples. RIN was not affected by either processing method ( $p=0.48$ ) or time at room temperature ( $p=0.18$ ). Additionally, we compared RIN of 89 breast specimens, stored as FF or OCT for longer periods of up to 5 years. The OCT embedded samples were laser micro dissected (LMD) before RNA processing. RIN was significantly different between these samples ( $p<0.001$ ) indicating possible effect of the LMD manipulation. No significant difference was observed between storage time of <30 days with 30-100 days ( $p=0.21$ ), 100-300 days ( $p=0.54$ ), or >300 days ( $p=0.23$ ). LMD samples were excluded from this comparison. Computing reliable metrics of RNA integrity is highly valuable. RIN can be utilized for quality assurance and incorporated into research reports to facilitate comparison of RNA-based bioassays across laboratories. Future studies will determine the effect of RIN on downstream research processes.



**Gene expression differences in normal breast tissue from disease-free African American and Caucasian women.**

Lori A. Field<sup>1\*</sup>, Brad Love<sup>2</sup>, Jennifer Kane<sup>1</sup>, Brenda Deyarmin<sup>1</sup>, Jeffrey A. Hooke<sup>3</sup>, Rachel E. Ellsworth<sup>4</sup>, and Craig D. Shriver<sup>3</sup>

<sup>1</sup>Clinical Breast Care Project, Windber Research Institute, Windber, PA, 15963, USA

<sup>2</sup>Invitrogen, Carlsbad, CA, 92008, USA

<sup>3</sup>Clinical Breast Care Project, Walter Reed Army Medical Center, Washington, DC, 20307, USA

<sup>4</sup>Clinical Breast Care Project, Henry M. Jackson Foundation for the Advancement of Military Medicine, Windber, PA, 15963, USA

\*Corresponding author

LAF: [l.field@wriwindber.org](mailto:l.field@wriwindber.org)

BL: [drblove@yahoo.com](mailto:drblove@yahoo.com)

JK: [j.kane@wriwindber.org](mailto:j.kane@wriwindber.org)

BD: [b.deyarmin@wriwindber.org](mailto:b.deyarmin@wriwindber.org)

JAH: [jeffrey.hooke1@us.army.mil](mailto:jeffrey.hooke1@us.army.mil)

REE: [r.ellsworth@wriwindber.org](mailto:r.ellsworth@wriwindber.org)

CDS: [craig.shriver@us.army.mil](mailto:craig.shriver@us.army.mil)

## **Abstract**

**Introduction:** The incidence of breast cancer in young women (<40 years of age) is higher in African American women than in Caucasian women. Tumors from African American women are also more often characterized by an aggressive phenotype, including high grade, ER- status, and BRCA1-like features, suggesting that differences in tumor behavior and outcome between African American and Caucasian women have a molecular component. How and why these differences arise is unknown.

**Methods:** Gene expression profiles from normal breast tissue from disease-free African American (n=25) and Caucasian women (n=22) were compared using Affymetrix GeneChip HG U133A 2.0 arrays. Significantly differentially expressed genes ( $P<0.005$ ) were identified using a Mann-Whitney U test with non-parametric assumptions. Differential expression of selected genes was verified with quantitative real-time PCR using TaqMan Gene Expression assays.

**Results:** Forty-nine genes were differentially expressed between African American and Caucasian women ( $P<0.005$ ) with 32 genes having higher expression and 17 genes having lower expression in African American women when compared to Caucasian women. These genes function in signal transduction, metabolism, immune response, transport, and transcription. Further validation by quantitative real-time PCR confirmed that both *ACSM1* and *PSPHL* were more highly expressed in normal breast tissue from African American women while expression of *NTRK3* was higher in Caucasian women.

**Conclusions:** These results show that normal breast tissue from African American and Caucasian women have different gene expression profiles, suggesting that breast tissue is biologically different between the two groups. These inherent differences in gene

expression may affect cellular growth and proliferation, making African American women more prone to developing aggressive tumors at an earlier age than Caucasian women.

## Introduction

Each year in the United States, over 200,000 women are diagnosed with invasive breast cancer and an estimated 40,000 will die of this disease. However, not all ethnic groups are impacted similarly. The highest overall incidence occurs among Caucasian women (CW) with 132.5 breast cancer cases per 100,000 women while incidence among African American women (AAW) is lower with 118.3 cases per 100,000 women. When stratified by age, however, the highest incidence among women under age 40 occurs in African Americans. Mortality rates from breast cancer are also higher in AAW than CW with 33.8 deaths per 100,000 AAW versus 25.0 deaths per 100,000 CW [1].

These differences in survival between CW and AAW with breast cancer have been attributed to disparities in socioeconomic status, healthcare access, and cultural beliefs. However, several studies that controlled for confounding factors such as income, education, patient age, tumor stage at diagnosis, and treatment regimen found that mortality among AAW remained higher than that of CW [2,3]. Moreover, AAW continued to have poorer prognoses and survival within equal-access healthcare systems [4,5].

In addition to differences in clinical outcomes, tumor characteristics also vary between AAW and CW. Tumors from AAW tend to be larger and more aggressive than those in CW [3,6,7], and they often share many features with tumors from women with inherited mutations in BRCA1 [8], including poor differentiation, pushing margins, high-grade nuclear atypia, high mitotic index, and lymphocytic infiltrate [3,6,7,9,10,11]. Tumors from AAW are also more likely to have a “triple-negative” phenotype (estrogen

receptor (ER), progesterone receptor (PR), and HER2 negative) and are therefore, more difficult to treat because hormonal and targeted therapies are ineffective [7,12].

Together, these differences in outcome and pathology suggest that molecular factors contribute to differences in tumor etiology between AAW and CW. Several studies comparing breast tumors from AAW and CW have identified molecular differences. For example, the cell cycle regulatory proteins p53, p16, and cyclin E were all found to have higher expression while cyclin D1 had lower expression in AAW compared to CW [6,11]. In addition, researchers have recently reported that breast tumors from AAW and CW have different gene expression profiles [13,14]. While these studies support the notion that breast tumors from AAW and CW are molecularly different, they do not provide information as to why these aggressive tumors develop at an early age in AAW. To determine genetic factors that may contribute to tumor etiology in AAW, gene expression analysis was performed on disease-free breast specimens from AAW and CW.

## **Materials and Methods**

### *Tissue Acquisition*

Breast specimens were obtained from 25 AAW and 22 CW enrolled in the Clinical Breast Care Project (CBCP) at Walter Reed Army Medical Center (WRAMC) who had undergone a reductive mammoplasty and/or a breast biopsy and had no evidence of malignant disease. All tissue samples were collected with approval from the WRAMC Human Use Committee and Institutional Review Board. All subjects enrolled in the

CBCP voluntarily agreed to participate and gave written informed consent. Clinical and pathological information for each patient is shown in Table 1.

An additional 10 normal breast samples from 5 AAW and 5 CW without breast disease were acquired from the Mary Ellen's Tissue Bank located at Indiana University, Indianapolis, IN ([https://iuberc.iupui.edu/ffl\\_admin/login.jsp](https://iuberc.iupui.edu/ffl_admin/login.jsp)) for use in the validation studies by quantitative real-time PCR.

#### *RNA Isolation*

RNA was isolated from frozen breast specimens using the RNeasy Lipid Tissue Midi Kit (Qiagen Inc., Valencia, CA) with minor modifications. Approximately 400 mg of tissue (12  $\mu$ m sections) were homogenized per sample. Following isolation, the RNA was purified using the RNeasy Mini Kit (Qiagen Inc.) cleanup protocol. The concentration of each sample was determined with the NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific Inc., Waltham, MA) while the RNA integrity was monitored with the Bioanalyzer (Agilent Technologies, Inc., Santa Clara, CA).

#### *RNA Amplification*

1  $\mu$ g of total RNA from each sample was amplified using the MessageAmp II aRNA Amplification kit (Applied Biosystems Inc., Foster City, CA). GeneChip Eukaryotic Poly-A Controls (Affymetrix, Inc., Santa Clara, CA) were spiked into the total RNA samples prior to amplification to serve as a quality control measure for the amplification process. The first and second strand cDNA synthesis and cDNA

purification as well as the in vitro transcription (IVT) reaction and aRNA purification were performed according to the amplification kit instruction manual. The aRNA was labeled during the IVT reaction with 10 mM biotin-11-UTP (PerkinElmer, Inc., Waltham, MA). The concentration and size distribution of the biotin-labeled aRNA was determined on the NanoDrop ND-1000 and Bioanalyzer, respectively.

#### *Fragmentation, Hybridization, Staining, and Scanning*

Biotin-labeled aRNA (15 µg) was fragmented by incubation with 5X Fragmentation buffer (Affymetrix, Inc.) for 35 min. at 94° C. 10 µg of each fragmented sample was prepared for hybridization as specified in the Affymetrix GeneChip Expression Analysis Technical Manual for 100 Format (Midi) expression arrays and then hybridized to Affymetrix HG U133A 2.0 arrays for 16 hours. The arrays were washed, stained, and scanned according to Affymetrix protocols.

#### *Quantitative real-time PCR*

Total RNA was reverse transcribed to single-stranded cDNA using the High Capacity cDNA Reverse Transcription kit (Applied Biosystems Inc.). TaqMan Gene Expression assays (Applied Biosystems Inc.) were used to validate the differential expression of *PSPHL* (Hs00863464\_m1), *ACSM1* (Hs00369504\_m1), and *NTRK3* (Hs00176797\_m1). All sample reactions were performed in duplicate using the iCycler Real-Time Detection System (Bio-Rad Laboratories, Inc., Hercules, CA). *GAPDH* (Hs99999905\_m1) served as the endogenous control while FirstChoice® Human Brain

Reference RNA (Applied Biosystems Inc.) was used as the calibrator. The Comparative  $C_t$  Method was used to determine relative levels of gene expression.

#### *Statistical Analysis*

The microarray data were quantile normalized. To determine if a particular probe set was differentially expressed, the median normalized signal of each probe set was compared between the AAW and CW samples using a Mann-Whitney U test.  $P$ -values were calculated using standard non-parametric assumptions.

The heat map and clustering were created using those probes that had at least a 1.5 fold difference in expression between AAW and CW and a  $P$ -value  $<0.005$ . For each of these probes, the normalized signal for each sample was divided by the median normalized signal for that probe and the  $\log_2$  of the quotient calculated. Hierarchical clustering was performed using complete linkage with a correlation metric.

For the analysis of the quantitative real-time PCR (qRT-PCR) data, the medians of the  $2^{-\Delta\Delta C_t}$  values for the AAW and CW samples were compared using a Mann-Whitney U test to determine if the relative fold change was significantly different between the two groups.

## **Results**

### *Clinical characteristics of patient samples.*

Patient specimens used for this study were obtained from women enrolled in the Clinical Breast Care Project (CBCP) at Walter Reed Army Medical Center (WRAMC) who had no evidence of malignant breast disease. Forty-one of the patients had

undergone a breast biopsy while the remaining six had received a reductive mammoplasty. All histological slides were reviewed and diagnoses were made by a single, dedicated breast pathologist. Ten of the patients, including all six who had undergone a reductive mammoplasty, had no evidence of abnormal changes in their breast tissue while the remaining patients' breast specimens contained benign changes such as fibrocystic changes, microcalcifications, and mild intraductal hyperplasia (Table 1). Twenty-five of the patients were self-described African Americans while 22 of the patients were Caucasian (non-Hispanic). The majority of the patients were under the age of 50; 68% (15/22) of the CW were younger than 50 while 84% (21/25) of the AAW were below age 50 (Table 1).

*Normal breast tissue from African American versus Caucasian women displays different gene expression profiles.*

A comparison of the gene expression profiles of the normal breast samples from the AAW to those of the CW showed that 121 probe sets representing 112 genes were expressed at significantly different levels ( $P < 0.005$ ) in the two ethnic groups. Of these genes, 49 had at least a 1.5 fold difference in expression between the AAW and CW and are listed along with their biological function(s) in Table 2.

Figure 1 shows a heat map and a hierarchical cluster analysis of the 47 breast specimens. Using 53 probe sets that were significantly differentially expressed at  $P < 0.005$  and had at least a 1.5 fold or greater difference in expression between AAW and CW, this analysis was able to separate the normal breast samples from the AAW and CW into two distinct clusters.

*Validation of gene expression differences by quantitative real-time PCR.*

*PSPHL* and *ACSM1* exhibited the largest fold differences in expression between the normal breast samples from AAW compared to those from CW and therefore, were selected for further validation by qRT-PCR. To verify these results, 45 specimens from the original set of 47 samples along with an additional 10 out-of-sample disease-free breast specimens provided by the Mary Ellen's Tissue Bank (Indianapolis, IN) were evaluated.

As shown in Figure 2a, the significantly higher expression of *PSPHL* in normal breast tissue from AAW was confirmed by qRT-PCR analysis. The median fold change in expression of the AAW samples relative to the calibrator was 17.15 compared to a median value of 0 or no detectable transcript for the samples from CW ( $P=8.55 \times 10^{-7}$ ). The *PSPHL* transcript was undetectable in 58% of the Caucasian samples (15 out of 26 samples) compared to less than 7% (2 out of 29 samples) of the African American samples. The higher expression of *ACSM1* in AAW that was observed by microarray was also verified by qRT-PCR (Figure 2b). The median fold change in expression of *ACSM1* was 207.94 in the African American samples compared to 51.16 in the Caucasian samples ( $P=0.003$ ).

Conversely, expression of *NTRK3* was found to be lower in AAW. Two different microarray probes representing *NTRK3* were expressed at greater than 1.5 fold higher levels in CW than in AAW. As shown in Figure 2c, this difference in expression was validated by qRT-PCR. Median fold change in expression of *NTRK3* in CW was 0.27 whereas in AAW it was 0.13 ( $P=0.04$ ).

## Discussion

The contributions of differences in socioeconomic status and healthcare access on disparities in breast cancer outcomes between AAW and CW have been extensively studied [15]. In recent years, efforts have focused on identifying molecular differences that contribute to the aggressive tumor phenotype and less favorable prognoses in AAW [16,17]. To our knowledge, this is the first study to identify genetic differences in normal breast tissue from AAW and CW.

Here, we found that normal breast tissues from disease-free AAW and CW have inherent differences in gene expression. Our expression studies identified 49 genes at  $P < 0.005$  with at least a 1.5 fold difference in expression in normal breast tissue from AAW compared to CW. These genes were found to be involved in cellular metabolism, signal transduction, immune response, transcription, and cellular transport. These results suggest that breast tissue from AAW may be more metabolically active with higher proliferation rates than that from CW and therefore, may make AAW more susceptible to DNA damage and ensuing malignancy.

Acyl-CoA synthetase medium-chain family member 1, encoded by the *ACSM1* gene, is one of 26 acyl-CoA synthetases identified in the human genome [18]. Experimental evidence in mice suggests that *ACSM1* may be important in the production of cellular energy in the form of ATP through the  $\beta$ -oxidation of medium-chain fatty acids in the mitochondrial matrix [19]. Consequently, changes in the expression of *ACSM1* could impact the formation of ATP and therefore, alter the rate of cell growth. The significantly higher expression of *ACSM1* in breast tissue from AAW relative to that

from CW could lead to increased energy production, resulting in increased cell growth and metabolism. This, in turn, may increase the likelihood that breast cells in AAW will acquire mutations that may lead to malignant transformation and subsequent tumor formation.

The phosphoserine phosphatase-like (*PSPHL*) gene was found to have the greatest fold difference in expression between AAW and CW with more than 10-fold higher expression in normal breast tissue from AAW. Recently, it has been reported that prostate tumors from African American men have higher expression of *PSPHL* than prostate tumors from Caucasian men [20]. In addition, the expression of *PSPHL* in the normal tissue surrounding the prostate tumor was also found to be higher in African American men than in Caucasian men. Similarly to what is observed for breast cancer, prostate tumors tend to be more aggressive with higher mortality in African American men than in Caucasian men. Thus, some genes, such as *PSPHL*, may contribute to the aggressive nature of both prostate and breast cancer in African American men and women.

Although the function of *PSPHL* is unknown, its higher expression in Fanconi anemia fibroblasts compared to normal fibroblasts [21] suggests a possible role in cell cycle progression and/or cellular proliferation. As such, higher expression of *PSPHL*, as is seen in breast tissue from AAW when compared to CW, may cause increased cellular proliferation that may support the development of tumors in the breast and/or increase the aggressiveness of breast tumors in AAW.

Unlike *ACSM1* and *PSPHL*, *NTRK3* expression was significantly lower in AAW. *NTRK3* encodes a membrane-bound neurotrophic tyrosine kinase receptor called TrkC.

Signaling through this receptor allows for the survival and differentiation of sensory neurons that sense limb movement and body position [22]. Somatic mutations in *NTRK3* have been detected in colorectal [23], lung [24], breast [25], and pancreatic [26] cancers. In addition, a study of 40 breast tumors found an inverse relationship between protein levels of TrkC and tumor grade, suggesting that higher expression of TrkC may serve as an indicator of good prognosis in breast cancer [27]. It is of note that the aggressive tumor phenotype associated with young AAW includes poor tumor differentiation, which may reflect decreased expression levels of TrkC.

The normal differentiation and development of tissues, including breast tissue, requires proper communication between the stroma and epithelia [28]. However, genetic and epigenetic changes in normal breast cells can alter this interaction between breast epithelial cells and the surrounding microenvironment and serve as an initiating event in tumor formation. Extensive work has chronicled the importance of the stroma in tumorigenesis [28,29,30]. Signaling from the stroma to the epithelia via secreted proteins such as chemokines and growth factors can promote cellular growth and proliferation, tumor initiation and progression, and metastasis [28,31]. However, normal stromal cells can also suppress tumor formation by inhibiting proliferation of cancer cells or preventing transformation to a fully malignant state [28,30]. Therefore, cross-talk between epithelial and stromal cells will determine whether a particular microenvironment is permissive for tumorigenesis. Thus, changes in gene expression in normal breast cells, such as those identified here, will presumably alter the tissue microenvironment and consequently affect its ability to suppress or promote tumor formation.

Each of the specimens used to generate the microarray data was collected from patients without breast cancer. The pathologic changes observed in the specimens (fibrocystic changes, microcalcifications, mild intraductal hyperplasia, etc.) are common findings in breast biopsies and are not associated with increased risk for the subsequent development of breast carcinoma. Thus, for purposes of this study, these specimens may be considered within the realm of physiologically normal tissue. To date, only one patient has received a subsequent diagnosis of breast cancer (Patient 22). Approximately 3 years after acquiring the normal breast specimen, this patient was diagnosed with invasive breast cancer in the contralateral breast.

Because aggressive breast cancer is associated with incidence in young AAW, we selected tissue from largely young (60% under age 40) or pre-menopausal (81% under age 50) patients for this study. Thus, the differences in normal breast biology between AAW and CW identified here may help determine what molecular factors predispose young AAW to a higher incidence of aggressive breast tumors.

## **Conclusions**

In this study, we identified gene expression differences in normal breast tissue from AAW and CW with no evidence of malignant breast disease. Many of the differentially expressed genes function in signal transduction, immune response, transport, and diverse metabolic pathways. The identification of these gene expression differences in normal breast tissue from AAW and CW is important to expanding our knowledge of basic breast biology and especially understanding the biological differences in breast tissue from AAW and CW. The gene expression differences identified here

may explain why AAW are more likely than CW to be diagnosed with breast cancer at a younger age and why tumors in AAW are often more aggressive. Further study of these genes in normal breast tissue as well as breast tumors may lead to the identification of new molecular targets and the development of novel therapies to treat the distinct molecular characteristics of breast tumors in AAW, resulting in improved clinical outcomes and minimizing the differences in breast cancer survival between AAW and CW.

#### **List of Abbreviations**

AAW – African American women

CBCP – Clinical Breast Care Project

C<sub>t</sub> – cycle threshold

CW – Caucasian women

ER – estrogen receptor

PR – progesterone receptor

qRT-PCR – quantitative real-time polymerase chain reaction

WRAMC – Walter Reed Army Medical Center

#### **Competing Interests**

The authors declare that they have no competing interests.

#### **Authors' Contributions**

LAF participated in the study design, performed the RNA isolations, microarrays, and qRT-PCR assays, and drafted the manuscript. BL performed the statistical analyses. JK and BD prepared the samples for RNA isolation. JAH pathologically reviewed and diagnosed all patient samples. REE conceived of the study and participated in its design and edited the manuscript. CDS participated in study design. All authors read and approved the final manuscript.

### **Acknowledgements**

Samples from the Mary Ellen's Tissue Bank (METB) were used in this study. We thank contributors, including Indiana University who collected samples used in this study, as well as patients and their families, whose help and participation made this work possible. We also thank Dr. Susan Clare who assisted us in obtaining samples from METB. This study was supported by the United States Department of Defense through the Henry M. Jackson Foundation of the Advancement of Military Medicine (MDA 905-00-1-0022 to CDS). Opinions and assertions contained herein are the private views of the authors and are not to be construed as official or as representing the views of the Department of the Army or the Department of Defense.

### **References**

1. **SEER Cancer Statistics Review, 1975-2004, National Cancer Institute**  
[[http://seer.cancer.gov/csr/1975\\_2004/](http://seer.cancer.gov/csr/1975_2004/)]
2. Joslyn SA, West MM: **Racial differences in breast carcinoma survival.** *Cancer* 2000, **88**:114-123.
3. Shavers VL, Harlan LC, Stevens JL: **Racial/ethnic variation in clinical presentation, treatment, and survival among breast cancer patients under age 35.** *Cancer* 2003, **97**:134-147.

4. Wojcik BE, Spinks MK, Optenberg SA: **Breast carcinoma survival analysis for African American and white women in an equal-access health care system.** *Cancer* 1998, **82**:1310-1318.
5. Jatoi I, Becher H, Leake CR: **Widening disparity in survival between white and African-American patients with breast carcinoma treated in the U.S. Department of Defense healthcare system.** *Cancer* 2003, **98**:894-899.
6. Porter PL, Lund MJ, Lin MG, Yuan X, Liff JM, Flagg EW, Coates RJ, Eley JW: **Racial differences in the expression of cell cycle-regulatory proteins in breast carcinoma.** *Cancer* 2004, **100**:2533-2542.
7. Amend K, Hicks D, Ambrosone CB: **Breast cancer in African-American women: differences in tumor biology from European-American women.** *Cancer Res* 2006, **66**:8327-8330.
8. Olopade OI, Fackenthal JD, Dunston G, Tainsky MA, Collins F, Whitfield-Broome C: **Breast cancer genetics in African Americans.** *Cancer* 2003, **97**:236-245.
9. Chen VW, Correa P, Kurman RJ, Wu X-C, Eley JW, Austin D, Muss H, Hunter CP, Redmond C, Sobhan M, Coates R, Reynolds P, Herman AA, Edwards BK: **Histological characteristics of breast carcinoma in blacks and whites.** *Cancer Epidemiol Biomarkers Prev* 1994, **3**:127-135.
10. Elledge RM, Clark GM, Chamness GC, Osborne CK: **Tumor biologic factors and breast cancer prognosis among white, Hispanic, and black women in the United States.** *J Natl Cancer Inst* 1994, **86**:705-712.
11. Jones BA, Kasl SV, Howe CL, Lachman M, Dubrow R, Curnen MM, Soler-Vila H, Beeghly A, Duan F, Wens P: **African American/white differences in breast carcinoma: p53 alterations and other tumor characteristics.** *Cancer* 2004, **101**:1293-1301.
12. Cleator S, Heller W, Coombes RC: **Triple-negative breast cancer: therapeutic options.** *Lancet Oncol* 2007, **8**:235-244.
13. Martin DN, Boersma BJ, Yi M, Reimers M, Yfantis H, Williams EH, Tsai YC, Stephens RM, Weinstein JN, Ambs S: **Gene expression analysis of African-American and European-American breast tumors [abstract].** In *First AACR Conference on the Science of Cancer Health Disparities in Racial/Ethnic Minorities and the Medically Underserved: 27-30 November 2007; Atlanta, GA.* Philadelphia, PA: AACR; 2007:35.
14. Baumbach LL, Ahearn ME, Gomez C, Jorda M, Silva O, Halsey TA, Yan J, Ellison K, Gluck S: **Identifying distinct multi-ethnic genome-wide alterations in breast cancer using paraffin embedded samples [abstract].** In *Proceedings of the 99th Annual Meeting of the American Association for Cancer Research: 12-16 April 2008; San Diego, CA.* Philadelphia, PA: AACR; 2008:208.

15. Newman LA, Griffith KA, Jatoi I, Simon MS, Crowe JP, Colditz GA: **Meta-analysis of survival in African American and white American patients with breast cancer: ethnicity compared with socioeconomic status.** *J Clin Oncol* 2006, **24**:1342-1349.
16. Morris GJ, Mitchell EP: **Higher incidence of aggressive breast cancers in African-American women: a review.** *J Natl Med Assoc* 2008, **100**:698-702.
17. Newman LA: **Breast cancer in African-American women.** *Oncologist* 2005, **10**:1-14.
18. Watkins PA, Maignel D, Jia Z, Pevsner J: **Evidence for 26 distinct acyl-coenzyme A synthetase genes in the human genome.** *J Lipid Res* 2007, **48**:2736-2750.
19. Fujino T, Takei YA, Sone H, Ioka RX, Kamataki A, Magoori K, Takahashi S, Sakai J, Yamamoto TT: **Molecular identification and characterization of two medium-chain acyl-CoA synthetases, MACS1 and the Sa gene product.** *J Biol Chem* 2001, **276**:35961-35966.
20. Wallace TA, Prueitt RL, Yi M, Howe TM, Gillespie JW, Yfantis HG, Stephens RM, Caporaso NE, Loffredo CA, Ambs S: **Tumor immunobiological differences in prostate cancer between African-American and European-American men.** *Cancer Res* 2008, **68**:927-936.
21. Planitzer SA, Machl AW, Rueckels M, Kubbies M: **Identification of a novel c-DNA overexpressed in Fanconi's anemia fibroblasts partially homologous to a putative L-3-phosphoserine-phosphatase.** *Gene* 1998, **210**:297-306.
22. Nakagawara A: **Trk receptor tyrosine kinases: a bridge between cancer and neural development.** *Cancer Lett* 2001, **169**:107-114.
23. Bardelli A, Parsons DW, Silliman N, Ptak J, Szabo S, Saha S, Markowitz S, Willson JKV, Parmigiani G, Kinzler KW, Vogelstein B, Velculescu VE: **Mutational analysis of the tyrosine kinome in colorectal cancers.** *Science* 2003, **300**:949.
24. Davies H, Hunter C, Smith R, Stephens P, Greenman C, Bignell G, Teague J, Butler A, Edkins S, Stevens C, Parker A, O'Meara S, Avis T, Barthorpe S, Brackenbury L, Buck G, Clements J, Cole J, Dicks E, Edwards K, Forbes S, Gorton M, Gray K, Halliday K, Harrison R, Hills K, Hinton J, Jones D, Kosmidou V, Laman R, et al: **Somatic mutations of the protein kinase gene family in human lung cancer.** *Cancer Res* 2005, **65**:7591-7595.
25. Stephens P, Edkins S, Davies H, Greenman C, Cox C, Hunter C, Bignell G, Teague J, Smith R, Stevens C, O'Meara S, Parker A, Tarpey P, Avis T, Barthorpe A, Brackenbury L, Buck G, Butler A, Clements J, Cole J, Dicks E, Edwards K, Forbes S, Gorton M, Gray K, Halliday K, Harrison R, Hills K, Hinton J, Jones D, et al: **A**

screen of the complete protein kinase gene family identifies diverse patterns of somatic mutations in human breast cancer. *Nat Genet* 2005, **37**:590-592.

26. Wood LD, Calhoun ES, Silliman N, Ptak J, Szabo S, Powell SM, Riggins GJ, Wang T-L, Yan H, Gazdar A, Kern SE, Pennacchio L, Kinzler KW, Vogelstein B, Velculescu VE: **Somatic mutations of *GUCY2F*, *EPHA3*, and *NTRK3* in human cancers.** *Hum Mutat* 2006, **27**:1060-1061.
27. Blasco-Gutierrez M, Jose-Crespo I, Zozaya-Alvarez E, Ramos-Sanchez R, Garcia-Atares N: **TrkC: a new predictive marker in breast cancer?** *Cancer Invest* 2007, **25**:405-410.
28. Hu M, Polyak K: **Microenvironmental regulation of cancer development.** *Curr Opin Genet Dev* 2008, **18**:27-34.
29. Polyak K: **Breast cancer: origins and evolution.** *J Clin Invest* 2008, **117**:3155-3163.
30. Proia DA, Kuperwasser C: **Tumor agonist or antagonist.** *Cell Cycle* 2008, **4**:1022-1025.
31. Molloy T, van't Veer LJ: **Recent advances in metastasis research.** *Curr Opin Genet Dev* 2008, **18**:35-41.

#### Figure Legends

**Figure 1. Heat map and hierarchical clustering of normal breast tissue specimens from African American and Caucasian women.** The heat map and clustering were performed using 53 probe sets that were identified by gene expression profiling to be significantly differentially expressed between AAW and CW ( $P < 0.005$ ) with at least a 1.5 fold difference in expression. Each branch of the dendrogram and each column of the heat map represent a single patient sample. Red branches represent AAW while green branches represent CW. Each row of the heat map represents a different probe set and the gene represented by the probe is listed on the right side of the heat map. Red squares on the heat map represent higher levels of gene expression while green represents lower

expression. Black squares indicate no difference in expression relative to the median normalized signal intensity for that probe.

**Figure 2. qRT-PCR analysis of *PSPHL*, *ACSM1*, and *NTRK3* in normal breast tissue from African American and Caucasian women.** (a) Expression of *PSPHL* is higher in normal breast tissue from AAW than in CW. The median fold change in expression of *PSPHL* in AAW and CW was 17.15 and 0, respectively ( $P=8.55 \times 10^{-7}$ ). (b) *ACSM1* is expressed at higher levels in normal breast tissue from AAW than in that from CW. The median fold change in expression of *ACSM1* was 207.94 in AAW compared to 51.16 in CW ( $P=0.003$ ). (c) Expression of *NTRK3* is higher in CW than in AAW with median fold changes in expression of 0.27 and 0.13, respectively ( $P=0.04$ ).

**Table 1** Clinical characteristics of patients.

Sample	Age	Ethnicity	Diagnosis
1	56	Caucasian	Microcalcifications, fibrocystic changes
2	71	Caucasian	Microcalcifications, fibrocystic changes
3	46	African American	Fibrocystic changes
4	36	African American	Fibrocystic changes
5	48	African American	No abnormalities
6	50	Caucasian	Microcalcifications, fibrocystic changes, mild intraductal hyperplasia, radial scar
7	50	Caucasian	Microcalcifications, fibrocystic changes
8	25	African American	Fibrocystic changes, mild intraductal hyperplasia
9	20	Caucasian	Microcalcifications, fibrocystic changes, mild intraductal hyperplasia
10	54	African American	Mild intraductal hyperplasia, duct ectasia
11	26	Caucasian	Fibrocystic changes
12	47	Caucasian	Fibrocystic changes
13	56	African American	Fibrocystic changes, mild intraductal hyperplasia, duct ectasia
14	55	Caucasian	Fibrocystic changes, mild intraductal hyperplasia
15	56	Caucasian	No abnormalities
16	29	African American	Fibrocystic changes, mild intraductal hyperplasia
17	26	African American	Microcalcifications, fibrocystic changes
18	48	African American	Fibrocystic changes, mild intraductal hyperplasia
19	32	Caucasian	Fibrocystic changes
20	37	Caucasian	No abnormalities
21	63	African American	Microcalcifications, fibrocystic changes
22	34	African American	No abnormalities
23	39	Caucasian	Fibrocystic changes, mild intraductal hyperplasia
24	37	African American	Fibrocystic changes, lymphocytic mastitis

25	58	African American	Microcalcifications, fibrocystic changes, mild intraductal hyperplasia, lactational metaplasia
26	39	African American	Microcalcifications, fibrocystic changes, mild intraductal hyperplasia
27	47	Caucasian	Fibrocystic changes
28	23	African American	No abnormalities
29	42	Caucasian	Fibrocystic changes
30	33	African American	Fibrocystic changes
31	29	African American	Fibrocystic changes
32	36	African American	Microcalcifications, fibrocystic changes, mild intraductal hyperplasia, lactational metaplasia
33	22	Caucasian	Fibrocystic changes
34	43	African American	Microcalcifications, fibrocystic changes
35	47	African American	Fibrocystic changes
36	37	Caucasian	No abnormalities
37	34	African American	Fibrocystic changes
38	22	Caucasian	Fibrocystic changes
39	33	African American	Fibrocystic changes, fat necrosis
40	63	Caucasian	Microcalcifications, fibrocystic changes
41	34	Caucasian	No abnormalities
42	25	African American	Fibrocystic changes
43	33	Caucasian	No abnormalities
44	27	African American	No abnormalities
45	36	Caucasian	Fibrocystic changes
46	38	Caucasian	No abnormalities
47	19	African American	Fibrocystic changes

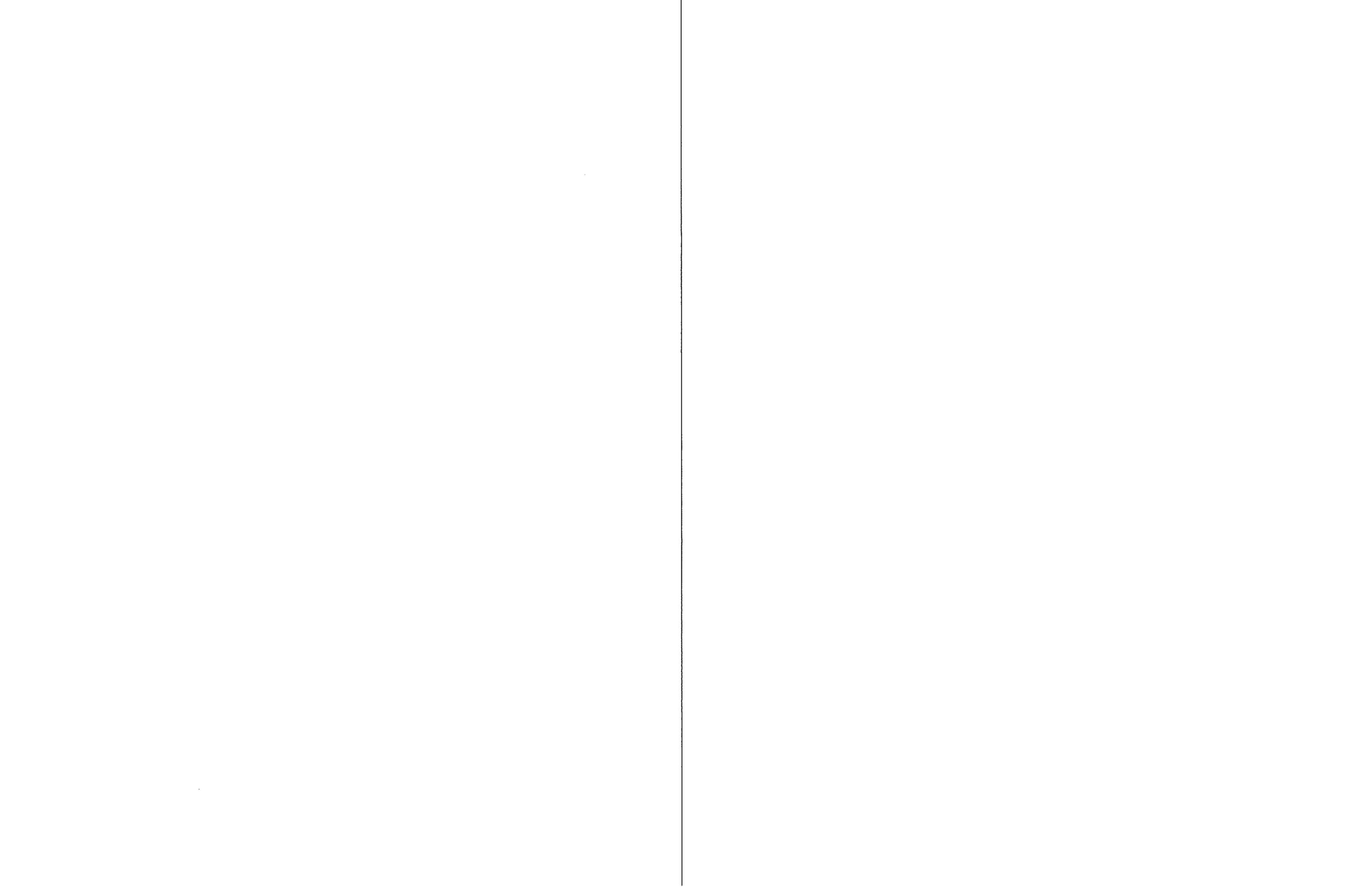
**Table 2** Differentially expressed genes in normal breast tissue from African American and Caucasian women (>1.5 fold difference). *GSTM1*, *IGH*, *POLR2J4*, and *NTRK3* are represented by multiple probes.

*Genes with higher expression in normal breast tissue from African American women*

Gene	Accession Number	P-value	Fold Change	Cell adhesion and motility	Cell growth and proliferation	Cell differentiation	Signal transduction	Transcription	Immuno response	Metabolism and catalytic activity	Transport	Other	Unknown
AAK1	BC002695	2.65x10 <sup>-3</sup>	1.68							X	X		
ABCA1	AK024328	3.49x10 <sup>-3</sup>	1.81				X			X	X	X	
ACSM1	AC003034	1.91x10 <sup>-4</sup>	4.82							X			
AMFR	NM_001144	8.01x10 <sup>-3</sup>	1.54				X			X	X	X	
ANXA8	NM_001630	3.74x10 <sup>-3</sup>	1.61						X	X			X
CABPA	NM_000715	9.55x10 <sup>-3</sup>	3.24	X						X			X
CMHAH	AF074480	1.28x10 <sup>-3</sup>	1.60									X	
COLEC10	NM_008438	4.89x10 <sup>-3</sup>	2.16								X		
CRYBB2	NM_000496	2.67x10 <sup>-3</sup>	2.60										
CUX2	AB006631	3.26x10 <sup>-3</sup>	2.85										
DKK1	NM_012242	1.44x10 <sup>-3</sup>	2.32				X			X		X	
GSTM1	X08020	2.84x10 <sup>-3</sup>	1.77										
GSST2	NM_000561	3.74x10 <sup>-3</sup>	1.86							X			
IGH	NM_000854	4.73x10 <sup>-3</sup>	1.97						X				
	M87789	2.67x10 <sup>-3</sup>	2.36								X		
	L14456	4.57x10 <sup>-3</sup>	2.25										
IGHM	BC001872	1.96x10 <sup>-3</sup>	1.71		X		X		X				
LILRA1	AF025529	4.00x10 <sup>-3</sup>	2.27				X		X			X	
MINDA	NM_002432	2.65x10 <sup>-3</sup>	1.64									X	
MTRF1L	AL049592	2.35x10 <sup>-3</sup>	1.75									X	
NLRP3	AK027194	3.15x10 <sup>-3</sup>	2.98				X		X			X	
PI3	NM_002638	1.99x10 <sup>-3</sup>	2.18						X			X	
PLA2G4C	AF065214	5.11x10 <sup>-3</sup>	1.63				X		X			X	
POLR2J	BE576209	1.85x10 <sup>-3</sup>	1.52					X		X			X
POLR2J4	AL047879	1.03x10 <sup>-3</sup>	1.57										
	AA601208	4.89x10 <sup>-3</sup>	1.51										X
PSPHL	NM_003832	4.12x10 <sup>-3</sup>	10.39										X
PTCH2	NM_003738	2.22x10 <sup>-3</sup>	1.97				X						
RPS26	NM_001029	4.28x10 <sup>-3</sup>	2.12									X	
SEZBL	AB047736	2.66x10 <sup>-3</sup>	2.71										X
SH2D1A	AF100540	1.38x10 <sup>-3</sup>	2.40				X		X				
SOS1	L13857	4.66x10 <sup>-3</sup>	2.58				X						X
SPINK2	NM_021114	1.14x10 <sup>-3</sup>	3.78								X		
TCN1	NM_001062	8.48x10 <sup>-3</sup>	2.82										
ZNF221	NM_013359	1.24x10 <sup>-3</sup>	3.34					X					

*Genes with higher expression in normal breast tissue from Caucasian women*

Gene	Accession Number	P-value	Fold Change	Cell adhesion and motility	Cell growth and proliferation	Cell differentiation	Signal transduction	Transcription	Immuno response	Metabolism and catalytic activity	Transport	Other	Unknown
AD1	NM_018269	2.26x10 <sup>-4</sup>	1.51							X			
CA2	M36532	2.30x10 <sup>-3</sup>	1.80							X			
CRYBB2P1	AA536000	2.46x10 <sup>-3</sup>	2.41							X			X
CYP26B1	NM_019885	1.99x10 <sup>-3</sup>	2.00							X			
DAPK1	BC003614	3.61x10 <sup>-3</sup>	2.31				X					X	
HTR3B	NM_006028	3.04x10 <sup>-3</sup>	1.99								X		
IL2ORA	NM_014432	2.84x10 <sup>-3</sup>	1.58				X						
LOC644450	AI050036	8.81x10 <sup>-3</sup>	3.17										X
LOC730092	AF054994	2.38x10 <sup>-3</sup>	1.70										X
NTRK3	S76476	9.51x10 <sup>-3</sup>	2.45			X							
PAX8	AL109696	9.51x10 <sup>-3</sup>	1.79			X							
PRL	BC001060	1.49x10 <sup>-3</sup>	3.54			X						X	
RGS11	NM_000948	1.92x10 <sup>-3</sup>	1.54				X					X	
SLC9A3R2	AF035771	3.61x10 <sup>-3</sup>	3.74		X		X				X		
TSGA10	NM_025244	5.10x10 <sup>-3</sup>	2.30										
WDR48	AI806628	3.55x10 <sup>-3</sup>	1.72									X	
ZBTB7A	AW027070	1.82x10 <sup>-3</sup>	1.65					X					X
		3.26x10 <sup>-3</sup>	2.22										



## **Molecular heterogeneity among sentinel lymph node metastases: Implications for the sentinel node hypothesis**

Darrell L. Ellsworth, Rachel E. Ellsworth, Heather L. Patney, Tyson E. Becker, Brenda Deyarmin, Jeffrey A. Hooke, Craig D. Shriver

Clinical Breast Care Project, Windber Research Institute, Windber, PA; Henry M. Jackson Foundation for the Advancement of Military Medicine, Rockville, MD; Clinical Breast Care Project, Walter Reed Army Medical Center, Washington, DC.

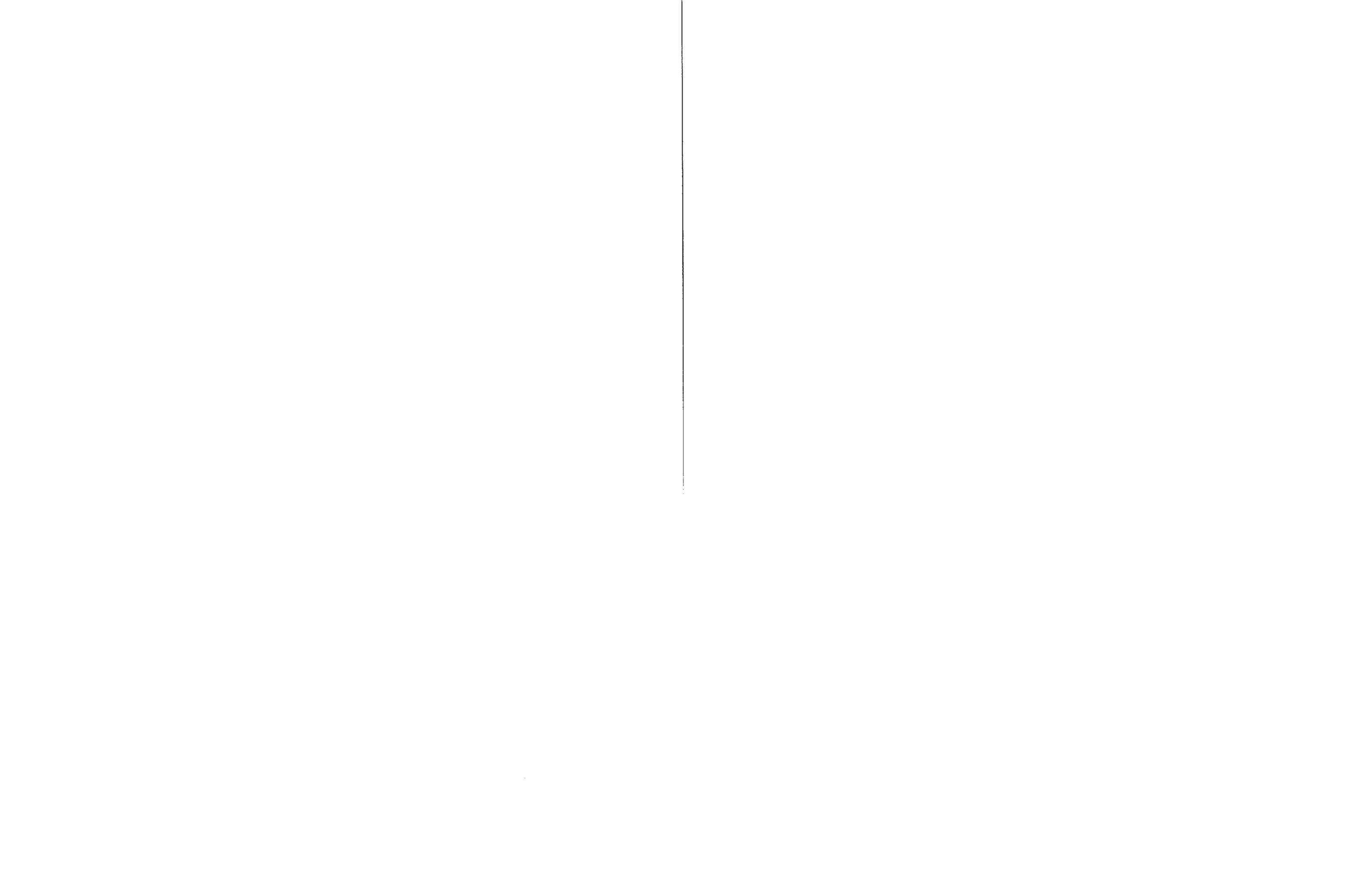
**Background:** The sentinel lymph node (SLN) hypothesis postulates that cancer cells initially spread from the primary tumor to the first-draining node(s), the sentinel node(s), before reaching higher echelon nodes. The sentinel node thus serves as a physical repository for the earliest tumor cells that are shed from the primary tumor and arrive in a nodal basin. Although SLN biopsy is clinically useful for staging breast cancer patients, little information exists regarding the origins of progenitor cells that define the genetic makeup and ultimately the clinical behavior of nodal metastases. This study used allelic imbalance (AI) to examine genomic relationships among metastases in the sentinel and non-sentinel axillary lymph nodes and multiple microdissected areas of the primary carcinoma in 30 patients with lymph node positive breast cancer.

**Methods:** Sentinel nodes were localized by standard scintigraphic and gamma probe techniques using 1.0 mCi unfiltered technetium-99m sulfur colloid (Tc-99) without blue dye. Pathologically positive nodes were identified by H&E histology/IHC and metastases were isolated by microdissection. Heterogeneity in the primary carcinoma was defined in three dimensions by laser microdissection of 6-15 areas (each ~2.86 sq mm in size) from multiple paraffin blocks. AI assessed at 26 chromosomal regions was used to examine molecular heterogeneity in sentinel and non-sentinel nodes as well as the primary tumor. Hierarchical clustering and phylogenetic analyses were then used to define patterns of molecular variation and relationships among samples.

### **Results:**

Lymph node metastases showed a high degree of genomic variability ranging from 0 – 54% AI, suggesting that some metastases progress with few genomic alterations while others carry numerous genomic changes. In particular, progenitor cells that colonize the sentinel nodes appeared to originate at different times during disease progression and may have developed by different molecular mechanisms in individual patients. Extensive heterogeneity in the primary tumor (0 – 38% AI) at the time of diagnosis may not fully explain patterns of molecular changes in the corresponding metastases.

**Conclusions:** Genomic diversity and timing of metastatic nodal spread may be important factors in determining outcomes of breast cancer patients. Although metastases develop in a sequential manner that maintains the known capability of SLN biopsy to accurately stage the axilla, these tumors develop from genomically-diverse progenitor cells. Growth of sentinel node metastases thus may be a consequence of stimulating factors that affect proliferation of previously disseminated cells rather than the timing of metastatic spread.



Genomic alterations associated with early stages of breast tumor metastasis

Running title: Genetic changes and breast cancer metastasis

Rachel E. Ellsworth<sup>1</sup>, Darrell L. Ellsworth<sup>2</sup>, Heather L. Patney<sup>2</sup>, Brenda Deyarmin<sup>2</sup>,  
Jeffrey A. Hooke<sup>3</sup>, Brad Love<sup>4</sup>, Craig D. Shriver<sup>3</sup>

<sup>1</sup>Clinical Breast Care Project, Henry M. Jackson Foundation for the Advancement of  
Military Medicine, Rockville, MD; <sup>2</sup>Clinical Breast Care Project, Windber Research  
Institute, Windber, PA; <sup>3</sup>Clinical Breast Care Project, Walter Reed Army Medical Center,  
Washington, DC; <sup>4</sup>Invitrogen, Carlsbad, CA

Contact for page proofs and correspondence:

Rachel E. Ellsworth  
Windber Research Institute  
620 Seventh Street  
Windber, PA 15963  
Phone: (814) 467-9844  
FAX: (814) 467-6334  
Email: [r.ellsworth@wriwindber.org](mailto:r.ellsworth@wriwindber.org)

Contact for reprints:

Ms. Kerri Cronin  
Clinical Breast Care Project  
Walter Reed Army Medical Center  
6900 Georgia Avenue, NW  
Washington, DC 20307  
Phone: (202) 782-0002  
[kerri.cronin@amedd.army.mil](mailto:kerri.cronin@amedd.army.mil)

Key words: allelic imbalance, primary breast tumor, metastasis

## ABSTRACT

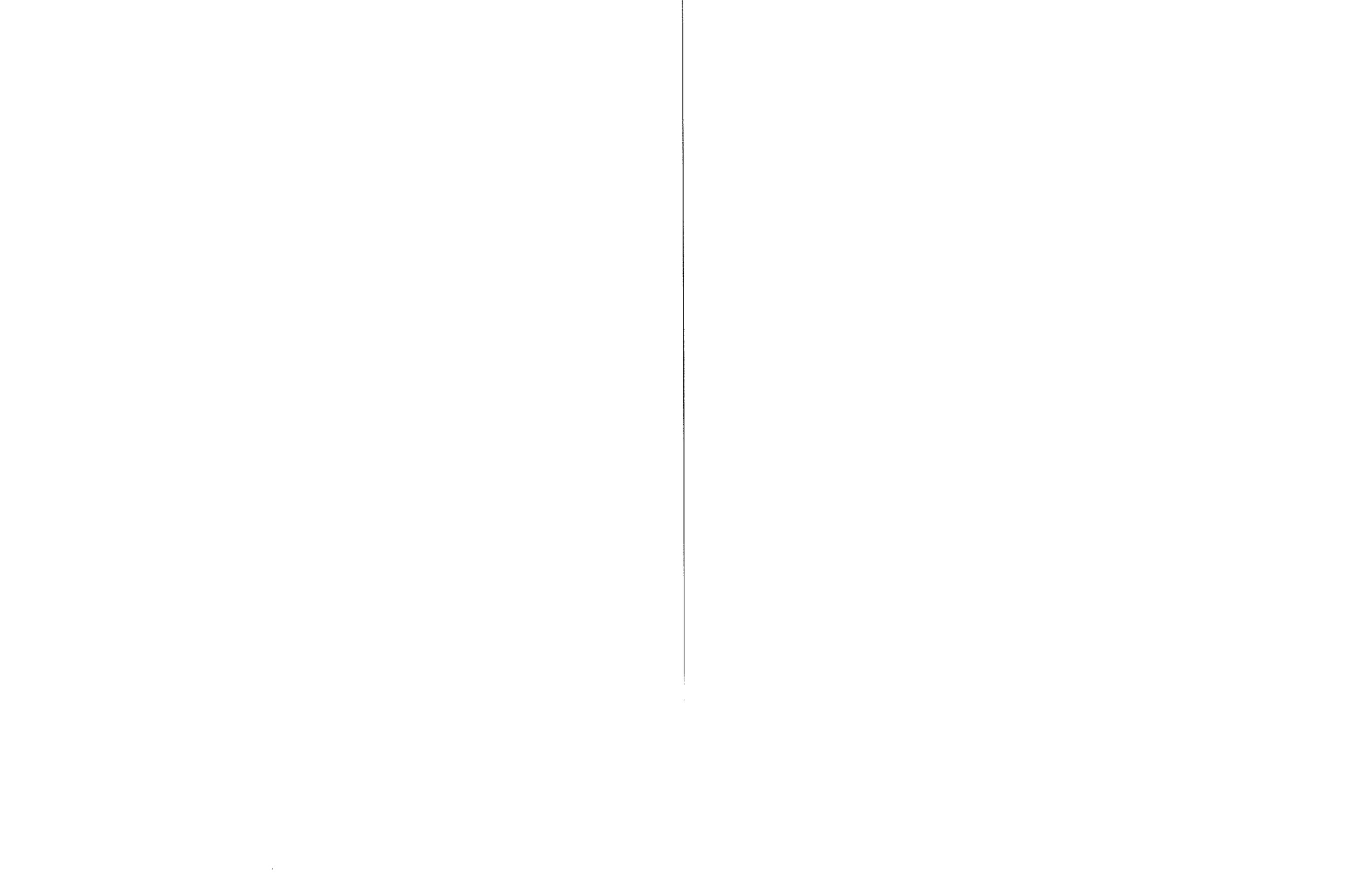
Background: Molecular studies suggest that acquisition of metastatic potential occurs early in the development of breast cancer; mechanisms by which cells disseminate from the primary carcinomas and successfully colonize foreign tissues are, however, largely unknown. Thus, we examined levels and patterns of chromosomal alterations in primary breast tumors from node-negative (n=114) and node-positive (n=115) patients to determine whether specific genomic changes are associated with tumor metastasis.

Methods: Fifty-two genetic markers representing 26 chromosomal regions commonly altered in breast cancer were examined in laser microdissected tumor samples to assess levels and patterns of allelic imbalance (AI). RT-PCR was performed to determine expression levels of candidate genes. Data was analyzed using exact unconditional and Student's t-tests significance value of  $P < 0.05$ .

Results: Overall levels of AI in primary breast tumors from node negative (20.8%) and node positive (21.9%) patients did not differ significantly ( $P = 0.291$ ). When data were examined by chromosomal region, only chromosome 8q24 showed significantly higher levels ( $P < 0.0005$ ) of AI in node-positive primary tumors (23%) versus node negative samples (6%). c-MYC showed significantly higher levels of gene expression in primary breast tumors from patients with lymph node metastasis.

Conclusions: Higher frequencies of AI at chromosome 8q24 in patients with positive lymph nodes suggest that genetic changes in this region are important to the process of metastasis. Because overexpression of c-MYC has been associated with cellular dissemination as well as development of the premetastatic niche, alterations of the 8q24

region, including c-MYC, may be key determinants in the development of lymph node metastasis.





ROYAL AUSTRALASIAN COLLEGE OF SURGEONS

## MEDIA RELEASE

### Star Wars defence battles breast cancer

Monday 12 May, 2008

American Star Wars technology developed to shoot down nuclear missiles could be used to save the lives of thousands of women with breast cancer, surgeons in Hong Kong will be told today.

American Colonel Craig Shriver will tell surgeons at the co-joint Annual Scientific Congress of the College of Surgeons of Australasia and Hong Kong, that a chance conversation at a cocktail party led to the discovery.

"I made the notes on a cocktail napkin while I was chatting to the Chief of the Army Space and Missile Defence Command; we discussed how defence technology could be used to fight breast cancer.

"In a missile strike there may be lots of dummy warheads to confuse star wars defence, what the technology does is to identify which missiles have the nuclear warhead. The same idea can be used to find the single cancer cell among the millions of normal ones.

"It is very exciting because we found that the technology which is basically an unclassified software algorithm is very accurate and it is a neat example of how defence technology can be turned towards another use that could help millions of people.

Colonel Craig Shriver is the Chief, Department of General Surgery at the Walter Reed Army Medical Centre, Washington DC and as a leader in breast cancer care will update surgeons on some of the latest developments in breast cancer care.

"When a woman has breast cancer generally she either has a lumpectomy to remove the cancerous tissue or if the cancer has spread throughout the breast, she may have to have the entire breast removed.

"This technique uses the unclassified defence software which can track the trajectory of missiles, to be used to better identify cancerous tumour cells on mammogram imaging, potentially tailoring the surgical approach and lessening the need for more disfiguring surgery.

**The research, which is being funded by the Department of Defence, is now in the middle stage and undergoing further refinements in the software algorithm for better breast cancer detection and analysis.**

This is the first time that surgeons from the Royal Australasian College of Surgeons have joined with surgeons from the Hong Kong College at one of the biggest ever surgical meetings held in the region.

More than 2000 surgeons are attending the co-joint Annual Scientific Congress, Achievement through Collaboration, at the Hong Kong Convention Centre where the latest innovations in surgery are being discussed.

World experts in surgical delivery and techniques will also update surgeons on what is happening at some of the leading surgical centres of the world, information that surgeons can use to save lives and improve patient care.

Presentations range from the ability of scientists to artificially grow blood vessels, a new treatment for back pain that could be used for other chronic pain, what is working in breast cancer treatments and the value of robotic surgery in Hong Kong, is it worth it?

The Vice Minister for health in China Jeifu Huang ( a former surgeon ) will be attending as well the well known Hong Kong engineer Sir Gordon Wu will detail the Chinese economic miracle from 1978 to 2008.

**The Congress will run from Tuesday 12 May until Friday 16 May, at the Hong Kong Convention and Exhibition Centre. Journalists are welcome to attend for more information please call**

**Fiona Gillies, RACS Media & PR Manager on 0407 339 556 or  
Ruth Charters, RACS Media & PR Officer on 0409 330274 or**

## The effect of HER2 expression on luminal A breast tumors

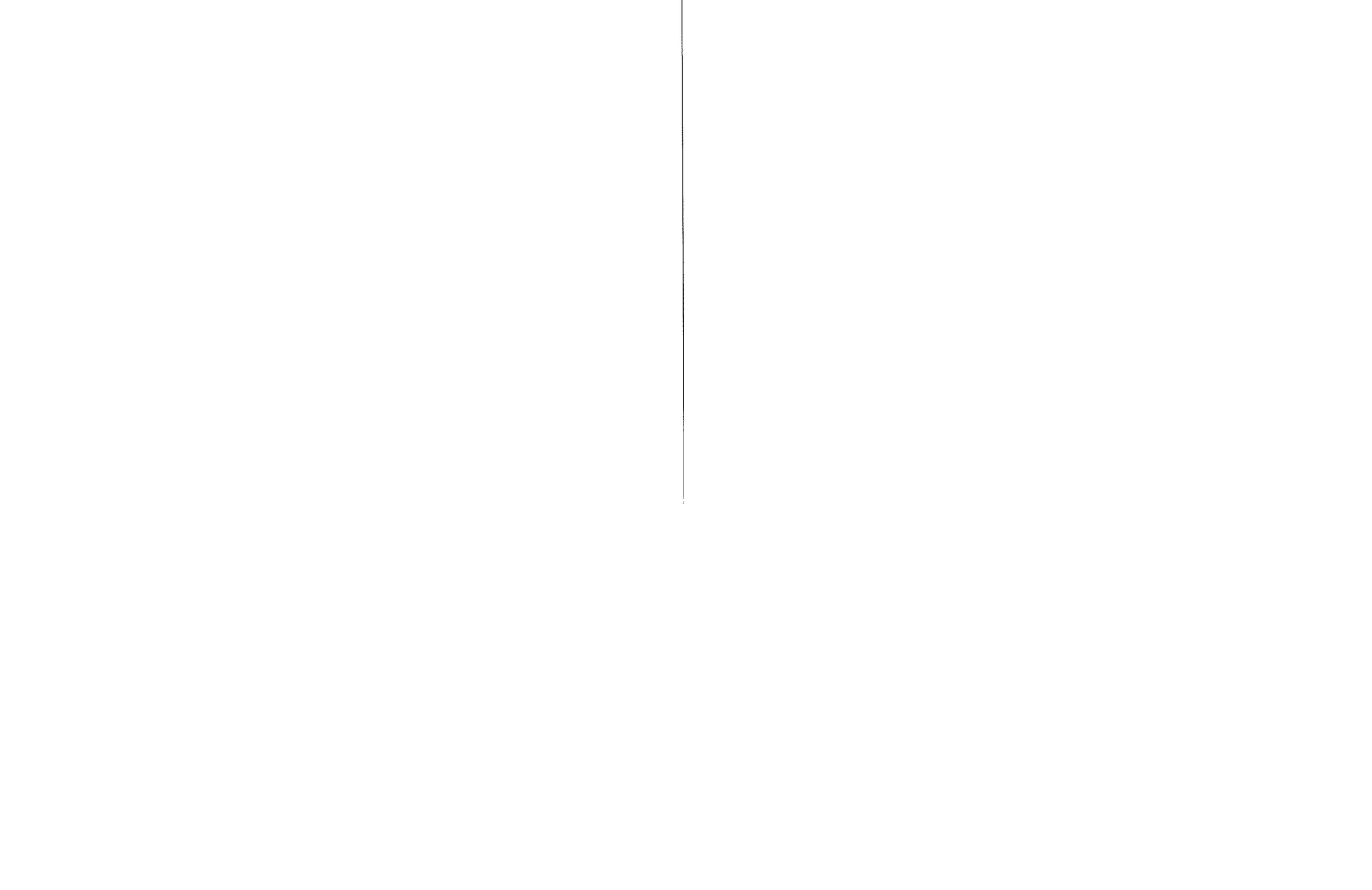
Rachel Ellsworth, Allyson Valente, Craig Shriver

**Background:** Luminal A (LumA) tumors are the most common subtype of breast cancer and have been associated with favorable prognosis. Although tumors can be considered LumA if they have  $\geq 1\%$  expression of ER, negative HER2 status ranges from 0+ (no expression) to 2+ (weak to moderate expression in  $>10\%$  of cells) without concurrent genomic amplification. It is not clear how moderate levels of HER2 expression influence tumor biology in LumA tumors.

**Methods:** The Clinical Breast Care Project database was queried to identify all hormone receptor positive cases with HER2 status defined as 0+ (n=89) or 2+/not amplified (n=154). Clinicopathological characteristics were compared between groups using chi-square analysis. Twenty-five 0+ and 25 2+ tumors were subjected to laser microdissection and hybridized to U133 2.0 microarrays (Affymetrix). Gene expression data was analyzed using Partek Genomics Suite.

**Results:** Age at diagnosis did not differ between patients with no (58.5 years) or moderate (59.3 years) HER2 expression. Patient groups did not differ by ethnicity, tumor stage, grade or size, or lymph node status. Although breast cancer mortality was low, all patients who died of disease (n=7) had moderate levels of HER2 expression ( $P=0.016$ ). At the molecular level, four genes were differentially expressed ( $P<0.01$ ,  $>2x$ -fold change): ESRRG expressed at significantly higher levels and DHRS2, SCUBE2 and SERPINA6 expressed at significantly lower levels in tumors with moderate HER2 expression.

**Discussion:** LumA (2+) tumors are dissimilar from LumA (0+) tumors. Increased expression of ESRRG has been associated with increased cellular proliferation and tamoxifen resistance, while decreased expression of SCUBE2 and SERPINA6 have been associated with inhibition of growth and proliferation. Together, these molecular differences suggest that LumA tumors with moderate HER2 expression, although classified as HER2 negative, have more aggressive characteristics than those without HER2 expression which may contribute to the significantly higher mortality rate seen in patients with LumA tumors expressing HER2.



Relationship between alterations of chromosome 8q24 and poorly-differentiated breast carcinoma

Heather L. Patney<sup>1</sup>, Brenda Deyarmin<sup>1</sup>, Jeffrey A. Hooke<sup>2</sup>, Rachel E. Ellsworth<sup>3</sup>, Brad Love<sup>4</sup>, Craig D. Shriver<sup>2</sup>

<sup>1</sup>Windber Research Institute, Windber, PA; <sup>2</sup>Walter Reed Army Medical Center, Washington, DC; <sup>3</sup>Henry M. Jackson Foundation for the Advancement of Military Medicine; <sup>4</sup>Invitrogen, Carlsbad, CA;

**Background:** Tumor grade can be used as a prognostic factor for patients with breast cancer, with poorly-differentiated tumors associated with less favorable outcome than well-differentiated tumors. Alteration of chromosome 8q24, including the c-MYC gene region, has been associated with high-grade disease. Here, we examined copy number changes of 8q24 in conjunction with gene expression levels from 12 candidate genes to determine how chromosomal alterations on 8q24 contribute to the unfavorable outcomes associated with poorly-differentiated breast cancer.

**Methods:** Copy number changes of c-MYC was assessed by FISH in a series of low- (n=56) and high-grade (n=61) breast tumors using commercial c-MYC and CEP8 probes (Vysis). Copy number changes were categorized as loss (copy number <1.7), low gain (3.0-4.0) and high gain (>4.0). Gene expression levels for 12 candidate genes were examined in a subset of low- (n=26) and high-grade (n=30) samples using TaqMan RT-PCR assays (Applied Biosystems). Data were analyzed using Mann-Whitney testing.

**Results:** Copy number gains of c-MYC were detected significantly more frequently ( $P < 0.0001$ ) in high-grade (mean = 3.92, range = 1.61-8.68) compared to low-grade samples (mean = 2.01, range = 1.02 – 4.87) with 40% of high-grade tumors having high copy number gains compared to <5% of low-grade tumors. Gene expression levels were significantly higher for NDUFB9, SQLE and MLZE ( $P < 0.005$ ); none of the genes had significantly lower levels in high-grade tumors.

**Conclusions:** While copy number gains of c-MYC were associated with high-grade disease, gene expression levels of c-MYC did not differ significantly between high- and low-grade tumors. In support of these data, previous studies in myeloid malignancies showed that although c-MYC was present on a common 4.3Mb amplicon on 8q24, expression levels of c-MYC were not increased, suggesting that other 8q24 genes may be the target of the 8q24 amplicon. Amplification of NDUFB9 and SQLE have been associated with poor survival in breast cancer patients, and expression of MLZE has been found to increase during the acquisition of metastatic potential of melanomas. Genomic amplification of the 8q24 region in conjunction with increased expression level of these three genes may, therefore, contribute to the less favorable prognosis seen in patients with poorly-differentiated breast cancer.



## Admixture mapping to identify breast cancer susceptibility loci in African American women

D Croft<sup>1</sup>, X Mao<sup>2</sup>, JA Hooke<sup>3</sup>, CD Shriver<sup>3</sup>, M Shriver<sup>2</sup>, RE Ellsworth<sup>4</sup>

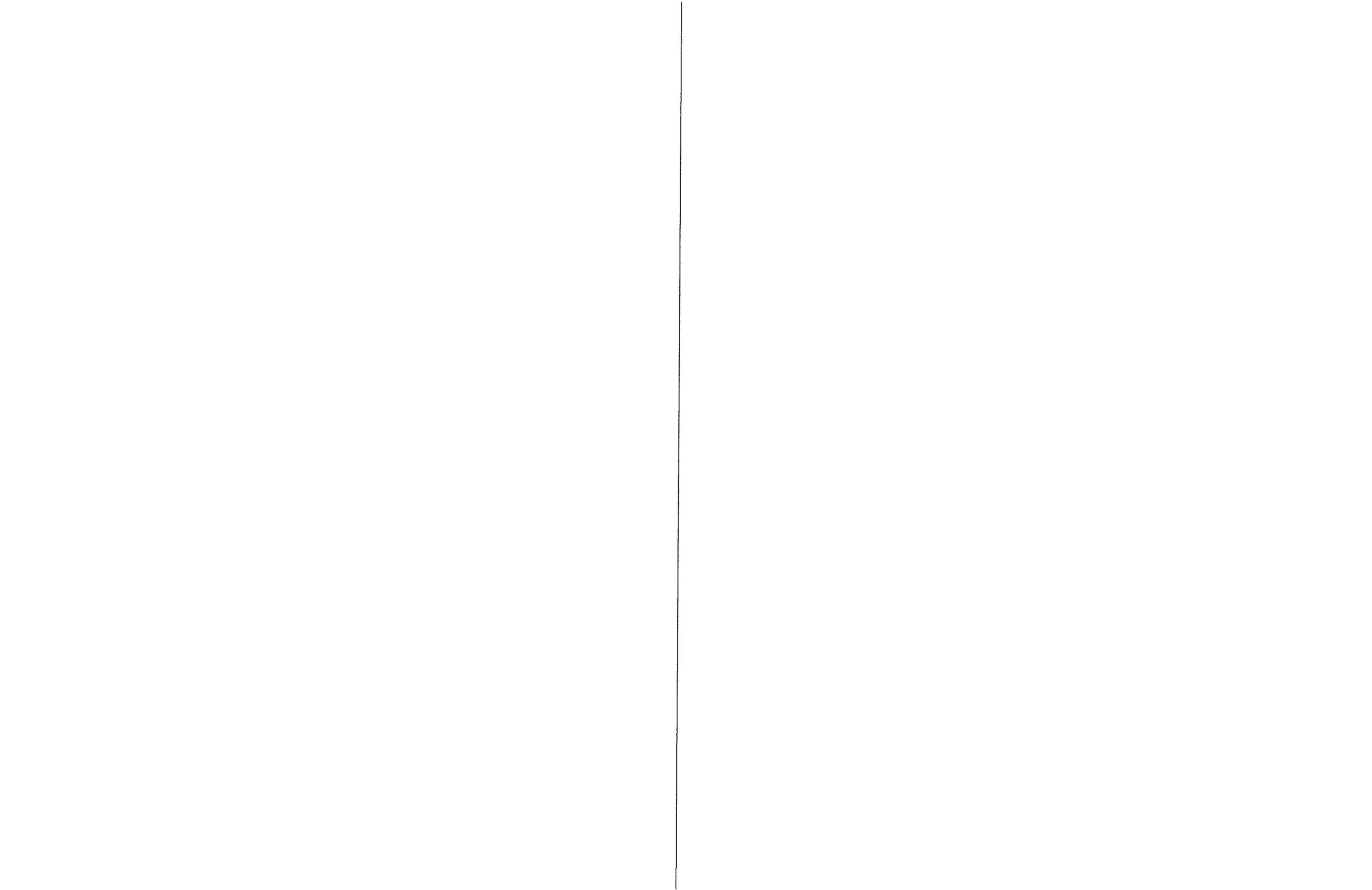
<sup>1</sup>Clinical Breast Care Project, Windber Research Institute, Windber, PA; <sup>2</sup>Department of Anthropology and Genetics, Pennsylvania State University, State College, PA; <sup>3</sup>Clinical Breast Care Project, Walter Reed Army Medical Center, Washington, DC; <sup>4</sup>Clinical Breast Care Project, Henry M. Jackson Foundation, Windber, PA

**Background:** The incidence of breast cancer in young women is higher in African American (AAW) compared to Caucasian (CW) and is associated with an aggressive phenotype, including high-grade, ER/PR and HER2 negative status and p53 positive characteristics. The aggressive tumor behavior and pathological features of AAW are also seen in women from the Bite of Biafra, an area of Western Africa that was the predominant source of the American slave trade. Similar tumor characteristics and shared ancestry suggest that population-specific molecular factors contribute to the aggressive tumor phenotypes associated with AAW. Therefore, a method of admixture mapping is proposed for this study. Admixture mapping is also known as mapping by admixture linkage disequilibrium. It assumes in a recent admixed population, if certain phenotype or the prevalence of a disease is mainly caused by one of the ancestral groups, the genetic variant can be located by the elevated ancestry from that group.

**Methods:** Pathological characterization was performed histology specimens from 91 self-described AAW diagnosed with invasive breast cancer enrolled in the Clinical Breast Care Project. Genomic DNA was isolated and SNP data generated using the Affymetrix Mapping 100K arrays. African ancestry was determined using maximum likelihood estimator and chromosomal regions associated with disease identified using AncestryMap and AdmixMap.

**Results:** The average age of diagnosis was 51 years. 25% of patients had poorly-differentiated, triple negative tumors. All 91 patients had significant African ancestry, with a mean level of Nigerian ancestry of 0.3857. Results from both AncestryMap and AdmixMap suggest loci on chromosomal regions 5q32.2, 9q13.1 and 17p13.3 are associated with the development of breast cancer.

**Discussion:** Recently, admixture mapping was used to successfully identify a prostate cancer gene to chromosome 8q24 in African American men. Using a similar approach, we identified three chromosomal regions associated with the presence of invasive breast cancer in AAW with significant Western African ancestry. Of note, loss of 9q13 has been associated with poor survival, and allelic imbalance of 17p13.3 has been associated with loss of hormone receptor expression and poor prognosis. Thus, the use of admixture mapping may aid in the identification of genes involved in the development of the aggressive form of breast cancer associated with AAW.



**Fingerprinting genomic heterogeneity in primary breast carcinomas and among sentinel lymph node metastases: Implications for clinical management of breast cancer patients**

Darrell L. Ellsworth, Rachel E. Ellsworth, Heather L. Patney, Allyson Oviedo, Alisha George, Daniel T. Croft, Jr., Brad Love, Rick M. Jordan, Brenda Deyarmin, Tyson E. Becker, Jeffrey A. Hooke, Craig D. Shriver

Clinical Breast Care Project, Windber Research Institute, Windber, PA; Henry M. Jackson Foundation for the Advancement of Military Medicine, Rockville, MD; Invitrogen, Carlsbad, CA; Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA; Clinical Breast Care Project, Walter Reed Army Medical Center, Washington, DC.

**Background:** Sentinel lymph node (SLN) status is a key prognostic factor for breast cancer patients. The SLN hypothesis postulates that cancer cells initially spread from the primary tumor to the first-draining (sentinel) node(s) before reaching higher echelon nodes. Although SLN biopsy is clinically useful for staging patients, little is known about the genomic heritage of cells that define the genetic makeup and clinical behavior of nodal metastases. We used allelic imbalance (AI) coupled with high-density SNP genotyping to examine the extent of genetic heterogeneity in primary tumors and relationships among metastases in sentinel and non-sentinel axillary lymph nodes in 30 patients with node positive breast cancer.

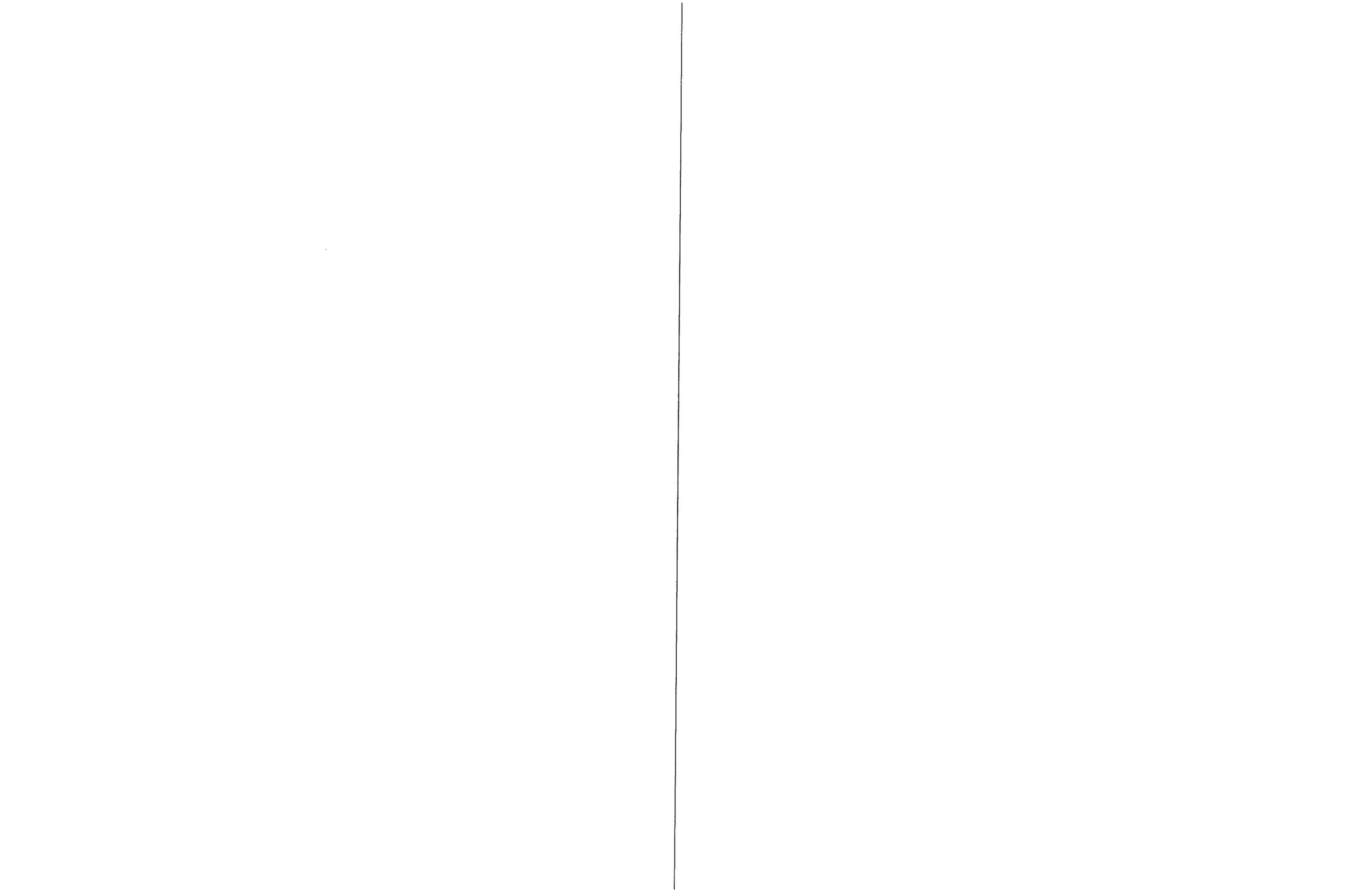
**Methods:** Pathologically positive nodes were identified by H&E/IHC, while sentinel nodes were localized by standard scintigraphic techniques. For each patient, nodal metastases and 6-15 areas from multiple paraffin blocks of the primary carcinoma were isolated by microdissection. AI was assessed by microsatellite typing at 26 chromosomal regions; genome-wide copy number variation was examined using Affymetrix GeneChip® Human Mapping 500K arrays and the Genotyping Consol. The genomic heritage of sentinel and non-sentinel nodal metastases in relation to multiple areas of the primary tumor was assessed by hierarchical clustering and phylogenetic analyses.

**Results:** Extensive genomic heterogeneity was observed in primary tumors and among nodal metastases (0–44% in primaries, 0–36% in metastases). Overall levels of variation were significantly higher ( $p < 0.01$ ) in primary tumors (17%) than metastases (10%), but did not differ between sentinel (10.2%) and non-sentinel (10.0%) metastases. Within patients, many (but not all) metastases appeared to be descended from different areas of the primary tumor, but patterns of descent were complex. As a group, sentinel node metastases were not genomically distinct from other axillary node metastases, and there was no evidence that SLN metastases are seeded by cells that colonize the axilla early in tumorigenesis.

**Conclusions:** Although metastases develop in a sequential manner that maintains the known capability of SLN biopsy to accurately stage the axilla, metastases appear to develop from genomically-diverse progenitor cells descended from different areas of the

primary tumor. Genomic diversity in primary tumors and metastases may be an important factor in prognosis and treatment resistance in breast cancer patients.

[http://www.sabcs.org/AbstractSubmission/Index\\_sh.asp](http://www.sabcs.org/AbstractSubmission/Index_sh.asp)



### **Abundance and distribution of polychlorinated biphenyls (PCBs) in breast tissue**

Ellsworth DL, Gillard D, Love B, Ellsworth RE, Deyarmin B, Hooke JA, Kostyniak PJ, Shriver CD. Windber Research Institute, Windber, PA; State University of New York at Buffalo, Buffalo, NY; Invitrogen, Carlsbad, CA; Walter Reed Army Medical Center, Washington, DC.

#### **Background:**

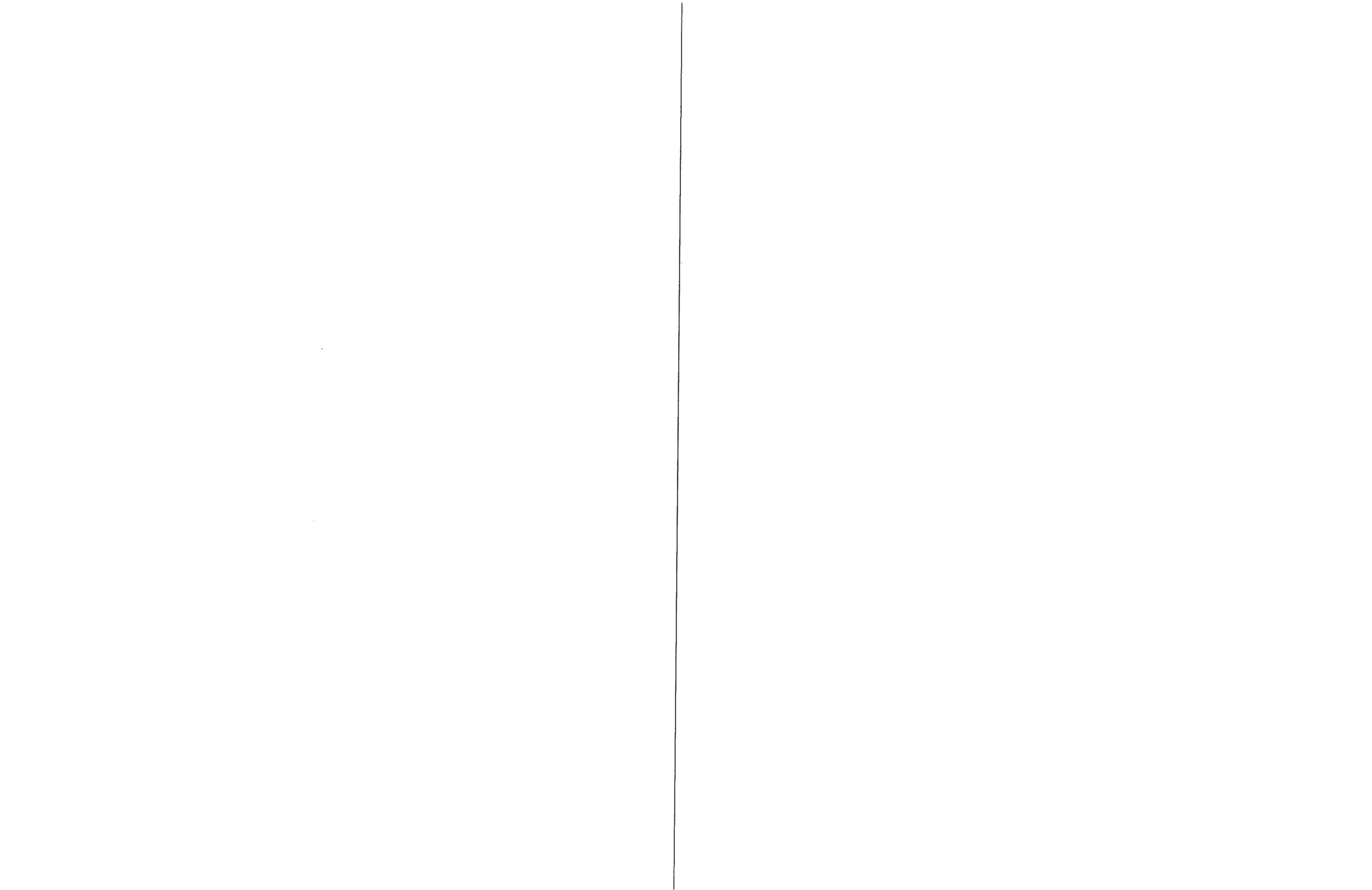
Breast cancer is a common disease with a complex etiology whose causative factors remain ambiguous and poorly defined. Many environmental chemicals act as carcinogens by causing damage to DNA, while other compounds have been shown to mimic the biological function of estrogen. Polychlorinated biphenyls (PCBs), which constitute a class of chlorinated compounds that are ubiquitous in the environment due to widespread industrial use, are known carcinogens believed to be associated with adverse health effects. Because the rising incidence of breast cancer in Westernized societies may result from exposure of the breast to increasing levels of exogenous chemicals, there is keen interest in determining relationships between exposure to environmental chemicals and breast cancer risk.

**Material and Methods:** Flash frozen sections of breast tissue were collected from each quadrant and the central region of the breast from 47 patients enrolled in the Clinical Breast Care Project (CBCP) at Walter Reed Army Medical Center. Pathological diagnoses ranged from disease free prophylactic mastectomy to metastatic breast cancer. From a portion of each tissue sample, 98 PCB congeners were assayed by pressurized liquid extraction followed by high resolution capillary gas chromatography with electron capture detection.

**Results:** All chemical concentrations were adjusted for the lipid content of each sample to account for variation in lipid composition in breast tissues. PCBs were abundant in the breast tissue sections – only one congener was not observed in any patient and four (153, 138+163, 180, 206) were seen in all patients in at least one quadrant. The mean total PCB concentration in 202 breast quadrants was 602 ng/g lipid (range 0–4030), while for individual congeners, PCB 153 had an average level of 107 ng/g lipid (range 8–773 ng/g, the maximum observed PCB in any quadrant). Total PCB concentration ( $P < 0.001$ ) and presence of 26 individual congeners ( $P < 0.05$ ) were highly correlated with patient age, particularly in the UIQ and LOQ, but not in the central region, suggesting that chemical concentrations may accumulate at different rates in different regions of the breast. Similarly, no significant differences were observed for PCB concentrations among the four traditional quadrants, but 23 PCBs were present at significantly lower levels ( $P < 0.05$ ) in the central region compared to the other quadrants. When examining PCB levels by disease status, seven individual congeners and median overall levels were significantly higher ( $P < 0.05$ ) in patients with invasive cancer compared to disease free patients.

**Discussion:** Population biomonitoring is beginning to assess exposures of the general population to environmental chemicals to evaluate long-term adverse health effects. This

study showed that many PCB residues are present in human breast tissue and these chemicals are not distributed evenly and are correlated with age and presence of invasive disease. Improved understanding of the true carcinogenic potential of environmental chemicals may be useful in efforts to reduce individual risk for breast cancer.



## Identification of breast cancer metastasis initiation and virulence genes

RE Ellsworth<sup>1</sup>, C Heckman<sup>2</sup>, B Love<sup>3</sup>, CD Shriver<sup>4</sup>

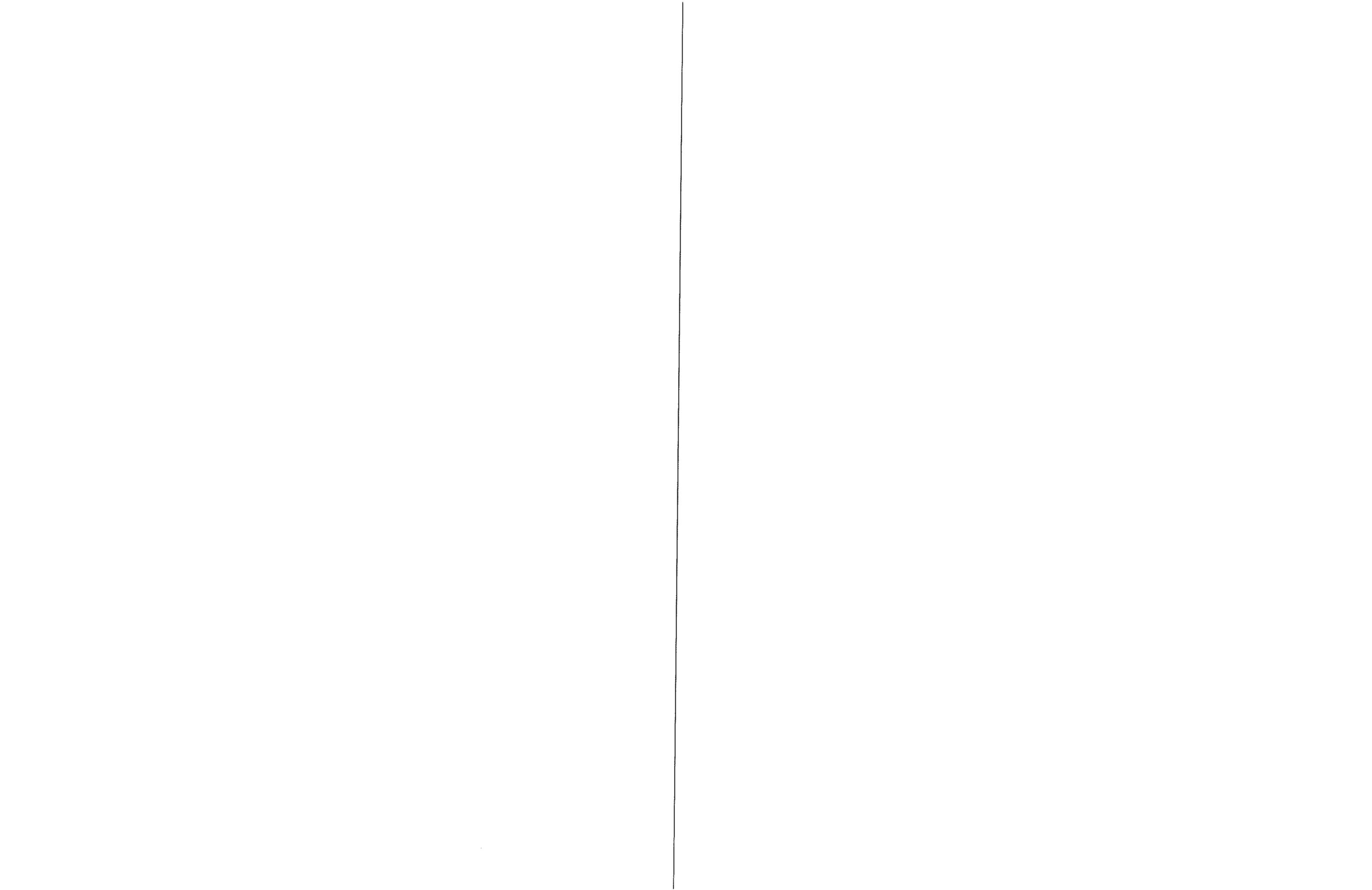
<sup>1</sup>Henry M. Jackson Foundation, <sup>2</sup>Windber, PA; Windber Research Institute, Windber, PA; <sup>3</sup>Invitrogen, Carlsbad, CA; <sup>4</sup>Walter Reed Army Medical Center, Washington, DC

**Background:** Metastasis involves processes such as invasion, angiogenesis, intravasation, extravasation and eventual proliferation in the target organ. Metastasis initiation genes are those involved in dissemination of cells from the primary tumor while metastasis virulence genes contribute to metastatic colonization, specifically at the secondary site. Identification of metastasis initiation and virulence genes is critical to improve our understanding of and the ability to effectively treat metastatic breast cancer.

**Methods:** A 51-gene signature was previously identified using RNA from 20 primary breast and matched metastatic lymph node tumors using the U133 2.0 microarrays (Affymetrix). Real-time PCR was performed for each assay using TaqMan assays (Applied Biosystems) using RNA from 25 pairs of primary breast and matched metastatic lymph node tumors. Data was evaluated using Mann-Whitney testing.

**Results:** Gene expression levels were significantly different for 40 genes including 25 with  $P < 0.001$  and 15 with  $P < 0.05$ . Thirteen of these genes had significantly higher levels and 27 had lower levels of expression in metastatic tumors. Ten genes had >50-fold difference in expression including WNT2, KRT14, TAC1, COL11A1, MMP13, GRP, and KERA with higher expression in primary tumors and EPHA3, PAX5 and NTS with higher expression in the metastasis.

**Discussion:** A breast cancer metastasis signature involving 40 genes has been identified and validated. Genes expressed at higher levels in primary breast tumors are largely involved in degradation of the extracellular matrix (ECM), likely enabling cells with metastatic potential to disseminate. Decreased expression of these genes in the metastatic tumors suggests that once tumor cells have disseminated, invasion into foreign tissues does not require active tissue remodeling. Rather, cells that have successfully metastasized are characterized by the expression of genes involved in transcription, metabolism and immune response, potentially blocking cellular differentiation and providing cells with proliferation and survival advantages. These data not only improve our understanding of the biological processes involved in successful metastatic but provide new targets to arrest tumor cells dissemination and metastatic colonization.



Is a diagnosis of invasive breast cancer an effective motivational factor for lifestyle change?

Nathaniel J. Sann<sup>1</sup>, Rachel E. Ellsworth<sup>1</sup>, Leonid Kvecher<sup>1</sup>, Craig D. Shriver<sup>2</sup>, Darrell L. Ellsworth<sup>1</sup>

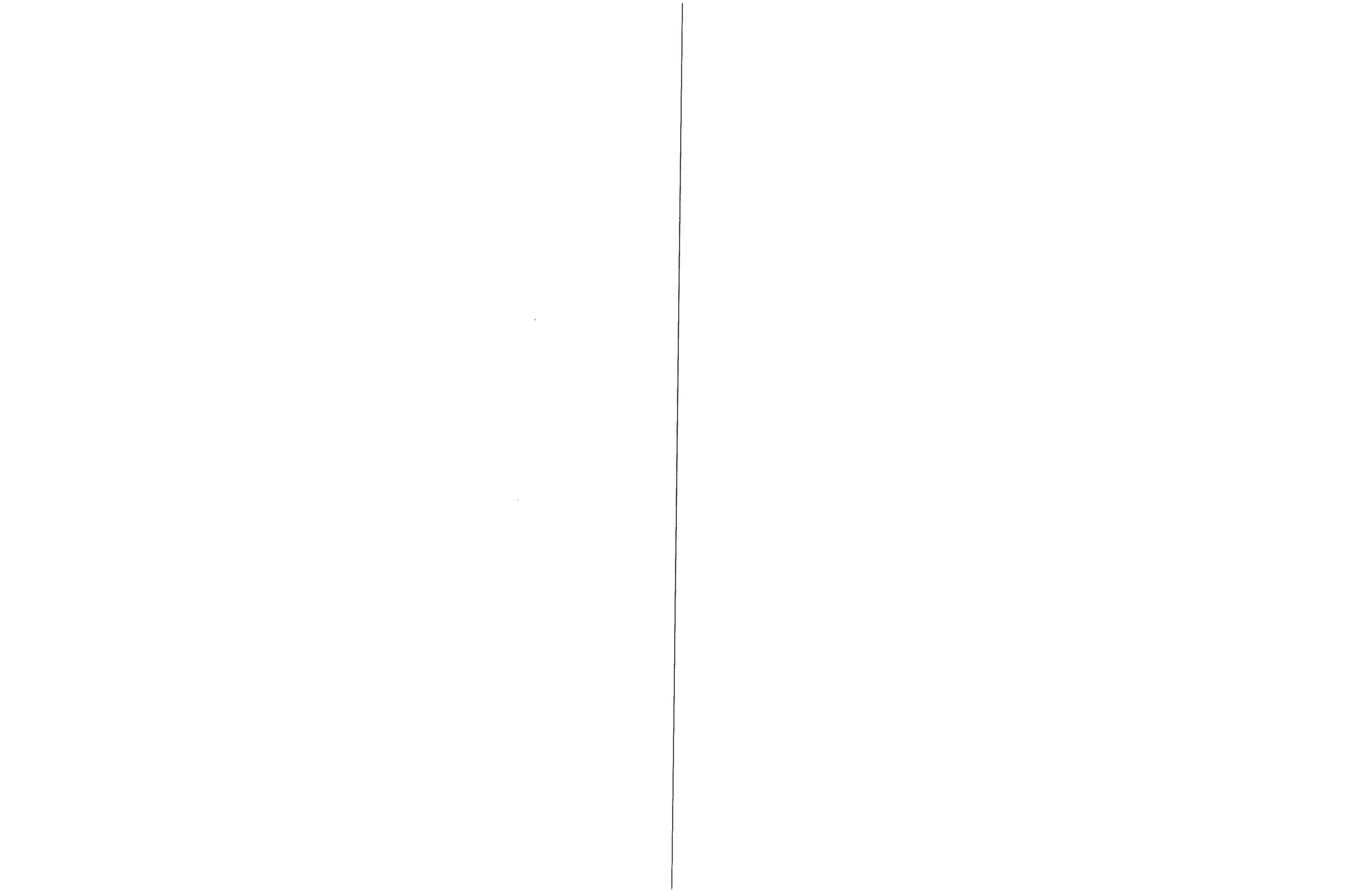
<sup>1</sup>Clinical Breast Care Project, Windber Research Institute, Windber, PA; <sup>2</sup>Clinical Breast Care Project, Walter Reed Army Medical Center, Washington, DC

**Background:** While survival rates for patients diagnosed with invasive breast cancer have increased dramatically, survivors often face a host of adverse health effects. Factors such as obesity, physical activity and tobacco may contribute to decreased survival and quality of life. Here, we evaluated behavioral risk factors in patients with and without breast cancer to determine whether a diagnosis of invasive disease was sufficient motivation to modify lifestyle choices.

**Methods:** The dataset included patients undergoing surgical procedures and diagnosed with invasive (n=299) or benign (n=130) breast disease between 2001 and 2006 and who had baseline and 1-year follow-up information available. Changes in BMI, fat intake, exercise frequency, alcohol and tobacco use, caffeine consumption and frequency of breast self exam (BSE) were assessed.

**Results:** Although significantly more invasive patients (70%) were menopausal compared to benign patients (48%), no other patient characteristics were significantly different at baseline; most were overweight, non (current) smokers, and consumed high fat diets, ~40% reported exercising  $\geq 3$  times/week, >60% had high ( $\geq 500$  mg/day) levels of caffeine and consumed low ( $\leq 1$  drink/month) levels of alcohol. The only modifiable behavior that showed a significant change ( $P < 0.05$ ) between baseline and 1-year was compliance with recommended BSE frequency in invasive patients, improving from 61% of patients at baseline to 72% one year later; a concomitant change was not seen in benign patients. No other behaviors changed either from baseline to 1-year or between invasive and benign patients.

**Conclusions:** The paucity of behavioral changes after a diagnosis of breast cancer suggests that invasive disease is not sufficient motivation to promote healthy lifestyles. Our data suggest a need for health-related behavioral counseling and support systems to successfully modify personal behaviors. Development and implementation of lifestyle recommendations have the potential to improve the health and quality of life of breast cancer survivors.



## Molecular characterization of early breast disease

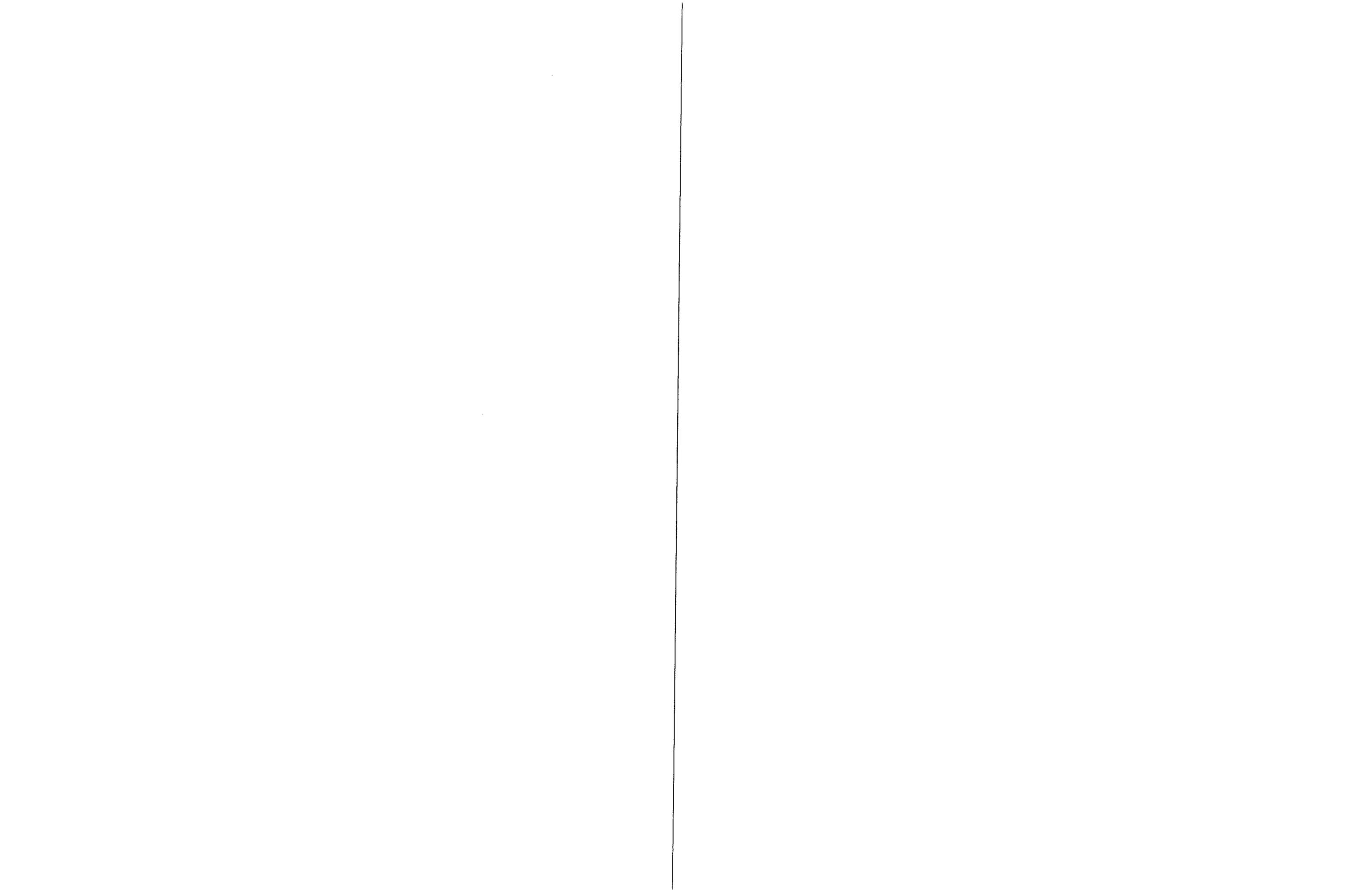
Jamie D. Weyandt, Rachel E. Ellsworth, Jamie L. Fantacone-Campbell, Brenda Deyarmin, Darrell L. Ellsworth, Jeffrey A. Hooke, Craig D. Shriver

**Background:** Columnar cell lesions (CCL) and atypical ductal hyperplasia (ADH) have been found to frequently coexist and share molecular changes with *in situ* and invasive breast cancer, suggesting that CCL and ADH may be precursors to invasive disease. These conclusions are, however, largely based on studies examining synchronous CCL and/or ADH from patients diagnosed with more advanced disease. In this study, we assessed allelic imbalance (AI) in pure CCL or ADH specimens to characterize molecular changes associated with the initiation and development of early breast lesions.

**Methods:** DNA samples were obtained from laser microdissected lesions from patients diagnosed with CCL with (n=23) or without (n=18) atypia or with ADH (n=32) without synchronous *in situ* or invasive disease. All specimens were characterized by a single breast pathologist. A microsatellite marker panel representing 26 chromosomal regions commonly altered in breast cancer was used to assess patterns of AI.

**Results:** The average AI frequency was 9% (range 0-20%) in CCL and 8% (range 0-25%) in ADH with no significant difference in levels of AI between CCL and ADH. The highest frequency of AI occurred at chromosomes 6q22-q23 and 22q13 (20%) in CCL and 8q24 (23%) in ADH. AI occurred at significantly higher levels at chromosome 13q14 ( $P<0.05$ ) in ADH compared to CCL and at 22q13 ( $P<0.01$ ) in CCL compared to ADH.

**Conclusions:** Differences in patterns of AI, especially at chromosome 22q13 suggest that CCL may not be a precursor to ADH and that the two diseases evolve along independent pathways. Low overall levels of AI in pure CCL and ADH suggest that the majority of these lesions are not genetically advanced. The variability in levels of AI in both CCL and ADH lesions may contribute to clinical heterogeneity in behavior and aggressive and suggest a need to assess molecular characteristics of early lesions to better predict outcomes.



## Metastatic molecular signature applied to primary breast tumors and their lymph node metastases

The molecular relationship between the best prognosticator for distant spread (i.e. lymph node metastasis) and the mechanism of distant spread (i.e. hematogenous dissemination) remains unclear. In an effort to clarify this molecular relationship we used a recently described metastatic solid organ molecular signature and applied it to primary breast cancers and their lymph node metastases. We found that in aggregate the expression pattern of the lymph node mirrored that of the solid organ metastases.

Despite recent advances in diagnostics the best prognosticator for survival remains the presence or absence of lymph node metastasis in breast cancer patients {Morabito, Magnani, et al. 2003 12 /id}. Thus, patients who have lymph node metastases at time of diagnosis will likely *ultimately* succumb to distant solid organ metastases without chemotherapy, whereas those that do not will likely enjoy long term survival even without systemic therapy {Moon 1984 18 /id}. Interestingly, the relative importance of lymphogenous versus hematogenous spread (and its resultant solid organ seeding) has been a central debate for nearly a century now yet their relationship remains unsettled {Fidler 2000 2 /id} {Ruoslahti 2002 24 /id}. In an effort to further clarify the relationship, at a molecular level, between lymphatic and solid organ metastasis we designed a molecular beacon based relative quantitative RT-PCR experiment using the majority of a recently described molecular signature identified by Ramaswamy et al from microarray analysis of unmatched primary adenocarcinomas and solid organ metastatic adenocarcinoma that consisted of seven up-regulated genes (SNRPF, EIF4EL3, HNRPAB, DHPS, PTTG1, COL1A2, and LMNB1) and eight down-regulated genes (ACTG2, MYLK, MYH11, CNN1, HLA-DPB1, MT3, NR4A1, and RBM5) {Ramaswamy, Ross, et al. 2002 3 /id}. From the IRB approved Clinical Breast Care Project tissue bank we identified five individuals with both invasive primary breast cancer and the corresponding lymph node metastasis. The clinical characteristics of these five patients are listed in Table 1. Samples underwent membrane based RNA isolation, two-step RT-PCR, and normalization to 18S ribosomal RNA. Primary tumor and lymph node metastasis expression was compared for each gene with 47 of 75 (63%) having greater than 2-fold difference in expression. These 47 data points were then subjected to binomial test to assess goodness-of-fit, with a correct fit defined as follows: those genes predicted by Ramaswamy et al to have increased or decreased expression in solid organ metastasis are found to be similarly expressed in the lymph node metastases. Table 2 details the accuracy of prediction of the binomial test for these 5 patients at the 47 points of data. Although, in no individual patient did the expression pattern of the lymph node metastasis perfectly match the signature for solid organ metastasis, when *all data points* were considered, calculation of the randomness versus accuracy revealed a p-value of .0395. That is to say, there is a 4% probability, that the frequency with which the gene expression pattern of lymph node metastases mimicked the pattern of solid organ metastases described by Ramaswamy et al, was by chance alone. This finding supports the hypothesis that solid organ metastases and lymph node metastases likely share a common molecular signature. This finding, in turn, allows us to better

conceptualize why the presence or absence of lymph node metastases has such a strong prognostic significance. Thus, if a patient has the phenotypic finding of lymph node metastasis then there is an increased probability cellular subpopulations within the primary tumor would have acquired the necessary genotype to allow for distant (i.e. solid organ) metastasis. Additionally, by extending the concept of shared molecular machinery between lymph node metastases and distant solid organ metastases, we hypothesize that it is possible that cells within the primary tumor could have evolved all the molecular machinery to permit systemic metastasis but failed to acquire those necessary for lymph node metastasis; just as the converse of this scenario is also possible. This could help explain why some patients with lymph node metastases will realize long-term survival whereas others without lymph node metastases will still succumb to distant disease. We plan to add additional primary tumor and lymph node metastasis samples as they become available and perform a similar experiment using primary breast tumors from patients with stage 1 and stage 4 disease.

**Table 1. Clinicopathologic data.** The ratio given in the LN row is the number of positive lymph nodes out of the total number of lymph nodes found in the specimens. The degree of differentiation was assessed on the samples processed for RNA isolation.

	TNM status	ER/PR	Differentiation	HER2-Neu
Primary 1	Stage 2b: T2N+M0	+/-	Poor	amplified
LN 1	1/12 +		Poor	
Primary 2	Stage 4: T2N+M+	-/-	Poor	not amplified
LN 2	30/54 +		Poor	
Primary 3	Stage 2b: T2N+M0	+/+	Moderate	not amplified
LN 3	26/38 +		Moderate	
Primary 4	Stage 2b: T2 N+M0	+/-	Well	not amplified
LN 4	4/24 +		Moderate	
Primary 5	Stage 2a: T1cN+M0	+/+	Poor	not amplified
LN 5	1/18 +		Poor	

Table 2. Goodness-of-fit to metastatic profile

	Correctly Predicted	Incorrectly Predicted	Patient Accuracy Prediction	Associated P-value
Patient 1	4	2	66.67%	.3437
Patient 2	6	6	50.00%	.6128
Patient 3	7	4	63.64%	.2744
Patient 4	5	2	71.43%	.2266
Patient 5	8	3	72.73%	.1133
Total	30	17	-	-

**GenBank accession numbers.** The following is a list of genes/GenBank accession numbers: SNRPF/AI032612, EIF4EL3/AF038957, HNRPAB/M65028, DHPS/U79262, PTTG1/AA203476, COL1A2/J03464, LMNB1/L37747, ACTG2/D00654, MYLK/U48959, MYH11/AF001548, CNN1/D17408, HLA-DPB1/M83664, MT3/S72043, NR4A1/L13740, and RBM5/AF091263.

**Acknowledgments**

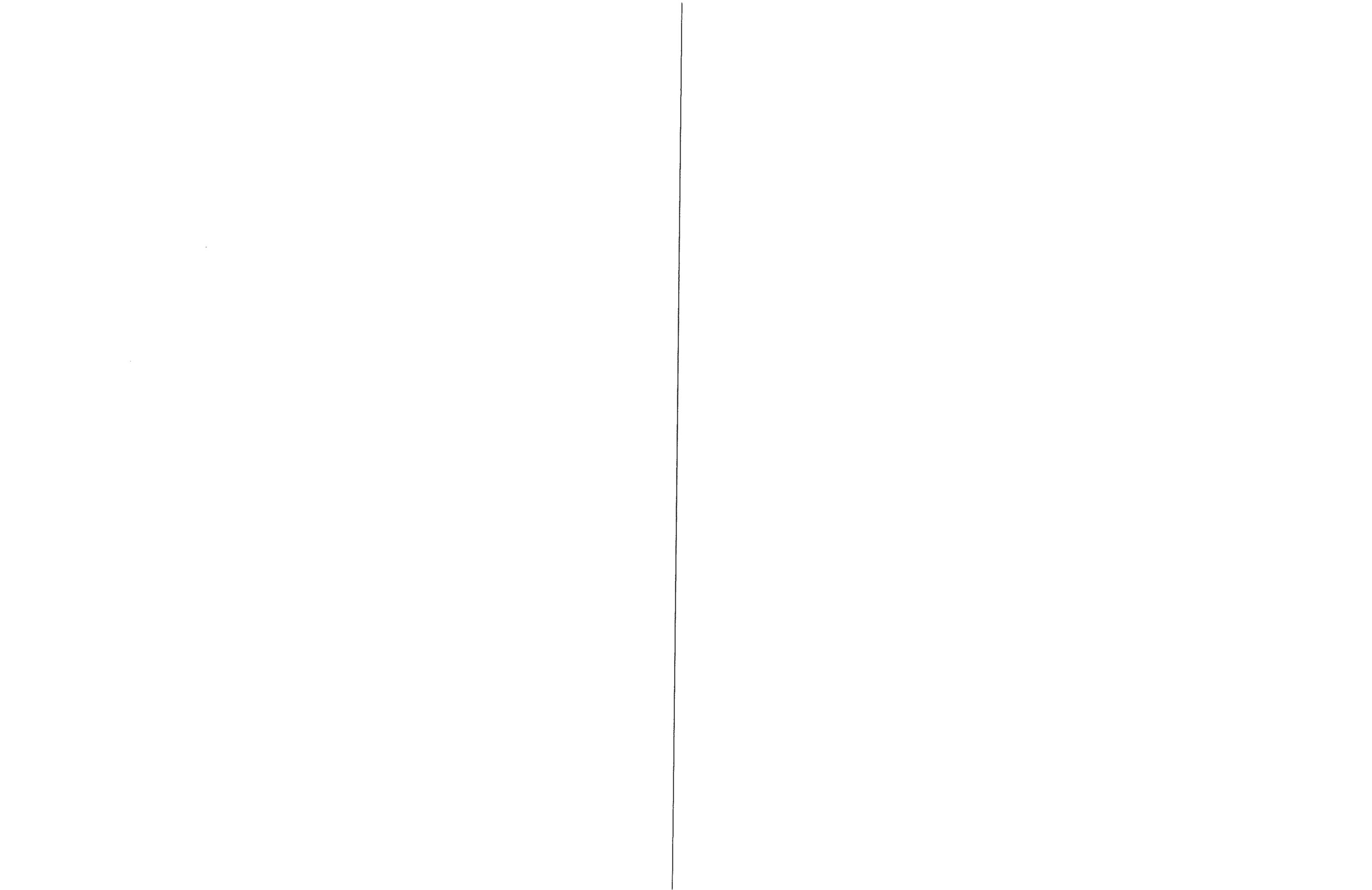
We would like to thank Jing He for her contribution to the study.

**Competing interest statement**

The authors declare that they have no competing financial interests.

**Shane Ottmann<sup>1,2</sup>, Rachel Ellsworth<sup>2</sup>, Brad Love<sup>3</sup>, Richard Somiari<sup>2</sup>, Jeffrey Hooke<sup>1</sup>, & Craig Shriver<sup>1,2</sup>**

<sup>1</sup>Clinical Breast Care Project, General Surgery Service, Walter Reed Army Medical Center, Washington, District of Columbia 20307, USA. <sup>2</sup>Windber Research Institute, Windber, Pennsylvania, 15963. <sup>3</sup>InforMax-An Invitrogen Company, Fredrick, Maryland, 21704. Correspondence should be addressed to S.O. ([ottmanns@mindspring.com](mailto:ottmanns@mindspring.com)).



**Gene expression differences in normal breast tissue from disease-free African American and Caucasian women.**

Lori A. Field<sup>1\*</sup>, Brad Love<sup>2</sup>, Jennifer Kane<sup>1</sup>, Brenda Deyarmin<sup>1</sup>, Jeffrey A. Hooke<sup>3</sup>, Rachel E. Ellsworth<sup>4</sup>, and Craig D. Shriver<sup>3</sup>

<sup>1</sup>Clinical Breast Care Project, Windber Research Institute, Windber, PA, 15963, USA

<sup>2</sup>Invitrogen, Carlsbad, CA, 92008, USA

<sup>3</sup>Clinical Breast Care Project, Walter Reed Army Medical Center, Washington, DC, 20307, USA

<sup>4</sup>Clinical Breast Care Project, Henry M. Jackson Foundation for the Advancement of Military Medicine, Windber, PA, 15963, USA

\*Corresponding author

LAF: [l.field@wriwindber.org](mailto:l.field@wriwindber.org)

BL: [drblove@yahoo.com](mailto:drblove@yahoo.com)

JK: [j.kane@wriwindber.org](mailto:j.kane@wriwindber.org)

BD: [b.deyarmin@wriwindber.org](mailto:b.deyarmin@wriwindber.org)

JAH: [jeffrey.hooke1@us.army.mil](mailto:jeffrey.hooke1@us.army.mil)

REE: [r.ellsworth@wriwindber.org](mailto:r.ellsworth@wriwindber.org)

CDS: [craig.shriver@us.army.mil](mailto:craig.shriver@us.army.mil)

## **Abstract**

**Introduction:** The incidence of breast cancer in young women (<40 years of age) is higher in African American women than in Caucasian women. Tumors from African American women are also more often characterized by an aggressive phenotype, including high grade, ER- status, and BRCA1-like features, suggesting that differences in tumor behavior and outcome between African American and Caucasian women have a molecular component. How and why these differences arise is unknown.

**Methods:** Gene expression profiles from normal breast tissue from disease-free African American (n=25) and Caucasian women (n=22) were compared using Affymetrix GeneChip HG U133A 2.0 arrays. Significantly differentially expressed genes ( $P<0.005$ ) were identified using a Mann-Whitney U test with non-parametric assumptions. Differential expression of selected genes was verified with quantitative real-time PCR using TaqMan Gene Expression assays.

**Results:** Forty-nine genes were differentially expressed between African American and Caucasian women ( $P<0.005$ ) with 32 genes having higher expression and 17 genes having lower expression in African American women when compared to Caucasian women. These genes function in signal transduction, metabolism, immune response, transport, and transcription. Further validation by quantitative real-time PCR confirmed that both *ACSM1* and *PSPHL* were more highly expressed in normal breast tissue from African American women while expression of *NTRK3* was higher in Caucasian women.

**Conclusions:** These results show that normal breast tissue from African American and Caucasian women have different gene expression profiles, suggesting that breast tissue is biologically different between the two groups. These inherent differences in gene

expression may affect cellular growth and proliferation, making African American women more prone to developing aggressive tumors at an earlier age than Caucasian women.

## Introduction

Each year in the United States, over 200,000 women are diagnosed with invasive breast cancer and an estimated 40,000 will die of this disease. However, not all ethnic groups are impacted similarly. The highest overall incidence occurs among Caucasian women (CW) with 132.5 breast cancer cases per 100,000 women while incidence among African American women (AAW) is lower with 118.3 cases per 100,000 women. When stratified by age, however, the highest incidence among women under age 40 occurs in African Americans. Mortality rates from breast cancer are also higher in AAW than CW with 33.8 deaths per 100,000 AAW versus 25.0 deaths per 100,000 CW [1].

These differences in survival between CW and AAW with breast cancer have been attributed to disparities in socioeconomic status, healthcare access, and cultural beliefs. However, several studies that controlled for confounding factors such as income, education, patient age, tumor stage at diagnosis, and treatment regimen found that mortality among AAW remained higher than that of CW [2,3]. Moreover, AAW continued to have poorer prognoses and survival within equal-access healthcare systems [4,5].

In addition to differences in clinical outcomes, tumor characteristics also vary between AAW and CW. Tumors from AAW tend to be larger and more aggressive than those in CW [3,6,7], and they often share many features with tumors from women with inherited mutations in BRCA1 [8], including poor differentiation, pushing margins, high-grade nuclear atypia, high mitotic index, and lymphocytic infiltrate [3,6,7,9,10,11]. Tumors from AAW are also more likely to have a “triple-negative” phenotype (estrogen

receptor (ER), progesterone receptor (PR), and HER2 negative) and are therefore, more difficult to treat because hormonal and targeted therapies are ineffective [7,12].

Together, these differences in outcome and pathology suggest that molecular factors contribute to differences in tumor etiology between AAW and CW. Several studies comparing breast tumors from AAW and CW have identified molecular differences. For example, the cell cycle regulatory proteins p53, p16, and cyclin E were all found to have higher expression while cyclin D1 had lower expression in AAW compared to CW [6,11]. In addition, researchers have recently reported that breast tumors from AAW and CW have different gene expression profiles [13,14]. While these studies support the notion that breast tumors from AAW and CW are molecularly different, they do not provide information as to why these aggressive tumors develop at an early age in AAW. To determine genetic factors that may contribute to tumor etiology in AAW, gene expression analysis was performed on disease-free breast specimens from AAW and CW.

## **Materials and Methods**

### *Tissue Acquisition*

Breast specimens were obtained from 25 AAW and 22 CW enrolled in the Clinical Breast Care Project (CBCP) at Walter Reed Army Medical Center (WRAMC) who had undergone a reductive mammoplasty and/or a breast biopsy and had no evidence of malignant disease. All tissue samples were collected with approval from the WRAMC Human Use Committee and Institutional Review Board. All subjects enrolled in the

CBCP voluntarily agreed to participate and gave written informed consent. Clinical and pathological information for each patient is shown in Table 1.

An additional 10 normal breast samples from 5 AAW and 5 CW without breast disease were acquired from the Mary Ellen's Tissue Bank located at Indiana University, Indianapolis, IN ([https://iubcrc.iupui.edu/ffl\\_admin/login.jsp](https://iubcrc.iupui.edu/ffl_admin/login.jsp)) for use in the validation studies by quantitative real-time PCR.

#### *RNA Isolation*

RNA was isolated from frozen breast specimens using the RNeasy Lipid Tissue Midi Kit (Qiagen Inc., Valencia, CA) with minor modifications. Approximately 400 mg of tissue (12  $\mu$ m sections) were homogenized per sample. Following isolation, the RNA was purified using the RNeasy Mini Kit (Qiagen Inc.) cleanup protocol. The concentration of each sample was determined with the NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific Inc., Waltham, MA) while the RNA integrity was monitored with the Bioanalyzer (Agilent Technologies, Inc., Santa Clara, CA).

#### *RNA Amplification*

1  $\mu$ g of total RNA from each sample was amplified using the MessageAmp II aRNA Amplification kit (Applied Biosystems Inc., Foster City, CA). GeneChip Eukaryotic Poly-A Controls (Affymetrix, Inc., Santa Clara, CA) were spiked into the total RNA samples prior to amplification to serve as a quality control measure for the amplification process. The first and second strand cDNA synthesis and cDNA

purification as well as the in vitro transcription (IVT) reaction and aRNA purification were performed according to the amplification kit instruction manual. The aRNA was labeled during the IVT reaction with 10 mM biotin-11-UTP (PerkinElmer, Inc., Waltham, MA). The concentration and size distribution of the biotin-labeled aRNA was determined on the NanoDrop ND-1000 and Bioanalyzer, respectively.

#### *Fragmentation, Hybridization, Staining, and Scanning*

Biotin-labeled aRNA (15 µg) was fragmented by incubation with 5X Fragmentation buffer (Affymetrix, Inc.) for 35 min. at 94° C. 10 µg of each fragmented sample was prepared for hybridization as specified in the Affymetrix GeneChip Expression Analysis Technical Manual for 100 Format (Midi) expression arrays and then hybridized to Affymetrix HG U133A 2.0 arrays for 16 hours. The arrays were washed, stained, and scanned according to Affymetrix protocols.

#### *Quantitative real-time PCR*

Total RNA was reverse transcribed to single-stranded cDNA using the High Capacity cDNA Reverse Transcription kit (Applied Biosystems Inc.). TaqMan Gene Expression assays (Applied Biosystems Inc.) were used to validate the differential expression of *PSPHL* (Hs00863464\_m1), *ACSM1* (Hs00369504\_m1), and *NTRK3* (Hs00176797\_m1). All sample reactions were performed in duplicate using the iCycler Real-Time Detection System (Bio-Rad Laboratories, Inc., Hercules, CA). *GAPDH* (Hs99999905\_m1) served as the endogenous control while FirstChoice® Human Brain

Reference RNA (Applied Biosystems Inc.) was used as the calibrator. The Comparative  $C_t$  Method was used to determine relative levels of gene expression.

#### *Statistical Analysis*

The microarray data were quantile normalized. To determine if a particular probe set was differentially expressed, the median normalized signal of each probe set was compared between the AAW and CW samples using a Mann-Whitney U test.  $P$ -values were calculated using standard non-parametric assumptions.

The heat map and clustering were created using those probes that had at least a 1.5 fold difference in expression between AAW and CW and a  $P$ -value  $<0.005$ . For each of these probes, the normalized signal for each sample was divided by the median normalized signal for that probe and the  $\log_2$  of the quotient calculated. Hierarchical clustering was performed using complete linkage with a correlation metric.

For the analysis of the quantitative real-time PCR (qRT-PCR) data, the medians of the  $2^{-\Delta\Delta C_t}$  values for the AAW and CW samples were compared using a Mann-Whitney U test to determine if the relative fold change was significantly different between the two groups.

## **Results**

### *Clinical characteristics of patient samples.*

Patient specimens used for this study were obtained from women enrolled in the Clinical Breast Care Project (CBCP) at Walter Reed Army Medical Center (WRAMC) who had no evidence of malignant breast disease. Forty-one of the patients had

undergone a breast biopsy while the remaining six had received a reductive mammoplasty. All histological slides were reviewed and diagnoses were made by a single, dedicated breast pathologist. Ten of the patients, including all six who had undergone a reductive mammoplasty, had no evidence of abnormal changes in their breast tissue while the remaining patients' breast specimens contained benign changes such as fibrocystic changes, microcalcifications, and mild intraductal hyperplasia (Table 1). Twenty-five of the patients were self-described African Americans while 22 of the patients were Caucasian (non-Hispanic). The majority of the patients were under the age of 50; 68% (15/22) of the CW were younger than 50 while 84% (21/25) of the AAW were below age 50 (Table 1).

*Normal breast tissue from African American versus Caucasian women displays different gene expression profiles.*

A comparison of the gene expression profiles of the normal breast samples from the AAW to those of the CW showed that 121 probe sets representing 112 genes were expressed at significantly different levels ( $P < 0.005$ ) in the two ethnic groups. Of these genes, 49 had at least a 1.5 fold difference in expression between the AAW and CW and are listed along with their biological function(s) in Table 2.

Figure 1 shows a heat map and a hierarchical cluster analysis of the 47 breast specimens. Using 53 probe sets that were significantly differentially expressed at  $P < 0.005$  and had at least a 1.5 fold or greater difference in expression between AAW and CW, this analysis was able to separate the normal breast samples from the AAW and CW into two distinct clusters.

*Validation of gene expression differences by quantitative real-time PCR.*

*PSPHL* and *ACSMI* exhibited the largest fold differences in expression between the normal breast samples from AAW compared to those from CW and therefore, were selected for further validation by qRT-PCR. To verify these results, 45 specimens from the original set of 47 samples along with an additional 10 out-of-sample disease-free breast specimens provided by the Mary Ellen's Tissue Bank (Indianapolis, IN) were evaluated.

As shown in Figure 2a, the significantly higher expression of *PSPHL* in normal breast tissue from AAW was confirmed by qRT-PCR analysis. The median fold change in expression of the AAW samples relative to the calibrator was 17.15 compared to a median value of 0 or no detectable transcript for the samples from CW ( $P=8.55 \times 10^{-7}$ ). The *PSPHL* transcript was undetectable in 58% of the Caucasian samples (15 out of 26 samples) compared to less than 7% (2 out of 29 samples) of the African American samples. The higher expression of *ACSMI* in AAW that was observed by microarray was also verified by qRT-PCR (Figure 2b). The median fold change in expression of *ACSMI* was 207.94 in the African American samples compared to 51.16 in the Caucasian samples ( $P=0.003$ ).

Conversely, expression of *NTRK3* was found to be lower in AAW. Two different microarray probes representing *NTRK3* were expressed at greater than 1.5 fold higher levels in CW than in AAW. As shown in Figure 2c, this difference in expression was validated by qRT-PCR. Median fold change in expression of *NTRK3* in CW was 0.27 whereas in AAW it was 0.13 ( $P=0.04$ ).

## Discussion

The contributions of differences in socioeconomic status and healthcare access on disparities in breast cancer outcomes between AAW and CW have been extensively studied [15]. In recent years, efforts have focused on identifying molecular differences that contribute to the aggressive tumor phenotype and less favorable prognoses in AAW [16,17]. To our knowledge, this is the first study to identify genetic differences in normal breast tissue from AAW and CW.

Here, we found that normal breast tissues from disease-free AAW and CW have inherent differences in gene expression. Our expression studies identified 49 genes at  $P < 0.005$  with at least a 1.5 fold difference in expression in normal breast tissue from AAW compared to CW. These genes were found to be involved in cellular metabolism, signal transduction, immune response, transcription, and cellular transport. These results suggest that breast tissue from AAW may be more metabolically active with higher proliferation rates than that from CW and therefore, may make AAW more susceptible to DNA damage and ensuing malignancy.

Acyl-CoA synthetase medium-chain family member 1, encoded by the *ACSM1* gene, is one of 26 acyl-CoA synthetases identified in the human genome [18]. Experimental evidence in mice suggests that ACSM1 may be important in the production of cellular energy in the form of ATP through the  $\beta$ -oxidation of medium-chain fatty acids in the mitochondrial matrix [19]. Consequently, changes in the expression of *ACSM1* could impact the formation of ATP and therefore, alter the rate of cell growth. The significantly higher expression of *ACSM1* in breast tissue from AAW relative to that

from CW could lead to increased energy production, resulting in increased cell growth and metabolism. This, in turn, may increase the likelihood that breast cells in AAW will acquire mutations that may lead to malignant transformation and subsequent tumor formation.

The phosphoserine phosphatase-like (*PSPHL*) gene was found to have the greatest fold difference in expression between AAW and CW with more than 10-fold higher expression in normal breast tissue from AAW. Recently, it has been reported that prostate tumors from African American men have higher expression of *PSPHL* than prostate tumors from Caucasian men [20]. In addition, the expression of *PSPHL* in the normal tissue surrounding the prostate tumor was also found to be higher in African American men than in Caucasian men. Similarly to what is observed for breast cancer, prostate tumors tend to be more aggressive with higher mortality in African American men than in Caucasian men. Thus, some genes, such as *PSPHL*, may contribute to the aggressive nature of both prostate and breast cancer in African American men and women.

Although the function of *PSPHL* is unknown, its higher expression in Fanconi anemia fibroblasts compared to normal fibroblasts [21] suggests a possible role in cell cycle progression and/or cellular proliferation. As such, higher expression of *PSPHL*, as is seen in breast tissue from AAW when compared to CW, may cause increased cellular proliferation that may support the development of tumors in the breast and/or increase the aggressiveness of breast tumors in AAW.

Unlike *ACSM1* and *PSPHL*, *NTRK3* expression was significantly lower in AAW. *NTRK3* encodes a membrane-bound neurotrophic tyrosine kinase receptor called TrkC.

Signaling through this receptor allows for the survival and differentiation of sensory neurons that sense limb movement and body position [22]. Somatic mutations in *NTRK3* have been detected in colorectal [23], lung [24], breast [25], and pancreatic [26] cancers. In addition, a study of 40 breast tumors found an inverse relationship between protein levels of TrkC and tumor grade, suggesting that higher expression of TrkC may serve as an indicator of good prognosis in breast cancer [27]. It is of note that the aggressive tumor phenotype associated with young AAW includes poor tumor differentiation, which may reflect decreased expression levels of TrkC.

The normal differentiation and development of tissues, including breast tissue, requires proper communication between the stroma and epithelia [28]. However, genetic and epigenetic changes in normal breast cells can alter this interaction between breast epithelial cells and the surrounding microenvironment and serve as an initiating event in tumor formation. Extensive work has chronicled the importance of the stroma in tumorigenesis [28,29,30]. Signaling from the stroma to the epithelia via secreted proteins such as chemokines and growth factors can promote cellular growth and proliferation, tumor initiation and progression, and metastasis [28,31]. However, normal stromal cells can also suppress tumor formation by inhibiting proliferation of cancer cells or preventing transformation to a fully malignant state [28,30]. Therefore, cross-talk between epithelial and stromal cells will determine whether a particular microenvironment is permissive for tumorigenesis. Thus, changes in gene expression in normal breast cells, such as those identified here, will presumably alter the tissue microenvironment and consequently affect its ability to suppress or promote tumor formation.

Each of the specimens used to generate the microarray data was collected from patients without breast cancer. The pathologic changes observed in the specimens (fibrocystic changes, microcalcifications, mild intraductal hyperplasia, etc.) are common findings in breast biopsies and are not associated with increased risk for the subsequent development of breast carcinoma. Thus, for purposes of this study, these specimens may be considered within the realm of physiologically normal tissue. To date, only one patient has received a subsequent diagnosis of breast cancer (Patient 22). Approximately 3 years after acquiring the normal breast specimen, this patient was diagnosed with invasive breast cancer in the contralateral breast.

Because aggressive breast cancer is associated with incidence in young AAW, we selected tissue from largely young (60% under age 40) or pre-menopausal (81% under age 50) patients for this study. Thus, the differences in normal breast biology between AAW and CW identified here may help determine what molecular factors predispose young AAW to a higher incidence of aggressive breast tumors.

## **Conclusions**

In this study, we identified gene expression differences in normal breast tissue from AAW and CW with no evidence of malignant breast disease. Many of the differentially expressed genes function in signal transduction, immune response, transport, and diverse metabolic pathways. The identification of these gene expression differences in normal breast tissue from AAW and CW is important to expanding our knowledge of basic breast biology and especially understanding the biological differences in breast tissue from AAW and CW. The gene expression differences identified here

may explain why AAW are more likely than CW to be diagnosed with breast cancer at a younger age and why tumors in AAW are often more aggressive. Further study of these genes in normal breast tissue as well as breast tumors may lead to the identification of new molecular targets and the development of novel therapies to treat the distinct molecular characteristics of breast tumors in AAW, resulting in improved clinical outcomes and minimizing the differences in breast cancer survival between AAW and CW.

#### **List of Abbreviations**

AAW – African American women

CBCP – Clinical Breast Care Project

C<sub>t</sub> – cycle threshold

CW – Caucasian women

ER – estrogen receptor

PR – progesterone receptor

qRT-PCR – quantitative real-time polymerase chain reaction

WRAMC – Walter Reed Army Medical Center

#### **Competing Interests**

The authors declare that they have no competing interests.

#### **Authors' Contributions**

LAF participated in the study design, performed the RNA isolations, microarrays, and qRT-PCR assays, and drafted the manuscript. BL performed the statistical analyses. JK and BD prepared the samples for RNA isolation. JAH pathologically reviewed and diagnosed all patient samples. REE conceived of the study and participated in its design and edited the manuscript. CDS participated in study design. All authors read and approved the final manuscript.

### **Acknowledgements**

Samples from the Mary Ellen's Tissue Bank (METB) were used in this study. We thank contributors, including Indiana University who collected samples used in this study, as well as patients and their families, whose help and participation made this work possible. We also thank Dr. Susan Clare who assisted us in obtaining samples from METB. This study was supported by the United States Department of Defense through the Henry M. Jackson Foundation of the Advancement of Military Medicine (MDA 905-00-1-0022 to CDS). Opinions and assertions contained herein are the private views of the authors and are not to be construed as official or as representing the views of the Department of the Army or the Department of Defense.

### **References**

1. **SEER Cancer Statistics Review, 1975-2004, National Cancer Institute**  
[[http://seer.cancer.gov/csr/1975\\_2004/](http://seer.cancer.gov/csr/1975_2004/)]
2. Joslyn SA, West MM: **Racial differences in breast carcinoma survival.** *Cancer* 2000, **88**:114-123.
3. Shavers VL, Harlan LC, Stevens JL: **Racial/ethnic variation in clinical presentation, treatment, and survival among breast cancer patients under age 35.** *Cancer* 2003, **97**:134-147.

4. Wojcik BE, Spinks MK, Optenberg SA: **Breast carcinoma survival analysis for African American and white women in an equal-access health care system.** *Cancer* 1998, **82**:1310-1318.
5. Jatoi I, Becher H, Leake CR: **Widening disparity in survival between white and African-American patients with breast carcinoma treated in the U.S. Department of Defense healthcare system.** *Cancer* 2003, **98**:894-899.
6. Porter PL, Lund MJ, Lin MG, Yuan X, Liff JM, Flagg EW, Coates RJ, Eley JW: **Racial differences in the expression of cell cycle-regulatory proteins in breast carcinoma.** *Cancer* 2004, **100**:2533-2542.
7. Amend K, Hicks D, Ambrosone CB: **Breast cancer in African-American women: differences in tumor biology from European-American women.** *Cancer Res* 2006, **66**:8327-8330.
8. Olopade OI, Fackenthal JD, Dunston G, Tainsky MA, Collins F, Whitfield-Broome C: **Breast cancer genetics in African Americans.** *Cancer* 2003, **97**:236-245.
9. Chen VW, Correa P, Kurman RJ, Wu X-C, Eley JW, Austin D, Muss H, Hunter CP, Redmond C, Sobhan M, Coates R, Reynolds P, Herman AA, Edwards BK: **Histological characteristics of breast carcinoma in blacks and whites.** *Cancer Epidemiol Biomarkers Prev* 1994, **3**:127-135.
10. Elledge RM, Clark GM, Chamness GC, Osborne CK: **Tumor biologic factors and breast cancer prognosis among white, Hispanic, and black women in the United States.** *J Natl Cancer Inst* 1994, **86**:705-712.
11. Jones BA, Kasl SV, Howe CL, Lachman M, Dubrow R, Curnen MM, Soler-Vila H, Beeghly A, Duan F, Wens P: **African American/white differences in breast carcinoma: p53 alterations and other tumor characteristics.** *Cancer* 2004, **101**:1293-1301.
12. Cleator S, Heller W, Coombes RC: **Triple-negative breast cancer: therapeutic options.** *Lancet Oncol* 2007, **8**:235-244.
13. Martin DN, Boersma BJ, Yi M, Reimers M, Yfantis H, Williams EH, Tsai YC, Stephens RM, Weinstein JN, Ambs S: **Gene expression analysis of African-American and European-American breast tumors [abstract].** In *First AACR Conference on the Science of Cancer Health Disparities in Racial/Ethnic Minorities and the Medically Underserved: 27-30 November 2007; Atlanta, GA.* Philadelphia, PA: AACR; 2007:35.
14. Baumbach LL, Ahearn ME, Gomez C, Jorda M, Silva O, Halsey TA, Yan J, Ellison K, Gluck S: **Identifying distinct multi-ethnic genome-wide alterations in breast cancer using paraffin embedded samples [abstract].** In *Proceedings of the 99th Annual Meeting of the American Association for Cancer Research: 12-16 April 2008; San Diego, CA.* Philadelphia, PA: AACR; 2008:208.

15. Newman LA, Griffith KA, Jatoi I, Simon MS, Crowe JP, Colditz GA: **Meta-analysis of survival in African American and white American patients with breast cancer: ethnicity compared with socioeconomic status.** *J Clin Oncol* 2006, **24**:1342-1349.
16. Morris GJ, Mitchell EP: **Higher incidence of aggressive breast cancers in African-American women: a review.** *J Natl Med Assoc* 2008, **100**:698-702.
17. Newman LA: **Breast cancer in African-American women.** *Oncologist* 2005, **10**:1-14.
18. Watkins PA, Maignel D, Jia Z, Pevsner J: **Evidence for 26 distinct acyl-coenzyme A synthetase genes in the human genome.** *J Lipid Res* 2007, **48**:2736-2750.
19. Fujino T, Takei YA, Sone H, Ioka RX, Kamataki A, Magoori K, Takahashi S, Sakai J, Yamamoto TT: **Molecular identification and characterization of two medium-chain acyl-CoA synthetases, MACS1 and the Sa gene product.** *J Biol Chem* 2001, **276**:35961-35966.
20. Wallace TA, Prueitt RL, Yi M, Howe TM, Gillespie JW, Yfantis HG, Stephens RM, Caporaso NE, A Loffredo CA, Ambs S: **Tumor immunobiological differences in prostate cancer between African-American and European-American men.** *Cancer Res* 2008, **68**:927-936.
21. Planitzer SA, Machl AW, Rueckels M, Kubbies M: **Identification of a novel c-DNA overexpressed in Fanconi's anemia fibroblasts partially homologous to a putative L-3-phosphoserine-phosphatase.** *Gene* 1998, **210**:297-306.
22. Nakagawara A: **Trk receptor tyrosine kinases: a bridge between cancer and neural development.** *Cancer Lett* 2001, **169**:107-114.
23. Bardelli A, Parsons DW, Silliman N, Ptak J, Szabo S, Saha S, Markowitz S, Willson JKV, Parmigiani G, Kinzler KW, Vogelstein B, Velculescu VE: **Mutational analysis of the tyrosine kinome in colorectal cancers.** *Science* 2003, **300**:949.
24. Davies H, Hunter C, Smith R, Stephens P, Greenman C, Bignell G, Teague J, Butler A, Edkins S, Stevens C, Parker A, O'Meara S, Avis T, Barthorpe S, Brackenbury L, Buck G, Clements J, Cole J, Dicks E, Edwards K, Forbes S, Gorton M, Gray K, Halliday K, Harrison R, Hills K, Hinton J, Jones D, Kosmidou V, Laman R, et al: **Somatic mutations of the protein kinase gene family in human lung cancer.** *Cancer Res* 2005, **65**:7591-7595.
25. Stephens P, Edkins S, Davies H, Greenman C, Cox C, Hunter C, Bignell G, Teague J, Smith R, Stevens C, O'Meara S, Parker A, Tarpey P, Avis T, Barthorpe A, Brackenbury L, Buck G, Butler A, Clements J, Cole J, Dicks E, Edwards K, Forbes S, Gorton M, Gray K, Halliday K, Harrison R, Hills K, Hinton J, Jones D, et al: **A**

**screen of the complete protein kinase gene family identifies diverse patterns of somatic mutations in human breast cancer.** *Nat Genet* 2005, **37**:590-592.

26. Wood LD, Calhoun ES, Silliman N, Ptak J, Szabo S, Powell SM, Riggins GJ, Wang T-L, Yan H, Gazdar A, Kern SE, Pennacchio L, Kinzler KW, Vogelstein B, Velculescu VE: **Somatic mutations of *GUCY2F*, *EPHA3*, and *NTRK3* in human cancers.** *Hum Mutat* 2006, **27**:1060-1061.
27. Blasco-Gutierrez M, Jose-Crespo I, Zozaya-Alvarez E, Ramos-Sanchez R, Garcia-Atares N: **TrkC: a new predictive marker in breast cancer?** *Cancer Invest* 2007, **25**:405-410.
28. Hu M, Polyak K: **Microenvironmental regulation of cancer development.** *Curr Opin Genet Dev* 2008, **18**:27-34.
29. Polyak K: **Breast cancer: origins and evolution.** *J Clin Invest* 2008, **117**:3155-3163.
30. Proia DA, Kuperwasser C: **Tumor agonist or antagonist.** *Cell Cycle* 2008, **4**:1022-1025.
31. Molloy T, van't Veer LJ: **Recent advances in metastasis research.** *Curr Opin Genet Dev* 2008, **18**:35-41.

#### Figure Legends

**Figure 1. Heat map and hierarchical clustering of normal breast tissue specimens from African American and Caucasian women.** The heat map and clustering were performed using 53 probe sets that were identified by gene expression profiling to be significantly differentially expressed between AAW and CW ( $P < 0.005$ ) with at least a 1.5 fold difference in expression. Each branch of the dendrogram and each column of the heat map represent a single patient sample. Red branches represent AAW while green branches represent CW. Each row of the heat map represents a different probe set and the gene represented by the probe is listed on the right side of the heat map. Red squares on the heat map represent higher levels of gene expression while green represents lower

expression. Black squares indicate no difference in expression relative to the median normalized signal intensity for that probe.

**Figure 2. qRT-PCR analysis of *PSPHL*, *ACSM1*, and *NTRK3* in normal breast tissue from African American and Caucasian women.** (a) Expression of *PSPHL* is higher in normal breast tissue from AAW than in CW. The median fold change in expression of *PSPHL* in AAW and CW was 17.15 and 0, respectively ( $P=8.55 \times 10^{-7}$ ). (b) *ACSM1* is expressed at higher levels in normal breast tissue from AAW than in that from CW. The median fold change in expression of *ACSM1* was 207.94 in AAW compared to 51.16 in CW ( $P=0.003$ ). (c) Expression of *NTRK3* is higher in CW than in AAW with median fold changes in expression of 0.27 and 0.13, respectively ( $P=0.04$ ).

**Table 1** Clinical characteristics of patients.

Sample	Age	Ethnicity	Diagnosis
1	56	Caucasian	Microcalcifications, fibrocystic changes
2	71	Caucasian	Microcalcifications, fibrocystic changes
3	46	African American	Fibrocystic changes
4	36	African American	Fibrocystic changes
5	48	African American	No abnormalities
6	50	Caucasian	Microcalcifications, fibrocystic changes, mild intraductal hyperplasia, radial scar
7	50	Caucasian	Microcalcifications, fibrocystic changes
8	25	African American	Fibrocystic changes, mild intraductal hyperplasia
9	20	Caucasian	Microcalcifications, fibrocystic changes, mild intraductal hyperplasia
10	54	African American	Mild intraductal hyperplasia, duct ectasia
11	26	Caucasian	Fibrocystic changes
12	47	Caucasian	Fibrocystic changes
13	56	African American	Fibrocystic changes, mild intraductal hyperplasia, duct ectasia
14	55	Caucasian	Fibrocystic changes, mild intraductal hyperplasia
15	56	Caucasian	No abnormalities
16	29	African American	Fibrocystic changes, mild intraductal hyperplasia
17	26	African American	Microcalcifications, fibrocystic changes
18	48	African American	Fibrocystic changes, mild intraductal hyperplasia
19	32	Caucasian	Fibrocystic changes
20	37	Caucasian	No abnormalities
21	63	African American	Microcalcifications, fibrocystic changes
22	34	African American	No abnormalities
23	39	Caucasian	Fibrocystic changes, mild intraductal hyperplasia
24	37	African American	Fibrocystic changes, lymphocytic mastitis

25	58	African American	Microcalcifications, fibrocystic changes, mild intraductal hyperplasia, lactational metaplasia
26	39	African American	Microcalcifications, fibrocystic changes, mild intraductal hyperplasia
27	47	Caucasian	Fibrocystic changes
28	23	African American	No abnormalities
29	42	Caucasian	Fibrocystic changes
30	33	African American	Fibrocystic changes
31	29	African American	Fibrocystic changes
32	36	African American	Microcalcifications, fibrocystic changes, mild intraductal hyperplasia, lactational metaplasia
33	22	Caucasian	Fibrocystic changes
34	43	African American	Microcalcifications, fibrocystic changes
35	47	African American	Fibrocystic changes
36	37	Caucasian	No abnormalities
37	34	African American	Fibrocystic changes
38	22	Caucasian	Fibrocystic changes
39	33	African American	Fibrocystic changes, fat necrosis
40	63	Caucasian	Microcalcifications, fibrocystic changes
41	34	Caucasian	No abnormalities
42	25	African American	Fibrocystic changes
43	33	Caucasian	No abnormalities
44	27	African American	No abnormalities
45	36	Caucasian	Fibrocystic changes
46	38	Caucasian	No abnormalities
47	19	African American	Fibrocystic changes

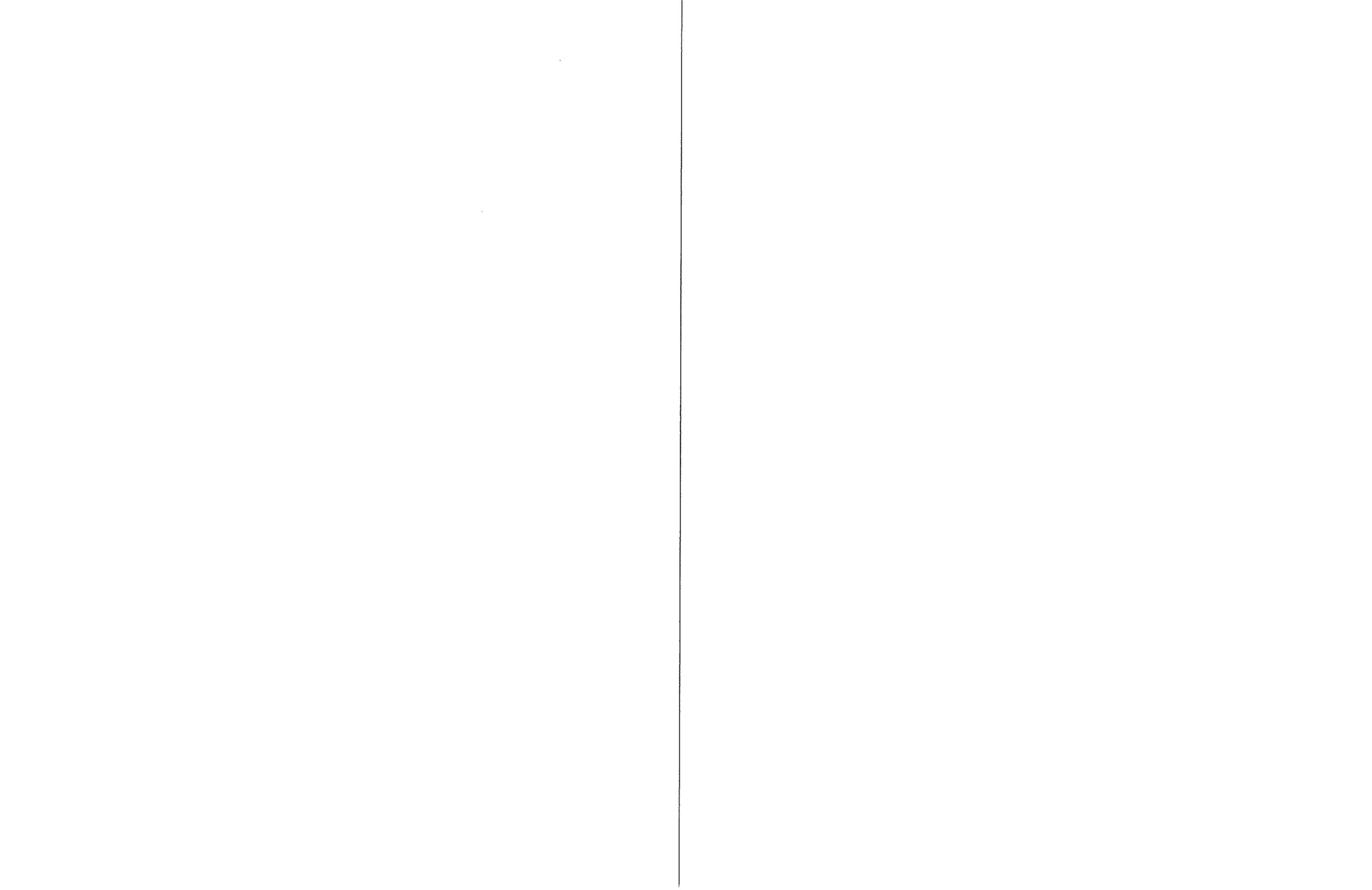
Table 2 Differentially expressed genes in normal breast tissue from African American and Caucasian women (>1.5 fold difference).  
 GSTM1, IGH, POLR2L14, and NTRK3 are represented by multiple probes.

*Genes with higher expression in normal breast tissue from African American women*

Gene	Accession Number	P-value	Fold Change	Cell adhesion and motility	Cell growth and proliferation	Cell differentiation	Signal transduction	Transcription	Immune response	Metabolism and catalytic activity	Transport	Other	Unknown
AAK1	BC002695	2.65x10 <sup>-3</sup>	1.68								X		
ABC11	AK024328	3.49x10 <sup>-3</sup>	1.81				X			X	X		
ACSM1	AC003024	1.91x10 <sup>-4</sup>	4.82				X			X	X		
ANFR	NM_001144	8.07x10 <sup>-3</sup>	1.54							X	X		
AUX46	NM_001630	3.74x10 <sup>-3</sup>	1.61	X						X			
C4BPA	NM_000715	9.56x10 <sup>-3</sup>	3.24						X				
CMAH	AF074480	1.26x10 <sup>-3</sup>	1.60										
COLEC10	NM_006438	4.89x10 <sup>-3</sup>	2.16									X	
CRYBB2	NM_000496	2.67x10 <sup>-3</sup>	2.60									X	
CUX2	AB006631	3.26x10 <sup>-3</sup>	2.85					X					
DKK1	NM_012242	1.44x10 <sup>-3</sup>	2.32				X			X			
GSTM1	X08020	2.84x10 <sup>-3</sup>	1.77							X			
GSTM1	NM_000561	3.74x10 <sup>-3</sup>	1.88							X			
GSTT2	NM_000854	4.79x10 <sup>-3</sup>	1.97						X				
IGH	M87789	2.67x10 <sup>-3</sup>	2.36										
IGH	L14486	4.57x10 <sup>-3</sup>	2.25		X						X		
IGHM	BC001872	1.99x10 <sup>-3</sup>	1.71				X						
LILRA1	AF025529	4.00x10 <sup>-3</sup>	2.27				X						
MNDA	NM_002432	2.65x10 <sup>-3</sup>	1.64					X				X	
MTRF1L	AL049952	2.55x10 <sup>-3</sup>	1.79									X	
NLRP3	AK027194	3.15x10 <sup>-3</sup>	2.98				X					X	
PI3	NM_002638	1.99x10 <sup>-3</sup>	2.18						X			X	
PLA2G4C	AF053214	5.11x10 <sup>-4</sup>	1.63				X					X	
POLR2J	BE676209	1.85x10 <sup>-3</sup>	1.52				X			X			
POLR2L14	AL047879	1.09x10 <sup>-3</sup>	1.57					X					X
POLR2L14	AA601208	4.98x10 <sup>-3</sup>	1.51										X
PSPHL	NM_009682	4.12x10 <sup>-3</sup>	10.39										X
PTCH2	NM_003738	2.22x10 <sup>-3</sup>	1.97				X						X
RPS26	NM_001029	4.26x10 <sup>-3</sup>	2.12									X	
SEZ6L	AB041736	2.06x10 <sup>-3</sup>	2.71									X	
SH2D1A	AF100540	1.36x10 <sup>-3</sup>	2.40				X						X
SOS1	L13857	4.66x10 <sup>-3</sup>	2.58				X		X				X
SPINK2	NM_021114	1.14x10 <sup>-4</sup>	3.78								X		
TCN1	NM_001062	8.48x10 <sup>-4</sup>	2.82					X					
ZNF221	NM_013359	1.24x10 <sup>-3</sup>	3.34					X					

*Genes with higher expression in normal breast tissue from Caucasian women*

Gene	Accession Number	P-value	Fold Change	Cell adhesion and motility	Cell growth and proliferation	Cell differentiation	Signal transduction	Transcription	Immune response	Metabolism and catalytic activity	Transport	Other	Unknown
ADI1	NM_018269	2.26x10 <sup>-4</sup>	1.51								X		
CA2	M36532	2.30x10 <sup>-3</sup>	1.80							X	X		X
CRYBB2P1	AA336200	2.48x10 <sup>-3</sup>	2.41							X			
CYP28B1	NM_019885	1.99x10 <sup>-3</sup>	2.00				X						
DAPK1	BC003674	3.61x10 <sup>-3</sup>	2.31							X			
HTR3B	NM_006028	3.04x10 <sup>-3</sup>	1.99				X				X		
IL20RA	NM_014432	2.84x10 <sup>-3</sup>	1.58				X						
LOC644450	AD50036	8.81x10 <sup>-4</sup>	3.17										X
LOC730092	AF054894	2.38x10 <sup>-3</sup>	1.70										X
NTRK3	S76476	9.51x10 <sup>-4</sup>	2.45			X						X	X
NTRK3	AL109696	9.51x10 <sup>-4</sup>	1.79			X						X	
PAX8	BC001060	1.49x10 <sup>-3</sup>	3.54			X						X	
PRL	NM_000948	1.92x10 <sup>-3</sup>	1.54				X					X	
RGS11	AB016929	3.61x10 <sup>-3</sup>	3.74		X		X					X	
SLC9A3R2	AF035771	5.10x10 <sup>-3</sup>	2.50								X		
TSGA10	NM_025244	1.72x10 <sup>-3</sup>	1.72									X	
VDR48	A806628	1.82x10 <sup>-3</sup>	1.65									X	
ZBTB7A	AW027070	3.26x10 <sup>-3</sup>	2.22					X					X



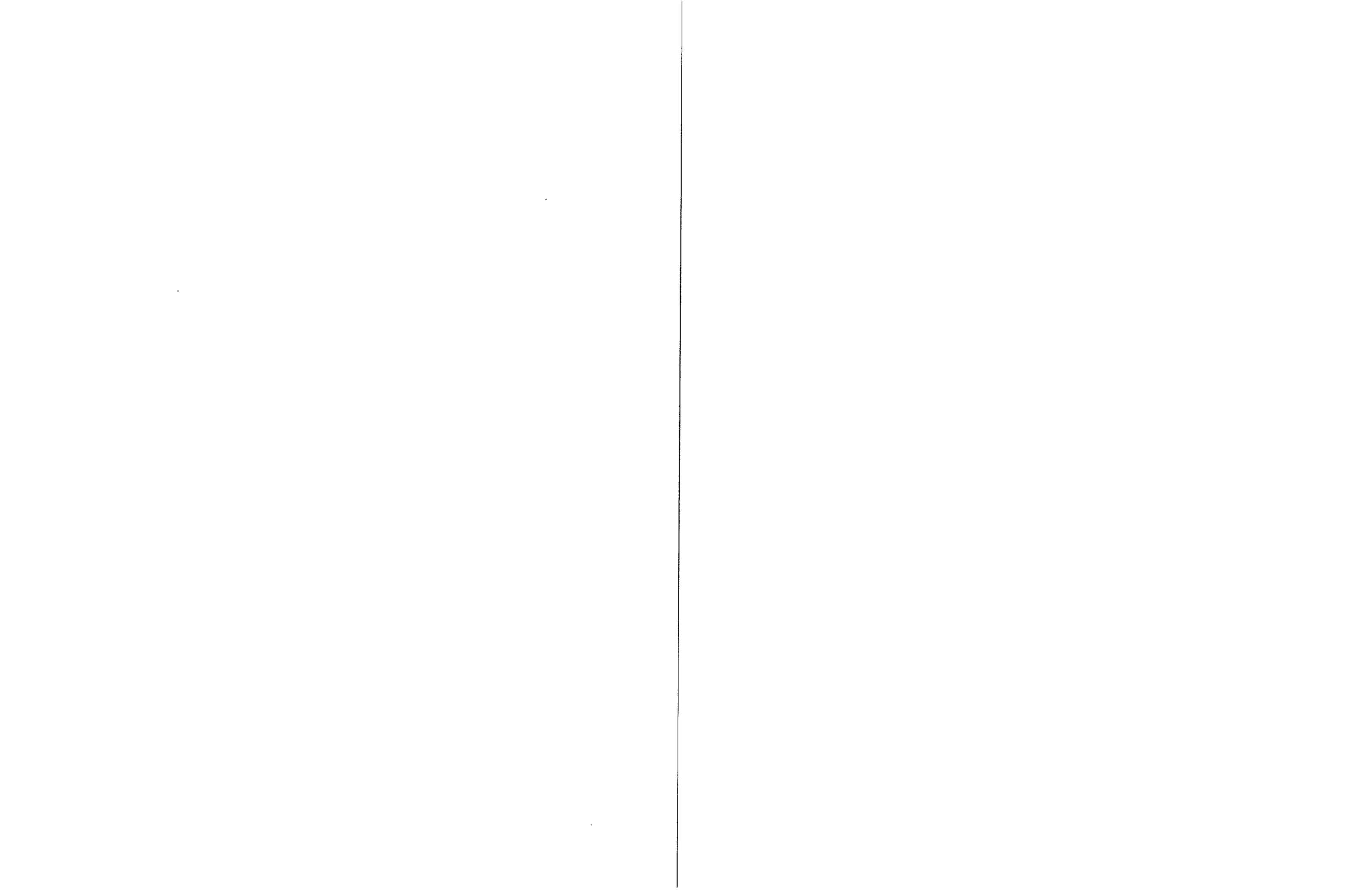
Genetic characterization of columnar cellular lesions and atypical ductal hyperplasia of the breast

Rachel E. Ellsworth<sup>1</sup>, Jamie D. Weyandt<sup>2</sup>, Jamie L. Fantacone-Campbell<sup>3</sup>, Brenda Deyarmin<sup>2</sup>, Darrell L. Ellsworth<sup>2</sup>, Jeffrey A. Hooke<sup>3</sup>, Craig D. Shriver<sup>3</sup>

**Aims:** Columnar cell lesions (CCL) and atypical ductal hyperplasia (ADH) frequently coexist and share molecular changes with *in situ* and invasive components, suggesting that CCL and ADH may be precursors to breast cancer. These conclusions are, however, largely based on studies examining CCL and/or ADH from patients diagnosed with more advanced disease. Thus, we assessed allelic imbalance (AI) in pure CCL or ADH specimens to characterize molecular changes in early breast lesions.

**Methods and results:** DNA samples were obtained from laser microdissected lesions from CCL with (n=42) or with ADH (n=31) without concurrent *in situ* or invasive disease. AI was assessed at 26 chromosomal regions commonly altered in breast cancer. The average AI frequency was 6.2% (range 0-20%) in CCL and 6.1% (range 0-25%) in ADH with no significant difference in levels of AI between CCL and ADH. The highest levels of AI were on chromosomes 8q24 and 17q21 in ADH (23%) and CCL (15%), respectively.

**Conclusions:** In conjunction with low overall levels of AI, chromosomal alterations that characterize low- and high-grade *in situ* and invasive disease were not frequent in pure CCL and ADH. Thus, pure lesions are not genetically advanced and are molecularly distinct from synchronous lesions.



A New Paradigm for Integrating Translational Research into Comprehensive Breast  
Care: The Walter Reed Army Medical Center / Windber Research Institute Experience

Lee Bronfman, RN, MA, CCRP  
Administrative Director, CBCP  
Walter Reed Army Medical Center  
6900 Georgia Avenue, NW  
Washington, D.C. 20307-5001

Craig D. Shriver, MD FACS, COL, MC  
Director and Principal Investigator, CBCP  
Program Director and Chief, General Surgery  
Walter Reed Army Medical Center  
6900 Georgia Avenue, NW  
Washington, D.C. 20307-5001

*The opinions and comments expressed in this article are those of the authors and do  
not necessarily represent the position of the Department of the Army or Department of  
Defense*

Reprint requests to:

Lee Bronfman, RN, MA, CCRP  
Administrative Director, CBCP  
Walter Reed Army Medical Center  
6900 Georgia Avenue, NW  
Washington, D.C. 20307-5001

**A New Paradigm for Integrating Translational Research into Comprehensive  
Breast Care: The Walter Reed Army Medical Center / Windber Research Institute  
Experience**

The Clinical Breast Care Project (CBCP) was established by direction of US Congress in January 2000. Human biospecimen collection (breast organ site) is a seminal pillar of the CBCP. To date, more than 33,000 total specimens have been collected and stored under a cooperative congressionally-sponsored agreement with the Windber Research Institute in Windber, Pennsylvania. The National Cancer Institute (NCI) sponsored a workshop in June 2008 entitled *Economic Considerations for Implementing the NCI Best Practices for Biospecimen Resources*. The authors were asked to present details from the CBCP experience on the cost of setting up and operating a high-level biorepository. These costs were meticulously gathered with input from our clinical, tissue banking and medical bioinformatics teams. Set up and operational costs were then computed for each facet of the operation to calculate the total expense of establishing this biorepository. Development and operating costs for other institutions will vary, but the chronology and costs presented here can serve as a guideline for other organizations, clinicians and scientists interested in establishing a biorepository for translational research.

Total word count - 173

**A New Paradigm for Integrating Translational Research into Comprehensive Breast Care: The Walter Reed Army Medical Center / Windber Research Institute Experience**

In establishing the basis and rationale behind our Clinical Breast Care Project (CBCP) biospecimen repository, we often refer to a paragraph from the 7<sup>th</sup> edition of the textbook, *Principles and Practice of Oncology* (1), which reads,

***"Identification of the important genetic derangements and the causally important genes and proteins depends on direct analysis of actual human cancer tissues, combined with insights gained using animal and cell culture methods".***

At the outset of our project development, we identified tissue banking for translational research as one of the five pillars upon which we would build the Clinical Breast Care Project. The four other interrelated pillars are Clinical Care, Targeted Research, Medical Bioinformatics and Risk Reduction. Envisioning a pristine collection of quality tissue with deep annotation of associated clinical, demographic and other epidemiologic data, we knew that rigid standard operating procedures for the collection, processing, shipping and storage of breast tissue, sera and lymph nodes would be an integral component of a robust and high-quality breast biorepository. In addition, the careful collection of extensive clinical data about each specimen and each donor would ensure the level of excellence required for the biomedical informatics pillar of the project (2).

The Clinical Breast Care Project at Walter Reed together with its partner the Windber Research Institute began collecting biospecimens under its Institutional Review Board (IRB) approved and novel research protocols in November 2001. As of February 13,

2009, 4056 individuals have donated 33,426 samples of tissue, lymph nodes and sera to the tissue biorepository of the Clinical Breast Care Project.

In June of 2007 the National Cancer Institute (NCI) published *Best Practices for Biospecimen Resources* (3). Recognizing the lack of standardization related to how specimens for research were collected, it was hoped that this publication would lead to greater uniformity in handling biospecimens. The response to this publication has been very positive. However, despite nearly unanimous agreement that such standards are necessary, there is widespread recognition that implementation comes with a price tag attached. To better understand the costs of operating a biorepository according to a strict set of standard operating procedures, and to further the process of developing an economic model for self-sustaining Biorepositories, the NCI Office of Biorepositories and Biospecimen Research sponsored a one day workshop in June 2008 entitled *Economic Considerations for Implementing the NCI Best Practices for Biospecimen Resources*. Experts and opinion leaders from government, industry, academia, and patient advocacy groups shared experiences and thoughts on how to build, maintain and operate the highest-quality (so called "platinum-level") biorepositories. Discussions were aimed at understanding the value proposition for institutions to procure high-quality biospecimens and explore the possibility for a platinum-level biorepository to operate on a cost-recovery, self-sustaining model.

Our group was asked to present our data and experience on the cost of setting up and operating a biorepository of the type and character of the CBCP. In tackling this task, we looked at what the CBCP had accomplished and determined set up costs over a two year period. We identified that during our developmental phase, we had devoted the

first six months to hiring a research-savvy individual with qualifications as a registered nurse or a post doctoral candidate. That individual (hereinafter referred to as the Administrator) would be charged with the responsibility, in consultation with the Principal Investigator, of writing the tissue banking protocol, the informed consent and HIPAA authorization. Once developed, those documents would begin their journey through a well respected Independent Review Board. The Administrator would then begin work on either developing or choosing an existing instrument to be used for clinical data acquisition. Salary and office supplies (computer, printer, fax, file cabinet, desk, chair, phone etc.) were computed as follows:

**Direct Cost: \$137,117**

**Indirect cost: \$20,842**

**Total cost: \$157,959**

In months six to twelve (6 – 12) a head nurse was added to staff. The Head Nurse would be responsible for:

- Developing Standard Operating Procedures in collaboration with the Administrator for such clinical activities as consenting subjects to the research, completing clinical questionnaires, blood acquisition etc.
- Developing an Orientation Program for new employees
- Developing list of supplies and equipment needed in collaboration with the Administrator

- Assist the Administrator in obtaining IRB approval (revisions of protocol, consent, impact statements, signature pages etc.)

The costs of this phase which included salary and office supplies (computer, printer, fax, file cabinet, desk, chair, phone etc.) were computed at:

**Direct Cost: \$123,649**

**Indirect Cost: \$18,795**

**Total Cost: \$142,444**

Months 12 -24 saw the addition to staff of another nurse, a part time pathologist and a data manager. Well organized office set ups were needed for additional staff including desks/cubicles, computers, printers, file cabinets, chairs, phones and office supplies as seen in Figure 2. Costs were computed at:

**Direct Cost: \$33,364**

**Indirect Costs: \$5,071**

**Total Cost: \$38,435**

A lab was set up for tissue acquisition and blood draw supplies.

**Direct Cost: \$44,299**

**Indirect Costs: \$6,733**

**Total Cost: \$51,032**

***Total cost for months 12-24 was \$646,070 and total CLINICAL set up costs for the first two years was calculated at \$946,474***

The next task was to determine the set up costs for the tissue bank. A senior management employee was required to oversee and manage banking operations. This person needs the assistance of an Associate/technician, and those direct and indirect costs resulted in a total for tissue banking personnel of \$256,143. Tissue bank capital equipment included a liquid nitrogen freezer and racks (see Figure 1), a liquid nitrogen freezer for back up, a centrifuge (refrigerated + rotors + adapters etc.), a minus 80 degree Centigrade freezer (includes back up) and a monitoring system. The direct and indirect costs amounted to a total of \$134,495. Other lab equipment and supplies such as refrigerator, liquid nitrogen tank for storage, racks for minus 80° freezer, small centrifuge (for microtubes, non refrigerated), pipettors, cryovials and reagents, factoring in direct and indirect costs totaled \$591,266. **Total Tissue Bank set up costs was calculated at:**

**Direct Cost: \$640,698**

**Indirect Costs: \$341,205**

**Total Cost: \$981,903**

The final set up costs to consider are the expenses associated with the development of a data warehouse. Personnel costs were figured at a direct cost of \$146,747, indirect costs of \$91,570 for a total of \$238,318. Four personal computers at \$8,000 with indirect costs of \$4992 added \$12,992 dollars to set up costs. Capital Equipment expenses for CLWS (Clinical Laboratory Workflow System) Genologics LIMS (tissue tracking system and clinical records) Inforsense (a clinical data tracking system and a database license from Oracle) were \$1,150,000. With indirect costs a total of \$1,165,600 was spent on capital equipment.

***Total Biomedical Informatics set up costs was calculated at:***

**Direct Cost: \$1,304,747**

**Indirect Costs: \$112,162**

**Total Cost: \$1,416,910**

Therefore, the overall set up costs for the clinical function, the tissue bank and the biomedical informatics components were calculated at a ***grand total of \$3,345,288.***

After learning what it cost to set up the program, our next challenge was to calculate operational (on-going) costs. We looked at tasks, amount of time to perform each task, the hourly rate of the person performing the task, the cost per action, and the indirect cost. For example, a nurse consents a patient to our core tissue protocol during the scheduled clinic visit. The patient is accompanied to an exam room, seen by a physician, scheduled for a biopsy, consented to the tissue protocol and receives pre-op teaching. This entire process was studied and found to take one hour, on average. The average hourly rate for a nurse in our clinic is \$42.41 and an indirect cost is applied in the amount of \$6.45. The Core Questionnaire, a 225 question instrument is completed by a nurse during the patient interview and takes approximately one additional hour to complete. Data managers perform a Quality Assurance review of each completed questionnaire, pick up and scan blood tubes and deliver blood specimens to the lab. Pathology researches the patient's prior history, picks up the tissue specimen from the operating room, performs gross dissection of tissue, performs slide reviews, completes a pathology checklist etc. Our careful and extensive analysis resulted in clinical

***Operational Costs being calculated at:***

**Direct Cost: \$342.79**

**Indirect Cost: \$52.10**

***Total Clinical Cost per patient accrued: \$394.90***

Operational costs for the bioinformatics function were calculated by identifying the percent of effort for personnel and their annual salaries. Data entry and Quality Assurance consumes 100 percent of effort for two data managers. The data tracking system consumed 30% of one individual's salary and 100% of time and effort for the bioinformatics specialist. The operational costs were calculated on 1000 patients per year and resulted in the following:

**Direct Cost: \$171,357**

**Indirect Cost: \$106,927**

**Total Cost: \$278,284**

**Divided by 1000 for *Total Bioinformatics Operational Costs per patient of \$278***

In calculating operational costs for the tissue bank we looked at the functions of receiving tissue, distributing and returning tissue, DNA extraction as requested for distribution, storage of specimens, amount of time for each function and the personnel involved. Direct and indirect costs for each function were calculated and the final numbers for each function are listed below:

**Total receiving costs (for one year with 826 total donors, 612 new donors, 2,553 total aliquots and 5,252 samples) \$18,737**

**Cost per patient: \$23**

**Total Storage Cost (represents storage of all CBCP specimens using 6 freezers as of June 2008) \$68,762**

Cost per patient: \$83

Tissue Distribution/DNA Isolation/Return: \$52,065

Cost per patient: \$63.

Tissue Bank Director (time calculated at 3hrs/week x 52 weeks) \$15,755

Cost per patient: \$19

Total Tissue Bank Operational Costs: \$155,319

***Tissue Bank Operational Cost per patient: \$188***

Therefore, our overall operational costs per patient are:

Clinical cost per patient: \$395

Biomedical Informatics Operational Cost per patient: \$278

Tissue Bank Operational Cost per patient: \$188

***Operational Cost per Patient \$861***

With 4056 patients enrolled at \$861 dollars per patient, we incurred a total expenditure of \$3,492,216 for operational costs, and \$3,345,288 for set up costs. The Clinical Breast Care Project spent, as of June 2008, a total of \$6,837,504 on its biospecimen breast biorepository over a course of 7 years.

Our experience is unique in its partnership of a military health care treatment facility and a civilian research institute associated with a community hospital. Set up and operating costs for other institutions will vary, but the chronology and costs presented here can serve as a guideline for other organizations, clinicians and scientists interested in

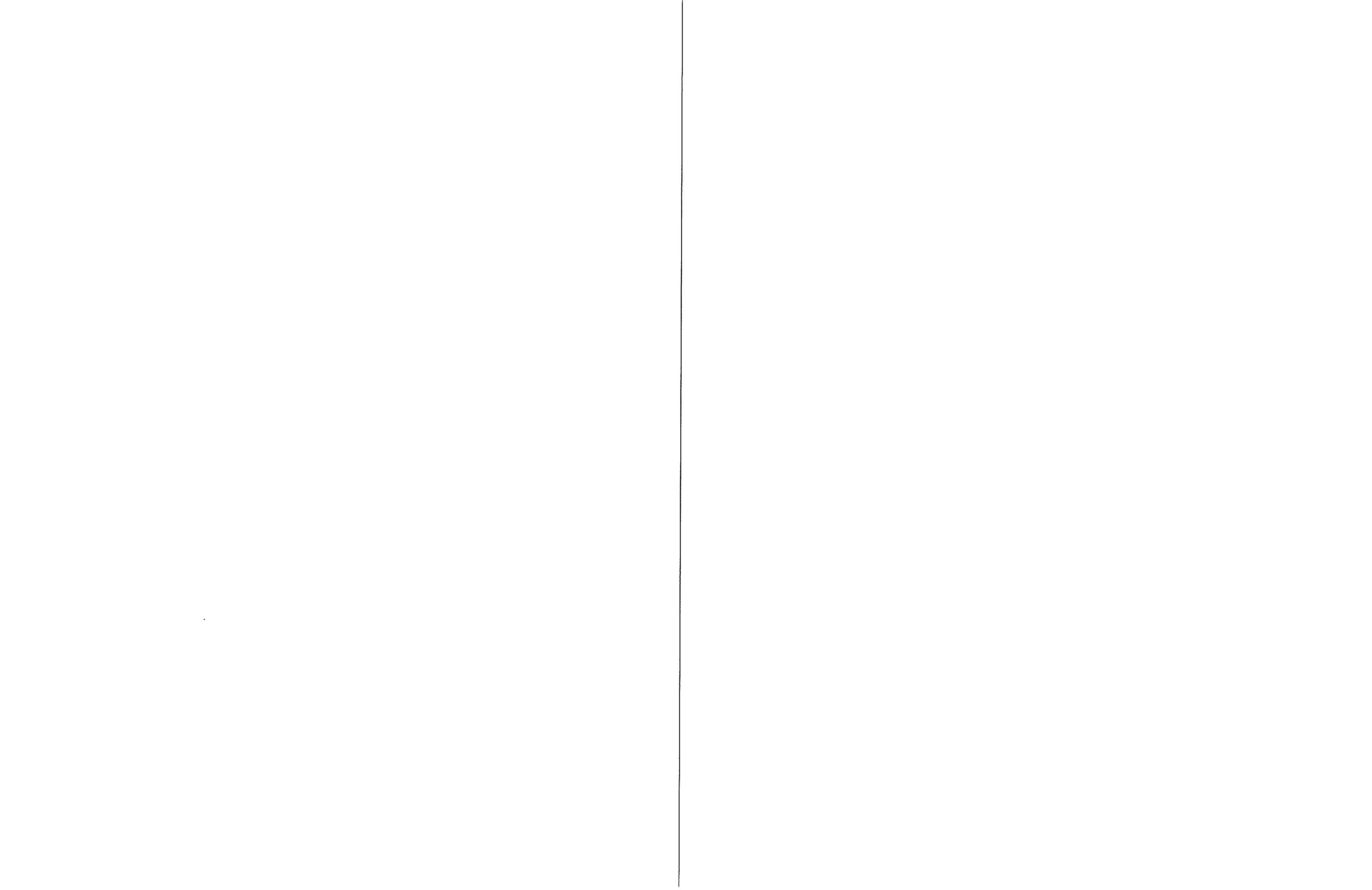
establishing a high-quality biorepository for translational research to accelerate progress in breast health and cancer care.

## References

1. DeVita V, Hellman S, Rosenberg S: Cancer Principles and Practice of Oncology 7<sup>th</sup> Edition. Philadelphia, PA, Lippincott, Williams & Wilkins, 2005
2. International Society for Biological and Environmental Repositories (ISBER). 2008 Best Practices for Repositories - Collection, Storage, Retrieval and Distribution of Biological Materials for Research, Second Edition, 2007. Available at: <http://www.isber.org/Pubs/BestPractices2008.pdf>. Last accessed March 28, 2009.
3. National Cancer Institute, NIH, US Department of Health and Human Services. NCI Best Practices for Biospecimen Resources, June 2007. Available at: [http://www.allirelandnci.org/pdf/NCI\\_Best\\_Practices\\_060507.pdf](http://www.allirelandnci.org/pdf/NCI_Best_Practices_060507.pdf). Last accessed March 28, 2009.

**Figure 1** Tissue Bank Equipment Including Liquid Nitrogen Freezers

**Figure 2 Biorespository Office Set Up at Windber Research Institute, Windber, PA**



Genomic instability of the breast tumor microenvironment: stromal alterations or technological artifact?

Seth Rummel, Allyson L. Valente, Jennifer L. Kane, Darrell L. Ellsworth, Craig D. Shriver, Rachel E. Ellsworth

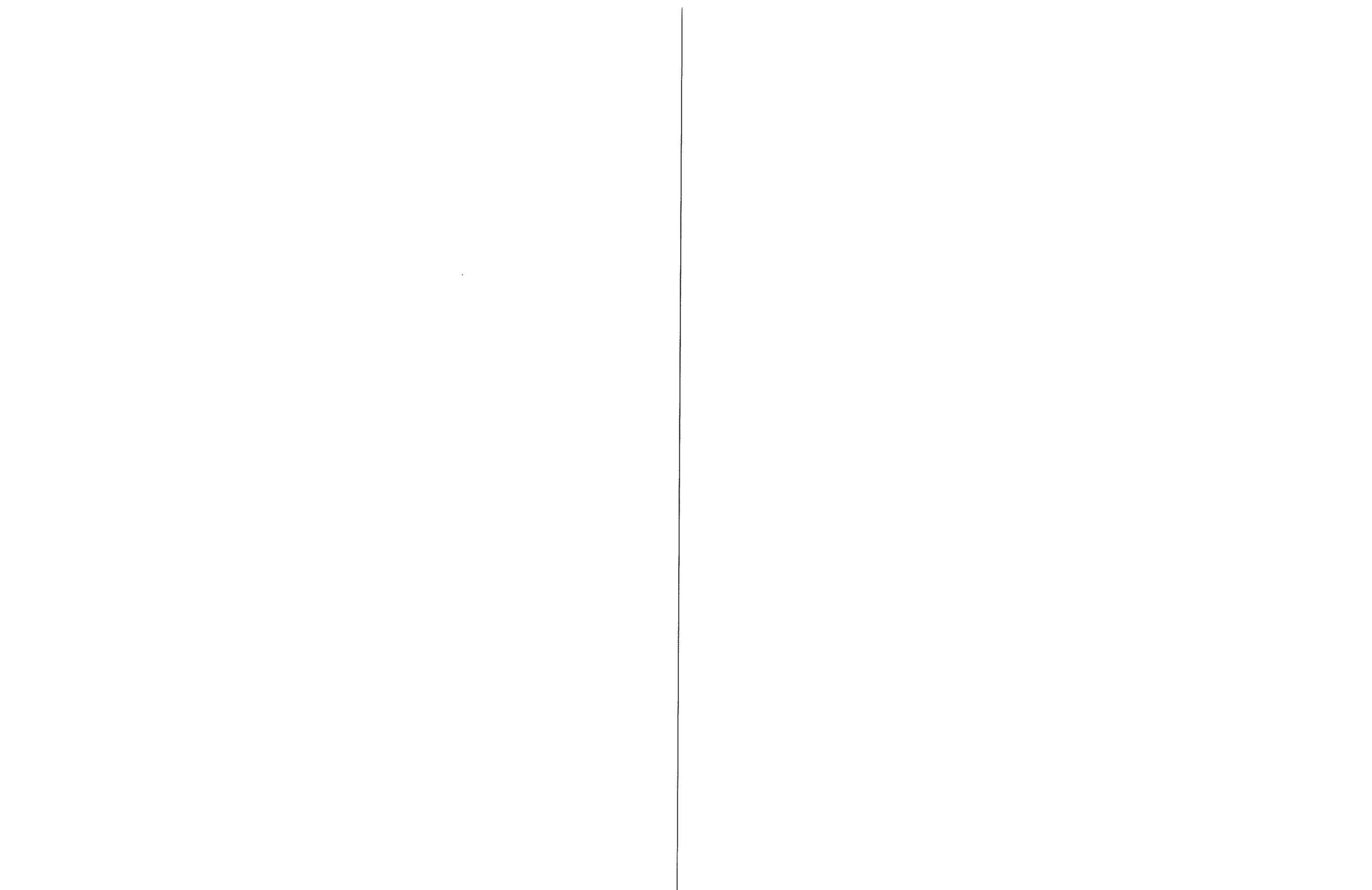
**Introduction:** The breast tumor microenvironment plays an active role in tumorigenesis. Molecular alterations, including epigenetic modifications to DNA, and changes in RNA and protein expression have been identified in tumor-associated stroma; however, there is considerable debate as to whether the stroma is characterized by genomic instability or whether detection of chromosomal alterations in the breast stroma is a reflection of technological artifact rather than the true genomic content of the tumor microenvironment.

**Methods:** Surgically-removed breast stroma specimens from 112 women undergoing reductive mammoplasty (n=7), prophylactic mastectomy (N=6) or mastectomy for a diagnosis of breast disease (n=99) were frozen in optimal cutting temperature medium. Allelic imbalance analysis was performed in 484 stromal specimens from 98 women using a panel of 52 microsatellite markers; SNP data was generated from a subset of 86 stromal specimens using 250K SNP arrays (Affymetrix). Copy number alterations were identified using Partek Genomics Suite.

**Results:** AI was not detected in 93% (444/484) of stroma specimens. When compared to AI generated from 77 FFPE stroma specimens, 32 (42%) of which harbored at least one detectable AI event, the frequency of AI in the FFPE specimens (4.62%) was

significantly higher ( $P < 0.0001$ ) than that found in frozen specimens (0.45%). Of the stroma specimens assayed using SNP arrays 95% (82/86) had no detectable alterations and the 11 copy number changes were small and not shared between specimens.

**Conclusions:** The data presented here support a model in which the tumor microenvironment is genetically stable. The direct comparison of copy number alterations between FFPE and frozen research-grade specimens using identical methodologies suggests that past reports of significant AI/LOH in breast stroma, both adjacent to and distant from the tumor, reflects artifact in the archival specimens caused by formalin-fixation, paraffin-embedding and tissue storage.



## Proteomic discrimination of well- from poorly-differentiated breast carcinomas

RE Ellsworth, EH Seeley, DL Ellsworth, B Deyarmin, JA Hooke, M Sanders, R Caprioli, CD Shriver

**Background:** Pathological grade is a useful prognostic factor for stratifying breast cancer patients into favorable (well-differentiated tumors) and less favorable (poorly-differentiated tumors) outcome groups. The current system of tumor grading, however, is highly subjective and a large proportion of tumors are characterized as intermediate-grade, making determination of optimal treatments difficult.

**Methods:** Primary breast tumor specimens from patients diagnosed with well- (n=27) and poorly-differentiated (n=51) invasive ductal carcinoma were obtained from patients enrolled in the Clinical Breast Care Project. Frozen tissues were sectioned and mounted on gold coated MALDI target plates for protein expression profiling. Hematoxylin and eosin (H&E) stained slides were prepared from serial sections for histological characterization. MALDI matrix was deposited as individual spots on the tissue sections in a histology directed manner to assay specific areas and tissue types of interest. Mass spectral data were then acquired from multiple sites across each tissue section.

**Results:** 129 features were observed in well-differentiated and 132 in poorly-differentiated tumors. While the majority of features detected were similar between the two groups, 6 protein features were expressed at significantly lower and 12 at significantly higher levels in the poorly-differentiated tumors, including increased expression of Calgranulin A and Calgizzarin.

**Conclusions:** Protein expression differences detected here suggest that well- and poorly-differentiated invasive breast tumors are molecularly distinct diseases and that these protein changes may contribute to the structural integrity of the tumor cell. In particular, calgranulin A and calgizzarin are members of the S100 protein family, and function in processes such as cell proliferation and differentiation. Further refinement of this differentiation protein signature may not only improve our understanding of the biological processes involved with tumor grade but provide pathologists with new molecular tools to classify breast tumors and reduce the subjectivity associated with current grading criteria.

## Copy Number Variation in Low- and High-Grade Breast Tumors

Rachel E. Ellsworth<sup>1</sup>, Heather L. Patney<sup>2</sup>, Jennifer L. Kane<sup>2</sup>, Jeffery A. Hooke<sup>3</sup> and Craig D. Shriver<sup>3</sup>

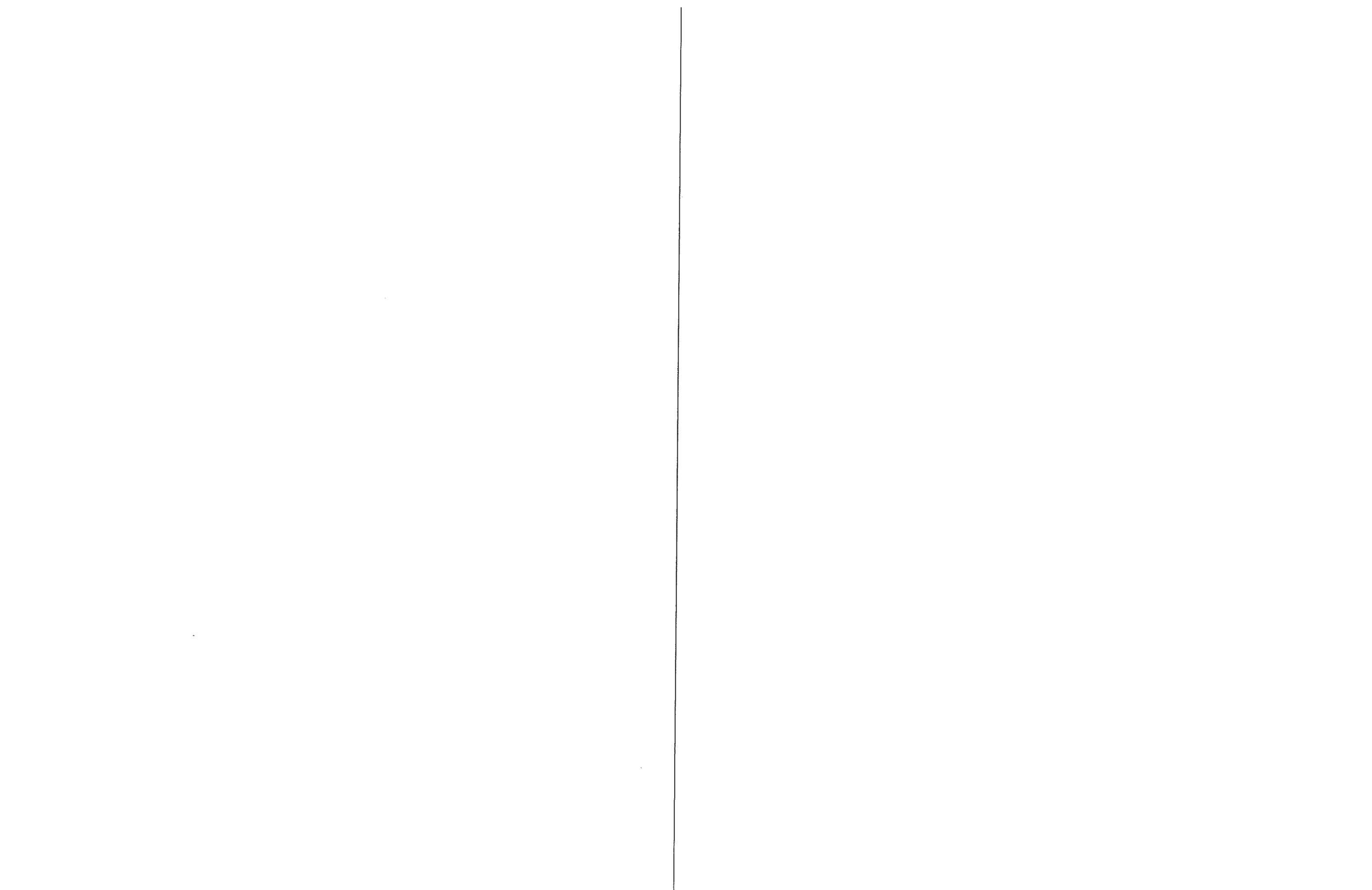
<sup>1</sup>Henry M. Jackson Foundation for the Advancement of Military Medicine, Windber, PA, <sup>2</sup>Windber Research Institute, Windber, PA and <sup>3</sup>Walter Reed Army Medical Center, Washington D.C.

**Objective** Histological grade classifies breast carcinomas into low-, intermediate- and high-grade. Developmental pathways for high-grade disease remain controversial; clinical data have been used to describe both linear models, in which high-grade tumors evolve from low-grade precursors, as well as models that depict low- and high-grade carcinomas as distinct molecular diseases. Genome-wide copy number alterations were examined in low- and high-grade breast tumors to identify genetic changes associated with the etiology of low- and high-grade breast carcinomas.

**Methods** Tumors were diagnosed and characterized by a single, dedicated breast pathologist. DNA was isolated after laser microdissection from low- (n=53) and high-grade (n=68) tumors. Genotype data were generated using Human Mapping 250K Sty arrays (Affymetrix). Copy number alterations and LOH were detected using GenomeConsole 3.0.2 (Affymetrix).

**Results** The most common alterations in low-grade carcinomas were gain of chromosomes 1q (75%) and 16p (58%) and loss of chromosome 16q (86%). These alterations were frequently simple and included the entire chromosomal arm. In contrast, patterns of alterations were complex in high-grade tumors, frequently with alternating gains and losses on a single chromosomal arm. In high-grade tumors, the most frequent alterations included gain of chromosome 1q (85%), 8q (84%), and loss of chromosomes 8p (54%), 14q (57%) and 17p (69%). Loss of 16q ( $P < 0.0000005$ ) and gain of 16p ( $P < 0.005$ ) occurred significantly more frequently in low- compared to high-grade tumors, while gain of 8q24 ( $P < 0.000005$ ) and 10p15-p14 ( $P < 0.00005$ ) and loss of chromosome 17p13 ( $P < 0.00005$ ) occurred significantly more frequently in high- compared to low-grade tumors. Only two samples, one low- and one high-grade, had a “flat” profile with no detectable chromosomal alterations.

**Conclusion** Low- and high-grade breast tumors are genetically different. Gain of chromosome 1q is common in both low- and high-grade tumors, and may be an early event in tumorigenesis. Given the significantly different patterns of alterations, especially significantly higher loss of chromosome 16q in high- compared to low-grade tumors, high-grade disease may not arise from dedifferentiation of low-grade but rather represent a separate disease defined by unique genetic alterations.



**OPEN ACCESS**  
Full open access to this and  
thousands of other papers at  
<http://www.la-press.com>.

## Genomic Heterogeneity of Breast Tumor Pathogenesis

Rachel E. Ellsworth<sup>1</sup>, Jeffrey A. Hooke<sup>2</sup>, Craig D. Shriver<sup>2</sup>, and Darrell L. Ellsworth<sup>3</sup>

<sup>1</sup>Clinical Breast Care Project, Henry M. Jackson Foundation for the Advancement of Military Medicine, 620 Seventh Street, Windber, PA, 15963, USA. <sup>2</sup>Clinical Breast Care Project, Walter Reed Army Medical Center, 6900 Georgia Avenue, NW, Washington, DC, 20307, USA. <sup>3</sup>Clinical Breast Care Project, Windber Research Institute, 620 Seventh Street, Windber, PA, 15963, USA. Email: [d.ellsworth@wriwindber.org](mailto:d.ellsworth@wriwindber.org)

**Abstract:** Pathological grade is a useful prognostic factor for stratifying breast cancer patients into favorable (low-grade, well-differentiated tumors) and less favorable (high-grade, poorly-differentiated tumors) outcome groups. Under the current system of tumor grading, however, a large proportion of tumors are characterized as intermediate-grade, making determination of optimal treatments difficult. In an effort to increase objectivity in the pathological assessment of tumor grade, differences in chromosomal alterations and gene expression patterns have been characterized in low-grade, intermediate-grade, and high-grade disease. In this review, we outline molecular data supporting a linear model of progression from low-grade to high-grade carcinomas, as well as contradicting genetic data suggesting that low-grade and high-grade tumors develop independently. While debate regarding specific pathways of development continues, molecular data suggest that intermediate-grade tumors do not comprise an independent disease subtype, but represent clinical and molecular hybrids between low-grade and high-grade tumors. Finally, we discuss the clinical implications associated with different pathways of development, including a new clinical test to assign grade and guide treatment options.

**Keywords:** breast cancer, tumor grade, genetic pathways

*Clinical Medicine: Oncology* 2009;3 77–85

This article is available from <http://www.la-press.com>.

© the authors, licensee Libertas Academica Ltd.

This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://www.creativecommons.org/licenses/by/2.0>) which permits unrestricted use, distribution and reproduction provided the original work is properly cited.



## Introduction

Breast cancer is the most common cancer in women worldwide, and in the United States (US) is estimated to account for ~26% of all new female cancer cases and 15% of all cancer deaths among women.<sup>1</sup> Incidence of breast cancer in the US has risen by approximately 1.2% per year since 1930,<sup>2</sup> such that one in eight American women now are expected to develop breast cancer during her lifetime. Research attempting to understand the molecular nature of breast cancer and its progression will have a tremendous impact on costs associated with disease, and importantly, on the nature of diagnosis, treatment, and prevention.

Pathological assessment of breast cancer is currently based on criteria such as tumor size, lymph node and hormone receptor status, and epidermal growth factor receptor 2 (HER2) expression, but pathology alone does not accurately predict outcomes, even for patients with similar tumor characteristics. Recent studies suggest that, despite use of identical treatment modalities in patients with similar pathological characteristics, clinical outcomes can be highly variable.<sup>3</sup> Differences in response to treatments such as Tamoxifen and Herceptin® likely reflect heterogeneity in pathological factors such as estrogen receptor (ER) and HER2 status. Breast carcinomas are heterogeneous at the molecular level, with at least five disease categories identified through differential patterns of gene expression.<sup>4-6</sup> This extensive clinical, pathological, and molecular heterogeneity complicates diagnosis, prognosis, and treatment of patients with breast cancer.

In this review, we examine use of the Nottingham histological score in assigning grade to breast carcinomas and the clinical utility of the Nottingham score in determining patient risk and outcome. We outline our understanding of how genomic alterations contribute to histological characteristics that define tumor grade and the importance of molecular changes in shaping tumor growth and differentiation in patients with breast cancer. An important focus of this review is the ongoing debate over development of high-grade and low-grade breast disease, specifically on whether low-grade and high-grade breast carcinomas represent separate and distinct diseases. We present molecular evidence supporting a linear model of progression from low-grade to high-grade carcinomas, as well as contradicting genetic data suggesting that low-grade

and high-grade tumors develop independently. While debate regarding specific pathways of development continues, molecular data suggest that intermediate-grade tumors do not comprise an independent disease subtype, but represent clinical and molecular hybrids between low-grade and high-grade tumors. Finally, we discuss the clinical implications of different pathways of development, including a new clinical test to assign grade and guide treatment options.

## Nottingham Histological Score

The Nottingham combined histological grading system, based on classification parameters developed by Bloom and Richardson<sup>7</sup> as modified by Elston and Ellis,<sup>8</sup> is currently the most widely used method for assessing breast tumor grade. The Nottingham score uses three components, tubule formation, nuclear pleomorphism, and mitotic count, each of which are scored independently using criteria described in Table 1.<sup>9</sup> Scores for the three components are then combined and the cumulative score serves as the classifier: low-grade (well-differentiated) tumors have a cumulative score of 3, 4, or 5; intermediate-grade (moderately-differentiated) tumors score 6 or 7; and high-grade (poorly-differentiated) tumors have cumulative scores of 8 or 9.

The Nottingham grading system has clinical utility in determining patient risk and outcome—patients with low-grade carcinomas have ~95% five-year survival compared to just 50% in patients with high-grade disease.<sup>8,10</sup> Although the prognostic power of the Nottingham score has prompted the College of American Pathologists to suggest using grade during staging,<sup>11</sup> grade has not yet been incorporated as a component of tumor staging.<sup>12</sup> Use of grade is impaired by 1) the inherent subjectivity associated with its assessment—concordance between pathologists ranges from 50%–85%,<sup>13</sup> and 2) the large number (30%–60%) of tumors classified as intermediate-grade (moderately-differentiated). These tumors have features of both low-grade and high-grade tumors, making it difficult to assess risk and determine the most appropriate treatment option for patients.<sup>14</sup>

## Genomic Discrimination of Low- and High-Grade Breast Carcinomas

Breast cancer progression can be defined by a non-obligatory sequence of histological changes from



normal epithelium through atypical hyperplasia, *in situ* carcinoma, and finally invasive malignancy.<sup>15</sup> The hypothesis of dedifferentiation posits that breast cancers evolve from well-differentiated to poorly-differentiated tumors following a linear model. The progressive sequence of dedifferentiation is: well-differentiated (grade 1) → moderately-differentiated (grade 2) → poorly-differentiated (grade 3). Support for a link between histological progression and tumor growth comes mainly from clinical studies, which have identified correlations between histological grade and tumor size,<sup>16</sup> or observed that palpable carcinomas detected by mammography tend to be well-differentiated.<sup>17</sup> In contrast, observations that recurrent carcinomas tend to exhibit the same level of cellular differentiation, and hence the same histological grade, as the original primary tumor<sup>18</sup> have led to the hypothesis that low-grade and high-grade carcinomas reflect different disease entities. Although certain DNA copy number changes defined by comparative genomic hybridization (CGH) correlate with the degree of histological differentiation,<sup>19</sup> several molecular studies suggest that the majority of

low-grade (well-differentiated) tumors do not progress to high-grade (poorly-differentiated) carcinomas. For example, Roylance et al<sup>20</sup> observed distinct genomic differences between grade I and grade III breast tumors; in particular, loss of chromosome 16q was significantly more frequent in grade I (65%) compared to grade III (16%) tumors. Likewise, Buerger et al<sup>21</sup> observed frequent loss of chromosome 16q in well-differentiated invasive breast carcinomas and concluded that sequential progression from low-grade to high-grade is unlikely because chromosomal alterations at 16q were not maintained in higher-grade tumors.

Since the initial models of disease progression were published, a number of studies examining levels and patterns of genomic variation in breast carcinomas have supported the hypothesis that low-grade and high-grade tumors represent separate genetic diseases, based largely on observations that the frequency of alterations at chromosome 16q was significantly higher in low-grade tumors. An allelic imbalance (AI) analysis using three microsatellite markers on chromosome 16q detected a significantly higher frequency of AI events in low-grade (grade 1) compared to high-grade (grade 3) tumors for two of the three markers.<sup>22</sup> Likewise, microsatellite-based data from our own group showed significantly higher levels of AI at chromosome 16q11-q22 in low-grade compared to high-grade breast carcinomas.<sup>23</sup> In addition, low-grade tumors contained larger alterations across the 23 Mb region of chromosome 16 compared to high-grade tumors. Only proximal markers (D16S409 and D16S2624) on 16q had a higher frequency of AI in grade 1 versus grade 3 tumors, suggesting that changes in the 16q11-q22 region are critical in the development of low-grade disease (Fig. 1). Similarly, an assessment of copy number status across chromosome 16q by CGH, AI, and fluorescence *in situ* hybridization (FISH) demonstrated that low-grade and high-grade disease were associated with different types of chromosomal alterations in the 16q region.<sup>24</sup> Physical loss of large portions of chromosome 16q was associated with low-grade disease, while small regions of loss of heterozygosity (LOH) were characteristic of high-grade tumors. Further, the timing of alterations at 16q appeared to differ between tumor grades, with physical loss of 16q being an early and critical event in the development of low-grade breast tumors, while smaller alterations of 16q occurred late in the development of high-grade carcinomas.<sup>24</sup> Together, these studies

**Table 1.** Criteria for histological grading of invasive breast carcinomas.<sup>a</sup>

Component	Score	Description
Tubule formation <sup>b</sup>	1	Majority of tumor (>75%)
	2	Moderate degree (10%–75%)
	3	Little or none (<10%)
Nuclear pleomorphism	1	Small, uniform nuclear size and shape
	2	Modest increase in size and variation
	3	Large with marked variation
Mitotic counts <sup>c</sup>	1	≤7
	2	8–16
	3	≥17

<sup>a</sup>Scores for the three components are combined and the cumulative score classifies breast tumors as: low-grade (well-differentiated) tumors, 3, 4, or 5; intermediate-grade (moderately-differentiated) tumors, 6 or 7; high-grade (poorly-differentiated) tumors, 8 or 9.

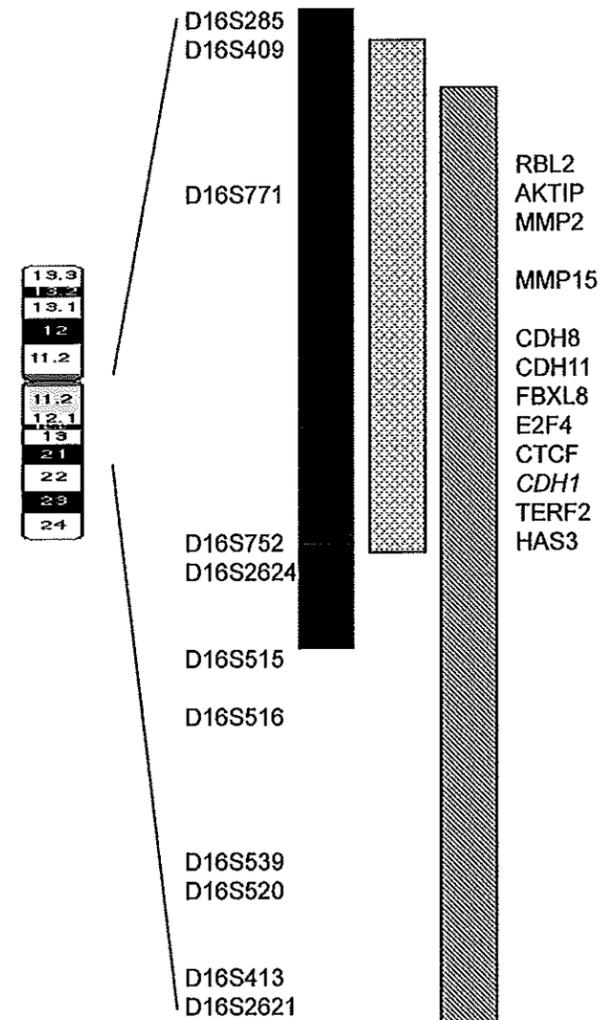
<sup>b</sup>Percent of carcinoma composed of tubular structures.

<sup>c</sup>Mitotic counts vary widely with microscope type. Scores provided here are per 10 high-power fields on an Olympus BX41 microscope with a field diameter 0.54 mm.



support a model by which low-grade and high-grade diseases develop along separate genetic pathways, with alterations of chromosome 16q serving as the critical genetic determinant between histological grades.

Molecular characterization of ductal carcinoma *in situ* (DCIS) lesions further supports a model of two distinct pathways of breast disease development.



**Figure 1.** Map of common region of LOH/AI on chromosome 16q in low-grade breast carcinomas. Slight variations in the boundaries of the region have been reported: black bar, (20); checked bar, (23); striped bar, (24). Candidate genes located in the region are shown on the right. Note that mutations in CDH1 have been associated with invasive lobular carcinoma, but not with low-grade or high-grade invasive ductal carcinoma.

Abbreviations: RBL2, retinoblastoma-like 2; AKTIP, akt-interacting protein; MMP, matrix metalloproteinase; CDH, cadherin; FBXL8, f-box and leucine-rich repeat protein 8; E2F4, e2f transcription factor 4; CTCF, CCCTC-binding factor; TERF2, telomeric repeat-binding factor 2; HAS3, hyaluronan synthase 3.

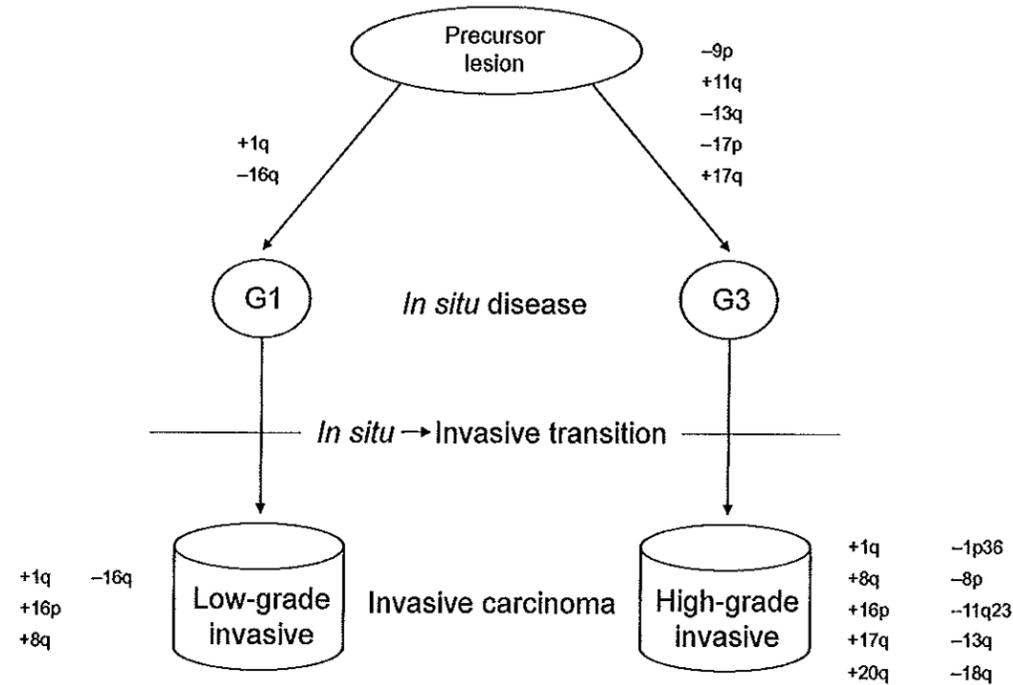
Higher levels of chromosomal alterations have been detected via CGH in high-grade compared to low-grade DCIS, with loss of 16q found almost exclusively in low-grade lesions.<sup>25,26</sup> AI analysis on 100 pure DCIS specimens (with no detectable invasive component) recently found significantly higher levels of AI in the high-grade compared to low-grade lesions—AI at chromosome 16q characterized low-grade lesions, while alterations at 6q25–q27, 8q24, 9p21, 13q14, and 17p13.1 were frequent in high-grade disease.<sup>27</sup> Similar patterns of chromosomal changes in *in situ* and invasive disease suggest that low-grade and high-grade invasive breast tumors evolve directly from low-grade and high-grade DCIS, respectively (Fig. 2).

#### Components of the Nottingham score

The Nottingham score uses three components, tubule formation, nuclear pleomorphism, and mitotic count, to assign histological grade, but mechanisms by which genomic changes in breast carcinomas specifically contribute to these underlying components are unknown. When patterns of AI were compared between tumors with favorable (=1) and unfavorable (=3) scores for each component, significantly higher levels of AI were observed in samples with unfavorable (high) scores for all components.<sup>28</sup> Tumors with reduced tubule formation (score = 3) showed higher levels of AI at chromosomal regions 11q23 and 13q12, those with high levels of nuclear atypia had frequent alterations at 9p21, 11q23, 13q14, 17p13, and 17q12, and carcinomas with high mitotic counts were commonly altered at 1p36, 11q23, and 13q14. Only region 16q11–q22 was altered more frequently in samples with low nuclear atypia. Alterations at 11q23 are common in breast tumors showing reduced tubule formation, high nuclear atypia, and high mitotic counts, suggesting that this is an early genetic change in the development of poorly-differentiated breast tumors; however, alterations at other chromosomal regions in poorly-differentiated tumors may specifically influence cell structure, nuclear morphology, and cellular proliferation.

#### Genomic heterogeneity and breast cancer

The identification of genomic signatures for low-grade and high-grade breast disease provides new insights into the heterogeneity of breast cancer. Under



**Figure 2.** Models of genomic changes depicting pathways of development for low-grade and high-grade breast carcinomas. Low-grade DCIS and IDCA are genetically distinct from high-grade *in situ* and invasive disease. Alterations of chromosomes 1q, 8q, 16p, and 16q are associated with low-grade disease, while high-grade tumors demonstrate higher levels of alterations at a number of regions throughout the genome. Abbreviations: G1, grade 1; G3, grade 3; IDCA, invasive ductal carcinoma.

current models of disease progression, low-grade and high-grade breast carcinomas develop independently along different genetic pathways, thus consideration of breast disease without regard to tumor grade may mask molecular (or environmental) factors specific to one grade.<sup>20</sup> For example, grade 1 DCIS has been shown to exhibit a significantly lower overall frequency of chromosomal changes than low-grade (well-differentiated) invasive carcinomas, but no individual chromosomal regions effectively differentiate low-grade *in situ* from invasive disease. In contrast, high-grade (poorly-differentiated) invasive tumors did not show significantly higher levels of AI than grade 3 DCIS, but AI events at specific chromosomal regions (1p36 and 11q23) were significantly more frequent in high-grade invasive tumors compared to high-grade DCIS.<sup>29</sup> Lower levels of AI in low-grade *in situ* lesions compared to low-grade invasive carcinomas may reflect the protracted time-to-progression associated with low-grade DCIS. Likewise, increased levels of AI at 1p36 and 11q23 in high-grade carcinomas suggest that these chromosomal regions may harbor

genes associated with invasiveness. Therefore, consideration of histological grade when analyzing genetic data has the potential to identify molecular changes associated with invasion and to define molecular signatures of aggressive behavior for low-grade and high-grade disease.

### Molecular Evidence for a Biological Continuum

Stratification of low-grade and high-grade breast carcinomas into separate molecular diseases is based on the high frequency of alterations observed for chromosome 16 in low-grade tumors and a low frequency of 16q alterations in high-grade tumors. To localize genes involved in low-grade IDCA, and to refine regions of chromosome 16 that may be important to the development of high-grade disease, Roylance et al characterized 40 low-grade (grade 1) and 17 high-grade (grade 3) IDCA using CGH with nearly contiguous coverage of chromosome 16q.<sup>30</sup> The majority of low-grade tumors showed large deletions of 16q, while high-grade tumors were more frequently



characterized by multiple, small chromosomal alterations, including copy number gains in this region. Because many regions demonstrated loss and gain of certain sections, chromosome 16q may be inherently unstable and many of these regions may contain secondary, rather than causative, alterations. Based on this data, and the identification of copy number gains not previously detected, Roylance et al<sup>30</sup> suggest that loss of chromosome 16q is an early event in the development of low-grade tumors, and postulate that high-grade carcinomas evolve from low-grade tumors by the accumulation of subsequent chromosomal alterations, such as small breaks and amplifications. These observations question the role of chromosome 16q deletions as the key to defining low- and high-grade genetic pathways of development.

The Nottingham grading system, used to assign histological grade to invasive carcinomas, does not adequately describe internal variation in the degree of differentiation within tumors. Although most pathologists rely on nuclear grade, either alone or in combination with central necrosis, to classify DCIS, one recent attempt to quantify histological diversity in 120 pure DCIS lesions found that ~46% of cases showed localized variability in histological grade. Nearly one-third of lesions with internal grade differences demonstrated further diversity for a panel of immunohistochemistry markers including ER, GATA-binding protein 3 (GATA3), and HER2.<sup>31</sup> The authors concluded that higher-grade DCIS gradually evolve from lower-grade *in situ* lesions by random accumulation of genetic mutations. These studies hypothesize that low-grade and high-grade breast carcinomas are not necessarily unique genetic diseases. Under this model, cells with the most aggressive/poorly-differentiated characteristics tend to become the dominant cell type during progression from low-grade to high-grade carcinomas.

### **Molecular Classification of Intermediate-Grade Tumors**

Molecular and pathological changes have been associated with low-grade and high-grade breast carcinomas, which represent the extremes of histological differentiation. Conversely, development of a model for intermediate-grade breast tumors presents a particular challenge because intermediate-grade carcinomas contain a blend of histological features common to

both low-grade (well-differentiated) and high-grade (poorly-differentiated) tumors. Because carcinomas with intermediate-grade histology represent 30%–60% of all invasive breast cancers, improved understanding of the genetics of these tumors is critical in determining the optimum course of treatment for the large group of patients with intermediate pathological features.<sup>32</sup>

Patterns of genetic changes usually do not differ significantly between intermediate-grade and high-grade carcinomas, supporting the idea that intermediate-grade invasive breast tumors develop from either grade 2 or grade 3 DCIS. Further studies of genomic alterations in breast tumors of different histological grades have shown that although genetic changes were more frequent in grade 3 tumors, alterations of one specific chromosomal region (16q) were significantly lower ( $P < 0.01$ ) in high-grade (26%) compared to intermediate-grade (54%) tumors.<sup>33</sup> Thus it appears that intermediate-grade carcinomas may represent a mixture of histological characteristics and may develop along two independent genetic pathways, one characterized by loss of chromosome 16q, few genomic alterations, and high rates of diploidy, while the other pathway is characterized by high homology with high-grade tumors.

In our ongoing studies of intermediate-grade breast carcinomas, we observed that clinicopathological characteristics and overall levels of genomic alterations in grade 2 tumors were generally intermediate compared to low-grade and high-grade disease.<sup>23</sup> Specifically, 47% of the intermediate-grade tumors showed patterns of genomic alterations similar to high-grade tumors, while 11% had a low-grade signature where AI was detected only at chromosome 16q. Of note, 24% of cases showed genetic features representing a mixture of low-grade and high-grade disease, while 18% had a unique genomic profile not observed in either high- or low-grade tumors. These data suggest that intermediate-grade carcinomas should not be classified as a discrete disease type, but represent a blend of low-grade and high-grade diseases.

Gene expression analysis has been widely used to identify genetic profiles associated with different stages of breast cancer development. Using laser microdissection to isolate pure populations of tumor cells and prevent cross-contamination from stroma or co-occurring lesions, histological grade rather than pathological stage, was found to correlate with



significantly different patterns of gene expression.<sup>34</sup> A subset of samples showed gene expression signatures that were distinctly grade-1-like or grade-3-like; most intermediate-grade tumors exhibited a mixed low- and high-grade gene expression profile. Similarly, an expression signature from only five genes—barren homologue—*Drosophila* (*BRRN1*), hypothetical protein FLJ11029, chromosome 6 open reading frame 173 (*C6orf173*), serine/threonine-protein kinase 6 (*STK6*), and maternal embryonic leucine zipper kinase (*MELK*)—has been shown to discriminate low-grade from high-grade tumors with ~95% accuracy.<sup>35</sup> When applied to intermediate-grade tumors, ~66% (83/126) were reclassified as low-grade-like (G2a) and 34% (43/126) as high-grade-like (G2b). Only five samples had true intermediate gene expression scores. Survival outcomes for the G2a and G2b groups were similar to those in patients with grade 1 and grade 3 tumors, respectively. Further investigation of intermediate-grade carcinomas classified as G2a and G2b demonstrated marked heterogeneity between the two groups, suggesting that intermediate grade tumors do not represent an independent disease subtype and that the G2a and G2b classifications should be considered separate pathobiological entities.<sup>35</sup>

Recognizing the inherent subjectivity in assigning histological grade and the need to better characterize intermediate-grade tumors, researchers have begun to analyze combined gene expression data sets from primary breast tumor samples derived from multiple sources. These approaches have led to the development of a gene expression grade index (GGI), based on 97 genes, which summarizes molecular differences between low-grade and high-grade breast tumors.<sup>14</sup> Similar to earlier results,<sup>34</sup> the GGI partitions intermediate-grade carcinomas into low-grade and high-grade clusters; with un-clustered cases representing a mixture of the two grades.

### Clinical Implications

Molecular data (DNA and RNA) suggest that intermediate-grade invasive breast cancer is not a discrete disease, but represents a blend between low-grade and high-grade tumors. However, whether poorly-differentiated tumors arise from well-differentiated carcinomas, or whether low-grade and high-grade tumors develop along independent genetic pathways remains unclear. Although multiple studies have

identified significant differences in gene expression between low-grade and high-grade disease,<sup>14,34,35</sup> gene and protein expression profiles are transient, reflecting biological conditions in the tumor at the time of excision, rather than an evolutionary history of tumor development. In contrast, chromosomal changes can be very useful for modeling disease progression. Continuing improvements in technologies to measure chromosomal alterations, such as copy number changes assessed by large-scale single-nucleotide polymorphism (SNP) arrays,<sup>36</sup> may provide the tools necessary to determine the role of chromosome 16q in the development of low-grade tumors and further examine the development of low-grade as well as high-grade breast carcinomas.

Determining relationships among tumors of different histological grades has important clinical implications for estimating risk and defining treatment options in patients with breast disease. For example, atypical ductal hyperplasia (ADH) specimens typically share a gene expression profile with grade 1 disease and tend to cluster with low-grade DCIS and well-differentiated invasive cancer.<sup>34</sup> Thus, ADH may represent a precursor lesion, specifically to low-grade breast cancer. Should low-grade disease be genetically distinct from high-grade, patients diagnosed with ADH could be considered lower-risk, reflecting the less aggressive phenotype of low-grade disease. Similarly, under a model of tumor progression from low-grade to high-grade through histological de-differentiation, identification of molecular changes that promote progression may provide molecular targets for the development of therapeutics to block the progression from low-grade to high-grade (aggressive) tumors.

Development of molecular signatures that closely correlate with histological differentiation may improve the assessment of tumor grade. At present, debate continues within the pathology community over the best way to assign histological grade. Some studies suggest that a two-tiered grading system comprised of nuclear pleomorphism and mitotic counts is superior to the current tripartite system that includes tubule formation.<sup>9,37</sup> In contrast, research suggests that a composite score based on a 7-point scale (range 3–9) is more accurate than the current system that converts the cumulative scores from tubule formation, nuclear pleomorphism, and mitotic



counts to a 3-grade system.<sup>38</sup> For example, while tumors with a composite score of 6 or 7 would be classified as intermediate-grade, those with a score of 7 have a prognosis similar to high-grade tumors. It is possible that tumors with a score of 6 correspond to the G2a tumor group and those with a score of 7 to the G2b tumor group defined by Ivshina et al<sup>35</sup> suggesting that tumors with scores of 6 and 7 should be considered separately when making treatment decisions.

Finally, molecular profiles such as the OncotypeDX™ (Genomic Health, Redwood City, CA) and MammaPrint® (Agendia, Amsterdam, The Netherlands) are now being used more frequently as clinical tools to determine treatment for certain groups of patients. For example, the OncotypeDX™ can be used to make decisions about chemotherapy after surgery for women with node-negative, ER-positive breast cancer. In 2008, Ipsogen (<http://www.ipsogen.com/>) developed the MapQuant Dx™ Genomic Grade test based on the GGI discussed above.<sup>14</sup> This test is being marketed as the first microarray-based diagnostic test to measure tumor grade. With the reported ability to classify 80% of intermediate-grade tumors as either low-grade or high-grade, the MapQuant assay may be useful in guiding treatment options, possibly sparing patients with grade 1 or grade 1-like tumors unnecessary treatments, while identifying patients who would benefit from chemotherapy.<sup>32</sup>

## Summary

Molecular characterization of breast tumors at both the DNA and RNA levels suggests that intermediate-grade carcinomas do not represent an independent disease subtype, but instead share clinical and molecular features of low-grade and high-grade tumors. In contrast, debate continues as to whether poorly-differentiated (high-grade) tumors evolve from well-differentiated (low-grade) tumors or whether low-grade and high-grade carcinomas represent discrete diseases that develop along separate genetic pathways. While efforts continue to improve our understanding of biological factors influencing the development of low-, intermediate-, and high-grade tumors, clinical uses of molecular assays are providing new ways to assign histological grade and guide treatments for patients with breast cancer.

## Acknowledgment

This work was performed under the auspices of the Clinical Breast Care Project, and supported by the United States Department of Defense (Military Molecular Medicine Initiative MDA W81XWH-05-2-0075). The opinion and assertions contained herein are the private views of the authors and are not to be construed as official or as representing the views of the Department of the Army or the Department of Defense.

## Disclosure

The authors report no conflicts of interest.

## References

1. Jemal A, Siegel R, Ward E, et al. Cancer statistics, 2008. *CA Cancer J Clin*. 2008;58(2):71–96.
2. Bowcock AM, ed. *Breast Cancer: Molecular Genetics, Pathogenesis, and Therapeutics*. Totowa, New Jersey: Humana Press; 1999.
3. Gort M, Broekhuis M, Otter R, Klazinga NS. Improvement of best practice in early breast cancer: actionable surgeon and hospital factors. *Breast Cancer Res Treat*. 2007;102(2):219–26.
4. Perou CM, Sorlie T, Eisen, et al. Molecular portraits of human breast tumours. *Nature*. 2000;406(6797):747–52.
5. Sorlie T, Perou CM, Tibshirani R, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A*. 2001;98(19):10869–74.
6. Sorlie T, Tibshirani R, Parker J, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci U S A*. 2003;100(14):8418–23.
7. Bloom HJ, Richardson WW. Histological grading and prognosis in breast cancer; a study of 1409 cases of which 359 have been followed for 15 years. *Br J Cancer*. 1957;11(3):359–77.
8. Elston CW, Ellis IO. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. *Histopathology*. 1991;19(5):403–10.
9. Tsuda H, Akiyama F, Kurosumi M, Sakamoto G, Watanabe T. Establishment of histological criteria for high-risk node-negative breast carcinoma for a multi-institutional randomized clinical trial of adjuvant therapy. Japan National Surgical Adjuvant Study of Breast Cancer (NSAS-BC) Pathology Section. *Jpn J Clin Oncol*. 1998;28(8):486–91.
10. Henson DE, Ries L, Freedman LS, Carriaga M. Relationship among outcome, stage of disease, and histologic grade for 22,616 cases of breast cancer: the basis for a prognostic index. *Cancer*. 1991;68(10):2142–9.
11. Fitzgibbons PL, Page DL, Weaver D, et al. Prognostic factors in breast cancer. College of American Pathologists Consensus Statement 1999. *Arch Pathol Lab Med*. 2000;124(7):966–78.
12. American Joint Committee on Cancer. *AJCC Cancer Staging Manual*. 6th ed. New York: Springer; 2002.
13. Robbins P, Pinder S, de Klerk N, et al. Histological grading of breast carcinomas: a study of interobserver agreement. *Hum Pathol*. 1995;26(8):873–9.
14. Sotiriou C, Wirapati P, Loi S, et al. Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. *J Natl Cancer Inst*. 2006;98(4):262–72.
15. Tabar L, Fagerberg G, Chen HH, Duffy SW, Gad A. Tumour development, histology and grade of breast cancers: prognosis and progression. *Int J Cancer*. 1996;66(4):413–9.
16. Duffy SW, Tabar L, Fagerberg G, et al. Breast screening, prognostic factors and survival—results from the Swedish two county study. *Br J Cancer*. 1991;64(6):1133–8.



17. Rajakariar R, Walker RA. Pathological and biological features of mammographically detected invasive breast carcinomas. *Br J Cancer*. 1995;71(1):150-4.
18. Millis RR, Barnes DM, Lampejo OT, Egan MK, Smith P. Tumour grade does not change between primary and recurrent mammary carcinoma. *Eur J Cancer*. 1998;34(4):548-53.
19. Tirkkonen M, Tanner M, Karhu R, Kallioniemi A, Isola J, Kallioniemi OP. Molecular cytogenetics of primary breast cancer by CGH. *Genes Chromosomes Cancer*. 1998;21(3):177-84.
20. Roylance R, Gorman P, Harris W, et al. Comparative genomic hybridization of breast tumors stratified by histological grade reveals new insights into the biological progression of breast cancer. *Cancer Res*. 1999;59(7):1433-6.
21. Buerger H, Otterbach F, Simon R, et al. Different genetic pathways in the evolution of invasive breast cancer are associated with distinct morphological subtypes. *J Pathol*. 1999;189(4):521-6.
22. Roylance R, Gorman P, Hanby A, Tomlinson I. Allelic imbalance analysis of chromosome 16q shows that grade I and grade III invasive ductal breast cancers follow different genetic pathways. *J Pathol*. 2002;196(1):32-6.
23. Ellsworth RE, Hooke JA, Love B, et al. Correlation of levels and patterns of genomic instability with histological grading of invasive breast tumors. *Breast Cancer Res Treat*. 2008;107(2):259-65.
24. Cleton-Jansen AM, Buerger H, ter Haar N, et al. Different mechanisms of chromosome 16 loss of heterozygosity in well- versus poorly differentiated ductal breast cancer. *Genes Chromosomes Cancer*. 2004;41(2):109-16.
25. Buerger H, Otterbach F, Simon R, et al. Comparative genomic hybridization of ductal carcinoma in situ of the breast—evidence of multiple genetic pathways. *J Pathol*. 1999;187(4):396-402.
26. Hwang ES, DeVries S, Chew KL, et al. Patterns of chromosomal alterations in breast ductal carcinoma in situ. *Clin Cancer Res*. 2004;10(15):5160-7.
27. Ellsworth RE, Ellsworth DL, Love B, et al. Correlation of levels and patterns of genomic instability with histological grading of DCIS. *Ann Surg Oncol*. 2007;14(11):3070-7.
28. Ellsworth RE, Hooke JA, Love B, Ellsworth DL, Shriver CD. Molecular changes in primary breast tumors and the Nottingham histologic score. *Pathol Oncol Res*. e-pub. 2009.
29. Ellsworth RE, Vertrees A, Love B, Hooke JA, Ellsworth DL, Shriver CD. Chromosomal alterations associated with the transition from in situ to invasive breast cancer. *Ann Surg Oncol*. 2008;15(9):2519-25.
30. Roylance R, Gorman P, Papior T, et al. A comprehensive study of chromosome 16q in invasive ductal and lobular breast carcinoma using array CGH. *Oncogene*. 2006;25(49):6544-53.
31. Allred DC, Wu Y, Mao S, et al. Ductal carcinoma *in situ* and the emergence of diversity during breast cancer evolution. *Clin Cancer Res*. 2008;14(2):370-8.
32. Ignatiadis M, Sotiriou C. Understanding the molecular basis of histologic grade. *Pathobiology*. 2008;75(2):104-11.
33. Buerger H, Momms EC, Littmann R, et al. Ductal invasive G2 and G3 carcinomas of the breast are the end stages of at least two different lines of genetic evolution. *J Pathol*. 2001;194(2):165-70.
34. Ma XJ, Salunga R, Tuggle JT, et al. Gene expression profiles of human breast cancer progression. *Proc Natl Acad Sci U S A*. 2003;100(10):5974-9.
35. Ivshina AV, George J, Senko O, et al. Genetic reclassification of histologic grade delineates new clinical subtypes of breast cancer. *Cancer Res*. 2006;66(21):10292-301.
36. Argos M, Kibriya MG, Jasmine F, et al. Genomewide scan for loss of heterozygosity and chromosomal amplification in breast carcinoma using single-nucleotide polymorphism arrays. *Cancer Genet Cytogenet*. 2008;182(2):69-74.
37. Komaki K, Sano N, Tangoku A. Problems in histological grading of malignancy and its clinical significance in patients with operable breast cancer. *Breast Cancer*. 2006;13(3):249-53.
38. Chowdhury N, Pai MR, Lobo FD, Kini H, Varghese R. Impact of an increase in grading categories and double reporting on the reliability of breast cancer grade. *APMIS*. 2007;115(4):360-6.

**Publish with Libertas Academica and every scientist working in your field can read your article**

*"I would like to say that this is the most author-friendly editing process I have experienced in over 150 publications. Thank you most sincerely."*

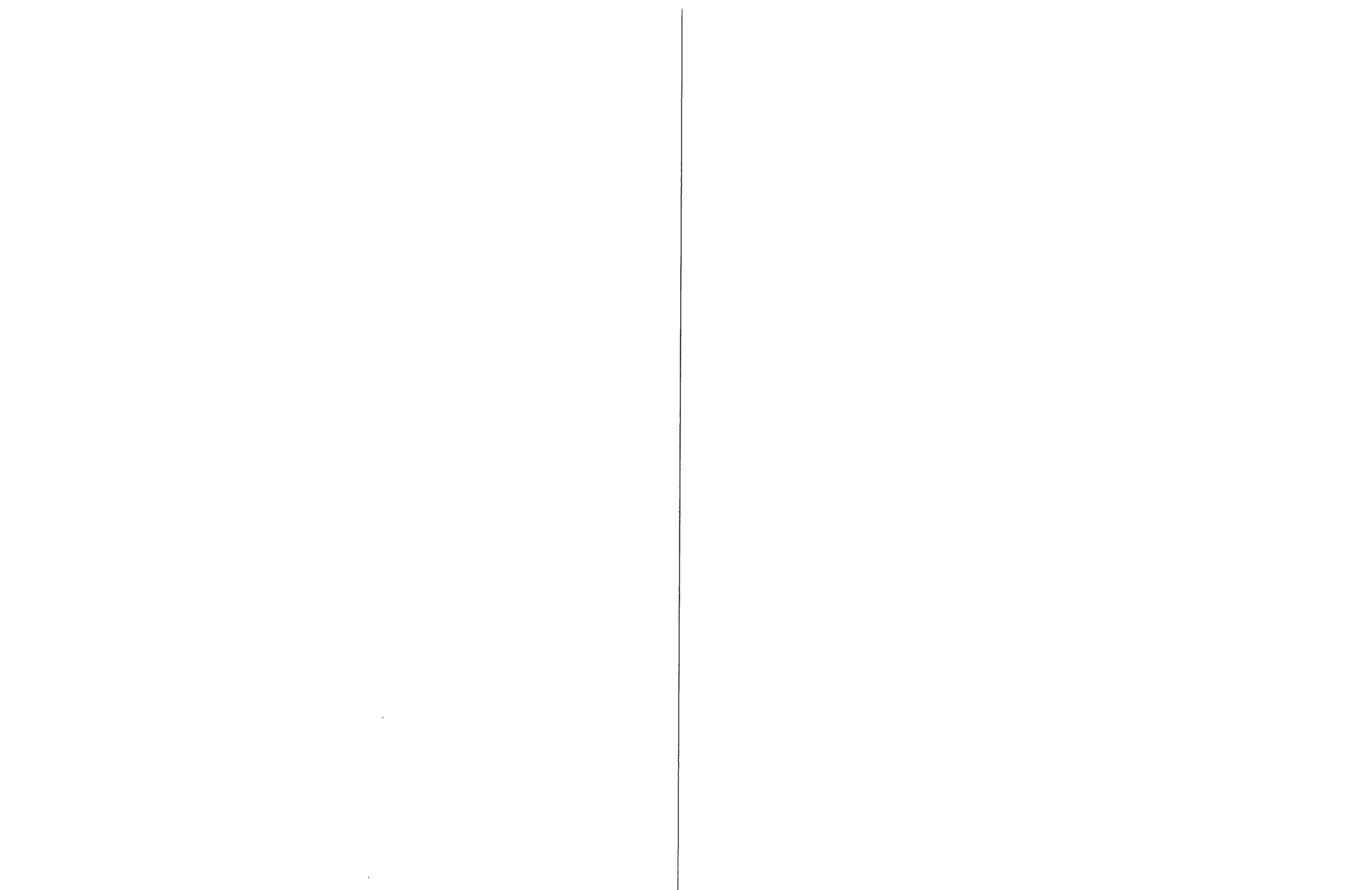
*"The communication between your staff and me has been terrific. Whenever progress is made with the manuscript, I receive notice. Quite honestly, I've never had such complete communication with a journal."*

*"LA is different, and hopefully represents a kind of scientific publication machinery that removes the hurdles from free flow of scientific thought."*

**Your paper will be:**

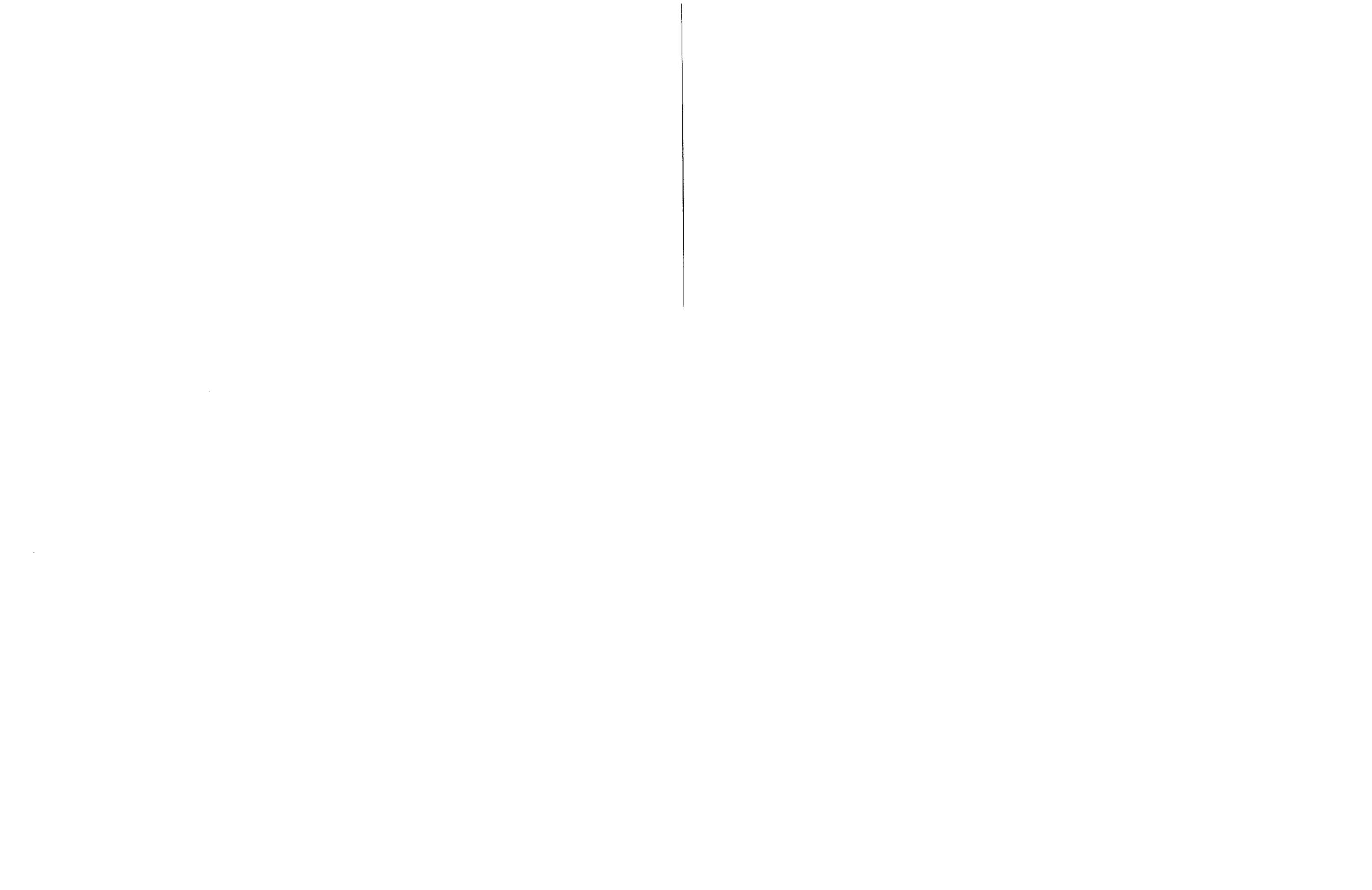
- Available to your entire community free of charge
- Fairly and quickly peer reviewed
- Yours! You retain copyright

**<http://www.la-press.com>**



Clinical implications of the molecular relationship between HER2 and BRCA1  
Rachel E. Ellsworth, Brenda Deyarmin, Jeffrey A. Hooke, Craig D. Shriver

While HER2 status is used prognostically, the effects of HER2 can be influenced by other genes; co-amplification of TOP2A modulates the effectiveness of treatment with Herceptin. BRCA1, deleted in >50% of sporadic breast tumors, is located distal to HER2 on chromosome 17q and is marked by structural instability. To determine how alterations at the BRCA1 locus may modify amplification of HER2, FISH analysis of BRCA1 was performed in breast tumors with HER2 amplification and/or allelic imbalance on chromosome 17q12. HER2 status (n=68) was determined using the PathVysion HER2 kit. For BRCA1 analysis, DNA probes were generated by nick translation from a BAC clone containing the entire BRCA1 gene in conjunction with a CEP17 probe. Copy number gains and deletions were defined as BRCA1:CEP17 >1.3 and <0.8, respectively; chromosomal content was defined using copy number values of the CEP17 probe: monosomy=<1.76, disomy=1.76-2.25, polysomy =>2.25. Clinical amplification of the HER2 gene was detected in 58% of tumors; an additional 16% of tumors were polysomic for chromosome 17q. The remaining 26% of tumors had previously detected allelic imbalance at the 17q12 region. The majority of samples (31%) were characterized by an amplification of HER2 with a downstream deletion of BRCA1, while HER2 and BRCA1 were amplified concomitantly in 8% of tumors, including those with polysomy of 17q. Fifteen percent of samples were polysomic for chromosome 17q; using current clinical scoring criteria, each of those samples were reported as having, despite the altered copy number, negative HER2 status. Levels of overall genomic instability, which were previously measured at 26 chromosomal regions, were significantly higher (P<0.05) in tumors with deletion of BRCA1 compared to those without loss of BRCA1 (38% and 24%, respectively). Alterations, especially deletions, of BRCA1 are highly correlated with copy number changes in HER2. Because deletion of BRCA1 has been associated with overall genomic instability, patients with amplification of HER2 and deletion of BRCA1 may prove unresponsive to treatment with Herceptin. Loss of BRCA1 should, therefore, be considered as a genetic modifier to HER2 amplification and the status of BRCA1 should be considered in selecting treatment options for patients with HER2 amplified tumors.



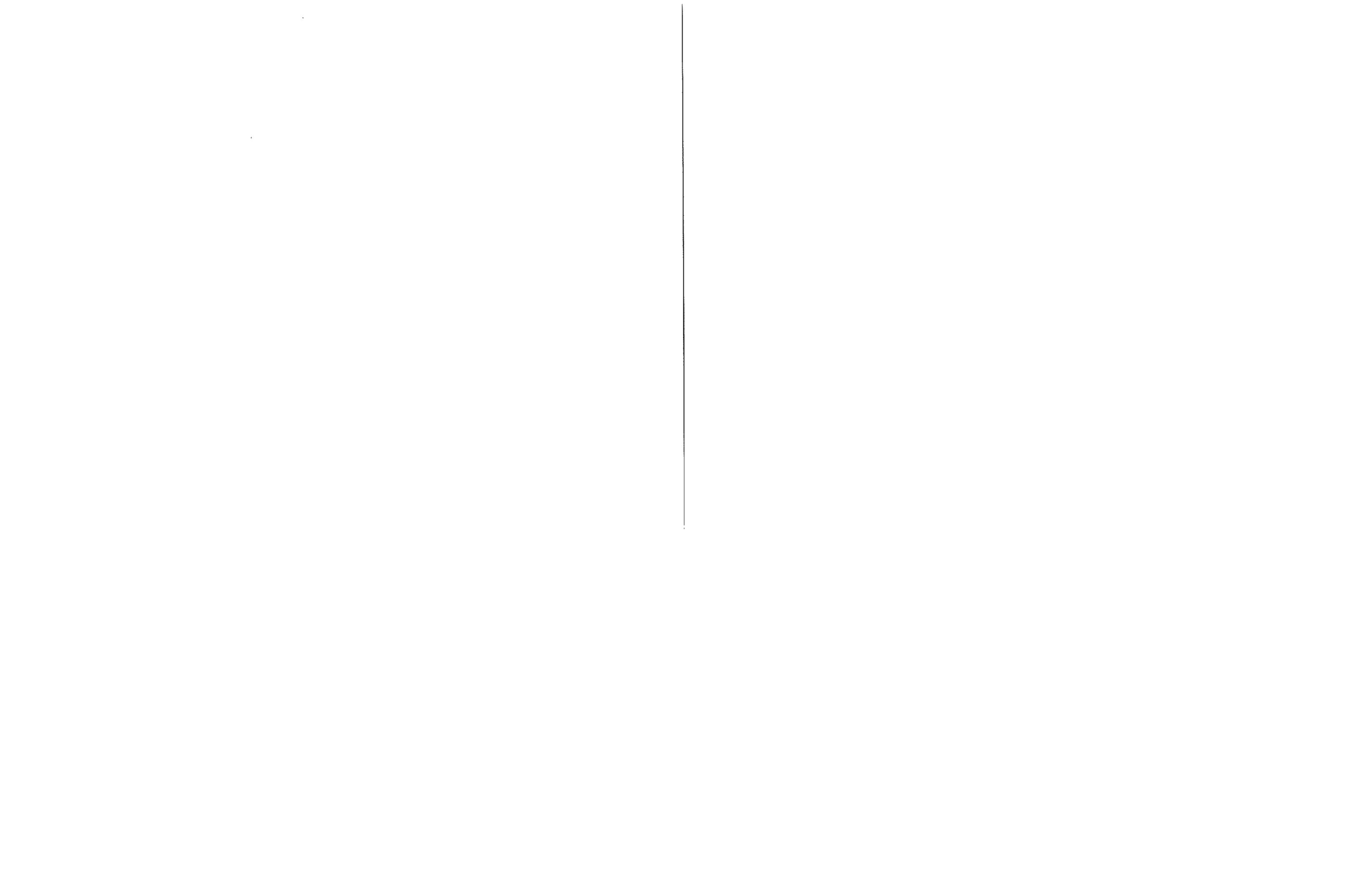
Identification of a molecular signature of breast cancer metastasis  
RE Ellsworth, LA Field, JA Hooke, JA Kane, BL Love, CD Shriver

**Background:** Lymph node status remains one of the most useful prognostic indicators in breast cancer. Current methods to assess nodal status, including sentinel lymph node biopsy and/or axillary dissection, can disrupt the lymphatic system and lead to secondary complications. In addition, 10-33% of women diagnosed with negative lymph nodes will later develop metastasis. Identification of molecular signatures discriminating primary tumors from patients with the propensity to metastasize from those without would not only allow for stratification of patients requiring surgical assessment of lymph node status but identification of node negative patients at high-risk of developing metastatic disease.

**Methods:** Research-grade frozen breast specimens were collected from women with negative (n=37) and positive (n=37) lymph node status. Patients with positive BRCA1 or BRCA2 status, neoadjuvant therapy or with negative lymph node status who later developed distant metastasis were not included in this study. RNA was isolated from pure tumor cell populations after laser microdissection and gene expression data was generated using HG U133A 2.0 arrays (Affymetrix). Data were randomized into training (n=17 node negative and 17 node positive) and test sets (n=20 node negative and 20 node positive). Mann-Whitney U testing was used to determine differential expression ( $P < 0.01$  and  $>2$  fold difference in expression).

**Results:** Data from the training set identified 226 probes representing 204 genes and 12 ESTs differentially expressed between node-negative and node-positive patients. Of these, 54 and 162 were expressed at higher levels in the node-positive and node-negative tumors, respectively. Using similar threshold values, five genes were validated using the test set, including BRF1, HMG20B, KIF23, MCL1 and TPSAB1.

**Discussion:** Although gene expression patterns were highly similar, a molecular signature of five genes that were differentially expressed between patients with and without metastasis was identified. Although MCL1 is involved in apoptosis, the function of the other genes in cancer metastasis has yet to be determined. These data may, therefore, provide new information into the metastatic process as well as allow identification of patients with propensity to metastasize.



Molecular characterization of breast cancer progression: early lesions are not genetically advanced

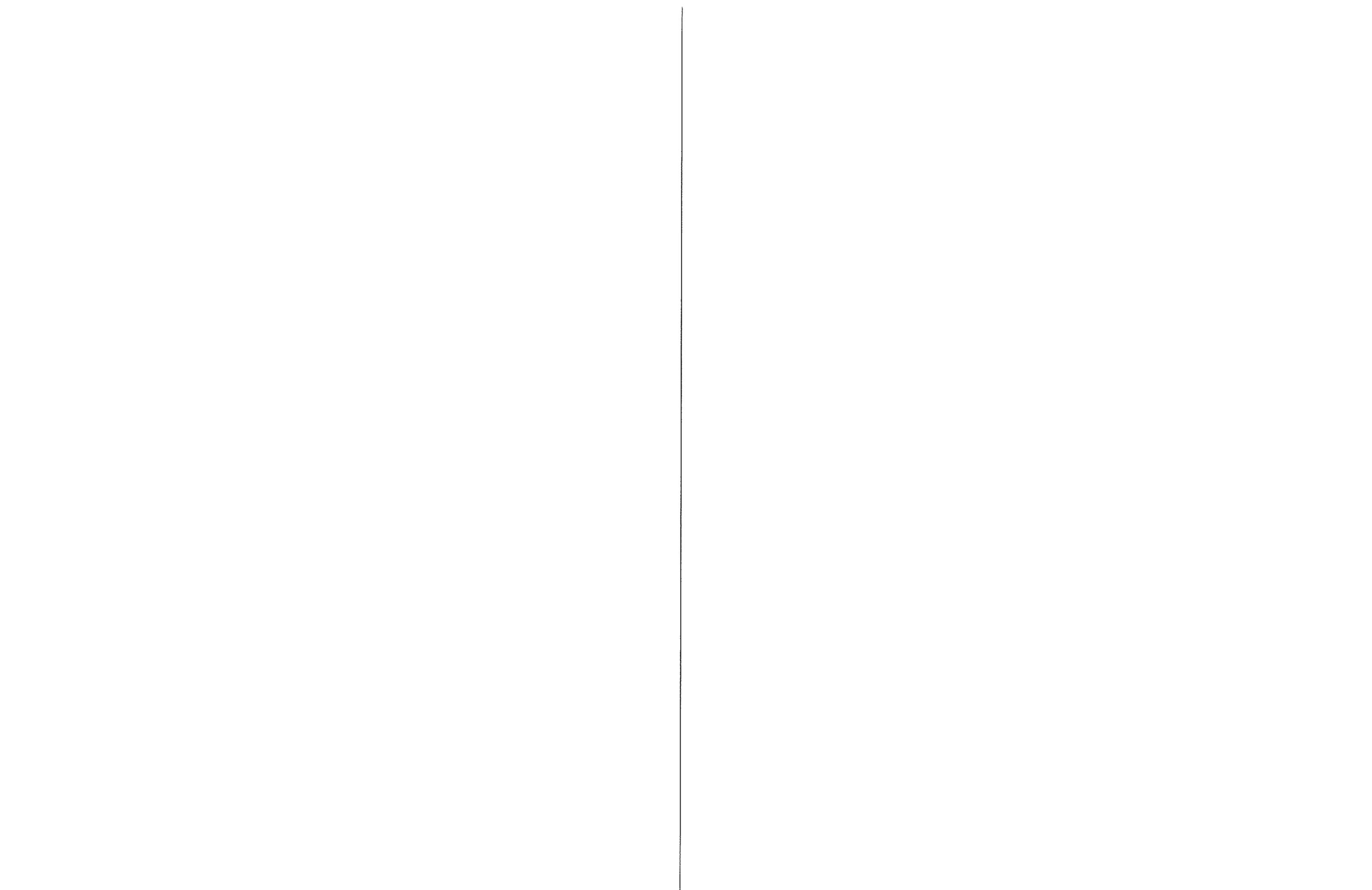
Rachel E. Ellsworth<sup>1</sup>, Jamie D. Weyandt<sup>2</sup>, Jamie L. Fantacone-Campbell<sup>3</sup>, Brenda Deyarmin<sup>2</sup>, Darrell L. Ellsworth<sup>2</sup>, Jeffrey A. Hooke<sup>3</sup>, Craig D. Shriver<sup>3</sup>

**Introduction:** Columnar cell lesions (CCL) and atypical ductal hyperplasia (ADH) frequently coexist and share molecular changes with *in situ* and invasive components, suggesting that CCL and ADH may be precursors to breast cancer. These conclusions are, however, largely based on studies examining CCL and ADH from patients diagnosed with more advanced disease. Thus, allelic imbalance (AI) was assessed in pure CCL or ADH specimens to characterize molecular changes in early breast lesions.

**Methods:** DNA samples were obtained from laser microdissected CCL (n=42) or ADH (n=31) lesions without concurrent *in situ* or invasive disease. AI was assessed at 26 chromosomal regions commonly altered in breast cancer. Data was analyzed using Fisher's exact and Student's t-tests using a cutoff of  $P < 0.05$  to define significance.

**Results:** The average AI frequency was 6.2% (range 0-20%) in CCL and 6.1% (range 0-25%) in ADH. The highest levels of AI were on chromosomes 17q12-q21 in CCL (15%) and 8q24 in ADH (23%); ~33% of early lesions did not have any detectable genetic changes. Levels of AI for both CCL and ADH were significantly ( $P < 0.0001$ ) lower than those found in *in situ* or invasive disease. Genetic changes characteristic of more advanced *in situ* and invasive disease, especially changes on chromosome 16q and 17p, were infrequent or non-detectable in pure early lesions.

**Conclusions:** Pure CCL and ADH lesions demonstrate lower levels of genetic alterations than DCIS, invasive carcinomas or early lesions from cancerous breasts. Alterations characteristic of advanced disease were not found in lesions from non-cancerous breasts. While data generated from synchronous lesions has been used as support for a model of evolution from CCL and ADH to *in situ* and invasive disease, pure early lesions are, in fact, not genetically advanced and molecular profiles do not support these lesions as obligatory precursors to more advanced disease. Given the molecular differences between pure and synchronous lesions, current models of disease initiation and progression based largely on data generated from cancerous breasts must be re-evaluated.



## Differential Benign Breast Disease Co-occurrence with Cancer in Caucasian and African American Women

Anthony Bekhash<sup>1</sup>, Jasmine Saini<sup>1</sup>, XiangYu Fan<sup>1</sup>, Jeffrey Hooke<sup>2</sup>, Richard J. Mural<sup>1</sup>, Craig D. Shriver<sup>2</sup>, Hai Hu<sup>1</sup>

<sup>1</sup>Windber Research Institute, Windber, PA

<sup>2</sup>Walter Reed Army Medical Center, Washington, DC

**Background:** Breast Cancers (BCs) in Caucasian (CA) and African American (AA) women have different characteristics. Recently, there have been reports, mostly focusing on Caucasians that Benign Breast Diseases (BBDs) may be risk factors for BCs. There are also a few reports of different co-occurring patterns of BBDs with BCs between the two populations. In the Clinical Breast Care Project (CBCP) more than 4,000 subjects have been enrolled following HIPAA-compliant, IRB-approved protocols, with more than 35,000 biospecimens collected. For patients requiring biopsies, an expanded pathology report for research is completed. The occurrences of any of the 131 pathologic conditions, including 66 BBDs, are recorded. All the results are reviewed by the same pathologist. The CBCP provides a uniformed pathology dataset for a comprehensive characterization of the association of BBDs with BCs.

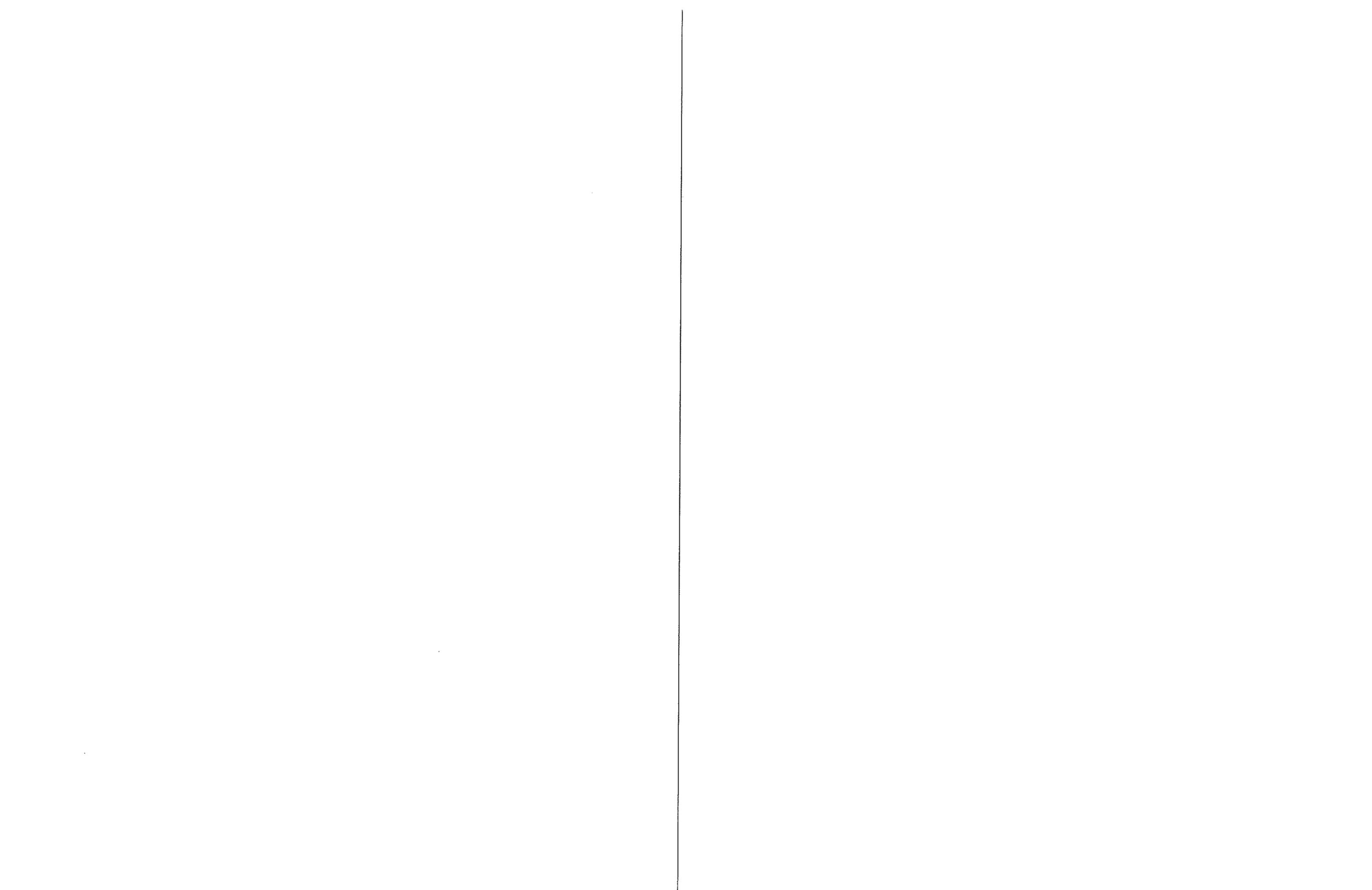
**Method:** The diagnoses were made from potentially multiple biopsies obtained over a short period of time, mostly within 60 days. CBCP subjects with BBD diagnosis and ethnicity information from their most recent visit were selected, giving a dataset totaling 1479 CA and 484 AA women. We studied the association between BBDs and BCs (including in situ, invasive, and malignant NOS) in these two groups using the chi-square test.

**Results:** In both populations 6 BBDs are positively associated with BCs, including atypical ductal hyperplasia (ADH) (4% vs 15% for AA  $p < 3.8E-05$ , meaning 4% cancer-free cases with ADH compared to 15% cancer cases co-occurring with it; 7% vs 16% for CA  $p < 1.9E-08$ ), microcalcifications (27% vs 56% for AA  $p < 1.2E-09$ , 35% vs 53% for CA  $p < 8.0E-12$ ), multiple (peripheral) papillomas (6% vs 16% for AA  $p < 0.0005$ , 4% vs 8% for CA  $p < 0.0002$ ), columnar cell hyperplasia (9% vs 21% for AA  $p < 0.0003$ , 6% vs 18% for CA  $p < 3.5E-12$ ), and moderate intraductal hyperplasia (IDH) (14% vs 28% for AA  $p < 0.0002$ , 14% vs 18% for CA  $p < 0.03$ ). On the contrary, 3 BBDs are negatively associated with BCs including fibroadenoma (30% vs 10% for AA  $p < 4.9E-07$ ; 21% vs 7% for CA  $p < 1.0E-13$ ), and mild IDH (10% vs 2% for AA  $p < 0.003$ ; 13% vs 7% for CA  $p < 0.0002$ ).

Eight BBDs differentially co-occur with BCs between AA and CA. BCs in AA are associated with sclerosing adenosis (13% vs 31%,  $p < 4.5E-06$ ), fibrocystic changes (45% vs 61%,  $p < 0.001$ ), fibroadenomatoid nodule (6% vs 12%,  $p < 0.042$ ), and cysts (42% vs 57%,  $p < 0.004$ ). BCs in CA are positively associated with columnar cell hyperplasia with atypia (3% vs 8%  $p < 3.6E-05$ ), atypical lobular

hyperplasia (1% vs 5%,  $p < 6.0E-05$ ), and radiation changes (0% vs 1.4%  $p < 0.004$ ), but negatively associated with duct ectasia (10% vs 2%,  $p < 7.9E-10$ ).

**Discussion** Our findings are in accordance with established BC risk factors such as ADH and moderate IDH for both ethnic groups. We further found that 8 BBDs differentially co-occur with BCs between AA and CA groups, and interestingly the 4 significant BBDs detected in the AA group typically show >10% increased co-occurrence rate in cancer cases compared to cancer-free cases, whereas the 4 BBDs detected in the CA group almost always show a <10% co-occurrence rate overall. More research is undergoing to understand the implications of this distinct co-occurrence pattern.



**Association of clinicopathologic characteristics with IHC-based breast cancer subtypes**

Jasmine Saini<sup>1</sup>, Al Kovatich<sup>3</sup>, Xing Li<sup>1</sup>, Anthony Bekash<sup>1</sup>, Jeff Hooke<sup>2</sup>, Richard Mural<sup>1</sup>, Craig D. Shriver<sup>2</sup>, Hai Hu<sup>1</sup>

<sup>1</sup>Biomedical Informatics, Windber Research Institute, Windber, PA

<sup>2</sup>Clinical Breast Care Project, Walter Reed Army Medical Center, Washington, DC

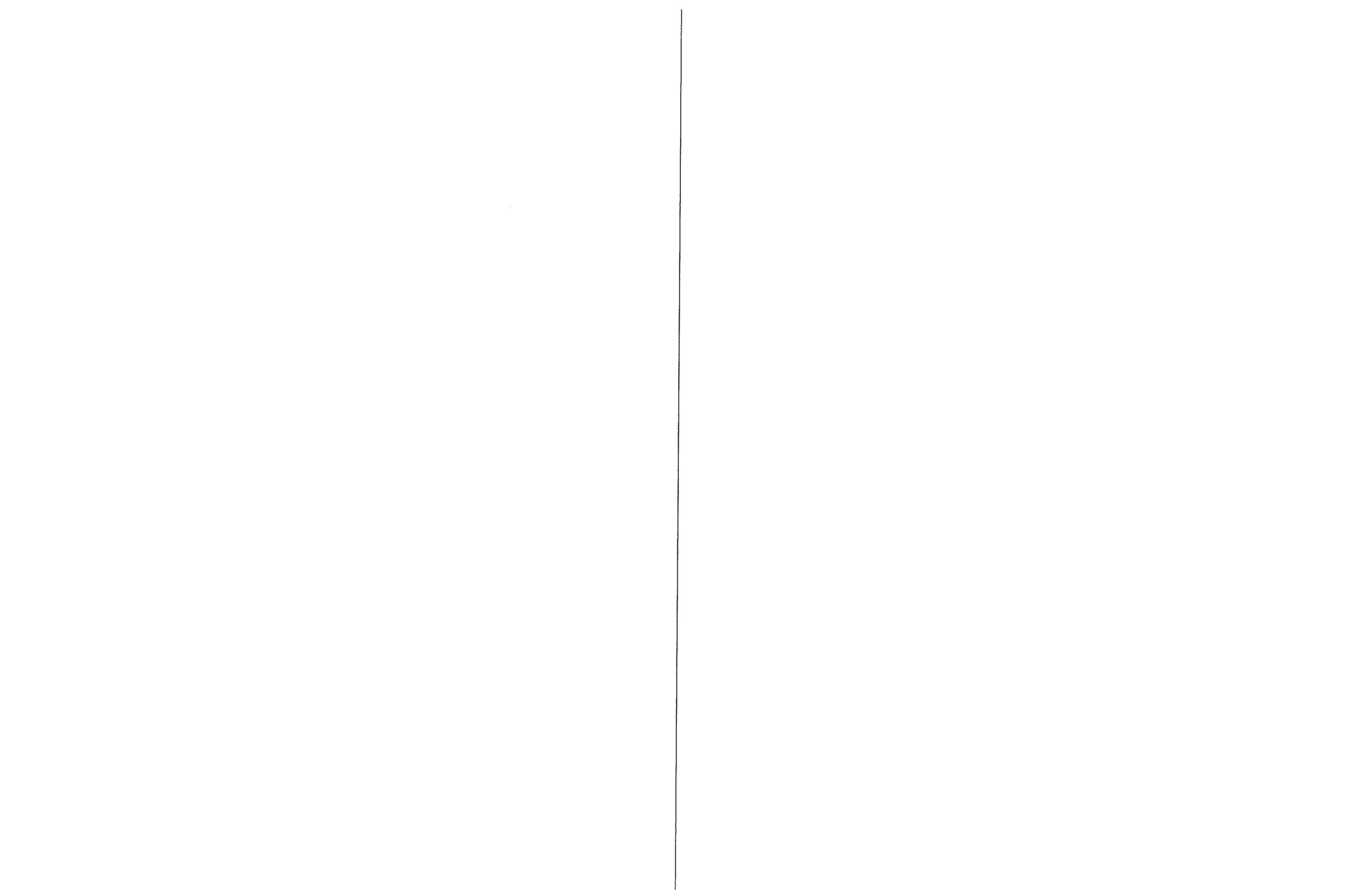
<sup>3</sup>MDR Global Systems, Windber, PA

**Background** Invasive breast cancers have been classified by gene expression into the following major subclasses – Luminal A, Luminal B, Triple Negative and Her2+. Recently, there have been reports to characterize these subclasses of cancer based on immunohistochemistry (IHC) assays of ER, PR and HER2. In the Clinical Breast Care Project (CBCP), we have enrolled >4000 subjects with comprehensive clinicopathologic data collected using a core questionnaire and a pathology checklist. The latter contains 131 pathologic conditions, including 66 Benign Breast Diseases (BBDs). IHC assay results for ER, PR, HER2, Ki67 and p53 were available on most biopsy-confirmed invasive breast cancer tissues. This complete data set was used in this study to understand the association between subclasses and clinicopathological characteristics.

**Methods** The study consists of a total of 419 invasive breast cancer patients classified based on IHC results for ER, PR and HER2, with a total of 240 Luminal A (ER and/or PR+, HER2-), 89 Luminal B (ER and/or PR+, HER2+), 40 Her2+ (ER-, PR-, HER2+) and 50 Triple Negative (ER-, PR-, HER2-). Clinical characteristics of age, ethnicity, menopausal status, tumor size, lymph node status, metastatic status, tumor stage, age of menarche, birth control pill usage, and history of hormone replacement therapy. Of the 66 BBDs, 21 were studied and 45 were removed due to null values. All the data were analyzed using chi-square test.

**Results** Majority of the cases were Luminal A subtype (57.3%), followed by Luminal B (21.2%), Triple Negative (12%) and Her2+ (9.5%). Luminal A had 54% patients older than 50 yrs of age in comparison to 42% in the non-Luminal A groups ( $p=0.01$ ). Tumor size distributions are also different among the subclasses ( $p=0.002$ ); 77% of all Luminal A tumors were in T1, Luminal B had 59% tumors in T1, and Her2+ had higher number of tumors in T4 in comparison to T3. Luminal B subtype had more post menopausal women when compared to the rest ( $p=0.05$ ). The protein expression of p53 showed a significant difference in the subclasses ( $p=0.03$ ), but no significance was obtained for Ki67. A trend was detected that there was a lower lymph node positive rate in Luminal A (28%,  $n=224$ ) when compared to the rest (36%,  $n=167$ ;  $p=0.10$ ). These cancer subclasses had different co-occurrences with Sclerosing Papilloma ( $p=0.009$ ), Juvenile Fibroadenoma ( $p=0.025$ ) and Intraductal Hyperplasia (mild) ( $p=0.016$ ). Triple Negative subtype also co-occurred with Apocrine Adenosis ( $p=0.021$ ) and Columnar Cell hyperplasia ( $p=0.018$ ), whereas Luminal B subtype co-occurred with Multiple (peripheral) Papillomas ( $p=0.014$ ) and Phyllodes Tumor ( $p=0.004$ ), and Her2+ co-occurred with Lactational Metaplasia ( $p=0.048$ ).

**Discussion** This study confirms reported differential association of subclasses with T-stage, menopausal status and age. We further found that in this patient population, there is a trend that Luminal A patients are less likely to be lymph node positive. Positive p53 expression is significantly associated with Luminal B, Her2+ and Triple Negative but not Luminal A subtypes. In addition, we found that several BBDs differentially associate with subtypes. This preliminary study points to new directions for further understanding of breast cancer subtypes.



**An approach to correlate the temporal information to facilitate specimen selection in the breast cancer research project**

Leonid Kvecher<sup>1</sup>, Weiqiang Wu<sup>1</sup>, Jeffrey Hooke<sup>2</sup>, Craig D. Shriver<sup>1,2</sup>, Richard J. Mural<sup>1</sup>, and Hai Hu<sup>1</sup>

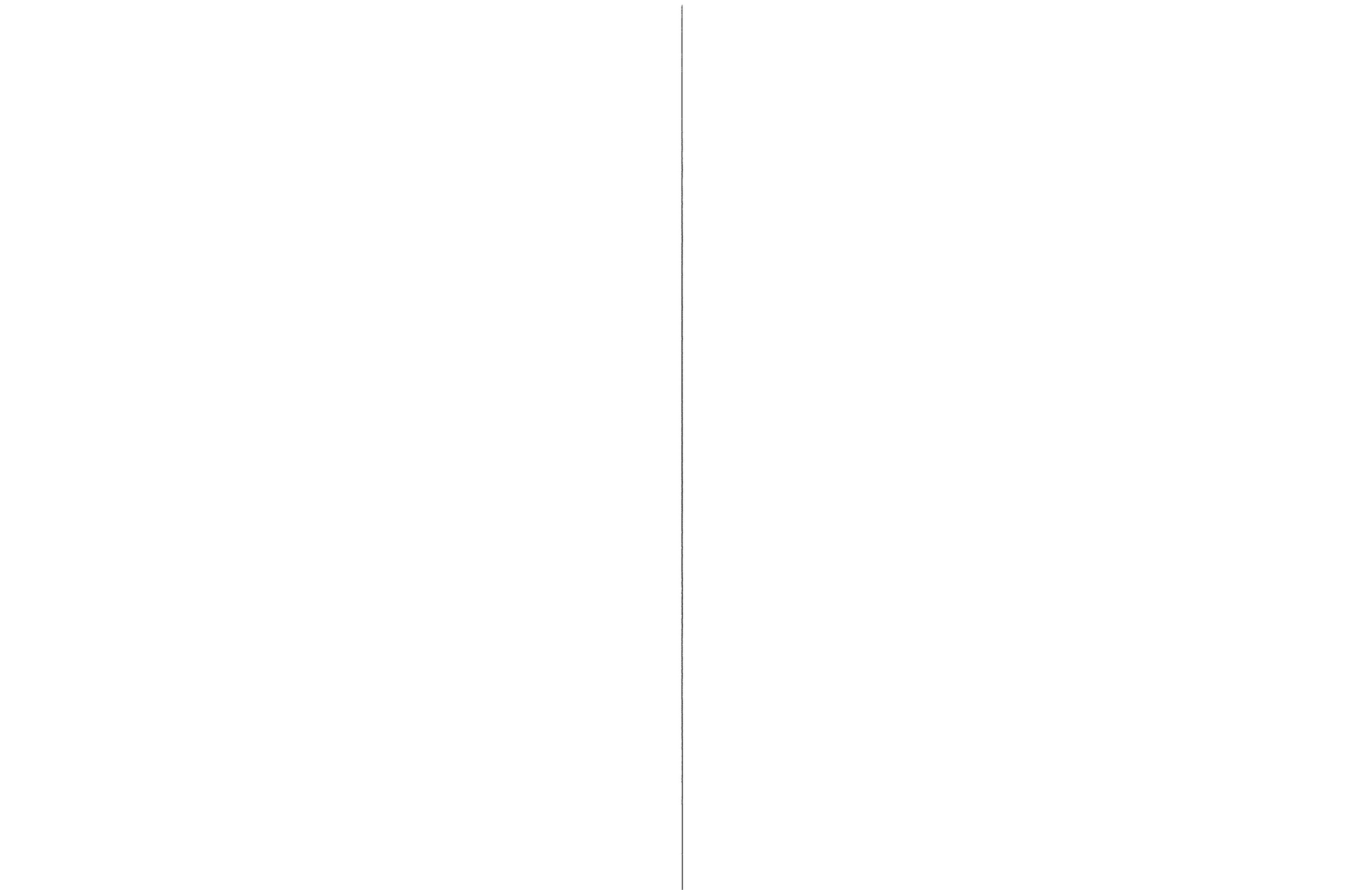
<sup>1</sup>Windber Research Institute, Windber, PA.

<sup>2</sup>Walter Reed Army Medical Center, Washington DC.

**Background:** Temporal information management is very important in translational research. In the Clinical Breast Care Project (CBCP), the information on subjects and their specimens may be collected at multiple time points using multiple instruments. All such information is stored in an in-house data warehouse. Currently, 4000+ subjects have been enrolled in the study following HIPAA-compliant IRB-approved protocols with 35,000+ specimens collected. Some of the patient's information is static but other data are time dependent. As a result, selecting samples for experiments is a challenge due to complicated temporal relationships between samples and information collected through various instruments.

**Methods and Results:** In the CBCP, the clinical information, blood, and solid tissues of a subject may be collected at multiple time points, associated with the completion of a Core Questionnaire (CQ) for clinical information, and/or a Pathology Checklist (PC) for pathology and sample information. We have designed and implemented an algorithm to use a set of pre-defined flags to precisely describe each sample related to patient's clinical and pathology information in the temporal domain. Five categories (flags) were created to describe the relationship between the sample date (SD) and the CQ date based on whether SD is within 60 days of the CQ date or there is missing data or not. The relationship between blood samples and pathology information is more complicated. Within 90 days, any of the 15 surgical procedures might be performed on a patient and blood samples might be collected before, at the time of, or between any procedures. For some experiments, it is crucial to select blood samples taken before tumor is impacted or severely impacted. Thus, we defined a dozen categories to describe the relationship between the SD and the procedure date (PD), including when the SD is earlier than any PD, equals to the first PD, or between certain procedures. Using these flags we have characterized the relationships between SDs and CQ dates, and between SDs and PDs for all the samples and all the subjects, and stored all the information into two relational tables. The temporal criteria for sample selection are now represented by the relationships between these flags, and can be implemented through several filtering processes. The described algorithm drastically reduces the time needed for precise sample selection from several days for manual efforts to several hours.

**Discussion:** We are in the process of developing a general data model for temporal information management. The method described here is a transitional solution that fulfills our current needs. As an initial effort some of the thresholds for categorizing different temporal conditions are arbitrary, and we are validating them with experimental results for future improvement. Nonetheless, this algorithm has greatly enhanced the efficiency of our subject and specimen selection for wet bench experiments. The same principle can be applied to the future temporal data model solution, for CBCP and other human disease studies.



## Converting paper medical records to electronic version to support breast cancer translational research and clinical practice

Mark Shay<sup>1</sup>, Joanna Duman<sup>2</sup>, Steve Eberly<sup>2</sup>, Hai Hu, PhD<sup>3</sup>, and Craig D. Shriver, MD<sup>4</sup>

<sup>1</sup>Concurrent Technologies Corporation (CTC), 100 CTC Drive, Johnstown, PA 15904;

<sup>2</sup>National Interest Security Company LLC (NISC), 3050 Chain Bridge Road, Suite 600, Fairfax, VA, 22030;

<sup>3</sup>Windber Research Institute, Windber, PA 15963, USA;

<sup>4</sup>Clinical Breast Care Project, Walter Reed Army Medical Center, Washington, DC 20307, USA

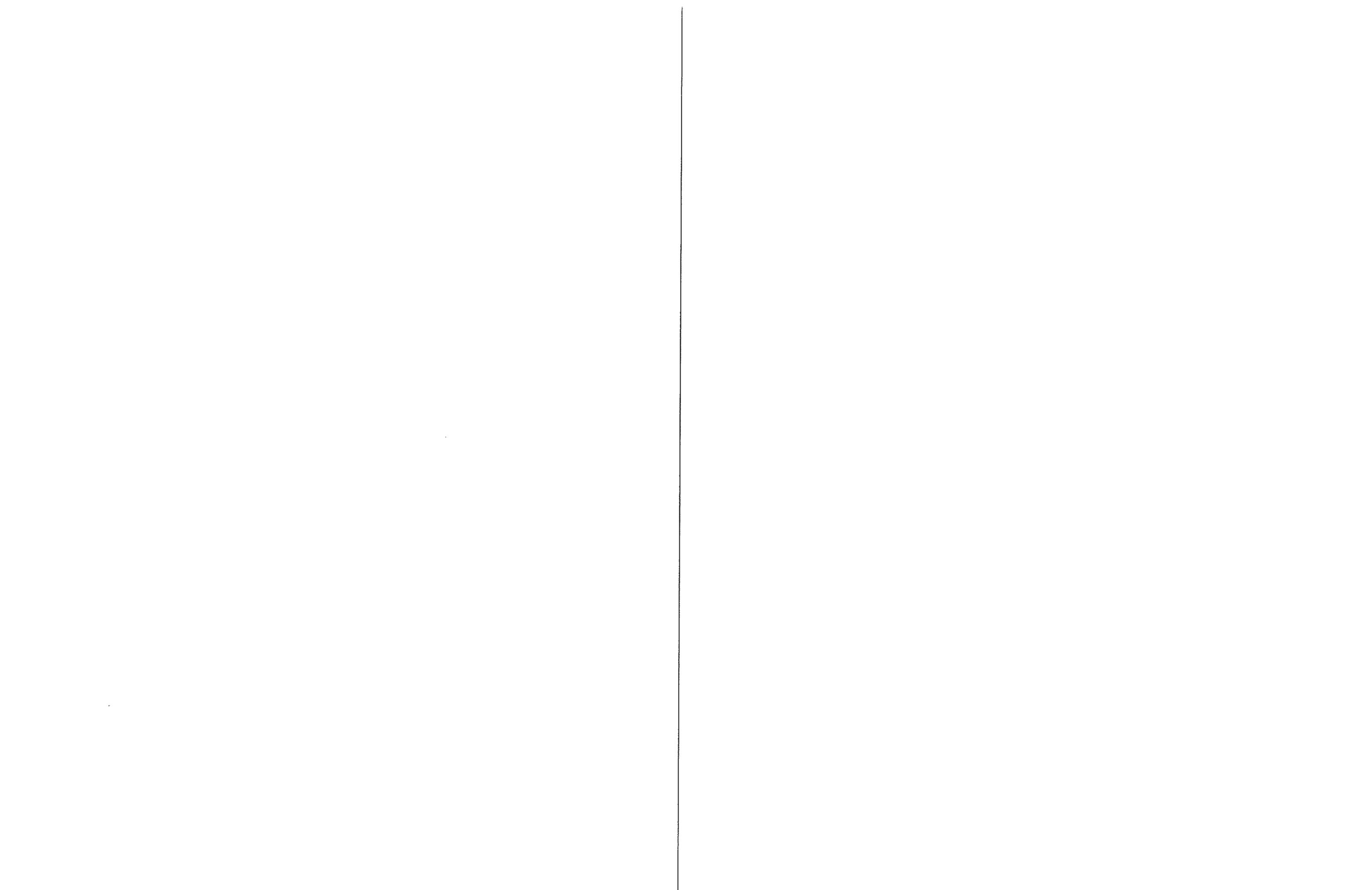
**Background:** The healthcare system has extremely large volumes of patients' paper medical records (PMRs) scattered throughout various medical facilities. Currently the industry is transitioning to Electronic Medical Records (EMR). Although each source of information is equally important, a complete longitudinal health record is rarely available or currently attainable. The goal of this effort is to convert the existing PMRs into searchable electronic equivalents and merge them with existing EMRs to create a more complete longitudinal health record to support clinical care and biomedical research.

**Method:** Using a subset of PMRs for subjects enrolled in the Clinical Breast Care Project at Walter Reed Army Medical Center, (WRAMC), we are developing an automated method to digitize and index the records, extract the biomedical information and prepare the data for delivery to clinicians and researchers. The electronic records were loaded into a database for immediate access by clinicians. To support translational research, MRs need to be de-identified, and information extracted and loaded into a research database; we used ten MRs to develop the operational method. Several methods were tested in the de-identification process: 1) manually, by striking out the protected health information (PHI) on paper before it was digitized and 2) "electronically", by redacting the record electronically on the computer after it was digitized. We are in the process of testing automated data extraction tools and natural language processors to automatically de-identify and extract data from the EMR.

**Result:** We quickly realized that only 10% of PMRs existed onsite at WRAMC; remaining MRs were held by the patients or other medical facilities. Approximately 300 PMRs, containing 66,600 pages, were scanned and digitized for this project. We have created a successful digitization process, which includes creating PDFs with hidden and searchable information. We have compared the effort and accuracy of various redaction methods. To de-identify 10 records, it took ~10 hours manually and ~20 hours electronically. It took longer electronically because of the preparations to ensure the removed information could not be retrieved. We expect that automated redaction tools will greatly reduce that effort. We also found that the electronic method had a 99% accuracy compared to 96% for the paper method. A portal prototype to allow access to medical records by clinicians and researchers is currently being tested and evaluated.

**Discussion:** Conversion of non-searchable data into an explorable and computable format will enable clinicians to acquire needed information more conveniently in their clinical care including treatment plan development. Similarly, properly de-identified, complete MRs will serve as a rich source of clinical information to support translational research. Although the method we are developing will initially satisfy the CBCP need for clinical

service and research, it will be further developed into a full solution to expand into other disease condition fields.



Role for CD44 in enhancing invasion, migration, and growth of triple negative (TN) breast cancer cells.

Joji Iida<sup>1</sup>, Matthew Nesbella<sup>1</sup>, John Lehman<sup>1</sup>, Richard Mural<sup>1</sup> and Craig Shriver<sup>2</sup>

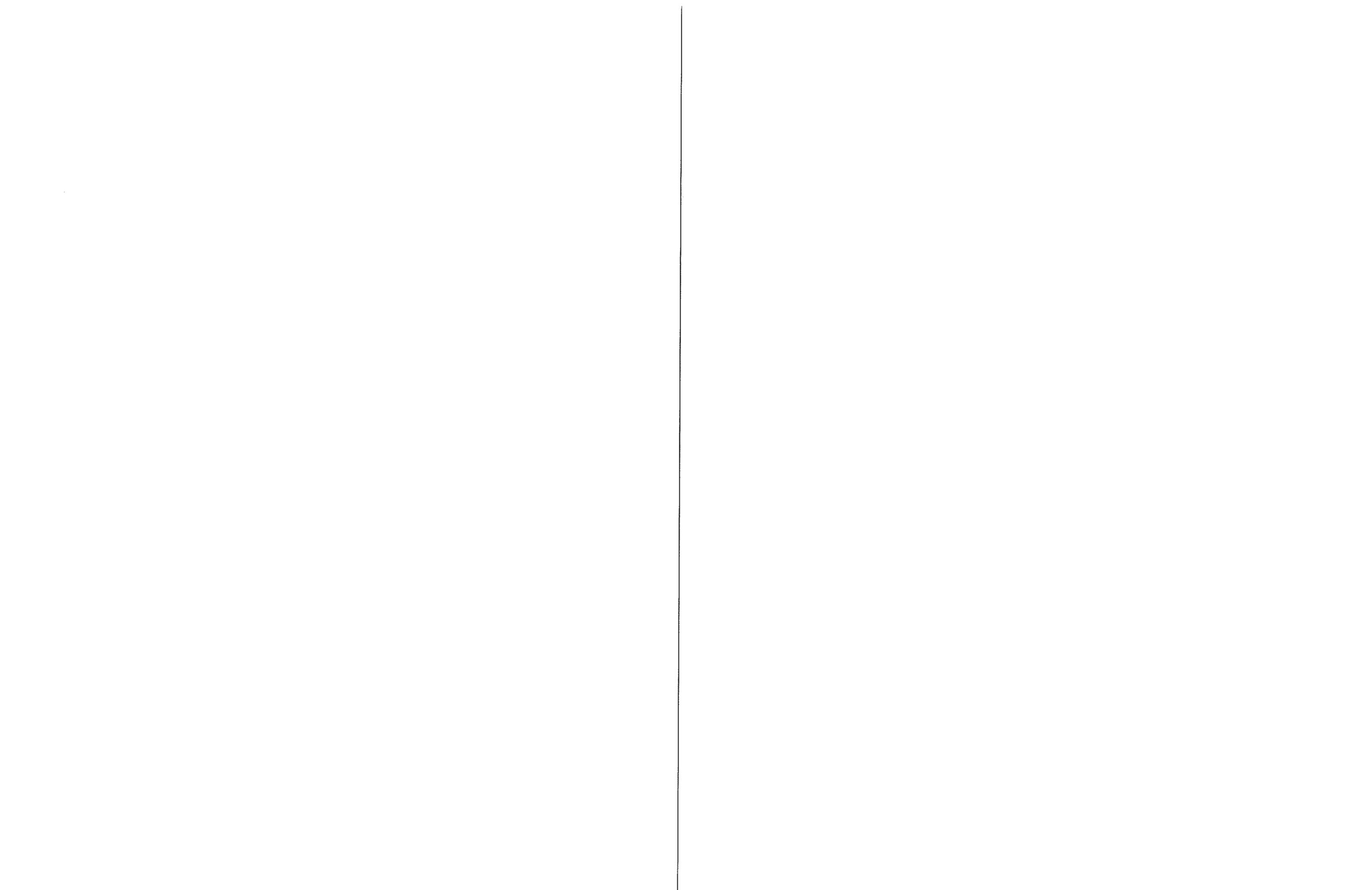
1: Department of Cell Biology, Windber Research Institute, Windber, PA 15963

2: Department of Surgery, Walter-Reed Army Medical Center, Washington, DC 20902

Neoplastic epithelial cells in breast carcinomas interact with various components in the tissue microenvironment including extracellular matrix (ECM) and mesenchymal cells. Recent studies identified CD44 as a metastasis-related molecule with multiple functions by promoting cell-cell and cell-ECM interactions. CD44 is an integral transmembrane protein encoded by a single 20-exon gene. In the standard form (CD44s), 10 of the 20 exons are translated. Multiple variant isoforms exist (CD44v1-10) which arises from alternate mRNA splicing of the remaining 10 exons. In contrast to the ubiquitous expression of the standard form of CD44, splice variants are highly restricted in their expression in normal or malignant tissues. Indeed, CD44 variants containing v3, v5, v6, v7-8, v10 exons are expressed in malignant breast cancer tissues. However, there is limited information regarding the biological functions of these exons to promote tumor invasion and metastasis.

Triple negative (TN) breast cancers are defined by a lack of expression of estrogen, progesterone, and her-2/neu receptors. It is widely recognized that TN breast cancers have a poorer prognosis than any other subtypes of breast cancer. Given the lack of effective targeted therapies for TN breast cancer patients, understanding of the mechanisms of growth and invasion in the tissues provides insight into developing novel approaches to lower the mortality from TN breast cancer. Importantly, recent studies suggest that more than half of TN patients had measurable CTCs. The goal of this study is to evaluate specific exon(s) of CD44 expressed on TN breast cancer cells as a model system for characterizing mechanisms of invasion, migration, and growth of breast cancer circulating tumor cells (CTCs).

In order to approach this goal, we utilized three TN cell lines (HCC38, HCC1937, and HCC1806) as model systems to evaluate CD44 in regulating invasion, migration, and growth in extracellular matrix (ECM) environments. Reverse transcriptase (RT)-PCR analysis using exon specific primers indicate that these cells expressed CD44v8-v10 and CD44s. We demonstrated that an inhibitory antibody against exon v10 of CD44 significantly inhibited  $\beta$ 1 integrin-mediated migration and invasion into Matrigel and type I collagen gel. Importantly, this antibody also inhibited three dimensional (3D) growth which is a  $\beta$ 1 integrin-independent process. The significant inhibition of these processes was also achieved when a FLAG-fusion exon v10 peptide (FLAG-v10, in which FLAG is tagged at the N-terminal of the peptide) was used as an inhibitor, implying that this exon would function to assemble molecular complexes on TN breast cancer cells that facilitate invasion, migration, and growth. Thus, these results suggest that generation of small synthetic molecules that block the functions of exon v10 of CD44 is promising approaches to inhibit metastasis of CTCs derived from TN breast cancer cells.



## Copy Number Variation in Low- and High-Grade Breast Tumors

Rachel E. Ellsworth<sup>1</sup>, Heather L. Patney<sup>2</sup>, Jennifer L. Kane<sup>2</sup>, Jeffery A. Hooke<sup>3</sup> and Craig D. Shriver<sup>3</sup>

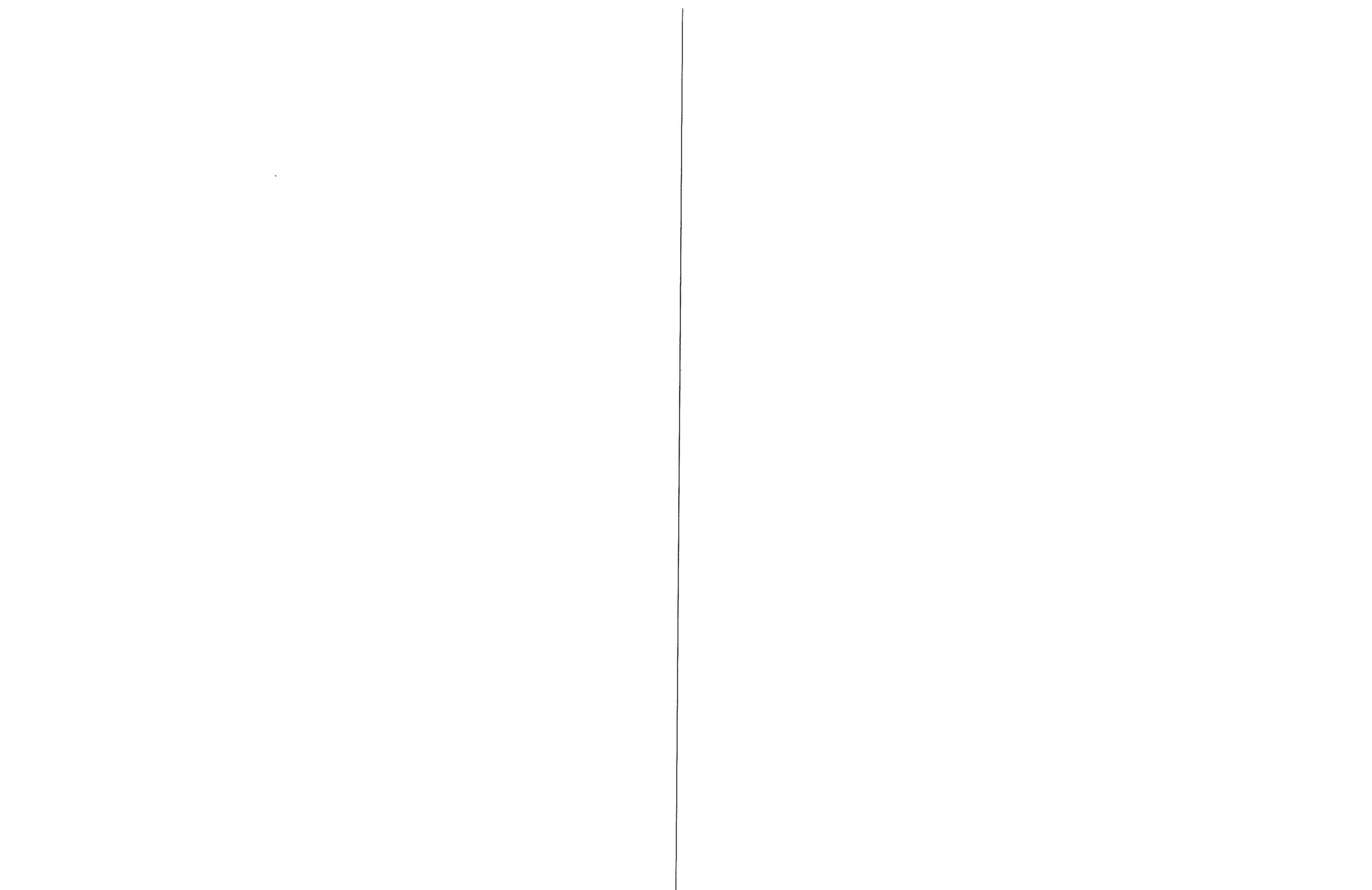
<sup>1</sup>Henry M. Jackson Foundation for the Advancement of Military Medicine, Windber, PA, <sup>2</sup>Windber Research Institute, Windber, PA and <sup>3</sup>Walter Reed Army Medical Center, Washington D.C.

**Objective** Histological grade classifies breast carcinomas into low-, intermediate- and high-grade. Developmental pathways for high-grade disease remain controversial; clinical data have been used to describe both linear models, in which high-grade tumors evolve from low-grade precursors, as well as models that depict low- and high-grade carcinomas as distinct molecular diseases. Genome-wide copy number alterations were examined in low- and high-grade breast tumors to identify genetic changes associated with the etiology of low- and high-grade breast carcinomas.

**Methods** Tumors were diagnosed and characterized by a single, dedicated breast pathologist. DNA was isolated after laser microdissection from low- (n= 53) and high-grade (n=68) tumors. Genotype data were generated using Human Mapping 250K Sty arrays (Affymetrix). Copy number alterations and LOH were detected using GenomeConsole 3.0.2 (Affymetrix).

**Results** The most common alterations in low-grade carcinomas were gain of chromosomes 1q (75%) and 16p (58%) and loss of chromosome 16q (86%). These alterations were frequently simple and included the entire chromosomal arm. In contrast, patterns of alterations were complex in high-grade tumors, frequently with alternating gains and losses on a single chromosomal arm. In high-grade tumors, the most frequent alterations included gain of chromosome 1q (85%), 8q (84%), and loss of chromosomes 8p (54%), 14q (57%) and 17p (69%). Loss of 16q ( $P < 0.0000005$ ) and gain of 16p ( $P < 0.005$ ) occurred significantly more frequently in low- compared to high-grade tumors, while gain of 8q24 ( $P < 0.000005$ ) and 10p15-p14 ( $P < 0.00005$ ) and loss of chromosome 17p13 ( $P < 0.00005$ ) occurred significantly more frequently in high- compared to low-grade tumors. Only two samples, one low- and one high-grade, had a "flat" profile with no detectable chromosomal alterations.

**Conclusion** Low- and high-grade breast tumors are genetically different. Gain of chromosome 1q is common in both low- and high-grade tumors, and may be an early event in tumorigenesis. Given the significantly different patterns of alterations, especially significantly higher loss of chromosome 16q in high- compared to low-grade tumors, high-grade disease may not arise from dedifferentiation of low-grade but rather represent a separate disease defined by unique genetic alterations.



Chromosomal alterations in pure early breast lesions: Implications for breast cancer progression

Rachel E. Ellsworth<sup>1</sup>, Darrell L. Ellsworth<sup>2</sup>, Jamie D. Weyandt<sup>2</sup>, Jamie L. Fantacone-Campbell<sup>3</sup>, Brenda Deyarmin<sup>2</sup>, Jeffrey A. Hooke<sup>3</sup>, Craig D. Shriver<sup>3</sup>

<sup>1</sup>Clinical Breast Care Project, Henry M. Jackson Foundation for the Advancement of Military Medicine, Windber, PA; <sup>2</sup>Clinical Breast Care Project, Windber Research Institute, Windber, PA; <sup>3</sup>Walter Reed Army Medical Center, Washington, DC

Contact for page proofs and correspondence:

Rachel E. Ellsworth  
Clinical Breast Care Project  
620 Seventh Street  
Windber, PA 15963  
Phone: (814) 361-6926  
FAX: (814) 467-6334  
Email: [r.ellsworth@wriwindber.org](mailto:r.ellsworth@wriwindber.org)

Short title: Genetics of early breast lesions

Keywords: columnar cell lesion, atypical ductal hyperplasia, allelic imbalance, breast cancer

**ABSTRACT**

**Introduction:** Columnar cell lesions (CCL) and atypical ductal hyperplasia (ADH) frequently coexist and share molecular changes with *in situ* and invasive components, suggesting that CCL and ADH may be precursors to breast cancer. These conclusions are, however, largely based on studies examining CCL and/or ADH from patients diagnosed with more advanced disease. Thus, we assessed allelic imbalance (AI) in pure CCL or ADH specimens to characterize molecular changes in early breast lesions.

**Methods:** DNA samples were obtained from laser microdissected CCL (n=42) or ADH (n=31) lesions without concurrent *in situ* or invasive disease. AI was assessed at 26 chromosomal regions commonly altered in breast cancer. Data was analyzed using Fisher's exact and Student's t-tests using a cutoff of  $P < 0.05$  to define significance.

**Results:** The average AI frequency was 6.2% (range 0-20%) in CCL and 6.1% (range 0-25%) in ADH. The highest levels of AI were observed on chromosomes 17q12-q21 in CCL (15%) and 8q24 in ADH (23%); ~33% of early lesions did not have any detectable genetic changes. Levels of AI in CCL and ADH were significantly ( $P < 0.0001$ ) lower than levels previously observed in either low- or high-grade DCIS lesions. Genetic changes characteristic of more advanced *in situ* and invasive disease, especially changes on chromosome 16q and 17p, were infrequent in pure early lesions.

**Conclusions:** Pure CCL and ADH lesions demonstrate lower levels of genetic alterations than DCIS, invasive carcinomas or early lesions from cancerous breasts. Alterations characteristic of advanced disease such as AI on chromosomes 16q and 17p were not detected in lesions from non-cancerous breasts. While chromosomal alterations in synchronous lesions have been used to develop a model of breast cancer progression

from CCL and ADH to *in situ* and invasive disease, pure early lesions do not appear to be genetically advanced and molecular profiles do not support these lesions as obligatory precursors to more advanced disease. Molecular differences between pure and synchronous lesions support re-evaluation of current models of disease initiation, progression and risk.

## INTRODUCTION

Each year in the United States alone, one million women are diagnosed with premalignant breast disease and more than half of all women over age 20 years will develop some form of benign breast disease during her lifetime <sup>1,2</sup>. Mammographic detection of columnar cell lesions (CCL) and atypical ductal hyperplasia (ADH) has increased in recent years as both are associated with mammographically-detectable microcalcifications <sup>3,4</sup>. The frequent finding of CCL with ADH and/or DCIS has led some investigators to suggest that detection of CCL warrants extensive pathological evaluation throughout the tumor specimen for more advanced disease <sup>3,5,6</sup>. In addition, frequent coexistence of CCL and ADH with *in situ* and invasive carcinomas, especially low-grade, suggests that these premalignant lesions are early precursors of invasive breast cancer <sup>2,3,5,7,8</sup>.

If CCL and ADH represent precursors of more advanced breast disease, these lesions should share at least some histologic and molecular characteristics with *in situ* and invasive carcinomas. CCL are cytologically similar to cells of co-existing DCIS and/or tubular carcinomas <sup>6,7,9</sup>, and ADH cytology is so similar to that of low-grade DCIS that classification of a lesion as ADH is dependent on the inability to diagnose DCIS <sup>4</sup>. Elevated estrogen receptor (ER), progesterone receptor (PR) and increased cellular proliferation, frequent markers of advanced disease, have been detected in CCL and ADH <sup>10-13</sup>. At the molecular level, alterations at chromosomes 11q, 16q and 3p have been found concomitantly in CCL and adjacent *in situ* and invasive breast cancer <sup>14,15</sup>, while alterations of chromosome 16q and 17p are frequently altered in both ADH and *in situ* and invasive breast carcinomas <sup>16-20</sup>.

While cytological, pathological and molecular characteristics shared between CCL/ADH and more advanced disease suggests that early lesions are precursors to breast cancer, the majority of these studies examined synchronous lesions from breasts containing *in situ* and/or invasive carcinomas. Given the advanced disease status, molecular characteristics of co-existing lesions may not accurately reflect early genetic alterations involved in breast pathogenesis and may be influenced by field cancerization effects. In this study, we assayed patterns of allelic imbalance (AI) in 73 pure CCL or ADH specimens from breasts without advanced disease to (1) identify chromosomal alterations associated with each type of early lesions, and (2) improve our understanding of molecular changes that occur during development of early breast lesions.

## **MATERIALS AND METHODS**

Paraffin-embedded specimens from patients diagnosed with CCL (n=42) or ADH (n=31) were obtained from the Clinical Breast Care Project (CBCP) Pathology Laboratory. Tissue and blood samples were collected with approval from the Walter Reed Army Medical Center Human Use Committee and Institutional Review Board (IRB). All subjects enrolled in the CBCP voluntarily agreed to participate and gave written informed consent. Clinical information was obtained for all CBCP samples using questionnaires designed by and administered under the auspices of the CBCP.

Diagnoses of all samples were made by a single breast pathologist from hematoxylin and eosin (H&E) stained slides. CCL were classified as either columnar cell changes or columnar cell hyperplasia, with or without atypia. The majority of cases had features of both columnar cell change and columnar cell hyperplasia. A diagnosis of

ADH was rendered when cells cytologically similar to low-grade DCIS coexisted with patterns of usual ductal hyperplasia and/or when there was partial involvement of the terminal duct lobular unit by characteristic morphology.

DNA was obtained from homogeneous populations of premalignant lesions following microdissection on an *ASLMD* laser microdissection system (Leica Microsystems, Wetzlar, Germany) (Figure 1) <sup>21</sup>. All microdissected sections were examined by the CBCP pathologist, who identified and marked regions of disease before microdissection. The integrity of multiple serial sections was established by pathological verification of the first and last sections stained with H&E. Referent DNA was obtained from blood clots using the Genra Clotspin and Puregene DNA purification kits (Qiagen Inc., Valencia, CA).

Fifty-two microsatellite markers from 26 chromosomal regions frequently altered in breast cancer were amplified and genotyped as previously described <sup>22</sup>. AI was detected using the formula  $(T1/T2)/(N1/N2)$  where T1 and N1 represent the peak heights of the less intense alleles and T2 and N2 represent the peak heights of the more intense alleles in the tumor and referent samples, respectively <sup>23</sup> (Figure 2). Samples with a normalized ratio of  $\leq 0.35$  were considered to show AI. Because each chromosomal region was represented by two polymorphic markers, AI in each region was classified according to the following criteria: (1) when at least one marker in a given region showed an allelic ratio  $\leq 0.35$ , the region was considered to show AI, (2) when neither marker had an allelic ratio  $\leq 0.35$  and at least one marker was informative, the region was considered normal, and (3) when both markers were homozygous, the region was considered uninformative.

Comparison of levels and patterns of AI were performed using Fisher's exact and Student's t-tests. A significance value of  $P < 0.05$  was used for all analyses.

## RESULTS

**Clinicopathologic features.** All samples were collected from female patients diagnosed with either CCL or ADH without concomitant *in situ* or invasive disease. No patients had a previous history of breast disease. The average age at diagnosis was 46.2 and 55.2 years ( $P < 0.005$ ) in patients with CCL and ADH, respectively. Samples were collected between 2001 and 2007; one patient diagnosed with columnar cell hyperplasia in 2001 was diagnosed with low-grade invasive ductal carcinoma in 2006.

**Overall AI frequencies.** Within the 72 lesions studied, 108 AI events were detected. The extent of AI ranged from 0 – 25%. Fourteen of the 42 (33%) CCL and 11/31 (35%) ADH lesions had no detectable AI events. When stratified by disease type, the average frequency of AI events - 6.2% in CCL and 6.1% in ADH – did not differ significantly.

**Frequency of AI by chromosomal region.** To investigate whether specific genetic alterations were associated with disease type, AI data were stratified by chromosomal location. The highest frequencies of AI were detected at chromosome 17q12-q21 in CCL (15%) and at chromosome 8q24 in ADH (23%). AI at chromosome 17p13.1 was not detectable in either CCL or ADH.

**Evaluation of chromosomal alterations within the biological continuum.** To determine whether the histological characterization of CCL and ADH as precursors to *in situ* and invasive breast cancer was supported at the molecular level, patterns and levels of AI in CCL and ADH were compared to AI data previously generated from disease-free, as well as low- and high-grade DCIS<sup>24,25</sup> (Table 1). The average level of AI in both CCL and ADH were significantly higher ( $P<0.005$ ) than in disease-free specimens (1.5%). AI frequencies did not differ between CCL and disease-free tissues at all 26 chromosomal regions, but were significantly higher ( $P<0.05$ ) at chromosome 8q24 in ADH compared to disease-free tissues.

When levels of AI in CCL and ADH were compared to those in low-( $n=19$ ) and high-grade ( $n=47$ ) DCIS, the average frequencies of AI in both low- (12%) and high-grade (26%) DCIS were significantly higher ( $P<0.0001$ ) than in CCL and ADH. Alterations of chromosome 16q are the most frequent alterations in low-grade DCIS (>50%), significantly fewer ( $P<0.005$ ) CCL and ADH lesions had detectable AI events. In addition, ADH had significantly fewer alterations at chromosome 13q12 than low-grade DCIS ( $P<0.05$ ). When compared to high-grade DCIS, levels of AI were significantly lower at 16/26 and 12/26 loci for CCL and ADH, respectively.

## DISCUSSION

Increased risk of developing invasive breast cancer has been associated with presence of CCL (~2-fold) and ADH (4-5-fold)<sup>26-29</sup>. Although detection of these precancerous lesions has increased with improvements in screening, diagnosis of CCL

and ADH is subjective, with features of CCL ranging from columnar cell alterations in luminal cells to flat/clinging DCIS, and diagnosis of ADH based on the inability to diagnose DCIS<sup>30</sup>. Molecular characterization of CCL and ADH may improve not only the ability to accurately diagnose, but to understand the etiology and progression of early breast lesions.

Studies evaluating genetic alterations in synchronous CCL and ADH lesions have observed genomic changes throughout the genome but frequently at 3p, 9 11q, 16q and 17p, leading to a model of breast cancer development in which that the columnar cell lesions are precursors to well-differentiated *in situ* and invasive breast carcinoma<sup>15,31</sup>. Similarly, recurrent alterations on chromosomes 1q, 8p, 16q, 17p and 17q in ADH lesions<sup>17-19</sup> suggest that ADH may be direct cancer precursors<sup>16</sup>. Together, these studies suggest that CCL and ADH share molecular changes with more advanced disease, especially loss of chromosome 16q, providing molecular support to a histological continuum of disease.

Shared genetic alterations between early lesions and advanced disease have largely been generated in synchronous lesions from cancerous breasts<sup>31</sup>. When synchronous lesions are compared to pure lesions from otherwise non-diseased breasts, it is clear that pure and synchronous lesions differ extensively at the molecular level. For example, in a study of 399 premalignant breast lesions, 26 ADH lesions were isolated from non-cancerous and 25 from cancerous breasts; ADH from cancerous breasts had detectable AI at 13/15 (87%) of loci and increased AI at chromosome 11p in synchronous (38%) compared to pure (15%) ADH. Because nearly 45% of the ADH lesions shared an AI event with the synchronous cancer, ADH could be viewed as a precursor to more advanced disease. In contrast, the frequency of AI in pure ADH was low<sup>20</sup>, suggesting

that pure ADH may be different at the molecular level from synchronous lesions. Similarly, benign proliferative lesions show few gross genomic alterations compared to *in situ* carcinomas, suggesting that benign lesions, such as ductal hyperplasias and papillomas, are not obligate direct precursors of more advanced disease<sup>32</sup>. In a study evaluating genetic changes in 14 pure ADH specimens, Rosenberg et al found that 40% of lesions had detectable alterations, suggesting ADH is comprised of monoclonal cell populations. These genetic alterations need not suggest that ADH is an obligate precursor of malignancy, rather, the presence of multiple monoclonal lesions in a single breast may reflect field defects within the breast<sup>33</sup>. Even within *in situ* carcinomas, which are genetically more advanced than CCL or ADH, pure DCIS demonstrates a divergent molecular profile from synchronous DCIS, the molecular profile of synchronous DCIS is quite similar to that from invasive lesions<sup>34,35</sup>. Thus, pure lesions are not equivalent to those found in cancerous breasts and models of tumor progression developed from data from synchronous lesions may not accurately reflect breast cancer etiology.

Our study, which included only pure early lesions, supports the idea that pure early breast lesions are molecularly distinct from synchronous lesions found in cancerous breasts. While studies of synchronous lesions have found high frequencies of AI on chromosomes 16q and 17p, similar to levels found in invasive disease<sup>30</sup>, AI in this study was infrequent at 16q and non-detectable at 17p13.1 in both CCL and ADH. Alterations of chromosome 16q, present in >50% of low-grade DCIS, were detected in only 7% of CCL and 10% of ADH and were not significantly different from background levels found in disease-free tissues. Physical loss of chromosome 16q is frequently the only detectable

genetic change in low-grade breast disease<sup>36</sup>; thus while a small subset of early lesions have with detectable AI at 16q may be genetically programmed to progress to low-grade disease, the majority of early CCL and ADH do not exhibit this hallmark of low-grade DCIS.

In addition, overall levels of AI were significantly higher in both CCL and ADH compared to disease-free tissues and significantly lower than low- and high-grade DCIS (Figure 3). Of note, although ~1/3 of patients with either CCL or ADH had no detectable AI events, 2-21% of early lesions had levels of AI similar to that seen in low-(12%) and high-grade DCIS (19%), respectively. These more “genetically advanced” lesions with high levels of AI may be more likely to progress, however, none of these patients have, to date, had a recurrence or progression of disease.

The differences in levels of genetic alterations may explain why early breast disease has a higher cure rate than the preinvasive DCIS. Patients with ADH treated with tamoxifen or other chemopreventive agents have a higher proportional risk reduction (86%) compared to patients diagnosed with DCIS (~40%)<sup>37,38</sup>. In addition, numerous studies have shown that the risk of recurrence does not increase when ADH is excised with positive margins while re-excision is clinically warranted for patients with DCIS. Clinical management of CCL as DCIS is not recommended as this is likely to result in the overtreatment of many patients<sup>39</sup>.

The low levels of genomic alterations of pure CCL and ADH may reflect the relatively young age of these lesions, as genetic alterations associated with growth and invasion have not yet accumulated. In contrast, synchronous lesions are genetically

advanced, sharing many alterations with *in situ* and invasive carcinomas. It is possible that these differences reflect the more advanced nature of synchronous lesions from cancerous breasts; alterations of chromosomal regions such as 16q, 17p and 17q may stimulate proliferation and invasion, providing these lesions with a genetically predetermined ability to progress quickly from preinvasive to invasive disease.

Alternatively, the similar chromosomal alterations found amongst lesions and carcinomas of the same breast may reflect field effects. First proposed in 1953 by Slaughter et al., field cancerization describes the presence of histologically abnormal tissues surrounding carcinomas and subsequent local recurrence of disease<sup>40</sup>. Briefly, a stem cell is thought to accumulate initiating genetic mutations, which are passed on to daughter cells. These cells escape the normal mechanisms of cell cycle control and expand, becoming a field of altered cells. As the field grows, additional genetic changes accumulate throughout the field, resulting in a variety of subclones, now developing independently from one another. These subclones have the ability to progress to different types of lesions<sup>41</sup>. Genomic changes shared between early lesions and synchronous invasive carcinomas, may, therefore reflect field effects and the unstable genome of breast tissue in a cancerous breast.

Overall levels of AI in ADH were more similar to those in disease-free tissues than DCIS at the molecular level, however, 23% of ADH specimens showed AI at chromosome 8q24, similar to DCIS. The v-myc myelocytomatosis viral oncogene homolog (MYC, NM\_002467) gene which maps to the 8q24 region, has been associated with altered cellular proliferation, growth, differentiation and metabolism<sup>42</sup>.

Amplification of c-MYC may contribute to progression of high-grade disease through

increased levels of proteolytic enzymes such as Cathepsin D and matrix metalloproteinases<sup>43,44</sup>, leading to disruption and degradation of the extracellular matrix and promoting cellular invasion. Larger retrospective studies measuring the extent of AI at chromosome 8q24 in ADH specimens from patients with disease recurrence will help determine whether assessment of c-MYC or other genes from this region can be used to identify those lesions likely to progress.

In conclusion, pure CCL and ADH lesions demonstrate significantly lower levels of genetic alterations than *in situ* or invasive carcinomas as well as synchronous early lesions from cancerous breasts. Molecular alterations previously reported to characterize early lesions such as alterations of chromosomes 16q and 17p were not found in lesions from non-cancerous breasts. While data generated from synchronous lesions has been used as support for a model of breast cancer progression from CCL and ADH to *in situ* and invasive disease, our data suggest that pure early lesions are not genetically advanced and may not represent obligatory precursors to more advanced disease. Given the molecular differences between pure and synchronous lesions, current models of disease initiation, progression and risk based largely on data generated from cancerous breasts should be re-evaluated.

**ACKNOWLEDGEMENTS**

This research was supported by a grant from the United States Department of Defense (Military Molecular Medicine Initiative MDA W81XWH-05-2-0075, Protocol 01-20006). The opinion and assertions contained herein are the private views of the authors and are not to be construed as official or as representing the views of the Department of the Army or the Department of Defense.

## References

1. MedTech Insight. MedTech Insight Report: U.S. breast disease detection and diagnostic technologies. A400. 2002. Newport Beach, CA.  
Ref Type: Report
2. Wellings SR, Jensen HM, Marcum RG. An atlas of subgross pathology of the human breast with special reference to possible precancerous lesions. *J Natl Cancer Inst* 1975; 55:231-273.
3. Fraser JL, Raza S, Chorny K, Connolly JL, Schnitt SJ. Columnar alteration with prominent apical snouts and secretions: a spectrum of changes frequently present in breast biopsies performed for microcalcifications. *Am J Surg Pathol* 1998; 22:1521-1527.
4. Pinder SE, Ellis IO. The diagnosis and management of pre-invasive breast disease: ductal carcinoma in situ (DCIS) and atypical ductal hyperplasia (ADH)--current definitions and classification. *Breast Cancer Res* 2003; 5:254-257.
5. Schnitt SJ, Vincent-Salomon A. Columnar cell lesions of the breast. *Adv Anat Pathol* 2003; 10:113-124.
6. Oyama T, Iijima K, Takei H, Horiguchi J, Iino Y, Nakajima T, Koerner F. Atypical cystic lobule of the breast: an early stage of low-grade ductal carcinoma in-situ. *Breast Cancer* 2000; 7:326-331.
7. Goldstein NS, O'Malley BA. Cancerization of small ectatic ducts of the breast by ductal carcinoma in situ cells with apocrine snouts: a lesion associated with tubular carcinoma. *Am J Clin Pathol* 1997; 107:561-566.
8. Abdel-Fatah TMA, Powe DG, Hodi Z, Lee AHS, Reis-Filho JS, Ellis IO. High frequency of coexistence of columnar cell lesions, lobular neoplasia, and low grade ductal carcinoma in situ with invasive tubular carcinoma and invasive lobular carcinoma. *Am J Surg Pathol* 2007; 31:417-426.
9. Moinfar F, Man YG, Arnold L, Bratthauer GL, Ratschek M, Tavassoli FA. Concurrent and independent genetic alterations in stromal and epithelial cells of mammary carcinoma: implications for tumorigenesis. *Cancer Res* 2000; 60:2562-2566.
10. Allred DC, Mohsin SK, Fuqua SAW. Histological and biological evolution of human premalignant breast disease. *Endocrine-Related Cancer* 2001; 8:47-61.
11. Anderson E, Clarke RB. Steroid receptors and cell cycle in normal mammary epithelium. *J Mammary Gland Biol Neoplasia* 2004; 9:3-13.
12. Conneely OM, Jericevic BM, Lydon JP. Progesterone receptors in mammary gland development and tumorigenesis. *J Mammary Gland Biol Neoplasia* 2003; 8:205-214.
13. Shoker BS, Jarvis C, Clarke RB, Anderson E, Hewlett J, Davies MP, Sibson DR, Sloane JP. Estrogen receptor-positive proliferating cells in the normal and precancerous breast. *Am J Pathol* 1999; 155:1811-1815.
14. Moinfar F, Man YG, Bratthauer GL, Ratschek M, Tavassoli FA. Genetic abnormalities in mammary ductal intraepithelial neoplasia-flat type ("clinging ductal carcinoma in situ"): a simulator of normal mammary epithelium. *88* 2000; 9:2081.

15. Simpson PT, Gale T, Reis-Filho JS, Jones C, Parry S, Sloane JP, Hanby A, Pinder SE, Lee AHS, Humphreys S, Ellis IO, Lakhani SR. Columnar cell lesions of the breast: the missing link in breast cancer progression? *Am J Surg Pathol* 2005; 29:734-746.
16. Amari M, Suzuki A, Monya T, Yoshinga K, Amano G, Sasano H, Ohuchi N, Satomi S, Horii A. LOH analyses of premalignant and malignant lesions of human breast: frequent LOH in 8p, 16q, and 17q in atypical ductal hyperplasia. *Oncol Rep* 1999; 6:1277-1280.
17. Gong G, DeVries S, Chew KL, Cha I, Ljung BM, Waldman FM. Genetic changes in paired atypical and usual ductal hyperplasia of the breast by comparative genomic hybridization. *Clin Cancer Res* 2001; 7:2410-2414.
18. Lakhani SR, Collins N, Stratton MR, Sloane JP. Atypical ductal hyperplasia of the breast: clonal proliferation with loss of heterozygosity on chromosomes 16q and 17p. *J Clin Pathol* 1995; 48:611-615.
19. Larson PS, de las Morenas A, Cerda SR, Bennett SR, Cupples LA, Rosenberg CL. Quantitative analysis of allele imbalance supports atypical ductal hyperplasia lesions as direct breast cancer precursors. *J Pathol* 2006; 209:307-316.
20. O'Connell P, Pekkel V, Fuqua SAW, Osborne CK, Allred DC. Analysis of loss of heterozygosity in 399 premalignant breast lesions at 15 genetic loci. *J Natl Cancer Inst* 1998; 90:697-703.
21. Ellsworth DL, Shriver CD, Ellsworth RE, Deyarmin B, Somiari RI. Laser capture microdissection of paraffin-embedded tissues. *Biotechniques* 2003; 34:42-46.
22. Ellsworth RE, Ellsworth DL, Lubert SM, Hooke J, Somiari RI, Shriver CD. High-throughput loss of heterozygosity mapping in 26 commonly deleted regions in breast cancer. *Cancer Epidemiol Biomarkers Prev* 2003; 12:915-919.
23. Medintz IL, Lee C-CR, Wong WW, Pirkola K, Sidransky D, Mathies RA. Loss of heterozygosity assay for molecular detection of cancer using energy-transfer primers and capillary array electrophoresis. *Genome Res* 2000; 10:1211-1218.
24. Ellsworth RE, Ellsworth DL, Deyarmin B, Hoffman LR, Love B, Hooke JA, Shriver CD. Timing of critical genetic changes in human breast disease. *Annals Surg Onc* 2005; 12:1054-1060.
25. Ellsworth RE, Ellsworth DL, Love B, Patney HL, Hoffman LR, Kane JL, Hooke JA, Shriver CD. Correlation of levels and patterns of genomic instability with histological grading of DCIS. *Annals Surg Onc* 2007; 14:3070-3077.
26. Dupont WD, Page DL. Risk factors for breast cancer in women with proliferative breast disease. *New England Journal of Medicine* 1985; 312:146-151.
27. Page DL, Dupont WD, Rogers LW. Breast cancer risk of lobular-based hyperplasia after biopsy: "ductal" pattern lesions. *Cancer Detect Prev* 1986; 9:441-448.
28. Shaaban AM, Sloane JP, West CR, Moore FR, Jarvis C, Williams EM, Foster CS. Histopathologic types of benign breast lesions and the risk of breast cancer: case-control study. *Am J Surg Pathol* 2002; 26:421-430.
29. McLaren BK, Gobbi H, Schuyler PA, Olson SJ, Parl FF, Dupont WD, Page DL. Immunohistochemical expression of estrogen receptor in enlarged lobular units with columnar alteration in benign breast biopsies: a nested case-control study. *Am J Surg Pathol* 2005; 29:105-108.

30. Reis-Filho JS, Lakhani SR. The diagnosis and management of pre-invasive breast disease: genetic alterations in pre-invasive lesions. *Breast Cancer Res* 2003; 5:313-319.
31. Dabbs DJ, Carter G, Fudge M, Peng Y, Swalsky P, Finklestein S. Molecular alterations in columnar cell lesions of the breast. *Mod Pathol* 2006; 19:344-349.
32. Boecker W, Buerger H, Schmitz K, Eliis IA, van Diest PJ, Sinn HP, Geradts J, Diallo R, Poremba C, Herbst H. Ductal epithelial proliferations of the breast: a biological continuum? Comparative genomic hybridization and high-molecular -weight cytokeratin expression patterns. *J Pathol* 2001; 195:415-421.
33. Rosenberg CL, de las Morenas A, Huang K, Cupples A, Faller DV. Detection of monoclonal microsatellite alterations in atypical breast hyperplasia. *Journal of Clinical Investigation* 1996; 98:1095-1100.
34. Castro NP, Osorio CABT, Torres C, Bastos EP, Mourao-Netto M, Soares FA, Brentani HP, Carraro DM. Evidence that molecular changes in cells occur before morphological alterations during the progression of breast ductal carcinoma. *Breast Cancer Res* 2008; 10:R87.
35. Iakovlev VV, Arneson NCR, Wong V, Wang C, Leung S, Iakovleva G, Warren K, Pintilie M, Done SJ. Genomic differences between pure ductal carcinoma in situ of the breast and that associated with invasive disease: a calibrated aCGH study. *Clin Cancer Res* 2008; 14:4446-4454.
36. Rakha EA, Green AR, Powe DG, Roylance R, Ellis IO. Chromosome 16 tumor-suppressor genes in breast cancer. *Genes Chromosomes Cancer* 2006; 45:527-535.
37. Fisher B, Costantino JP, Wickerham DL, Redmond CK, Kavanah M, Cronin WM, Vogel V, Robidoux A, Dimitrov N, Atkins J, Daly M, Wieand S, Tan-Chiu E, Ford L, Wolmark N, National Surgical Adjuvant Breast and Bowel Project Investigators. Tamoxifen for prevention of breast cancer: report of the National Surgical Adjuvant Breast and Bowel Project P-1 Study. *J Natl Cancer Inst* 1998; 90:1371-1388.
38. Fisher B, Dignam J, Wolmark N, Wickerham DL, Fisher ER, Mamounas E, Smith R, Begovic M, Dimitrov NV, Margolese RG, Kardinal CG, Kavanah MT, Fehrenbacher L, Oishi RH. Tamoxifen in treatment of intraductal breast cancer: National Surgical Adjuvant Breast and Bowel Project B-24 randomised controlled trial. *The Lancet* 1999; 353:1993-2000.
39. Schnitt SJ. The diagnosis and management of pre-invasive breast disease: flat epithelial atypia- classification, pathologic features and clinical significance. *Breast Cancer Res* 2003; 5:263-268.
40. Slaughter DP, Southwick HW, Smejkal W. "Field cancerization" in oral stratified squamous epithelium. *Cancer* 1953; 6:963-968.
41. Braakhuis BJ, Tabor MP, Kummer JA, Leemans CR, Brakenhoff RH. A genetic explanation of Slaughter's concept of field cancerization: evidence and clinical implications. *Cancer Res* 2003; 63:1727-1730.
42. Levens D. Disentangling the MYC web. *Proc Natl Acad Sci USA* 2002; 99:5757-5759.
43. Brouillet JP, Theillet C, Maudelonde T, Defrenne A, Simony-Lafontaine J, Sertour J, Pujol H, Jeanteur P, Rochefort H. Cathepsin D assay in primary breast cancer and lymph nodes: relationship with c-myc, c-erb-B-2 and int-2 oncogene amplification and node invasiveness. *Eur J Cancer* 1990; 26:437-441.

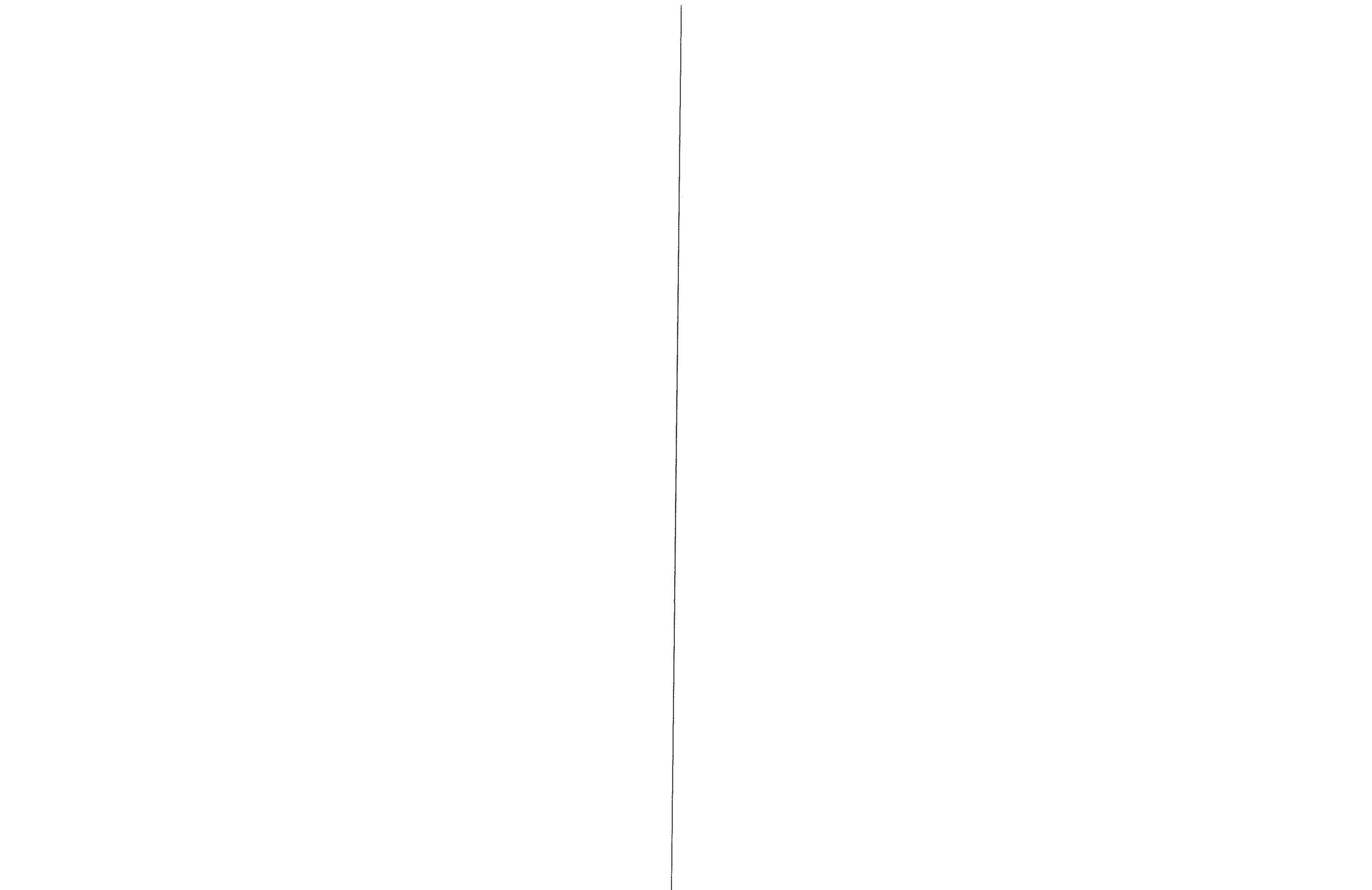
44. Giambernardi TA, Grant GM, Taylor GP, Hay RJ, Maher VM, McCormick JJ, Klebe RJ. Overview of matrix metalloproteinase expression in cultured human cells. *Matrix Biol* 1998; 16:483-496.

## FIGURE LEGEND

Figure 1. Hematoxylin and eosin stained sections showing the specificity of laser microdissection in obtaining homogenous populations of diseased cells. Panels A-B, CCL showing regions of both columnar cell change and columnar cell hyperplasia, without atypia; panels C-D; atypical ductal hyperplasia.

Figure 2. Detection of allelic imbalance in breast disease with fluorescence-based genotyping. Alleles for marker D3S1566 on chromosome 3p14 were detected as fluorescent peaks in referent DNA (top) and microdissected CCH DNA (bottom). Two peaks were detected in the reference sample at 162 and 164 bp, while the 164 bp peak is missing in the CCH sample.

Figure 3. Genomic instability (% informative markers showing AI) across a continuum of pre-invasive breast disease. Significant differences in AI were observed between disease-free (DF) tissues and early lesions ( $P < 0.005$ ) and between early lesions and grade 1 (G1) DCIS and grade 3 (G3) DCIS ( $P < 0.0001$ ).



# Chromosomal Alterations in Pure Nonneoplastic Breast Lesions: Implications for Breast Cancer Progression

Rachel E. Ellsworth PhD, Darrell L. Ellsworth PhD, Jamie D. Weyandt BS, Jamie L. Fantacone-Campbell MS, Brenda Deyarmin HT(ASCP), Jeffrey A. Hooke MD, Craig D. Shriver MD

Translational Research and Biomarkers

Online First™ - January , 2010

[View full article HTML](#) | [View full article PDF](#) | [Download Citation](#)

## Abstract

### Introduction

Columnar cell lesions (CCL) and atypical ductal hyperplasia (ADH) frequently coexist and share molecular changes with in situ and invasive components, suggesting that CCL and ADH may be precursors to breast cancer. These conclusions are largely based on studies examining CCL and/or ADH from patients diagnosed with more advanced disease. We assessed allelic imbalance (AI) in pure CCL or ADH specimens to characterize molecular changes in nonneoplastic breast lesions.

### Methods

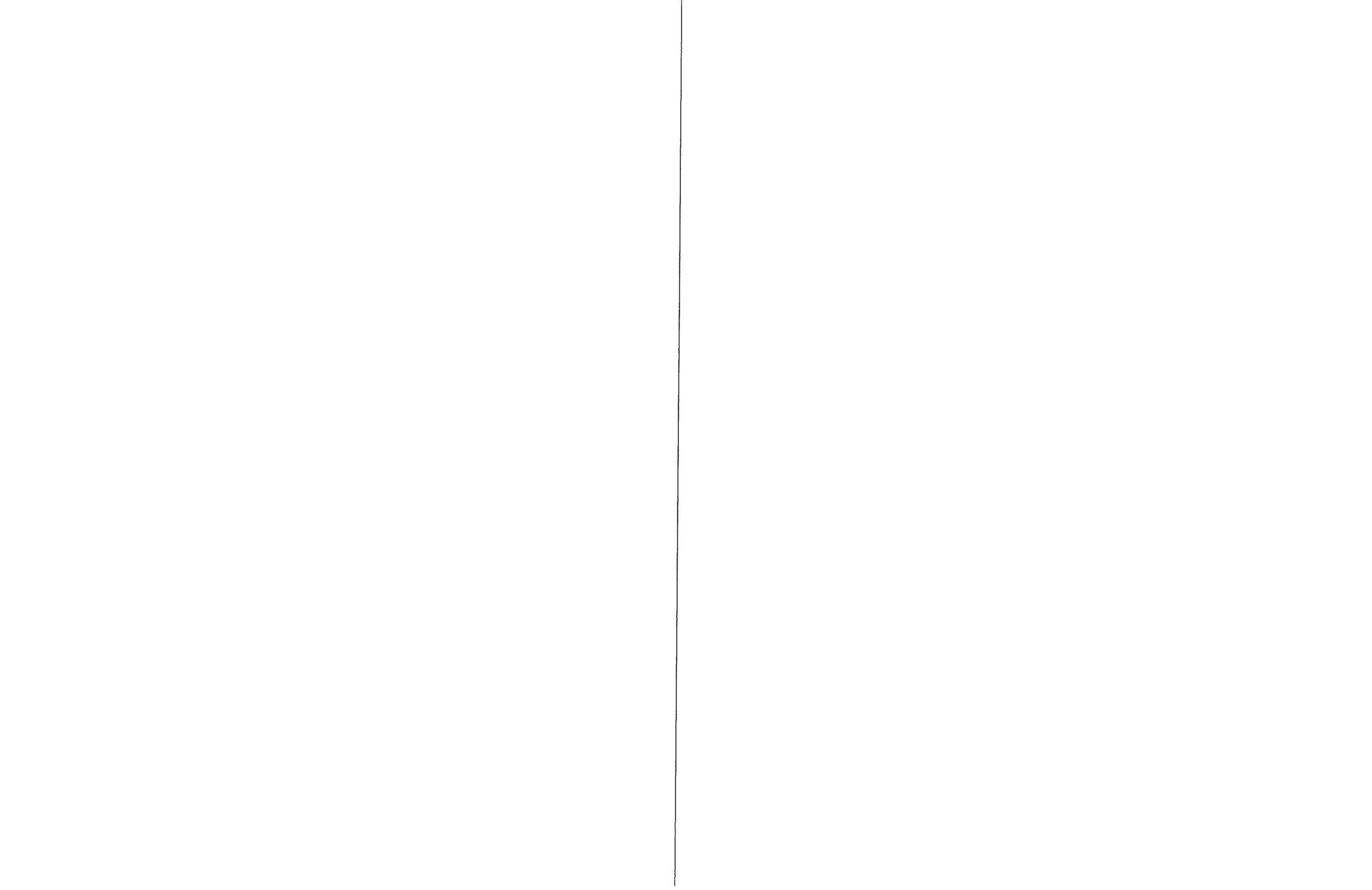
DNA samples were obtained from laser-microdissected pure CCL ( $n = 42$ ) or ADH ( $n = 31$ ). AI was assessed at 26 chromosomal regions commonly altered in breast cancer. Data were analyzed using Fisher's exact and Student's  $t$ -tests using a cutoff of  $P < 0.05$ .

### Results

The average AI frequency was 6.2% in CCL and 6.1% in ADH; ~33% of nonneoplastic lesions had no detectable genetic changes. Levels of AI in CCL and ADH were significantly ( $P < 0.0001$ ) lower than observed in either low- or high-grade ductal carcinoma in situ (DCIS) lesions. Genetic changes characteristic of in situ and invasive disease, especially on chromosomes 16q and 17p, were infrequent in pure nonneoplastic lesions.

### Conclusions

Pure CCL and ADH lesions demonstrate lower levels of genetic alterations than DCIS, invasive carcinomas or CCL/ADH lesions from cancerous breasts; alterations of chromosomes 16q and 17p were not detected. Pure CCL and ADH lesions are not genetically advanced, and molecular profiles do not support these lesions as obligatory precursors to more advanced disease. Molecular differences between pure and synchronous lesions support re-evaluation of current models of disease initiation, progression, and risk.



## Breast Cancer in the Personal Genomics Era

Rachel E. Ellsworth<sup>1</sup>, David J. Decewicz<sup>2</sup>, Craig D. Shriver<sup>3</sup> and Darrell L. Ellsworth<sup>\*,2</sup>

<sup>1</sup>Clinical Breast Care Project, Henry M. Jackson Foundation for the Advancement of Military Medicine, Windber, PA, USA; <sup>2</sup>Clinical Breast Care Project, Walter Reed Army Medical Center, Washington, DC, USA and <sup>3</sup>Clinical Breast Care Project, Windber Research Institute, Windber, PA, USA

**Abstract:** Breast cancer is a heterogeneous disease with a complex etiology that develops from different cellular lineages, progresses along multiple molecular pathways, and demonstrates wide variability in response to treatment. The "standard of care" approach to breast cancer treatment in which all patients receive similar interventions is rapidly being replaced by personalized medicine, based on molecular characteristics of individual patients. Both inherited and somatic genomic variation is providing useful information for customizing treatment regimens for breast cancer to maximize efficacy and minimize adverse side effects. In this article, we review (1) hereditary breast cancer and current use of inherited susceptibility genes in patient management; (2) the potential of newly-identified breast cancer-susceptibility variants for improving risk assessment; (3) advantages and disadvantages of direct-to-consumer testing; (4) molecular characterization of sporadic breast cancer through immunohistochemistry and gene expression profiling and opportunities for personalized prognostics; and (5) pharmacogenomic influences on the effectiveness of current breast cancer treatments. Molecular genomics has the potential to revolutionize clinical practice and improve the lives of women with breast cancer.

Received on: November 30, 2009 - Revised on: January 24, 2010 - Accepted on: January 26, 2010

**Keywords:** Breast cancer, personal genomics, genetic tests, gene expression, risk assessment.

### INTRODUCTION

Breast cancer is the most frequently occurring cancer and the leading cause of death in women between 20 and 59 years of age in the United States [1]. Over the last fifty years, the incidence of breast cancer has increased dramatically, such that today one in eight women are expected to develop breast cancer during her lifetime. Last year in the United States, more than 40,000 women died from breast cancer [2] and the cost of treatment for breast cancer patients exceeded \$8 billion [3].

Breast cancer has a complex etiology where susceptibility is influenced by both environmental and genetic factors. Considerable experimental and epidemiological evidence suggests that lifetime exposure to endogenous hormones, notably estrogens and androgens, promotes breast carcinogenesis. In population-based studies, factors related to increased estrogen exposure throughout a woman's lifetime, such as early menarche, late menopause, use of oral contraceptives, and hormone replacement therapy, have been associated with a ~2-fold increase in breast cancer risk among premenopausal women [4-6]. Other risk factors including age, family history, late age (>30 years) at first pregnancy or never being pregnant, and high breast density [7], as well as modifiable risk factors such as nutrition [8], exercise, and alcohol/tobacco use are also important in defining risk for breast cancer.

Heterogeneity in clinical, pathological, and molecular characteristics makes breast cancer a challenging disease to

manage. Pathological characterization of breast carcinomas includes a number of variables such as tissue architecture, cellular differentiation, and size/presence of local or distant metastasis, which influence disease progression, risk assessment, and prognosis. Other variables including estrogen receptor (ER), progesterone receptor (PR), and epidermal growth factor receptor-2 (HER2) expression vary widely among breast cancer patients and thus are routinely assessed as a component of standard care for determining the most effective treatment. At the molecular level, breast cancer has been found to exhibit extensive variability, with at least five tumor subtypes identified by patterns of gene expression [9, 10]. In women with similar pathological characteristics, molecular heterogeneity of breast disease may cause clinical outcomes to vary widely even though patients receive identical treatments [11].

The extensive molecular and pathological diversity observed in breast cancer patients suggests that breast cancer is not a homogeneous disease that can be effectively managed by a "standard of care" approach. Breast cancer may be more appropriately defined as a myriad of diseases characterized by variability in developmental pathways, propensity to metastasize, and response to treatment that can only be successfully treated by regimens targeted to individual patients. Personalized medicine provides care and treatment based on fixed and modifiable risk factors unique to each patient, as well as pathological and molecular characteristics that make each breast carcinoma unique. In this review, we critically examine the role of inherited and somatic genomic variation in breast cancer, outlining the predictive utility of susceptibility variants, extent and potential information content of cancer genomics, and the promise of personalized medicine and personalized oncology.

\*Address correspondence to this author at the Clinical Breast Care Project, Windber Research Institute, 620 Seventh Street, Windber, PA 15963, USA; Tel: 814-361-6911; Fax: 814-467-6334; E-mail: d.ellsworth@wriwindber.org

## HEREDITARY BREAST CANCER

### Identification of Susceptibility Genes

The idea that breast cancer has a familial or inherited component was proposed as early as 1757 when Le Dran described a 19-year old woman with breast cancer whose grandmother and maternal uncle died of breast cancer [12]. In 1866, Broca characterized a family with ten women across four generations who were affected with breast cancer, and provided sufficient data to create a family pedigree that clearly showed the heritable nature of the disease [13]. More recently, twin studies and segregation/risk analysis have provided additional evidence that the development of breast cancer has a genetic component [14-17].

Identification of genes associated with the development of breast cancer is complicated by the co-occurrence of sporadic and heritable breast cancer within families. Because breast cancer affects one in eight women, approximately 11% of families will contain more than one female with breast cancer [18], but it may be challenging to distinguish disease attributable to an inherited cancer susceptibility gene from chance clustering of sporadic breast cancer cases within a family. To identify breast cancer susceptibility genes, large extended families meeting stringent criteria for defining heritable breast cancer were needed. Using this approach, 23 extended Caucasian families containing 146 individuals with breast cancer including early onset cases, bilateral disease, and/or male breast cancer were assembled. Forty percent of these families showed strong linkage to a marker on chromosome 17q21 [19]. In 1994, a novel gene with a zinc-finger domain and wide tissue expression, including breast and ovary, was identified as the breast cancer 1 gene (BRCA1) [20].

Despite the strong contribution of BRCA1 to hereditary breast and ovarian cancer, mutations in BRCA1 do not account for all cases of inherited breast cancer, implicating the existence of a second major susceptibility gene. Using techniques similar to those used in the discovery of BRCA1, a novel gene on chromosome 13q12-q13 was identified as BRCA2 in 1995 [21, 22].

Following the identification of BRCA1 and BRCA2, screening tests were developed to identify mutation carriers and families at risk for hereditary breast and ovarian cancer. Screening is now recommended for women with breast cancer diagnosed at an early age, women with bilateral breast cancer, women with a family history of breast and ovarian cancer involving multiple family members (including males), and women of Ashkenazi Jewish ancestry [23]. Currently in the United States, Myriad Genetics ([www.myriad.com](http://www.myriad.com)) is the sole provider of BRCA<sup>Analysis</sup>, a direct sequencing approach to detect mutations in BRCA1 and BRCA2.

### Patient Management

Clinical management of BRCA mutation carriers may include surgical intervention, chemoprevention, or increased surveillance. Although prophylactic mastectomy is the most effective intervention for BRCA1 and BRCA2 carriers, reducing the risk of developing breast cancer by 90% [24], mastectomy is a highly invasive procedure with adverse

physical and potentially devastating psychological effects. Thus, other non-surgical methods have been developed for detection and risk reduction [25]. As tamoxifen in BRCA mutation carriers with breast cancer has been shown to significantly decrease the rate of contralateral disease [26], chemoprevention may be considered as part of an overall risk management program. In addition, magnetic resonance imaging (MRI) is more sensitive than mammography for detecting malignancy in BRCA carriers [27]. The International Consensus Conference on Breast Cancer Risk, Genetics, and Risk Management recently advocated use of both mammography and MRI at alternating six-month intervals for screening BRCA mutation carriers [28].

BRCA status appears to confer a unique tumor phenotype that may contribute to poorer outcomes in patients with breast cancer. Primary breast tumors from BRCA1 mutation carriers have unique pathology, including high histological grade, atypical medullary histotype, high rates of cellular proliferation, pushing margins and infiltrating lymphocytes, and are frequently ER, PR, and HER2 negative (triple negative) [29]. These clinical attributes contribute to the "basal-like" phenotype of BRCA1 breast carcinomas and may influence tumor behavior and aggressiveness. Some studies have observed lower survival in BRCA1 mutation carriers compared to non-carriers [30, 31], suggesting poorer outcomes in patients with BRCA1 mutations; however, other studies have found no evidence of significant differences in long-term survival [32, 33].

The availability of genetic testing for BRCA mutations is potentially valuable in surgical decision making for newly diagnosed breast cancer patients, as approximately one-half of patients who are discovered to carry BRCA1/2 mutations are likely to choose bilateral mastectomy compared to only one-fourth of non-carriers [34]. At present, BRCA mutation status is not considered when making recommendations for systemic therapy because response to such therapies has not been shown to differ between carriers and non-carriers [28]. However, alterations in DNA repair caused by defects in BRCA1 and BRCA2 may enhance the sensitivity of BRCA-positive tumors to platinum agents such as cisplatin and carboplatin, which cross-link DNA and interfere with replication [35]. Similarly, BRCA dysfunction sensitizes cells to poly(ADP-ribose) polymerase (PARP) inhibitors, leading to chromosomal instability, cell cycle arrest, and apoptosis in BRCA1 and BRCA2 deficient cells [36]. Clinical trials are currently underway to determine the efficacy of PARP inhibitors in treating BRCA deficient breast tumors, which may open new avenues for less toxic therapies that target particular DNA repair pathways in BRCA1 and BRCA2 mutation carriers [37].

### Genomic Variability Among BRCA Carriers

Identification of BRCA1 and BRCA2 has improved clinical management of some individuals with hereditary breast and ovarian cancer, but personalized care is not yet available for mutation carriers. Important barriers to effective treatment include allelic heterogeneity, environmental effects, and genetic modifiers. Allelic heterogeneity, where different mutations in the BRCA1 and BRCA2 genes show variability in penetrance, has been associated with different risks for developing disease [38, 39]. Although many women

with BRCA mutations have a high probability of developing breast cancer, 15% of BRCA1-positive women and 20% of BRCA2-positive women will never develop breast cancer [40]. Likewise, environmental factors such as exercise and body weight, environmental exposures [41], contraceptive and hormone use, and reproductive history may influence tumor development in BRCA carriers [42]. Modifier genes are believed to contribute to variability in cancer risk, but the identification of genetic modifiers has been difficult. To date, only a single variant (-135G>C) in the recombination protein A (RECA or RAD51) gene has been confirmed as a genetic modifier in BRCA2 carriers [43]. The androgen receptor (AR), amplified in breast cancer 1 (AIB1), and aurora kinase (AURKA) genes have been implicated as genetic modifiers in BRCA carriers, but remain only candidates as positive results have yet to be replicated [44].

To hasten the identification of genetic modifier genes, the Consortium of Investigators of Modifiers of BRCA1 and BRCA2 (CIMBA) was established in 2005 as a "consortium of consortia". With the collection of clinical data and genetic material from more than 15,000 BRCA1 and BRCA2 patients from around the world, genetic studies from CIMBA should have sufficient power to identify modifier genes [45]. Identification of genetic modifiers that influence risk of developing breast and/or ovarian cancer is needed to refine individual risk estimates and guide treatment options, including the need for prophylactic mastectomy and/or oophorectomy, in a personalized fashion.

Mutations in the BRCA1 and BRCA2 genes account for the majority of families with six or more cases of early-onset breast and/or ovarian cancer; however, many families with a high incidence of breast cancer have no detectable mutations

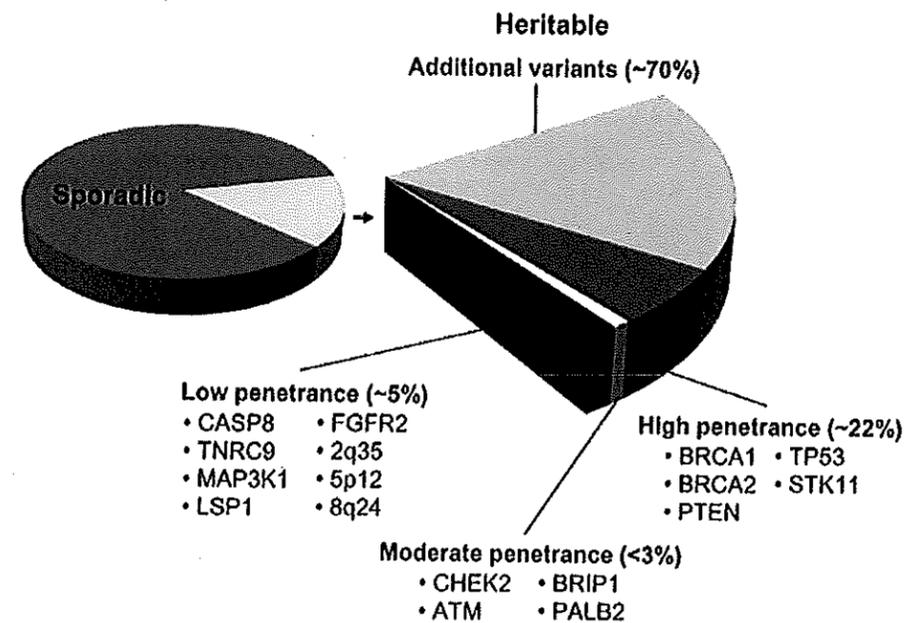
in BRCA1 or BRCA2 [46], suggesting the existence of additional breast cancer susceptibility genes. Although genes associated with increased susceptibility have been identified, the majority of causative breast cancer genes remain unknown [47] — see Fig. (1). Confounding factors such as (1) heterogeneous phenotypes attributable to several unidentified BRCA genes, each accounting for a small number of cancer cases; (2) genes influencing development of other types of cancer in addition to breast that may not be linked to breast cancer; and (3) weakly penetrant genes that are heritable but resemble sporadic breast cancer in appearance [48] make the identification of high-risk breast cancer genes difficult, but critical to providing personalized and effective care to patients with hereditary and/or familial breast cancer.

#### RISK ASSESSMENT FOR SPORADIC BREAST CANCER

##### Risk Estimation

Inherited mutations have been associated with 10-15% of all breast cancer cases; however, disease etiology in the majority of women appears to be sporadic, lacking a significant family history. Because sporadic breast cancer may be influenced by a number of lifestyle and environmental factors as well as common low-risk variants in a number of genes, several models have been developed in an attempt to quantify individualized breast cancer risk:

- The Gail model measures risk based on patient age, age at menarche, number of prior breast biopsies, age at first live birth, and number of first degree relatives affected by breast cancer [49].



**Fig. (1).** Complexity and largely unknown molecular etiology of breast cancer. The majority of breast cancer cases (~70%) are considered sporadic in nature because individuals with disease do not have an extensive family history of breast cancer and molecular alterations contributing to the disease have not been identified. Familial breast cancer (~30% of patients), often seen in families with a high incidence of breast cancer, has been associated with a number of high-, moderate-, and low-penetrance susceptibility genes.

- The Claus model estimates risk based on the number of affected relatives and their respective age at diagnosis [50].
- The BRCAPRO model calculates risk of developing breast cancer based on the probability of carrying a BRCA1 or BRCA2 mutation [51].

These models have been widely used to predict risk and direct patient care, but each model has limitations because no model accounts for the spectrum of risk factors influencing breast cancer. For example, the Gail model considers only first-degree relatives without regard to age at diagnosis or presence of ovarian cancer, thus potentially underestimating genetic risk. The Claus and BRCAPRO models only consider family history, potentially underestimating risk in women with other risk factors [52]. In addition, these models were developed 10-20 years ago when incidence of breast cancer in the general population was lower than it is today — use of lower baseline risk estimates may contribute to an underestimation of current risk [53]. More recent models, such as the Tyrer-Cuzick model, utilize family history, endogenous estrogen exposure, and presence of benign disease to model breast cancer risk [54], but contributions from other factors such as mammographic breast density, weight gain, steroid hormone levels, and susceptibility genes have not been incorporated [55].

The discovery of the BRCA1 and BRCA2 genes advanced risk assessment in families affected by hereditary breast and ovarian cancer, but identification of molecular markers associated with increased breast cancer risk in patients without a family history of breast cancer has remained far more challenging. Without a strong family history, linkage approaches involving large pedigrees such as those used to identify BRCA1 and BRCA2 are not applicable. Sporadic breast cancer is not usually associated with other cancers such as ovarian or male breast cancer and unlike BRCA1-positive carcinomas, which exhibit specific histological characteristics, sporadic breast cancer cases comprise a vast array of phenotypes. Early approaches to identify sporadic breast cancer susceptibility genes compared the frequency of DNA variants in genes from molecular pathways believed to be involved in breast cancer development between cases

with disease and healthy matched controls. An association study using candidate genes recently identified caspase-8 (CASP8) as a low-risk susceptibility gene where the major (H) allele of the D302H polymorphism had a protective effect on the development of breast cancer [56]. Despite success in identifying CASP8, candidate gene approaches have not been widely successful in identifying additional breast cancer susceptibility genes [57].

#### Whole-Genome Approaches

Candidate gene approaches are rapidly giving way to genome-wide association studies (GWAS), which evaluate a dense array of genetic markers representing common variation throughout the genome. Completion of the human genome sequence and subsequent identification of single nucleotide polymorphisms (SNPs) now permits millions of informative SNPs across the genome to be assayed simultaneously. GWAS are useful for mapping genes of interest to small, localized regions of the genome and for detecting the effects of common (>5% minor allele frequency) alleles on disease risk [58]. Moreover, GWAS are performed without *a priori* knowledge of the underlying genetic defect(s), which may be advantageous since many genes identified through whole genome approaches were not previously suspected to influence the disease under investigation [59].

Recent GWAS have identified a number of loci that appear to be associated with breast cancer susceptibility (Table 1). For example, the fibroblast growth factor receptor 2 (FGFR2), mitogen-activated kinase kinase kinase 1 (MAP3K1), lymphocyte-specific protein (LSP1), and trinucleotide repeat-containing 9 (TNRC9/LOC643714) genes, along with a 110 kb region of chromosome 8q24 have been associated with breast cancer in large studies involving thousands of subjects [60, 61]. Associations with other chromosomal regions — 2q35, 5p12, 6q22, and 16q12 — also have been reported [62-64]. Further analysis has shown that allelic variation at FGFR2, TNRC9, 8q24, 2q35, and 5p12 is associated with physiological characteristics of breast tumors, such as ER status [62, 64, 65], and specific FGFR2, MAP3K1, and TNRC9 variants may interact with

Table 1. Low-Penetrance Variants that may Influence Sporadic Breast Cancer Identified through Genome-Wide Association Studies

SNP	Chromosome	Candidate Genes	MAF <sup>a</sup>	OR <sup>b</sup>	Reference
rs889312	5q11	MAP3K1	0.28	1.13	[60]
rs2180341	6q22	ECHDC1, RNF146	0.27	1.41	[63]
rs2981582	10q26	FGFR2	0.38	1.26	[60, 61]
rs3803662	16q12	TNRC9, LOC643714	0.25	1.20	[60, 62]
rs3817198	11p15	LSP1	0.30	1.07	[60]
rs10941679	5p12	MRPS30	0.25	1.19	[64]
rs13281615	8q24		0.40	1.08	[60]
rs13387042	2q35		0.50	1.21	[62]

<sup>a</sup>MAF = minor allele frequency.  
<sup>b</sup>OR = odds ratio per allele.

BRCA1 and BRCA2 mutations to increase breast cancer risk [66].

Despite recent success in identifying genetic determinants of breast cancer, susceptibility alleles identified through GWAS are believed to account for only ~5% of breast cancer risk [67]. If future studies are to be successful in identifying additional low-risk susceptibility alleles and low-frequency, highly-penetrant variants [68], interactions between genes and environmental exposures must be assessed [69] and methods must be developed to evaluate mechanisms by which DNA variants in intronic or intergenic regions contribute to disease. As risk associated with susceptibility alleles may vary between racial/ethnic populations due to differences in frequency, patterns of disequilibrium, and interactions with environmental factors [60, 62, 70], sufficiently powered genetic studies in women from various ethnic groups are needed to improve risk reduction strategies for all women.

#### Direct-to-Consumer Testing

New susceptibility variants identified by GWAS have not yet been incorporated into genetic tests with beneficial clinical utility for breast cancer patients. However, genetic analysis and risk assessment are available commercially through direct-to-consumer (DTC) testing. A number of for-profit companies offer personal genetic information based on DTC tests — the largest and most recognized companies include 23andMe ([www.23andme.com](http://www.23andme.com)), deCODEme ([www.decodeme.com](http://www.decodeme.com)), Navigenics® ([www.navigenics.com](http://www.navigenics.com)), and Knome®, Inc. ([www.knome.com/home](http://www.knome.com/home)) (Table 2). For a fee of \$99 to \$99,500 consumers provide a blood, buccal, or saliva sample for targeted SNP analysis or whole-genome sequencing. Genetic information provided to the consumer varies greatly among companies, from trivial facts such as ear wax type to ancestry information to information on risk for disease [71, 72]. Although DTC tests epitomize “personalized genomics” by providing consumers with individual genotypes, critics note that the clinical utility of such tests is limited and often incongruent with marketing claims. Because information on family history and environmental exposures is usually not accounted for, DTC risk estimates may not be sufficiently accurate to enable consumers to make appropriate medical decisions [73, 74].

The majority of genetic risk assessments developed thus far focus on DNA variants; however, a new RNA-based sig-

nature has been developed for non-invasive breast cancer screening using peripheral blood samples. Although based on a small number of cases (n=24) and controls (n=32), a subset of 37 genes in the assay correctly classified 82% of patients [75]. Despite a relatively high misclassification rate, DiaGenic ([www.diagenic.no](http://www.diagenic.no)) has since developed this gene expression signature into a clinical screening tool, currently available only in India as BCtect™ India.

#### MOLECULAR CHARACTERIZATION OF BREAST TUMORS – PERSONALIZED PROGNOSTICS

##### Pathological Characterization

Human breast carcinomas exhibit diverse pathological characteristics that are associated with different clinical outcomes, and thus are routinely used to guide treatment options. Accordingly, an accurate definition of prognosis is dependent on the ability to detect and quantify differences in tumor attributes, such as rates of proliferation and propensity to metastasize. Routine tumor evaluation currently includes: (1) histopathological classification; (2) grade determination; and (3) quantification of tumor size, surgical margin status, and lymph node involvement. Histopathological characterization, based on microscopic cellular morphology, classifies breast carcinomas into common subtypes (ductal or lobular carcinoma), which tend to have similar prognoses [76], or less common forms such as mucinous, tubular, and papillary (favorable prognosis) [77] or inflammatory breast cancer (poor prognosis) [78]. Increasing tumor size has long been associated with poor prognosis [79], but improved mammographic detection of smaller tumors has decreased the prognostic utility of tumor size [80]. Presence of positive surgical margins has been associated with local recurrence, but only 27% of patients with extensively positive margins will have recurrent disease [81, 82]. Likewise, the Nottingham Histological Score, widely used for assessing histological grade, is clinically useful for stratifying patients into low risk (low-grade disease, 95% five-year survival) and high risk (high-grade disease, 50% five-year survival) groups [83, 84], but the reliability of breast tumor grade in predicting survival is hampered by subjectivity associated with its assessment [85]. Axillary lymph node status is the most reliable predictor of survival, differentiating women who are likely to have >90% five-year survival (patients with negative nodes) from those who are likely to have <70% survival (women with nodal metastasis) [86]. Although these clinical attributes are cur-

Table 2. Leading Direct-to-Consumer Genetic Testing Companies

Company	Headquarters	Website	Cost (USD)	Genetic Counseling	Breast Cancer Susceptibility Variants
23andMe	Mountain View, CA	<a href="http://www.23andme.com">www.23andme.com</a>	\$399	No	2 SNPS
deCODEme	Reykjavik, Iceland	<a href="http://www.decodeme.com">www.decodeme.com</a>	\$985 <sup>a</sup>	Yes	11 variants <sup>b</sup>
Knome	Cambridge, MA	<a href="http://www.knome.com">www.knome.com</a>	Custom <sup>c</sup>	Yes	DNA sequence
Navigenics	Foster City, CA	<a href="http://www.navigenics.com">www.navigenics.com</a>	\$999 <sup>d</sup>	Yes	unknown

<sup>a</sup>Complete scan.

<sup>b</sup>For women of European descent.

<sup>c</sup>KnomeSELECT™ is \$24,500 for complete sequence of 20,000 genes; KnomeCOMPLETE™ is \$99,500 for complete genome sequence.

<sup>d</sup>Option for ongoing subscription (\$199 per year) for updates.

rently the standard of care for breast cancer patients, many are imprecise in their ability to accurately predict outcomes.

### Immunohistochemistry

Molecular markers have the potential to provide additional prognostic information to supplement traditional pathological assessments for disease management in breast cancer patients. As mentioned above, traditional immunohistochemistry (IHC) markers routinely used in the classification of breast cancer include ER, PR, and HER2. Tumors positive for ER and PR expression frequently have low cellular proliferation rates, tend to exhibit lower histological grade, and are associated with more favorable prognosis [87]. ER and PR expression also is useful for identifying patients who will likely benefit from hormonal therapy, as women with ER and PR negative breast cancer do not gain a survival benefit from anti-estrogen tamoxifen [88].

The HER2 gene is a member of the epidermal growth factor receptor family with tyrosine kinase activity and is amplified at the DNA level and/or over-expressed in 15-25% of breast cancers. Carcinomas with amplified/over-expressed HER2 exhibit high histological grade and usually have a poor prognosis [89, 90]. Some patients with positive HER2 status (15-20%) are eligible to receive trastuzumab, a monoclonal antibody targeting HER2, in combination with standard chemotherapy [91].

Rigorous clinical studies have shown that evaluating ER, PR, and HER2 status provides additional prognostic information beyond that normally achieved by histological assessment alone. For example, breast carcinomas that are ER-negative, PR-negative, and do not have HER2 over expressed (triple negative) are marked by aggressive behavior, but because women with triple-negative disease are not eligible for tamoxifen or trastuzumab treatment, they usually have relatively low long-term survival [92]. Other markers such as nuclear antigen Ki67 are not routinely used to guide treatment selection, but hold great promise for monitoring the effectiveness of neoadjuvant chemotherapy and predicting recurrence-free survival [93-95].

Individual estimates of outcome using clinical and pathological characteristics of breast tumors, including age,

menopausal status, co-morbid conditions, tumor size, number of positive lymph nodes, and ER status have been incorporated into a computer program, Adjuvant! Online ([www.adjuvantonline.com/index.jsp](http://www.adjuvantonline.com/index.jsp)), which is available over the Internet as a decision aid for patients and their physicians [96]. The program estimates the efficacy of endocrine therapy and chemotherapy as well as overall and disease-free survival in a user-friendly format that effectively brings patients into the decision-making process regarding personalized treatments.

Although IHC analysis of ER, PR, and HER2 is widely used in the pathological evaluation of breast tumors, additional molecular signatures involving multiple genes and/or proteins are desperately needed to more accurately classify tumors and guide treatment selection. Recently, a multi-gene IHC-based test known as MammoStrat<sup>®</sup> (Applied Genomics, Huntsville, AL; [www.applied-genomics.com/mammostrat.html](http://www.applied-genomics.com/mammostrat.html)) was developed to classify breast cancer patients into low-, moderate-, or high-risk categories for disease recurrence [97] (Table 3). MammoStrat<sup>®</sup> uses conventional paraffin-embedded tissue to assay five markers by IHC: tumor protein p53 (TP53) — known to play a central role in cell cycle regulation; *Hpa* II tiny fragments locus 9C (HTF9C) — involved in DNA replication and cell cycle control; carcinoembryonic antigen-related cell adhesion molecule 5 (CEACAM5) — aberrantly expressed in some cancers; *nmyc* downstream-regulated gene 1 (NDRG1) — may function as a signaling protein in growth arrest and cellular differentiation; and solute carrier family 7 (cationic amino acid transporter, y<sup>+</sup> system), member 5 (SLC7A5) — mediates amino acid transport. The MammoStrat<sup>®</sup> test may have utility for predicting patient outcomes, but currently requires five separate slides (one slide per antibody), which has the potential to show variability in staining intensity and scoring between patients.

### Gene Expression Signatures and Tumor Classification

With the sequencing of the human genome and identification of many human genes, whole-genome approaches using array-based methods have been developed to evaluate expression levels for thousands of genes in a single experi-

Table 3. Selected Molecular Diagnostic Tests for Breast Cancer

Test	Company	Assay Type <sup>a</sup>	Number of Genes/Proteins	Classification	Reference
Breast Bioclassifier™	University Genomics	qRT-PCR	55	Tumor subtype Therapeutic guidance	[9]
MammaPrint™	Agendia	Microarray	70	Prognostic Therapeutic guidance	[105, 106]
MammoStrat <sup>®</sup>	Applied Genomics	IHC	5	Prognostic	[97]
MapQuant DX™	Ipsogen	Microarray	97	Tumor grade	[103, 104]
Oncotype DX™	Genomic Health	qRT-PCR	21	Prognostic Therapeutic guidance	[108, 109]
Rotterdam signature	Veridex	Microarray	76	Prognostic	[107]

<sup>a</sup>qRT-PCR = quantitative real-time PCR; IHC = immunohistochemistry.

ment. Initial studies of human breast carcinomas identified several distinct subtypes of breast cancer using large-scale gene expression profiling: (1) ER positive (luminal-like), characterized by high expression of many genes expressed in breast luminal cells; (2) basal-like; (3) HER2 positive; and (4) normal-like [9]. Further experiments subdivided the luminal (ER positive) tumors into luminal A and luminal B subtypes, and found that these five subtypes are useful in predicting relapse-free and overall survival, with the HER2 positive and basal-like subtypes having the shortest survival times [10].

To improve the clinical utility of molecular signatures for predicting outcomes in women with breast cancer, a new gene signature encompassing a larger number of genes has been identified. The Single Sample Predictor (SSP) signature is reported to identify the original five intrinsic subtypes, plus a new "IFN-regulated" subtype, characterized by high expression of interferon-regulated genes [98]. Likewise, the Breast Bioclassifier™ (University Genomics, Saint Louis, MO) is a commercial assay for classifying breast carcinomas by subtype. Expression levels for 55 genes provide subtype information and a continuous risk score to guide treatment options ([www.bioclassifier.com/](http://www.bioclassifier.com/)).

As gene expression signatures for classifying breast tumors likely reflect underlying biological characteristics of disease, a number of gene expression signatures focusing on specific molecular pathways have been developed. For example, tumorigenesis and wound healing share many similar physiological and molecular processes, including recruitment of inflammatory cells, stimulation of fibroblast and epithelial cell proliferation, cell migration, and angiogenesis. A "wound response" gene expression signature has been associated with metastatic progression and mortality in breast cancer patients [99], and may improve risk stratification independently of established clinicopathological risk factors [100]. If the wound response signature proves to be effective in identifying patients at high-risk of recurrence after breast conserving therapy, it may be useful as a diagnostic tool to identify patients who should be offered more aggressive treatments, such as increased radiation, or who should consider mastectomy rather than breast conserving therapy [101]. Likewise, breast carcinomas showing high levels of expression for hypoxia-related genes tend to exhibit p53 mutations, negative ER status, and high histological grade, and have been associated with lower overall and disease-free survival [102].

Gene expression profiles are increasingly being used as molecular tools to complement pathological evaluation and guide treatment options. To overcome inherent subjectivity in histological grading of breast tumors, the Gene expression Grade Index (GGI) was developed to summarize the similarity between gene expression profiles and tumor grade. The score attempts to classify low-grade and high-grade tumors, and to subdivide intermediate-grade tumors into low-grade, high-grade, or mixed-grade groups [103]. Further refinement of the gene expression grade index led to the MapQuant DX™ Genomic Grade assay (Ipsogen, Marseille, France; [www.ipsogen.com/](http://www.ipsogen.com/)), marketed as the first microarray-based diagnostic test to measure tumor grade [104]. With the reported ability to classify ~80% of intermediate-grade breast

tumors as either low-grade or high-grade, the MapQuant DX™ Genomic Grade assay may be useful in guiding treatment options, possibly sparing patients with low-grade (grade 1 or grade 1-like) tumors unnecessary treatments, while identifying those patients who would most likely benefit from chemotherapy.

#### Gene Expression Signatures and Disease Risk

Molecular profiles are now being used more frequently as clinical tools to determine treatment for certain groups of patients by categorizing them into low-risk and high-risk groups. The MammaPrint™ assay (Agendia, Amsterdam, The Netherlands; [www.agendia.com](http://www.agendia.com)) is a 70-gene signature developed using tumor tissue from young women (<55 years of age) with node-negative disease, who either developed distant metastasis or remained disease-free after five-years [105]. Overall 10-year survival for the "poor-prognosis" signature is ~55%, while 10-year survival in women with the "good-prognosis" signature is 95%. The probability of being free from distant metastasis after 10 years is 51% for the poor prognosis and 85% for the good prognosis profile [106]. A second group of researchers subsequently developed a 76-gene profile (Rotterdam signature) that could identify breast cancer patients at high risk for distant recurrence. The signature could identify patients who developed distant metastases within five years when traditional prognostic factors were considered (hazard ratio 5.55, 95% CI 2.46–12.5) and could predict metastasis in both premenopausal and postmenopausal patients [107].

The gene expression signatures outlined above were refined from global expression profiling experiments involving thousands of genes and flash-frozen tumor specimens. An alternative approach relied on an extensive literature search to identify candidate genes (n=250) believed to be involved in disease development based on known function. Gene expression levels were assayed in 447 patients with ER-positive, node-negative breast cancer to identify a small subset of 16 genes (plus five reference genes) amenable to analysis by real-time-PCR (RT-PCR) on RNA isolated from formalin-fixed, paraffin-embedded (FFPE) specimens. The resulting 21-gene signature, known as Oncotype DX® (Genomic Health, Redwood, CA; [www.genomichealth.com/](http://www.genomichealth.com/)) provides a probability of recurrence score for women with early stage (Stage I or II), ER-positive, node-negative breast cancer, and categorizes patients as low-, intermediate-, or high-risk. In validation studies using patients from the National Surgical Adjuvant Breast and Bowel Project (NSABP) clinical trial B-14 who received tamoxifen, the probability of distant recurrence at 10 years for the three risk categories was: low-risk — 6.8% (95% CI 4.0-9.6); intermediate-risk — 14.3% (95% CI 8.3-20.3); and high-risk — 30.5% (95% CI 23.6-37.4). Recurrence scores also correlated significantly with relapse-free interval and overall survival [108]. In a subsequent study, Oncotype DX™ was used to assess the benefit of adjuvant chemotherapy in ER-positive, node-negative patients. Because the highest benefit was observed in patients with high-risk scores, while women with low-risk recurrence scores did not benefit from chemotherapy [109], Oncotype DX™ may be useful in guiding treatment options in ER-positive, node-negative patients.

Clinical trials of the MammaPrint™ and Oncotype DX™ assays are currently in progress. In the Microarray In Node-negative Disease may Avoid ChemoTherapy (MINDACT) trial, 6,000 node-negative women will be assigned to treatment groups based on risk stratification by traditional clinical-pathological factors (ADJUVANT! Online) and the MammaPrint™ molecular signature [110]. Patients classified as low risk by both methods will not receive chemotherapy, while those considered high risk for relapse by both methods will be given the opportunity to receive adjuvant chemotherapy. Patients of primary interest, those with discordant results, will be randomized to treatment based on either Adjuvant! Online or MammaPrint™ to determine which test is more effective in defining treatment in node-negative patients. The Trial Assigning Individualized Options for Treatment (TAILORx) is examining whether hormone receptor-positive patients with an intermediate Oncotype DX™ risk recurrence score benefit from chemotherapy. The trial is recruiting 10,000 hormone-receptor-positive patients with HER2-negative and lymph-node-negative disease. Treatment will be based on the risk recurrence score as follows:  $\leq 10$  — hormone therapy alone;  $\geq 26$  — hormone and chemotherapy; intermediate scores — randomization to either hormone therapy alone or to hormone therapy and chemotherapy. The goal is to integrate Oncotype DX™ into the clinical decision-making process and refine the utility of the assay in clinical practice [111].

Molecular signatures have improved the ability to predict outcome and identify breast cancer patients who would most likely benefit from systemic therapy, thus providing an additional layer of personalized medicine. However, no current molecular signature is 100% accurate, and 5-10% of patients now classified as low-risk are likely to relapse. Furthermore,

current classification systems were developed to predict only short-term (<5 years) outcomes; thus there is a need to develop signatures that identify patients with protracted disease progression who may benefit from prolonged therapy [112]. Although outcome prediction tends to be similar between gene-expression signatures, overlap among genes comprising the signatures is relatively low, suggesting that these profiles assess common biological pathways, but have not identified the actual genes driving tumor behavior and outcome [113]. Finally, some multigene predictor assays are being adopted and marketed before they have been properly validated and proven to be clinically informative, thus the degree to which expression-based tests will alter the course of patient treatment remains unclear [114, 115].

#### PHARMACOGENOMICS OF BREAST CANCER

Pharmacogenomics in breast cancer evaluates the effect of inherited genomic variation on patient response or resistance to treatment. Genetic variability is commonly measured at the DNA level in the form of chromosomal alterations or DNA sequence variants (Table 4). Conversely, somatic genomic changes (DNA variants and gene expression profiles) in breast tumors can influence rates of apoptosis, cell proliferation, and DNA damage repair, which may have direct effects on response to treatment and survival. To be most effective, personalized medicine must incorporate information from innate genetic variation as well as somatic mutations in diseased tissue [116].

#### Endocrine Therapy

Estrogens play an important role in the etiology of breast cancer by stimulating growth and proliferation of ductal

Table 4. Selected Genetic Polymorphisms Affecting Response to Therapy in Breast Cancer Patients

Treatment	Gene	Variant	Functional Change	Response to Treatment	Reference
Chemotherapy					
Doxorubicin	CBR3	11G>A	Decreased enzyme activity	Hematological toxicity	[142]
Anthracyclines	MnSOD	Ala <sup>16</sup>	Higher levels of reactive oxygen species	Decreased mortality	[143]
	MPO	—463GG	Higher levels of reactive oxygen species	Decreased mortality	[143]
	GSTP1	313A>G	Altered drug transport	Hematological toxicity	[150]
	MTHFR	1298A>C	Altered drug metabolism	Non-hematological toxicity	[150]
Endocrine therapy					
Tamoxifen	CYP2D6	*3, *4, *5, *10, *41	Reduced function/nonfunctional enzyme	Poor clinical outcome	[120, 121]
Aromatase inhibitors	CYP19A1	Cys <sup>264</sup> , Thr <sup>364</sup>	Decreased enzyme activity	Reduced benefit	[128]
Radiotherapy					
	TP53	Arg72Pro, PIN3	Decreased apoptosis	Risk of telangiectasia	[149]
Targeted therapy					
Trastuzumab	HER2	Heterodimer	Prevents disruption by trastuzumab	Poor response to treatment	[132]

epithelial cells in the breast, thus the status of the estrogen receptor in breast carcinomas provided one of the earliest avenues for personalized medicine. Fortunately, hormone-receptor-positive tumors usually are responsive to agents such as Tamoxifen that block the function of estrogen. Tamoxifen is a potent antagonist of the ER with inhibitory effects on tumor growth that has become the gold standard for endocrine treatment of estrogen-receptor-positive breast cancer in premenopausal and postmenopausal women [117]. Tamoxifen is associated with side effects such as blood clots, stroke, and increased risk of endometrial and uterine cancer, but five-year use of tamoxifen has been shown to reduce risk of cancer recurrence by ~50% [118]. For most patients, the benefit of using tamoxifen for hormone-receptor-positive disease outweighs the risk of serious side effects; however, a small subgroup of hormone-receptor-positive patients who carry specific variants in the cytochrome P450 2D6 (CYP2D6) gene do not benefit from tamoxifen. The CYP2D6 gene is a key enzyme in the metabolism of tamoxifen to its active metabolite endoxifen. Several DNA variants in CYP2D6 result in poor metabolism of tamoxifen and lower levels of endoxifen [119]. Patients who carry reduced-function or nonfunctional CYP2D6 alleles have been found to derive inferior therapeutic benefit from tamoxifen and thus are at increased risk of breast cancer recurrence [120] or have significantly shorter disease-free survival than non-carriers [121]. Studies are underway to determine the utility of CYP2D6 genotyping for making clinical decisions about tamoxifen and the potential to optimize breast cancer therapy [122, 123].

Alternate forms of directed anti-estrogen therapies do exist for patients with hormone-receptor-positive breast cancer including aromatase inhibitors that block the production of estrogen, and compounds such as fulvestrant (Faslodex<sup>®</sup>) that down-regulate and degrade the ER protein. Aromatase inhibitors such as anastrozole (Arimidex<sup>®</sup>), letrozole (Femara<sup>®</sup>), and exemestane (Aromasin<sup>®</sup>) target cytochrome P450 19 (CYP19A1 or aromatase), an enzyme involved in estrogen synthesis in peripheral organs. Premenopausal women with functional ovaries do not receive aromatase inhibitor therapy because first and second generation aromatase inhibitors did not effectively suppress estrogen levels and because decreased estrogen levels in peripheral tissues could be counteracted by increased estrogen synthesis in the ovaries [124]. In postmenopausal women, aromatase inhibitors are well-tolerated and improve both disease-free and recurrence-free survival [125-127]. Similar to CYP2D6, the Cys<sup>264</sup> and Thr<sup>364</sup> variants in aromatase are associated with decreased activity and lower levels of immunoreactive protein, which may contribute to variation among patients in response to aromatase inhibitor therapy [128]. Although directed endocrine therapies provide treatments specific for patients with hormone-receptor-positive breast cancer, factors such as menopausal status and innate genetic variability may alter the effectiveness of treatment.

#### Treatment for HER2 Positive Breast Cancer

Therapies directed at the HER2 protein provide a second avenue of targeted treatment for some patients with breast cancer. Trastuzumab (Herceptin<sup>®</sup>, Genentech, South San

Francisco, CA; www.gene.com/) is a humanized monoclonal antibody that binds to the extracellular domain of the HER2 protein, blocking tumor cell growth. Trastuzumab is the current standard of care in adjuvant therapy for HER2-positive breast cancer, effective as a single agent or in combination with chemotherapeutics for the 20-25% of patients with HER2-positive cancer [129]. However, many patients with HER2-positive disease do not derive tangible benefit from trastuzumab. Given that the cost per patient for trastuzumab ranges from \$20,000-\$80,000 per year with the potential for significant adverse side effects [130], a more precise classification of HER2-positive patients who will derive benefit from trastuzumab and improved understanding of how amplification and/or over-expression of HER2 contribute to aggressive tumor biology are critical to improving patient treatment.

The major oncogenic unit in HER2-positive breast cancer appears to be a heterodimer between the HER2 and epidermal growth factor receptor-3 (HER3) proteins, where HER3 functions as a necessary dimerization partner for HER2 to achieve full oncogenic signaling potential [131]. Recent studies have shown that HER2/HER3 heterodimers promote cellular proliferation in both *in vitro* and *in vivo* models, suggesting that HER3 may be an important therapeutic target in HER2-positive patients [132]. Pertuzumab has been shown to bind to the dimerization arm of HER2, blocking HER2/HER3 heterodimerization and attenuating growth of solid tumors in model systems [133]. Thus, combining pertuzumab with trastuzumab may augment therapeutic benefit by blocking HER2/HER3 signaling. Monogram Biosciences (South San Francisco, CA; www.monogrambio.com/) has developed the commercially-available HERmark<sup>™</sup> test to measure total HER2 levels and HER2 homodimers in FFPE tissue and is developing a *VeraTag*<sup>™</sup> assay to quantify levels of HER2/HER3 heterodimers. These assays may allow patients with HER2-positive breast cancer to receive the most efficacious combination of new drugs targeting HER2.

#### Chemotherapeutics

Chemotherapy involves use of chemical agents as part of a systemic treatment targeting proliferative cancer cells. Adjuvant chemotherapy is used to reduce risk of recurrence after primary therapy in women with localized breast cancer and to provide palliative care in patients with advanced (metastatic) disease. In contrast, neoadjuvant chemotherapy is normally used to shrink moderate- to large-sized breast carcinomas prior to surgical resection, which permits use of less aggressive surgical options, including breast conservation, and may be useful in guiding longer-term treatment based on tumor response to specific drug combinations [134]. Obviously, the ability to predict which patients will benefit from adjuvant therapy and identify who will respond favorably to neoadjuvant regimens would provide an additional level of personalized care.

#### Gene Expression and Chemotherapeutic Agents

Gene expression profiling has been used to study the biological responses of human breast carcinomas to optimize chemotherapeutic treatments. Cell lines derived from luminal and basal epithelium have been observed to respond differently to agents commonly used in chemotherapy, such as

doxorubicin (DOX) and 5-fluorouracil (5FU). In culture, luminal cell lines show low levels of expression for genes regulating cellular proliferation and the cell cycle, while basal cell lines tend to repress genes involved in cellular differentiation when exposed to DOX and 5FU [135]. Similarly, different molecular subtypes of breast cancer defined by gene expression profiling respond differently to preoperative chemotherapy, with basal-like and HER2-positive subtypes being more sensitive to paclitaxel and doxorubicin than luminal and normal-like cancers [136]. Expression signatures also have been used to predict clinical response of breast cancer patients receiving either cyclophosphamide-adriamycin or epirubicin-5FU as part of their adjuvant chemotherapy regimen [137] and to distinguish primary breast tumors that are responsive or resistant to docetaxel chemotherapy [138]. These observations further highlight the vast amount of molecular variability among breast carcinomas and emphasize the need for additional molecular signatures to more effectively guide treatment.

#### DNA Variation and Chemotherapeutic Agents

Clinical responses in breast cancer patients to commonly used chemotherapeutic agents vary considerably, from optimum therapeutic response to partial (beneficial) response to severe adverse events. Variation at the DNA level in an increasing number of genes is now known to affect the pharmacokinetics and pharmacodynamics of many chemotherapeutic drugs [139, 140], thus influencing toxicity and patient response. To improve the safety and efficacy of current treatments, therapies could be tailored to individual patients based on their genetic makeup [141]. For example, the carbonyl reductase 3 (CBR3) gene contributes to the reduction of DOX to doxorubicinol, a less potent metabolite, and the extent of metabolism is believed to be a source of variability in doxorubicin chemotherapy. The 11G>A variant (rs8133052) in CBR3 has been shown to influence tumor tissue expression of CBR3 and is associated with inter-individual variability in clinical outcomes. Women with the 11GG genotype experience greater leukocyte toxicity and are less likely to show a reduction in tumor size than women carrying 11AA [142].

A number of chemotherapeutics generate reactive oxygen species that function by damaging DNA and triggering the apoptotic cascade. Women carrying variants in genes associated with oxidative stress, such as manganese superoxide dismutase (MnSOD), catalase (CAT), and myeloperoxidase (MPO) that result in higher levels of reactive oxygen species, tend to have better overall survival than women with genotypes associated with lower levels of reactive oxygen species when treated with chemotherapy [143]. Due to the large number of drug-metabolizing enzymes and drug transporters containing polymorphisms that affect chemotherapy-related toxicity and treatment outcomes in breast cancer patients, improved pharmacogenetic information is needed to identify individuals at risk for toxicity and poor response.

#### Genomics in Clinical Practice

Recent developments in the clinical arena are indicative of the emerging importance of personal genomics in the prevention, surveillance, and treatment of breast cancer. Professional organizations such as the American Society of Clinical

Oncology (ASCO) have issued recommendations on the use of molecular markers for guiding therapy and determining prognosis in breast cancer patients [144]:

- CA 15-3 and CA 27.29 (assays to detect circulating MUC-1 antigen in peripheral blood) — contributes to decisions regarding therapy for metastatic breast cancer in conjunction with diagnostic imaging, history, and physical examination
- Carcinoembryonic antigen (CEA) — contributes to decisions regarding therapy for metastatic breast cancer in conjunction with diagnostic imaging, history, and physical examination
- ER/PR — should be measured on every primary invasive breast cancer to identify patients most likely to benefit from endocrine therapy
- HER2 — should be measured on every primary invasive breast cancer at diagnosis or recurrence to guide trastuzumab therapy
- Urokinase plasminogen activator (uPA) and plasminogen activator inhibitor 1 (PAI-1) — measured by ELISA on fresh or frozen tissue for determining prognosis in newly-diagnosed, node-negative breast cancer patients
- *Oncotype DX*<sup>®</sup> — in newly-diagnosed patients with node-negative, ER-positive breast cancer, can be used to predict risk of recurrence in women treated with tamoxifen

Large cancer centers such as Massachusetts General Hospital (MGH) and Memorial Sloan-Kettering Cancer Center (MSKCC) are now embracing the importance of genomics in clinical practice, recently implementing policies to routinely assay a number of breast cancer-related genes — v-akt murine thymoma viral oncogene homolog 1 (AKT1) and HER2 at MSKCC, phosphatase and tensin homolog (PTEN) and TP53 at MGH, and phosphatidylinositol 3-kinase, catalytic, alpha (PIK3CA) at both institutions [145]. As genomic medicine becomes an integrated part of health care delivery, use of personalized genomics in the clinical treatment of breast cancer will increase.

#### CONCLUSIONS

The era of personalized molecular medicine for breast cancer is on the horizon. Identification of strongly penetrant genes such as BRCA1 and BRCA2 has improved risk assessment for women with hereditary forms of breast cancer, but sporadic breast cancer presents additional challenges due to the influence on disease risk of lifestyle and environmental factors as well as common low-risk DNA variants. Substantial progress has been made in applying genomic discoveries to breast cancer treatment, as gene expression profiles are now being used to partition heterogeneous breast carcinomas into specific groups associated with different prognosis, pathological features, and developmental behavior. Customized treatments based on genetic susceptibility of the patient and molecular characteristics of the tumor allow more effective treatments that minimize adverse drug reactions. Yet despite these advances, personalized genomics in breast cancer is still in its infancy and genomic technologies

have just begun to realize their full potential in clinical practice. Currently, the majority of genes contributing to sporadic breast cancer have not been identified. Whole-genome association studies have identified some DNA variants contributing to breast cancer risk that provide new insights into the pathophysiology of disease and may ultimately prove useful for developing targeted interventions [146], but many variants will never have clinical utility. Although molecular information may estimate risk and guide treatment in certain populations, expression profiles may not be applicable to other high-risk groups. To maximize the effectiveness of genomic data, other factors that influence breast cancer risk, such as the impact of environmental exposures, lifestyle factors, and personal behaviors must become integral components of risk prediction models.

Recently proposed health-care reform legislation does not advocate personalized genomic medicine directly, but has the potential to profoundly affect genomics in medicine by adopting a shared decision making model that would incorporate patient preferences and values into their medical treatment plan. As educated patients become more knowledgeable about personalized genomics, demand for molecular-based tests may increase dramatically. Expanded use of genetic information for tailoring treatments in breast cancer patients will present several challenges, such as managing the impact of genetic testing on healthcare delivery and cost [147]. In order to maximize the potential of personal genomics in medicine, physicians must be fully prepared to deal with issues related to genomic tests, including the ability to critically evaluate and interpret genomic results [148] and issues such as cost-effectiveness, predictive limitations, and impact on quality of life must be considered.

#### ACKNOWLEDGEMENTS

This research was supported by the United States Department of Defense (Military Molecular Medicine Initiative MDA W81XWH-05-2-0075, protocol #01-20006) and was performed under the auspices of the Clinical Breast Care Project, a joint effort of many investigators and staff members whose contributions are gratefully acknowledged. The opinion and assertions contained herein are the private views of the authors and are not to be construed as official or as representing the views of the Department of the Army or the Department of Defense.

#### REFERENCES

- [1] Jemal, A.; Siegel, R.; Ward, E.; Hao, Y.; Xu, J.; Thun, M.J. Cancer statistics, 2009. *CA Cancer J. Clin.*, 2009, 59, 225-249.
- [2] *Breast Cancer Facts and Figures: 2007-2008*. American Cancer Society: Atlanta, GA, 2008.
- [3] *Cancer Trends Progress Report – 2007 Update*. National Cancer Institute, National Institutes of Health, Department of Health and Human Services: Bethesda, MD, 2007, <http://progressreport.cancer.gov>.
- [4] Key, T.; Appleby, P.; Barnes, I.; Reeves, G.; Enogenous Hormones and Breast Cancer Collaborative Group. Endogenous sex hormones and breast cancer in postmenopausal women: reanalysis of nine prospective studies. *J. Natl. Cancer Inst.*, 2002, 94, 606-616.
- [5] Kaaks, R.; Berrino, F.; Key, T.; Rinaldi, S.; Dossus, L.; Biessy, C.; Secreto, G.; Amiano, P.; Bingham, S.; Boeing, H.; Bueno de Mesquita, H.B.; Chang-Claude, J.; Clavel-Chapelon, F.; Fournier, A.; van Gils, C.H.; Gonzalez, C.A.; Gurrea, A.B.; Critselis, E.; Khaw, K.T.; Krogh, V.; Lahmann, P.H.; Nagel, G.; Olsen, A.; Onland-Moret, N.C.; Overvad, K.; Palli, D.; Panico, S.; Peeters, P.; Quirós, J.R.; Roddam, A.; Thiebaut, A.; Tjonneland, A.; Chirlaque, M.D.; Trichopoulou, A.; Trichopoulos, D.; Tumino, R.; Vineis, P.; Norat, T.; Ferrari, P.; Slimani, N.; Riboli, E. Serum sex steroids in premenopausal women and breast cancer risk within the European Prospective Investigation into Cancer and Nutrition (EPIC). *J. Natl. Cancer Inst.*, 2005, 97, 755-765.
- [6] Kaaks, R.; Rinaldi, S.; Key, T.J.; Berrino, F.; Peeters, P.H.M.; Biessy, C.; Dossus, L.; Lukanova, A.; Bingham, S.; Khaw, K.-T.; Allen, N.E.; Bueno-de-Mesquita, H.B.; van Gils, C.H.; Grobbee, D.; Boeing, H.; Lahmann, P.H.; Nagel, G.; Chang-Claude, J.; Clavel-Chapelon, F.; Fournier, A.; Thiébaud, A.; González, C.A.; Quirós, J.R.; Tormo, M.-J.; Ardanaz, E.; Amiano, P.; Krogh, V.; Palli, D.; Panico, S.; Tumino, R.; Vineis, P.; Trichopoulou, A.; Kalapothaki, V.; Trichopoulos, D.; Ferrari, P.; Norat, T.; Saracci, R.; Riboli, E. Postmenopausal serum androgens, oestrogens and breast cancer risk: the European prospective investigation into cancer and nutrition. *Endocr. Relat. Cancer*, 2005, 12, 1071-1082.
- [7] Rim, A.; Chellman-Jeffers, M. Trends in breast cancer screening and diagnosis. *Cleve. Clin. J. Med.*, 2008, 75, S2-S9.
- [8] Bissonauth, V.; Shatenstein, B.; Ghadirian, P. Nutrition and breast cancer among sporadic cases and gene mutation carriers: An overview. *Cancer Detect. Prev.*, 2008, 32, 52-64.
- [9] Perou, C.M.; Sorlie, T.; Eisen, M.B.; van de Rijn, M.; Jeffrey, S.S.; Rees, C.A.; Pollack, J.R.; Ross, D.T.; Johnsen, H.; Akslen, L.A.; Fluge, Ø.; Pergamenschikov, A.; Williams, C.; Zhu, S.X.; Lonning, P.E.; Borresen-Dale, A.L.; Brown, P.O.; Botstein, D. Molecular portraits of human breast tumours. *Nature*, 2000, 406, 747-752.
- [10] Sorlie, T.; Perou, C.M.; Tibshirani, R.; Aas, T.; Geisler, S.; Johnsen, H.; Hastie, T.; Eisen, M.B.; van de Rijn, M.; Jeffrey, S.S.; Thorsen, T.; Quist, H.; Matese, J.C.; Brown, P.O.; Botstein, D.; Eystein Lonning, P.; Borresen-Dale, A.L. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl. Acad. Sci. USA*, 2001, 98, 10869-10874.
- [11] Gort, M.; Broekhuis, M.; Otter, R.; Klazinga, N.S. Improvement of best practice in early breast cancer: actionable surgeon and hospital factors. *Breast Cancer Res. Treat.*, 2007, 102, 219-226.
- [12] Eisinger, F.; Sobol, H.; Serin, D.; Whorton, J.C. Hereditary breast cancer, circa 1750. *Lancet*, 1998, 351, 1366.
- [13] Lynch, H.T. Introduction to Cancer Genetics. In *Cancer Genetics*, Charles C. Thomas: Springfield, IL, 1976; pp. 3-31.
- [14] Holm, N.V.; Hauge, M.; Harvald, B. Etiologic factors of breast cancer elucidated by a study of unselected twins. *J. Natl. Cancer Inst.*, 1980, 65, 285-298.
- [15] Williams, W.R.; Anderson, D.E. Genetic epidemiology of breast cancer: segregation analysis of 200 Danish pedigrees. *Genet. Epidemiol.*, 1984, 1, 7-20.
- [16] Newman, B.; Austin, M.A.; Lee, M.; King, M.-C. Inheritance of human breast cancer: Evidence for autosomal dominant transmission in high-risk families. *Proc. Natl. Acad. Sci. USA*, 1988, 85, 3044-3048.
- [17] Houlston, R.S.; McCarter, E.; Parbhoo, S.; Scurr, J.H.; Slack, J. Family history and risk of breast cancer. *J. Med. Genet.*, 1992, 29, 154-157.
- [18] Gelehrter, T.D.; Collins, F.S.; Ginsburg, D. Cancer Genetics. In *Principles of Medical Genetics*, Lippincott, Williams, and Wilkins: Philadelphia, PA, 1998; pp. 245-272.
- [19] Hall, J.M.; Lee, M.K.; Newman, B.; Morrow, J.E.; Anderson, L.A.; Huey, B.; King, M.C. Linkage of early-onset familial breast cancer to chromosome 17q21. *Science*, 1990, 250, 1684-1689.
- [20] Miki, Y.; Swensen, J.; Shattuck-Eidens, D.; Futreal, P.A.; Harshman, K.; Tavtigian, S.; Liu, Q.; Cochran, C.; Bennett, L.M.; Ding, W.; Bell, R.; Rosenthal, J.; Hussey, C.; Tran, T.; McClure, M.; Frye, C.; Hattier, T.; Phelps, R.; Haugen-Strano, A.; Katcher, H.; Yakumo, K.; Gholami, Z.; Shaffer, D.; Stone, S.; Bayer, S.; Wray, C.; Bogden, R.; Dayananth, P.; Ward, J.; Tonin, P.; Narod, S.; Bristow, P.K.; Norris, F.H.; Helvering, L.; Morrison, P.; Rosteck, P.; Lai, M.; Barrett, J.C.; Lewis, C.; Neuhausen, S.; Cannon-Albright, L.; Goldgar, D.; Wiseman, R.; Kamb, A.; Skolnick, M.H. A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science*, 1994, 266, 66-71.
- [21] Wooster, R.; Neuhausen, S.L.; Mangion, J.; Quirk, Y.; Ford, D.; Collins, N.; Nguyen, K.; Seal, S.; Tran, T.; Averill, D.; Fields, P.; Marshall, G.; Narod, S.; Lenoir, G.M.; Lynch, H.; Feunteun, J.; Devilee, P.; Cornelisse, C.J.; Menko, F.H.; Daly, P.A.; Ormiston, W.; McManus, R.; Pye, C.; Lewis, C.M.; Cannon-Albright, L.A.; Peto, J.; Ponder, B.A.J.; Skolnick, M.H.; Easton, D.F.; Goldgar,

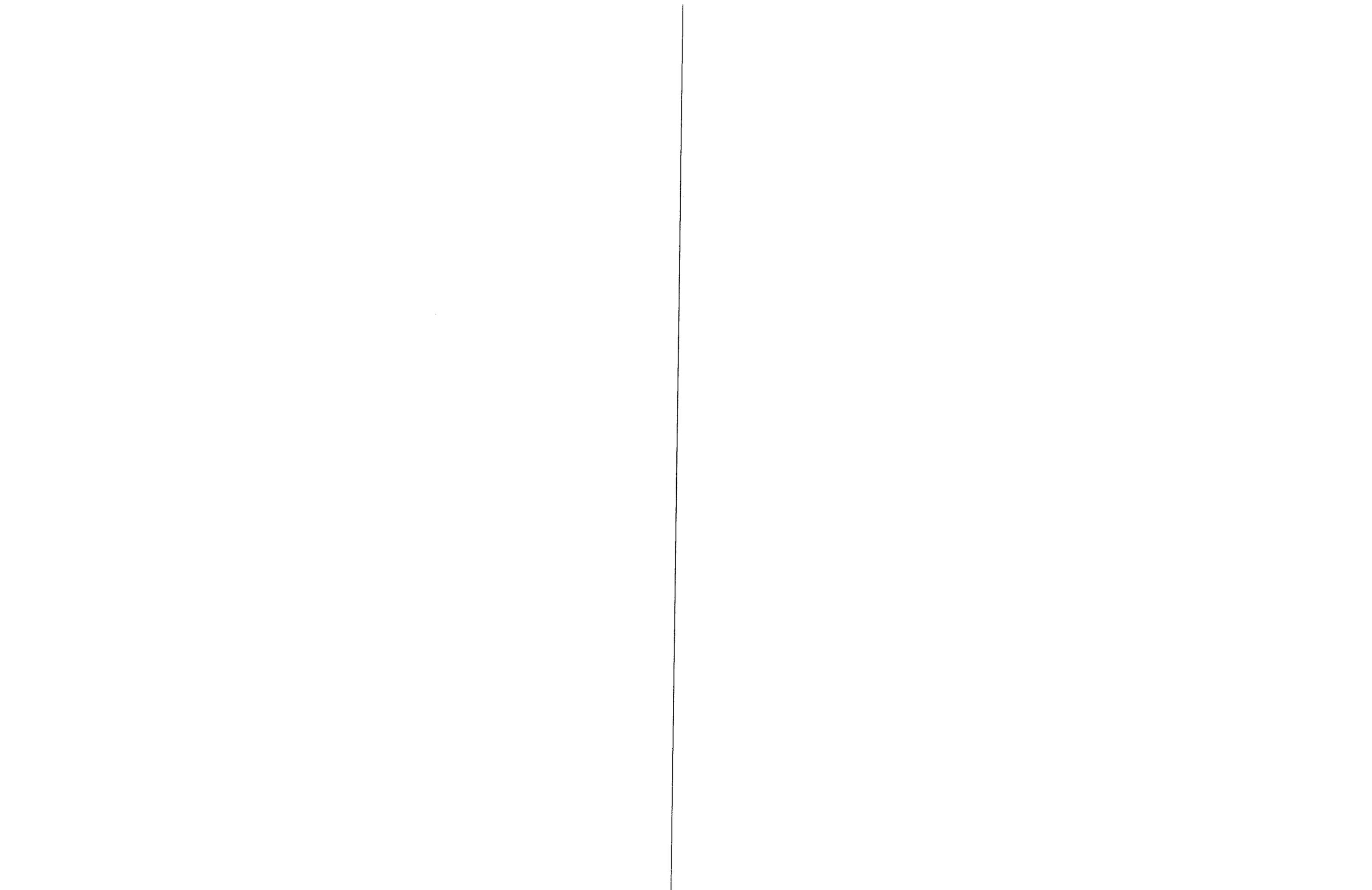
- D.E.; Stratton, M.R. Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12-13. *Science*, 1994, 265, 2088-2090.
- [22] Wooster, R.; Bignell, G.; Lancaster, J.; Swift, S.; Seal, S.; Mangion, J.; Collins, N.; Gregory, S.; Gumbs, C.; Micklem, G.; Barfoot, R.; Hamoudi, R.; Patel, S.; Rice, C.; Biggs, P.; Hashim, Y.; Smith, A.; Connor, F.; Arason, A.; Gudmundsson, J.; Ficenece, D.; Kelsell, D.; Ford, D.; Tonin, P.; Bishop, D.T.; Spurr, N.K.; Ponder, B.A.J.; Eeles, R.; Peto, J.; Devilee, P.; Cornelisse, C.; Lynch, H.; Narod, S.; Lenoir, G.; Eglisson, V.; Barkadottir, R.B.; Easton, D.F.; Bentley, D.R.; Futreal, P.A.; Ashworth, A.; Stratton, M.R. Identification of the breast cancer susceptibility gene BRCA2. *Nature*, 1995, 378, 789-792.
- [23] Nelson, H.D.; Huffman, L.H.; Fu, R.; Harris, E.L. U.S. Preventive Services Task Force. Genetic risk assessment and BRCA mutation testing for breast and ovarian cancer susceptibility: systematic evidence review for the U.S. Preventive Services Task Force. *Ann. Intern. Med.*, 2005, 143, 362-379.
- [24] Hartmann, L.C.; Schaid, D.J.; Woods, J.E.; Crotty, T.P.; Myers, J.L.; Arnold, P.G.; Petty, P.M.; Sellers, T.A.; Johnson, J.L.; McDonnell, S.K.; Frost, M.H.; Jenkins, R.B. Efficacy of bilateral prophylactic mastectomy in women with a family history of breast cancer. *N. Engl. J. Med.*, 1999, 340, 77-84.
- [25] Domchek, S.M.; Weber, B.L. Clinical management of BRCA1 and BRCA2 mutation carriers. *Oncogene*, 2006, 25, 5825-5831.
- [26] Narod, S.A.; Brunet, J.-S.; Ghadirian, P.; Robson, M.; Heimdal, K.; Neuhausen, S.L.; Stoppa-Lyonnet, D.; Lerman, C.; Pasini, B.; de los Rios, P.; Weber, B.; Lynch, H. Hereditary Breast Cancer Clinical Study Group. Tamoxifen and risk of contralateral breast cancer in BRCA1 and BRCA2 mutation carriers: a case-control study. Hereditary Breast Cancer Clinical Study Group. *Lancet*, 2000, 356, 1876-1881.
- [27] Kriege, M.; Brekelmans, C.T.M.; Boetes, C.; Besnard, P.E.; Zonderland, H.M.; Obdeijn, I.M.; Manolliu, R.A.; Kok, T.; Peterse, H.; Tilanus-Linthorst, M.M.A.; Muller, S.H.; Meijer, S.; Oosterwijk, J.C.; Beex, L.V.A.; Tollenaar, R.A.E.; de Koning, H.J.; Rutgers, E.J.T.; Klijn, J.G.M. Magnetic Resonance Imaging Screening Study Group. Efficacy of MRI and mammography for breast-cancer screening in women with a familial or genetic predisposition. *N. Engl. J. Med.*, 2004, 351, 427-437.
- [28] Schwartz, G.F.; Hughes, K.S.; Lynch, H.T.; Fabian, C.J.; Fentiman, I.S.; Robson, M.E.; Domchek, S.M.; Hartmann, L.C.; Holland, R.; Winchester, D.J. The Consensus Conference Committee. Proceedings of the International Consensus Conference on Breast Cancer Risk, Genetics, and Risk Management, April, 2007. *Cancer*, 2008, 113, 2627-2637.
- [29] Turner, N.C.; Reis-Filho, J.S. Basal-like breast cancer and the BRCA1 phenotype. *Oncogene*, 2006, 25, 5846-5853.
- [30] Robson, M.E.; Chappuis, P.O.; Satagopan, J.; Wong, N.; Boyd, J.; Goffin, J.R.; Hudis, C.; Roberge, D.; Norton, L.; Bégin, L.R.; Offit, K.; Foulkes, W.D. A combined analysis of outcome following breast cancer: differences in survival based on BRCA1/BRCA2 mutation status and administration of adjuvant treatment. *Breast Cancer Res.*, 2004, 6, R8-R17.
- [31] Moller, P.; Evans, D.G.; Reis, M.M.; Gregory, H.; Anderson, E.; Maehle, L.; Laloo, F.; Howell, A.; Apold, J.; Clark, N.; Lucassen, A.; Steel, C.M. Surveillance for familial breast cancer: Differences in outcome according to BRCA mutation status. *Int. J. Cancer*, 2007, 121, 1017-1020.
- [32] Liebens, F.P.; Carly, B.; Pastijn, A.; Rozenberg, S. Management of BRCA1/2 associated breast cancer: A systematic qualitative review of the state of knowledge in 2006. *Eur. J. Cancer*, 2007, 43, 238-257.
- [33] Rennert, G.; Bisland-Naggan, S.; Barnett-Griness, O.; Bar-Joseph, N.; Zhang, S.; Rennert, H.S.; Narod, S.A. Clinical outcomes of breast cancer in carriers of BRCA1 and BRCA2 mutations. *N. Engl. J. Med.*, 2007, 357, 115-123.
- [34] Schwartz, M.D.; Lerman, C.; Brogan, B.; Peshkin, B.N.; Halbert, C.H.; DeMarco, T.; Lawrence, W.; Main, D.; Finch, C.; Magnant, C.; Pennanen, M.; Tsangaris, T.; Willey, S.; Isaacs, C. Impact of BRCA1/BRCA2 counseling and testing on newly diagnosed breast cancer patients. *J. Clin. Oncol.*, 2004, 22, 1823-1829.
- [35] Tutt, A.N.; Lord, C.J.; McCabe, N.; Farmer, H.; Turner, N.; Martin, N.M.; Jackson, S.P.; Smith, G.C.; Ashworth, A. Exploiting the DNA repair defect in BRCA mutant cells in the design of new therapeutic strategies for cancer. *Cold Spring Harb. Symp. Quant. Biol.*, 2005, 70, 139-148.
- [36] Farmer, H.; McCabe, N.; Lord, C.J.; Tutt, A.N.J.; Johnson, D.A.; Richardson, T.B.; Santarosa, M.; Dillon, K.J.; Hickson, I.; Knights, C.; Martin, N.M.B.; Jackson, S.P.; Smith, G.C.M.; Ashworth, A. Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature*, 2005, 434, 917-921.
- [37] Tutt, A.; Ashworth, A. Can genetic testing guide treatment in breast cancer? *Eur. J. Cancer*, 2008, 44, 2774-2780.
- [38] Thompson, D.; Easton, D. Breast Cancer Linkage Consortium. Variation in cancer risks, by mutation position, in BRCA2 mutation carriers. *Am. J. Hum. Genet.*, 2001, 68, 410-419.
- [39] Thompson, D.; Easton, D. Breast Cancer Linkage Consortium. Variation in BRCA1 cancer risks by mutation position. *Cancer Epidemiol. Biomarkers Prev.*, 2002, 11, 329-336.
- [40] Easton, D.F.; Ford, D.; Bishop, D.T. Breast Cancer Linkage Consortium. Breast and ovarian cancer incidence in BRCA1-mutation carriers. *Am. J. Hum. Genet.*, 1995, 56, 265-271.
- [41] Weyandt, J.; Ellsworth, R.E.; Hooke, J.A.; Shriver, C.D.; Ellsworth, D.L. Environmental chemicals and breast cancer risk—a structural chemistry perspective. *Curr. Med. Chem.*, 2008, 15, 2680-2701.
- [42] King, M.-C.; Marks, J.H.; Mandell, J.B. New York Breast Cancer Study Group. Breast and ovarian cancer risks due to inherited mutations in BRCA1 and BRCA2. *Science*, 2003, 302, 643-646.
- [43] Antoniou, A.C.; Sinilnikova, O.M.; Simard, J.; Léoné, M.; Dumont, M.; Neuhausen, S.L.; Struwing, J.P.; Stoppa-Lyonnet, D.; Barjhoux, L.; Hughes, D.J.; Coupier, I.; Belotti, M.; Lasset, C.; Bonadona, V.; Bignon, Y.-J.; Genetic Modifiers of Cancer Risk in BRCA1/2 Mutation Carriers Study (GEMO); Rebbeck, T.R.; Wagner, T.; Lynch, H.T.; Domchek, S.M.; Nathanson, K.L.; Garber, J.E.; Weitzel, J.; Narod, S.A.; Tomlinson, G.; Olopade, O.I.; Godwin, A.; Isaacs, C.; Jakubowska, A.; Lubinski, J.; Gronwald, J.; Górski, B.; Byrski, T.; Huzarski, T.; Peock, S.; Cook, M.; Baynes, C.; Murray, A.; Rogers, M.; Daly, P.A.; Dorkins, H.; Epidemiological Study of BRCA1 and BRCA2 Mutation Carriers (EMBRACE); Schmutzler, R.K.; Versmold, B.; Engel, C.; Meindl, A.; Arnold, N.; Neideracher, D.; Deissler, H.; German Consortium for Hereditary Breast and Ovarian Cancer (GCHBOC); Spurdle, A.B.; Chen, X.; Waddell, N.; Cloonan, N.; The Kathleen Cuninghame Consortium for Research into Familial Breast Cancer (kConFab); Kirschhoff, T.; Offit, K.; Friedman, E.; Kaufmann, B.; Laitman, Y.; Galore, G.; Rennert, G.; Lejbkiewicz, F.; Raskin, L.; Andrulis, I.L.; Ilyushik, E.; Ozcelik, H.; Devilee, P.; Vreeswijk, M.P.G.; Greene, M.H.; Prindiville, S.A.; Osorio, A.; Benitez, J.; Zikan, M.; Szabo, C.I.; Kilpivaara, O.; Nevanlinna, H.; Hamann, U.; Durocher, F.; Arason, A.; Couch, F.J.; Easton, D.F.; Chenevix-Trench, G. Consortium of Investigators of Modifiers of BRCA1/2 (CIMBA). RAD51 135G->C modifies breast cancer risk among BRCA2 mutation carriers: results from a combined analysis of 19 studies. *Am. J. Hum. Genet.*, 2007, 81, 1186-1200.
- [44] Hughes, D.J. Use of association studies to define genetic modifiers of breast cancer risk in BRCA1 and BRCA2 mutation carriers. *Fam. Cancer*, 2008, 7, 233-244.
- [45] Chenevix-Trench, G.; Milne, R.L.; Antoniou, A.C.; Couch, F.J.; Easton, D.F.; Goldgar, D.E. CIMBA. An international initiative to identify genetic modifiers of cancer risk in BRCA1 and BRCA2 mutation carriers: the Consortium of Investigators of Modifiers of BRCA1 and BRCA2 (CIMBA). *Breast Cancer Res.*, 2007, 9, 104.
- [46] Ford, D.; Easton, D.F.; Stratton, M.; Narod, S.; Goldgar, D.; Devilee, P.; Bishop, D.T.; Weber, B.; Lenoir, G.; Chang-Claude, J.; Sobol, H.; Teare, M.D.; Struwing, J.; Arason, A.; Scherneck, S.; Peto, J.; Rebbeck, T.R.; Tonin, P.; Neuhausen, S.; Barkadottir, R.; Eyfjord, J.; Lynch, H.; Ponder, B.A.J.; Gayther, S.A.; Birch, J.M.; Lindblom, A.; Stoppa-Lyonnet, D.; Bignon, Y.; Borg, A.; Hamann, U.; Haites, N.; Scott, R.J.; Maugard, C.M.; Vasen, H.; Seitz, S.; Cannon-Albright, L.A.; Schofield, A.; Zelada-Hedman, M. Breast Cancer Linkage Consortium. Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families. *Am. J. Hum. Genet.*, 1998, 62, 676-689.
- [47] Nathanson, K.L.; Weber, B.L. "Other" breast cancer susceptibility genes: searching for more holy grail. *Hum. Mol. Genet.*, 2001, 10, 715-720.
- [48] Lacroix, M.; Leclercq, G. BreastMed Consortium. The "portrait" of hereditary breast cancer. *Breast Cancer Res. Treat.*, 2005, 89, 297-304.

- [49] Gail, M.H.; Brinton, L.A.; Byar, D.P.; Corle, D.K.; Green, S.B.; Schairer, C.; Mulvihill, J.J. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *J. Natl. Cancer Inst.*, 1989, *81*, 1879-1886.
- [50] Claus, E.B.; Risch, N.; Thompson, W.D. Autosomal dominant inheritance of early-onset breast cancer. Implications for risk prediction. *Cancer*, 1994, *73*, 643-651.
- [51] Berry, D.A.; Parmigiani, G.; Sanchez, J.; Schildkraut, J.; Winer, E. Probability of carrying a mutation of breast-ovarian cancer gene BRCA1 based on family history. *J. Natl. Cancer Inst.*, 1997, *89*, 227-237.
- [52] Euhus, D.M. Understanding mathematical models for breast cancer risk assessment and counseling. *Breast J.*, 2001, *7*, 224-232.
- [53] Jacobi, C.E.; de Bock, G.H.; Siegerink, B.; van Asperen, C.J. Differences and similarities in breast cancer risk assessment models in clinical practice: which model to choose? *Breast Cancer Res. Treat.*, 2009, *115*, 381-390.
- [54] Tyrer, J.; Duffy, S.W.; Cuzick, J. A breast cancer prediction model incorporating familial and personal risk factors. *Stat. Med.*, 2004, *23*, 1111-1130.
- [55] Evans, D.G.R.; Howell, A. Breast cancer risk-assessment models. *Breast Cancer Res.*, 2007, *9*, 213.
- [56] MacPherson, G.; Healey, C.S.; Teare, M.D.; Balasubramanian, S.P.; Reed, M.W.R.; Pharoah, P.D.P.; Ponder, B.A.J.; Meuth, M.; Bhattacharyya, N.P.; Cox, A. Association of a common variant of the CASP8 gene with reduced risk of breast cancer. *J. Natl. Cancer Inst.*, 2004, *96*, 1866-1869.
- [57] Dunning, A.M.; Healey, C.S.; Pharoah, P.D.P.; Teare, M.D.; Ponder, B.A.J.; Easton, D.F. A systematic review of genetic polymorphisms and breast cancer risk. *Cancer Epidemiol. Biomarkers Prev.*, 1999, *8*, 843-854.
- [58] Hirschhorn, J.N.; Daly, M.J. Genome-wide association studies for common diseases and complex traits. *Nat. Rev. Genet.*, 2005, *6*, 95-108.
- [59] Frazer, K.A.; Murray, S.S.; Schork, N.J.; Topol, E.J. Human genetic variation and its contribution to complex traits. *Nat. Rev. Genet.*, 2009, *10*, 241-251.
- [60] Easton, D.F.; Pooley, K.A.; Dunning, A.M.; Pharoah, P.D.P.; Thompson, D.; Ballinger, D.G.; Struwing, J.P.; Morrison, J.; Field, H.; Luben, R.; Wareham, N.; Ahmed, S.; Healey, C.S.; Bowman, R.; the SEARCH collaborators; Meyer, K.B.; Haiman, C.A.; Kolonel, L.K.; Henderson, B.E.; Le Marchand, L.; Brennan, P.; Sangrajrang, S.; Gaborieau, V.; Odefrey, F.; Shen, C.-Y.; Wu, P.-E.; Wang, H.-C.; Eccles, D.; Evans, D.G.; Peto, J.; Fletcher, O.; Johnson, N.; Seal, S.; Stratton, M.R.; Rahman, N.; Chenevix-Trench, G.; Bojesen, S.E.; Nordestgaard, B.G.; Axelsson, C.K.; Garcia-Closas, M.; Brinton, L.; Chanock, S.; Lissowska, J.; Peplonska, B.; Nevanlinna, H.; Fagerholm, R.; Eerola, H.; Kang, D.; Yoo, K.-Y.; Noh, D.-Y.; Ahn, S.-H.; Hunter, D.J.; Hankinson, S.E.; Cox, D.G.; Hall, P.; Wedren, S.; Liu, J.; Low, Y.-L.; Bogdanova, N.; Schürmann, P.; Dörk, T.; Tollenaar, R.A.E.; Jacobi, C.E.; Devilee, P.; Klijn, J.G.M.; Sigurdson, A.J.; Doody, M.M.; Alexander, B.H.; Zhang, J.; Cox, A.; Brock, I.W.; MacPherson, G.; Reed, M.W.R.; Couch, F.J.; Goode, E.L.; Olson, J.E.; Meijers-Heijboer, H.; van den Ouweland, A.; Uitterlinden, A.; Rivadeneira, F.; Milne, R.L.; Ribas, G.; Gonzalez-Neira, A.; Benítez, J.; Hopper, J.L.; McCredie, M.; Southey, M.; Giles, G.G.; Schroen, C.; Justenhoven, C.; Brauch, H.; Hamann, U.; Ko, Y.-D.; Spurdle, A.B.; Beesley, J.; Chen, X.; kConFab; AOCs Management Group; Mannerman, A.; Kosma, V.-M.; Kataja, V.; Hartikainen, J.; Day, N.E.; Cox, D.R.; Ponder, B.A.J. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature*, 2007, *447*, 1087-1093.
- [61] Hunter, D.J.; Kraft, P.; Jacobs, K.B.; Cox, D.G.; Yeager, M.; Hankinson, S.E.; Wacholder, S.; Wang, Z.; Welch, R.; Hutchinson, A.; Wang, J.; Yu, K.; Chatterjee, N.; Orr, N.; Willett, W.C.; Colditz, G.A.; Ziegler, R.G.; Berg, C.D.; Buys, S.S.; McCarty, C.A.; Spencer Feigelson, H.; Calle, E.E.; Thun, M.J.; Hayes, R.B.; Tucker, M.; Gerhard, D.S.; Fraumeni, J.F., Jr.; Hoover, R.N.; Thomas, G.; Chanock, S.J. A genome-wide association study identifies alleles in *FGFR2* associated with risk of sporadic postmenopausal breast cancer. *Nat. Genet.*, 2007, *39*, 870-874.
- [62] Stacey, S.N.; Manolescu, A.; Sulem, P.; Rafnar, T.; Gudmundsson, J.; Gudjonsson, S.A.; Masson, G.; Jakobsdottir, M.; Thorlacius, S.; Helgason, A.; Aben, K.K.; Strobbe, L.J.; Albers-Akkers, M.T.; Swinkels, D.W.; Henderson, B.E.; Kolonel, L.N.; Le Marchand, L.; Millastre, E.; Andres, R.; Godino, J.; Garcia-Prats, M.D.; Polo, E.; Tres, A.; Mouy, M.; Saemundsdottir, J.; Backman, V.M.; Gudmundsson, L.; Kristjansson, K.; Bergthorsson, J.T.; Kostic, J.; Frigge, M.L.; Geller, F.; Gudbjartsson, D.; Sigurdsson, H.; Jonsdottir, T.; Hrafnkelsson, J.; Johannsson, J.; Sveinsson, T.; Myrdal, G.; Grimsson, H.N.; Jonsson, T.; von Holst, S.; Werelius, B.; Margolin, S.; Lindblom, A.; Mayordomo, J.I.; Haiman, C.A.; Kiemeny, L.A.; Johannsson, O.T.; Gulcher, J.R.; Thorsteinsdottir, U.; Kong, A.; Stefansson, K. Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat. Genet.*, 2007, *39*, 865-869.
- [63] Gold, B.; Kirchhoff, T.; Stefanov, S.; Lautenberger, J.; Viale, A.; Garber, J.; Friedman, E.; Narod, S.; Olshen, A.B.; Gregersen, P.; Kosarin, K.; Olsh, A.; Bergeron, J.; Ellis, N.A.; Klein, R.J.; Clark, A.G.; Norton, L.; Dean, M.; Boyd, J.; Offit, K. Genome-wide association study provides evidence for a breast cancer risk locus at 6q22.33. *Proc. Natl. Acad. Sci. USA*, 2008, *105*, 4340-4345.
- [64] Stacey, S.N.; Manolescu, A.; Sulem, P.; Thorlacius, S.; Gudjonsson, S.A.; Jonsson, G.F.; Jakobsdottir, M.; Bergthorsson, J.T.; Gudmundsson, J.; Aben, K.K.; Strobbe, L.J.; Swinkels, D.W.; van Engelenburg, K.C.A.; Henderson, B.E.; Kolonel, L.N.; Le Marchand, L.; Millastre, E.; Andres, R.; Saez, B.; Lambea, J.; Godino, J.; Polo, E.; Tres, A.; Picelli, S.; Rantala, J.; Margolin, S.; Jonsson, T.; Sigurdsson, H.; Jonsdottir, T.; Hrafnkelsson, J.; Johannsson, J.; Sveinsson, T.; Myrdal, G.; Grimsson, H.N.; Sveinsdottir, S.G.; Alexiusdottir, K.; Saemundsdottir, J.; Sigurdsson, A.; Kostic, J.; Gudmundsson, L.; Kristjansson, K.; Masson, G.; Fackenthal, J.D.; Adebamowo, C.; Ogundiran, T.; Olopade, O.I.; Haiman, C.A.; Lindblom, A.; Mayordomo, J.I.; Kiemeny, L.A.; Gulcher, J.R.; Rafnar, T.; Thorsteinsdottir, U.; Johannsson, O.T.; Kong, A.; Stefansson, K. Common variants on chromosome 5p12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat. Genet.*, 2008, *40*, 703-706.
- [65] Garcia-Closas, M.; Hall, P.; Nevanlinna, H.; Pooley, K.; Morrison, J.; Richesson, D.A.; Bojesen, S.E.; Nordestgaard, B.G.; Axelsson, C.K.; Arias, J.I.; Milne, R.L.; Ribas, G.; Gonzalez-Neira, A.; Benítez, J.; Zamora, P.; Brauch, H.; Justenhoven, C.; Hamann, U.; Ko, Y.-D.; Bruening, T.; Haas, S.; Dörk, T.; Schürmann, P.; Hillemanns, P.; Bogdanova, N.; Bremer, M.; Karstens, J.H.; Fagerholm, R.; Aaltonen, K.; Aittomäki, K.; von Smitten, K.; Blomqvist, C.; Mannerman, A.; Uusitupa, M.; Eskelinen, M.; Tengström, M.; Kosma, V.-M.; Kataja, V.; Chenevix-Trench, G.; Spurdle, A.B.; Beesley, J.; Chen, X.; Australian Ovarian Cancer Management Group; Kathleen Cumingham Foundation Consortium For Research into Familial Breast Cancer; Devilee, P.; van Asperen, C.J.; Jacobi, C.E.; Tollenaar, R.A.E.; Huijts, P.E.A.; Klijn, J.G.M.; Chang-Claude, J.; Kropp, S.; Slanger, T.; Flesch-Janys, D.; Mutschelknauss, E.; Salazar, R.; Wang-Gohrke, S.; Couch, F.; Goode, E.L.; Olson, J.E.; Vachon, C.; Fredericksen, Z.S.; Giles, G.G.; Baglietto, L.; Severi, G.; Hopper, J.L.; English, D.R.; Southey, M.C.; Haiman, C.A.; Henderson, B.E.; Kolonel, L.N.; Le Marchand, L.; Stram, D.O.; Hunter, D.J.; Hankinson, S.E.; Cox, D.G.; Tamimi, R.; Kraft, P.; Sherman, M.E.; Chanock, S.J.; Lissowska, J.; Brinton, L.A.; Peplonska, B.; Hooning, M.J.; Meijers-Heijboer, H.; Collee, J.M.; van den Ouweland, A.; Uitterlinden, A.G.; Liu, J.; Lin, L.Y.; Yuqing, L.; Humphreys, K.; Czene, K.; Cox, A.; Balasubramanian, S.P.; Cross, S.S.; Reed, M.W.R.; Blows, F.; Driver, K.; Dunning, A.; Tyrer, J.; Ponder, B.A.J.; Sangrajrang, S.; Brennan, P.; McKay, J.; Odefrey, F.; Gabrielle, V.; Sigurdson, A.; Doody, M.; Struwing, J.P.; Alexander, B.; Easton, D.F.; Pharoah, P.D. Heterogeneity of breast cancer associations with five susceptibility loci by clinical and pathological characteristics. *PLoS Genet.*, 2008, *4*, e1000054.
- [66] Antoniou, A.C.; Spurdle, A.B.; Sinilnikova, O.M.; Healey, S.; Pooley, K.A.; Schmutzler, R.K.; Versmold, B.; Engel, C.; Meindl, A.; Arnold, N.; Hofmann, W.; Sutter, C.; Niederacher, D.; Deissler, H.; Caldes, T.; Kämpjärvi, K.; Nevanlinna, H.; Simard, J.; Beesley, J.; Chen, X.; Kathleen Cumingham Consortium for Research into Familial Breast Cancer (kConFab); Neuhausen, S.L.; Rebbeck, T.R.; Wagner, T.; Lynch, H.T.; Issacs, C.; Weitzel, J.; Granz, P.A.; Daly, M.B.; Tomlinson, G.; Olopade, O.I.; Blum, J.L.; Couch, F.J.; Peterlongo, P.; Manoukian, S.; Barile, M.; Radice, P.; Szabo, C.I.; Pereira, L.H.; Greene, M.H.; Rennert, G.; Lejbkovicz, F.; Barnett-Griness, O.; Andrulis, I.L.; Ozelik, H.; OCGN; Gerdes, A.M.; Caligo, M.A.; Laitman, Y.; Kaufman, B.; Milgrom, R.; Friedman, E.; Swedish BRCA1 and BRCA2 study collaborators; Domcheck,

- S.M.; Nathanson, K.L.; Osorio, A.; Llorca, G.; Milne, R.L.; Benitez, J.; Hamann, U.; Hogervorst, F.B.; Manders, P.; Ligtenberg, M.J.; van den Ouweland, A.M.; DNA-HEBON collaborators; Peock, S.; Cook, M.; Platte, R.; Evans, D.G.; Eeles, R.; Pichert, G.; Chu, C.; Eccles, D.; Davidson, R.; Douglas, F.; EMBRACE; Godwin, A.K.; Barjhoux, L.; Mazoyer, S.; Sobol, H.; Bourdon, V.; Eisinger, F.; Chompret, A.; Capoulade, C.; Bressac-de Paillerets, B.; Lenoir, G.M.; Gauthier-Villars, M.; Houdayer, C.; Stoppa-Lyonnet, D.; GEMO; Chenevix-Trench, G.; Easton, D.F.; CIMBA. Common breast cancer-predisposition alleles are associated with breast cancer risk in BRCA1 and BRCA2 mutation carriers. *Am. J. Hum. Genet.*, 2008, 82, 937-948.
- [67] Pharoah, P.D.P.; Antoniou, A.C.; Easton, D.F.; Ponder, B.A.J. Polygenes, risk prediction, and targeted prevention of breast cancer. *N. Engl. J. Med.*, 2008, 358, 2796-2803.
- [68] Bodmer, W.; Bonilla, C. Common and rare variants in multifactorial susceptibility to common diseases. *Nat. Genet.*, 2008, 40, 695-701.
- [69] Ambrosone, C.B. The promise and limitations of genome-wide association studies to elucidate the causes of breast cancer. *Breast Cancer Res.*, 2007, 9, 114.
- [70] Olopade, O.I.; Grushko, T.A.; Nanda, R.; Huo, D. Advances in breast cancer: Pathways to personalized medicine. *Clin. Cancer Res.*, 2008, 14, 7988-7999.
- [71] Kaye, J. The regulation of direct-to-consumer genetic tests. *Hum. Mol. Genet.*, 2008, 17, R180-R183.
- [72] Lenzer, J.; Brownlee, S. Knowing me, knowing you. *BMJ*, 2008, 336, 858-860.
- [73] Geransar, R.; Einsiedel, E. Evaluating online direct-to-consumer marketing of genetic tests: informed choices or buyers beware? *Genet. Test.*, 2008, 12, 13-23.
- [74] Hogarth, S.; Javitt, G.; Melzer, D. The current landscape for direct-to-consumer genetic testing: Legal, ethical, and policy issues. *Annu. Rev. Genomics Hum. Genet.*, 2008, 9, 161-182.
- [75] Sharma, P.; Sahni, N.S.; Tibshirani, R.; Skaane, P.; Urdal, P.; Berghagen, H.; Jensen, M.; Kristiansen, L.; Moen, C.; Sharma, P.; Zaka, A.; Arnes, J.; Sauer, T.; Aksten, L.A.; Schlichting, E.; Borresen-Dale, A.-L.; Lønneborg, A. Early detection of breast cancer based on gene-expression patterns in peripheral blood cells. *Breast Cancer Res.*, 2005, 7, R634-R644.
- [76] Sastre-Garau, X.; Jouve, M.; Asselain, B.; Vincent-Salomon, A.; Beuzeboac, P.; Dorval, T.; Durand, J.-C.; Fourquet, A.; Pouillart, P. Infiltrating lobular carcinoma of the breast. Clinicopathologic analysis of 975 cases with reference to data on conservative therapy and metastatic patterns. *Cancer*, 1996, 77, 113-120.
- [77] Louwman, M.W.J.; Vriezen, M.; van Beek, M.W.P.; Nolthenius-Puyllaert, M.C.; van der Sangen, M.J.C.; Roumen, R.M.; Kiemeny, L.A.; Coebergh, J.W.W. Uncommon breast tumors in perspective: Incidence, treatment and survival in the Netherlands. *Int. J. Cancer*, 2007, 121, 127-135.
- [78] Kleer, C.G.; van Golen, K.L.; Merajver, S.D. Molecular biology of breast cancer metastasis. Inflammatory breast cancer: clinical syndrome and molecular determinants. *Breast Cancer Res.*, 2000, 2, 423-429.
- [79] Engel, J.; Eckel, R.; Aydemir, U.; Aydemir, S.; Kerr, J.; Schlesinger-Raab, A.; Dirschedl, P.; Hölzel, D. Determinants and prognoses of locoregional and distant progression in breast cancer. *Int. J. Radiat. Oncol. Biol. Phys.*, 2003, 55, 1186-1195.
- [80] Ross, J.S.; Harbeck, N. Prognostic and Predictive Factors Overview. In *Molecular Oncology of Breast Cancer*, Ross, J.S.; Hortobagyi, G.N., Eds.; Jones and Bartlett Publishers: Sudbury, MA, 2005, pp. 128-141.
- [81] Park, C.C.; Mitsumori, M.; Nixon, A.; Recht, A.; Connolly, J.; Gelman, R.; Silver, B.; Hetelekidis, S.; Abner, A.; Harris, J.R.; Schnitt, S.J. Outcome at 8 years after breast-conserving surgery and radiation therapy for invasive breast cancer: influence of margin status and systemic therapy on local recurrence. *J. Clin. Oncol.*, 2000, 18, 1668-1675.
- [82] Kunos, C.; Latson, L.; Overmoyer, B.; Silverman, P.; Shenk, R.; Kinsella, T.; Lyons, J. Breast conservation surgery achieving >2 mm tumor-free margins results in decreased local-regional recurrence rates. *Breast J.*, 2006, 12, 28-36.
- [83] Elston, C.W.; Ellis, I.O. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. *Histopathology*, 1991, 19, 403-410.
- [84] Henson, D.E.; Ries, L.; Freedman, L.S.; Carriaga, M. Relationship among outcome, stage of disease, and histologic grade for 22,616 cases of breast cancer: the basis for a prognostic index. *Cancer*, 1991, 68, 2142-2149.
- [85] Robbins, P.; Pinder, S.; de Klerk, N.; Dawkins, H.; Harvey, J.; Sterrett, G.; Ellis, I.; Elston, C. Histological grading of breast carcinomas: a study of interobserver agreement. *Hum. Pathol.*, 1995, 26, 873-879.
- [86] Ries, L.A.G.; Melbert, D.; Krapcho, M.; Mariotto, A.; Miller, B.A.; Feuer, E.J.; Clegg, L.; Horner, M.J.; Howlader, N.; Eisner, M.P.; Reichman, M.; Edwards, B.K. *SEER Cancer Statistics Review, 1975-2004*. National Cancer Institute: Bethesda, MD, 2007, [http://seer.cancer.gov/csr/1975\\_2004/](http://seer.cancer.gov/csr/1975_2004/).
- [87] Elledge, R.M.; Fuqua, S.A. Estrogen and Progesterone Receptors. In *Diseases of the Breast*, Harris, J.R., Ed.; Lippincott, Williams, and Wilkins: Philadelphia, PA, 2000, pp. 471-488.
- [88] Early Breast Cancer Trialists' Collaborative Group (EBCTCG). Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. *Lancet*, 2005, 365, 1687-1717.
- [89] Slamon, D.J.; Godolphin, W.; Jones, L.A.; Holt, J.A.; Wong, S.G.; Keith, D.E.; Levin, W.J.; Stuart, S.G.; Udove, J.; Ullrich, A.; Press, M.F. Studies of the HER-2-neu proto-oncogene in human breast and ovarian cancer. *Science*, 1989, 244, 707-712.
- [90] Wolff, A.C.; Hammond, M.E.H.; Schwartz, J.N.; Hagerty, K.L.; Allred, D.C.; Cote, R.J.; Dowsett, M.; Fitzgibbons, P.L.; Hanna, W.M.; Langer, A.; McShane, L.M.; Paik, S.; Pegram, M.D.; Perez, E.A.; Press, M.F.; Rhodes, A.; Sturgeon, C.; Taube, S.E.; Tubbs, R.; Vance, G.H.; van de Vijver, M.; Wheeler, T.M.; Hayes, D.F. American Society of Clinical Oncology; College of American Pathologists. American Society of Clinical Oncology/College of American Pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer. *J. Clin. Oncol.*, 2007, 25, 118-145.
- [91] Slamon, D.J.; Leyland-Jones, B.; Shak, S.; Fuchs, H.; Paton, V.; Bajamonde, A.; Fleming, T.; Eiermann, W.; Wolter, J.; Pegram, M.; Baselga, J.; Norton, L. Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *N. Engl. J. Med.*, 2001, 344, 783-792.
- [92] Irvin, W.J., Jr.; Carey, L.A. What is triple-negative breast cancer? *Eur. J. Cancer*, 2008, 44, 2799-2805.
- [93] Urruticoechea, A.; Smith, I.E.; Dowsett, M. Proliferation marker Ki-67 in early breast cancer. *J. Clin. Oncol.*, 2005, 23, 7212-7220.
- [94] Dowsett, M.; Smith, I.E.; Ebbs, S.R.; Dixon, J.M.; Skene, A.; A'hem, R.; Salter, J.; Detre, S.; Hills, M.; Walsh, G.; IMPACT Trialists Group. Prognostic value of Ki67 expression after short-term presurgical endocrine therapy for primary breast cancer. *J. Natl. Cancer Inst.*, 2007, 99, 167-170.
- [95] Jones, R.L.; Salter, J.; A'hem, R.; Nerurkar, A.; Parton, M.; Reis-Filho, J.S.; Smith, I.E.; Dowsett, M. The prognostic significance of Ki67 before and after neoadjuvant chemotherapy in breast cancer. *Breast Cancer Res. Treat.*, 2009, 116, 53-68.
- [96] Ravdin, P.M.; Siminoff, L.A.; Davis, G.J.; Mercer, M.B.; Hewlett, J.; Gerson, N.; Parker, H.L. Computer program to assist in making decisions about adjuvant therapy for women with early breast cancer. *J. Clin. Oncol.*, 2001, 19, 980-991.
- [97] Ring, B.Z.; Seitz, R.S.; Beck, R.; Shasteen, W.J.; Tarr, S.M.; Cheang, M.C.; Yoder, B.J.; Budd, G.T.; Nielsen, T.O.; Hicks, D.G.; Estopinal, N.C.; Ross, D.T. Novel prognostic immunohistochemical biomarker panel for estrogen receptor-positive breast cancer. *J. Clin. Oncol.*, 2006, 24, 3039-3047.
- [98] Hu, Z.; Fan, C.; Oh, D.S.; Marron, J.S.; He, X.; Qaqish, B.F.; Livasy, C.; Carey, L.A.; Reynolds, E.; Dressler, L.; Nobel, A.; Parker, J.; Ewend, M.G.; Sawyer, L.R.; Wu, J.; Liu, Y.; Nanda, R.; Treiakova, M.; Ruiz Orrico, A.; Dreher, D.; Palazzo, J.P.; Perreard, L.; Nelson, E.; Mone, M.; Hansen, H.; Mullins, M.; Quackenbush, J.F.; Ellis, M.J.; Olopade, O.I.; Bernard, P.S.; Perou, C.M. The molecular portraits of breast tumors are conserved across microarray platforms. *BMC Genomics*, 2006, 7, 96.
- [99] Chang, H.Y.; Sneddon, J.B.; Alizadeh, A.A.; Sood, R.; West, R.B.; Montgomery, K.; Chi, J.-T.; van de Rijn, M.; Botstein, D.; Brown, P.O. Gene expression signature of fibroblast serum response predicts human cancer progression: similarities between tumors and wounds. *PLoS Biology*, 2004, 2, 206-214.
- [100] Chang, H.Y.; Nuyten, D.S.A.; Sneddon, J.B.; Hastie, T.; Tibshirani, R.; Sorlie, T.; Dai, H.; He, Y.D.; van't Veer, L.J.;

- Bartelink, H.; van de Rijn, M.; Brown, P.O.; van de Vijver, M.J. Robustness, scalability, and integration of a wound-response gene expression signature in predicting breast cancer survival. *Proc. Natl. Acad. Sci. USA*, 2005, 102, 3738-3743.
- [101] Nuyten, D.S.A.; Kreike, B.; Hart, A.A.M.; Chi, J.-T.A.; Sneddon, J.B.; Wessels, L.F.A.; Peterse, H.J.; Bartelink, H.; Brown, P.O.; Chang, H.Y.; van de Vijver, M.J. Predicting a local recurrence after breast-conserving therapy by gene expression profiling. *Breast Cancer Res.*, 2006, 8, R62.
- [102] Chi, J.-T.; Wang, Z.; Nuyten, D.S.A.; Rodriguez, E.H.; Schaner, M.E.; Salim, A.; Wang, Y.; Kristensen, G.B.; Helland, A.; Borresen-Dale, A.-L.; Giaccia, A.; Longaker, M.T.; Hastie, T.; Yang, G.P.; van de Vijver, M.J.; Brown, P.O. Gene expression programs in response to hypoxia: cell type specificity and prognostic significance in human cancers. *PLoS Med.*, 2006, 3, e47.
- [103] Sotiriou, C.; Wirapati, P.; Loi, S.; Harris, A.; Fox, S.; Smeds, J.; Nordgren, H.; Farmer, P.; Praz, V.; Haibe-Kains, B.; Desmedt, C.; Larsimont, D.; Cardoso, F.; Peterse, H.; Nuyten, D.; Buyse, M.; van de Vijver, M.J.; Bergh, J.; Piccart, M.; Delorenzi, M. Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. *J. Natl. Cancer Inst.*, 2006, 98, 262-272.
- [104] Loi, S.; Haibe-Kains, B.; Desmedt, C.; Lallemand, F.; Tutt, A.M.; Gillet, C.; Ellis, P.; Harris, A.; Bergh, J.; Foekens, J.A.; Klijn, J.G.M.; Larsimont, D.; Buyse, M.; Bontempi, G.; Delorenzi, M.; Piccart, M.J.; Sotiriou, C. Definition of clinically distinct molecular subtypes in estrogen receptor-positive breast carcinomas through genomic grade. *J. Clin. Oncol.*, 2007, 25, 1239-1246.
- [105] van't Veer, L.J.; Dai, H.; van de Vijver, M.J.; He, Y.D.; Hart, A.A.M.; Mao, M.; Peterse, H.L.; van der Kooy, K.; Marton, M.J.; Witteveen, A.T.; Schreiber, G.J.; Kerkhoven, R.M.; Roberts, C.; Linsley, P.S.; Bernards, R.; Friend, S.H. Gene expression profiling predicts clinical outcome of breast cancer. *Nature*, 2002, 415, 530-536.
- [106] van de Vijver, M.J.; He, Y.D.; van't Veer, L.J.; Dai, H.; Hart, A.A.M.; Voskuil, D.W.; Schreiber, G.J.; Peterse, J.L.; Roberts, C.; Marton, M.J.; Parrish, M.; Atsma, D.; Witteveen, A.; Glas, A.; Delahaye, L.; van der Velde, T.; Bartelink, H.; Rodenhuis, S.; Rutgers, E.T.; Friend, S.H.; Bernards, R. A gene-expression signature as a predictor of survival in breast cancer. *N. Engl. J. Med.*, 2002, 347, 1999-2009.
- [107] Wang, Y.; Klijn, J.G.M.; Zhang, Y.; Sieuwerts, A.M.; Look, M.P.; Yang, F.; Talantov, D.; Timmermans, M.; Meijer-van Gelder, M.E.; Yu, J.; Jatkoe, T.; Berns, E.M.J.; Atkins, D.; Foekens, J.A. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet*, 2005, 365, 671-679.
- [108] Paik, S.; Shak, S.; Tang, G.; Kim, C.; Baker, J.; Cronin, M.; Baehner, F.L.; Walker, M.G.; Watson, D.; Park, T.; Hiller, W.; Fisher, E.R.; Wickerham, D.L.; Bryant, J.; Wolmark, N. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N. Engl. J. Med.*, 2004, 351, 2817-2826.
- [109] Paik, S.; Tang, G.; Shak, S.; Kim, C.; Baker, J.; Kim, W.; Cronin, M.; Baehner, F.L.; Watson, D.; Bryant, J.; Costantino, J.P.; Geyer, C.E., Jr.; Wickerham, D.L.; Wolmark, N. Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. *J. Clin. Oncol.*, 2006, 24, 3726-3734.
- [110] Cardoso, F.; Piccart-Gebhart, M.; van't Veer, L.; Rutgers, E.; TRANSBIG Consortium. The MINDACT trial: the first prospective clinical validation of a genomic tool. *Mol. Oncol.*, 2007, 1, 246-251.
- [111] Sparano, J.A.; Paik, S. Development of the 21-gene assay and its application in clinical practice and clinical trials. *J. Clin. Oncol.*, 2008, 26, 721-728.
- [112] Desmedt, C.; Ruffz-García, E.; André, F. Gene expression predictors in breast cancer: current status, limitations and perspectives. *Eur. J. Cancer*, 2008, 44, 2714-2720.
- [113] Fan, C.; Oh, D.S.; Wessels, L.; Weigelt, B.; Nuyten, D.S.A.; Nobel, A.B.; van't Veer, L.J.; Perou, C.M. Concordance among gene-expression-based predictors for breast cancer. *N. Engl. J. Med.*, 2006, 355, 560-569.
- [114] Ross, J.S.; Hatzis, C.; Fraser Symmans, W.; Pusztai, L.; Hortobágyi, G.N. Commercialized multigene predictors of clinical outcome for breast cancer. *Oncologist*, 2008, 13, 477-493.
- [115] Ross, J.S. Multigene predictors in early-stage breast cancer: moving in or moving out? *Expert Rev. Mol. Diagn.*, 2008, 8, 129-135.
- [116] Pusztai, L.; Stec, J.; Ayers, M.; Ross, J.S.; Wagner, P.; Rouzier, R.; Symmans, F.; Hortobágyi, G.N. Pharmacogenetics, Pharmacogenomics, and Predicting Response to Therapy. In *Molecular Oncology of Breast Cancer*, Ross, J.S.; Hortobágyi, G.N., Eds.; Jones and Bartlett Publishers: Sudbury, MA, 2005, pp. 439-456.
- [117] Jordan, V.C. Tamoxifen: a most unlikely pioneering medicine. *Nat. Rev. Drug Discov.*, 2003, 2, 205-213.
- [118] Fisher, B.; Costantino, J.P.; Wickerham, D.L.; Redmond, C.K.; Kavanah, M.; Cronin, W.M.; Vogel, V.; Robidoux, A.; Dimitrov, N.; Atkins, J.; Daly, M.; Wieand, S.; Tan-Chiu, E.; Ford, L.; Wolmark, N.; National Surgical Adjuvant Breast and Bowel Project Investigators. Tamoxifen for prevention of breast cancer: report of the National Surgical Adjuvant Breast and Bowel Project P-1 Study. *J. Natl. Cancer Inst.*, 1998, 90, 1371-1388.
- [119] Bradford, L.D. CYP2D6 allele frequency in European Caucasians, Asians, Africans and their descendants. *Pharmacogenomics*, 2002, 3, 229-243.
- [120] Schroth, W.; Goetz, M.P.; Hamann, U.; Fasching, P.A.; Schmidt, M.; Winter, S.; Fritz, P.; Simon, W.; Suman, V.J.; Ames, M.M.; Safgren, S.L.; Kuffel, M.J.; Ulrich Ulmer, H.; Boländer, J.; Strick, R.; Beckmann, M.W.; Koelbl, H.; Weinsilboun, R.M.; Ingle, J.N.; Eichelbaum, M.; Schwab, M.; Brauch, H. Association between CYP2D6 polymorphisms and outcomes among women with early stage breast cancer treated with tamoxifen. *JAMA*, 2009, 302, 1429-1436.
- [121] Goetz, M.P.; Rae, J.M.; Suman, V.J.; Safgren, S.L.; Ames, M.M.; Visscher, D.W.; Reynolds, C.; Couch, F.J.; Lingle, W.L.; Flockhart, D.A.; Desta, Z.; Perez, E.A.; Ingle, J.N. Pharmacogenetics of tamoxifen biotransformation is associated with clinical outcomes of efficacy and hot flashes. *J. Clin. Oncol.*, 2005, 23, 9312-9318.
- [122] Tan, S.-H.; Lee, S.-C.; Goh, B.-C.; Wong, J. Pharmacogenetics in breast cancer therapy. *Clin. Cancer Res.*, 2008, 14, 8027-8041.
- [123] Higgins, M.J.; Rae, J.M.; Flockhart, D.A.; Hayes, D.F.; Stearns, V. Pharmacogenetics of tamoxifen: who should undergo CYP2D6 genetic testing? *J. Natl. Compr. Canc. Netw.*, 2009, 7, 203-213.
- [124] Dowsett, M.; Haynes, B.P. Hormonal effects of aromatase inhibitors: focus on premenopausal effects and interaction with tamoxifen. *J. Steroid Biochem. Mol. Biol.*, 2003, 86, 255-263.
- [125] Breast International Group (BIG) 1-98 Collaborative Group; Thürlimann, B.; Keshaviah, A.; Coates, A.S.; Mouridsen, H.; Mauriac, L.; Forbes, J.F.; Paridaens, R.; Castiglione-Gertsch, M.; Gelber, R.D.; Rabaglio, M.; Smith, I.; Wardley, A.; Price, K.N.; Goldhirsch, A. A comparison of letrozole and tamoxifen in postmenopausal women with early breast cancer. *N. Engl. J. Med.*, 2005, 353, 2747-2757.
- [126] Boccardo, F.; Rubagotti, A.; Puntoni, M.; Guglielmini, P.; Amoroso, D.; Fini, A.; Paladini, G.; Mesiti, M.; Romeo, D.; Rinaldini, M.; Scali, S.; Porpiglia, M.; Benedetto, C.; Restuccia, N.; Buzzi, F.; Franchi, R.; Massidda, B.; Distanti, V.; Amadori, D.; Sismondi, P. Switching to anastrozole versus continued tamoxifen treatment of early breast cancer: preliminary results of the Italian Tamoxifen Anastrozole Trial. *J. Clin. Oncol.*, 2005, 23, 5138-5147.
- [127] Goss, P.E.; Ingle, J.N.; Martino, S.; Robert, N.J.; Muss, H.B.; Piccart, M.J.; Castiglione, M.; Tu, D.; Shepherd, L.E.; Pritchard, K.I.; Livingston, R.B.; Davidson, N.E.; Norton, L.; Perez, E.A.; Abrams, J.S.; Cameron, D.A.; Palmer, M.J.; Pater, J.L. Randomized trial of letrozole following tamoxifen as extended adjuvant therapy in receptor-positive breast cancer: updated findings from NCIC CTG MA.17. *J. Natl. Cancer Inst.*, 2005, 97, 1262-1271.
- [128] Ma, C.X.; Adjei, A.A.; Salavaggi, O.E.; Coronel, J.; Pelleymounter, L.; Wang, L.; Eckloff, B.W.; Schaid, D.; Wieben, E.D.; Adjei, A.A.; Weinsilboun, R.M. Human aromatase: gene resequencing and functional genomics. *Cancer Res.*, 2005, 65, 11071-11082.
- [129] Nahta, R.; Esteva, F.J. Herceptin: mechanisms of action and resistance. *Cancer Lett.*, 2006, 232, 123-138.
- [130] Harris, M. Monoclonal antibodies as therapeutic agents for cancer. *Lancet Oncol.*, 2004, 5, 292-302.
- [131] Holbro, T.; Beerli, R.R.; Maurer, F.; Koziczak, M.; Barbas, C.F., III; Hynes, N.E. The ErbB2/ErbB3 heterodimer functions as an

- oncogenic unit: ErbB2 requires ErbB3 to drive breast tumor cell proliferation. *Proc. Natl. Acad. Sci. USA*, 2003, 100, 8933-8938.
- [132] Lee-Hoeflich, S.T.; Crocker, L.; Yao, E.; Pham, T.; Munroe, X.; Hoeflich, K.P.; Sliwkowski, M.X.; Stern, H.M. A central role for HER3 in *HER2*-amplified breast cancer: implications for targeted therapy. *Cancer Res.*, 2008, 68, 5878-5887.
- [133] Agus, D.B.; Akita, R.W.; Fox, W.D.; Lewis, G.D.; Higgins, B.; Pisacane, P.I.; Lofgren, J.A.; Tindell, C.; Evans, D.P.; Maiese, K.; Scher, H.I.; Sliwkowski, M.X. Targeting ligand-activated ErbB2 signaling inhibits breast and prostate tumor growth. *Cancer Cell*, 2002, 2, 127-137.
- [134] Moulder, S.; Hortobagyi, G.N. Advances in the treatment of breast cancer. *Clin. Pharmacol. Ther.*, 2008, 83, 26-36.
- [135] Troester, M.A.; Hoadley, K.A.; Sorlie, T.; Herbert, B.-S.; Borresen-Dale, A.-L.; Lonning, P.E.; Shay, J.W.; Kaufmann, W.K.; Perou, C.M. Cell-type-specific responses to chemotherapeutics in breast cancer. *Cancer Res.*, 2004, 64, 4218-4226.
- [136] Rouzier, R.; Perou, C.M.; Symmans, W.F.; Ibrahim, N.; Cristofanilli, M.; Anderson, K.; Hess, K.R.; Stec, J.; Ayers, M.; Wagner, P.; Morandi, P.; Fan, C.; Rabiul, I.; Ross, J.S.; Hortobagyi, G.N.; Pusztai, L. Breast cancer molecular subtypes respond differently to preoperative chemotherapy. *Clin. Cancer Res.*, 2005, 11, 5678-5685.
- [137] Li, L.-F.; Xu, X.-J.; Zhao, Y.; Liu, Z.-B.; Shen, Z.-Z.; Jin, W.-R.; Shao, Z.-M. Integrated gene expression profile predicts prognosis of breast cancer patients. *Breast Cancer Res. Treat.*, 2009, 113, 231-237.
- [138] Chang, J.C.; Wooten, E.C.; Tsimelzon, A.; Hilsenbeck, S.G.; Gutierrez, M.C.; Elledge, R.; Mohsin, S.; Osborne, C.K.; Chamness, G.C.; Allred, D.C.; O'Connell, P. Gene expression profiling for the prediction of therapeutic response to docetaxel in patients with breast cancer. *Lancet*, 2003, 362, 362-369.
- [139] Bosch, T.M.; Meijerman, I.; Beijnen, J.H.; Schellens, J.H.M. Genetic polymorphisms of drug-metabolising enzymes and drug transporters in the chemotherapeutic treatment of cancer. *Clin. Pharmacokinet.*, 2006, 45, 253-285.
- [140] Marsh, S.; Liu, G. Pharmacokinetics and pharmacogenomics in breast cancer chemotherapy. *Adv. Drug Deliv. Rev.*, 2009, 61, 381-387.
- [141] Hoskins, J.M.; Carey, L.A.; McLeod, H.L. *CYP2D6* and tamoxifen: DNA matters in breast cancer. *Nat. Rev. Cancer*, 2009, 9, 576-586.
- [142] Fan, L.; Goh, B.-C.; Wong, C.-I.; Sukri, N.; Lim, S.-E.; Tan, S.-H.; Guo, J.-Y.; Lim, R.; Yap, H.-L.; Khoo, Y.-M.; Iau, P.; Lee, H.-S.; Lee, S.-C. Genotype of human carbonyl reductase *CBR3* correlates with doxorubicin disposition and toxicity. *Pharmacogenet. Genomics*, 2008, 18, 621-629.
- [143] Ambrosone, C.B.; Ahn, J.; Singh, K.K.; Rezaishiraz, H.; Furberg, H.; Sweeney, C.; Coles, B.; Trovato, A. Polymorphisms in genes related to oxidative stress (*MPO*, *MnSOD*, *CAT*) and survival after treatment for breast cancer. *Cancer Res.*, 2005, 65, 1105-1111.
- [144] Harris, L.; Fritsche, H.; Mennel, R.; Norton, L.; Ravdin, P.; Taube, S.; Somerfield, M.R.; Hayes, D.F.; Bast, R.C., Jr. American Society of Clinical Oncology 2007 update of recommendations for the use of tumor markers in breast cancer. *J. Clin. Oncol.*, 2007, 25, 5287-5312.
- [145] Check Hayden, E. Personalized cancer therapy gets closer. *Nature*, 2009, 458, 131-132.
- [146] Manolio, T.A.; Brooks, L.D.; Collins, F.S. A HapMap harvest of insights into the genetics of common disease. *J. Clin. Invest.*, 2008, 118, 1590-1605.
- [147] Phillips, K.A.; Veenstra, D.L.; Ramsey, S.D.; Van Bebber, S.L.; Sakowski, J. Genetic testing and pharmacogenomics: issues for determining the impact to healthcare delivery and costs. *Am. J. Manag. Care*, 2004, 10, 425-432.
- [148] Salari, K. The dawning era of personalized medicine exposes a gap in medical education. *PLoS Med.*, 2009, 6, e1000138.
- [149] Chang-Claude, J.; Ambrosone, C.B.; Lilla, C.; Kropp, S.; Helmbold, I.; von Fournier, D.; Haase, W.; Sautter-Bühl, M.-L.; Wenz, F.; Schmezer, P.; Popanda, O. Genetic polymorphisms in DNA repair and damage response genes and late normal tissue complications of radiotherapy for breast cancer. *Br. J. Cancer*, 2009, 100, 1680-1686.
- [150] Zárate, R.; González-Santiago, S.; de la Haba, J.; Bandres, E.; Morales, R.; Salgado, J.; Gómez, A.; Aranda, E.; García-Foncillas, J. *GSTP1* and *MTHFR* polymorphisms are related with toxicity in breast cancer adjuvant anthracycline-based treatment. *Curr. Drug Metab.*, 2007, 8, 481-486.



# Loss of Nuclear Localized and Tyrosine Phosphorylated Stat5 in Breast Cancer Predicts Unfavorable Clinical Outcome and Increased Risk of Anti-Estrogen Therapy Failure

Amy Ryder<sup>1</sup>, Agnieszka K. Witkiewicz<sup>2</sup>, Chengbao Liu<sup>1</sup>, Ginger A. Stringer<sup>3</sup>, Alexander C. Klimowicz<sup>4</sup>, Thai H. Tran<sup>1</sup>, Joachim Torhorst<sup>5</sup>, Ning Yang<sup>1</sup>, Anne L. Rosenberg<sup>6</sup>, Jeffrey A. Hooke<sup>7</sup>, Albert J. Kovatich<sup>8</sup>, Marja T. Nevalainen<sup>1</sup>, Craig D. Shriver<sup>7</sup>, Terry Hyslop<sup>3</sup>, Guido Sauter<sup>9</sup>, David L. Rimm<sup>10</sup>, Anthony M. Magliocco<sup>4</sup>, and Hallgeir Rui<sup>1</sup>

<sup>1</sup>Department of Cancer Biology, Kimmel Cancer Center, Philadelphia, Pennsylvania; <sup>2</sup>Department of Pathology, Thomas Jefferson University, Philadelphia, Pennsylvania;

<sup>3</sup>Division of Biostatistics, Thomas Jefferson University, Philadelphia, Pennsylvania; <sup>4</sup>Department of Oncology, Tom Baker Cancer Centre, Alberta Cancer Board, Alberta, Canada; <sup>5</sup>Institute for Pathology, University of Basel, Basel Switzerland;

<sup>6</sup>Department of Surgery, Division of General Surgery, Thomas Jefferson University, Philadelphia, Pennsylvania;

<sup>7</sup>Walter Reed Army Medical Center, Washington, DC, USA;

<sup>8</sup>MDR Global Systems, LLC., Windber, Pennsylvania;

<sup>9</sup>Department of Pathology, Yale University School of Medicine, New Haven, CT 06520, USA;

<sup>10</sup>Department of Pathology, University Medical Center, Hamburg-Eppendorf, Germany

Supported by NIH grants R01-CA101841 and R01-CA118740 (H.R.), Komen for the Cure Promise Grant KG091116 (H.R., A.K.W., C.L., J.A.H., A.J.K., C.D.S., T.H.) DOD CDMRP Predoctoral Fellowship (A.R.), and NCI Support Grant 1P30CA56036 to the Kimmel Cancer Center. The Project is funded, in part, under a Commonwealth University Research Enhancement Program grant with the Pennsylvania Department of Health (H.R.). The Department specifically disclaims responsibility for any analyses, interpretations or conclusions. The views expressed in this article are those of the authors and do not reflect the official policy of the Department of the Army (DOA), Department of Defense (DOD), or US Government.

\*Address correspondence to: Hallgeir Rui, Department of Cancer Biology, Thomas Jefferson University, 233 S. 10<sup>th</sup> St, Philadelphia, PA 19107. hallgeir.rui@jefferson.edu

**Running title:** Stat5 and response to anti-estrogen therapy.

**Key words:** breast cancer, Stat5, hormone therapy, prognosis, response prediction

**Statement of Translational Relevance**

This study provides novel support for nuclear localized and tyrosine phosphorylated Stat5 (Nuc-pYStat5) as an independent prognostic marker in lymph node-negative breast cancer. The study also suggests that loss of Nuc-pYStat5 is an independent predictive marker of increased risk of failure of anti-estrogen adjuvant therapy regardless of node status. Measuring levels of Nuc-pYStat5 may prove informative in identifying patients who are likely to fail anti-estrogen therapy.

**Abstract**

**Purpose:** To examine the utility of nuclear localized and tyrosine phosphorylated Stat5 (Nuc-pYStat5) as a marker of prognosis in node-negative breast cancer and as a predictor of response to anti-estrogen therapy.

**Experimental Design:** Levels of Nuc-pYStat5 were analyzed in five archival materials of breast cancer, by traditional diaminobenzidine-chromogen immunostaining and pathologist scoring in whole tissue sections or by immunofluorescence-based detection and automated quantitative analysis (AQUA) in tissue microarrays.

**Results:** Nuc-pYStat5 was an independent marker of prognosis in node-negative breast cancer when adjusted for other common pathology parameters in multivariate analyses both by standard chromogen detection with pathologist scoring of whole tissue sections (n=301) and by quantitative immunofluorescence of tissue microarrays (n=158). Corresponding results were achieved by two distinct monoclonal antibodies. In a breast cancer progression material (n=130), quantitative AQUA revealed markedly reduced levels of Nuc-pYStat5 in invasive carcinoma compared to normal breast epithelia or DCIS, which were further reduced in lymph node metastases. In a fourth (n=166) and fifth (n=65) breast cancer material from patients who received adjuvant anti-estrogen therapy, loss of Nuc-pYStat5 predicted increased risk of therapy failure. Quantitative AQUA analysis revealed a cohort of patients with low Nuc-pYStat5 at especially elevated risk for failure of adjuvant anti-estrogen therapy (multivariate hazard ratio=7.9, 95% CI=2.5-25.0,  $p<0.001$ ).

**Conclusions:** Nuc-pYStat5 is an independent prognostic marker in node-negative breast cancer, and may be useful as an independent predictive marker of response to adjuvant anti-estrogen therapy regardless of node status. Prospective studies to explore Nuc-pYStat5 as a diagnostic in breast cancer are now justified.

## Introduction

Signal transducer and activator of transcription (Stat5) is a latent cytoplasmic transcription factor and a primary mediator of prolactin signaling in breast epithelia (Gouilleux, Wakao et al. 1994; Liu, Robinson et al. 1997). Upon phosphorylation of Stat5 on a conserved tyrosine residue by Jak2, Stat5 dimers translocate to the cell nucleus and binds to the DNA of target genes (Gouilleux, Wakao et al. 1994). The prolactin-Stat5 pathway promotes growth and differentiation of mammary epithelia, and Stat5 is highly activated in terminally differentiated breast epithelial cells during lactation (Liu, Robinson et al. 1997; Teglund, McKay et al. 1998; Xie, LeBaron et al. 2002; Wagner, Krempler et al. 2004). Stat5 is also activated at a basal level in non-pregnant and non-lactating mouse and human epithelia (Nevalainen, Xie et al. 2002). Stat5 has been implicated as a mammary tumor promoter in mice, since tumor development was delayed in Stat5-deficient mice and was induced in mice expressing a hyper-active Stat5 transgene (Humphreys and Hennighausen 1999; Iavnilovitch, Groner et al. 2002; Ren, Cai et al. 2002). On the other hand, *in vitro* laboratory studies have indicated that active Stat5 promotes cellular differentiation and inhibits invasive characteristics of human breast cancer cell lines (Sultan, Xie et al. 2005; Nouhi, Chughtai et al. 2006; Sultan, Brim et al. 2008). Based on these and other data, a working model where Stat5 acts as a promoter of breast cancer initiation and a suppressor of progression of established breast cancer has been proposed (Wagner and Rui 2008).

Consistent with the notion of a pro-differentiation effect of Stat5 in established human breast cancer, several immunohistochemical studies have reported that reduced levels of Stat5 protein or reduced levels of tyrosine phosphorylated and nuclear localized Stat5 (Nuc-pYStat5) were associated with less well-differentiated morphology, higher histological grade, and more advanced breast cancer (Cotarla, Ren et al. 2004; Nevalainen, Xie et al. 2004; Bratthauer, Strauss et al. 2006; Yamashita, Nishio et al. 2006). Importantly, loss of staining for Nuc-pYStat5 was a marker of poor prognosis in human breast cancer, particularly in node-negative breast cancer (Nevalainen, Xie et al. 2004). A limitation of that initial report was that only tissue microarrays were studied, and it therefore remained to be determined whether Nuc-pYStat5 was an independent marker of prognosis when whole tissue sections were evaluated. Furthermore, reduced levels of Stat5 protein in the primary tumor correlated with resistance to post-relapse anti-estrogen therapy (Yamashita, Nishio et al. 2006). However, Yamashita and colleagues did not examine the levels of nuclear localized and tyrosine phosphorylated Stat5 protein, which is considered to be the DNA-binding and transcriptionally active protein fraction, nor did that report address the predictive value of Stat5 on outcome following standard anti-estrogen adjuvant therapy.

Here we report systematic analyses of Nuc-pYStat5 in five separate archival materials of breast cancer. First, we interrogated a clinical material of node-negative breast cancer represented by whole tissue sections and analyzed by traditional diaminobenzidine (DAB)-chromogen immunohistochemistry (IHC) and pathologist scoring. We then developed an immunofluorescence-based assay using a second monoclonal antibody for quantification of levels of Nuc-pYStat5 in tissues by the objective Automated Quantitative Analysis (AQUA) platform (Camp, Chung et al. 2002; McCabe, Dolled-Filhart et al. 2005), and applied this methodology to a separate material of node-negative breast cancer patients. In both materials, Nuc-pYStat5 remained an independent marker of prognosis when adjusted for other standard pathology parameters in multivariate analyses. Furthermore, quantitative AQUA analysis of a third clinical material representing a progression array of unmatched normal breast tissue, DCIS, invasive breast cancer, and lymph node metastases, revealed that levels of Nuc-pYStat5 were markedly reduced in invasive carcinoma compared to normal breast epithelia or DCIS, and were further reduced in lymph node metastases. Importantly, in a fourth and fifth archival material of breast cancer from patients who received adjuvant anti-estrogen therapy, loss of Nuc-pYStat5 predicted an increased risk of failure to respond to anti-estrogen therapy regardless of lymph node status. Quantitative AQUA analysis specifically revealed the presence of a cohort of patients with low Nuc-pYStat5 who were at especially high risk for failure of anti-estrogen therapy. The present work supports the prognostic and predictive value of Nuc-pYStat5 in breast cancer, and initiation of prospective studies to determine the efficacy of Nuc-pYStat5 as a predictive marker of response to adjuvant anti-estrogen therapy is now warranted.

## Materials and Methods

**Breast Tumor Materials.** Archival and deidentified formalin-fixed paraffin-embedded breast cancer specimens representing five independent clinical materials were obtained for these studies, and included both tissue microarrays and whole tissue sections. Demographic and clinical characteristics of patients in Materials I, II, IV, and V are presented in **Table 1A-B**. Material III is a progression series of normal breast, ductal carcinoma in situ (DCIS), invasive breast cancer and lymph node metastases for which follow-up data or demographics are not available. The use of all tissues was approved by the ethics committee of the respective institutions. Within the various materials, cases with missing data points were not over-represented by any characteristic.

**Material I.** Whole tissue sections from unselected node-negative invasive ductal carcinomas diagnosed from 1974-1996 with corresponding clinical and time to recurrence data were obtained from the National Cancer Institute Cooperative Breast Cancer Tissue Resource (CBCTR). Informative data for each of the three immunohistochemical markers, Nuc-pYStat5, ER, and PR, were obtained for 301 patients.

**Material II.** A tissue microarray was used that contained 651 human breast cancer specimens from the Yale University Department of Pathology archives diagnosed between 1961 and 1983 with corresponding clinical and death of disease data. The patient population and array construction have been previously described in multiple reports (Kang, Dolled-Filhart *et al.* 2003; Dolled-Filhart, Ryden *et al.* 2006; Giltane, Moeder *et al.* 2009). Briefly, the tissue microarray was constructed with single 0.6-mm cores spaced 0.8 mm apart in a grid format using a Tissue Microarrayer (Beecher Instruments, Sun Prairie, WI). The tissue microarray blocks were cut to 5- $\mu$ m sections with a microtome and placed on slides with an adhesive tape-transfer method (Instrumedics, Inc., Hackensack, NJ), and UV cross-linked. Node-negative patients represented 324 tumor spots on the array and, among those, valid Nuc-pYStat5 AQUA scores, clinical markers and follow-up data were available for 158 patients. Complete treatment information was not available for the entire cohort, but the majority of node-negative patients was treated only by surgical resection with local radiation and did not receive adjuvant chemotherapy or tamoxifen.

**Material III.** A human breast cancer progression array was constructed using cutting edge matrix assembly (CEMA) technology (LeBaron, Crismon *et al.* 2005) from deidentified pathology blocks from the Thomas Jefferson University Hospital archive dating to 1970-72 and contained a total of 180 patient specimens, including 40 unmatched normal breast tissues, 20 DCIS, 20 invasive ductal carcinomas, histological grade I, 40 invasive ductal carcinomas, histological grade II, 40 invasive ductal carcinomas,

histological grade III, and 20 lymph node breast cancer metastases. Invasive ductal carcinomas were graded by pathologist (A.W.). Hormone receptor status of the invasive ductal carcinomas was determined by standard IHC and was identified as 20% Her2-positive, 59% ER-positive, and 42% PR-positive. For the Nuc-pYStat5 marker, 130 tissue spots provided valid AQUA scores.

Material IV. A tissue microarray containing 0.6-mm tumor cores from breast cancer patients with corresponding clinical and breast cancer-specific survival data was obtained from the Institute of Pathology at the Kantonsspital Basel (Basel, Switzerland). Nuc-pYStat5 was shown to be a favorable marker of prognosis in this material, described as Material C (Nevalainen, Xie et al. 2004), but no survival analysis related to anti-estrogen adjuvant therapy was presented. Of the patients represented on the array, a cohort of 166 patients who received anti-estrogen therapy and had complete clinical follow-up and Nuc-pYStat5 data is the focus of this analysis.

Material V. The Tamoxifen 50/50 tissue microarray was constructed from a random series of breast cancer patients identified through the Alberta Cancer Registry (Calgary, Alberta, Canada). Patients were diagnosed from 1987-2000 and treated with up to 60 months (average 30 months) of tamoxifen in the adjuvant setting. Medical histories were reviewed to verify treatment information and clinical characteristics, including age at diagnosis, tumor type, tumor size, tumor grade, hormone receptor status, and lymph node status, and sample sections from tumor block were reviewed. One hundred patients were selected for the array, including 50 patients who died of breast cancer and were selected as "resistant to tamoxifen", as well as 50 patients with greater than 5-years follow-up without breast cancer recurrence and classified as "good response". Three 0.6mm punches from each tumor specimen were placed in the array. The majority of patients were ER-positive but a subset of ER-negative patients was represented on the array as tamoxifen was used in ER-negative patients during this time interval. Analysis of 65 spots yielded informative data on Nuc-pYStat5.

*Detection of nuclear localized and tyrosine phosphorylated Stat5 (Nuc-pYStat5) by DAB-chromogen immunohistochemistry (Materials I and IV).* Mouse monoclonal anti-phosphoStat5 antibody AX1 was provided by Advantex BioReagents (El Paso, TX), and its specificity for the phosphorylated conserved tyrosine residue Y694/Y699 of Stat5a/b in immunohistochemistry and immunoblotting has been previously validated by peptide competition, inducible phosphorylation studies, site-directed mutagenesis, and Stat5 knock-out model analyses (Nevalainen, Xie et al. 2002). Tissue array sections containing paraffin-embedded, formalin-fixed tissues from malignant human breast were deparaffinized in xylene,

rehydrated in graded ethanol, and heated in a pressure-cooker with pH 10 antigen retrieval solution (AR10; Biogenex, San Ramon, CA). Following antigen retrieval, endogenous peroxidase activity was blocked by incubating slides in 0.3% H<sub>2</sub>O<sub>2</sub> for 10 min at room temperature, and non-specific binding of immunoglobulin was minimized by preincubation in normal goat serum for 2 h at room temperature. For staining of Material IV, AX1 was diluted in 1% bovine serum albumin in phosphate buffered saline and incubated the tissue array section at a final concentration of 0.6 µg/ml for 16 h as recommended by the provider. Staining of Material I was performed on a Dako Autostainer in 10 batches of 40 slides and incubation time for AX1 was shortened to 45 min at a final concentration of 1.2 µg/ml. Antigen-antibody complexes were detected using biotinylated goat anti-mouse secondary antibody (Biogenex) followed by streptavidin-horseradish-peroxidase complex, using 3,3-(prime) DAB as chromogen and Mayer hematoxylin as counterstain. Individual breast tumor samples were scored by a pathologist blinded to clinical outcome for levels of transcriptionally active Stat5 as defined by levels of nuclear localized and tyrosine phosphorylated Stat5 by estimating percent of malignant cells with detectable staining and overall staining intensity on a scale ranging from 0-3.

***Quantification of Active Stat5 by Immunofluorescence (Materials II, III, and V).*** Rabbit monoclonal anti-phosphoStat5 antibody, E208, (clone E208, Cat#1208-1, Epitomics, Burlingame, CA) was used for automated quantitative analysis (AQUA). This antibody is generated to the same phospho-antigen (Y694/Y699 Stat5a/b) as the mouse monoclonal AX1 antibody and showed similar specificity in side-by-side DAB-chromogen IHC testing (see results). Antigen retrieval was performed using the DAKO PT-module with citric acid buffer, pH 6.0, and primary antibody was diluted 1:1,600 from the supplier-provided stock and incubated for 30 min. Co-detection of cytokeratin for tumor-masking was achieved using mouse monoclonal anti-pancytokeratin (Clone AE1/AE3, Cat#M3515, DAKO, Carpinteria, CA) at 1:100 dilution and co-incubated with the E208 antibody. Staining of the tissue array slides representing Materials II, III, and V was performed on a Dako Autostainer. After washing off unbound primary antibodies, HRP-conjugated anti-rabbit IgG antibody and Alexa-488-conjugated antimouse IgG antibody were added for 30 min, followed by washing and incubation with Cy5-tyramide (Cat#NEL745B001KT, Perkin Elmer, Waltham, MA) for 10 min. Following further washing, slides were coverslipped using DAPI-containing mounting media.

***Ex vivo human breast tissue.*** Control experiments compared rabbit monoclonal E208 with mouse monoclonal AX1 using DAB-based IHC. For these studies, freshly isolated and viable surgical human breast tissues were obtained from Thomas Jefferson University Hospital. Explants of healthy mammary

tissue were cultured *ex vivo* at 37°C in RPMI media containing 10% FBS and 1mM sodium pyruvate and incubated with or with 100 nM human prolactin for 1 h to induce phosphorylation of Stat5. Tissue explants were formalin-fixed and embedded in paraffin for sectioning, and stained under optimal conditions for each antibody. DAB-chromogen IHC was performed as described above.

**Automated Quantitative Analysis (AQUA).** The AQUA/PM2000 platform (HistoRx) (Camp, Chung *et al.* 2002) was used to quantify fluorescence-based immunostaining for Nuc-pYStat5 on Materials II, III and V. The tissue array slides were automatically scanned and fluorescent images were captured in three channels, FITC/Alexa-488 (cytokeratin), Cy5 (Nuc-pYStat5) or DAPI (nuclei). AQUA scores were blinded to clinical data and objectively generated by calculating the mean logarithmic signal intensities within the cell nuclei of the epithelial compartment only, defined by cytokeratin-positive and DAPI-positive mapping.

**X-Tile Cutpoint Analysis and Statistical Methods.** Analytic survival-based cutpoints for Nuc-pYStat5 in Materials II and V were derived using the X-tile software (Camp, Dolled-Filhart *et al.* 2004). The X-tile software selects optimal cutpoints by calculating and plotting log-rank  $X^2$  statistics at for each possible division, then dividing into groups that are optimal indicators of risk to correct for a high false-positive rate associated with making multiple comparisons when finding the minimum *P* value using the log-rank test. Cutpoint selection was validated using automatically generated training and validation sets.

Survival analyses (Materials I, II, IV, V) were performed by constructing Kaplan-Meier curves and adjusted Cox proportional hazards models. Endpoints for analysis were breast cancer recurrence (Material I) and breast cancer-specific death (Materials II, IV, V). Hazard ratios were estimated using both univariate and multivariate Cox proportional hazards regression models. Proportional hazards assumptions were assessed graphically using a log-log survival plot and statistically with Schoenfeld residuals. When available, variables included in the adjusted Cox models were Nuc-pYStat5 status, patient race, age at diagnosis, tumor grade, tumor size, estrogen receptor status, progesterone receptor status, and Her2 receptor status. Material I data analyses were stratified by treatment location to control for variability in treatment and outcome between centers. Survival analyses of Materials I, II, and IV were performed using Stata/SE version 11.0 (StataCorp, College Station, Texas) and SPSS version 16.0 (SPSS Inc, Chicago, IL) for Material V. One-way ANOVA (SPSS Inc, Chicago, IL) was used to test for differences between across breast histology groups in progression array of Material III, followed by

Ryder *et al.* Stat5 and response to anti-estrogen therapy.

Dunnett's T3 pairwise post-hoc test for unequal variances to identify differences between individual breast histology groups.

## Results

*Loss of nuclear localized and tyrosine phosphorylated Stat5 as reviewed by pathologist on whole tissue sections predicts breast cancer recurrence.* To determine whether levels of nuclear localized and tyrosine phosphorylated (Nuc-pYStat5) would predict outcome in lymph node-negative breast cancer patients, which we previously had reported solely based on tissue microarray analyses (Nevalainen, Xie et al. 2004), we examined the more clinically relevant setting of whole tissue sections. Material I represented 301 primary invasive ductal carcinomas from patients with node-negative disease and for which data on recurrence was available. Immunohistochemistry using the same anti-pYStat5 antibody, AX1, and DAB-chromogen detection as in the original tissue microarray study (Nevalainen, Xie et al. 2004) was applied to Material I. To limit inter-slide variability in staining, the slides were stained with an automated staining system (DAKO). Levels of Nuc-pYStat5 were assessed, blinded to clinical outcome, by an independent pathologist (A.W.) based on nuclear staining intensity (scored 0-3) and percent nuclei with positive staining within the tumor. For survival analyses, a cutpoint was used that defined positive Nuc-pYStat5 status when staining intensity was greater than 0 in more than ten percent of the carcinoma cells. Univariate analysis indicated that low levels of Nuc-pYStat5 were associated with a 2.9-fold increased risk of breast cancer recurrence (95% CI 1.4-6.0;  $p=0.005$ ; univariate Cox proportional hazards regression, **Figure 1; Table 2A**).

*Objective and quantitative immunofluorescence analysis of nuclear localized and tyrosine phosphorylated Stat5 and association with breast cancer survival.* The mouse monoclonal AX1 antibody used to detect the conserved phosphorylated tyrosine residue Y694/Y699 of Stat5a/b gives best results after over-night incubation at 4°C and specificity has previously been carefully validated in immunohistochemistry (Nevalainen, Xie et al. 2002; Nevalainen, Xie et al. 2004). Optimal sensitivity with AX1 is not achieved with the 30-45 min incubation time required for autostainers. To improve detection of active Stat5 on an autostainer platform and to extend the analysis to a second antibody, the rabbit monoclonal anti-phosphoStat5 antibody, E208 (Epitomics), was tested in immunostaining of formalin fixed, paraffin embedded tissues. To compare anti-phosphoStat5 antibodies, we stained human breast tissue explants that had been incubated *ex vivo* with or without prolactin for 1h, which provided relevant negative and positive control tissues, respectively (**Figure 2A**). While the two antibodies provided very similar results under their respective optimized staining conditions, E208 was chosen for the quantitative analyses based on its ability to stain effectively within the 30 min incubation time required by autostainers.

E208 was used for fluorescence-based detection of Nuc-pYStat5 in an independent Material II, which included 158 primary breast carcinomas from lymph node-negative patients represented in tissue microarray format. Extended follow-up times up to 38 years were available for patients represented in Material II. Automated Quantitative Analysis (AQUA) was used to objectively quantify levels of fluorescently labeled tyrosine phosphorylated Stat5 within the nuclei of cytokeratin-positive cancer cells and X-Tile software was used to generate a cutpoint for positive and negative cases (Camp, Dolled-Filhart *et al.* 2004). AQUA quantification of Nuc-pYStat5 predicted breast cancer survival in Material II, and was highly consistent with DAB-chromogen and pathologist scoring of Material I (**Figure 2B**). Univariate analysis revealed that patients expressing low levels of Nuc-pYStat5 had a 1.9 fold-increased risk of breast cancer-related death compared to patients expressing high levels of Nuc-pYStat5 (95% CI 1.1-3.2;  $p=0.016$ ; univariate Cox proportional hazards regression, **Figure 2B**; **Table 2B**). Thus, loss of Nuc-pYStat5 was associated with poor clinical outcome by univariate analysis of two independent materials and by independent analytical methods.

*Nuclear localized and tyrosine phosphorylated Stat5 is an independent marker of node-negative breast cancer prognosis when adjusted for standard pathology and clinical markers.* Importantly, in multivariate Cox's proportional hazard regression analyses, Nuc-pYStat5 was an independent marker of disease prognosis in both Materials I and II, when adjusting for standard clinical and pathological markers, including age at diagnosis, race, grade, tumor size, and hormone receptor status. In Material I, patients with low Nuc-pYStat5 expression, as detected by DAB-chromogen staining using AX1 mouse monoclonal antibody on whole tissue sections, had a 2.5-fold greater risk of developing recurrent disease after adjusting for other factors (95% CI 1.2-5.3;  $p=0.019$ ; multivariate Cox proportional hazards regression, **Table 2A**). Likewise, when quantified using E208 rabbit monoclonal antibody and fluorescence-based AQUA, Nuc-pYStat5 remained an independent marker of breast cancer specific survival in Material II, as revealed by an 1.8-fold increased risk of dying from breast cancer (95% CI 1.03-3.3;  $p=0.039$ ; multivariate Cox proportional hazards regression, **Table 2B**). We conclude that Nuc-pYStat5 is an independent marker of outcome in patients with lymph node-negative breast cancer.

*Levels of nuclear localized and tyrosine phosphorylated Stat5 are diminished during breast cancer progression.* Previous reports are contradictory with regard to levels of Stat5 activation in human invasive breast cancer compared to normal breast epithelia. Elevated levels of Stat5 activity in breast cancer were first reported based on nuclear extracts and electrophoretic mobility shift assay (EMSA)

(Watson and Miller 1995). Cotarla and colleagues reported frequent activation of Stat5 in invasive breast cancer, detecting nuclear localized, tyrosine phosphorylated Stat5 by IHC in 78% of breast cancer specimens examined (Cotarla, Ren et al. 2004). In contrast, a significant loss of Nuc-pYStat5 detected by IHC in invasive breast cancer and in metastases has been suggested by our group and others (Nevalainen, Xie et al. 2004; Bratthauer, Strauss et al. 2006). However, none of the previous analytical approaches were quantitative, and we therefore used AQUA to quantify Nuc-pYStat5 levels in a breast tissue progression material, Material III. This progression material was represented in tissue array format and included normal breast tissue, ductal carcinoma in situ (DCIS), invasive breast cancer (IDC), and breast cancer lymph node metastases. Levels of Nuc-pYStat5 measured by immunofluorescence and quantitative AQUA analysis were significantly different between groups (ANOVA  $F=22.87$ ,  $p\leq 0.0001$ ) and were markedly reduced over breast cancer progression (**Figure 3**). Specifically, while levels of Nuc-pYStat5 were high and remained unchanged between normal and DCIS, Nuc-pYStat5 was significantly reduced in IDC compared to normal tissue (Dunnett's T3 post-hoc test  $p<0.0001$ ) with the greatest loss of Nuc-pYStat5 observed in lymph node metastases (Dunnett's T3 post-hoc test  $p<0.0001$ ; **Figure 3**). These quantitative data provide novel insight supporting the notion of progressive loss of Nuc-pYStat5 during breast cancer progression.

*Nuclear localized and tyrosine phosphorylated Stat5 predicts responsiveness to anti-estrogen therapy.*

Previous studies have indicated cross-talk between Stat5 and estrogen receptor (ER) (Bjornstrom, Kilic et al. 2001; Faulds, Pettersson et al. 2001). A relationship has been proposed between Stat5 expression levels and survival time in patients who received anti-estrogen therapy following relapse based on a limited dataset of 70 patients (Yamashita, Nishio et al. 2006). However, the study used a Stat5b-specific antibody and did not specifically detect the presumably transcriptionally active, nuclear localized and tyrosine phosphorylated form of Stat5, but instead reported total cellular Stat5b. To specifically address the predictive value of Nuc-pYStat5 in adjuvant anti-estrogen treatment of patients with breast cancer, we initially examined levels of Nuc-pYStat5 expression using DAB-chromogen staining with AX1 and standard pathologist scoring on a cohort of 166 patients, Material IV. Loss of Nuc-pYStat5 expression was a significant predictive marker of increased risk of failure of anti-estrogen therapy in Material IV (HR 1.8;  $p=0.043$ ; univariate Cox regression analysis; **Figure 4A, Table 3A**).

The analysis was extended to quantify Nuc-pYStat5 expression by AQUA analysis using the E208 antibody in an independent tissue microarray material, Material V, which comprised 65 patients treated with tamoxifen. X-tile determined a cutpoint for low and high Nuc-pYStat5 that revealed a cohort of

patients whose tumors had low levels of Nuc-pYStat5 and were at markedly increased risk of failing tamoxifen treatment and dying from breast cancer. In fact, a hazard ratio of 4.8 (95% CI 2.1-11.2;  $p < 0.001$ , univariate Cox regression) was detected (**Figure 4B**, **Table 3B**). In multivariate analysis adjusting for age at diagnosis, tumor size, tumor grade, node status, ER/PR status, and Her2 overexpression, Nuc-pYStat5 remained an independent marker of breast cancer survival in Material V (Cox regression HR=7.9, 95% CI 2.5-25.0,  $p < 0.001$ , **Table 3B**). Positive node status (Cox regression HR=4.4, 95% CI 1.6-12.0,  $p = 0.003$ ) and expression of ER/PR (Cox regression HR=0.23, 95% CI 0.06-0.87,  $p = 0.03$ ) were also independent markers of survival, while Her2 status, tumor grade, and age were not (**Table 3B**). These observations suggest that low levels of Nuc-pYStat5 are an independent marker of increased risk of failure of tamoxifen treatment.

Materials IV and V analyzed above included both lymph node-negative and lymph node-positive disease. Based on previous observations that Nuc-pYStat5 was particularly effective at predicting prognosis in node-negative patients (Nevalainen, Xie *et al.* 2004) and current analysis of Materials I and II, the predictive value of Nuc-pYStat5 in anti-estrogen therapy was also explored in the subgroups of patients with node-negative disease of Materials IV and V (**Figure 4C-D**). Despite limited sample sizes, loss of Nuc-pYStat5 predicted increased risk of death from breast cancer in node-negative patients both by DAB-chromogen IHC in Material IV (Log Rank  $p = 0.035$ ; **Figure 4C**) and by fluorescence-based AQUA of Material V (Log Rank  $p < 0.001$ ; **Figure 4C**). Collectively, studies of two clinical materials of patients treated with anti-estrogen in the adjuvant setting, suggest that low levels of Nuc-pYStat5 predicts elevated risk of therapy failure.

## Discussion

The present study documents that low levels of nuclear localized and tyrosine phosphorylated Stat5 (Nuc-pYStat5) represent an independent marker of poor breast cancer prognosis in patients with lymph node-negative breast cancer, based on examination of two independent clinical materials analyzed by traditional pathologist scoring of whole tissue sections (Material I) and by quantitative AQUA analysis (Material II). Consistent outcomes were found using two distinct monoclonal anti-phosphotyrosine-Stat5 antibodies. The data represents the first evidence from analysis of Nuc-pYStat5 in whole tumor tissue sections, and provides novel support for the previously proposed prognostic value of Nuc-pYStat5 in breast cancer based on tissue array specimens (Nevalainen, Xie et al. 2004). Furthermore, quantitative immunofluorescence-based AQUA analysis of a breast cancer progression material revealed markedly lower levels of Nuc-pYStat5 in invasive breast cancer and metastases compared to normal breast epithelia or DCIS. Most importantly, analysis of a fourth and fifth breast cancer material representing patients who received anti-estrogen adjuvant therapy suggested that levels of Nuc-pYStat5 constitutes a new independent predictive marker of response to anti-estrogen adjuvant therapy.

Loss of Nuc-pYStat5 may become a clinical tool useful for predicting increased risk of failure of anti-estrogen therapy. The predictive value of Nuc-pYStat5 was particularly pronounced in quantitative analysis of Material V. Loss of Nuc-pYStat5 in this material as quantified by AQUA was an independent predictor of failure to respond to tamoxifen therapy when adjusting for age at diagnosis, tumor grade, lymph node status, Her2 overexpression, and ER/PR status. Quantitative AQUA revealed a population of patients whose tumors had particularly low levels of Nuc-pYStat5 using the objective X-Tiles software to determine the optimal cutpoint. Notably, 15% of patients whose tumors displayed the lowest levels of Nuc-pYStat5 had particularly elevated risk of failing to respond to adjuvant tamoxifen therapy. The predictive value of Nuc-pYStat5 in Material IV, as measured by DAB-chromogen and pathologist scoring was not as pronounced as that detected by AQUA quantification in Material V. DAB-chromogen does not provide the same dynamic range of signal compared to immunofluorescence (Camp, Chung et al. 2002). Consistent with this, as many as 56% of cases in Material IV were negative for Nuc-pYStat5 when detected by DAB-chromogen IHC. In contrast to node-negative Materials I and II, Materials IV and V included node-negative and node-positive cases. Reduced levels of Nuc-pYStat5 also predicted resistance to anti-estrogen therapy when node-negative breast cancer patients of Materials IV and V were analyzed separately. The increased efficacy of Nuc-pYStat5 to predict response to anti-estrogen treatment in Material V compared to Material IV might be related to sensitivity of Nuc-pYStat5 detection by

AQUA, but may also reflect differences between materials such as year of diagnosis and cotreatment. However, the number of events within the node-negative subgroups of breast cancer patients in Materials IV and V were low and follow-up data in larger materials will be needed for more conclusive interpretation in node-negative patients.

Anti-estrogen therapy of breast cancer patients is currently guided by positive tumor expression of estrogen receptor-alpha (ER) by immunohistochemistry. However, approximately 30% of patients with ER-positive breast cancer fail to respond to anti-estrogen therapy due to inherent or acquired resistance (Osborne 1998; Clarke, Leonessa *et al.* 2001; Schiff, Massarweh *et al.* 2004; Prat and Baselga 2008). AQUA quantification of levels of epidermal growth factor receptor (EGFR) in 564 patients enrolled in a randomized clinical trial for adjuvant tamoxifen therapy indicated that high levels of EGFR predicted resistance to tamoxifen (Giltane, Ryden *et al.* 2007). Several reports have also indicated that low levels of p27 expression predict failure to respond to hormone therapy (Pohl, Ruda *et al.* 2003; McCallum, Baker *et al.* 2004; Porter, Barlow *et al.* 2006; Filipits, Ruda *et al.* 2009). A meta-analysis of 12 studies implicated Her2 as a modest predictor of resistance to anti-estrogen therapy (De Laurentiis, Arpino *et al.* 2005). This has been supported by experimental studies indicating that Her2 overexpression may lead to loss of ER expression or ER-independent growth and tamoxifen resistance of breast cancer cells (Benz, Scott *et al.* 1992; Liu, el-Ashry *et al.* 1995; Kurokawa, Lenferink *et al.* 2000; Osborne, Bardou *et al.* 2003; Shou, Massarweh *et al.* 2004). However, current American Society of Clinical Oncology guidelines for the use of tumor markers in breast cancer concluded there was insufficient data to include Her2 in the current clinical guidelines as a predictor of response to endocrine therapy (Harris, Fritsche *et al.* 2007). In our multivariate analyses, Her2 status had only borderline significant value in predicting outcome in the anti-estrogen treated patients of Materials IV and V. It will be of interest to expand the analyses of Nuc-pYStat5 to include p27 and EGFR in future work.

Stat5 and ER are both hormone-activated transcription factors critical for normal development of breast epithelia. Positive interactions between Stat5 and ER have been previously demonstrated based on data from breast cancer and other cell lines by several groups (Bjornstrom, Kilic *et al.* 2001; Yamashita, Iwase *et al.* 2003; Wang and Cheng 2004), however negative interaction between Stat5 and ER have also been reported (Stoecklin, Wissler *et al.* 1999; Faulds, Pettersson *et al.* 2001). Stat5 and ER have also been shown to directly interact (Bjornstrom, Kilic *et al.* 2001; Wang and Cheng 2004). Furthermore, ER-alpha was reported to be a direct target gene of Stat5 and Stat5 upregulated ER-alpha expression in rat ovaries (Frasor, Barkai *et al.* 2001; Frasor, Park *et al.* 2001). Initial data has indicated that Stat5 activation could

restore ER expression in the ER-negative BT20 human breast cancer cell line (Sultan, Brim et al. 2008). Because the ER gene is frequently silenced in breast cancer, a positive role of Stat5 on ER expression levels would constitute one possible biological mechanism to explain the observed effect of Nuc-pYStat5 as a predictive marker of response to anti-estrogen therapy. Alternatively, Stat5 may cooperate with ER in transcriptional regulation of a subset of common target genes.

There are two highly homologous human Stat5 genes, Stat5a and Stat5b, which encode proteins that share greater than 90% amino acid identity (Liu, Robinson et al. 1995). Because the conserved phosphotyrosine motifs of Stat5a and Stat5b are identical, anti-phosphoStat5 antibodies, including AX1 and E208 used in this study, do not distinguish between phosphorylated Stat5a and Stat5b. Both Stat5a and Stat5b are activated by prolactin receptors in human breast cancer cell lines (Neilson, Zhu et al. 2007) and the two proteins have partially overlapping but distinct effects and regulation (Grimley, Dong et al. 1999). For instance, we have recently reported that in breast cancer cells, prolactin-activated Stat5a but not prolactin-activated Stat5b, repressed the BCL6 oncogene (Tran, Utama et al.). Further work is needed to determine the individual contribution of Stat5a and Stat5b to the observed prognostic and predictive values of Nuc-pYStat5. Yamashita and colleagues reported increased risk of tamoxifen treatment failure associated with loss of Stat5b protein expression in patients with breast cancer relapse (Yamashita, Nishio et al. 2006). However, the report did not differentiate nuclear from cytoplasmic Stat5b protein and therefore it is unclear if the protein was transcriptionally active. In fact only 10% of the tumors expressing Stat5b protein had nuclear localization of Stat5b

Collectively, the present work supports the notion of frequent loss of Stat5 signaling in progressive breast cancer, and that loss of Nuc-pYStat5 is associated with decreased survival and elevated risk of failure on anti-estrogen therapy. The present data supports the initiation of prospective studies on the utility of Nuc-pYStat5 as a prognostic and predictive marker in breast cancer.

**Acknowledgements.** We thank Eva Andersson for expert technical help with immunostaining of sections of Material I and Melanie Gironde for editorial assistance. We thank Markus Zuber, Department of Surgery, University of Basel, Basel, Switzerland, Ossi R. Köchli, Universitäts-Frauenklinik, University of Basel, Basel, Switzerland, Frank Mross, Kreiskrankenhaus, Lörrach, Germany, and Holger Dieterich, Frauenklinik, Rheinfelden, Germany, for extensive support with clinical data analysis.

**Conflicts of interest.** The following authors or their immediate family members have indicated a financial interest. Consultant and ownership interest: Hallgeir Rui, Advantex BioReagents (El Paso, Texas); Marja T. Nevalainen, Advantex BioReagents; David L. Rimm, HistoRx (New Haven, CT). Advantex BioReagents provided material support for this study and holds intellectual property rights on the primary observations.

## References

- Benz, C. C., G. K. Scott, et al. (1992). "Estrogen-dependent, tamoxifen-resistant tumorigenic growth of MCF-7 cells transfected with HER2/neu." *Breast Cancer Res Treat* **24**(2): 85-95.
- Bjornstrom, L., E. Kilic, et al. (2001). "Cross-talk between Stat5b and estrogen receptor-alpha and -beta in mammary epithelial cells." *J Mol Endocrinol* **27**(1): 93-106.
- Bratthauer, G. L., B. L. Strauss, et al. (2006). "STAT 5a expression in various lesions of the breast." *Virchows Arch* **448**(2): 165-71.
- Camp, R. L., G. G. Chung, et al. (2002). "Automated subcellular localization and quantification of protein expression in tissue microarrays." *Nat Med* **8**(11): 1323-7.
- Camp, R. L., M. Dolled-Filhart, et al. (2004). "X-tile: a new bio-informatics tool for biomarker assessment and outcome-based cut-point optimization." *Clin Cancer Res* **10**(21): 7252-9.
- Clarke, R., F. Leonessa, et al. (2001). "Cellular and molecular pharmacology of antiestrogen action and resistance." *Pharmacol Rev* **53**(1): 25-71.
- Cotarla, I., S. Ren, et al. (2004). "Stat5a is tyrosine phosphorylated and nuclear localized in a high proportion of human breast cancers." *Int J Cancer* **108**(5): 665-71.
- De Laurentiis, M., G. Arpino, et al. (2005). "A meta-analysis on the interaction between HER-2 expression and response to endocrine treatment in advanced breast cancer." *Clin Cancer Res* **11**(13): 4741-8.
- Dolled-Filhart, M., L. Ryden, et al. (2006). "Classification of breast cancer using genetic algorithms and tissue microarrays." *Clin Cancer Res* **12**(21): 6459-68.
- Faulds, M. H., K. Pettersson, et al. (2001). "Cross-talk between ERs and signal transducer and activator of transcription 5 is E2 dependent and involves two functionally separate mechanisms." *Mol Endocrinol* **15**(11): 1929-40.
- Filipits, M., M. Rudas, et al. (2009). "Low p27 expression predicts early relapse and death in postmenopausal hormone receptor-positive breast cancer patients receiving adjuvant tamoxifen therapy." *Clin Cancer Res* **15**(18): 5888-94.
- Frasor, J., U. Barkai, et al. (2001). "PRL-induced ERalpha gene expression is mediated by Janus kinase 2 (Jak2) while signal transducer and activator of transcription 5b (Stat5b) phosphorylation involves Jak2 and a second tyrosine kinase." *Mol Endocrinol* **15**(11): 1941-52.
- Frasor, J., K. Park, et al. (2001). "Differential roles for signal transducers and activators of transcription 5a and 5b in PRL stimulation of ERalpha and ERbeta transcription." *Mol Endocrinol* **15**(12): 2172-81.
- Giltane, J. M., C. B. Moeder, et al. (2009). "Quantitative multiplexed analysis of ErbB family coexpression for primary breast cancer prognosis in a large retrospective cohort." *Cancer* **115**(11): 2400-9.
- Giltane, J. M., L. Ryden, et al. (2007). "Quantitative measurement of epidermal growth factor receptor is a negative predictive factor for tamoxifen response in hormone receptor positive premenopausal breast cancer." *J Clin Oncol* **25**(21): 3007-14.
- Gouilleux, F., H. Wakao, et al. (1994). "Prolactin induces phosphorylation of Tyr694 of Stat5 (MGF), a prerequisite for DNA binding and induction of transcription." *Embo J* **13**(18): 4361-9.
- Grimley, P. M., F. Dong, et al. (1999). "Stat5a and Stat5b: fraternal twins of signal transduction and transcriptional activation." *Cytokine Growth Factor Rev* **10**(2): 131-57.
- Harris, L., H. Fritsche, et al. (2007). "American Society of Clinical Oncology 2007 update of recommendations for the use of tumor markers in breast cancer." *J Clin Oncol* **25**(33): 5287-312.
- Humphreys, R. C. and L. Hennighausen (1999). "Signal transducer and activator of transcription 5a influences mammary epithelial cell survival and tumorigenesis." *Cell Growth Differ* **10**(10): 685-94.

- Iavnilovitch, E., B. Groner, et al. (2002). "Overexpression and forced activation of stat5 in mammary gland of transgenic mice promotes cellular proliferation, enhances differentiation, and delays postlactational apoptosis." Mol Cancer Res **1**(1): 32-47.
- Kang, J. Y., M. Dolled-Filhart, et al. (2003). "Tissue microarray analysis of hepatocyte growth factor/Met pathway components reveals a role for Met, matriptase, and hepatocyte growth factor activator inhibitor 1 in the progression of node-negative breast cancer." Cancer Res **63**(5): 1101-5.
- Kurokawa, H., A. E. Lenferink, et al. (2000). "Inhibition of HER2/neu (erbB-2) and mitogen-activated protein kinases enhances tamoxifen action against HER2-overexpressing, tamoxifen-resistant breast cancer cells." Cancer Res **60**(20): 5887-94.
- LeBaron, M. J., H. R. Crismon, et al. (2005). "Ultrahigh density microarrays of solid samples." Nat Methods **2**(7): 511-3.
- Liu, X., G. W. Robinson, et al. (1995). "Cloning and expression of Stat5 and an additional homologue (Stat5b) involved in prolactin signal transduction in mouse mammary tissue." Proc Natl Acad Sci U S A **92**(19): 8831-5.
- Liu, X., G. W. Robinson, et al. (1997). "Stat5a is mandatory for adult mammary gland development and lactogenesis." Genes Dev **11**(2): 179-86.
- Liu, Y., D. el-Ashry, et al. (1995). "MCF-7 breast cancer cells overexpressing transfected c-erbB-2 have an in vitro growth advantage in estrogen-depleted conditions and reduced estrogen-dependence and tamoxifen-sensitivity in vivo." Breast Cancer Res Treat **34**(2): 97-117.
- McCabe, A., M. Dolled-Filhart, et al. (2005). "Automated quantitative analysis (AQUA) of in situ protein expression, antibody concentration, and prognosis." J Natl Cancer Inst **97**(24): 1808-15.
- McCallum, M., C. Baker, et al. (2004). "A prognostic index for operable, node-negative breast cancer." Br J Cancer **90**(10): 1933-41.
- Neilson, L. M., J. Zhu, et al. (2007). "Coactivation of janus tyrosine kinase (Jak)1 positively modulates prolactin-Jak2 signaling in breast cancer: recruitment of ERK and signal transducer and activator of transcription (Stat)3 and enhancement of Akt and Stat5a/b pathways." Mol Endocrinol **21**(9): 2218-32.
- Nevalainen, M. T., J. Xie, et al. (2002). "Basal activation of transcription factor signal transducer and activator of transcription (Stat5) in nonpregnant mouse and human breast epithelium." Mol Endocrinol **16**(5): 1108-24.
- Nevalainen, M. T., J. Xie, et al. (2004). "Signal transducer and activator of transcription-5 activation and breast cancer prognosis." J Clin Oncol **22**(11): 2053-60.
- Nouhi, Z., N. Chughtai, et al. (2006). "Defining the role of prolactin as an invasion suppressor hormone in breast cancer cells." Cancer Res **66**(3): 1824-32.
- Osborne, C. K. (1998). "Tamoxifen in the treatment of breast cancer." N Engl J Med **339**(22): 1609-18.
- Osborne, C. K., V. Bardou, et al. (2003). "Role of the estrogen receptor coactivator AIB1 (SRC-3) and HER-2/neu in tamoxifen resistance in breast cancer." J Natl Cancer Inst **95**(5): 353-61.
- Pohl, G., M. Rudas, et al. (2003). "High p27Kip1 expression predicts superior relapse-free and overall survival for premenopausal women with early-stage breast cancer receiving adjuvant treatment with tamoxifen plus goserelin." J Clin Oncol **21**(19): 3594-600.
- Porter, P. L., W. E. Barlow, et al. (2006). "p27(Kip1) and cyclin E expression and breast cancer survival after treatment with adjuvant chemotherapy." J Natl Cancer Inst **98**(23): 1723-31.
- Prat, A. and J. Baselga (2008). "The role of hormonal therapy in the management of hormonal-receptor-positive breast cancer with co-expression of HER2." Nat Clin Pract Oncol **5**(9): 531-42.
- Ren, S., H. R. Cai, et al. (2002). "Loss of Stat5a delays mammary cancer progression in a mouse model." Oncogene **21**(27): 4335-9.
- Schiff, R., S. A. Massarweh, et al. (2004). "Cross-talk between estrogen receptor and growth factor pathways as a molecular target for overcoming endocrine resistance." Clin Cancer Res **10**(1 Pt 2): 331S-6S.

- Shou, J., S. Massarweh, et al. (2004). "Mechanisms of tamoxifen resistance: increased estrogen receptor-HER2/neu cross-talk in ER/HER2-positive breast cancer." *J Natl Cancer Inst* **96**(12): 926-35.
- Stoecklin, E., M. Wissler, et al. (1999). "Interactions in the transcriptional regulation exerted by Stat5 and by members of the steroid hormone receptor family." *J Steroid Biochem Mol Biol* **69**(1-6): 195-204.
- Sultan, A. S., H. Brim, et al. (2008). "Co-overexpression of Janus kinase 2 and signal transducer and activator of transcription 5a promotes differentiation of mammary cancer cells through reversal of epithelial-mesenchymal transition." *Cancer Sci* **99**(2): 272-9.
- Sultan, A. S., J. Xie, et al. (2005). "Stat5 promotes homotypic adhesion and inhibits invasive characteristics of human breast cancer cells." *Oncogene* **24**(5): 746-60.
- Teglund, S., C. McKay, et al. (1998). "Stat5a and Stat5b proteins have essential and nonessential, or redundant, roles in cytokine responses." *Cell* **93**(5): 841-50.
- Tran, T. H., F. E. Utama, et al. "Prolactin Inhibits BCL6 Expression in Breast Cancer through a Stat5a-Dependent Mechanism." *Cancer Res.*
- Wagner, K. U., A. Krempler, et al. (2004). "Impaired alveologenesis and maintenance of secretory mammary epithelial cells in Jak2 conditional knockout mice." *Mol Cell Biol* **24**(12): 5510-20.
- Wagner, K. U. and H. Rui (2008). "Jak2/Stat5 signaling in mammaryogenesis, breast cancer initiation and progression." *J Mammary Gland Biol Neoplasia* **13**(1): 93-103.
- Wang, Y. and C. H. Cheng (2004). "ERalpha and STAT5a cross-talk: interaction through C-terminal portions of the proteins decreases STAT5a phosphorylation, nuclear translocation and DNA-binding." *FEBS Lett* **572**(1-3): 238-44.
- Watson, C. J. and W. R. Miller (1995). "Elevated levels of members of the STAT family of transcription factors in breast carcinoma nuclear extracts." *Br J Cancer* **71**(4): 840-4.
- Xie, J., M. J. LeBaron, et al. (2002). "Role of tyrosine kinase Jak2 in prolactin-induced differentiation and growth of mammary epithelial cells." *J Biol Chem* **277**(16): 14020-30.
- Yamashita, H., H. Iwase, et al. (2003). "Naturally occurring dominant-negative Stat5 suppresses transcriptional activity of estrogen receptors and induces apoptosis in T47D breast cancer cells." *Oncogene* **22**(11): 1638-52.
- Yamashita, H., M. Nishio, et al. (2006). "Stat5 expression predicts response to endocrine therapy and improves survival in estrogen receptor-positive breast cancer." *Endocrine-Related Cancer.*

### Figure Legends

**Figure 1.** Nuc-pYStat5 and risk of breast cancer recurrence in whole tissue sections of node-negative breast cancer. Risk of breast cancer recurrence was significantly elevated in patients whose tumors expressed low Nuc-pYStat5 as detected by mouse monoclonal anti-pYStat5 AX1 antibody and standard DAB IHC and pathologist review. Kaplan-Meier analysis of Disease Free Survival in Material I according to high or low Nuc-pYStat5 expression. Censored cases (+) and number of patients per group are indicated.

**Figure 2.** Validation of specificity of rabbit monoclonal anti-pYStat5 antibody E208 and immunofluorescence-based quantitative AQUA analysis of Nuc-pYStat5 and breast cancer specific survival. Detection of nuclear localized and tyrosine phosphorylated Stat5 by DAB-chromogen IHC in human breast tissue surgical explants treated with (positive control) or without (negative control) human prolactin (100nM; 60min) using AX1 (upper panels) and E208 (lower panels) anti-pYStat5 antibodies (a). Patients with high levels of Nuc-pYStat5 had significantly better breast cancer specific survival than patients with low Nuc-pYStat5 as measured by immunofluorescence and AQUA analysis (b). Kaplan-Meier analysis of breast cancer specific survival in Material II according to high or low Nuc-pYStat5 expression. Censored cases (+) and number of patients per group are indicated.

**Figure 3.** Loss of Nuc-pYStat5 during breast cancer progression. Levels of Nuc-pYStat5 as detected by immunofluorescence and quantified by AQUA were significantly reduced in invasive ductal carcinoma (IDC) (n=72) and lymph node metastases (n=17) when compared to normal (n=27) and ductal carcinoma in situ (DCIS) (n=14) in Material III.

**Figure 4.** Nuc-pYStat5 and breast cancer survival in patients receiving anti-estrogen adjuvant therapy. Nuc-pYStat5, detected by DAB-chromogen IHC and pathologist scoring (a) and immunofluorescence quantified by AQUA analysis (b), predicted breast cancer survival in patients treated with anti-estrogen therapy. Nuc-pYStat5 was also a significant prognostic marker in the node-negative subsets of both Material IV (c) and Material V (d). Censored cases (+) and number of patients per group are indicated.

**Table 1A.** Characteristics of Material I, II, IV, and V. ER - estrogen receptor; PR - progesterone receptor.

Discrete Variables		Material I Number (%)	Material II Number (%)	Material IV Number (%)	Material V Number (%)
Center	Fox Chase Cancer Center	94 (31.2)	--	--	--
	Kaiser Permanente	124 (41.2)	--	--	--
	University of Miami	42 (14.0)	--	--	--
	Washington University	41 (13.6)	--	--	--
Race	Asian	3 (1.0)	--	--	--
	Black	27 (9.0)	1 (0.6)	--	--
	White	270 (89.7)	156 (98.7)	--	--
	Other	1 (0.3)	1 (0.6)	--	--
Age	<50	60 (19.9)	46 (29.1)	22 (13.3)	8 (8.0)
	≥50	241 (80.1)	112 (70.9)	144 (86.7)	92 (92.0)
Tumor Size	<2cm	152 (50.5)	60 (38.0)	43 (26.1)	41 (41.0)
	≥2cm, <5cm	141 (46.8)	80 (50.6)	101 (61.2)	44 (44.0)
	≥5cm	8 (2.7)	18 (11.4)	21 (12.7)	15 (15.0)
Tumor/Nuclear Grade	1	71 (23.6)	34 (21.5)	35 (21.1)	15 (15.8)
	2	132 (43.9)	90 (57.0)	79 (47.6)	47 (49.5)
	3	98 (32.6)	34 (21.5)	52 (31.3)	33 (34.7)
ER Status	Negative	51 (16.9)	50 (31.6)	29 (21.8)	9 (10.3)
	Positive	250 (83.1)	108 (68.4)	104 (78.2)	78 (89.7)
PR Status	Negative	95 (31.6)	57 (36.1)	57 (34.3)	31 (36.5)
	Positive	206 (68.4)	101 (63.9)	109 (65.7)	54 (63.5)
Her2 Status	Non-overexpressed	--	138 (87.3)	131 (85.6)	86 (86.0)
	Over-expressed	--	20 (12.7)	22 (14.4)	14 (14.0)
Node Status	Negative	301 (100)	158 (100)	54 (34.8)	44 (50.6)
	Positive	0 (0)	0 (0)	101 (65.2)	43 (94.4)
Chemotherapy	Untreated	259 (86.0)	--	128 (79.0)	98 (99.0)
	Treated	42 (14.0)	--	34 (21.0)	1 (1.0)
Hormone Therapy	Untreated	226 (75.1)	--	0 (0)	0 (0)
	Treated	75 (24.9)	--	166 (100)	100 (100)
Radiation Therapy	Untreated	212 (70.4)	--	--	38 (39.2)
	Treated	89 (29.6)	--	--	59 (60.8)
Disease Recurrence/ Death of Disease	Disease Free/No Death	235 (78.1)	97 (61.4)	115 (69.3)	34 (38.6)
	Disease Recurrence/Death	66 (21.9)	61 (38.6)	51 (30.7)	54(61.4)
Nuc-pYStat5 Status	Low	218 (72.4)	83 (52.5)	93 (56.0)	10 (15.4)
	High	83 (27.6)	75 (47.5)	73 (44.0)	57 (84.6)

**Table 1B.** Characteristics of Material I, II, IV, and V. SD - standard deviation.

Continuous Variables		Mean	Median	Range	SD
<b>Material I</b>	Age at Diagnosis	61.8	62	29-88	12.9
	Tumor Size (cm)	2.1	1.9	0.5-7.5	1.0
	Followup (months)	106.8	100	2-326	65.4
	Nuc-pYStat5 score	5.8	0.0	0-100	11.0
	Year of Diagnosis			1974-1996	
<b>Material II</b>	Age at Diagnosis	58.1	58	28-86	12.2
	Tumor Size (cm)	2.6	2	0.4-11	1.7
	Followup (months)	187.0	167.4	4.2-465.0	123.7
	Nuc-pYStat5 score	856.2	776.7	264-2236	418.0
	Year of Diagnosis			1961-1983	
<b>Material IV</b>	Age at Diagnosis	62.9	63	35-97	12.0
	Tumor Size (cm)	3.1	2.5	0.5-13.0	18.6
	Followup (months)	57.6	58.5	2-137	27.1
	Nuc-pYStat5 score	1	0	0-4	1.3
	Positive Nodes	3.7	2	0-28	5.7
Year of Diagnosis			1985-1995		
<b>Material V</b>	Age at Diagnosis	69.2	71.4	37.9-89.3	11.4
	Tumor Size (cm)	2.6	2.0	0.4-11.0	1.7
	Followup (months)	123.2	112.0	81.1-225.8	35.2
	Nuc-pYStat5 score	1314.7	1030.9	533-4367	842.2
	Positive Nodes	2.4	0	0-22	3.8
Year of Diagnosis			1987-2000		

**Table 2.** Univariate and multivariate survival analyses of breast cancer recurrence, Material I, (A) and breast cancer specific survival, Material II (B). Cox proportional hazards regression was used to evaluate prognostic factors. HR - hazard ratio; CI - confidence interval.

<b>A. Material I</b>			Multivariate Adjusted			Unadjusted		
Variable		n	HR	p	95% CI	HR	p	95% CI
Age at Diagnosis	< 50	60	1	---	---	1	---	---
	≥ 50	241	0.755	0.357	0.416-1.373	0.793	0.405	0.460-1.368
Race	Caucasian	270	1	---	---	1	---	---
	Non-Caucasian	31	1.201	0.616	0.586-2.463	1.440	0.305	0.718-2.888
Tumor Grade	1	71	1	---	---	1	---	---
	2	132	2.443	0.032	1.078-5.534	2.521	0.024	1.128-5.637
	3	98	2.747	0.023	1.153-6.545	3.249	0.003	1.475-7.158
Tumor Size	< 2cm	152	1	---	---	1	---	---
	≥ 2cm, < 5cm	141	1.652	0.071	0.957-2.851	1.966	0.010	1.174-3.291
	≥ 5cm	8	1.290	0.698	0.356-4.672	2.376	0.162	0.707-7.979
ER Status	Negative	51	1	---	---	1	---	---
	Positive	250	2.382	0.043	1.028-5.517	0.945	0.861	0.504-1.771
PR Status	Negative	95	1	---	---	1	---	---
	Positive	206	0.575	0.093	0.302-1.097	0.724	0.206	0.439-1.195
Nuc-pYStat5	Low	218	2.493	0.019	1.164-5.340	2.883	0.005	1.377-6.039
	High	83	1	---	---	1	---	---

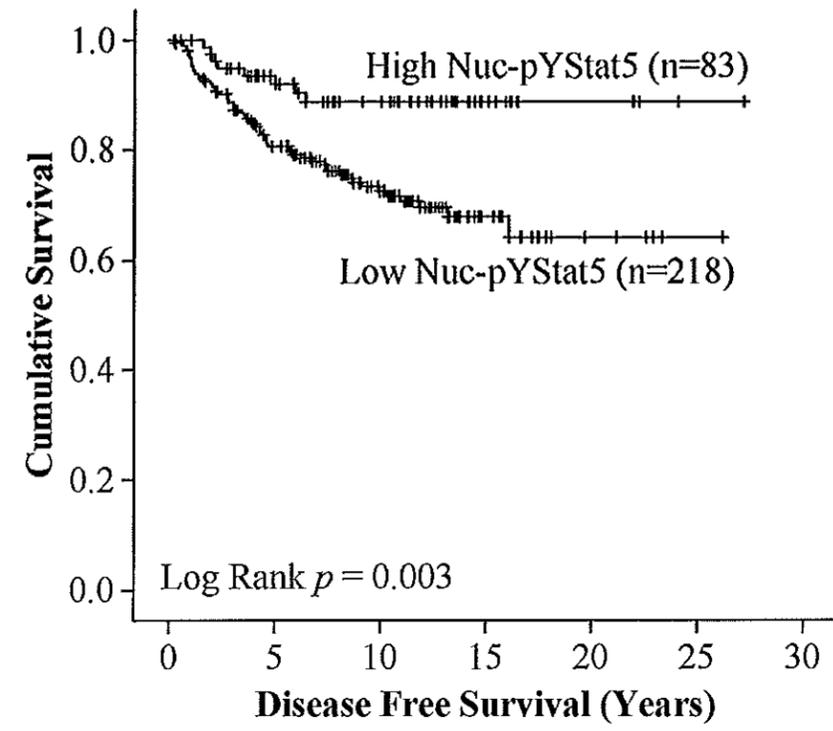
<b>B. Material II</b>			Multivariate Adjusted			Unadjusted		
Variable		n	HR	p	95% CI	HR	p	95% CI
Age at Diagnosis	< 50	46	1	----	----	1	----	----
	≥ 50	112	1.317	0.364	0.727-2.386	1.245	0.441	0.713-2.175
Race	Caucasian	156	1	----	----	1	----	----
	Non-Caucasian	2	5.867	0.024	1.269-27.124	4.417	0.042	1.058-18.449
Tumor Grade	1	34	1	----	----	1	----	----
	2	90	1.129	0.712	0.592-2.154	1.107	0.752	0.588-2.084
	3	34	0.732	0.504	0.294-1.825	0.923	0.557	0.705-1.207
Tumor Size	< 2cm	60	1	----	----	1	----	----
	≥ 2cm, < 5cm	80	2.386	0.006	1.290-4.413	2.143	0.012	1.184-3.876
	≥ 5cm	18	4.078	0.001	1.753-9.498	2.896	0.008	1.312-6.391
ER Status	Negative	50	1	----	----	1	----	----
	Positive	108	0.780	0.474	0.394-1.542	0.933	0.800	0.546-1.595
PR Status	Negative	57	1	----	----	1	----	----
	Positive	101	1.064	0.839	0.588-1.926	0.875	0.613	0.521-1.469
Her2 Status	Negative	138	1	----	----	1	----	----
	Over-expressed	20	1.218	0.633	0.541-2.744	1.112	0.780	0.528-2.339
Nuc-pYStat5	Low	83	1.835	0.039	1.03-3.270	1.898	0.016	1.128-3.194
	High	75	1	----	----	1	----	----

**Table 3.** Univariate and multivariate analysis of Nuc-pYStat5 and breast cancer specific survival in breast cancer patients treated with adjuvant anti-estrogen in Materials IV (A) and V (B). Cox proportional hazards regression was used to evaluate prognostic factors. HR - hazard ratio; CI - confidence interval.

<b>A. Material IV</b>		Multivariate Adjusted				Unadjusted			
Variable		n	HR	p	95% CI	n	HR	p	95% CI
Age at Diagnosis	< 50	20	1	---	---	22	1	---	---
	≥ 50	124	0.873	0.731	0.402-1.894	144	0.582	0.127	0.291-1.166
Tumor Size	< 2cm	36	1	---	---	43	1	---	---
	2 - <5cm	88	1.094	0.828	0.486-2.464	101	1.352	0.410	0.660-2.766
	≥ 5cm	20	3.520	0.019	1.232-10.060	21	3.350	0.006	1.414-7.938
Tumor Grade	1	28	1	---	---	35	1	---	---
	2	68	1.186	0.773	0.373-3.776	79	2.698	0.072	0.914-7.958
	3	48	2.868	0.070	0.917-8.964	52	7.188	<0.001	2.507-20.606
ER/PR Status	Negative	21	1	---	---	23	1	---	---
	Positive	123	0.350	0.007	0.162-0.755	143	0.270	<0.001	0.145-0.505
Her2 Status	Normal	124	1	---	---	131	1	---	---
	Over-expressed	20	1.827	0.153	0.799-4.176	22	3.086	0.001	1.620-5.877
Lymph Node Status	Negative	49	1	---	---	54	1	---	---
	Positive	95	2.177	0.068	0.945-5.014	101	2.951	0.009	1.314-6.627
Nuc-pYStat5	Low	83	1.290	0.460	0.656-2.538	93	1.833	0.043	1.020-3.293
	High	61	1	---	---	73	1	---	---

<b>B. Material V</b>		Multivariate Adjusted				Unadjusted			
Variable		n	HR	p	95% CI	n	HR	p	95% CI
Age at Diagnosis	< 50	5	1	---	---	7	1	---	---
	≥ 50	44	1.479	0.725	0.167-13.073	90	6.673	0.060	0.921-48.354
Tumor Size	< 2cm	22	1	---	---	39	1	---	---
	2 - <5cm	22	1.008	0.992	0.197-5.153	43	2.836	0.001	1.509-5.330
	≥ 5cm	5	1.245	0.789	0.251-6.173	15	2.027	0.112	0.848-4.846
Tumor Grade	1	6	1	---	---	14	1	---	---
	2	24	1.415	0.668	0.290-6.904	46	0.815	0.624	0.358-1.851
	3	19	0.804	0.659	0.306-2.115	32	2.089	0.073	0.934 -4.672
ER/PR Status	Negative	5	1	---	---	8	1	---	---
	Positive	44	0.232	0.030	0.062-0.871	76	0.349	0.011	0.155-0.783
Her2 Status	Normal	41	1	---	---	83	1	---	---
	Over-expressed	8	2.512	0.091	0.865-7.301	14	2.000	0.050	1.001-3.996
Lymph Node Status	Negative	26	1	---	---	43	1	---	---
	Positive	23	4.447	0.003	1.645-12.025	41	3.667	<0.001	1.990-6.756
Nuc-pYStat5	Low	7	7.937	<0.001	2.506-20.00	10	4.802	<0.001	2.054-11.226
	High	42	1	---	---	55	1	---	---

**Figure 1.**



**Figure 2.**

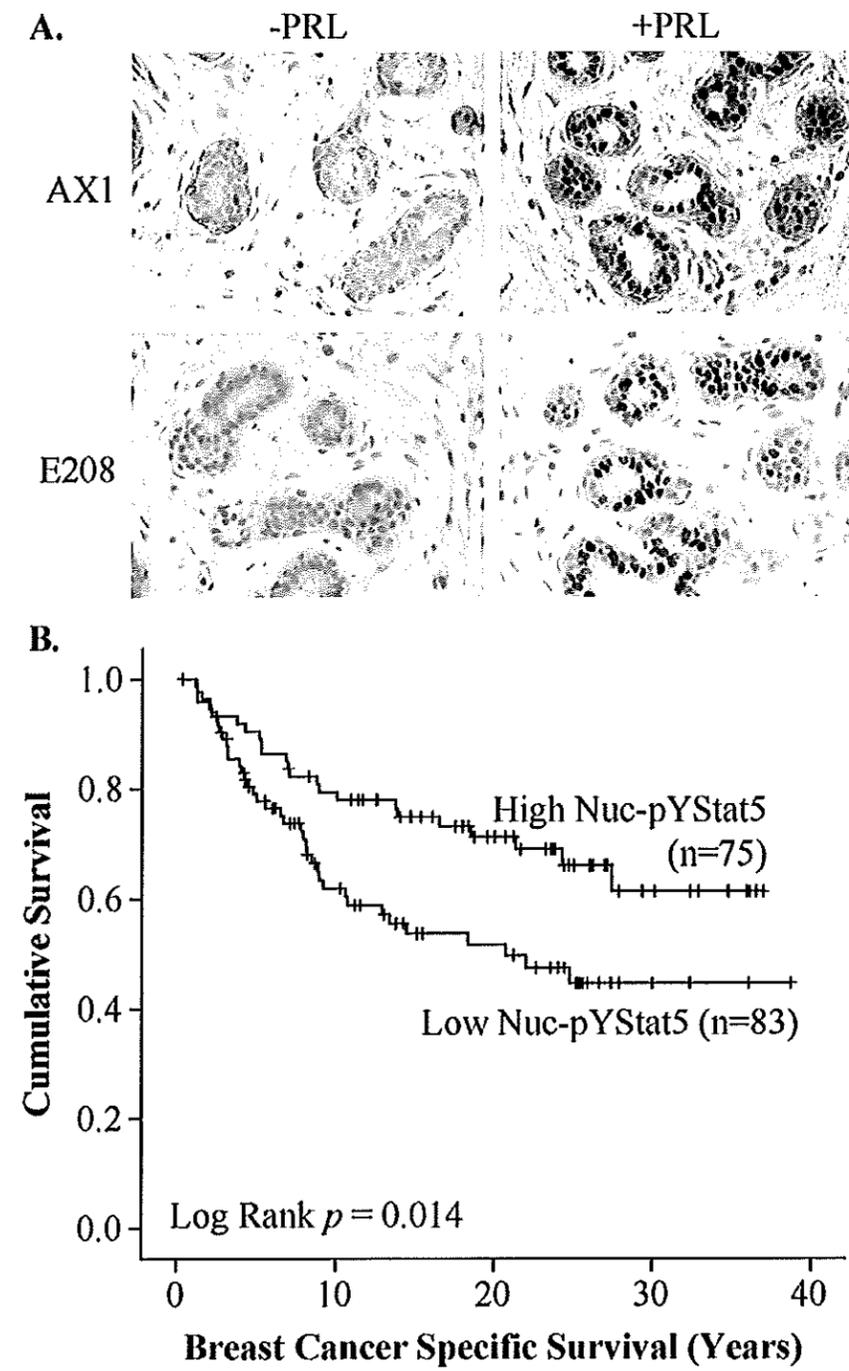
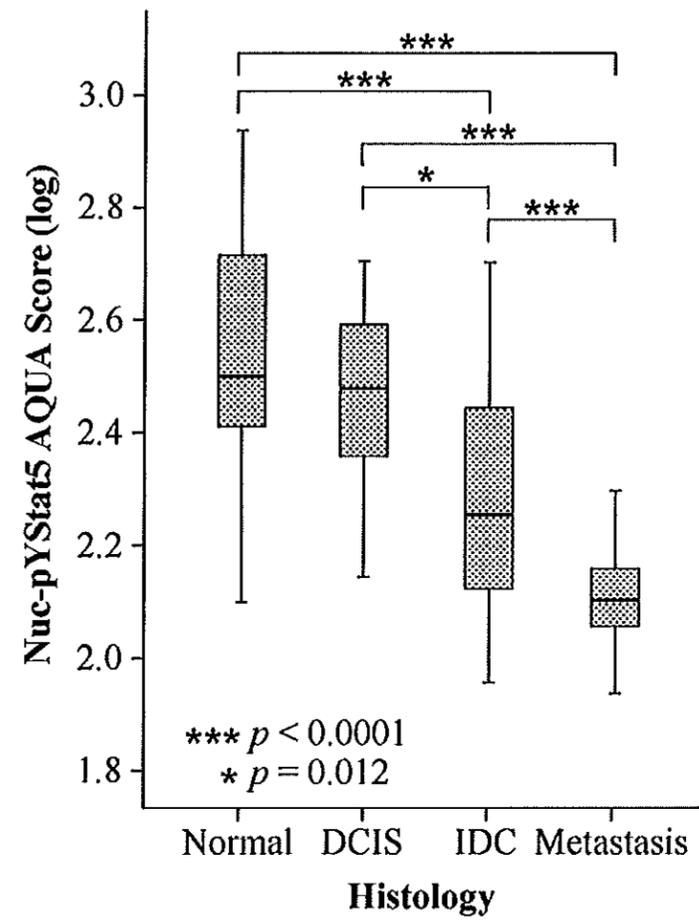
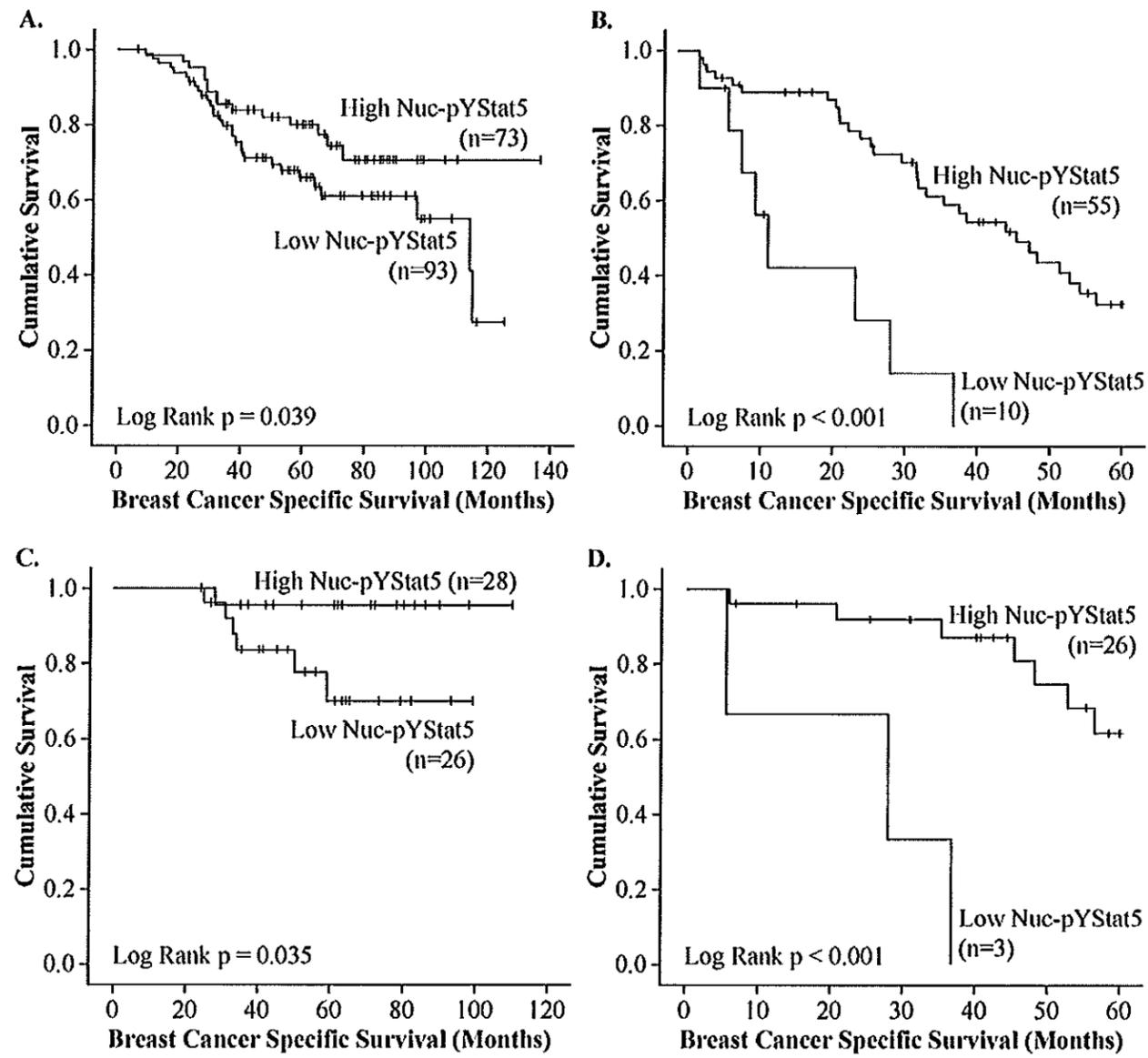
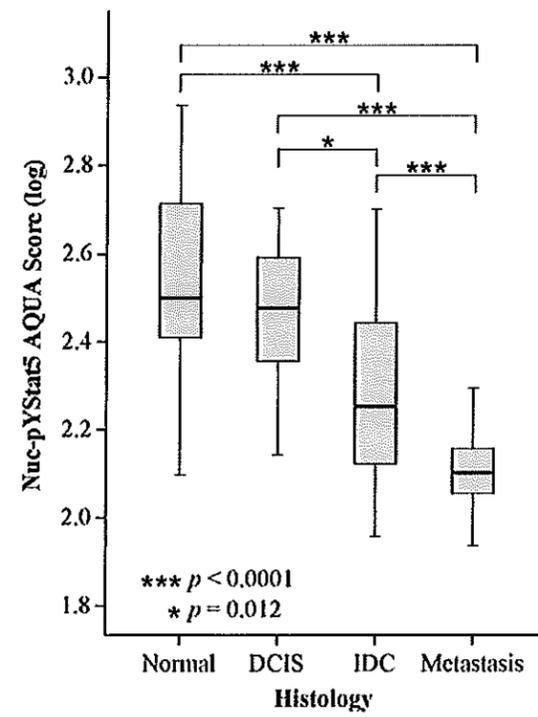
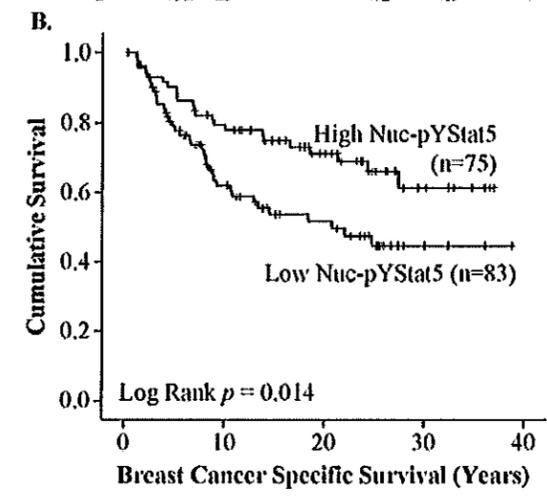
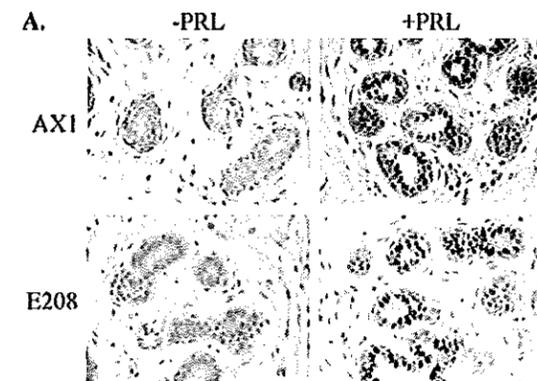
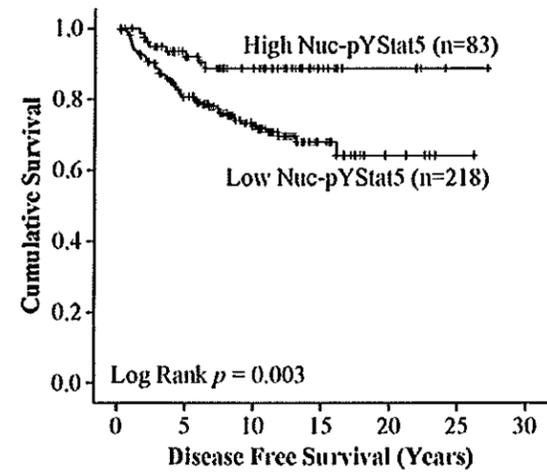


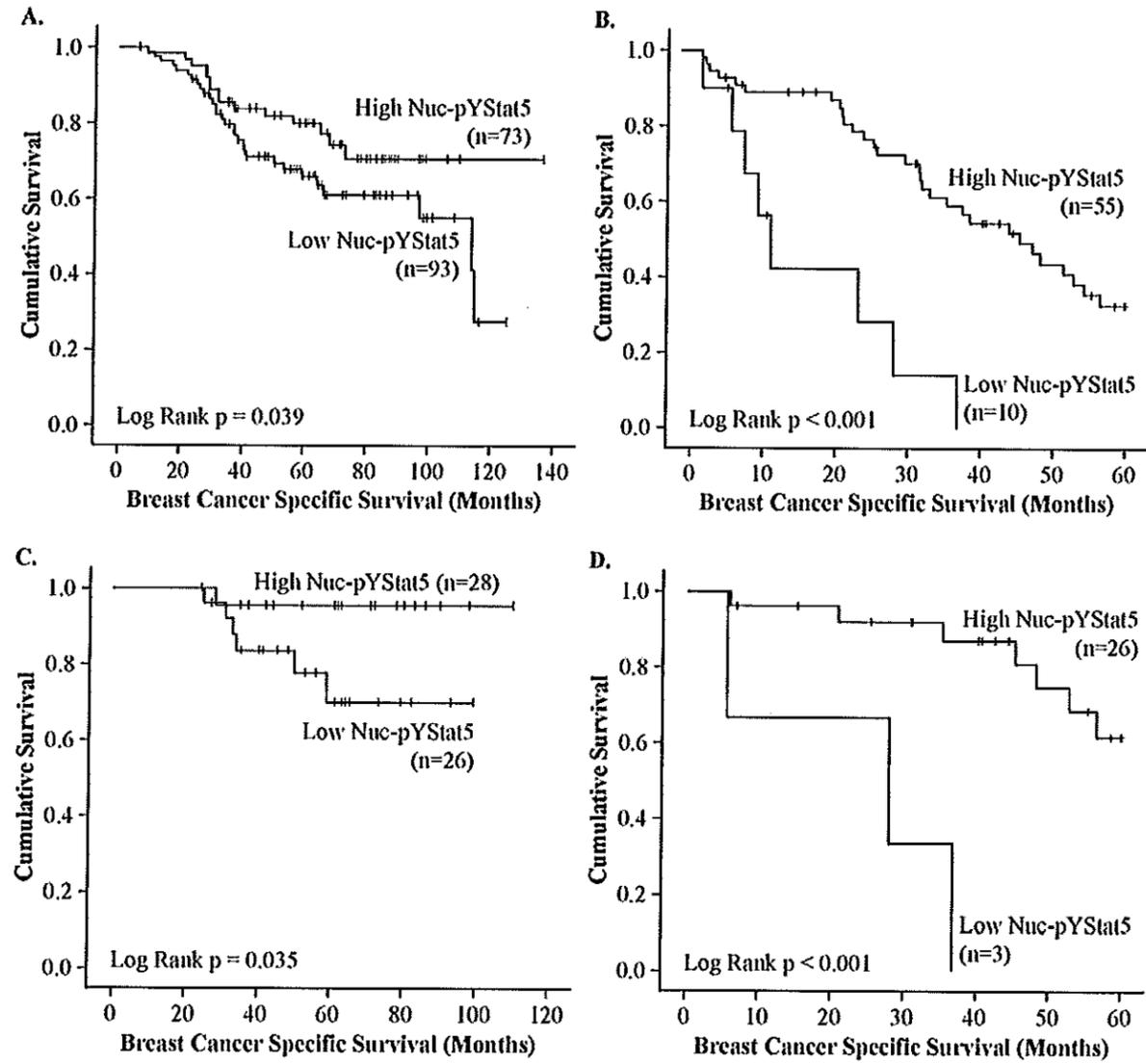
Figure 3.

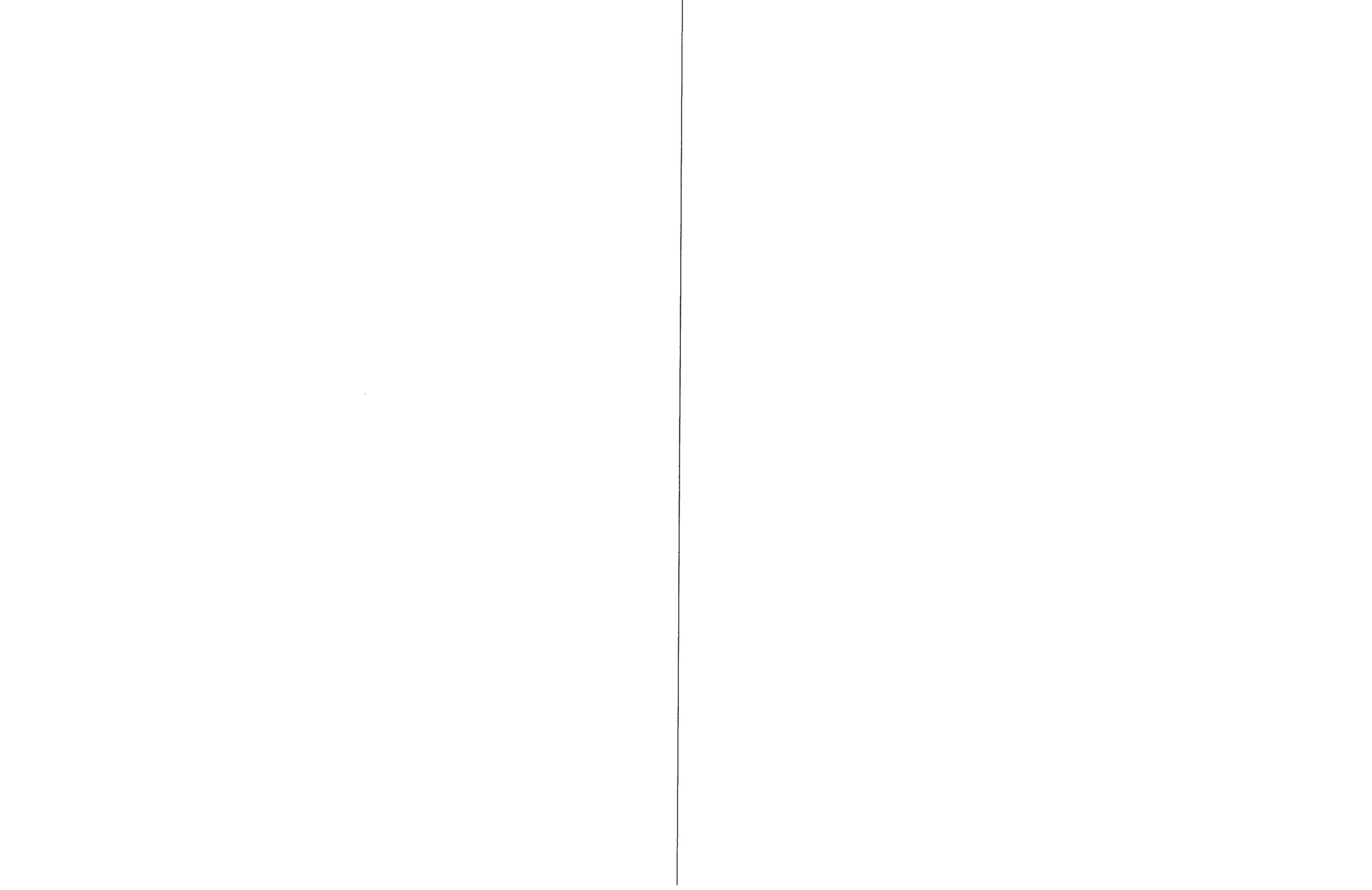


**Figure 4**









Molecular characterization of breast cancer progression: early lesions are not genetically advanced

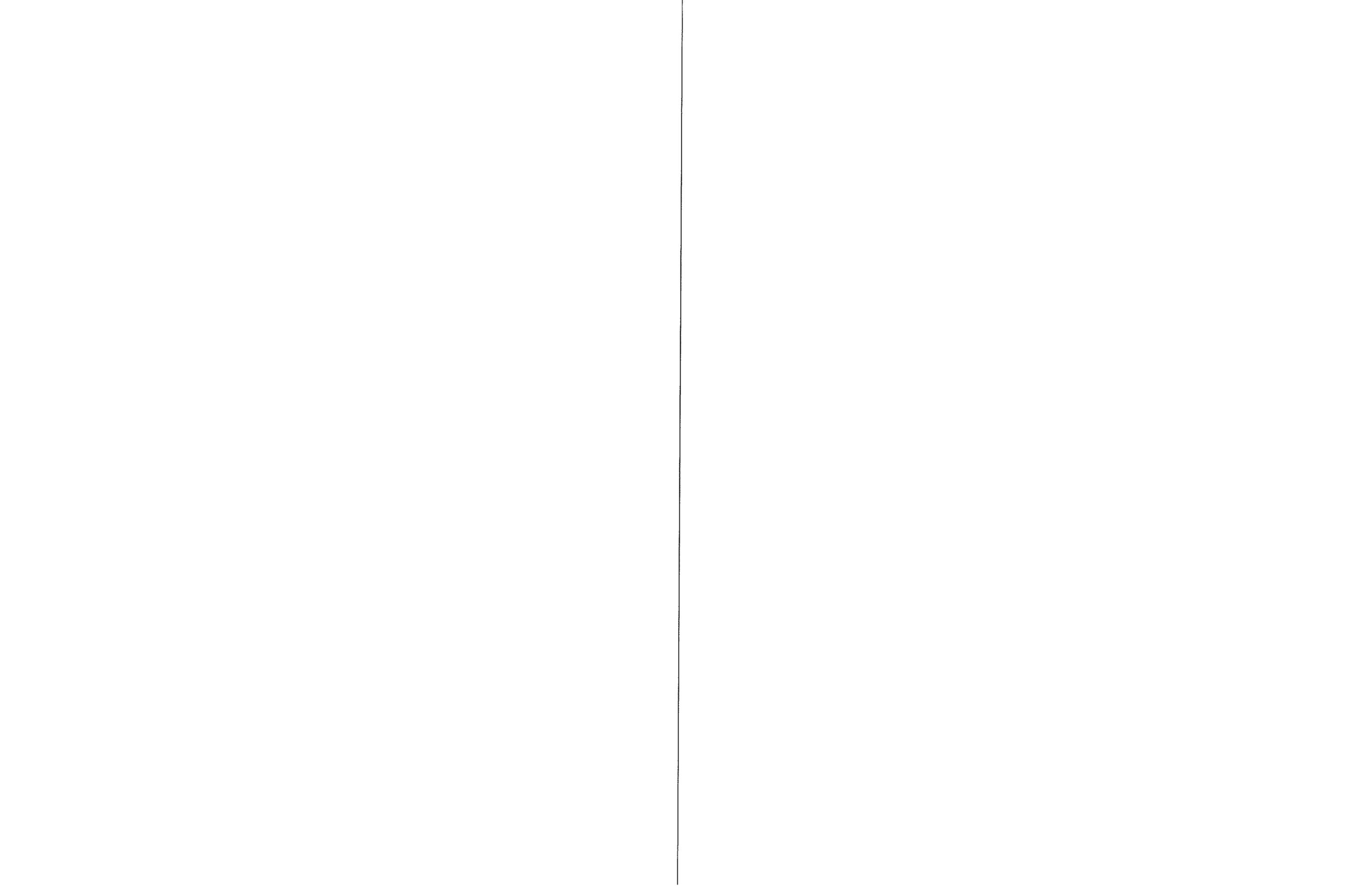
Rachel E. Ellsworth<sup>1</sup>, Jamie D. Weyandt<sup>2</sup>, Jamie L. Fantacone-Campbell<sup>3</sup>, Brenda Deyarmin<sup>2</sup>, Darrell L. Ellsworth<sup>2</sup>, Jeffrey A. Hooke<sup>3</sup>, Craig D. Shriver<sup>3</sup>

**Introduction:** Columnar cell lesions (CCL) and atypical ductal hyperplasia (ADH) frequently coexist and share molecular changes with *in situ* and invasive components, suggesting that CCL and ADH may be precursors to breast cancer. These conclusions are, however, largely based on studies examining CCL and ADH from patients diagnosed with more advanced disease. Thus, allelic imbalance (AI) was assessed in pure CCL or ADH specimens to characterize molecular changes in early breast lesions.

**Methods:** DNA samples were obtained from laser microdissected CCL (n=42) or ADH (n=31) lesions without concurrent *in situ* or invasive disease. AI was assessed at 26 chromosomal regions commonly altered in breast cancer. Data was analyzed using Fisher's exact and Student's t-tests using a cutoff of  $P < 0.05$  to define significance.

**Results:** The average AI frequency was 6.2% (range 0-20%) in CCL and 6.1% (range 0-25%) in ADH. The highest levels of AI were on chromosomes 17q12-q21 in CCL (15%) and 8q24 in ADH (23%); ~33% of early lesions did not have any detectable genetic changes. Levels of AI for both CCL and ADH were significantly ( $P < 0.0001$ ) lower than those found in *in situ* or invasive disease. Genetic changes characteristic of more advanced *in situ* and invasive disease, especially changes on chromosome 16q and 17p, were infrequent or non-detectable in pure early lesions.

**Conclusions:** Pure CCL and ADH lesions demonstrate lower levels of genetic alterations than DCIS, invasive carcinomas or early lesions from cancerous breasts. Alterations characteristic of advanced disease were not found in lesions from non-cancerous breasts. While data generated from synchronous lesions has been used as support for a model of evolution from CCL and ADH to *in situ* and invasive disease, pure early lesions are, in fact, not genetically advanced and molecular profiles do not support these lesions as obligatory precursors to more advanced disease. Given the molecular differences between pure and synchronous lesions, current models of disease initiation and progression and risk assessment based largely on data generated from cancerous breasts must be re-evaluated.



Laser Microdissection for Gene Expression Profiling

<sup>1</sup>Lori A. Field, <sup>1</sup>Brenda Deyarmin, <sup>2</sup>Craig D. Shriver, <sup>1</sup>Darrell L. Ellsworth, <sup>3</sup>Rachel E. Ellsworth

<sup>1</sup>Windber Research Institute, Windber, PA; <sup>2</sup>Walter Reed Army Medical Center, Washington, DC; and <sup>3</sup>Henry M. Jackson Foundation for the Advancement of Military Medicine, Windber, PA

Contact information for corresponding author:

Rachel E. Ellsworth, PhD

Director, Translational Breast Research

Clinical Breast Care Project

620 Seventh Street

Windber, PA 15963

(814) 361-6926

r.ellsworth@wriwindber.org

## **Abstract**

Microarray-based gene expression profiling is revolutionizing biomedical research by allowing expression profiles of thousands of genes to be interrogated in a single experiment. In cancer research, the use of laser microdissection (LM) to isolate RNA from tissues provides the ability to accurately identify molecular profiles from different cell types that comprise the tumor and its surrounding microenvironment. Because RNA is an unstable molecule, the quality of RNA extracted from tissues can be affected by sample preparation and processing. Thus, special protocols have been developed to isolate research-quality RNA after LM. This chapter provides detailed descriptions of protocols used to generate microarray data from high-quality frozen breast tissue specimens, as well as challenges associated with formalin-fixed paraffin-embedded (FFPE) specimens.

**Key Words:** laser microdissection; gene expression; microarray; frozen tissue; FFPE; molecular signature; breast cancer.

## 1. Introduction

Tumorigenesis is a complex process, involving structural changes at multiple chromosomal locations and altered expression of numerous genes and proteins. Early efforts to identify genes involved in cancer development evaluated single genes with known or putative roles in cellular processes such as growth, proliferation, angiogenesis, and apoptosis. While these efforts have resulted in the identification of more than 350 genes (1), additional genes of unknown or presumably unrelated function likely play critical roles in cancer development and progression (2). cDNA microarrays, which allow quantitative, large-scale analysis of gene expression, provide a global approach to identifying genes involved in tumorigenesis and metastasis without *a priori* knowledge of the underlying molecular pathways (3). Microarrays have been used to develop molecular signatures that correlate with tumor characteristics or outcomes and are being used in clinical diagnostic tests to guide treatments for patients with breast cancer (4,5).

Despite the successful development of clinical assays and the publication of hundreds of microarray-based papers, the majority of microarray studies have used RNA isolated from tissue by homogenization or manual microdissection. Because the majority of human tumors are highly heterogeneous, with numerous cell types comprising the primary tumor and surrounding microenvironment, laser microdissection (LM) is necessary to isolate specific cells. For example, RNA isolated from laser-microdissected breast tumor cells will be free from contamination from normal epithelial, stromal, and vascular cells, which could compromise the accuracy of the resulting gene expression profiles.

Because RNA is sensitive to degradation, isolation of RNA after LM requires a defined protocol that includes careful cleaning of all equipment with RNase inhibitors, special histological stains, and rapidity (less than 30 minutes) in cutting, mounting, and microdissecting the tissues. In this chapter, we present protocols for performing microarray analysis using RNA isolated after LM and describe alternate protocols for gene expression analysis of formalin-fixed paraffin-embedded (FFPE) archival specimens.

## **2. Materials**

### **2.1 Sectioning, Staining, and Laser Microdissection**

1. Membrane-based laser microdissection slides (Catalog #11505158; W. Nuhsbaum)
2. Disposable microtome blades, HP35n, non-coated (#3152765; Thermo Fisher Scientific)
3. 0.5 ml PCR tubes (#951010057; Eppendorf)
4. RNaseZap<sup>®</sup> (#AM9780; Applied Biosystems)
5. Nuclease-free water (#AM9932; Applied Biosystems)
6. LCM Staining Kit (#AM1935; Applied Biosystems) — *store Cresyl Violet at 4°C*
7. 50% ethanol
8. 75% ethanol
9. 95% ethanol
10. 100% ethanol
11. Xylene (used only for FFPE samples)

12. Tissue-Tek<sup>®</sup> Cryomold<sup>®</sup> Standard, 25x20x5mm (#62534-25; Electron Microscopy Sciences)
13. Cryomatrix Optimal Cutting Temperature (OCT) compound (#6769006; Thermo Fisher Scientific)

## 2.2. RNA Isolation from Frozen Tissue

1. RNAqueous<sup>®</sup>-Micro kit (#AM1931; Applied Biosystems) — *store LCM Additive, 10X DNase I Buffer, DNase I, and DNase Inactivation Reagent at -20°C*
2. Nuclease-free water
3. 100% ethanol
4. Agilent RNA 6000 Pico kit (#5067-1513; Agilent Technologies) — *store Pico Dye Concentrate, Pico Marker, Pico Conditioning Solution, Pico Gel Matrix, and Spin Filters at 4°C; store Pico Ladder at -20°C*
5. Agilent 2100 Bioanalyzer (Agilent Technologies)
6. RNaseZap<sup>®</sup>

## 2.3. Amplification and Fragmentation of RNA from Frozen Tissue

1. MessageAmp<sup>™</sup> II aRNA Amplification kit (#AM1751; Applied Biosystems) — *store Box #2 at -20°C*
2. GeneChip<sup>®</sup> Eukaryotic Poly-A RNA Control kit (#900433; Affymetrix) — *store at -20°C*
3. 75 mM Bio-11-UTP (#AM8451; Applied Biosystems) — *store at -80°C*
4. Nuclease-free water

5. 5X Fragmentation Buffer, component of the GeneChip® Sample Cleanup Module (#900371; Affymetrix)
6. Agilent RNA 6000 Pico kit
7. Agilent RNA 6000 Nano kit (#5067-1511; Agilent Technologies) — *store Nano Dye Concentrate, Nano Marker, Nano Gel Matrix, and Spin Filters at 4°C; store Nano Ladder at -20°C*
8. Agilent 2100 Bioanalyzer
9. NanoDrop ND-1000 Spectrophotometer (Thermo Fisher Scientific) — Note: the current model is the NanoDrop 2000.

#### **2.4. Hybridization of aRNA to Microarrays**

1. GeneChip® Expression 3' Amplification Reagents containing 20X Eukaryotic Hybridization Controls and Control Oligonucleotide B2 (#900454; Affymetrix) — *store at -20°C*
2. Herring Sperm DNA (#D1811; Promega) — *store at -20°C*
3. Bovine Serum Albumin (BSA) (#15561-020; Invitrogen) — *store at -20°C*
4. MES Hydrate (#M5287; Sigma-Aldrich)
5. MES Sodium Salt (#M5057; Sigma-Aldrich)
6. 5M NaCl (#S5150; Sigma-Aldrich)
7. 0.5M EDTA (#E7889; Sigma-Aldrich)
8. Tween 20 (#H5152; Promega)
9. DMSO (#D5879; Sigma-Aldrich)
10. Nuclease-free water

11. GeneChip<sup>®</sup> Human Genome U133A 2.0 Arrays (HG U133A 2.0) (#900469; Affymetrix) — *store at 4°C*
12. Hybridization oven (Affymetrix)

### **2.5. Washing, Staining, and Scanning Microarrays**

1. Bovine Serum Albumin (BSA)
2. Streptavidin Phycoerythrin (SAPE) (#S866; Invitrogen) — *cover with foil and store at 4°C*
3. Goat IgG (#I5256; Sigma-Aldrich) — *store at 4°C*
4. Biotinylated anti-streptavidin (#BA-0500; Vector Laboratories) — *store at 4°C*
5. 20X SSPE (#S2015; Sigma-Aldrich)
6. 5M NaCl
7. Tween-20
8. Nuclease-free water
9. Tough Spots (#T-SPOTS; Diversified Biotech)
10. Fluidics Station (Affymetrix)
11. Scanner (Affymetrix)

### **2.6. Quantitative Real-Time Polymerase Chain Reaction of RNA from Frozen**

#### **Tissue**

1. High-capacity cDNA Reverse Transcription Kit (#4374966; Applied Biosystems) — *store at -20°C*

2. TaqMan<sup>®</sup> Universal PCR Master Mix (#4304437 or #4364340; Applied Biosystems) — *store at 4°C*
3. TaqMan<sup>®</sup> Gene Expression Assays (Inventoried: #4331182 or Made to Order: #4351372; Applied Biosystems) — *store at -20°C shielded from light*
4. FirstChoice<sup>®</sup> Human Brain Reference RNA (#AM6050; Applied Biosystems) — *store at -80°C*
5. iCycler iQ<sup>™</sup> PCR plates (#223-9441; Bio-Rad Laboratories)
6. iCycler iQ<sup>™</sup> thermal seals (#223-9444; Bio-Rad Laboratories)
7. iCycler iQ<sup>™</sup> Real-Time PCR Detection System (Bio-Rad Laboratories)

#### **2.7. RNA Isolation from Formalin-Fixed Paraffin Embedded Specimens**

1. RecoverAll<sup>™</sup> Total Nucleic Acid Isolation Kit (#AM1975; Applied Biosystems)
2. 100% ethanol

#### **2.8. Quantitative Real-Time Polymerase Chain Reaction of RNA from FFPE**

1. See list of materials in **Section 2.6. – “Quantitative Real-Time Polymerase Chain Reaction of RNA from Frozen Tissue”**.

### **3. Methods**

RNA is extremely susceptible to degradation by RNase enzymes in the environment. To generate high-quality microarray or quantitative real-time polymerase chain reaction (qRT-PCR) data, it is critical to obtain RNA of the highest possible quality by preventing RNase contamination during tissue collection and processing, RNA isolation, and

downstream applications. Several general precautions should be taken when working with RNA in the laboratory. All equipment and laboratory benches should be thoroughly cleaned with RNaseZap<sup>®</sup> and then rinsed with nuclease-free or deionized water. All pipette tips, tubes, reagents, and other consumables must be RNase-free. Pipette tips should contain barriers and should be changed each time you pipette, even if you are pipetting the same reagent, to avoid potential cross-contamination between samples and to prevent RNase contamination. For most procedures, it is advisable to use nuclease-free, hydrophobic, non-stick tubes to minimize loss of sample that may otherwise adhere to the tube walls. Gloves should be worn at all times and changed frequently, especially after coming into contact with liquids or surfaces that may be contaminated with RNases.

### **3.1 Sectioning, Staining, and Laser Microdissection**

To prevent RNA degradation, tissue sectioning, staining, and LM must be performed as quickly as possible (typically within 30 minutes). In our laboratory, 2 individuals perform these steps and process one slide at a time. The LCM Staining Kit employs a novel staining procedure that avoids exposing the tissue sections to pure water at any step, thus minimizing the potential for RNA degradation.

#### **3.1.1. Reagent Preparation**

1. In the bottles provided with the LCM Staining Kit, prepare 95%, 75%, and 50% ethanol solutions by diluting 100% ethanol with nuclease-free water. Add the dehydration beads to the bottle labeled 100% ethanol and add absolute ethanol. Do not use the ethanol in this container to make any of the diluted solutions.

2. Clean the staining containers included in the LCM Staining Kit with RNaseZap<sup>®</sup>.

For FFPE samples, a glass staining dish should also be cleaned. Spray the containers generously with RNaseZap<sup>®</sup> and allow them to sit for 10 minutes. Rinse twice with distilled water and then perform a final rinse with nuclease-free water. Allow the containers to dry under a hood and then fill with the appropriate solutions.

### 3.1.2. Sectioning

1. Set the temperature of the cryostat to -30°C.
2. Clean the knife holder (*not* the knife blade itself) with 100% ethanol and treat the brushes that will be used to manipulate the tissue sections with RNaseZap<sup>®</sup>.
3. Cool the specimen and brushes in the cryostat.
4. Inside the cryostat, remove the frozen OCT-embedded tissue from its cryomold and mount securely to the metal specimen stage with OCT compound, orienting the tissue according to regions of interest (**Note 1**).
5. Using a fresh disposable blade, shave OCT from the block until the tissue becomes visible. Set the cutting thickness to 8 µm.
6. Section the tissue and use a small brush to straighten out the newly-cut sections.
7. Manipulate sections onto the foil slides (**Note 2**).

### 3.1.3. Staining

Perform staining under a hood used only for RNA procedures. Change all containers and blade surfaces between each patient sample.

#### **A. OCT-embedded Tissue**

1. Cut sections at 8  $\mu\text{m}$  and mount onto a membrane-based laser microdissection slide.
2. Wash slide in 95% ethanol for 30 seconds.
3. Wash in 75% ethanol for 30 seconds.
4. Wash in 50% ethanol for 30 seconds (**Note 3**).
5. Pipette Cresyl Violet (~50  $\mu\text{l}$ ) onto the slide to completely cover the tissue sections; allow the slide to sit for 15 seconds.
6. Rinse in 95% ethanol for 5 seconds.
7. Rinse in 100% ethanol for 5 seconds.
8. Rinse in a second container of 100% ethanol for 30 seconds (**Note 4**).
9. Allow slide to air dry.

#### **B. Formalin-Fixed Paraffin-Embedded Tissue**

1. Fill the clean staining dish with nuclease-free water and warm on a hot plate to the desired temperature for the paraffin being used (typically 37-42°C). Change the water bath between each sample.
2. Cut sections at 8  $\mu\text{m}$  and lay out ribbon onto the warm water bath.
3. Mount sections onto a membrane-based LM slide.
4. Place slides in an incubator set at 56°C for 15 minutes (**Note 5**).
5. Wash in xylene for 1 minute; repeat twice for a total of 3 washes.
6. Wash in 95% ethanol for 30 seconds.

7. Wash in 75% ethanol for 30 seconds.
8. Wash in 50% ethanol for 30 seconds.
9. Pipette Cresyl Violet stain onto the slide using enough volume to cover the sections; allow to sit for 15 seconds.
10. Rinse in 95% ethanol for 5 seconds.
11. Rinse in 100% ethanol for 5 seconds.
12. Rinse in 100% ethanol for 30 seconds.

#### **3.1.4. Laser Microdissection**

1. Use a cover-slipped H&E section to orient the tissue for microdissection. Estimate the number of cells — in our experience, ~10,000 cells usually yields sufficient RNA for downstream applications.
2. Locate the area on the Cresyl Violet-stained section to be microdissected (**Note 6**).
3. Pipette 60  $\mu$ l of Lysis Solution for OCT-embedded tissues, or 60  $\mu$ l of Digestion Buffer for FFPE-embedded tissues, into the cap of a clean 0.5 ml Eppendorf tube. Place the cap into the cap holder apparatus of the Laser Microdissection System.
4. Microdissect the area of interest (**Fig. 1**) and drop the sample into the buffer (**Note 7**).
5. Add the remaining 40  $\mu$ l of Lysis Solution (OCT tissues) or 340  $\mu$ l of Digestion Buffer (FFPE tissues) to the tube and carefully close the lid.

### **3.2. RNA Isolation from Frozen Tissue**

#### **3.2.1. Reagent Preparation**

1. Before first use, add 10.5 ml of 100% ethanol to Wash Solution 1 and 22.4 ml of 100% ethanol to Wash Solution 2/3 and mix well (**Note 8**).
2. On first use, thaw the Pico Ladder on ice, centrifuge briefly, and transfer to an RNase-free tube. Heat-denature the ladder for 2 minutes at 70°C in a heat block, then immediately place on ice. Add 90 µl of nuclease-free water, pipette up and down several times, and flick the tube to mix. Briefly centrifuge the tube and aliquot 5-10 µl to RNase-free tubes. Store at -70°C (**Note 9**).

### 3.2.2. RNA Isolation

1. Place a tube containing nuclease-free water (at least 50 µl per sample) in a heat block at 95°C.
2. Pre-warm an air incubator to 42°C.
3. Thaw LCM Additive and 10X DNase I Buffer on ice.
4. Flick the tube containing the microdissected sample from **Section 3.1.4. – “Laser Microdissection”** several times and centrifuge briefly. Place the sample in the 42°C incubator for 30 minutes (**Note 10**).
5. Approximately 6-7 minutes prior to completion of the 30-minute incubation, pre-wet the Micro Filter by adding 30 µl of Lysis Solution to the filter, which is placed in a Micro Elution Tube (Micro Filter Cartridge Assembly). After 5 minutes, centrifuge the Micro Filter Cartridge Assembly for 30 seconds at 16,000 rcf to remove the Lysis Solution from the filter.

6. Remove the microdissected sample from the 42°C incubator, vortex on maximum speed by pulsing 3 times, and centrifuge briefly. Add 3 µl of LCM Additive, mix by vortexing, and centrifuge briefly.
7. Add 52 µl of 100% ethanol and mix completely into the sample by pipetting up and down (**Note 11**). Transfer the sample to the center of the filter in the Micro Filter Cartridge Assembly. Centrifuge for 1 minute at 10,000 rcf (**Note 12**).
8. Add 180 µl of Wash Solution 1 to the filter and centrifuge for 1 minute at 10,000 rcf.
9. Add 180 µl of Wash Solution 2/3 and centrifuge for 30 seconds at 16,000 rcf. Repeat this step one time.
10. Remove the filter from the collection tube and discard the flow-through. Recap the assembly and centrifuge for 1 minute to remove trace amounts of liquid.
11. Remove the filter containing the sample and place in a new Micro Elution Tube.
12. Add 10 µl of nuclease-free water heated to 95°C in Step 1 above to the center of the filter (**Note 13**). Incubate the assembly for 5 minutes at room temperature, then centrifuge for 1 minute at 16,000 rcf to elute the RNA. Repeat this step with a second 10 µl volume of 95°C nuclease-free water, incubate, and centrifuge.
13. Remove the filter and place the sample on ice.

### 3.2.3. DNase Treatment

1. Add 2 µl of 10X DNase I Buffer and 1 µl of DNase I to the sample and mix by gently flicking the tube. Centrifuge briefly and incubate for 20 minutes in a heat

block at 37°C. During the incubation, remove the DNase Inactivation Reagent from the freezer and thaw at room temperature.

2. Remove the sample from the heat block. Vigorously vortex the DNase Inactivation Reagent and add 2.3 µl to the sample. Gently tap the side of the tube to mix and incubate for 2 minutes at room temperature. After 1 minute, vortex the sample, tap the tube to move all contents to the bottom, and continue the incubation for 1 minute.
3. Centrifuge the sample for 1 minute 30 seconds at 16,000 rcf to pellet the DNase Inactivation Reagent. Transfer the supernatant containing the RNA to a new tube without disturbing the pellet, then place the RNA on ice (**Note 14**).

#### **3.2.4. Assessing RNA Integrity**

1. Remove an aliquot of the Pico Ladder from the freezer and thaw on ice. Remove the Pico Gel Matrix, Pico Dye Concentrate, Pico Conditioning Solution, and Pico Marker from 4°C and allow the reagents to warm to room temperature for at least 30 minutes. Ensure that the Dye Concentrate is shielded from light (**Note 15**).
2. Add 550 µl of Gel Matrix to a Spin Filter and centrifuge for 10 minutes at 1500 rcf. Aliquot 65 µl of filtered gel into the tubes provided with the kit (produces 7-8 tubes of filtered gel). The filtered gel may be stored at 4°C for up to 2 months.
3. Vortex the tube of Dye Concentrate for 10 seconds and then centrifuge briefly. Add 1 µl of Dye Concentrate to a tube of filtered gel (warmed to room temperature), vortex for 10 seconds, then centrifuge for 10 minutes at 16,000 rcf. One tube of gel-dye mix can be used to run 2 chips per day.

4. Transfer 1.25-1.5  $\mu$ l of each RNA sample into a 0.65 ml tube. Heat the sample for 2 minutes in a heat block at 65-70°C. Place on ice for ~5 minutes to cool, then centrifuge briefly to collect the RNA at the bottom of the tube.
5. Start the 2100 Expert Software and turn on the Bioanalyzer. Place an electrode cleaner containing 350  $\mu$ l of nuclease-free water in the instrument and close the lid (**Note 16**). On the instrument menu, select “Assays”, “Electrophoresis”, “RNA”, and finally “Eukaryotic Total RNA Pico Series”. Select the number of samples (from 1-11) to be assayed. Enter the sample information and any additional comments pertaining to that sample.
6. Place the Pico Chip on the chip priming station (ensuring that the base plate is on “C”) and pull the syringe back to 1 ml. Add 9  $\mu$ l of gel-dye mix to the well labeled with an encircled “G” (**Note 17**). Close the chip priming station until you hear a click, then press the syringe down until it is secured beneath the syringe clip. After 30 seconds, release the clip, wait 5 seconds, and pull the syringe back to the 1 ml mark.
7. Add 9  $\mu$ l of gel-dye mix to the 2 remaining wells marked “G”. Add 9  $\mu$ l of Pico Conditioning Solution to the well marked “CS”. Add 5  $\mu$ l of Pico Marker to the ladder well and to each well that will contain an RNA sample. Add 6  $\mu$ l of Pico Marker to any empty sample wells.
8. Add 1  $\mu$ l of diluted Pico Ladder to the ladder well and 1  $\mu$ l of sample to the appropriate sample well. After loading all wells, vortex the chip using the manufacturer-supplied vortex for 1 minute at 2,400 rpm. During this time, remove

the electrode cleaner from the instrument. Place the Pico Chip on the Agilent 2100 Bioanalyzer and begin the run by pressing “Start” (Notes 18 and 19) (Fig. 2).

### **3.3. Amplification and Fragmentation of RNA from Frozen Tissue**

When using small amounts of RNA for gene expression analysis, it is often necessary to first amplify the RNA to generate sufficient material for hybridization to the microarray. For RNA isolated from laser-microdissected tissues, 2 rounds of amplification are normally required. All frozen reagents for the amplification protocol should be thawed on ice; enzymes should be stored at -20°C immediately prior to and after use. All master mixes should be prepared in excess (generally ~5%) to avoid running short of master mix when working with large numbers of samples.

#### **3.3.1. Preparation of Poly-A Spike-in Controls**

1. Completely thaw the Poly-A Control Stock on ice, then add 2 µl to a small tube. Add 38 µl of nuclease-free water and mix well by vortexing or flicking the tube. Centrifuge briefly to collect the liquid at the bottom of the tube. This is the First Dilution and can be stored at -80°C for up to 6 weeks (or 8 freeze-thaw cycles).
2. Remove 2 µl of the First Dilution and place in a new tube. Add 98 µl of nuclease-free water to make the Second Dilution. Mix well and centrifuge briefly.
3. Combine 2 µl of the Second Dilution with 98 µl of nuclease-free water to make the Third Dilution. Mix well and centrifuge.
4. Combine 2 µl of the Third Dilution with 18 µl of nuclease-free water to prepare the Fourth Dilution. Mix well and centrifuge.

5. Combine 2  $\mu$ l of the Fourth Dilution with 18  $\mu$ l of nuclease-free water to prepare the Fifth Dilution. Mix well and centrifuge (**Note 20**).

### 3.3.2. Preparation of Samples for Amplification

1. Using the estimated RNA concentration obtained from the Bioanalyzer, calculate the volume of sample containing 10 ng of RNA (**Note 21**). Transfer this volume to a 0.2 ml PCR tube and adjust the total volume to 9  $\mu$ l with nuclease-free water. If the volume needed for 10 ng of RNA is greater than 9  $\mu$ l, transfer this amount to a hydrophobic, non-stick microcentrifuge tube and centrifuge in a vacuum concentrator until the volume is  $\leq 9$   $\mu$ l. Transfer the concentrated sample to a 0.2 ml PCR tube and adjust the volume to 9  $\mu$ l with nuclease-free water.
2. Flick the tubes to mix and centrifuge briefly to collect the liquid at the bottom of the tube.

### 3.3.3. First Round Amplification

1. Add 2  $\mu$ l of the Fifth Dilution of the Poly-A Controls to each sample containing 10 ng of RNA (**Note 22**). Flick the tubes to mix and centrifuge briefly.
2. Add 1  $\mu$ l of Oligo(dT) Primer to each sample, flick the tubes to mix, and centrifuge briefly. Incubate samples for 10 minutes at 70°C in a thermal cycler.
3. Remove samples from the thermal cycler, centrifuge briefly, and place on ice.
4. In a small tube, prepare a master mix containing the following for each sample:
  - a. 2  $\mu$ l 10X First Strand Buffer
  - b. 4  $\mu$ l dNTP mix

c. 1  $\mu$ l RNase Inhibitor

d. 1  $\mu$ l ArrayScript™

Vortex the tube to mix and centrifuge briefly to collect the contents at the bottom of the tube. Add 8  $\mu$ l of the master mix to each sample, flick the tubes to mix, and centrifuge. Incubate samples for 2 hours at 42°C in an air incubator or hybridization oven, then centrifuge briefly, and place on ice.

5. Prepare a master mix on ice containing the following reagents for each sample:

a. 63  $\mu$ l nuclease-free water

b. 10  $\mu$ l 10X Second Strand Buffer

c. 4  $\mu$ l dNTP mix

d. 2  $\mu$ l DNA Polymerase

e. 1  $\mu$ l of RNase H

Vortex to mix and centrifuge briefly to collect the master mix at the bottom of the tube. Add 80  $\mu$ l of master mix to each sample, flick the samples to mix, and centrifuge briefly. Incubate the samples in a pre-cooled thermal cycler for 2 hours at 16°C (**Note 23**), then centrifuge briefly and place on ice.

6. Place a tube containing at least 30  $\mu$ l of nuclease-free water per sample in a heat block set to 50-55° C. For each sample, place a filter inside a cDNA Elution tube.

Note: add 24 ml of 100% ethanol to the Wash Buffer before using for the first time.

7. Transfer the samples from the 0.2 ml tubes to 1.5 ml microcentrifuge tubes. Add 250  $\mu$ l of cDNA Binding Buffer to each sample, mix by pipetting up and down and then flicking the tubes several times. Centrifuge samples briefly, then transfer each

sample to the filter of a cDNA Filter Cartridge. Centrifuge samples for 1 minute at 10,000 rcf, then discard the flow-through.

8. Add 500  $\mu$ l of Wash Buffer to each filter. Centrifuge for 1 minute at 10,000 rcf and discard the flow-through.
9. Centrifuge the cDNA Filter Cartridges for 1 minute at 10,000 rcf to remove any residual liquid from the filter. Transfer filters to new cDNA Elution tubes and discard the old tubes.
10. Add 10  $\mu$ l of nuclease-free water warmed to 50-55° C to the center of each filter. Incubate for 2 minutes at room temperature. Elute samples by centrifuging for 1 minute 30 seconds at 10,000 rcf. Repeat this step using a second 10  $\mu$ l volume of warm nuclease-free water.
11. Discard filters and place tubes containing the eluted cDNA on ice.
12. Prepare the *in vitro* transcription (IVT) master mix at room temperature. Note that for the first round of amplification, the IVT reactions contain only unmodified dNTPs. For each sample include:
  - a. 4  $\mu$ l T7 ATP
  - b. 4  $\mu$ l T7 CTP
  - c. 4  $\mu$ l T7 GTP
  - d. 4  $\mu$ l T7 UTP
  - e. 4  $\mu$ l T7 10X Reaction Buffer
  - f. 4  $\mu$ l T7 Enzyme mix

Vortex the master mix and centrifuge briefly to collect the contents at the bottom of the tube. Aliquot 24  $\mu$ l of master mix to each sample, flick the tubes to mix, and

centrifuge briefly. Incubate samples for 14 hours in an air incubator or hybridization oven at 37°C.

13. Place a tube containing nuclease-free water in a heat block at 50-55°C — we recommend heating at least 120 µl of nuclease-free water per sample.
14. For each sample, place an aRNA Filter Cartridge in an aRNA Collection Tube.
15. Remove the IVT reactions from the incubator. Add 60 µl of nuclease-free water to each sample, mix by flicking the tube, and centrifuge briefly. Add 350 µl of aRNA Binding Buffer followed by 250 µl of 100% ethanol to each sample. Mix the samples by pipetting up and down at least 5 times, then transfer each sample to an aRNA Filter Cartridge. Centrifuge samples for 1 minute at 10,000 rcf, then discard the flow-through and remount the filter on the collection tube.
16. Add 650 µl of Wash Buffer to each Filter Cartridge and centrifuge for 1 minute at 10,000 rcf. Discard the flow-through and place the Filter Cartridge back inside the collection tube. Centrifuge samples for an additional 1 minute at 10,000 rcf to remove residual Wash Buffer. Discard the flow-through and place the Filter Cartridge in a new collection tube.
17. Apply 100 µl of nuclease-free water warmed to 50-55°C to the center of each filter. Incubate at room temperature for 2 minutes, then centrifuge for 1 minute at 10,000 rcf to elute the aRNA.
18. Remove 3 µl of the aRNA and transfer to a small tube. Heat the samples for 2 minutes in a heat block at 65-70°C. Place samples on ice to cool, then centrifuge the samples briefly, and return to ice.

19. Run 1  $\mu\text{l}$  of the first round aRNA samples from Step 18 on the Bioanalyzer using a Pico Chip following the instructions outlined in **Section 3.2.4. – “Assessing RNA Integrity” (Fig. 3A)**. Use 1.5  $\mu\text{l}$  of the remaining aRNA to measure the concentration of each sample on a NanoDrop ND-1000 Spectrophotometer.

### 3.3.4. Second Round Amplification

1. Calculate the volume of first round aRNA (using the concentration obtained on the NanoDrop) needed to obtain 1  $\mu\text{g}$  of starting material for the second round of amplification (**Note 24**). If this volume exceeds 10  $\mu\text{l}$  for any sample, concentrate those samples in a vacuum concentrator to less than 10  $\mu\text{l}$ . In a 0.2 ml PCR tube, adjust the volume of all samples to 10  $\mu\text{l}$  using nuclease-free water.
2. Add 2  $\mu\text{l}$  of Second Round Primers to each aRNA sample. Flick the tubes to mix and centrifuge briefly. Place samples in a thermal cycler heated to 70°C for 10 minutes, then centrifuge briefly and place on ice.
3. Prepare a master mix containing the following for each sample:
  - a. 2  $\mu\text{l}$  10X First Strand Buffer
  - b. 4  $\mu\text{l}$  dNTP mix
  - c. 1  $\mu\text{l}$  RNase Inhibitor
  - d. 1  $\mu\text{l}$  ArrayScript™

Vortex the master mix and centrifuge briefly. Add 8  $\mu\text{l}$  of master mix to each sample and flick the tubes to mix. Centrifuge briefly and incubate for 2 hours at 42°C in an air incubator or hybridization oven.

4. Following incubation, centrifuge the samples briefly and place on ice. Add 1  $\mu$ l of RNase H to each sample, flick the tubes to mix, and centrifuge briefly to collect the contents at the bottom of the tube. Incubate samples for 30 minutes at 37°C in an air incubator or hybridization oven, then centrifuge briefly and place on ice.
5. Add 5  $\mu$ l of the Oligo(dT) Primer to each sample, flick the tubes to mix, and centrifuge briefly. Incubate samples for 10 minutes at 70°C in a thermal cycler, then centrifuge and place on ice.
6. Prepare a master mix on ice for the second strand synthesis that includes the following for each sample:
  - a. 58  $\mu$ l nuclease-free water
  - b. 10  $\mu$ l 10X Second Strand Buffer
  - c. 4  $\mu$ l dNTP mix
  - d. 2  $\mu$ l DNA PolymeraseVortex to mix and add 74  $\mu$ l to each sample. Flick the tubes to mix and centrifuge briefly. Incubate samples for 2 hours in a thermal cycler that has been pre-cooled to 16°C. Remove samples from the thermal cycler, centrifuge briefly and place on ice.
7. Purify the cDNA following the exact procedure outlined in Steps 6-11 from **Section 3.3.3. – “First Round Amplification”** above.
8. Prepare a master mix at room temperature that contains for each sample:
  - a. 4  $\mu$ l T7 ATP
  - b. 4  $\mu$ l T7 CTP
  - c. 4  $\mu$ l T7 GTP
  - d. 2.6  $\mu$ l T7 UTP

e. 1.4  $\mu$ l biotin-11-UTP

f. 4  $\mu$ l T7 Reaction Buffer

g. 4  $\mu$ l T7 Enzyme Mix

Vortex and centrifuge briefly to collect the contents at the bottom of the tube. Add

24  $\mu$ l of master mix to each sample, flick the tubes to mix, and centrifuge briefly.

Incubate the samples for 14 hours at 37°C in an air incubator or hybridization oven.

9. Purify the second round, labeled aRNA using the same procedure for aRNA purification outlined in **Section 3.3.3. – “First Round Amplification”** (Steps 13-17).
10. Run the second round, labeled aRNA on the Bioanalyzer using the Agilent RNA 6000 Nano kit. Reagent and sample preparation for the Nano Chip is very similar to that for the Pico Chip with minor exceptions. Warm all refrigerated Nano reagents to room temperature. Prepare the filtered gel and gel-dye mix using the Nano Gel Matrix and Nano Dye Concentrate as outlined above in Steps 2-3 of **Section 3.2.4. – “Assessing RNA Integrity”**.
11. Thaw the Nano Ladder on ice. Flick the tube several times and centrifuge briefly. Transfer 2.5  $\mu$ l of the Nano Ladder to a new tube. Prepare 5-10  $\mu$ l aliquots of the remaining ladder and store at -20°C for future use. Place 3  $\mu$ l of each aRNA sample in a small tube. Heat the samples and ladder for 2 minutes in a heat block at 65-70°C. Place the samples on ice to cool, then centrifuge briefly to collect any condensation at the bottom of the tube.

12. Start the 2100 Expert Software and clean the electrodes as previously described.  
Select the “Eukaryotic Total RNA Nano Series Assay” and select the number of samples (from 1-12) that will be run. Enter the sample information.
13. Load the Nano Chip using the same procedure for loading the Pico Chip (**Section 3.2.4. – “Assessing RNA Integrity”**), but note that the Nano Chip does not use Conditioning Solution, allowing 12 samples to be run (**Fig. 3B**).
14. Measure the concentration of 1.5  $\mu$ l of the remaining second round aRNA on the NanoDrop.

### **3.3.5. Fragmentation of Labeled aRNA**

1. For each sample, transfer 15  $\mu$ g of labeled aRNA to a 0.2 ml PCR tube. The aRNA yield after 2 rounds of amplification typically exceeds 1,500 ng/ $\mu$ l and often is ~2,000 ng/ $\mu$ l; therefore, you should not have to vacuum-concentrate the samples prior to fragmentation. Add nuclease-free water to each sample to bring the total volume to 24  $\mu$ l. Flick the tubes to mix and centrifuge briefly.
2. Add 6  $\mu$ l of 5X Fragmentation Buffer to each sample, flick the tubes to mix, and briefly centrifuge. Fragment the aRNA samples by incubating for 35 minutes at 94°C in a thermal cycler. Place the samples on ice to cool, then centrifuge briefly to collect condensation at the bottom of the tube.
3. Run the fragmented samples on a Bioanalyzer Nano Chip (**Note 25**) (**Fig. 4**).

### **3.4. Hybridization of Fragmented aRNA to Microarrays**

Microarrays for gene expression analysis are available from several commercial vendors, including Affymetrix, Agilent Technologies, and Illumina. In this chapter, we provide protocols for hybridization of labeled, fragmented aRNA to the Affymetrix HG U133A 2.0 arrays. The protocol is the same for all eukaryotic arrays manufactured by Affymetrix; however, the starting amount of fragmented aRNA and volume of reagents used in certain steps will vary across the different array formats.

### 3.4.1. Reagent Preparation

1. Thaw the 20X Eukaryotic Hybridization Controls on ice. Mix well by flicking the tube or vortexing gently and then centrifuge briefly to collect the contents at the bottom of the tube. Prepare aliquots containing the volume used during a typical hybridization set-up (we usually hybridize 8 samples at a time) and store at -20°C.
2. Prepare 1 L of 12X MES Stock Buffer by dissolving 64.61g of MES hydrate and 193.3 g of MES Sodium Salt in 800 ml of nuclease-free water. When completely dissolved, adjust the volume to 1,000 ml with nuclease-free water and filter through a 0.22 µm filter. The pH should be between 6.5 and 6.7. Store at 4°C wrapped in foil to shield from light. Discard the solution if it turns yellow.
3. Prepare 50 ml of 2X Hybridization Buffer by combining:
  - a. 8.3 ml 12X MES Stock Buffer
  - b. 17.7 ml 5M NaCl
  - c. 4.0 ml 0.5M EDTA
  - d. 19.9 ml nuclease-free water
  - e. 100 µl 10% Tween-20

Filter the solution through a 0.22  $\mu\text{m}$  filter and store wrapped in foil at 4°C.

### 3.4.2. Hybridization

1. Approximately 1 hour before setting up the hybridization, remove the HG U133A 2.0 arrays from the refrigerator and allow them to equilibrate to room temperature. Thaw the Control Oligonucleotide B2, 20X Eukaryotic Hybridization Controls, Herring Sperm DNA, and BSA on ice. Set the temperature of the Hybridization Oven to 45°C.
2. Transfer 20  $\mu\text{l}$  (10  $\mu\text{g}$ ) of fragmented aRNA to a 1.5 ml microcentrifuge tube.
3. Prepare a 1X solution of Hybridization Buffer by combining equal volumes of 2X Hybridization Buffer and nuclease-free water; vortex to mix well.
4. Heat the 20X Eukaryotic Hybridization Controls for 5 minutes at 65°C in a heat block before adding to the hybridization master mix.
5. Prepare the hybridization master mix (at 5% excess) by combining for each sample:
  - a. 3.3  $\mu\text{l}$  Control Oligonucleotide B2
  - b. 10  $\mu\text{l}$  20X Eukaryotic Hybridization Controls
  - c. 2  $\mu\text{l}$  Herring Sperm DNA
  - d. 2  $\mu\text{l}$  BSA
  - e. 100  $\mu\text{l}$  2X hybridization buffer
  - f. 20  $\mu\text{l}$  DMSO
  - g. 42.7  $\mu\text{l}$  nuclease-free water

Vortex to mix well, centrifuge briefly, and add 180  $\mu$ l of the master mix to each fragmented sample (from Step 2 above). Vortex the samples, centrifuge briefly, and incubate for 5 minutes at 99°C in a heat block.

6. Place the arrays face down on a laboratory tissue. Insert a pipette tip into the top right septum on the back of the array to permit air to vent when filling the array. Fill the arrays with 160  $\mu$ l of 1X Hybridization Buffer, remove the pipette tip vent, and incubate the arrays for 10 minutes at 60 rpm in the Hybridization Oven set to 45°C.
7. Transfer the samples from 99°C to a 45°C heat block. Incubate for 5 minutes, then centrifuge the samples for 5 minutes at 16,000 rcf.
8. Remove the arrays from the Hybridization Oven. Insert the pipette tip vent and remove the hybridization buffer. Load 130  $\mu$ l of the sample into the array, making sure that a bubble is present that can freely move when the array is slowly tilted from side to side. Incubate the arrays for 16 hours at 45°C in the Hybridization Oven while rotating at 60 rpm (**Notes 26 and 27**).

### **3.5. Washing, Staining, and Scanning of Microarrays**

#### **3.5.1. Reagent Preparation**

1. Prepare 250 ml of 2X Stain Buffer by combining:
  - a. 41.7 ml 12X MES Stock Buffer (see **Section 3.4.1. – “Reagent Preparation”**)
  - b. 92.5 ml 5M NaCl
  - c. 2.5 ml 10% Tween-20
  - d. 113.3 ml nuclease-free water

Filter through a 0.22  $\mu\text{m}$  filter, wrap in foil to shield from light, and store at 4°C.

2. Reconstitute the goat IgG by adding 150 mM NaCl to make a 10 mg/ml solution.

For example, add 1 ml of 150 mM NaCl to 10 mg of lyophilized goat IgG. Aliquots should be stored at -20°C, but once thawed for use, store at 4°C.

3. Add 1 ml of nuclease-free water to reconstitute the biotinylated anti-streptavidin antibody to 0.5 mg/ml. Gently pipette up and down to mix, then store at 4°C.

4. Prepare 1 L of Wash A solution by combining:

- a. 300 ml 20X SSPE
- b. 699 ml deionized water
- c. 1 ml 10% Tween-20

Filter through a 0.22  $\mu\text{m}$  filter and store at room temperature.

5. Prepare 1 L of Wash B solution by combining:

- a. 83.3 ml 12X MES Stock Buffer
- b. 5.2 ml 5M NaCl
- c. 1 ml 10% Tween-20
- d. 910.5 ml deionized water

Filter through a 0.22  $\mu\text{m}$  filter, cover with foil, and store at 4°C.

### **3.5.2. Washing, Staining, and Scanning**

1. Prepare a master mix (in 5% excess) of the SAPE Solution (Stains 1 and 3) by

combining for each sample:

- a. 600  $\mu\text{l}$  2X Stain Buffer
- b. 48  $\mu\text{l}$  BSA

c. 12  $\mu$ l SAPE (vortex well before pipetting)

d. 540  $\mu$ l of deionized water

Vortex to mix. For each sample, aliquot 600  $\mu$ l of the SAPE Solution into two 1.5 ml microcentrifuge tubes.

2. Prepare a master mix for the Antibody Solution (Stain 2) in 5% excess. For each sample, combine:

a. 300  $\mu$ l 2X Stain Buffer

b. 24  $\mu$ l BSA

c. 6  $\mu$ l IgG

d. 3.6  $\mu$ l biotinylated anti-streptavidin

e. 266.4  $\mu$ l deionized water

Mix well by vortexing and aliquot 600  $\mu$ l of the Antibody Solution master mix into a 1.5 ml microcentrifuge tube for each sample.

3. After 16 hours, remove the arrays from the Hybridization Oven. Insert the pipette tip vent and remove the sample solution from the arrays. Fill the arrays with 160  $\mu$ l of Wash A solution.

4. Start the GeneChip<sup>®</sup> Operating System (GCOS) or Affymetrix<sup>®</sup> GeneChip<sup>®</sup> Command Console<sup>®</sup> (AGCC) and enter the sample and experiment information.

5. Replace the appropriate water bottles on the fluidics station with Wash A solution and Wash B solution. Prime all modules on the fluidics station.

6. Select the appropriate wash and stain protocol for your arrays and press "Run". Load the arrays and staining tubes onto the fluidics station. The protocol will

automatically run until staining of the arrays is complete (usually ~1 hour and 15 minutes).

7. Remove the arrays from the fluidics station. Place Tough Spots over the septa on the back of the arrays to prevent leakage during scanning. Once the scanner has warmed up (10-15 minutes), scan the arrays using the GCOS or AGCC software. When scanning is complete, zoom in on each scanned image and check the entire image for any abnormalities. The scanned images can now be analyzed to generate signal intensities for the probes on the arrays as well as a QC report.
8. Once the wash and stain protocol is complete, replace the Wash A solution and Wash B solution with deionized water. Run the Shutdown protocol on all modules, and when finished, turn off the fluidics station.

### **3.6. Quantitative Real-Time Polymerase Chain Reaction Assays Using RNA from Frozen Tissue**

Gene expression differences identified by microarray analysis are frequently confirmed using qRT-PCR. Due to the low yield of RNA following laser microdissection of small tissue sections, RNA is amplified for a single round prior to performing qRT-PCR. To amplify the RNA, follow the procedure detailed in **Section 3.3.3. – “First Round Amplification”**; however, it is not necessary to add the Poly-A controls.

Note that there are several different chemistries available for performing qRT-PCR, including TaqMan<sup>®</sup> probes, Molecular Beacons, Scorpions<sup>®</sup>, or SYBR<sup>®</sup> Green. The following protocol uses TaqMan<sup>®</sup> Gene Expression Assays, which utilize the same PCR conditions so there is no need to design PCR primers or optimize PCR conditions. In

order to achieve accurate results from qRT-PCR, be careful and precise when performing each step in the protocol, especially pipetting.

### 3.6.1. Reverse Transcription of RNA from Frozen Tissue

1. Reverse transcribe the aRNA (one round of amplification) using the High-Capacity cDNA Reverse Transcription Kit. Up to 2 µg of RNA (or aRNA) can be reverse transcribed (**Note 28**). In addition, reverse transcribe an appropriate amount of FirstChoice<sup>®</sup> Human Brain Reference RNA or other reference RNA which will be used to calibrate the relative levels of gene expression in the samples of interest.
2. If necessary, vacuum concentrate the samples to  $\leq 10$  µl and increase the volume to 10 µl with nuclease-free water in a 0.2 ml PCR tube.
3. Prepare the Reverse Transcription master mix (in 5% excess) containing the following for each sample:
  - a. 2 µl 10X RT Buffer
  - b. 0.8 µl 25X dNTP Mix
  - c. 2 µl 10X RT Random Primers
  - d. 1 µl MultiScribe Reverse Transcriptase
  - e. 1 µl RNase Inhibitor
  - f. 3.2 µl nuclease-free water

Vortex to mix, centrifuge briefly, and add 10 µl of the master mix to each sample.

4. Flick the sample tubes to mix and centrifuge briefly to collect the contents at the bottom of the tube. Incubate the samples in a thermal cycler using the following

program: 10 minutes at 25°C, 2 hours at 37°C, 5 seconds at 85°C, and then hold at 4°C.

5. When the reverse transcription reaction is complete, add nuclease-free water to adjust the concentration of the samples and the reference to 5 ng/μl. For example, if 1 μg of aRNA was reverse transcribed in a 20 μl reaction, add 180 μl of water to bring the concentration to 1 μg/200 μl or 5 ng/μl.
6. Mix the diluted samples well and centrifuge briefly. Place on ice or store at -20°C.

### 3.6.2. qRT-PCR of RNA from Frozen Tissue

1. Perform the qRT-PCR using the TaqMan<sup>®</sup> Gene Expression Assay of interest (**Note 29**). These assays can be run on a variety of real-time instruments, including the Bio-Rad iCycler<sup>™</sup>. Turn the iCycler<sup>™</sup> on and allow it to warm up for at least 15 minutes. Enter the plate set-up sample information, select and load the appropriate fluorophore (FAM-490), and save the file.
2. qRT-PCR reactions should be performed in duplicate, but if sufficient starting material is available, triplicate or quadruplicate reactions are recommended. The following protocol is based on 10 ng of cDNA per reaction (**Note 30**). Pipette 2 μl of cDNA (5 ng/μl) into the appropriate well of an iCycler<sup>™</sup> PCR plate.
3. Prepare a master mix containing the following per reaction:
  - a. 2.5 μl TaqMan<sup>®</sup> Gene Expression Assay (20X)
  - b. 20.5 μl nuclease-free water
  - c. 25 μl TaqMan<sup>®</sup> Universal PCR Master Mix

When running a full 96-well plate, make a master mix for 100 reactions (excess of 4 reactions). Vortex the master mix and centrifuge briefly to bring the contents to the bottom of the tube. Add 48  $\mu$ l of master mix to each sample well. Place a thermal seal on the plate, centrifuge the plate for ~10 seconds at ~3,000 ref, and then load the plate on the iCycler™. Set the reaction volume to 50  $\mu$ l and run the following protocol: 95°C for 10 minutes, 50 cycles of 95°C for 15 seconds followed by 60°C for 1 minute, then hold at 4°C.

4. Average the  $C_t$  values for each sample and the reference (**Note 31**). Calculate the relative transcript levels for each sample using the Comparative  $C_t$  Method (**Note 32**).

### **3.7. RNA Isolation from Formalin-Fixed Paraffin Embedded Specimens**

Archived FFPE tissues represent a valuable source of molecular information for the study of cancer (6-8). Although formalin fixation and paraffin embedding preserves tissue structure and cellular morphology, protein-protein and protein-nucleic acid cross links form during the preservation process, which chemically modify and damage the nucleic acids. RNA is particularly susceptible to fragmentation by FFPE, often making RNA isolated from FFPE tissues unusable for molecular analysis. The following protocol is specifically designed to recover RNA from FFPE samples that can be used for downstream applications such as qRT-PCR.

#### **3.7.1. Reagent Preparation**

1. Add 42 ml of 100% ethanol to the Wash 1 Concentrate and 48 ml of 100% ethanol to the Wash 2/3 Concentrate. Mix well.

### 3.7.2. RNA Isolation from FFPE Tissue

1. Place the laser microdissected FFPE samples from **Section 3.1.4. – “Laser Microdissection”** in 400 µl of Digestion Buffer. Add 4 µl of Protease to each sample and mix the tubes gently ensuring that the samples are entirely immersed in buffer. Incubate the samples in a heat block set to 50°C for 3 hours (**Note 33**).
2. Add 480 µl of Isolation Additive to each sample and vortex to mix. The solution should be white and cloudy (**Note 33**).
3. Add 1.1 ml of 100% ethanol and mix by carefully pipetting the sample up and down. The sample should become clear after mixing (**Note 33**).
4. Place a Filter Cartridge inside a collection tube provided with the kit. Apply 700 µl of the sample to the filter, centrifuge for 1 minute at 10,000 rcf, then discard the flow-through and return the filter to the same collection tube.
5. Apply an additional 700 µl of sample to the filter and centrifuge for 1 minute at 10,000 rcf. Continue this process until the entire sample has been passed through the filter.
6. Add 700 µl of Wash 1 solution to the Filter Cartridge and centrifuge for 30 seconds at 10,000 rcf. Discard the flow-through.
7. Add 500 µl of Wash 2/3 solution to the filter and centrifuge for 30 seconds at 10,000 rcf. Discard the flow-through. Centrifuge the Filter Cartridge assembly an additional 30 seconds at 10,000 rcf to remove residual liquid from the filter.

8. Treat the samples in the filters with DNase by combining for each sample:

a. 6  $\mu$ l 10X DNase Buffer

b. 4  $\mu$ l DNase

c. 50  $\mu$ l nuclease-free water

Add this solution to the center of each filter and incubate for 30 minutes at room temperature.

9. Add 700  $\mu$ l of Wash 1 to the filters, incubate for 1 minute at room temperature, then centrifuge for 30 seconds at 10,000 rcf. Discard the flow-through.

10. Add 500  $\mu$ l of Wash 2/3 to the filters and centrifuge for 30 seconds at 10,000 rcf. Discard the flow-through. Add an additional 500  $\mu$ l of Wash 2/3 to the filters, centrifuge for 1 minute at 10,000 rcf, then transfer the filters to new collection tubes.

11. Add 30  $\mu$ l of nuclease-free water heated to 95°C to the center of each filter (**Note 33**). Incubate at room temperature for 1 minute and then centrifuge for 1 minute at 16,000 rcf. Repeat this step using a second 30  $\mu$ l volume of heated nuclease-free water.

12. Read the concentration of the eluted samples on the NanoDrop and run the samples on a Bioanalyzer Pico Chip as described in **Section 3.2.4. – “Assessing RNA Integrity” (Fig. 5)**.

13. Store samples at -80°C.

### **3.8. Quantitative Real-Time Polymerase Chain Reaction Assays Using RNA from**

#### **FFPE**

RNA extracted from FFPE tissues is usually not conducive to amplification using an oligo(dT) primer due to fragmentation and degradation. Therefore, qRT-PCR is performed directly on RNA from FFPE tissues — there is no RNA amplification step.

### **3.8.1. Reverse Transcription of RNA from FFPE**

1. Reverse transcribe the FFPE RNA as described in **Section 3.6.1. – “Reverse Transcription of RNA from Frozen Tissue”** (Steps 2-5).

### **3.8.2. qRT-PCR of RNA from FFPE**

1. Following reverse transcription of the FFPE RNA to cDNA, perform qRT-PCR as outlined for frozen tissue in **Section 3.6.2. – “qRT-PCR of RNA from Frozen Tissue”**, but increase the number of PCR cycles to 60 cycles.
2. Average the  $C_t$  values for each sample and the reference as described above. Calculate the relative transcript levels for each sample using the Comparative  $C_t$  Method.

## **4. Notes**

1. Flash-frozen tissues can only be sectioned after embedding in OCT. To embed a flash-frozen tissue sample in OCT, place the tissue and plastic cryomold in the cryostat set to  $-30^{\circ}\text{C}$ . Cover the bottom of the cryomold with OCT, orient the tissue so that the region of interest is facing up, then fill the mold with additional OCT. Allow the compound to solidify — the block will turn white in color when completely frozen.

2. The foil slides should be room temperature before applying a tissue section. Tissues will not properly adhere to cold slides. Note: If you decide to place more than 1 tissue section on a slide to maximize resources, keep the slide in the cryostat while placing the additional tissue section(s) onto the slide. If the additional section does not adhere to the cold slide, warm the area of the slide where the tissue will be deposited by pressing your thumb to the reverse side. Work as quickly as possible.
3. If the OCT is not completely dissolved after the 50% ethanol wash, make sure that the solution is mixed properly. Shake the original container, add fresh solution, and wash the slide again for 30 seconds.
4. When staining multiple slides per patient sample, change the second 100% ethanol solution when it is noticeably purple in color, indicating that it is no longer a 100% solution. A 100% solution prevents leaching of stain from the slide and allows for faster drying.
5. Incubation at 56°C heat-fixes the section to the slide. The section will fall off the slide during staining if not properly fixed to the slide.
6. Because the slide has no cover-slip, histologic features may not be easily identified. Cellular details can be better visualized by pipetting 100% ethanol onto the tissue section.
7. Large microdissected areas can be lifted directly from the slide with a clean pair of forceps and placed into 100 µl of buffer. We generally perform LM at 4X magnification (after careful inspection at 10X) when working with frozen breast tissue. If the laser tends to drift from the user-defined cut lines, calibrate the laser periodically. With Laser Microdissection LMD software version 4.4, the laser

control settings we typically use at 4X are: aperture, 15; intensity, 44; speed, 6 for clean cuts; offset, 18; bridge, medium; aperture differential, 6.

8. The Lysis Solution and Wash Solutions may be stored at 4°C or room temperature, but be sure to warm the reagents to room temperature before beginning the protocol.
9. The Pico Ladder should only be heated when first diluted and aliquotted (see Step 2 of **Section 3.2.1. – “Reagent Preparation”**) and may not run properly if heated multiple times.
10. Completely submerge the foil containing the microdissected tissue in the Lysis Solution to ensure complete cell lysis and inactivation of endogenous RNases. If this is not achieved by centrifugation, push the foil into the buffer using a pipette tip.
11. To isolate both large and small RNA species including tRNAs or microRNAs, add 129 µl of 100% ethanol rather than 52 µl.
12. When transferring the sample to the filter, be sure not to transfer large pieces of foil that may block the flow of liquid through the filter.
13. To increase accuracy when pipetting hot liquids, pre-wet the pipette tip before pipetting the 95°C water and applying it to the filter.
14. To easily remove the DNase Inactivation Reagent from the RNA and maximize RNA recovery, centrifuge the sample to pellet the Reagent. Transfer ~20 µl of RNA to a new tube, centrifuge the tube containing the Reagent for an additional 30 seconds to 1 minute, then transfer the remaining RNA using a small bore pipette tip (10XL tips work well) without disturbing the DNase Inactivation Reagent pellet.

15. When running a Pico or Nano Chip, ensure that the reagents have not expired and that no more than 2 months have passed since the gel was filtered. Expired reagents can adversely affect the run and can cause aberrations to the baseline of the electropherogram.
16. When cleaning the Bioanalyzer electrodes prior to running a Pico or Nano Chip, fill the electrode cleaner with nuclease-free water, not RNaseZap<sup>®</sup>. The Pico Chip is very sensitive and even minute amounts of RNaseZap<sup>®</sup> can interfere with proper function.
17. When loading the Pico or Nano Chips, do not use the blowout function of the pipette as this may introduce air bubbles that may interfere with the chip running properly.
18. Most RNA samples isolated from frozen tissue via laser microdissection have a RIN (RNA Integrity Number) greater than 7, which is acceptable for downstream applications. Samples with a lower RIN (between 6 and 7) may be usable. Repeat the RNA isolation if significant degradation has occurred. Remove the sample from the study if subsequent isolations do not show improved RNA quality.
19. The pin set on the Bioanalyzer will need periodic maintenance — remove and clean thoroughly with RNaseZap<sup>®</sup>, then rinse thoroughly with nuclease-free water.
20. The final volume of this dilution may be adjusted based on the number of samples that will be amplified at one time. For example, if amplifying 10 or more samples, prepare 40  $\mu$ l of the Fifth Dilution by mixing 4  $\mu$ l of the Fourth Dilution with 36  $\mu$ l of nuclease-free water.

21. We typically use 10 ng of RNA as input for the first round of amplification; however, we have amplified as little as 2 ng with more than sufficient yields of second round aRNA.
22. If the starting amount for amplification is greater or less than 10 ng, adjust the volume of the Poly-A Controls proportionately. For example, if starting with 5 ng of RNA, add 1  $\mu$ l of the Fifth Dilution to the sample.
23. Ensure that the thermal cycler lid is also cooled to 16°C. Exposure of the reaction to higher temperatures will lead to inefficient second-strand cDNA synthesis and will compromise aRNA yield. If the lid temperature cannot be adjusted to 16°C, turn the lid temperature off or incubate with the lid off.
24. We have started the second round of amplification with as little as 350 ng of first round aRNA with successful results. Any remaining aRNA from the first round of amplification can be stored at -80°C for future use.
25. Because the samples have just been heated to 94°C, it is not necessary to heat denature them at 70°C. The heat-denatured ladder from Step 11 of **Section 3.3.4. – “Second Round Amplification”** can be used in this run without reheating. The same gel-dye mix that was prepared to run the Nano Chip for the second round aRNA can be used to run the Nano Chip for the fragmented aRNA if both chips are run on the same day.
26. Before placing the arrays in the hybridization oven overnight, cover each septum with a small piece of tape to prevent the sample from leaking out of the array.
27. Arrange the arrays in the hybridization oven with a balanced configuration to prevent undue stress to the motor.

28. We typically reverse transcribe sufficient aRNA to run all of the TaqMan<sup>®</sup> Gene Expression Assays that we have defined for the study so that additional reverse transcription reactions on the same sample are not necessary.
29. We recommend using TaqMan<sup>®</sup> Gene Expression Assays with the suffix “\_ml” if possible, as these assays amplify regions spanning exon junctions and will not amplify genomic DNA that may contaminate the RNA sample. For “\_g1” or “\_s1” assays, include a control with no reverse transcriptase when performing the reverse transcription reaction and subsequent PCR to ensure that amplification is due to the presence of RNA transcripts and not genomic DNA contamination.
30. 10 ng of cDNA usually produces good results; however, when working with certain low abundance transcripts, the starting amount of cDNA may need to be increased and the volume of water in the master mix adjusted accordingly.
31. In some cases, the chosen reference RNA may not express a particular transcript of interest and an alternative reference RNA will need to be selected.
32. To calculate relative transcript levels using the Comparative C<sub>t</sub> Method:

- a. Determine the  $\Delta C_t$  as follows:

$$\Delta C_t = C_{t \text{ target}} - C_{t \text{ reference}}$$

where the target is the gene of interest and the reference represents an endogenous control such as actin or GAPDH.

- b. Calculate the  $\Delta\Delta C_t$  using the following formula:

$$\Delta\Delta C_t = \Delta C_{t \text{ test sample}} - \Delta C_{t \text{ calibrator}}$$

- c. The fold change relative to the calibrator is  $2^{-\Delta\Delta C_t}$ .

As an alternative to the Comparative C<sub>t</sub> Method, relative levels of gene expression can be determined using the Relative Standard Curve Method. The following document describes both the Comparative C<sub>t</sub> Method and the Relative Standard Curve Method:

Applied Biosystems. Guide to Performing Relative Quantitation of Gene Expression Using Real-Time Quantitative PCR, available at:  
[http://www3.appliedbiosystems.com/cms/groups/mcb\\_support/documents/generaldocuments/cms\\_042380.pdf](http://www3.appliedbiosystems.com/cms/groups/mcb_support/documents/generaldocuments/cms_042380.pdf)

33. This protocol has been modified by the manufacturer. The current RecoverAll™ protocol reduces the volume of Digestion Buffer from 400 µl to 100 µl, shortens the incubation from 3 hours at 50°C to 15 minutes at 50°C followed by 15 minutes at 80°C, reduces the volume of Isolation Additive from 480 µl to 120 µl and the volume of 100% ethanol from 1.1 ml to 275 µl, and lowers the temperature of the eluant from 95°C to room temperature (22-25°C).

## 5. Commercial Vendor Information

1. Affymetrix — <http://www.affymetrix.com>
2. Agilent Technologies — <http://www.agilent.com>
3. Applied Biosystems — <http://www.appliedbiosystems.com>
4. Bio-Rad Laboratories — <http://www.bio-rad.com>
5. Diversified Biotech — <http://divbio.com/>
6. Electron Microscopy Sciences — <http://emsdiasum.com/microscopy/>
7. Eppendorf — <http://www.eppendorf.com>

8. Invitrogen — <http://www.invitrogen.com>
9. Promega — <http://www.promega.com>
10. Sigma-Aldrich — <http://www.sigmaaldrich.com>
11. Thermo Fisher Scientific — <http://www.thermofisher.com>
12. Vector Laboratories — <http://vectorlabs.com/>
13. W. Nuhsbaum — <http://www.nuhsbaum.com/>

### **Acknowledgments**

This work was supported by the United States Department of Defense (Military Molecular Medicine Initiative MDA W81XWH-05-2-0075, protocol #01-20006) and was performed under the auspices of the Clinical Breast Care Project, a joint effort of many investigators and staff members. The opinion and assertions contained herein are the private views of the authors and are not to be construed as official or as representing the views of the Department of the Army or the Department of Defense.

### **References**

1. Collins, F. S. and Barker, A. D. (2007) Mapping the cancer genome. Pinpointing the genes involved in cancer will help chart a new course across the complex landscape of human malignancies. *Sci. Am.* **296**, 50-57.
2. Manolio, T. A., Brooks, L. D., and Collins, F. S. (2008) A HapMap harvest of insights into the genetics of common disease. *J. Clin. Invest.* **118**, 1590-1605.

3. Ellsworth, R. E., Seebach, J., Field, L. A., Heckman, C., Kane, J., Hooke, J. A., et al. (2009) A gene expression signature that defines breast cancer metastases. *Clin. Exp. Metastasis* **26**, 205-213.
4. Perou, C. M., Sorlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., et al. (2000) Molecular portraits of human breast tumours. *Nature* **406**, 747-752.
5. Sørlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., et al. (2001) Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl. Acad. Sci. USA* **98**, 10869-10874.
6. Ellsworth, R. E., Ellsworth, D. L., Deyarmin, B., Hoffman, L. R., Love, B., Hooke, J. A., et al. (2005) Timing of critical genetic changes in human breast disease. *Ann. Surg. Oncol.* **12**, 1054-1060.
7. Becker, T. E., Ellsworth, R. E., Deyarmin, B., Patney, H. L., Jordan, R. M., Hooke, J. A., et al. (2008) The genomic heritage of lymph node metastases: implications for clinical management of patients with breast cancer. *Ann. Surg. Oncol.* **15**, 1056-1063.
8. Ellsworth, R. E., Hooke, J. A., Love, B., Kane, J. L., Patney, H. L., Ellsworth, D. L., et al. (2008) Correlation of levels and patterns of genomic instability with histological grading of invasive breast tumors. *Breast Cancer Res. Treat.* **107**, 259-265.

## Figure Legends

Fig. 1. Staining and laser microdissection of formalin-fixed paraffin-embedded breast tissue containing a ductal carcinoma *in situ* (DCIS). (A) Standard hematoxylin and eosin (H&E) stain of the DCIS on a glass slide. (B) DCIS on a foil slide stained with Cresyl Violet. (C) Breast tissue after removal of the DCIS by laser microdissection.

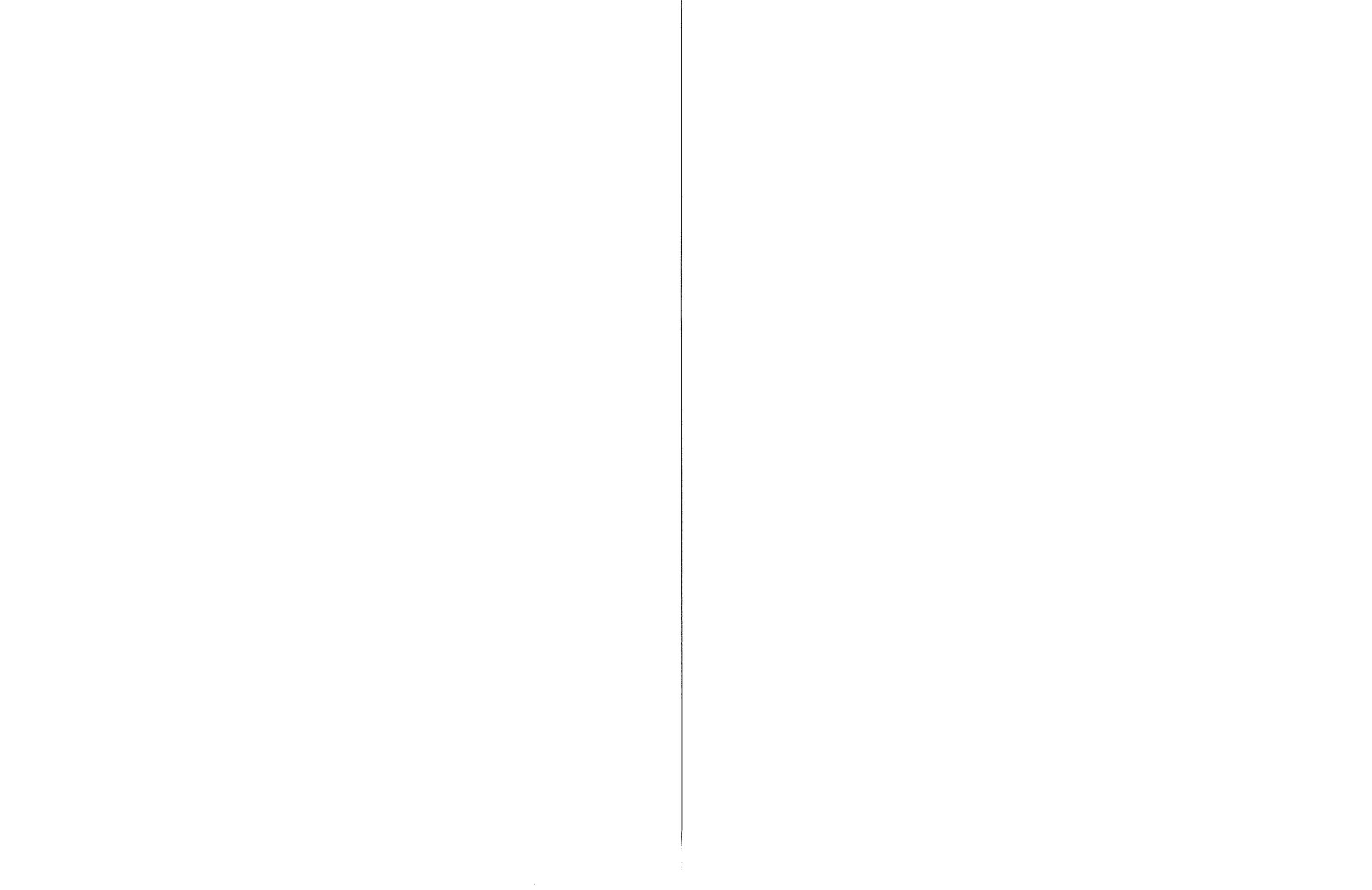
Fig. 2. Electropherogram of total RNA isolated from frozen breast tissue collected via laser microdissection using the RNAqueous<sup>®</sup>-Micro kit. The RNA (RIN = 8.7) was assayed on the Bioanalyzer using a Pico Chip. The 18S rRNA and 28S rRNA peaks are visible near 2,000 and 4,000 nucleotides, respectively.

Fig. 3. Electropherograms of RNA isolated from frozen breast tissue following one and two rounds of amplification. (A) Total RNA amplified using the MessageAmp<sup>™</sup> II aRNA Amplification kit and assayed on the Bioanalyzer using a Pico Chip. (B) Second round aRNA assayed using a Nano Chip. The majority of the second round aRNA product should be >500 nucleotides in length.

Fig. 4. Electropherogram of fragmented, second round aRNA assayed on the Bioanalyzer using a Nano Chip. The aRNA fragments should be ~35 to 200 nucleotides in length.

Fig. 5. Electropherogram of total RNA isolated from formalin-fixed paraffin-embedded breast tissue following laser microdissection. RNA was isolated using the RecoverAll<sup>™</sup>

Total Nucleic Acid Isolation Kit and assayed on the Bioanalyzer using a Pico Chip. Note that the 18S rRNA and 28S rRNA are completely degraded and thus not distinguishable.



Identification of differentially expressed genes in breast tumors from African American compared to Caucasian women

Running title: Genetics of African American breast tumors

Lori A. Field, PhD<sup>1</sup>, Brad Love, PhD<sup>2</sup>, Brenda Deyarmin, HT(ASCP)<sup>1</sup>, Jeffrey A. Hooke, MD<sup>3</sup>, Craig D. Shriver, MD<sup>3</sup>, Rachel E. Ellsworth, PhD<sup>4</sup>

<sup>1</sup>Windber Research Institute, Windber, PA; <sup>2</sup>BioReka, LLC, Timonium, MD, <sup>3</sup>Walter Reed Army Medical Center, Washington, DC; <sup>4</sup>Henry M. Jackson Foundation for the Advancement of Military Medicine

Corresponding Author and Reprints

Rachel E. Ellsworth  
620 Seventh Street  
Windber, PA 15905  
Phone: (814) 467-9844  
FAX: (814) 467-6334  
r.ellsworth@wriwindber.org

Number Text Pages = 25

Number Tables = 4

Number Figures = 3

This research was supported by a grant from the United States Department of Defense (Military Molecular Medicine Initiative MDA W81XWH-05-2-0075, Protocol 01-20006).

There are no financial disclosures for any of the authors.

Condensed abstract:

Molecular analysis of invasive breast tumors as well as non-malignant breast tissues revealed differences in gene expression between AAW and CW. These differences in expression of genes involved in immunity, cell growth and differentiation and invasion and metastasis may contribute to aggressive breast cancer in African American women.

## **ABSTRACT**

**Background:** Breast tumors from African American women (AAW) have less favorable pathological characteristics and higher mortality rates are higher than those of Caucasian women. While socioeconomic status may influence prognosis, biological factors are also likely to contribute to tumor behavior.

**Methods:** Patients with invasive breast cancer were matched by age, grade and ER status; patients with benign disease were matched by age and diagnosis type. RNA from laser microdissected tumors and whole sectioned non-malignant breast tissues was hybridized to HG U133A 2.0 microarrays. Data was analyzed using Partek Genomics Suite using a cutoff of  $P < 0.001$ , 1.5-fold change and results validated by qRT-PCR.

**Results:** Clinicopathological factors did not differ significantly for age at diagnosis, tumor size or stage, lymph node or HER2 status, intrinsic subtype or mortality. Two-way analysis of the tumor specimens revealed 25 probes representing 23 genes differentially expressed between populations; hierarchical clustering classified 24/26 AAW and 25/26 CW correctly. In the non-malignant specimens, 15 probes representing 13 genes were differentially expressed, including five genes that also differed in the tumor specimens; these genes were able to correctly classify non-malignant breast specimens from 20/22 of AAW and all of the CW.

**Conclusions:** Despite matching of tumors by pathological characteristics, molecular profiles differed between AAW and CW in both invasive tumors as well as benign breast tissues. These differentially expressed genes, including *CRYBB2*, *PSPHL* and *SOS1*, are involved in cellular growth and differentiation, invasion and metastasis and immune response and thus may contribute to the poor outcome in AAW.

Keywords: breast cancer, African American, gene expression

## INTRODUCTION

Breast cancer places an exceptional burden on the African American population. Although the overall incidence is not as high as in Caucasian women (CW), the age-adjusted breast cancer mortality rate in African American women (AAW) is highest (33/100,000) of any ethnic group in the United States (1). Contributions to outcome disparities are varied and may include differences in cultural and religious beliefs, perceptions of the medical establishment, and socioeconomic factors, such as lower income and decreased likelihood to have medical insurance, all of which may affect access to quality health care (2). AAW are also more likely to be obese and to have diabetes and less likely to breastfeed, all of which are risk factors for breast cancer (3-5).

The poor prognosis of AAW with breast cancer can also be attributed to biological factors. Breast tumors from AAW are larger, more likely to be poorly differentiated and more likely to be estrogen receptor (ER) negative, all of which are associated with poor prognosis (6). Especially troubling is the higher frequency of triple negative (ER, progesterone receptor (PR) and HER2 negative) in AAW (20-30%) compared to CW (10-13%) as these tumors are not amenable to customized treatments such as tamoxifen, aromatase inhibitors and trastuzumab, leaving chemotherapy as the only treatment option for these patients (7-10). AAW have significantly higher frequency of basal-like breast tumors, a subtype associated with the highest propensity to metastasize to brain and lung, and lowest survival rates (11-13).

The Clinical Breast Care Project (CBCP) is in a unique position to investigate molecular contributions to breast cancer in the African American population. The CBCP has been successful in recruitment of African Americans into research programs,

including the collection of research-quality frozen tissue specimens, such that to date, ~25% of patients enrolled are African American (14). In addition, the primary clinical site of the CBCP is located at Walter Reed Army Medical Center (WRAMC) in Washington DC, a military hospital that provides access to timely and standardized medical care to all military personnel and dependents, regardless of rank, status or ability to pay. The CBCP thus affords the opportunity to identify molecular differences contributing to disease in African Americans within an equal-access health-care setting. To this end, gene expression analysis was performed using RNA from laser-microdissected primary breast tumors from AAW and CW matched by age, grade and ER status, as well as in disease-free breast tissue specimens from AAW and CW without malignancy.

## **METHODS**

### *Tissue Samples*

Diagnosis of every specimen in the Clinical Breast Care Project was performed by a dedicated breast pathologist from hematoxylin and eosin (H&E) stained slides; staging was performed using guidelines defined by the *AJCC Cancer Staging Manual* sixth edition (15) and grade assigned using the Nottingham Histologic Score (16;17). ER and PR status were determined by immunohistochemical (IHC) analysis at a clinical laboratory (MDR Global, Windber, PA) and HER2 status was assayed using the PathVysion® HER-2 DNA Probe kit (Abbott Laboratories, Abbott Park, IL) according to manufacturer's protocols. The CBCP database was queried to identify all female African American patients diagnosed with invasive breast cancer at Walter Reed Army Medical

Center between 2001 and 2007 with frozen tumor specimens available. These patient specimens were then matched by age, grade and ER status to those from CW diagnosed within the same time period. Clinicopathological characteristics for these cases are shown in Table 1.

The database was also queried to identify AAW diagnosed with benign disease, including those patients undergoing reduction mammoplasty or diagnosed after a biopsy with no evidence of disease or benign (e.g., fibrocystic change) disease. Those patients with frozen tissue specimens were age-matched to CW with benign disease.

#### *RNA Isolation, Amplification, aRNA Labeling and Hybridization*

For each case, hematoxylin and eosin stained slides were examined by a breast pathologist and tumor areas marked for laser microdissection. Two to five serial sections (8  $\mu$ m thick) were cut, mounted on glass PEN foil slides (Leica Microsystems, Wetzlar, Germany), stained using the LCM staining kit (Applied Biosystems, Foster City, CA) and laser microdissected on an *ASLMD* laser microdissection system (Leica Microsystems, Wetzlar, Germany). Slide preparation, staining and cutting were performed within 15 minutes to preserve RNA integrity.

RNA was isolated from laser microdissected tumor cells using the RNAqueous-Micro kit (Applied Biosystems, Foster City, CA) and treated with DNase I to remove any contaminating genomic DNA. RNA integrity was assessed using the 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA). RNA was converted to biotin-labeled aRNA using two rounds of amplification with the MessageAmpII aRNA Amplification kit (Applied Biosystems, Foster City, CA), labeled with biotin-11-UTP (PerkinElmer, Waltham, MA) and the concentration and quality of the samples measured with the

NanoDrop 1000 (NanoDrop Products, Wilmington, DE) and the 2100 Bioanalyzer.

Hybridization, washing, staining and scanning were performed using the HG U133A 2.0 arrays (Affymetrix, Santa Clara, CA) according to manufacturer's protocols.

#### *Quantitative Real-time PCR*

For qRT-PCR analysis, 500 ng RNA from each of the original cases plus additional out-of-sample tumors (14 AAW and 15 CW tumor specimens and five disease-free specimens from both AAW and CW) were reverse transcribed using the High Capacity cDNA reverse transcription kit (Applied Biosystems, Foster City, CA). qRT-PCR was performed for those genes with a FDR<0.05, >2-fold change using commercially available TaqMan gene expression assays, except for CRYBB2, for which custom primers were designed. PCR amplification was performed in duplicate using TaqMan Universal PCR Master Mix as recommended by the manufacturer (Applied Biosystems, Foster City, CA) on the iCycler Real-Time Detection System (Bio-Rad, Hercules, CA). GAPDH was used as the endogenous control for normalization of all assays. Relative quantification of gene expression levels was determined using the Comparative C<sub>t</sub> method (18).

#### *Analysis and Statistics*

Affymetix gene expression data was imported into Partek<sup>®</sup> Genomics Suite™ 6.5 (Partek, Inc, St. Louis, MO) as CEL files using default Partek parameters. Raw data was pre-processed, including background correction, normalization and summarization using robust multi-array average (RMA) analysis and expression data log<sub>2</sub> transformed. Principal component analysis was performed to identify outliers and evaluate whether batch effects, ethnicity or intrinsic subtype significantly affected the data.

Given the inherent differences in gene expression between intrinsic subtypes and separation of basal-like and luminal A clusters using PCA analysis (Figure 1), differential expression analysis for the tumor specimens was performed using both one-way (ethnicity alone) and two-way ANOVA (ethnicity and subtype). Intrinsic subtypes were defined as luminal A = ER and or PR+/HER2-; luminal B = ER and or/PR+/HER2+; HER2-enriched = ER and PR-/HER2+ and triple negative = ER, PR and HER2- (11). For the benign specimens, one-way ANOVA was performed using ethnicity as the variable. Gene lists were created using a cut-off of  $P < 0.001$ , 1.5-fold change for both tumor and benign specimens. Second level analysis using a False-detection Rate (FDR)  $< 0.05$ , 2-fold change was performed for both datasets. Hierarchical clustering was performed using the Gene Expression module. For the analysis of the quantitative real-time PCR (qRT-PCR) data, the medians of the  $2^{-\Delta\Delta C_t}$  values for both the primary tumor as well as disease-free specimen data were compared using a Mann-Whitney U test to determine if the relative fold change was significantly different between the two groups.

## RESULTS

### *Tumor characteristics*

The average age at diagnosis did not differ significantly between AAW (51.9 years, range 34-73 years) and CW (51.5 years, range 31-76 years). Tumor characteristics did not differ significantly between the two groups for tumor size, grade or stage, or for hormone receptor, HER2 or lymph node status (Table 1). The frequency of intrinsic subtype did not differ significantly between the two populations. Likewise, mortality rates did not differ significantly with 20% of AAW and 15% of CW having died of

disease after an average time between diagnosis and death of 30 and 41 months in AAW and CW, respectively.

#### *Gene expression data*

Gene expression analysis using only ethnicity as a variable identified 23 probes from 22 genes differentially expressed between AAW and CW; three of which – PSPHL, SOS1, and WDR48 - had a false detection rate (FDR) $<0.05$ ,  $>2.0$  fold-change. Using two-way ANOVA, 25 probes representing 23 genes were differentially expressed between AAW and CW. In this model, CRYBB2, LTF, PSPHL, SOS1 and WDR48 had a FDR $<0.05$ ,  $>2$ -fold-change. Comparison of the two models revealed a common set of 18 genes (Table 2). LTF had the highest fold-change of any gene ( $>8$ -fold higher expression in CW) in both models. Hierarchical clustering based on the 25 probes identified in two-way ANOVA data (Figure 2) resulted in the most accurate classification (94%), misclassifying two AAW and one CW.

#### *qRT-PCR validation*

Sufficient RNA was not available to perform qRT-PCR to validate all genes identified as differentially expressed. Those genes identified in both the one-way and two-way models with  $>2$  fold-change were evaluated by qRT-PCR. Differential expression was confirmed for LTF and PSD3 with  $>3.5$ -fold higher expression in CW and for CRYBB2, PSHPL and SOS1 with  $>2.5$ -fold higher expression in AAW. Using TaqMan assay Hs00961273\_m1, differential expression of WDR48 was not confirmed.

#### *Gene expression analysis of non-malignant breast tissue*

Patients with non-malignant breast cancer either underwent reductive mammoplasty (n=3 AAW and CW) or biopsy (n=19 AAW and CW) and were diagnosed with no

evidence of disease or with benign conditions (Table 3). Average age at diagnosis and pathologic diagnosis did not differ significantly between populations.

Gene expression analysis revealed 15 probes representing 13 genes were differentially expressed between benign breast tissue from AAW and CW (Table 4). Three genes – C4BPA, CRYBB2 and PSPHL - had a FDR  $<0.05$  and  $>2.0$  fold-change in the non-malignant specimens. In addition to CRYBB2 and PSPHL, NPIPL2, SOS1 and WDR48 were also differentially expressed in non-malignant as well as tumor specimens. Hierarchical clustering with these 15 probes was able to correctly classify 91% (20/22) of AAW and 100% of CW (Figure 3).

For the non-malignant breast specimens, genes with FDR  $<0.05$ ,  $>2.0$  fold-change were further evaluated by qRT-PCR. Differential expression was validated by qRT-PCR for CRYBB2 at  $>1.5$ -fold and C4BPA and PSPHL at  $>$  five-fold higher levels in breast tissue from AAW.

## **DISCUSSION**

African American women suffer disproportionately from aggressive breast cancer. To reduce the incidence of developing breast cancer and improve the survival of AAW, it is critical to identify population-specific factors associated with breast cancer. Investigation of epidemiological factors have identified factors including disparities in socioeconomic status affecting access to health care, reproductive factors such as never having breastfed, and anthropomorphic characteristics such as obesity, that are all associated with increased risk of breast cancer (2-5). While these studies can lead to effective interventions such as the encouragement of breastfeeding and maintenance of a

healthy weight in AAW, a complete understanding of the biological differences associated with poor prognosis in AAW is necessary to improve diagnosis and prognosis as well as the development of effective, population-specific therapeutics.

Because tumors in AAW are more frequently diagnosed at a younger age, of higher grade and more frequently ER negative, we evaluated gene expression patterns in primary breast tumors from AAW and CW matched by age, grade and ER status, all treated within an equal-access health care system. PCA did not effectively cluster AAW primary tumor samples from those of CW, suggesting that primary breast tumors from are more similar than different at the gene expression level. Hierarchical clustering with 25 differentially expressed probes, however, was able to classify 94% of patient samples correctly. It should be noted that both misclassified AAW do have significant West African ancestry as confirmed using admixture analysis (unpublished data).

To our knowledge, only one other publication utilized gene expression to identify molecular differences between tumors from AAW and CW. Martin et al. compared gene expression profiles from tumor epithelium from 18 AAW and 17 CW. In this study, PSPHL and CRYBB2 were also found to be expressed at significantly higher levels in tumors from AAW (19). Using only these two genes, Martin et al. found that they could accurately classify 94% of African American and 100% of Caucasian samples used in the original microarray study, and 93% of AAW and 86% of CW of 55 breast tumors evaluated by qRT-PCR. This two-gene classifier was less effective in our data set, in fact, hierarchical clustering using only PSPHL and CRYBB2 resulting in three clusters: one with low PSPHL and low CRYBB2 expression comprised of 16 CW and 2 AAW samples, one with high PSPHL and high CRYBB2 comprised of 15 AAW and one CW

and a mixed cluster represented by high PSPHL and low CRYBB2 composed of nine AAW and nine CW.

Martin et al. also performed gene expression analysis on tumor stroma, collected after laser microdissection from the specimens utilized in their tumor analysis. Two of the 19 most significantly expressed genes in the Martin study, PSPHL and SOS1, were also differentially expressed in our benign breast tissues. It must be noted that the non-tumor tissues used in these two studies are not equivalent. Martin et al. evaluated breast stroma surrounding invasive breast tumors, while our study evaluated non-malignant breast tissue from patients without current or past invasive breast cancer. The breast tumor microenvironment plays an active role in tumorigenesis, and cross-talk between the tumor epithelium and surrounding stroma affects gene expression within the stroma (20). Thus the differences identified in tumor-associated stroma may reflect differences in risk of recurrence or progression between AAW and CW. In contrast, gene signatures in our study the genetic profiles were derived from breast tissues without malignancy and thus may reflect intrinsic differences in breast tissue that may influence the risk or subtype of tumor development.

In a recent study of patients enrolled within the Southwest Oncology Group clinical trials, African American ethnicity was associated with less favorable survival only for sex-specific cancers, such as breast, ovarian and prostate cancer (21). Comparisons of gene expression in tumors from African Americans and Caucasians have been performed for prostate and endometrial cancer, each of which are not only influenced by sex-hormones but also have more aggressive tumors and less favorable outcomes in African Americans compared to Caucasians. In a study comparing gene expression profiles from

laser microdissected prostate tumors from African American and Caucasian men, 47 genes were differentially expressed including, PSPHL and CRYBB2. As with the breast cancer study from Martin et al, the authors evaluated the utility of the two-gene PSPHL and CRYBB2 signature to discriminate prostate tumors from African Americans and Caucasians and found they could correctly classify 91% and 88% of African American and 94% and 100% of Caucasian tumors from the original data set and validation set, respectively (22). A second study of gene expression profiles in primary cell cultures from prostate tumors from 14 African American and 13 Caucasian men found differential expression of PSPHL, CRYBB2, SOS1 and ADI1, each of which were differentially expressed in our breast tumors. In a study evaluating gene expression in endometrial tumors from AAW compared to CW, 16 genes, including PSPHL and SOS1, were found to be differentially expressed ( $P < 0.001$ ); hierarchical clustering based on these genes, however, was ineffective in classifying tumors by ethnicity (23). In contrast, a second study evaluating gene expression profiles of endometrial tumors found that PSPHL was expressed at 9-fold higher levels in tumors from AAW compared to CW (24).

Increased expression of PSPHL is common in sex-specific tumors from African Americans. The function of the L-3-phosphoserine phosphatase homolog (PSPHL) gene is not well understood. PSPHL was originally identified in fibroblasts from patients with Fanconi's anemia (FA) (25) and was later found to be correlated with recurrence of pterygia, a disease of the ocular surface characterized by tumor-like changes including remodeling of the extracellular matrix, changes in expression of cytokines and growth factors, and decreased apoptosis (26). IHC staining for PSPHL revealed protein expression at the superficial epithelium of the primary but not recurrent pterygia,

suggesting that decreased PSPHL is correlated with increasing alterations in the basement membrane.

Expression differences of PSPHL may be attributed to population-specific frequencies of SNP rs6700, located in the 3'UTR of PSPHL. The minor allele frequency (MAF) for rs6700 is higher in African Americans (0.34) compared to Caucasians (0.10) ([http://www.ncbi.nlm.nih.gov/projects/SNP/snp\\_ref.cgi?rs=6700](http://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?rs=6700)) and is associated with higher expression of PSPHL. In fact, PSPHL has the highest natural variation of expression of any gene studied in humans with 8-fold differences between individuals homozygous for the minor allele versus those homozygous for the major allele of SNP rs6700 (27). The increased expression of PSPHL may reflect differences in allele frequencies of rs6700.

Probe 206777\_s\_at was derived from 135 base pair sequence 100% identical to gene sequences from both the crystallin, beta B2 (CRYBB2) and its pseudogene CRYBB2P1. CRYBB2 is one of the major proteins of the vertebrate eye lens, and mutations in CRYBB2 are associated with cataracts (28). Although it is not clear how a lens protein would contribute to breast tumorigenesis, in a rat model, the expression of crybb2 increases in damaged optic nerves, promoting regrowth of axons (29). Thus, it is possible that overexpression of CRYBB2 in AAW may promote or enhance tumor growth.

Son of sevenless, drosophila homolog 1 (SOS1) is a member of the RAS gene family that functions in signal transduction, controlling cell growth and differentiation. SOS1 is an upstream regulator of the Ras/MAPK signaling pathway (30). In prostate cancer cell lines, knockdown of SOS1 resulted in growth inhibition through cell cycle arrest and cell

death. In addition, the SOS1 protein was expressed at higher levels in prostate tumors with high Gleason scores, suggesting that SOS1 is a marker of aggressive tumor biology (31). Higher expression of SOS1 in breast tumors may also contribute to the more aggressive phenotype and poor prognosis in AAW.

The WD repeat domain 48 (WDR48) gene functions as an activator of deubiquitinating enzymes (32). qRT-PCR was performed using the TaqMan assay Hs00961273\_m1 that spans exons 17 and 18 and did not confirm the microarray results. WDR48 is represented by four probes on the U133 2.0 microarrays, all corresponding to the 3' most exon. Probes 222157\_s\_at, 65591\_at, and 221735\_at demonstrated P-values >0.10 and fold changes <1.10. WDR48 is represented by multiple transcripts, at least four of which are protein coding ([http://www.ensembl.org/Homo\\_sapiens/Gene/Summary?g=ENSG00000114742;r=3:39093489-39138155](http://www.ensembl.org/Homo_sapiens/Gene/Summary?g=ENSG00000114742;r=3:39093489-39138155)), thus differences in gene expression for WDR48 may be attributable to a specific alternate transcript.

Lactotransferrin (LTF) had the highest (>8-fold) expression level differences in tumors between AAW and CW. The functions of LTF are diverse and include anti-bacterial and anti-viral activities, regulation of cellular differentiation and growth, and modulation of the immune and inflammatory responses (33). In breast cancer cell lines, LTF was found to induce growth arrest at the G1 to S transition by altering the expression of genes involved in the MAPK and Akt pathways (34;35). As with PSPHL, the MAF for SNP rs1126477 within the LTF gene is significantly different between populations with a MAF of 0.25 for CW and 0.587 in African Americans ([http://www.ncbi.nlm.nih.gov/projects/SNP/snp\\_ref.cgi?rs=1126477](http://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?rs=1126477)). This SNP, which

leads to an A29T alteration of LTF has been associated with aggressive periodontitis in African American but not Caucasian patients (36).

Like PSPHL, the function of the pleckstrin and Sec7 domain containing 3 (PSD3) gene is not well understood; however, decreased expression of PSD3 has been detected in a number of types of cancer (37;38). In addition, decreased expression of PSD3 was detected in metastatic compared to non-metastatic primary breast tumors, suggesting the PSD3 may be a metastasis suppressor gene (39). Finally, the G allele of SNP rs1487745 within the PSD3 gene has been associated with indolent rather than aggressive prostate cancer. Although these results were not validated in other study cohorts, the majority of the latter populations were exclusively or mostly of European descent (40), thus this SNP may, in fact, be associated with aggressive prostate cancer in the African American population. The MAF for rs1487745 is, 0.083 in Caucasians but 0.568 in African Americans (<http://www.ncbi.nlm.nih.gov/projects/SNP/snpref.cgi?rs=1487745>), and these population differences may correlate with differences in gene expression levels.

A number of genes differentially expressed in the breast tumors were also differentially expressed in the non-malignant tissues, including CRYBB2, NPIIPL2, PSPHL, SOS1 and WDR48. Although tempting to speculate that differential expression of these genes may increase the risk of developing aggressive cancer in AAW, within our non-malignant study population, only one AAW has developed malignant disease. This patient had a biopsy at age 34, which was found to be benign (no abnormalities detected). Six years later, she was diagnosed with stage IIIA invasive breast cancer in the contralateral breast. Given the invasive diagnosis was in the contralateral breast and that

none of the other non-malignant patients have progressed to more advanced diagnoses, it is not clear how these genes impact cancer risk.

Non-malignant breast tissues were also characterized by differential expression of genes involved in detoxification of carcinogens, suppression of the RAS/MAPK pathway and immune response. The complement component 4-binding protein alpha (C4BPA) gene, which was expressed at >6-fold higher expression in AAW tissues, encodes the alpha chains of the C4b-binding protein, which inhibits the complement cascade. The complement system has a number of functions, including removal of waste products, protection from pathogens and guidance of the immune response (41). Higher expression of C4BPA may produce a microenvironment susceptible to tumor formation by increased inhibition of the complement cascade.

In conclusion, breast tumors and non-malignant breast tissues demonstrate ethnicity-specific differences in gene expression. Population-specific differences in the MAF for SNPs that have been associated with expression levels and/or disease states for genes such as LTF, PSD3 and PSPHL may be responsible for these genetic differences. Altered expression levels of these genes are involved in cellular growth and differentiation, invasion and metastasis and immune response. These results not only provide new data to account for the biological differences and poor outcome in AAW with breast cancer but also provide new molecular targets for the development of ethnically-appropriate treatments for this disadvantaged population.

## ACKNOWLEDGEMENTS

The opinion and assertions contained herein are the private views of the authors and are not to be construed as official or as representing the views of the Department of the Army or the Department of Defense.

## REFERENCES

1. American Cancer Society. Breast cancer facts and figures: 2009-2010. 2009. Atlanta, GA, American Cancer Society.
2. US Census Bureau. The American Community- Blacks: 2004. 2007. U.S. Department of Commerce, Economics and Statistics Administration.
3. Centers for Disease Control and Prevention (CDC). Racial and ethnic differences in breastfeeding initiation and duration, by state - National Immunization Survey, United States, 2004-2008. *MMWR Morb.Mortal.Wkly.Rep.* 2010;59:327-34.
4. Ogden CL, Carroll MD, Curtin LR, McDowell MA, Tabak CJ, Flegal KM. Prevalence of overweight and obesity in the United States, 1999-2004. *J.A.M.A.* 2006;295:1549-55.
5. Rose DP, Haffner SM, Baillargeon J. Adiposity, the metabolic syndrome, and breast cancer in African-American and white American women. *Endocr.Rev.* 2007;28:763-77.

6. Elledge RM, Clark GM, Chamness GC, Osborne CK. Tumor biologic factors and breast cancer prognosis among white, Hispanic and black women in the United States. *J.Natl.Cancer Inst.* 1994;86:705-12.
7. Amend K, Hicks D, Ambrosone CB. Breast cancer in African-American women: differences in tumor biology from European-American women. *Cancer Res.* 2006;66:8327-30.
8. Stead LA, Lash TL, Sobieraj JE, Chi DD, Westrup JL, Charlot M et al. Triple-negative breast cancers are increased in black women regardless of age or body mass index. *Breast Cancer Res.* 2009;11:R18.
9. Lund MJ, Butler EN, Bumpers HL, Okoli J, Rizzo M, Hatchett N et al. High prevalence of triple-negative tumors in an urban cancer center. *Cancer* 2008;113:608-15.
10. Morris GJ, Naidu S, Topham AK, Guiles F, Xu Y, McCue P et al. Differences in breast carcinoma characteristics in newly diagnosed African-American and Caucasian patients: a single-institution compilation compared with the National Cancer Institute's Surveillance, Epidemiology, and End Results database. *Cancer* 2007;110:876-84.
11. Carey LA, Perou CM, Livasy CA, Dressler LG, Cowan D, Conway K et al. Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. *J.A.M.A.* 2006;295:2492-502.

12. Smid M, Wang Y, Zhang Y, Sieuwerts AM, Yu J, Klijn JG et al. Subtypes of breast cancer show preferential site of relapse. *Cancer Res.* 2008;68:3108-14.
13. Sorlie T, Perou CM, Tibshirania R, Aas T, Geisler S, Johnsen H et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc.Natl.Acad.Sci.USA* 2001;98:10869-74.
14. Ellsworth RE, Zhu K, Bronfman L, Gutchell V, Hooke JA, Shriver CD. The Clinical Breast Care Project: an important resource in investigating environmental and genetic contributions to breast cancer in African American women. *Cell Tissue Bank* 2008;9:109-20.
15. American Joint Committee on Cancer. AJCC Cancer Staging Manual, 6th ed. New York, NY: Springer, 2002.
16. Bloom HJ, Richardson WW. Histological grading and prognosis in breast cancer. *Br.J.Cancer* 1957;11:359-77.
17. Elston CW, Ellis IO. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow up. *Histopathology* 1991;19:403-10.
18. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 2001;25:402-8.

19. Martin DN, Boersma BJ, Yi M, Reimers M, Howe TM, Yfantis HG et al. Differences in the tumor microenvironment between African-American and European-American breast cancer patients. *PLoS ONE* 2009;4:e4531.
20. Allinen M, Beroukhi R, Cai L, Brennan C, Lahti-Domenici J, Huang H et al. Molecular characterization of the tumor microenvironment in breast cancer. *Cancer Cell* 2004;6:17-32.
21. Albain KS, Unger JM, Crowley JJ, Coltman CA, Hershman DL. Racial disparities in cancer survival among randomized clinical trials patients of the Southwest Oncology Group. *J.Natl.Cancer Inst.* 2009;101:984-92.
22. Wallace TA, Prueitt RL, Yi M, Howe TM, Gillespie JW, Yfantis HG et al. Tumor immunobiological differences in prostate cancer between African-American and European-American men. *Cancer Res.* 2008;68:927-36.
23. Ferguson SE, Olshen AB, Levine DA, Viale A, Barakat RR, Boyd J. Molecular profiling of endometrial cancers from African-American and Caucasian women. *Gynecol.Oncol.* 2006;101:209-13.
24. Maxwell, G. L. Gynecologic Cancer Center for racial disparities. W81XWH-05-2-0065. 2008.
25. Planitzer SA, Wachl AW, Rueckels M, Kubbies M. Identification of a novel c-DNA overexpressed in Fanconi's anemia fibroblasts partially homologous to a putative L-3-phosphoserine-phosphatase. *Gene* 1998;210:297-306.

26. Kuo C-H, Miyazaki D, Nawata N, Tominaga T, Yamasaki A, Sasaki Y et al.  
Prognosis-determinant candidate genes identified by whole genome scanning in eyes with pterygia. *Invest.Ophthalmol.Vis.Sci.* 2007;48:3566-75.
27. Morley M, Molony CM, Weber TM, Devlin JL, Ewens KG, Spielman RS et al.  
Genetic variation of genome-wide variation in human gene expression. *Nature* 2004;430:743-7.
28. Graw J. Genetics of crystallins: cataract and beyond. *88* 2009;2:189.
29. Liedtke T, Schwamborn JC, Schroer U, Thanos S. Elongation of axons during regeneration involves retinal crystallin beta b2 (crybb2). *Mol.Cell.Proteomics* 2007;6:895-907.
30. Porfiri E, McCormick F. Regulation of epidermal growth factor receptor signaling by phosphorylation of the ras exchange factor hSOS1. *J.Biol.Chem.* 1996;271:5871-7.
31. Timofeeva OA, Zhang X, Ransom HW, Varghese RS, Kallakruy BV, Wang K et al.  
Enhanced expression of SOS1 is detected in prostate cancer epithelial cells from African-American men. *Int.J.Oncol.* 2009;35:751-60.
32. Cohn MA, Kee Y, Haas W, Gygi SP, D'Andrea AD. UAF1 is a subunit of multiple deubiquitinating enzyme complexes. *J.Biol.Chem.* 2009;284:5343-51.
33. Baker EN. Lactoferrin: a multi-tasking protein par excellence. *Cell.Mol.Life Sci.* 2005;62:2529-30.

34. Damiens E, El Yazidi I, Mazurier J, Duthille I, Spik G, Boilly-Marer Y. Lactoferrin inhibits G1 cyclin-dependent kinases during growth arrest of human breast carcinoma cells. *J.Cell Biochem.* 1999;74:486-98.
35. Ward PP, Paz E, Conneely OM. Multifunctional roles of lactoferrin: a critical overview. *Cell.Mol.Life Sci.* 2005;62:2540-8.
36. Jordan WJ, Eskdale J, Lennon GP, Pestoff R, Wu L, Fine DH et al. A non-conservative, coding single-nucleotide polymorphism in the N-terminal region of lactoferrin is associated with aggressive periodontitis in an African-American, but not a Caucasian population. *Genes Immun.* 2005;6:632-5.
37. van den Boom J, Wolter M, Blaschke B, Knobbe CB, Reifenberger G. Identification of novel genes associated with astrocytoma progression using suppression subtractive hybridization and real-time reverse transcription-polymerase chain reaction. *Int.J.Cancer* 2006;119:2330-8.
38. Pils D, Horak P, Gleiss A, Sax C, Fajjani G, Moebus VJ et al. Five genes from chromosomal band 8p22 are significantly down-regulated in ovarian carcinoma: N33 and EFA6R have a potential impact on overall survival. *Cancer* 2005;104:2417-29.
39. Thomassen M, Tan Q, Kruse TA. Gene expression meta-analysis identifies chromosomal regions and candidate genes involved in breast cancer metastasis. *Breast Cancer Res.Treat.* 2009;113:239-49.

40. Xu J, Zheng SL, Isaacs SD, Wiley KE, Wiklund F, Sun J et al. Inherited genetic variant predisposes to aggressive but not indolent prostate cancer.

*Proc.Natl.Acad.Sci.USA* 2010;107:2136-40.

41. Walport MJ. Complement. First of two parts. *N.Engl.J.Med.* 2001;344:1058-66.

## Figure Legends

Figure 1. Principal component analysis (PCA) of gene expression from 52 primary tumor samples from 26 AAW and 26 CW. Although there were few luminal B (green balls) or HER2- enriched (red balls) primary tumors in this dataset, triple negative (purple balls) and luminal A (blue balls) formed two clusters.

Figure 2. Hierarchical clustering of 52 primary breast tumors. African American samples are noted by red bars, Caucasian by blue bars. Red squares indicate high expression, blue indicate low expression. The misclassified CW (A) was diagnosed at age 37 with a stage IIA, ER and PR negative, HER2 positive, high-grade IDCA. One of the misclassified AAW (B) was diagnosed at age 34 with a stage IIIA, ER, PR and HER2 positive (luminal B) IDCA while the other (C) was diagnosed at age 53 with a stage IIB, ER positive, PR and HER2 negative (luminal A), low-grade ILCA. Both AAW had metastatic lymph nodes while the CW had negative lymph node status.

Figure 3. Hierarchical clustering of 44 patients with non-malignant breast tissue. African American samples are noted by red bars, Caucasian by blue bars. Red squares indicate high expression, blue indicate low expression. One misclassified AAW (A) was diagnosed at age 26 with fibrocystic changes with microcalcifications, while the other (B) was diagnosed at age 29 with fibrocystic changes and mild ductal hyperplasia. Neither woman has had a subsequent breast diagnosis.

Table 1. Clinical and pathological characteristics of 52 primary breast tumors from 26 AAW and 26 CW evaluated by microarray analysis

	AAW (n=26)	CW (n=26)	P-value
<b>Age</b>			0.567
<40 years	15%	8%	
40-49 years	35%	46%	
≥50 years	50%	46%	
<b>Tumor Size</b>			0.942
T1	38%	42%	
T2	54%	50%	
T3	8%	8%	
<b>Tumor Grade</b>			0.773
Well (Grade 1)	12%	19%	
Moderate (Grade 2)	44%	38%	
Poor (Grade 3)	44%	42%	
<b>Hormone Receptor Status</b>			1.000
ER+ and/or PR+	62%	62%	
ER-/PR-	38%	38%	
<b>HER2 Status</b>			1.000
Amplified	15%	12%	
Not amplified	85%	88%	
<b>Subtype<sup>a</sup></b>			0.673
Luminal A	46%	54%	
Luminal B	12%	4%	
HER2 enriched	4%	8%	
Triple negative	38%	35%	
<b>Lymph Node Status</b>			0.762
Negative	68%	62%	
Positive	32%	38%	
<b>TNM Stage</b>			0.493
Stage I	27%	35%	
Stage II	46%	42%	
Stage III	15%	15%	
Stage IV	12%	8%	
<b>Status</b>			0.529
Died of disease	20%	15%	
Died other causes	0%	4%	
Alive with disease	4%	12%	
Alive, disease-free	76%	69%	

<sup>a</sup>Subtype was determined using the surrogate IHC markers: luminal A=ER/PR+, HER2-; luminal B=ER/PR+, HER2+; HER2 enriched=ER/PR-, HER2+, triple negative=ER/PR/HER2-

Table 2. Genes differentially expressed between tumors from AAW and CW. Genes are separated into those identified by one-way (Ethnicity) only, two-way (Ethnicity and subtype) only and those in common between one- and two-way ANOVA. For the genes in common, P-values and fold-change from the two-way analysis are reported. Fold-change is calculated as AAW/CW. Genes in bold had a FDR<0.05, 2-fold change.

Gene Symbol	Accession Number	Gene Name	Probe ID	P-value	Fold-change	Real-time P-value
ADII	NM_018269	Acireductone dioxygenase 1	217761_at	0.00076	-1.88	
C1orf115	NM_024709	Chromosome 1 open reading frame 115	218546_at	0.00067	-1.93	
<b>CRYBB2///CRYBB2P1</b>	<b>NM_000496</b>	<b>Crystallin, beta B2///crystallin beta B2 pseudogene</b>	<b>206777_s_at</b>	<b>4.63E-06</b>	<b>3.57</b>	<b>0.00008</b>
CRYBB2P1	NR_033734.1	Crystallin beta B2 pseudogene	222048_at	0.00027	-1.59	
EIF2S1	NM_004094	Eukaryotic translation initiation factor 2, subunit 1 alpha	201143_s_at	0.00025	-1.53	
GOLPH3L	NM_018178	Golgi phosphoprotein 3-like	218361_at	0.00051	-1.57	
<b>LTF</b>	<b>NM_002343</b>	<b>Lactotransferrin</b>	<b>202018_s_at</b>	<b>1.55E-05</b>	<b>-8.22</b>	<b>0.00018</b>
NP1L2	XM_002344099	Nuclear pore complex interacting protein-like2	221992_at	5.36E-05	2.00	
PRDM10	NM_199438	PR domain containing 10	219515_at	0.00066	-1.55	
PSD3	NM_015310	Pleckstrin and Sec7 domain containing 3	203354_s_at	0.00014	-2.74	0.00207
			203355_s_at	0.00049	-2.89	

PSPHL <sup>b</sup>	AJ001612	L-3-Phosphoserine-phosphatase homologue	218613_at <sup>a</sup>	0.00057	-2.04
RABEP1	NM_004703	Rabaptin, RAB GTPase binding effector protein 1	205048_s_at	3.72E-06	3.97 <0.00001
RPL13	NM_033251	Ribosomal protein L13	203223_at	0.00012	-1.57
RYBP	NM_012234	RING1 and YY1 binding protein	212191_x_at	0.00011	1.58
SOS1	NM_005633	Son of sevenless homolog 1	212777_at	2.96E-07	2.79 0.00107
WDR48	NM_020839	WD repeat domain 48	56919_at	1.43E-06	-2.15 0.64379
ZFP112	NM_001083335	Zinc finger protein 112 homolog	222237_s_at	4.91E-06	-1.73
ZNF395	NM_018660	Zinc finger protein 395	218149_s_at	0.00072	-1.62
<b>Genes identified by two-way ANOVA only</b>					
AR	NM_000044	Androgen receptor	211621_at	0.00025	-1.78
COIL	NM_004645	Coilin	203653_s_at	0.00051	1.53
COL4A5	NM_033380	Type IV collagen alpha 5	213110_s_at	0.00061	-2.13
GPD1L	NM_015141	Glycerol-3-phosphate dehydrogenase 1-like	212510_at	0.00056	-1.61
IGSF3	NM_001007237	Immunoglobulin superfamily, member 3	202421_at	0.00088	-1.56
<b>Genes identified by one-way ANOVA only</b>					

ABLIM1	NM_001003408	Actin binding LIM protein 1	200965_s_at	0.00099	-1.56
INSR	NM_000208	Insulin receptor	213792_s_at	0.00066	-1.53
MRPL48	NM_016055	Mitochondrial ribosomal protein L48	218281_at	0.00031	1.51
SHQ1	NM_018130	SHQ1 homolog	63009_at	2.26E-06	-1.53

<sup>a</sup>The third probe from PSD3 218613\_at was differentially expressed in two-way ANOVA but not one-way

<sup>b</sup>Probe 205048\_s\_at is listed in the Affymetrix database as PSPH but in fact is from PSPHL

Table 3. Clinical and pathological characteristics of 22 patients with non-malignant breast specimens matched by grade and diagnosis

	AAW (n=22)	CW (n=22)	P-value
<b>Age</b>			0.507
<40 years	59%	55%	
40-49 years	23%	14%	
≥50 years	18%	31%	
<b>Benign Diagnosis</b>			0.564
No abnormalities	18%	32%	
Fibrocystic change	36%	36%	
Fibrocystic change w/microcalcifications	14%	18%	
Mild intraductal hyperplasia	14%	9%	
Other	18%	5%	

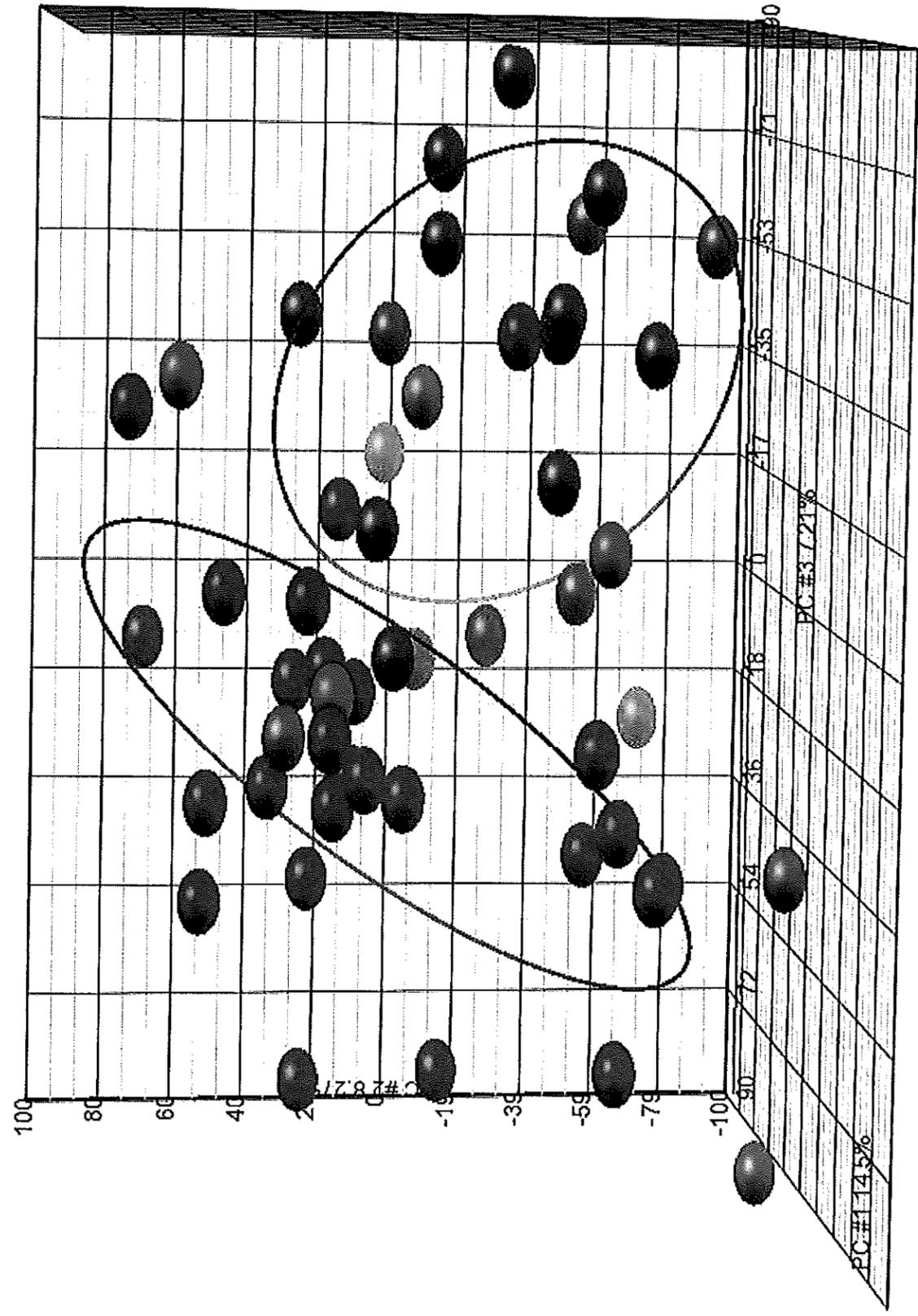
<sup>a</sup>Other includes patients diagnosed with fat necrosis (n=1), lactational metaplasia (n=1) and duct ectasia (n=2) in AAW and radial scar (n=1) in CW

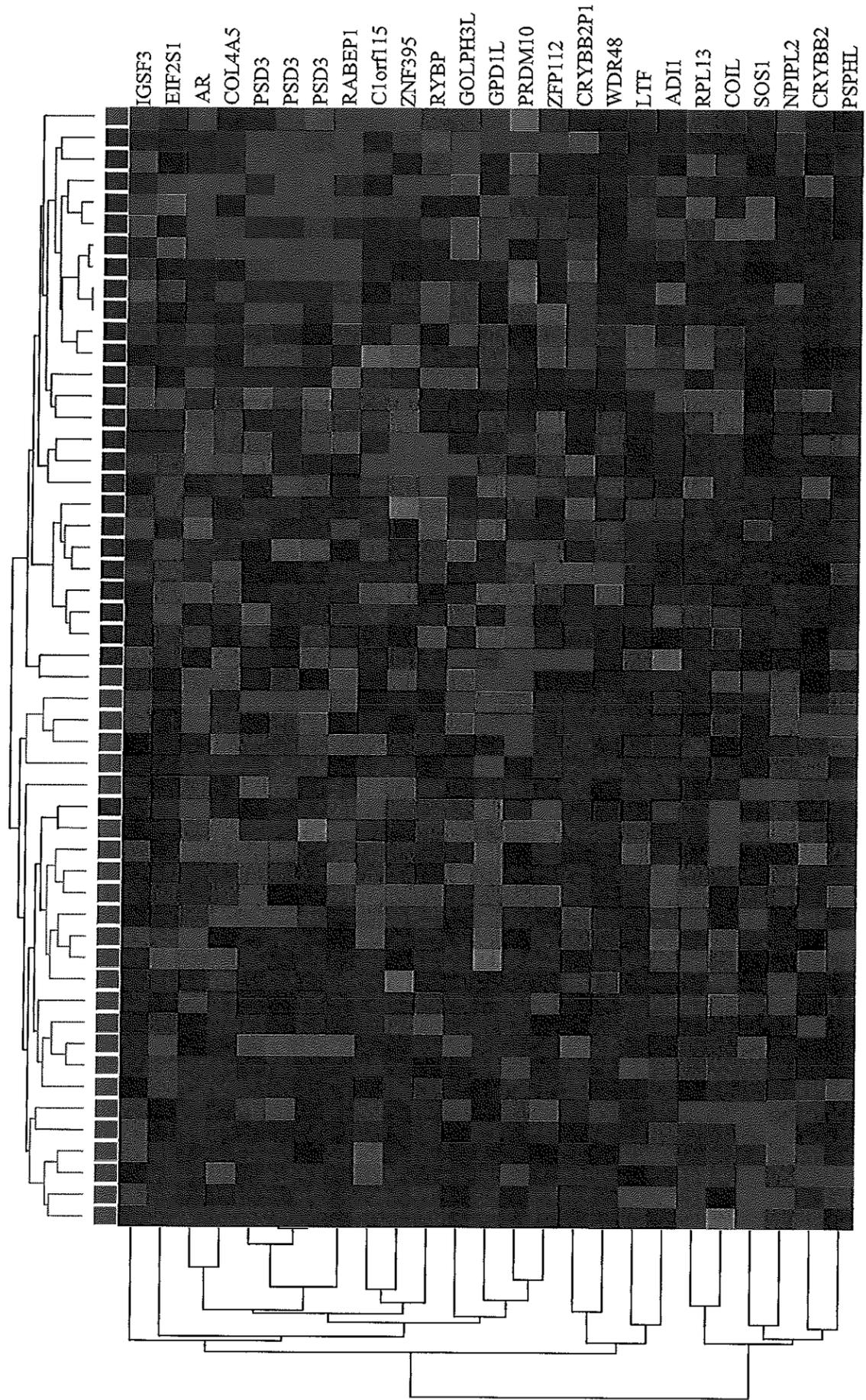
Table 4. Genes differentially expressed between non-malignant tissues from AAW and CW. Fold-change is calculated as AAW: CW. Genes in bold had a FDR < 0.05, 2-fold change. Genes in italics were differentially expressed in both tumors and non-malignant breast tissues from AAW compared to CW. HLA-DPA1 and PI3 were represented by two differentially expressed probes.

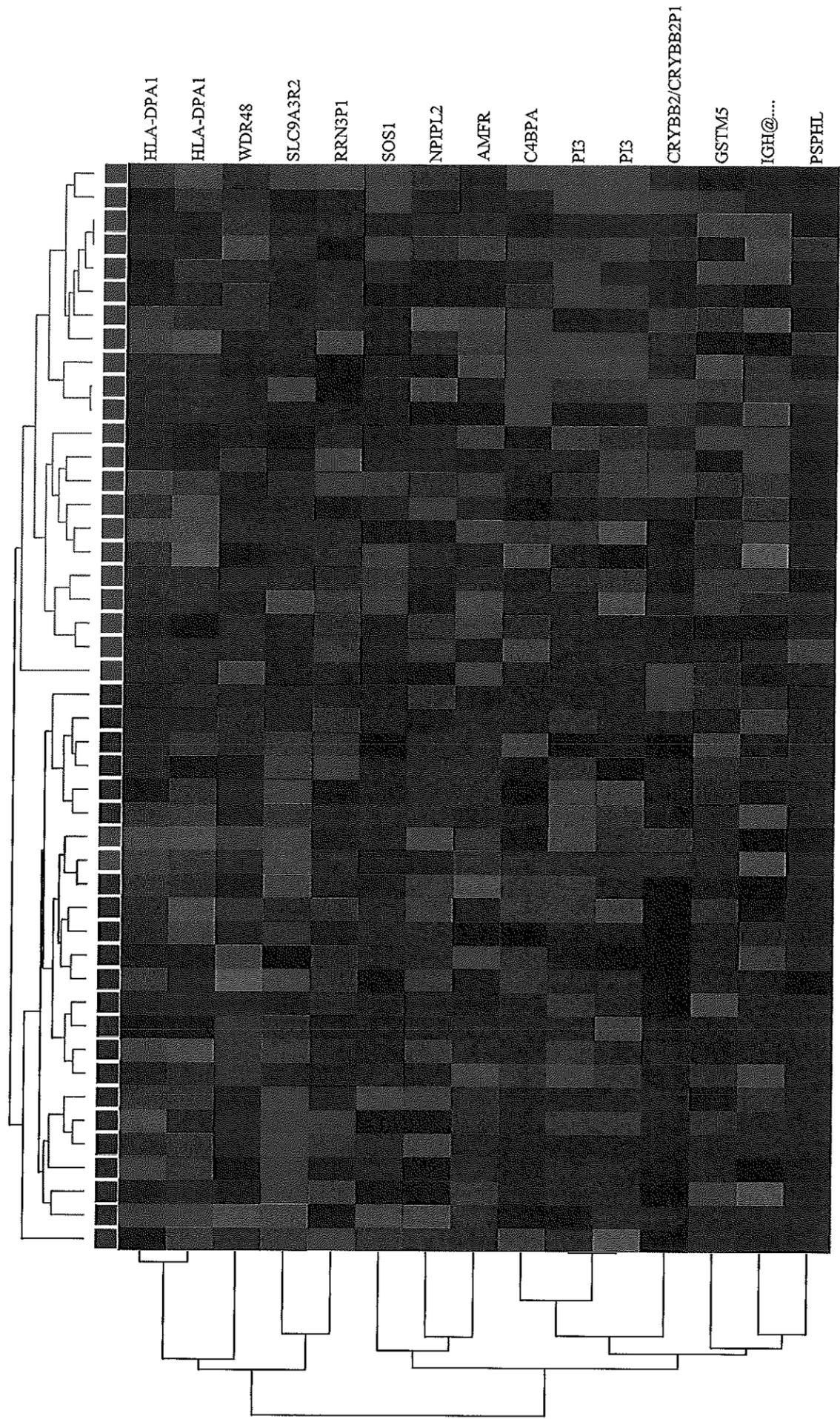
Gene Symbol	Accession Number	Gene Name	Probe ID	P-value	Fold-change	Real-time P-value
AMFR	NM_001144	Autocrine motility factor receptor	202203_s_at	2.73E-05	1.67	
<b>C4BPA</b>	<b>NM_000715</b>	<b>Complement component 4 binding protein, alpha</b>	<b>205654_at</b>	<b>4.41E-06</b>	<b>2.30</b>	<b>&lt;0.00005</b>
<b>CRYBB2//CRYBB2P1</b>	<b>NM_000496</b>	<b><i>Crystallin, beta B2//crystallin beta B2 pseudogene</i></b>	<b>206777_s_at</b>	<b>2.56E-06</b>	<b>2.60</b>	<b>0.04621</b>
GSTM5	NM_000851	Glutathione S-transferase mu 5	205752_s_at	0.00059	1.64	
HLA-DPA1	NM_033554	Major histocompatibility complex, class II, DP	213537_at	0.00072	-2.16	
Immunoglobulin heavy chain locus	XM_001718220/// XM_002348257	Parts of antibodies, mostly variable regions	211990_at 211430_s_at	0.00039 0.00055	-2.49 2.51	
<i>NP1PL2</i>	<i>XM_002344099</i>	<i>Nuclear pore complex interacting protein-like 2</i>	<i>221992_at</i>	<i>0.00012</i>	<i>1.71</i>	
PI3	NM_002638	Elafin preprotein	203691_at	0.00041	1.65	
<b>PSPHL</b>	<b>AJ001612</b>	<b><i>L-3-Phosphoserine-phosphatase homologue</i></b>	<b>205048_s_at</b>	<b>1.68E-07</b>	<b>4.66</b>	<b>&lt;0.00001</b>
RRN3P1	NR_003370	RNA polymerase I transcription factor homolog psuedogene 1	215211_at	0.00027	-1.51	

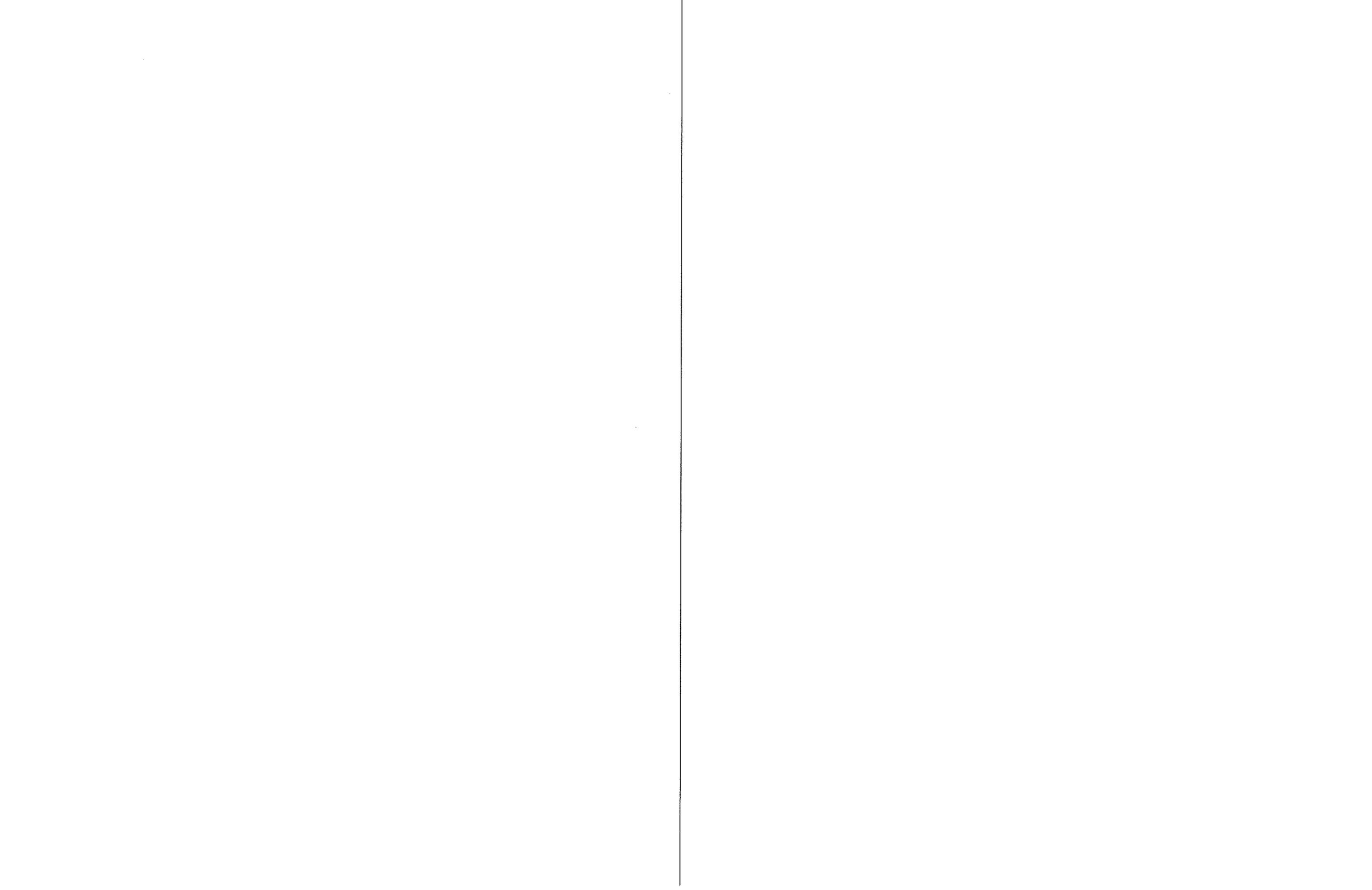
SLC9A3R2	NM_001130012	Solute carrier family 9 isoform 3 regulator 2	209830_s_at	2.50E-06	-1.65
SOS1	NM_005633	<i>Son of sevenless homolog 1</i>	212777_at	1.68E-06	2.14
WDR48	NM_020839	<i>WD repeat domain 48</i>	56919_at	6.49E-05	-1.64

PCA Mapping (30%)









Milburn M MD, Rosman M MD, Mylander C PhD, Liang, Wen DO, Hooke J MD, Kovatich A PhD, Campbell L, Shriver C, M.D., Tafra L MD

Title: Does Tumor Immunohistochemistry Discordance Between Core Needle Biopsy and Mastectomy Raise Concerns for Treatment Recommendations?

Intro: In current practice, core needle biopsies of breast cancers undergo immunohistochemistry (IHC) testing and the results (ER, PR, HER2, and Ki67 expression) direct patient treatment. The purpose of this study was to compare pathology and IHC results from the core needle biopsy of the primary tumor with those of the surgically resected tumor and metastatic lymph nodes.

Methods: This study analyzed prospectively archived tissue. Patients who presented with a diagnosis of invasive breast carcinoma from 06/23/2010 to 10/1/2010 were chosen from a multi-center, IRB-approved, tissue, blood and data banking project. Core needle biopsy was performed and pathological analysis included tumor type and grade, and ER, PR, HER2, and Ki67 status by IHC. After definitive surgical resection of the tumor and lymph nodes, the specimens were examined by site pathology and sections sent to central pathology for independent analysis (pathology and IHC). A breast-only central pathology team supervised the project and ensured strict adherence to protocol which included timed specimen removal, placement on ice to minimize anoxic tissue time, and standardized tissue processing. SAS Inc's JMP version 7 was used for statistical analysis.

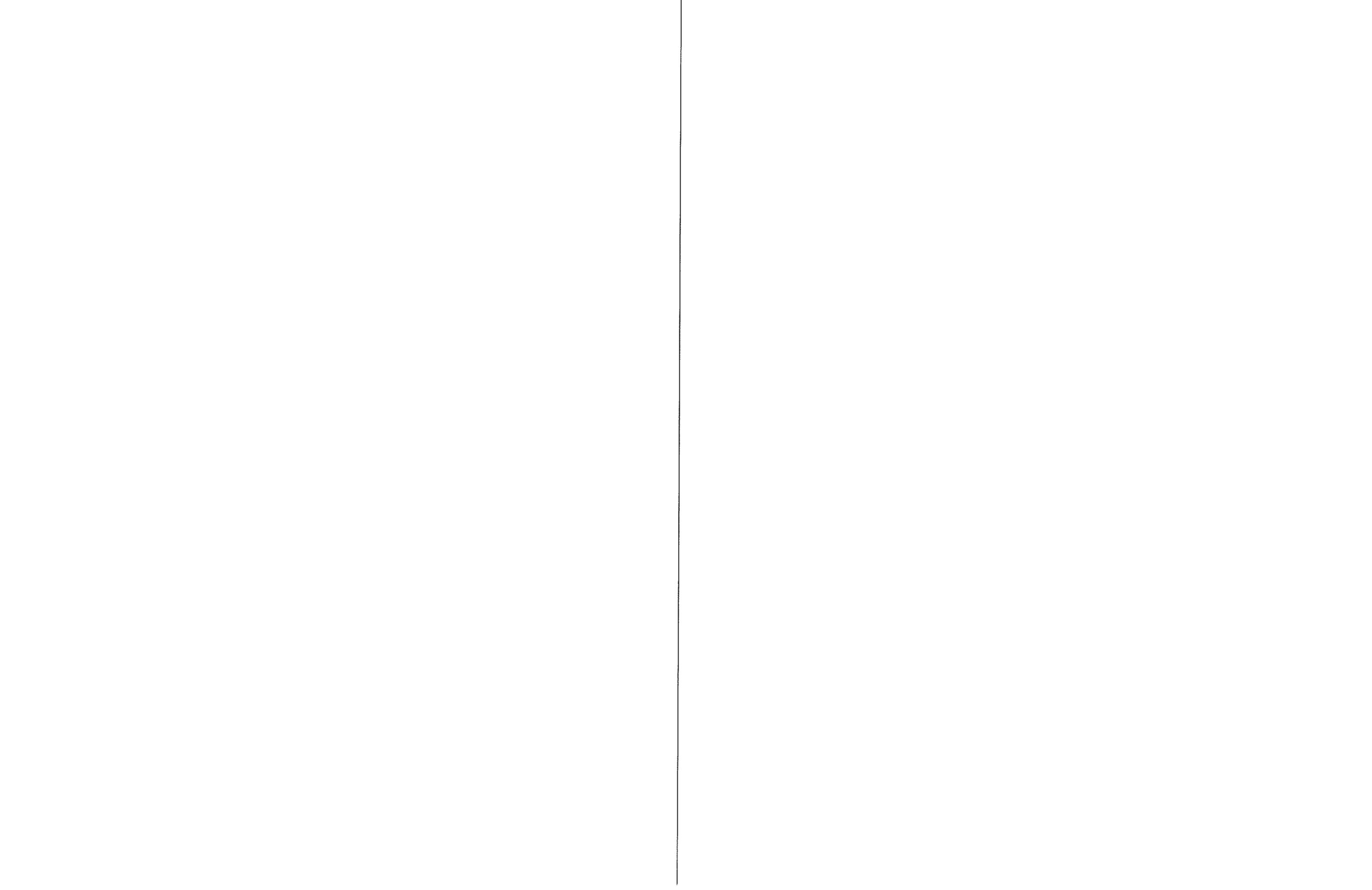
Results: Samples from 26 patients were analyzed, including 12 patients with metastatic lymph nodes. There was excellent concordance of tumor type and substantial agreement in tumor grading between the core biopsy and the surgical resection ( $\kappa=0.71$ ). Only four of the samples differed but only by one grade-level (3 were intermediate- vs. high-grade and 1 was well- vs. intermediate-grade).

There were moderate differences in ER-expression between the core biopsy and the surgical specimens ( $\kappa=0.44$ ). Most significantly, 15% of ER-positive samples on the core biopsy ( $n=4$ ) were ER-negative on the final surgical specimen as seen in Table 1. There was poorer agreement in Ki67 status between the core biopsy and the surgical samples ( $\kappa=0.28$ ). Of note, most of the core biopsies had higher Ki67 percentages, some dramatically higher, than found in the surgical resection as shown in Figure 1.

There was also good concordance of ER, PR, HER2, and Ki67 expression between the 12 surgical resections with their corresponding metastatic lymph nodes.

Conclusion: There was substantial concordance between tumor pathology (type and grade) of the core and surgical specimens and good concordance of IHC results between the surgical resections and their metastatic lymph nodes. Both the tumor and the lymph node surgical specimens were handled similarly and performed at the same time by the same laboratory. A trend towards lower IHC marker values was seen from core biopsy to surgical specimen and could be explained by differences in tissue volume, fixation and process techniques, as well as anoxic tissue time. These data also raise concerns regarding ER and Ki67 discordance between core biopsy and final surgical resection and their relationship to long-term therapeutic management for patients. This may have

potential implications for genomic expression profiles on surgical specimens used for patient decision making.



Identification of CDH1 mutations in patients with lobular carcinoma with and without a family history of gastric cancer

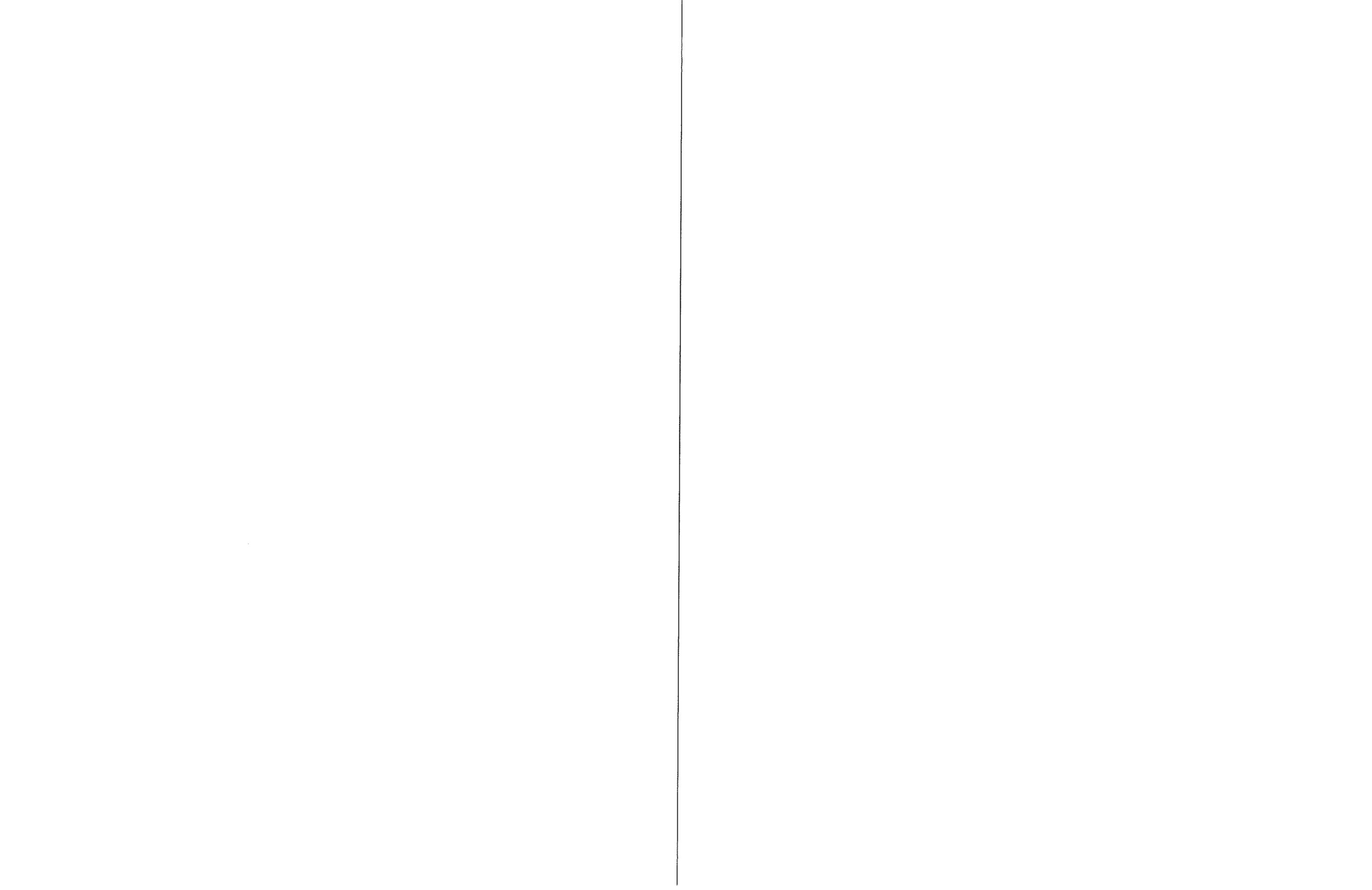
Allyson L. Valente, Rachel E. Ellsworth, Susan Lubert, Charvonne Holliday, Craig D. Shriver

**Background:** The E-cadherin (CDH1) gene is involved in cell adhesion and maintenance of tissue architecture. Loss of expression of CDH1 has been found in both invasive lobular carcinoma (ILCA) of the breast as well as hereditary diffuse gastric cancer (HDGC). While mutations in CDH1 have been found in patients with hereditary diffuse gastric cancer, less is known about the role of CDH1 mutations in patients with breast cancer, thus mutation status was determined in a breast cancer cohort with and without family histories of gastric cancer.

**Methods:** The Clinical Breast Care Project database was queried to identify all patients with ILCA with or without a family history of gastric or stomach cancer as well as those with non-lobular invasive cancer with a family history of gastric or stomach cancer. Genomic DNA was isolated from peripheral blood samples. DNA variants were detected for each exon using high-resolution melting technology and the underlying change identified by direct sequencing.

**Results:** Of the 72 patients with ILCA, five had a first degree relative with stomach cancer; 14 patients with non-lobular invasive cancer (12 IDCA, one mixed lobular and ductal, and one tubular) had first degree relatives with stomach or gastric cancer. Three mutations previously associated with gastric cancer were identified: A617T and V832M were identified in a patient with ILCA with no history of gastric cancer and in a patient with IDCA with a family history of gastric cancer, respectively, while one patient with ILCA and no family history harbored the splice site mutation 1137G-T. Three novel variants were also identified: G879S, in a patient with ILCA and no family history, P825Q in a patient with mixed lobular and ductal features with a family history of stomach cancer and L230F in a patient with IDCA and a family history of stomach cancer.

**Conclusions:** The frequency of CDH1 mutations in this patient population was low (~7%) and were split between patients with ILCA and no family history of gastric cancer and those with non-lobular cancers and at least one family member with gastric or stomach cancer. Identification of known mutations in patients with non-lobular carcinomas suggests that while these specific CDH1 mutations are associated with increased risk of hereditary diffuse gastric cancer, they do not confer increased risk for lobular cancer. In those patients with ILCA, the low frequency of germline mutations in CDH1 suggests alternate methods, such as epigenetic modification or miRNA silencing, may be driving the suppression of expression of E-cadherin and the development of ILCA.



**Loss of Nuclear Localized and Tyrosine Phosphorylated Stat5: A Predictor of Poor Clinical Outcome and Increased Risk of Antiestrogen Therapy Failure in Breast Cancer.** Amy R. Peck<sup>1</sup>, Agnieszka K. Witkiewicz<sup>2</sup>, Chengbao Liu<sup>1</sup>, Ginger A. Stringer<sup>3</sup>, Alexander C. Klimowicz<sup>4</sup>, Edward Pequignot<sup>3</sup>, Boris Freydin<sup>3</sup>, Thai H. Tran<sup>1</sup>, Ning Yang<sup>1</sup>, Anne L. Rosenberg<sup>5</sup>, Jeffrey A. Hooke<sup>6</sup>, Albert J. Kovatich<sup>7</sup>, Marja T. Nevalainen<sup>1</sup>, Craig D. Shriver<sup>6</sup>, Terry Hyslop<sup>3</sup>, Guido Sauter<sup>8</sup>, David L. Rimm<sup>9</sup>, Anthony M. Magliocco<sup>4</sup>, and Hallgeir Rui<sup>1</sup>

<sup>1</sup>Department of Cancer Biology, Kimmel Cancer Center, Philadelphia, Pennsylvania

<sup>2</sup>Department of Pathology, Thomas Jefferson University, Philadelphia, Pennsylvania

<sup>3</sup>Division of Biostatistics, Thomas Jefferson University, Philadelphia, Pennsylvania

<sup>4</sup>Department of Oncology, Tom Baker Cancer Centre, Alberta Cancer Board, Alberta, Canada

<sup>5</sup>Department of Surgery, Thomas Jefferson University, Philadelphia, Pennsylvania

<sup>6</sup>Walter Reed Army Medical Center, Washington, DC, USA

<sup>7</sup>MDR Global Systems, LLC, Windber, Pennsylvania

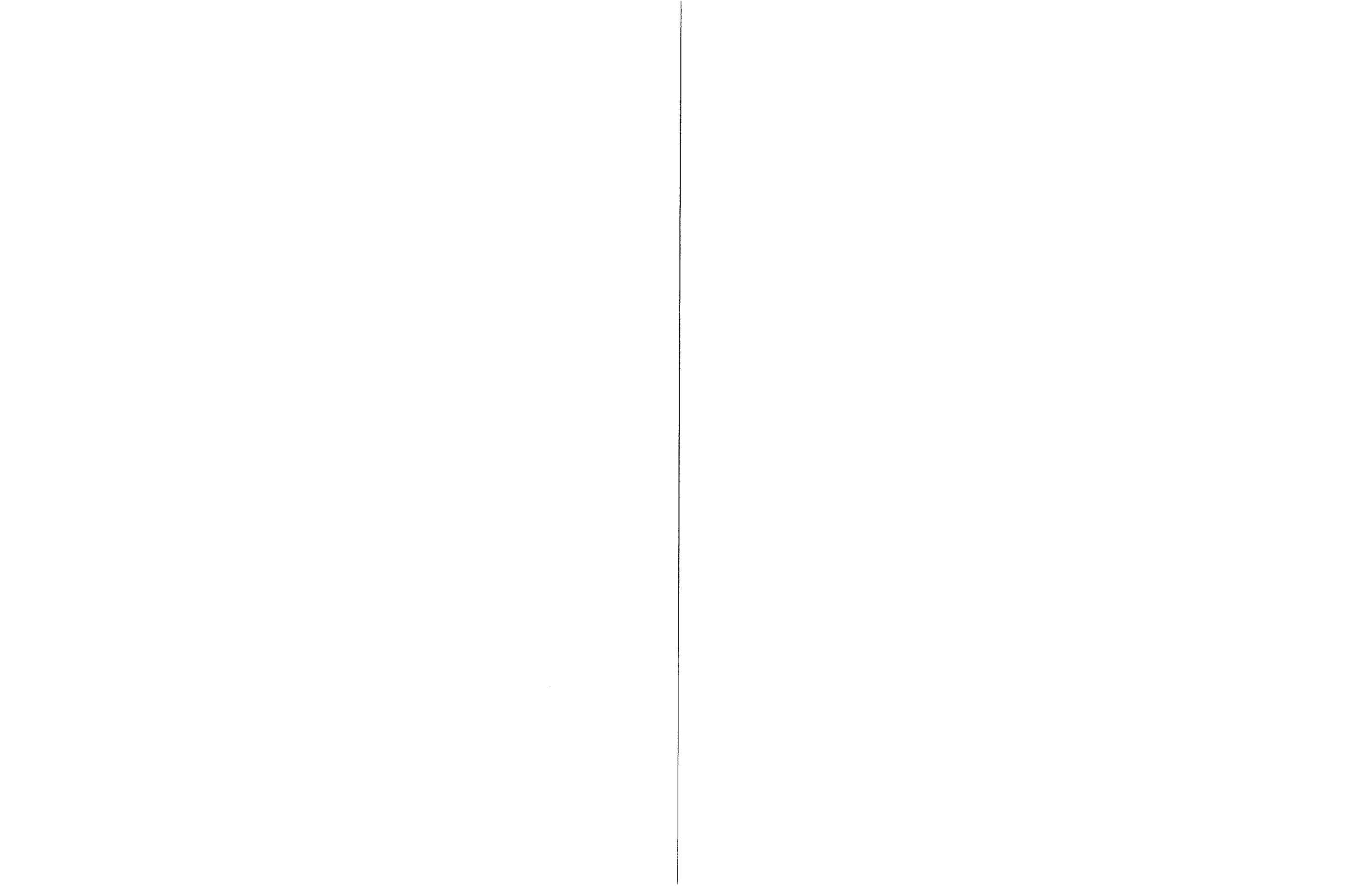
<sup>8</sup>Department of Pathology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

<sup>9</sup>Department of Pathology, Yale University School of Medicine, New Haven, Connecticut

Previous work based on tissue microarrays has suggested that nuclear localized and tyrosine phosphorylated Stat5 (Nuc-pYStat5) is a marker of prognosis in node-negative breast cancer. The purpose of this study was to validate the prognostic value of Nuc-pYStat5 in whole tissue sections and expand the analyses to a quantitative immunofluorescence-based assay both for prognosis and to explore Nuc-pYStat5 as a predictor of response to antiestrogen therapy. Levels of Nuc-pYStat5 were analyzed in five archival materials of breast cancer by traditional diaminobenzidine-chromogen immunostaining and pathologist scoring of whole tissue sections or by immunofluorescence and automated quantitative analysis (AQUA) of tissue microarrays. Nuc-pYStat5 was an independent prognostic marker as measured by cancer-specific survival (CSS) in node-negative breast cancer patients that did not receive systemic adjuvant therapy, when adjusted for common pathology parameters in multivariate analyses both by standard chromogen detection with pathologist scoring of whole tissue sections (Material I; n=233) and by quantitative immunofluorescence of a tissue microarray (Material II; n=291). Two distinct monoclonal antibodies gave corresponding results. Progression Material III (n=130) revealed frequent loss of Nuc-pYStat5 in invasive carcinoma compared to normal breast epithelia or DCIS, and general loss of Nuc-pYStat5 in lymph node metastases. In Material IV (n=221), loss of Nuc-pYStat5 was associated with increased risk of antiestrogen therapy failure as measured by CSS and time to recurrence (TTR). More sensitive AQUA quantification of Material V (n=91) of antiestrogen treated patients identified by multivariate analysis a cohort with low Nuc-pYStat5 at elevated risk for therapy failure (CSS HR=21.55 (5.61,82.77),  $p<0.001$ ; TTR HR=7.30 (2.34,22.78),  $p=0.001$ ). We conclude that Nuc-pYStat5 is an independent prognostic marker in node-negative breast cancer. If confirmed in prospective studies, Nuc-pYStat5 may become useful as a predictive marker of response to adjuvant hormone therapy.

Supported by NIH grants R01-CA101841 and R01-CA118740 (H.R.), Komen for the Cure Promise Grant KG091116 (H.R., A.K.W., C.L., J.A.H., A.J.K., C.D.S., T.H.), DOD CDMRP Predoctoral Fellowship (A.R.), and NCI Support Grant 1P30CA56036 to the Kimmel Cancer Center. The Project is funded, in part, under a Commonwealth University Research Enhancement Program grant with the Pennsylvania Department of Health (H.R.). The Department specifically disclaims responsibility for any analyses, interpretations or conclusions. The views expressed in this article are those of the authors and do not reflect the official policy of the Department of the Army (DOA), Department of Defense (DOD), or US Government.

\*Address correspondence to: Hallgeir Rui, Department of Cancer Biology, Thomas Jefferson University, 233 S. 10<sup>th</sup> St, Philadelphia, PA 19107. [hallgeir.rui@jefferson.edu](mailto:hallgeir.rui@jefferson.edu)



## The role of *PSPHL* in breast cancer in African American women.

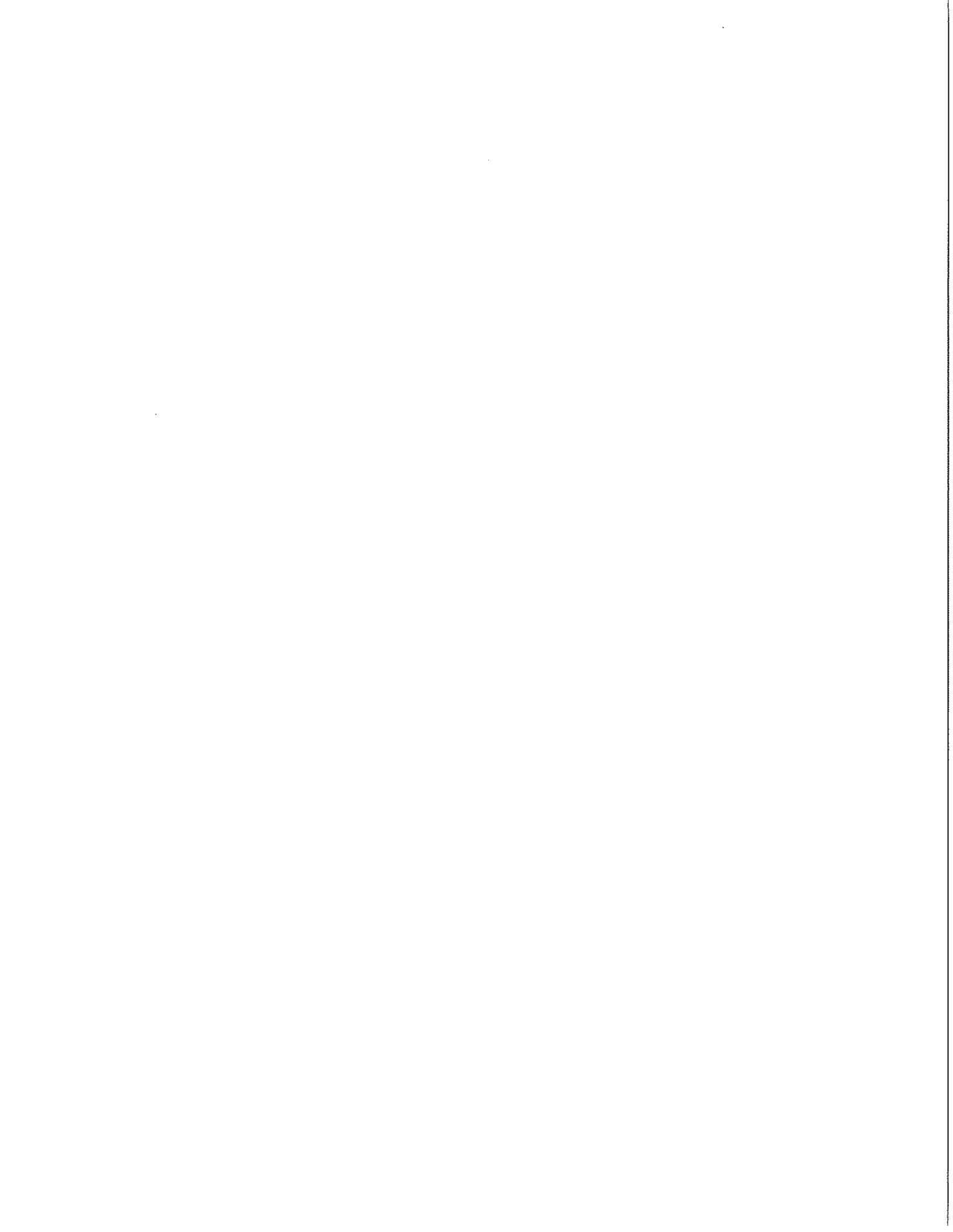
Field L, Rummel S, Shriver CD, Ellsworth RE

**Background:** Younger age at diagnosis and poorer prognosis in African American women (AAW) with breast cancer have led to a number of studies evaluating gene expression differences in tumors from AAW compared to those from Caucasian women (CW). Studies have found that *PSPHL* expression is significantly higher not only in breast but also in prostate and endometrial tumors from African Americans compared to Caucasians. SNP rs6700 in the 3' UTR of *PSPHL* has been found to correlate with expression. Here, we investigated how rs6700 affects *PSPHL* expression levels in AAW and CW with and without breast cancer.

**Methods:** Genomic DNA was isolated from 74 women with (35 AAW and 39 CW) and 37 women without breast cancer (23 AAW and 14 CW). rs6700 was genotyped using TaqMan SNP Genotyping assay C\_\_314922\_10. *PSPHL* expression levels in breast tissue were measured in these same women using TaqMan Gene Expression assay Hs00863464\_m1 using RNA isolated from laser microdissected breast tumor cells or normal breast tissue from women with or without breast cancer, respectively.

**Results:** *PSPHL* expression in breast tumors from AAW was significantly higher than in CW ( $P = 6 \times 10^{-6}$ ). The median fold change in expression in AAW was 42.2; while only 2 (6%) AAW had no detectable expression of *PSPHL*, 21/39 (54%) CW did not express *PSPHL*. The minor allele frequency (T) in AAW with breast cancer was 0.41 compared to 0.06 in CW with breast cancer. In both AAW and CW, median levels of *PSPHL* expression were lowest (15.45 in AAW, 0 in CW) in women with the C/C genotype. In addition, all patients whose tumors had no detectable *PSPHL* expression had the C/C genotype. In women without breast cancer, the C/C genotype was also correlated with lower *PSPHL* expression and was more frequent among CW compared to AAW (79% and 39%, respectively).

**Conclusions:** *PSPHL* is more highly expressed in both diseased and normal breast tissue from AAW compared to CW and expression of *PSPHL* correlated with rs6700 genotypes. Although the function of *PSPHL* is unknown, it may function in cellular proliferation. Higher expression in AAW may, therefore, promote the development of tumors in AAW. It must be noted, however, that the frequency of the C allele is significantly lower in the African American population as a whole (53% vs. 92%); therefore, differences described here may reflect population differences rather than any specific contribution to breast cancer etiology. However, due to the higher expression of *PSPHL* in African Americans with other hormone-dependent cancers, its role in tumorigenesis requires further study.



## Breastfeeding, Baby Boomers and Breast Cancer

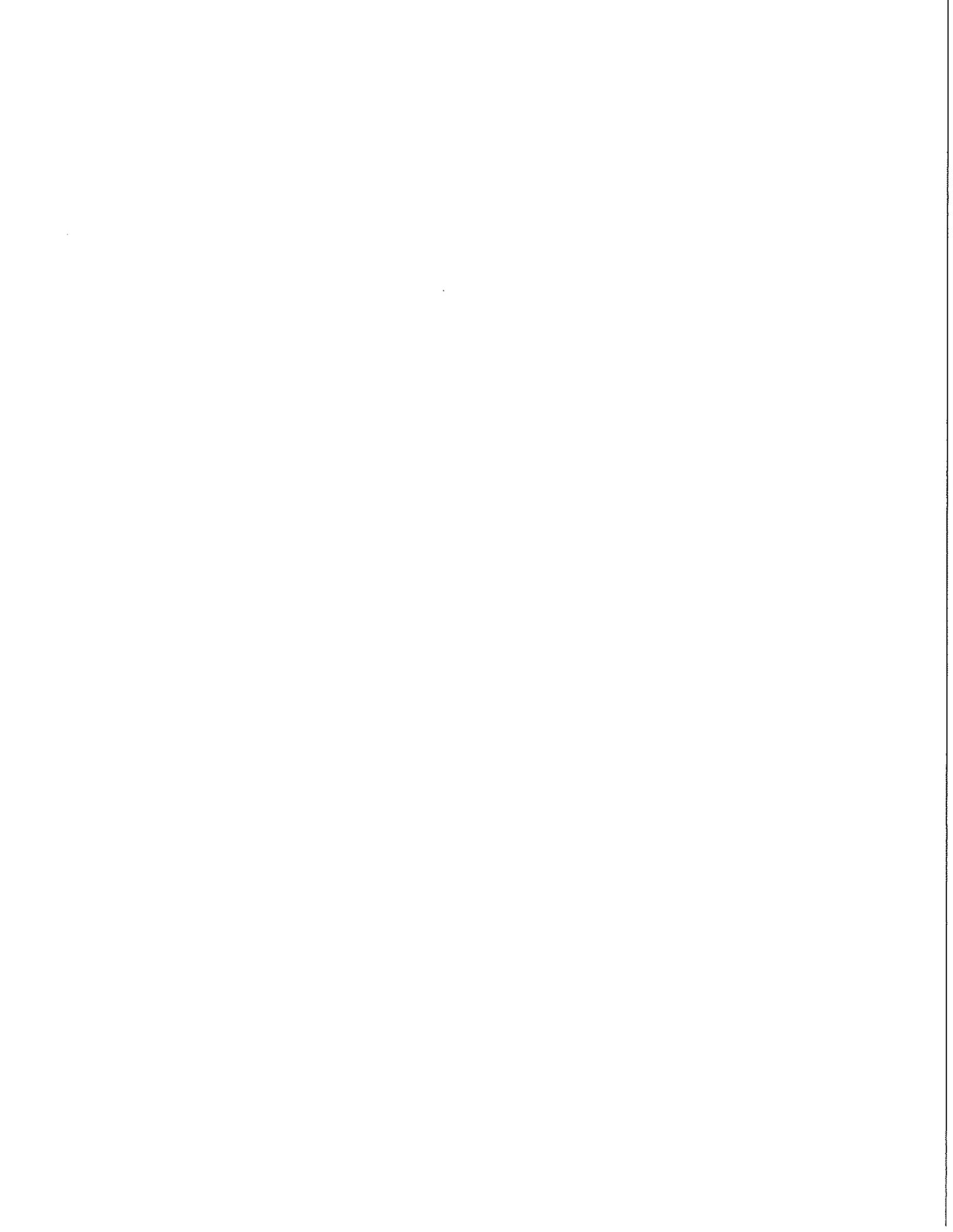
RE Ellsworth, AL Oviedo, JL Kane, CD Shriver

**INTRODUCTION:** Breastfeeding has been associated with an overall decreased risk of developing breast cancer such that for every 12 months of breast feeding, the relative risk of breast cancer decreases by 4.3%. In addition, the protective effect of breastfeeding has been associated with decreased risk of triple negative and estrogen-responsive breast cancer. Although the mechanism by which breastfeeding provides a protect effect is not well-defined, physical changes in the mammary epithelium reflecting maximal differentiation, may be involved.

**METHODS:** The database of the Clinical Breast Care Project was queried to identify all patients born between 1946 and 1964 (“Baby Boomers”) with invasive breast cancer who had given birth to at least one living child. Clinicopathologic characteristics were compared between patients who breastfed (both ever and those who breastfed at least 6 months) and those who never breast fed using chi-square and Fisher’s exact tests. Gene expression data from microdissected tumors was generated using U133A 2.0 microarrays and analyzed using the Partek Genomics Suite.

**RESULTS:** Of the 188 women who ever breastfed, 135 (72%) breastfed >6 months; an additional 120 parous women born within the same time period never breastfed. Age at diagnosis and menarche did not differ between groups; however, women who breastfed gave birth at a later age (25.2 vs. 22.3 years) and had more children (2.47 vs. 2.09). No differences were detected between patients who breastfed (either ever or >6 months) and those who did not for any tumor characteristics, including hormone and HER2 status, grade, size, lymph node status or mortality. At the molecular level, significant gene expression differences were not detected between tumors from patients who breastfed and those who did not. When evaluated by tumor subtype, however, significantly more ( $P<0.009$ ) patients who had never breastfed (50%) had basal-like breast tumors compared to those who had (14%).

**CONCLUSIONS:** Although a molecular signature that effectively discriminated tumors from patients who breastfed from those who did not was not identified, never breastfeeding was associated with an increase in basal-like breast tumors. These data suggest that the beneficial effects of breastfeeding extend beyond decreased risk of developing breast cancer but provide a protective effect against aggressive breast disease as well.



# **The Potential Role of MMPs as Prognostic Markers in Breast Cancer**

**Jing Zhou<sup>1</sup>, Richard Mural<sup>1</sup>, Craig D. Shriver<sup>2</sup> and Stella B. Somiari<sup>1</sup>**

<sup>1</sup>Clinical Breast Care Project, Windber Research Institute, 620 7th Street, Windber, PA 15963, USA

<sup>2</sup>Clinical Breast Care Project, Walter Reed Army Medical Center, 6900 Georgia Avenue, NW Washington, DC 20307, USA

Address correspondence to [s.somiar@wriwindber.org](mailto:s.somiar@wriwindber.org)

## **ABSTRACT**

The metalloproteinase enzymes are involved in tumor development and progression. Aberrant metalloproteinase expression has been implicated in breast cancer onset, progression and prognosis. Here we review the expression profile of some of the matrix metalloproteinase (MMP) enzymes and discuss their expression pattern and role in breast cancer. Since the polymorphisms of these MMPs have also been studied as possible disease markers, we discuss their potential clinical usefulness as predictors of breast cancer/breast cancer severity. The review has special emphasis on three of the most commonly studied MMPs; MMP-1, MMP-2, and MMP-9.

## **1. Introduction**

The Extracellular matrix (ECM) including both the interstitial matrix and the basement membrane are composed of collagens, elastins, gelatin, matrix glycoproteins, and proteoglycans. Extracellular matrix plays an essential role not only in physical processes such as growth, wound healing and fibrosis but also in pathological events such as tumor invasion and metastasis. Metastasis often involves the destruction of the ECM by enzymes such as serine and threonine proteases, and matrix metalloproteinases [1].

The matrix metalloproteinases (MMPs) are a large family of calcium-dependent zinc-containing endopeptidases, which are very important enzymes that actively participate in remodeling the ECM by degrading certain of its constituents along with promoting cell proliferation, migration, differentiation, apoptosis and angiogenesis [2-6]. The matrix metalloproteinases consist of soluble and membrane type MMPs. Based on the specificity of their substrates, the soluble MMPs are classified into collagenases (MMP-1, MMP-8 MMP-13, and MMP18), gelatinases (MMP-2, MMP-9), stromelysins (MMP-3, MMP-10, MMP-11, MMP-19, MMP-27), matrilysins (MMP-7, MMP-26) and others (MMP-12, MMP-20, MMP-23, MMP-28). They share common structural motifs including a pro-peptide domain, a catalytic domain, a hinge region and a hemopexin-like domain, which may be responsible for the macomolecular substrate recognition as well as for interaction with tissue specific inhibitors of MMPs (TIMPs). Membrane-type MMPs (MMP-14, MMP-15, MMP-16, MMP-17, MMP-24, and MMP-25), on the other hand are associated with the cell membrane through their additional transmembrane and intracellular domain, the membrane linker domain [7-9].

In breast cancer, the spreading of the tumor to the distant organs (bone, liver, kidney, thyroid, and rarely brain) constitutes the main cause of the lethal cases. Thus, early detection of a breast tumor can play an important role in decreasing the mortality rate. Pathological features like tumor size, differentiation grade, lymph node status and molecular markers such as estrogen and progesterone receptor level, HER-2/neu gene amplification, Bcl-2 or p53 over expression, and BRCA1 and BRCA2 mutation have been utilized for breast cancer diagnosis. However, the identification of more sensitive markers for early detection of breast cancer, the detection of micrometastasis and the determination of the appropriate disease prognosis remain the key to eliminating the mortality associated with this disease. As a potential metastasis indicator, MMPs have been recently studied in human and animal models and in breast cancer cell lines [10-15]. Increased expression of MMPs in breast tumors and elevated levels in the plasma of breast cancer patients has been reported [16-18]. This review will discuss current findings that provide new insights into the prognostic role of MMPs in breast cancer and the role of MMPs in breast cancer progression. The potential usefulness of MMPs as clinical indicators of breast cancer severity and aggressiveness will be reviewed including their role in the prediction and prognosis of breast cancer. In view of the large number of MMPs that have been studied, we will streamline this review by focusing more on MMP-1, MMP-2 and MMP-9.

## **2. Expression pattern of MMPs and their correlation with breast cancer**

Matrix metalloproteinases, are important mediators of ECM degradation, and have been implicated in tumor migration and metastasis of almost all human cancers [19-29]. Both the

MMPs and their inhibitors, the tissue inhibitors of matrix metalloproteinases (TIMPs) have been extensively investigated in breast cancer [13, 14, 30-36]. In these studies, increased expression of certain MMPs is often correlated with tumor progression, increased invasiveness and poor prognosis. For example, tissue expression of MMP-1, MMP-2, MMP-9, MMP-11, membrane type MMP-1 (MMP-14), TIMP-1 and TIMP-2 have been correlated with poor outcome in breast cancer patients [37-44]. In the MDA-MB-231 xenograft model of breast cancer, MMP-1 and MMP-2 have been identified as potential gene signatures for lung metastasis of primary breast tumor [45, 46]. Tumor-derived, rather than stromal fibroblast-derived MMP-13 has also been associated with aggressive breast tumor types and poor overall survival of breast cancer patients [47]. Recently, Köhrmann's group systemically analyzed the expression of all known human MMPs at both the mRNA and protein level in primary human breast cancer and breast cancer cell lines and found stronger expression of MMP-1, 2, 8, 9, 10, 11, 12, 13, 15, 19, 23, 24, 27 and 28 in breast cancer tissue compared to normal breast tissue [48]. These observations provide insight into the role of MMPs in breast cancer development and progression. Using a publicly available database of 295 consecutive women diagnosed with breast cancer, McGowan and Duffy [49] carried out a comprehensive investigation to relate the mRNA expression levels of 17 different MMPs (MMP-1, 2, 3, 7-17, 19, 20, and 24) to tumor characteristics and outcome. Using univariate analysis they found that high mRNA expression of MMP-1, 9, 12, 14, and 15 was associated with poor overall survival. However, among the studied MMPs, only MMP-14 predicted outcome independent of tumor size, tumor grade, lymph node status and ER status. Interestingly, none of the MMPs analyzed in this study were associated with favorable patient outcome. Although this finding may appear inconsistent with the protective role reported for specific MMPs in human cancer [50, 51] they do not rule

out such a role because some of the MMPs investigated in this study have been observed to be protective during the early phase of carcinogenesis. Such protective effect could be lost following formation of an invasive malignancy [49]. ADAM (a disintegrin and metalloprotease) is a related enzyme of the family of MMPs. The prognostic role of ADAM has been recently explored in breast cancer [52]. Like MMPs, they have zinc binding consensus sequence at their catalytic site and are involved in ectodomain shedding, regulation of growth factor availability and in cell-cell or cell-matrix interactions through their proteolytic activity in both normal and disease related states [53-55]. Increasing levels of ADAM12 has been reported in urine of patients with breast cancer in parallel with the progression of the disease [56]. ADAM17 is also elevated dramatically in high-grade breast tumor and predicts patients with shorter overall survival [57].

### **3. Tissue Expression of MMP-1, MMP-2 and MMP-9 and correlation with Breast Cancer.**

*3.1. Tissue Expression of MMP-1:* Matrix metalloproteinase-1 is the most ubiquitously expressed interstitial collagenase. It cleaves fibrillar collagens including collagen type I, II, and III. Its cleavage products are then rapidly denatured at body temperature to gelatin, which can be further processed by gelatinases MMP-2 and MMP-9 [58]. MMP-1 can be produced by a wide variety of normal cells (for example, stromal fibroblast, macrophages and endothelial cells) and it is involved in tissue remodeling and repair. Over expression of MMP-1 has been shown to play a role in initiating mammary tumorigenesis by degrading stroma and releasing growth factors and other mitogens for epithelial cells [59, 60]. Specifically, MMP-1 has been

observed to degrade the following growth factor binding proteins: insulin-like growth factor (IGF), fibroblast growth factor (FGF), and transforming growth factors (TGF and TGF- $\beta$ ) [61]. In breast cancer, MMP-1 expression is associated with adverse outcome [40, 62, 63] and is implicated in bone metastasis and also in the multiple steps leading to the development of lung metastasis [46, 64]. Using real time RT-PCR, Cheng et al [65] detected higher mRNA expression levels of MMP-1 in breast tumor tissue specimens compared to the adjacent normal tissue. This increased mRNA level correlated inversely with estrogen receptor levels and was significantly correlated with high frequency of recurrence and fatal outcome. The study was based on data from 85 patients with a median follow-up of 38 months [65]. By immunohistochemistry, Poola et al [40] identified MMP-1 as a predictive marker for patients with precancerous breast lesions. They observed significant association of MMP-1 expression with precancerous tissues obtained from patients with a history of cancer. However, in the group of patients having a diagnosis of ADH (atypical ductal hyperplasia), who never had cancer before and who did not develop cancer in 5-7 years after the ADH diagnosis, 26 out of 30 of these patients were negative for MMP-1 [40]. In general, there was more intense staining of MMP-1 in stroma cells compared to epithelial cells [40]. They also observed that elevated mRNA expression of MMP-1 was evident in ductal lavage samples diagnosed as atypia compared with samples diagnosed as benign by RT-PCR [40]. The successful identification of MMP-1 mRNA in ductal lavage could promote its use as a screening tool for women without visible lesions or to identify lesions that may develop into cancer. Other similar procedures such as random periareolar fine needle aspiration could become useful as additional screening tools for early detection of breast cancer onset. These data point to the potential role of MMP-1 as a clinical marker for breast cancer risk or early diagnosis.

Bohn et al [66] studied biomarker profiles by tissue microarray for bone metastasis in breast carcinomas. Although they observed high expression of MMP-1 in both primary and metastatic tumors of patients, they however could not distinguish the tumor groups (metastatic and primary) based on the expression of MMP-1 (and CXCR4, CD44, TFF-1, PTHrP, FGFR3 and IL-11) but observed that tumors associated with bone metastasis tended to be larger than 2 cm [66]. In an earlier study however, Massagué's group using athymic nude mice, observed over expression of IL-11, MMP-1 and CXCR4 genes at the mRNA levels in the metastatic cells of bone, thus supporting the idea of MMP-1 as a gene signature for bone metastasis in breast cancer[64]. later on, by transcriptomic micro-array analysis of the highly and weakly lung-metastatic cell populations, they also found that MMP1 along with other genes as signature genes for lung metastasis of breast cancer[46].

Based on gene expression profiling studies employing microarray analysis, the following distinct breast cancer subgroups have been identified- luminal A, luminal B, human epidermal growth factor receptor-2 (HER-2), normal breast like and basal like type cancers [67, 68]. The basal type is comprised almost entirely of tumors lacking the estrogen receptor (ER), progesterone receptor (PR) and HER-2 (triple negative phenotype). Hormone therapy is effective for luminal A and B subtypes, both of which express ER. Trastuzumab (Herceptin) may be administered to HER-2 positive patients but no targeted therapies exist for basal subtype except chemotherapy. However, McGowna and Duffy have shown that the expression of MMP-1 mRNA is significantly elevated in the basal type breast cancer compared with non

basal tumors [49]. This finding suggests that MMP-1 might be used as a target for the treatment of the basal subtype of breast cancer.

*3.2. Tissue Expression of MMP-2 and MMP-9:* Matrix metalloproteinase-2 and MMP-9 are gelatinases (also known as gelatinase A and B) that primarily hydrolyze gelatine and type IV collagen, the major structural component of basement membrane [69]. Both of the gelatinases are involved in tumor invasion and metastasis by their capacity for degradation of extracellular matrix and induction of angiogenesis [70, 71]. Unlike other MMPs, MMP-2 is constitutively expressed by most connective tissue cells including endothelial cells, osteoblasts, fibroblasts and myoblasts. It can also cleave insulin-like growth factor-(IGF)-binding proteins and release IGFs, thus having a strong effect on simulating cell proliferation and inhibiting apoptosis. Matrix metalloproteinase-2 may also be involved in mitogenic and angiogenic activities by targeting fibroblast growth factor receptor 1 and produce the active soluble ectodomain of the receptor [59]. All these activities of MMP-2 may contribute to both cancer development and progression.

Some studies show MMP-2 expression to be associated with poorer prognosis [72-74], while others fail to relate MMP-2 expression to a poorer prognosis [75-77]. Using *in situ* hybridization Tetu et al observed overexpression of MMP-2 in close to 600 cases of breast cancer, but they did not find significant association between increased MMP-2 expression with greater risk of recurrence [42]. Using real time RT-PCR to assess the expression of MMP-2 mRNA in breast tumor tissue, Decock et al found significant correlation of high expression of MMP-2 mRNA in luminal A invasive ductal carcinomas of premenopausal patients compared

to those with postmenopausal status, but no correlation was found between MMP-2 expression and other clinicopathological parameters, like tumor size, differentiation grade, and lymph node involvement [78]. In a tissue array study of 278 invasive breast cancers specimens, Sivula et al [73] observed coexpression of MMP-2 and cyclooxygenase-2 (COX-2) may be directly involved in mammary carcinogenesis, since in transgenic mouse, expression of COX-2 is sufficient for inducing breast cancer [79]. In this study, they also found that high expression of either COX-2 or MMP-2 was associated with decreased disease specific survival when compared with the group of low COX-2 or MMP-2 expression [73]. Positive MMP-2 has also been correlated with shortened survival in postmenopausal patients with node-positive breast carcinoma of a low tumor burden [72]. These data point to the role of MMP-2 as a prognostic marker in specific breast cancer cases.

Pellikainen et al [80] investigated the expression and the prognostic role of MMP-2 and MMP-9 in a prospective series of 421 breast cancer patients diagnosed and treated within a five year period. Immunohistochemistry was performed using a primary antibody for MMP-2 that recognizes the human proform of MMP-2 and a primary antibody for MMP-9 that recognizes both the pro- and active forms of human MMP-9. Both MMP-2 and MMP-9 were expressed in the cytoplasm of malignant and stromal cells. Positive stromal expression of the MMPs was associated with aggressive factors and high MMP expression in carcinoma cells was related to small tumors (T1, stage I). The expression of MMP-9 at high levels in carcinoma cells was found to be associated with node-negative patients and this expression favored survival, whereas positive expression of MMP-9 in stromal cells was associated with poorly differentiated tumors and HER-2 over expression predicted poor survival in ER positive breast

cancers. The study also found that positive stromal MMP-2 expression was associated with ER negativity, HER-2 over expression, poor tumor differentiation and a diagnosis of ductal carcinoma [80]. Previously, Duffy et al reported an inverse relationship of both MMP-8 and MMP-9 with ER levels in breast carcinoma based on enzyme linked immunosorbent assays (ELISA), thus suggesting that high levels of these enzymes are related to poor patient outcome since the presence of ER is generally correlated with good prognosis [81]. *In situ* hybridization studies show the expression of MMP-2 and MMP-9 mRNA mainly in the stromal cells with some in the carcinoma cells [76, 82, 83]. Matrix metalloproteinase-2 mRNA has also been detected in some benign glandular cells [84], which is consistent with studies that have detected MMP-2 and MMP-9 activity in benign breast tissues by zymography [83, 85, 86].

A recent study by Shah [87] compared both pro- and active levels of MMP-9 and MMP-2 in tissue extracts of malignant and adjacent normal breast tissue. They found that the active levels of both MMP-2 and MMP-9 were higher in malignant breast tissue and that pro-MMP-2, active-MMP2 and active-MMP9 could significantly discriminate between malignant and adjacent normal breast tissue. Further more, they found that MMP-2 activation ratio was significantly associated with increased risk of lymph node metastasis in node-negative patients and it showed discriminatory efficacy between patients with and without lymph node metastasis [87]. Metastasis is a major adverse step in the progression of breast carcinoma and auxiliary lymph node metastasis has been considered the most reliable predictor of poor prognosis, including early recurrence and short survival duration. Using cDNA and tissue microarray analyses to study the gene and protein expression profile changes associated with metastasis, Hao et al [88] studied 9 paired primary breast cancer and their lymph node

metastases. In this study, the gene expression patterns were highly heterogeneous among the set of sample pairs. But they observed decreased mRNA and increased protein expression of MMP-2 along with fibronectin, another key extracellular matrix protein, in the metastases group compared to the corresponding primary tumor. This finding suggests a negative feedback control of gene expression in the metastasis process and both cDNA microarray and tissue microarray analysis are important for obtaining a complete picture of molecular events associated with metastasis [88].

The role of metalloproteases in breast cancer prognosis continues to pose complications in interpretation of results. The complexity of reports may be due to the use of different technological methods which introduces the issue of sensitivity. For example, *in situ* hybridization can not differentiate between the active and inactive forms of MMPs, but can more accurately provide information about localization/expression patterns. On the other hand, a technique such as an immunoassay can detect enzyme activity, but not expression pattern. Contradictory results may also occur as a result of comparing studies utilizing homogenous specimens (e.g. tumor population only) with those utilizing heterogeneous specimens. Heterogeneous specimens such as one where the tumor has a mixture of benign and/or non invasive sections could provide data different from one using purely tumor (100-90% tumor or laser micro-dissected tumor) specimens. The use of different models (animal versus human) may also produce varying results even when experiments are performed using the same technological platform. These discrepancies/contradictions throw light on the need for careful assessment of each experimental design before comparisons are made between experiments.

#### **4. Circulating MMP-1, MMP-2 and MMP-9 and correlation with Breast Cancer.**

Decock et al [89] reported that median plasma levels of MMP-1 were higher in patients with no breast cancer (controls) compared to those with breast cancer. There was also a trend of lower MMP-1 levels and increasing tumor size. Using 4.24ng/ml as the cut-off value for MMP-1 level in plasma, the sensitivity of MMP-1 was 80%, and the specificity was 24%. These data indicate that circulating MMP-1 may be a good candidate biomarker for breast cancer. It has been observed that the median plasma MMP-1 levels in non-inflammatory breast cancer patients are twice as high as that of inflammatory breast cancer, a rare but very aggressive disease [89]. This is contradictory to previous findings of increased tissue expression of MMP-1 as a marker of poor diagnosis in breast cancer [40, 46]. This discrepancy may be explained by the hypothesis that lower plasma MMP-1 could reflect a higher concentration of MMP-1 in the tumor microenvironment, thus providing a reverse correlation of plasma MMP-1 expression with tumor growth [89]. We have previously reported that preoperative plasma concentration and activity of total MMP-2 are significantly higher in breast cancer patients than in control individuals [90], whereas the serum concentration of active-MMP-2 were significantly lower in breast cancer patient than in controls, which suggests that active-MMP-2 may be associated with normal tissue remodeling, and it is the balance between the active and total forms of MMP that maybe important in breast cancer classification [91]. In our study we did not observe significant difference in the concentration of active-MMP-2 in serum between breast cancer and benign patients, a possible indication that MMP-2 may not be actively involved in the complex events that initiate or promote cancer progression from the benign to the malignant state. Unlike MMP-2, both active- and total-

MMP-9 were significantly higher in breast cancer patients compared to the benign group, suggesting a more active role of MMP-9 in the events that characterize the benign and cancer conditions. On the other hand, the activity of total-MMP-2 in serum was significantly lower in the control group compared to people who were at a high risk of developing breast cancer (Gail Model 5-year risk ratio of less than 1.67), or benign and cancer patients. Such distinction would have significant impact on early detection, monitoring and management of breast cancer and could lead to the establishment of proper intervention measures [91].

Measurement of serum concentrations of MMP-2 by the quantitative sandwich enzyme immunoassay technique on 57 consecutive patients with invasive breast cancer before surgery and 12 patients with benign breast tumor showed the mean value of serum MMP-2 in patients with invasive breast cancer ( $694.3 \pm 140.5$  ng/ml) to be significantly higher ( $P = 0.026$ ) than those of the control group ( $593.3 \pm 134.0$  ng/ml) [92]. The study also demonstrated that high preoperative serum MMP-2 levels reflect more advanced Tumor-Node-Metastasis (TNM) staging and thus might reflect the severity of invasive breast cancer [92]. Leppa [93] studied 133 node-positive breast cancer patients and showed a significant inverse relationship of postoperative serum MMP-2 levels with favorable prognosis. The overall survival and disease-free survival rates were better among patients with low MMP-2 levels than in patients with high levels. The bone and visceral metastases frequency was also significantly lower in patients with low serum MMP-2 levels and the prognostic value was more prominent in the ER positive subgroup [93]. These results support MMP-2 as a strong and independent indicator of the aggressiveness of breast cancer and its potential as a convenient circulating marker for prognosis in node positive breast cancer. High level of serum MMP-9 has been associated with

increased risk of lymph node metastasis; poor relapse-free and overall survival and thus may be used as a predictive marker for breast cancer metastasis and outcome [94]. Monitoring plasma MMP-9 may also be useful to evaluate therapeutic efficacy and predict outcome, since progressive decrease in the levels of MMP-9 in plasma was observed in patients who respond well to adjuvant therapy after surgery [95]. Gradually increasing MMP-9 activity can be detected 1 to 8 months before clinical diagnosis of recurrence in relapse cases [95]. Studies by Di carlo and his collaborators also noted an enhanced serum MMP-9/MMP-2 ratio in cancer patients compared with those with benign disease and healthy individuals [96].

Studies indicate that MMP-9 can form a complex with neutrophil gelatinase-associated lipocalin (NGAL), an inducible acute phase protein that has recently been implicated in human cancer. Binding of NGAL to MMP-9 protects MMP from autodegradation [97-99]. It has been shown that circulating levels of MMP-9, NGAL and MMP-9/NGAL complex increased significantly in patients with invasive ductal carcinoma of the breast compared with healthy controls[17]. Elevation of serum MMP-9 and NGAL levels was also observed in ADH and ductal carcinoma (DCIS) patients. Significant correlation was observed only between serum levels of MMP-9 and NGAL, but not their complex in more severe cases of breast disease [17]. The MMP-9/NGAL complex can also be detected in urine samples of breast cancer patients, but not in those from healthy controls [98]. Dramatic increase of urine MMP-9 have also been observed in patients with ADH and lobular carcinoma in situ (LCIS) compared to normal individuals [100]. Together, these data suggest that serum measurement of MMP-9 and NGAL or urine detection of MMP-9 or MMP-9/NGAL complex may be useful as non-invasive methods of predicting breast disease status and monitoring breast cancer progression.

## **5. Genetic polymorphisms of MMP and its prognostic role in cancer**

A number of regulatory mechanisms can impact the role of MMP on the extracellular matrix, such as degradation, regulation of transcription, activation of latent MMPs, and inhibition of MMP activity. Transcriptional regulation, however, for most MMPs, is one of its most important role because the expression of most MMPs is a response to physiological and pathological changes [101]. An exception is MMP-2, which is ubiquitously expressed at significant levels even in apparently quiescent tissues [101]. Consensus transcriptional cis-elements like AP-1 and PEA-3 have been found in the promoter regions of MMPs. These elements and other cis elements in MMP promoters can interact with transcription factors, like Fos, Jun and members of the Ets families, thus playing an important role in regulating MMP gene expression [102]. Genetic polymorphisms, the naturally occurring sequence variation in the promoter regions of MMP genes may create changes in these transcription factor binding sites and influence the expression of MMPs, thus resulting in differential levels of expression of MMPs in individuals and providing different levels of susceptibility to a variety of diseases including coronary heart disease, abdominal aortic aneurysm, and cancers [103-105]. Single-nucleotide polymorphisms (SNP) in MMP-1 (insertion of G at position -1607), MMP-2 (substitution of c to T at 1306), MMP-3 (deletion of A at position 1171), MMP-7 (substitution of A to G at -181), MMP-9 (C to T substitution at-1562), MMP-12 (A to G substitution at -82), and MMP-13 (A to G substitution at -77) have been reported to have effects on transcription activity of these MMPs. In addition, the polymorphisms of MMP-1, -2, -3, and -9 have been

implicated in the susceptibility, invasive capacity and prognosis of cancers [105-107], Insertion of an adenosine in the MMP-3 promoter at position -1612 bp relative to the transcriptional start site creates a 5A/6A polymorphism. Carriers of the 5A allele are observed to have a relatively higher promoter activity presumably because of the higher affinity of 6A allelic promoter to a putative transcriptional repressor [108]. Matrix metalloproteinase-3 expression has been studied in atherosclerotic plaque where rupture is most commonly detected, and in increased risk of acute myocardial infarction, a clinical event commonly caused by rupture of coronary atherosclerotic plaque has been associated with individuals carrying the 5A allele [109]. A significantly higher frequency of the presence of the 5A polymorphism in MMP-3 promoter was also observed among patients with cancer (mammary, ovarian, pulmonary but not colorectal) than among individuals without cancer in an Italian population, suggesting that it is a risk factor for developing cancer [110]. The MMP-1 2G polymorphism has been linked to early-onset of lung cancer [113], increased risk of nasopharyngeal carcinoma, oral squamous cell carcinoma, colorectal cancer susceptibility [114-116], aggressiveness of head and neck cancer [117, 118] and negative prognosis in patients with ovarian cancer [119]. On the other hand, the 2G polymorphism of MMP-1 appears to play a protective role against prostate cancer [120].

Breast cancer is a very heterogeneous disease based on the various clinical outcomes observed in patients [111, 112]. This makes it very essential and necessary to predict the aggressive potential of the disease at an early stage and to tailor treatment accordingly. Given the role of MMPs in breast cancer progression and the correlation of high MMP levels with poor prognosis, increasing effort has been made to explore the relationship between polymorphism

of MMPs and breast cancer progression, based on the hypothesis that individuals carrying high expressing MMP genotypes may be more susceptible to tumor spread compared with patients who lack these alleles [106]. These experiments as expected have been performed using different ethnically and geographically distinct populations, and the scope and scale including the statistical methods applied are different. However, several of these experiments did show that the polymorphism of MMP-1, MMP-2, MMP-3, and MMP-9 are promising biomarkers for breast cancer. Although a few of the reports showed weak association of these promoter polymorphisms in MMPs with breast cancer, the definite role of MMP polymorphisms in breast cancer prognosis and their clinical significance should be considered within specific clinical situations. There is the need to further investigate the role of MMP polymorphisms as essential tools for clinical prediction especially in combination with other predictors of breast cancer.

#### *5.1. Polymorphism in MMP-1 gene promoter and its relationship with breast cancer:*

Brinckerhoff and her group first reported that a single nucleotide polymorphism in MMP-1 promoter was produced by the insertion of guanine at -1607bp [105]. This polymorphism creates a binding site for Ets transcription factor and increases the transcription activity of MMP-1 gene. They also found that the frequency of the 2G polymorphism dramatically increased in eight tumor cell lines from 30% to 62.5% [105]. Hughes[106] reported a 3.9 fold increased risk of lymph node metastasis for patients carrying the 2G/2G genotype of MMP-1 compared with subjects carrying the 1G/1G genotype in 221 mixed ethnicity patients. However, when they restricted the analysis to the Caucasian-only group, the risk decreased to 2.6 fold [106]. This could explain discrepancies reported in literature and point to the fact that

the role of MMP polymorphisms will vary based on ethnic differences and research results should be viewed based on such differences. Przybyowska et al [121] reported the association between MMP-1 2G/2G genotype and lymph node metastasis in patients with breast cancer supporting previous observation that increased expression of MMP-1 correlates with lymph node metastasis and poor prognosis. These reports all point to the potential prognostic role of MMP-1 polymorphisms in identifying patients at risk of more aggressive disease [121]. In the study by Hughes [106], patients with MMP-1 2G/2G genotype also had a significant shorter overall survival ratio than the 1G/1G carrying patients, however, after adjusting for node status, this association was lost. This observation Hughes says, supports their hypothesis that elevated levels of MMPs promote spread of breast cancer and high expressing MMP-1 2G/2G genotypes promote lymph node metastasis. Thus, MMP-1 may play a role in distant tumor dissemination [106]. A study on an ethnically homogenous north Sweden population however failed to find significant link of the SNP of MMP-1 and other MMPs (MMP-2, -3, -9, -13) with increased breast cancer risk and decreased survival [122]. Also, studies reported by McColgan and Sharman did not identify polymorphism of MMPs-1, -2, -3, -9 as prognostic markers for breast cancer, although they did observe significant association of MMP-1 nt-1607 polymorphism with colorectal cancer [113]. These findings once again indicate that the role of MMP-1 or other MMP polymorphisms in disease will vary within different ethnic populations. Therefore the usefulness of MMP polymorphisms as a clinical tool will be population specific.

Hughes et al [109] also studied the haplotype (a set of SNPs on a single\_chromatid that are statistically associated) effect of MMP-1, MMP-3, MMP-7 MMP-12 and MMP1-3 genes on breast cancer progression since they are clustered on the same chromosome (11q22.3). A

positive link was identified between haplotype 2G-5A-AAA and 2G-5A-G-A-A and the incidence of lymph node metastasis. The latter haplotype was further linked to poor overall survival even after adjustment for lymph node status [106]. The results indicate that the combination effect of multiple high-expressing MMP-alleles may provide a stronger predictor of prognosis than individual SNPs [106].

*5.2. Polymorphism in MMP-2 gene promoter and its relationship with breast cancer:* Several SNPs have been identified in the MMP-2 promoter region and the C to T transition located at nucleotide -1306 have been intensively studied. This SNP abolishes the Sp-1-binding site and diminishes its promoter activity, resulting in a significantly greater gene expression which is driven by the C allelic MMP-2 promoter rather than by the T allelic counterpart [107]. The MMP-2 - 1306CC genotype has been associated with increased risk of cancer development including lung cancer, prostate cancer, and esophageal squamous cell carcinoma [123-126]. A study based on the analysis of 462 breast cancer patients in China demonstrated a significantly reduced risk of breast cancer in patients carrying the T allele genotype, especially in subjects who were younger at diagnosis [32]. Similarly, increased risk of breast cancer for MMP-2 - 1306CC allele genotype carriers was observed in a Mexican population [127]. The increased risk is also more pronounced among women 50 years old or younger [127]. This is consistent with the idea that genetic susceptibility correlates with early age of disease onset [127]. Lei et al only observed a marginal correlation of MMP-2 1306 polymorphism with prognostic indicators of breast cancer in a North Sweden population [122]. An investigation into the role of MMP-2 in breast cancer progression showed that TT carriers have significantly smaller breast tumors compared with patients carrying the CT or CC genotype [128]. When these

patients were subgrouped according to their ER status, favorable survival for TT patients was observed in the ER-positive group, however, this trend was not statistically significant. But in the ER-negative group, TT patients had poorer survival compared to CT or CC patients ( $P= 0.002$ ) [128].

*5.3. Polymorphism of MMP-9 gene promoter and its relationship with breast cancer:* A single nucleotide polymorphism (C→T) at position -1562 in the MMP-9 promoter is reported to affect expression of this gene. The C→T substitution at the polymorphic site increases transcription activity of MMP-9 promoter as a result of loss of binding site for a repressor protein [129]. The T allele has been reported to be significantly associated with 2.6 fold increased risk of breast cancer and the TT genotype with a non-significant 1.9 fold risk in a larger study [130]. Increased T allele frequency is not only associated with the occurrence of breast cancer but is also related with lymphatic invasion in breast cancer [106]. However, Grieu et al [128] report that the MMP-9 -1562 T allele was associated with features of good prognosis including non-ductal type histology, positive ER status and the absence of TP53 mutation. In their study, patients with MMP-9 -1562 CT or TT genotypes showed marginally better prognosis compared to CC homozygotes ( $p=0.06$ ) [128].

In general, these reports indicate the potential usefulness of MMP-1, MMP-2 and MMP-9 polymorphisms in clinical prognosis. The significance of MMP polymorphisms in breast cancer prediction and prognosis is expected to be different in different ethnic groups. Studies that focus on different geographical regions and ethnic populations may provide information

on their precise role as biomarkers for breast cancer and the foundation for their use in clinical practice in those geographic regions.

## **6. The potential role of MMPs as therapeutic targets for cancer treatment**

With the increasing acknowledgment of the role of MMPs in tumor growth and metastasis, and the dysregulated angiogenesis associated with these events, significant efforts have been made to design and develop diagnostic and therapeutic products using MMPs. Some of these include the development of MMP inhibitors (MMPIs) such as Marinastat. Marinastat is a second generation MMP inhibitor that can mimic the structure of collagen at the MMP cleavage site and block the activity of different MMPs by occupying the substrate-binding site and chelating the zinc [131, 132]. Although MMP inhibitors have produced some promising results, for example, in a phase II clinical trial, Neovastat, a compound that selectively inhibits MMP-2, -9 and -12 and the vascular endothelial growth factor (VEGF) was administered to patients with refractory renal cell and this increased patient's median survival time [133]. In many other cases, positive clinical effects failed to be obtained. Factors contributing to the unsuccessful clinical results include for example, the lack of specificity of MMP administered thus resulting in off-target effects, such as musculoskeletal adverse effects which limits the maximum-tolerated dose as observed with Marinastat [132]. Like many other anticancer drugs, the effective dose of MMP inhibitors required during treatment at the early stage of cancer may be lower than that required at the later or end stage of the disease, thus limiting the toxic effects

that hamper the utilization of these drugs. It therefore would be more effective to treat patients at the early stages of cancer progression rather than at the end stage of the disease. Interestingly, the drug efficacy of an early MMPI drug, INCB7839, a small molecule that specifically inhibits ADAM, a family of enzyme related with MMPs, did not induce musculoskeletal adverse effects in initial animal studies and is currently being used in phase II clinical trials against breast carcinoma and several other solid tumors [134]. The multifaceted role of MMPs has been a major hindrance in the therapeutic use of MMPs such that nonspecific blocking of different MMP activities of the same MMP at different disease stages may also block the protective effects of some MMPs against cancer. For example, both animal and human studies support the protective role of MMP-12 in tumor progression [135, 136]; in a transgenic mouse model of invasive squamous cell carcinoma, MMP9 stimulates tumor proliferation at early stages but restricts further malignant progression at later stages [137]. To fully understand the biological mechanism underlying MMPs' positive and negative role in tumor progression, studies that provide more accurate evaluation of the role of these MMPs in vivo will need to be designed. This may be possible through noninvasive imaging techniques that are capable of monitoring MMP activity at different stages of cancer progression [138]. Also resolving the 3D structures of more MMPs will enable the design of effective specific MMPIs that will provide more successful clinical application of MMPIs as anticancer therapeutic tools.

## **7. Summary**

We have reviewed the expression profile of MMPs and specifically MMP-1, MMP-2 and MMP-9 in both solid tissue and blood of breast cancer patients. We also discussed MMP polymorphisms and their clinical significance in the early onset of breast cancer and its progression. Although some discrepancies exist between studies from different groups, which may partially be due to differences in the technological platforms employed, experimental models utilized and the specific populations (of humans) studied, still a lot of these results point to the potential usefulness of MMPs as biomarkers and the fact that they hold some promise in the clinic for early detection, diagnosis, prognosis and prediction. Below, we have summarized the results from some of these studies (Table 1) to provide a snapshot for reference. The multifaceted role of the MMPs in cancer onset and progression continues to plague our progress in MMP research. However, we believe that in addition to the development of sophisticated imaging systems for accurate evaluation of MMP efficacy *in vivo*, the use of a panel of biomarkers which include MMPs and other molecules that work in sequence with the MMPs in the process of cancer development (such as angiogenesis) could also speed up the identification of more clinically useful biomarkers for breast cancer diagnosis and prognosis.

**Table 1. Summary of selected MMPs and their potential role as clinical indicators for breast cancer**

	solid tissue biomarkers	Ref.	circulating biomarker	Ref.	polymorphism	Ref.	
Early detection	↑MMP-1:stroma	40	↑total MMP-2 & activity: plasma	90	MMP-2: 1306CC in Mexican, Chinese population	32,127	
	↑MMP-1 mRNA: ductal lavage	40	↑total MMP-2 & ↓active MMP-2:serum	91	MMP-9: 1562T allele	130	
			↑MMP-9:serum	91			
			↑MMP-9/MMP2 ratio: serum	96			
			↑MMP-9:urine	98,100			
LN metastasis	↓ MMP-2 mRNA & ↓ protein	88	↑serum MMP-9	94	MMP-1: 1607 2G/2G mixed ethnicity	106,121	
	Active MMP-2	87			MMP-1,3, 7, 12,13: 2G-5A-A-A-A haplotype	106	
					MMP-9: 1562 T allele in mix ethnicity	106	
Severity			↓↓plasma MMP-1	89			
			↑ preoperative MMP-2: IBC serum	92			
			↑MMP-9:IBC &ADH & DCIS	17			
Prognosis							
	poor	↑ MMP-1 mRNA: tumor	65	↑postoperative MMP-2:serum/ LN+ patient	93	MMP-1: 1607 2G/2G mixed ethnicity	121
		+MMP-2 : stroma	80	↑MMP-9:serum	94	MMP-1,3, 7, 12,13: 2G-5A-G-A-A haplotype	106
		+ MMP-9: stroma/ER+ group	80			MMP-2: 1306TT/ ER- group	128
	good	+ MMP-2: carcinoma	80			MMP-2: 1306TT/ ER+ group	128
		↑ MMP-9: carcinoma	80			MMP-9: 1562 T allele in Australian population	128

ER: estrogen receptor; LN: lymph node; Negative (-), Positive (+).

IBC: invasive breast cancer; ADH: atypical ductal hyperplasia; DCIS: ductal carcinoma in situ.

↑: increased tissue expression/circulating level; ↓: decreased tissue expression/circulating level; ↓↓: greatly decreased level.

## References

1. Fingleton, B., *Matrix metalloproteinases: roles in cancer and metastasis*. Front Biosci, 2006. 11: p. 479-91.
2. Menon, B., M. Singh, and K. Singh, *Matrix metalloproteinases mediate beta-adrenergic receptor-stimulated apoptosis in adult rat ventricular myocytes*. Am J Physiol Cell Physiol, 2005. 289(1): p. C168-76.
3. McCawley, L.J., et al., *Keratinocyte expression of MMP3 enhances differentiation and prevents tumor establishment*. Am J Pathol, 2008. 173(5): p. 1528-39.
4. Wang, W., et al., *Matrix metalloproteinase-1 promotes muscle cell migration and differentiation*. Am J Pathol, 2009. 174(2): p. 541-9.
5. Zeng, Z.Z., et al., *alpha(5)beta(1) Integrin Ligand PHSRN Induces Invasion and alpha(5) mRNA in Endothelial Cells to Stimulate Angiogenesis*. Transl Oncol, 2009. 2(1): p. 8-20.
6. Shan, B., et al., *Activation of proMMP-2 and Src by HHV8 vGPCR in human pulmonary arterial endothelial cells*. J Mol Cell Cardiol, 2007. 42(3): p. 517-25.
7. Visse, R. and H. Nagase, *Matrix metalloproteinases and tissue inhibitors of metalloproteinases: structure, function, and biochemistry*. Circ Res, 2003. 92(8): p. 827-39.
8. Verma, R.P. and C. Hansch, *Matrix metalloproteinases (MMPs): chemical-biological functions and (Q)SARs*. Bioorg Med Chem, 2007. 15(6): p. 2223-68.
9. Sounni, N.E., et al., *Membrane type-1 matrix metalloproteinase and TIMP-2 in tumor angiogenesis*. Matrix Biol, 2003. 22(1): p. 55-61.

10. Song, B., et al., *MicroRNA-21 regulates breast cancer invasion partly by targeting tissue inhibitor of metalloproteinase 3 expression*. J Exp Clin Cancer Res, 2010. 29: p. 29.
11. Zheng, S.Q., R.Q. Huang, and Y.J. Zhang, *[Role of matrix metalloproteinase (MMP)-2 and -9 and vascular endothelial growth factor C in lymph node metastasis of breast cancer]*. Zhonghua Bing Li Xue Za Zhi, 2010. 39(4): p. 240-4.
12. Lee, H.S., M.H. Na, and W.K. Kim, *alpha-Lipoic acid reduces matrix metalloproteinase activity in MDA-MB-231 human breast cancer cells*. Nutr Res, 2010. 30(6): p. 403-9.
13. Lee, C.C., et al., *Sesamin Inhibits Macrophage-Induced Vascular Endothelial Growth Factor and Matrix Metalloproteinase-9 Expression and Proangiogenic Activity in Breast Cancer Cells*. Inflammation, 2010.
14. Radisky, E.S. and D.C. Radisky, *Matrix metalloproteinase-induced epithelial-mesenchymal transition in breast cancer*. J Mammary Gland Biol Neoplasia, 2010. 15(2): p. 201-12.
15. Sun, Y., et al., *Oroxylin A suppresses invasion through down-regulating the expression of matrix metalloproteinase-2/9 in MDA-MB-435 human breast cancer cells*. Eur J Pharmacol, 2009. 603(1-3): p. 22-8.
16. Laudanski, P., et al., *Increased serum level of membrane type 1-matrix metalloproteinase (MT1-MMP/MMP-14) in patients with breast cancer*. Folia Histochem Cytobiol, 2010. 48(1): p. 101-3.
17. Provatopoulou, X., et al., *Circulating levels of matrix metalloproteinase-9 (MMP-9), neutrophil gelatinase-associated lipocalin (NGAL) and their complex MMP-9/NGAL in breast cancer disease*. BMC Cancer, 2009. 9: p. 390.

18. Farias, E., et al., *Plasma metalloproteinase activity is enhanced in the euglobulin fraction of breast and lung cancer patients*. Int J Cancer, 2000. 89(4): p. 389-94.
19. Gonzalez Rodriguez, I., et al., *Expression of matrix metalloproteinase-9 in prostate cancer. Preliminary experience*. Arch Esp Urol, 2010. 63(2): p. 119-24.
20. Szarvas, T., et al., *Matrix metalloproteinase-7 as a marker of metastasis and predictor of poor survival in bladder cancer*. Cancer Sci, 2010. 101(5): p. 1300-8.
21. Yi, Y.C., et al., *Matrix metalloproteinase-7 (MMP-7) polymorphism is a risk factor for endometrial cancer susceptibility*. Clin Chem Lab Med, 2010. 48(3): p. 337-44.
22. Kim, Y.W., et al., *A comparative study of protein expression in primary colorectal cancer and synchronous hepatic metastases: the significance of matrix metalloproteinase-1 expression as a predictor of liver metastasis*. Scand J Gastroenterol, 2010. 45(2): p. 217-25.
23. Garg, P., et al., *Matrix metalloproteinase-9 functions as a tumor suppressor in colitis-associated cancer*. Cancer Res, 2010. 70(2): p. 792-801.
24. Li, M., et al., *Prognostic significance of matrix metalloproteinase-1 levels in peripheral plasma and tumour tissues of lung cancer patients*. Lung Cancer, 2010. 69(3): p. 341-7.
25. Luukkaa, H., et al., *Matrix metalloproteinase (MMP)-7 in salivary gland cancer*. Acta Oncol, 2010. 49(1): p. 85-90.
26. Luukkaa, H., et al., *Expression of matrix metalloproteinase-1, -7, -9, -13, Ki-67, and HER-2 in epithelial-myoepithelial salivary gland cancer*. Head Neck, 2010. 32(8): p. 1019-27.

27. Yeh, Y.C., et al., *Elevated serum matrix metalloproteinase-3 and -7 in H. pylori-related gastric cancer can be biomarkers correlating with a poor survival*. Dig Dis Sci, 2010. 55(6): p. 1649-57.
28. Scherf, D.B., et al., *Single nucleotide polymorphisms in matrix metalloproteinase genes and lung cancer chemotherapy response and prognosis*. Eur Respir J, 2010. 35(2): p. 381-90.
29. Zohny, S.F. and S.T. Fayed, *Clinical utility of circulating matrix metalloproteinase-7 (MMP-7), CC chemokine ligand 18 (CCL18) and CC chemokine ligand 11 (CCL11) as markers for diagnosis of epithelial ovarian cancer*. Med Oncol, 2009.
30. Shen, Q., et al., *Tissue inhibitor of metalloproteinase-2 regulates matrix metalloproteinase-2-mediated endothelial barrier dysfunction and breast cancer cell transmigration through lung microvascular endothelial cells*. Mol Cancer Res, 2010. 8(7): p. 939-51.
31. Almholt, K., et al., *Metastasis is strongly reduced by the matrix metalloproteinase inhibitor Galardin in the MMTV-PyMT transgenic breast cancer model*. Mol Cancer Ther, 2008. 7(9): p. 2758-67.
32. Zhou, Y., et al., *Substantial reduction in risk of breast cancer associated with genetic polymorphisms in the promoters of the matrix metalloproteinase-2 and tissue inhibitor of metalloproteinase-2 genes*. Carcinogenesis, 2004. 25(3): p. 399-404.
33. Susskind, H., et al., *Increased plasma levels of matrix metalloproteinase-9 and tissue inhibitor of metalloproteinase-1 in lung and breast cancer are altered during chest radiotherapy*. Int J Radiat Oncol Biol Phys, 2003. 56(4): p. 1161-9.

34. Rucci, N., et al., *Receptor activator of NF-kappaB ligand enhances breast cancer-induced osteolytic lesions through upregulation of extracellular matrix metalloproteinase inducer/CD147*. *Cancer Res*, 2010. 70(15): p. 6150-60.
35. Gach, K., et al., *The influence of opioids on matrix metalloproteinase-2 and -9 secretion and mRNA levels in MCF-7 breast cancer cell line*. *Mol Biol Rep*, 2010.
36. Karroum, A., et al., *Matrix metalloproteinase-9 is required for tubular network formation and migration of resistant breast cancer cells MCF-7 through PKC and ERK1/2 signalling pathways*. *Cancer Lett*, 2010. 295(2): p. 242-51.
37. Przybyłowska, K., et al., *Polymorphisms of the promoter regions of matrix metalloproteinases genes MMP-1 and MMP-9 in breast cancer*. *Breast Cancer Res Treat*, 2006. 95(1): p. 65-72.
38. Sieuwerts, A.M., et al., *Concentrations of TIMP1 mRNA splice variants and TIMP-1 protein are differentially associated with prognosis in primary breast cancer*. *Clin Chem*, 2007. 53(7): p. 1280-8.
39. Mendes, O., et al., *MMP2 role in breast cancer brain metastasis development and its regulation by TIMP2 and ERK1/2*. *Clin Exp Metastasis*, 2007. 24(5): p. 341-51.
40. Poola, I., et al., *Identification of MMP-1 as a putative breast cancer predictive marker by global gene expression analysis*. *Nat Med*, 2005. 11(5): p. 481-3.
41. Jinga, D.C., et al., *MMP-9 and MMP-2 gelatinases and TIMP-1 and TIMP-2 inhibitors in breast cancer: correlations with prognostic factors*. *J Cell Mol Med*, 2006. 10(2): p. 499-510.

42. Tetu, B., et al., *The influence of MMP-14, TIMP-2 and MMP-2 expression on breast cancer prognosis*. Breast Cancer Res, 2006. 8(3): p. R28.
43. Leppa, S., et al., *Clodronate treatment influences MMP-2 associated outcome in node positive breast cancer*. Breast Cancer Res Treat, 2005. 90(2): p. 117-25.
44. Chenard, M.P., et al., *High levels of stromelysin-3 correlate with poor prognosis in patients with breast carcinoma*. Int J Cancer, 1996. 69(6): p. 448-51.
45. Gupta, G.P., et al., *Mediators of vascular remodelling co-opted for sequential steps in lung metastasis*. Nature, 2007. 446(7137): p. 765-70.
46. Minn, A.J., et al., *Genes that mediate breast cancer metastasis to lung*. Nature, 2005. 436(7050): p. 518-24.
47. Zhang, B., et al., *Tumor-derived matrix metalloproteinase-13 (MMP-13) correlates with poor prognoses of invasive breast cancer*. BMC Cancer, 2008. 8: p. 83.
48. Kohrmann, A., et al., *Expression of matrix metalloproteinases (MMPs) in primary human breast cancer and breast cancer cell lines: New findings and review of the literature*. BMC Cancer, 2009. 9: p. 188.
49. McGowan, P.M. and M.J. Duffy, *Matrix metalloproteinase expression and outcome in patients with breast cancer: analysis of a published database*. Ann Oncol, 2008. 19(9): p. 1566-72.
50. Martin, M.D. and L.M. Matrisian, *The other side of MMPs: protective roles in tumor progression*. Cancer Metastasis Rev, 2007. 26(3-4): p. 717-24.
51. Lopez-Otin, C. and L.M. Matrisian, *Emerging roles of proteases in tumour suppression*. Nat Rev Cancer, 2007. 7(10): p. 800-8.

52. Roy, R., J. Yang, and M.A. Moses, *Matrix metalloproteinases as novel biomarkers and potential therapeutic targets in human cancer*. J Clin Oncol, 2009. 27(31): p. 5287-97.
53. Stone, A.L., M. Kroeger, and Q.X. Sang, *Structure-function analysis of the ADAM family of disintegrin-like and metalloproteinase-containing proteins (review)*. J Protein Chem, 1999. 18(4): p. 447-65.
54. Edwards, D.R., M.M. Handsley, and C.J. Pennington, *The ADAM metalloproteinases*. Mol Aspects Med, 2008. 29(5): p. 258-89.
55. Murphy, G., *The ADAMs: signalling scissors in the tumour microenvironment*. Nat Rev Cancer, 2008. 8(12): p. 929-41.
56. Roy, R., et al., *ADAM 12 cleaves extracellular matrix proteins and correlates with cancer status and stage*. J Biol Chem, 2004. 279(49): p. 51323-30.
57. McGowan, P.M., et al., *ADAM-17 predicts adverse outcome in patients with breast cancer*. Ann Oncol, 2008. 19(6): p. 1075-81.
58. Stamenkovic, I., *Matrix metalloproteinases in tumor invasion and metastasis*. Semin Cancer Biol, 2000. 10(6): p. 415-33.
59. Duffy, M.J., et al., *Metalloproteinases: role in breast carcinogenesis, invasion and metastasis*. Breast Cancer Res, 2000. 2(4): p. 252-7.
60. Hojilla, C.V., F.F. Mohammed, and R. Khokha, *Matrix metalloproteinases and their tissue inhibitors direct cell fate during cancer development*. Br J Cancer, 2003. 89(10): p. 1817-21.
61. Vu, T.H. and Z. Werb, *Matrix metalloproteinases: effectors of development and normal physiology*. Genes Dev, 2000. 14(17): p. 2123-33.

62. Behrens, P., et al., *The Ets-1 transcription factor is up-regulated together with MMP 1 and MMP 9 in the stroma of pre-invasive breast cancer*. J Pathol, 2001. 194(1): p. 43-50.
63. Park, Y.H., et al., *Ets-1 upregulates HER2-induced MMP-1 expression in breast cancer cells*. Biochem Biophys Res Commun, 2008. 377(2): p. 389-94.
64. Kang, Y., et al., *A multigenic program mediating breast cancer metastasis to bone*. Cancer Cell, 2003. 3(6): p. 537-49.
65. Cheng, S., et al., *High MMP-1 mRNA expression is a risk factor for disease-free and overall survivals in patients with invasive breast carcinoma*. J Surg Res, 2008. 146(1): p. 104-9.
66. Bohn, O.L., et al., *Biomarker profile in breast carcinomas presenting with bone metastasis*. Int J Clin Exp Pathol, 2009. 3(2): p. 139-46.
67. Sorlie, T., et al., *Repeated observation of breast tumor subtypes in independent gene expression data sets*. Proc Natl Acad Sci U S A, 2003. 100(14): p. 8418-23.
68. Sorlie, T., *Molecular portraits of breast cancer: tumour subtypes as distinct disease entities*. Eur J Cancer, 2004. 40(18): p. 2667-75.
69. Liotta, L.A., W.G. Stetler-Stevenson, and P.S. Steeg, *Cancer invasion and metastasis: positive and negative regulatory elements*. Cancer Invest, 1991. 9(5): p. 543-51.
70. Egeblad, M. and Z. Werb, *New functions for the matrix metalloproteinases in cancer progression*. Nat Rev Cancer, 2002. 2(3): p. 161-74.
71. Sternlicht, M.D. and Z. Werb, *How matrix metalloproteinases regulate cell behavior*. Annu Rev Cell Dev Biol, 2001. 17: p. 463-516.
72. Talvensaaari-Mattila, A., et al., *Matrix metalloproteinase-2 (MMP-2) is associated with the risk for a relapse in postmenopausal patients with node-positive breast carcinoma*

- treated with antiestrogen adjuvant therapy.* Breast Cancer Res Treat, 2001. 65(1): p. 55-61.
73. Sivula, A., et al., *Association of cyclooxygenase-2 and matrix metalloproteinase-2 expression in human breast cancer.* Breast Cancer Res Treat, 2005. 89(3): p. 215-20.
74. Pacheco, M.M., et al., *Expression of gelatinases A and B, stromelysin-3 and matrilysin genes in breast carcinomas: clinico-pathological correlations.* Clin Exp Metastasis, 1998. 16(7): p. 577-85.
75. Brown, P.D. and R. Giavazzi, *Matrix metalloproteinase inhibition: a review of anti-tumour activity.* Ann Oncol, 1995. 6(10): p. 967-74.
76. Tetu, B., et al., *Prognostic significance of stromelysin 3, gelatinase A, and urokinase expression in breast cancer.* Hum Pathol, 1998. 29(9): p. 979-85.
77. Nakopoulou, L., et al., *MMP-2 protein in invasive breast cancer and the impact of MMP-2/TIMP-2 phenotype on overall survival.* Breast Cancer Res Treat, 2003. 77(2): p. 145-55.
78. Decock, J., et al., *Matrix metalloproteinase expression patterns in luminal A type breast carcinomas.* Dis Markers, 2007. 23(3): p. 189-96.
79. Ghosh, M., et al., *PPARdelta is pro-tumorigenic in a mouse model of COX-2-induced mammary cancer.* Prostaglandins Other Lipid Mediat, 2009. 88(3-4): p. 97-100.
80. Pellikainen, J.M., et al., *Expression of matrix metalloproteinase (MMP)-2 and MMP-9 in breast cancer with a special reference to activator protein-2, HER2, and prognosis.* Clin Cancer Res, 2004. 10(22): p. 7621-8.
81. Duffy, M.J., et al., *Assay of matrix metalloproteases types 8 and 9 by ELISA in human breast cancer.* Br J Cancer, 1995. 71(5): p. 1025-8.

82. Soini, Y., et al., *72 KD and 92 KD type IV collagenase, type IV collagen, and laminin mRNAs in breast cancer: a study by in situ hybridization*. J Histochem Cytochem, 1994. 42(7): p. 945-51.
83. Davies, B., et al., *Activity of type IV collagenases in benign and malignant breast disease*. Br J Cancer, 1993. 67(5): p. 1126-31.
84. Dalberg, K., et al., *Gelatinase A, membrane type 1 matrix metalloproteinase, and extracellular matrix metalloproteinase inducer mRNA expression: correlation with invasive growth of breast cancer*. World J Surg, 2000. 24(3): p. 334-40.
85. Garbett, E.A., M.W. Reed, and N.J. Brown, *Proteolysis in human breast and colorectal cancer*. Br J Cancer, 1999. 81(2): p. 287-93.
86. Remacle, A.G., et al., *Assay of matrix metalloproteinases types 1, 2, 3 and 9 in breast cancer*. Br J Cancer, 1998. 77(6): p. 926-31.
87. Shah, F.D., et al., *Clinical significance of matrix metalloproteinase 2 and 9 in breast cancer*. Indian J Cancer, 2009. 46(3): p. 194-202.
88. Hao, X., et al., *Differential gene and protein expression in primary breast malignancies and their lymph node metastases as revealed by combined cDNA microarray and tissue microarray analysis*. Cancer, 2004. 100(6): p. 1110-22.
89. Decock, J., et al., *Plasma MMP1 and MMP8 expression in breast cancer: protective role of MMP8 against lymph node metastasis*. BMC Cancer, 2008. 8: p. 77.
90. Somiari, S.B., et al., *Plasma concentration and activity of matrix metalloproteinase 2 and 9 in patients with breast disease, breast cancer and at risk of developing breast cancer*. Cancer Lett, 2006. 233(1): p. 98-107.

91. Somiari, S.B., et al., *Circulating MMP2 and MMP9 in breast cancer -- potential role in classification of patients into low risk, high risk, benign disease and breast cancer categories*. Int J Cancer, 2006. 119(6): p. 1403-11.
92. Sheen-Chen, S.M., et al., *Serum levels of matrix metalloproteinase 2 in patients with breast cancer*. Cancer Lett, 2001. 173(1): p. 79-82.
93. Leppa, S., et al., *A high serum matrix metalloproteinase-2 level is associated with an adverse prognosis in node-positive breast carcinoma*. Clin Cancer Res, 2004. 10(3): p. 1057-63.
94. Wu, Z.S., et al., *Prognostic significance of MMP-9 and TIMP-1 serum and tissue expression in breast cancer*. Int J Cancer, 2008. 122(9): p. 2050-6.
95. Ranuncolo, S.M., et al., *Plasma MMP-9 (92 kDa-MMP) activity is useful in the follow-up and in the assessment of prognosis in breast cancer patients*. Int J Cancer, 2003. 106(5): p. 745-51.
96. Di Carlo, A., et al., *Matrix metalloproteinase-2 and matrix metalloproteinase-9 type IV collagenases in serum of patients with pleural effusions*. Int J Oncol, 2005. 26(5): p. 1363-8.
97. Yan, L., et al., *The high molecular weight urinary matrix metalloproteinase (MMP) activity is a complex of gelatinase B/MMP-9 and neutrophil gelatinase-associated lipocalin (NGAL). Modulation of MMP-9 activity by NGAL*. J Biol Chem, 2001. 276(40): p. 37258-65.
98. Fernandez, C.A., et al., *The matrix metalloproteinase-9/neutrophil gelatinase-associated lipocalin complex plays a role in breast tumor growth and is present in the urine of breast cancer patients*. Clin Cancer Res, 2005. 11(15): p. 5390-5.

99. Kubben, F.J., et al., *Clinical evidence for a protective role of lipocalin-2 against MMP-9 autodegradation and the impact for gastric cancer*. Eur J Cancer, 2007. 43(12): p. 1869-76.
100. Pories, S.E., et al., *Urinary metalloproteinases: noninvasive biomarkers for breast cancer risk assessment*. Cancer Epidemiol Biomarkers Prev, 2008. 17(5): p. 1034-42.
101. Matrisian, L.M., *Metalloproteinases and their inhibitors in matrix remodeling*. Trends Genet, 1990. 6(4): p. 121-5.
102. Ye, S., *Polymorphism in matrix metalloproteinase gene promoters: implication in regulation of gene expression and susceptibility of various diseases*. Matrix Biol, 2000. 19(7): p. 623-9.
103. Ye, S., et al., *Preliminary report: genetic variation in the human stromelysin promoter is associated with progression of coronary atherosclerosis*. Br Heart J, 1995. 73(3): p. 209-15.
104. Yoon, S., et al., *Genetic analysis of MMP3, MMP9, and PAI-1 in Finnish patients with abdominal aortic or intracranial aneurysms*. Biochem Biophys Res Commun, 1999. 265(2): p. 563-8.
105. Rutter, J.L., et al., *A single nucleotide polymorphism in the matrix metalloproteinase-1 promoter creates an Ets binding site and augments transcription*. Cancer Res, 1998. 58(23): p. 5321-5.
106. Hughes, S., et al., *Matrix metalloproteinase single-nucleotide polymorphisms and haplotypes predict breast cancer progression*. Clin Cancer Res, 2007. 13(22 Pt 1): p. 6673-80.

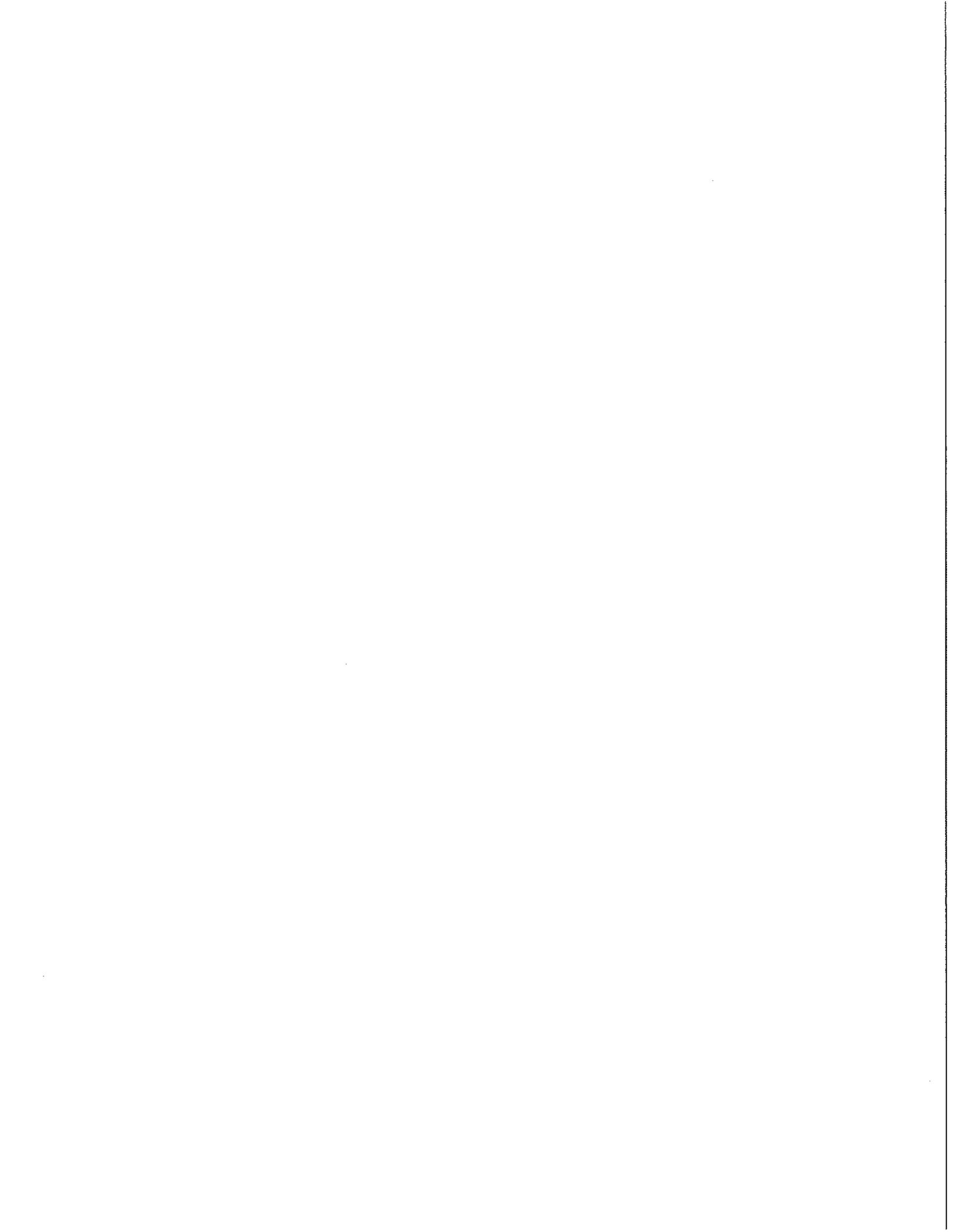
107. Price, S.J., D.R. Greaves, and H. Watkins, *Identification of novel, functional genetic variants in the human matrix metalloproteinase-2 gene: role of Sp1 in allele-specific transcriptional regulation*. J Biol Chem, 2001. 276(10): p. 7549-58.
108. Ye, S., et al., *Progression of coronary atherosclerosis is associated with a common genetic variant of the human stromelysin-1 promoter which results in reduced gene expression*. J Biol Chem, 1996. 271(22): p. 13055-60.
109. Davies, M.J., *Acute coronary thrombosis--the role of plaque disruption and its initiation and prevention*. Eur Heart J, 1995. 16 Suppl L: p. 3-7.
110. Biondi, M.L., et al., *MMP1 and MMP3 polymorphisms in promoter regions and cancer*. Clin Chem, 2000. 46(12): p. 2023-4.
111. Andre, F. and L. Pusztai, *Molecular classification of breast cancer: implications for selection of adjuvant chemotherapy*. Nat Clin Pract Oncol, 2006. 3(11): p. 621-32.
112. Andre, F. and L. Pusztai, *Heterogeneity of breast cancer among patients and implications for patient selection for adjuvant chemotherapy*. Pharm Res, 2006. 23(9): p. 1951-8.
113. McColgan, P. and P. Sharma, *Polymorphisms of matrix metalloproteinases 1, 2, 3 and 9 and susceptibility to lung, breast and colorectal cancer in over 30,000 subjects*. Int J Cancer, 2009. 125(6): p. 1473-8.
114. Kondo, S., et al., *Epstein-Barr virus latent membrane protein 1 induces the matrix metalloproteinase-1 promoter via an Ets binding site formed by a single nucleotide polymorphism: enhanced susceptibility to nasopharyngeal carcinoma*. Int J Cancer, 2005. 115(3): p. 368-76.
115. Ghilardi, G., et al., *Matrix metalloproteinase-1 promoter polymorphism 1G/2G is correlated with colorectal cancer invasiveness*. Clin Cancer Res, 2001. 7(8): p. 2344-6.

116. Vairaktaris, E., et al., *Association of matrix metalloproteinase-1 (-1607 1G/2G) polymorphism with increased risk for oral squamous cell carcinoma*. *Anticancer Res*, 2007. 27(1A): p. 459-64.
117. P, O.c., P. Leksriskul, and S. Sangruchi, *A functional polymorphism in the matrix metalloproteinase-1 gene promoter is associated with susceptibility and aggressiveness of head and neck cancer*. *Int J Cancer*, 2006. 118(10): p. 2548-53.
118. Hashimoto, T., et al., *Association of matrix metalloproteinase (MMP)-1 promoter polymorphism with head and neck squamous cell carcinoma*. *Cancer Lett*, 2004. 211(1): p. 19-24.
119. Six, L., et al., *A polymorphism in the matrix metalloproteinase-1 gene promoter is associated with the prognosis of patients with ovarian cancer*. *Gynecol Oncol*, 2006. 100(3): p. 506-10.
120. Tsuchiya, N., et al., *Clinical significance of a single nucleotide polymorphism and allelic imbalance of matrix metalloproteinase-1 promoter region in prostate cancer*. *Oncol Rep*, 2009. 22(3): p. 493-9.
121. Przybylowska, K., et al., *An association between the matrix metalloproteinase 1 promoter gene polymorphism and lymphnode metastasis in breast cancer*. *J Exp Clin Cancer Res*, 2004. 23(1): p. 121-5.
122. Lei, H., et al., *Promoter polymorphisms in matrix metalloproteinases and their inhibitors: few associations with breast cancer susceptibility and progression*. *Breast Cancer Res Treat*, 2007. 103(1): p. 61-9.

123. Yu, C., et al., *Correlation between a single nucleotide polymorphism in the matrix metalloproteinase-2 promoter and risk of lung cancer*. *Cancer Res*, 2002. 62(22): p. 6430-3.
124. Dos Reis, S.T., et al., *Matrix metalloproteinase-2 polymorphism is associated with prognosis in prostate cancer*. *Urol Oncol*, 2008.
125. Chen, X.B., et al., *[Genetic polymorphisms in STK15 and MMP-2 associated susceptibility to esophageal cancer in Mongolian population]*. *Zhonghua Yu Fang Yi Xue Za Zhi*, 2009. 43(7): p. 559-64.
126. Li, Y., et al., *Association of functional polymorphisms in MMPs genes with gastric cardia adenocarcinoma and esophageal squamous cell carcinoma in high incidence region of North China*. *Mol Biol Rep*, 2010. 37(1): p. 197-205.
127. Delgado-Enciso, I., et al., *Matrix metalloproteinase-2 promoter polymorphism is associated with breast cancer in a Mexican population*. *Gynecol Obstet Invest*, 2008. 65(1): p. 68-72.
128. Grieu, F., W.Q. Li, and B. Iacopetta, *Genetic polymorphisms in the MMP-2 and MMP-9 genes and breast cancer phenotype*. *Breast Cancer Res Treat*, 2004. 88(3): p. 197-204.
129. Zhang, B., et al., *Functional polymorphism in the regulatory region of gelatinase B gene in relation to severity of coronary atherosclerosis*. *Circulation*, 1999. 99(14): p. 1788-94.
130. O'Mara, T.A., J.A. Clements, and A.B. Spurdle, *The use of predictive or prognostic genetic biomarkers in endometrial and other hormone-related cancers: justification for extensive candidate gene single nucleotide polymorphism studies of the matrix metalloproteinase family and their inhibitors*. *Cancer Epidemiol Biomarkers Prev*, 2009. 18(9): p. 2352-65.

131. Betz, M., et al., *1.8-A crystal structure of the catalytic domain of human neutrophil collagenase (matrix metalloproteinase-8) complexed with a peptidomimetic hydroxamate primed-side inhibitor with a distinct selectivity profile*. Eur J Biochem, 1997. 247(1): p. 356-63.
132. Saghatelian, A., et al., *Activity-based probes for the proteomic profiling of metalloproteases*. Proc Natl Acad Sci U S A, 2004. 101(27): p. 10000-5.
133. Batist, G., et al., *Neovastat (AE-941) in refractory renal cell carcinoma patients: report of a phase II trial with two dose levels*. Ann Oncol, 2002. 13(8): p. 1259-63.
134. Fridman, J.S., et al., *Selective inhibition of ADAM metalloproteases as a novel approach for modulating ErbB pathways in cancer*. Clin Cancer Res, 2007. 13(6): p. 1892-902.
135. Acuff, H.B., et al., *Analysis of host- and tumor-derived proteinases using a custom dual species microarray reveals a protective role for stromal matrix metalloproteinase-12 in non-small cell lung cancer*. Cancer Res, 2006. 66(16): p. 7968-75.
136. Yang, W., et al., *Human macrophage metalloelastase gene expression in colorectal carcinoma and its clinicopathologic significance*. Cancer, 2001. 91(7): p. 1277-83.
137. Coussens, L.M., et al., *MMP-9 supplied by bone marrow-derived cells contributes to skin carcinogenesis*. Cell, 2000. 103(3): p. 481-90.
138. Bremer, C., C.H. Tung, and R. Weissleder, *In vivo molecular target assessment of matrix metalloproteinase inhibition*. Nat Med, 2001. 7(6): p. 743-8.

Disclaimer: The views expressed in this article are those of the authors and do not reflect the official policy of the Department of the Army, Department of Defense, or the US Government



Breast cancer metastasis gene signature: the search for the Holy Grail  
RE Ellsworth, LA Field, B Love, JL Kane, JA Hooke, CD Shriver

**Background:** Lymph node status remains one of the most useful prognostic indicators in breast cancer, however, current methods to assess nodal status disrupt the lymphatic system lead to secondary complications, moreover, 30% of women diagnosed with negative lymph nodes will develop metastases. Identification of molecular signatures discriminating primary tumors from patients with the propensity to metastasize from those without would allow for stratification of patients requiring surgical assessment of lymph node status as well as identification of node negative patients at high-risk of developing metastatic disease.

**Methods:** Frozen breast specimens were collected from women with negative (n=39) and positive (n=36) lymph node status. Patients with positive BRCA1/BRCA2 status, neoadjuvant therapy or negative lymph node status who later developed distant metastasis were not included in this study. RNA was isolated from pure tumor cell populations after laser microdissection and gene expression data generated using HG U133A 2.0 arrays (Affymetrix). Data was randomly split into a training set (29 Negative and 26 Positive) and a validation set (10 Negative and 10 Positive) and a classifier was built using a forward-building Bayesian naïve classifier.

**Results:** Amongst all the pathological variables examined, including tumor type, grade, hormonal and HER2 status, only tumor size differed significantly, with patients with positive lymph nodes having significantly larger ( $P < 0.01$ ) tumors than those with negative lymph nodes. ANOVA analysis failed to identify a reliable signature discriminating metastatic from non-metastatic primary tumors. A median prediction probability  $> 95\%$  was achieved with the Bayesian Classifier using 17 markers (mean probability 94.31%, median 95.05%, min 74.62% and max 99.91%) in the training set. however, this gene set was only able to correctly classify 9 /20 samples in validation set, resulting in approximately random classification

**Conclusions:** The inability to derive molecular profiles of metastasis may reflect tumor heterogeneity (e.g., different methods of metastasis for each tumor subtype), the presence of only a few cells within the primary tumor that have metastatic potential, the influence of the microenvironment on metastasis, or an inherited host susceptibility to metastasis, rather than molecular properties of the tumor itself. Together, these data suggest that a molecular signature that can discriminate between primary breast tumors with and without lymph node metastases does not exist.

## Breast Cancer Metastasis Gene Signature: the Search for the Holy Grail

Rachel E. Ellsworth<sup>a</sup>, Lori A. Field<sup>b</sup>, Brad Love<sup>c</sup>, Jennifer L. Kane<sup>b</sup>, Jeffrey A. Hooke<sup>d</sup>,  
Craig D. Shriver<sup>d</sup>

<sup>a</sup>Clinical Breast Care Project, Henry M. Jackson Foundation for the Advancement of Military Medicine, Windber, PA USA; <sup>b</sup>Clinical Breast Care Project, Windber Research Institute, Windber, PA USA; <sup>c</sup>BioReka, LLC, Timonium, MD USA; <sup>d</sup>Clinical Breast Care Project, Walter Reed Army Medical Center, Washington, DC USA

Corresponding Author:

Rachel Ellsworth

Clinical Breast Care Project

Henry M. Jackson Foundation for the Advancement of Military Medicine

620 Seventh Street

Windber, PA 15963

Phone: (814) 467-9844

Fax: (814) 467-6334

Email: [r.ellsworth@wriwindber.org](mailto:r.ellsworth@wriwindber.org)

Clinical Breast Care Project

Windber Research Institute

620 Seventh Street

Windber, PA 15963

LA Field: [l.field@wriwindber.org](mailto:l.field@wriwindber.org)

JL Kane: [j.kane@wriwindber.org](mailto:j.kane@wriwindber.org)

BioReka, LLC

211 Locknell Rd

Timonium, MD 21093

B Love: [brad@bioreka.com](mailto:brad@bioreka.com)

Clinical Breast Care Project

Walter Reed Army Medical Center

6900 Georgia Avenue, NW

Washington, DC 20307

JA Hooke: [Jeffrey.hooke@us.army.mil](mailto:Jeffrey.hooke@us.army.mil)

CD Shriver: [craig.shriver@us.army.mil](mailto:craig.shriver@us.army.mil)

## **ABSTRACT**

Lymph node status remains one of the most useful prognostic indicators in breast cancer; however, current methods to assess nodal status disrupt the lymphatic system and may lead to secondary complications. Moreover, ~30% of women diagnosed with negative lymph nodes will develop metastatic disease. Identification of molecular signatures discriminating primary tumors from patients with the propensity to metastasize from those without would allow for the stratification of patients requiring surgical assessment of lymph node status as well as identification of patients at risk of developing metastatic disease. Thus, frozen breast specimens were collected from women with negative (n=39) and positive (n=35) lymph node status and after laser microdissection of pure tumor cell populations, gene expression data generated. Data was split into training and validation sets; a median prediction probability >95% was achieved with a Bayesian Classifier, however, this signature was only able to correctly classify 9/20 samples in a validation set, resulting in approximately random classification. The inability to derive molecular profiles of metastasis may reflect tumor heterogeneity, paucity of cells within the primary tumor with metastatic potential, influence of the microenvironment, or inherited host susceptibility to metastasis. A molecular signature discriminating primary breast tumors with and without lymph node metastases does not exist.

Keywords: breast cancer metastasis, molecular signature

Abbreviations:

ALND: axillary lymph node dissection

CBCP: Clinical Breast Care Project

ER: estrogen receptor

PCR: polymerase chain reaction

PR: progesterone receptor

qRT-PCR: quantitative real-time polymerase chain reaction

SLNB: sentinel lymph node biopsy

## 1. INTRODUCTION

Breast cancer is the most common cancer in women from Western countries. In 2009, approximately 190,000 women in the United States were diagnosed with and more than 40,000 died from breast cancer (American Cancer Society 2009). The progression of malignant breast cancer from localized to systemic disease can lead to impaired organ function, widespread systemic failure, and eventually, death. Five-year survival rates differ dramatically between women with negative lymph nodes (>90%) compared to those with axillary lymph node metastasis (<70%) (National Cancer Institute 1994).

Axillary lymph node status is not only the most reliable predictor of survival but also critical in developing treatment regimens (Eifel et al. 2001). Axillary lymph node dissection (ALND) remains the most accurate method of evaluating nodal status, however, ALND is associated with significant morbidities including disruption of nerves of the axilla, shoulder pain and weakness, and lymphedema (Newman and Newman 2007). Furthermore, ALND has not been associated with significant survival advantage (Fisher et al. 1985), thus alternate methods of evaluating lymph node status have been developed. Sentinel lymph node biopsy (SLNB) assesses lymph node status in the sentinel or first-draining nodes along the axillary lymph node chain; on average, two-three lymph nodes are removed and those patients with negative lymph node status are spared complete axillary dissection. While this less invasive procedure is associated with lower morbidities, patients may still experience complications including decreased range of motion, numbness and lymphedema, as well as a false negative rate of 8-10% (Newman and Newman 2007; Wilke et al. 2006).

Using lymph node status as a prognostic tool is not only impacted by potentially serious side effects and significant false negative rates, but also 30% of patients who have negative lymph node status at the time of diagnosis, develop distant metastasis and ~20% die within 15 years (Rosen et al. 1993; Michaelson et al. 2003). The development of new methods to predict metastatic propensity would allow for women at low-risk for developing metastasis to avoid surgical axillary lymph node evaluation while identifying those patients at high risk for developing metastasis. In this study, microarray-based gene expression analysis was performed on primary breast tumors from patients with and without metastatic lymph nodes to identify molecular signatures associated with breast cancer metastasis.

## **2. MATERIALS AND METHODS**

### *2.1 Tissue Samples*

Tissue and blood samples in the Clinical Breast Care Project (CBCP) tissue bank were collected with approval from the Walter Reed Army Medical Center Human Use Committee and Institutional Review Board. All subjects enrolled in the CBCP voluntarily agreed to participate and gave written informed consent. Clinical information was obtained for all CBCP samples using questionnaires designed by and administered under the auspices of the CBCP. The CBCP database was queried to identify all patients diagnosed with invasive cancer between 2001 and 2008. Patients with a previous history of breast cancer, documented BRCA1 or BRCA2 mutations, or who underwent neoadjuvant therapy were not eligible for this study. Patients with isolated tumor cells or micrometastases as well as those diagnosed with negative lymph node status who later

died of disease were excluded from analysis. Primary tumors from patients with positive (n=35) and negative (n=39) lymph node status were selected for gene expression analysis. Clinicopathological characteristics for these cases are shown in Table 1. An additional 56 primary tumors were selected for validation studies.

## *2.2 RNA Isolation, Amplification, aRNA Labeling and Hybridization*

For each case, hematoxylin and eosin stained slides were examined by a dedicated breast pathologist and tumor areas marked for laser microdissection. One to six serial sections (8  $\mu$ m thick) were cut, mounted on glass PEN foil slides (Leica Microsystems, Wetzlar, Germany), stained using the LCM staining kit (Applied Biosystems, Foster City, CA) and laser microdissected on an *ASLMD* laser microdissection system (Leica Microsystems, Wetzlar, Germany). Slide preparation, staining and cutting were performed within 15 minutes to preserve RNA integrity.

RNA was isolated from laser microdissected tumor cells using the RNAqueous-Micro kit (Applied Biosystems, Foster City, CA) and treated with DNase I to remove contaminating genomic DNA. RNA integrity was assessed using the 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA), and then converted to biotin-labeled aRNA using two rounds of amplification with the MessageAmpII aRNA Amplification kit according to the manufacturer's protocols (Applied Biosystems, Foster City, CA). Poly-A Controls (Affymetrix, Santa Clara, CA) were added to each RNA sample prior to the first round of amplification to monitor efficiency of amplification. Samples were labeled with biotin-11-UTP (Applied Biosystems, Foster City, CA) during the second *in vitro* transcription reaction. The concentration and quality of the aRNA samples were measured with the NanoDrop 1000 (Thermo Fisher Scientific, Waltham, MA) and the

2100 Bioanalyzer, respectively. Hybridization was performed using the HG U133A 2.0 arrays (Affymetrix, Santa Clara, CA) according to manufacturer's recommendations. Following hybridization, the arrays were washed and stained using the GeneChip Fluidics Station 450 and scanned with the GeneChip Scanner 3000.

### *2.3 Quantitative Real-time PCR*

For quantitative real-time PCR (qRT-PCR) analysis, 500 ng of laser microdissected RNA from 18 of the original samples plus 56 out-of-sample cases were reverse transcribed into cDNA using the High Capacity cDNA reverse transcription kit (Applied Biosystems, Foster City, CA). PCR amplification was performed using TaqMan Universal PCR Master Mix using standard conditions as recommended by the manufacturer (Applied Biosystems, Foster City, CA). Each TaqMan Gene Expression Assay was performed on duplicate samples using the iCycler Real-Time Detection System (Bio-Rad, Hercules, CA). B-actin was used as the endogenous control for normalization of all assays. Relative quantification of gene expression levels was determined using the Comparative  $C_t$  method.

### *2.4 Analysis and Statistics*

All Affymetrix chip data was normalized using a quantile normalization scheme, where the empirical cumulative distribution functions of the signals on the chip are made equal to the average across all chips. Each marker was then normalized across all experiments by taking the log base 2 of the quotient of the signal by the median marker signal.

For the first set of 40 samples (20 node negative, 20 node positive), a Mann-Whitney U-test was used to find differentially expressed markers, and p-values were

calculated using the standard combinatoric approach. Fold change was calculated by taking the log base 2 of the geometric mean of the node negative versus node positive primary tumor sample ratios. Genes were considered as differentially expressed if there was at least a 2-fold difference in levels of expression and  $P < 0.01$ . All clustering was done with a hierarchical clustering using complete linkage using a correlation coefficient metric. For the analysis of qRT-PCR data, the medians of the  $2^{-\Delta\Delta Ct}$  values for node negative and node positive tumors were compared using a Mann-Whitney U test to determine if the relative fold change was significantly different between the two groups.

For the model building approach, normalized data was randomly split into a training (29 negative and 25 positive) and a validation set (10 negative and 10 positive). A forward-building Bayesian naïve classifier was used to build a classifier by taking the training set and randomly splitting it into a preliminary training (24 negative and 20 positive) and testing (five negative and five positive) set. In the preliminary training set, classification probabilities were calculated based on intensities for each marker from the training set on the testing set such that for each of the preliminary markers a probability of observing this intensity based on the preliminary training data was calculated. For each of the 22,277 markers, there are ten classification probabilities; calculation of the minimum probability of the ten should be the minimum observed probability of correctly classifying the sample based on the marker intensity. This random splitting of the training data set was repeated and the minimum probability calculated 10,000 times. After the 10,000 random splits and predictions, the mean, median, minimum and maximum probabilities were calculated, and the marker with the highest median correct prediction probability selected. This process was repeated where a naïve Bayesian

classifier was performed with the selected marker and then tested with the remaining 22,276 markers; the next marker was then selected based on the 10,000 bootstraps based on the highest median probability prediction. This process was continued until the median prediction probability was over 95%. The final model with the set of predictors was then used to predict the validation set.

### **3. RESULTS**

#### *3.1 Clinicopathological characteristics*

Tumor samples used in this study were from women diagnosed with primary breast cancer between 2001 and 2008. The average age at diagnosis did not differ significantly between those with (55.1 years) and those without (57.7 years) lymph node metastasis. Amongst all the pathological variables examined (Table 1), including tumor type, grade, hormonal and HER2 status, only tumor size differed significantly, with patients with positive lymph nodes having significantly larger ( $P < 0.01$ ) tumors than those with negative lymph nodes. To date, none of the node negative patients have died of disease, and 24/39 (62%) have been disease-free for  $\geq 5$  years. In contrast, 6/35 (17%) of node positive patients died of disease, one (3%) has progressed from stage IIIa to stage IV, and 47% have remained disease-free for at least five years.

#### *3.2 Data analysis*

Gene expression data from a preliminary set of 40 tumors (20 lymph node negative and 20 lymph node positive) were assessed to identify genes differentially expressed between the two patient groups. Using a cutoff of  $\geq$ two-fold difference in

expression and  $P < 0.01$ , 224 probes corresponding to 198 characterized and 14 uncharacterized or putative genes were differentially expressed. Attempts to validate the ten top rank-ordered genes using qRT-PCR on 18 specimens from the original data set plus an additional 56 independent primary tumor samples failed to support the microarray results (Table 2).

Microarray-based gene expression data was generated for an additional 34 primary tumors (19 node negative and 15 node positive). Samples were randomized into test (20 node negative and 20 node positive) and training sets (19 node negative and 15 node positive) and Mann-Whitney analysis performed. This method failed to produce a stable molecular signature.

Given that the contribution of each individual gene may be small but together a multi-gene signature might be effective in discriminating node-negative from node-positive patients, an alternate statistical approach was employed to analyze the data. A forward-building Bayesian naïve classifier was used to identify genes that differed between the two groups in an iterative fashion, resulting in a group of genes that discriminate tumor types. Within the training set, the optimal classifier was developed using 17 genes with a mean of probability 94.31%, median of 95.05%, minimum of 74.62% and maximum of 99.91%. As with the earlier analysis, however, this 17-marker panel set, when applied to the validation set, could only predict 9/20 samples correctly (six node-negative and three node-positive samples) which approximates randomness.

### *3.3 Tumor phenotypes*

Given the marked tumor heterogeneity at both the pathologic and molecular levels, we evaluated whether subclasses of tumors existed within the 17-gene classifier. Tumor size, grade and histology did not influence sample clustering. Estrogen receptor (ER), progesterone receptor (PR) and HER2 status were used as surrogate markers for four intrinsic subtypes: luminal A = ER and/or PR+/HER2-, luminal B = ER and/or PR+/HER2+, HER2+ = ER and PR-/HER2+ and triple negative = ER, PR and HER2- (Carey, Perou et al. 2006). Luminal A was the predominant tumor type in both the LN positive (56%) and LN negative (69%) clusters (Figure 1), with smaller proportions of luminal B, HER2+ and triple negative tumors ( 8% vs. 7%; 15% vs. 17%, and 19% vs. 7% in lymph node positive and lymph node negative, respectively). Of note, all the lymph node positive triple negative tumors clustered together within the 17-gene metastasis signature, allowing for the possibility that triple negative tumors metastasize via different molecular pathways than tumors with luminal origins.

#### **4. DISCUSSION**

Breast cancer research has benefited tremendously from the development of whole-genome approaches such as array-based gene expression analysis. Whole-genome technologies not only allow for the evaluation of thousands of genes in a single experiment, but are also agnostic, requiring no *a priori* knowledge of genes differentially expressed between two disease states. Molecular signatures have been developed from whole-genome gene expression arrays that can classify breast tumors by intrinsic subtype, tumor grade, and risk of recurrence (Perou et al. 2000; Sotiriou et al. 2006; van't Veer et al. 2002), each of which can be used as a prognostic tool. The MammaPrint®

assay, based on a 70-gene signature developed using microarray data, is now available as a clinical test to predict outcome and identify those patients who would benefit from systemic therapy. While this 70-gene signature can predict outcome, it cannot predict lymph node status (van de Vijer et al. 2002)

While the 70-gene poor prognosis signature does not effectively classify primary tumors by lymph node status, efforts have been made to develop a breast tumor molecular signature that differs between patients with and without lymph node metastasis. Evaluation of gene expression patterns of 176 candidate genes between primary tumors without lymph node metastasis and those with 10 or more positive lymph nodes revealed differences in gene expression, with significantly higher expression of ERBB2 ( $P < 0.0001$ ) in tumors from node positive compared to node negative tumors (Bertucci et al. 2000). From a pool of 89 primary tumors, data from 19 primary tumors without lymph node metastasis and 18 with  $\geq 10$  positive lymph nodes were compared to generate a metagene profile, enriched for genes involved in cellular immunity, capable of predicting lymph node status with 90% accuracy (Huang et al. 2003). The use of Serial Analysis of Gene Expression in 27 invasive ductal carcinomas revealed 245 genes differentially expressed ( $P < 0.05$ ) between lymph node negative and lymph node positive primary tumors; results were validated in an independent set of tumors for seven genes including homeobox protein hox-c10 (HOXC10), tumor protein D52 like-1 (TPD52L1) and zinc finger protein 36 like-1 (ZFP36L1) (Abba et al. 2007).

In contrast, a number of research groups have failed to develop molecular signatures predictive of lymph node metastasis. Gene expression data from 129 primary

breast tumors was used to successfully develop signatures correlating expression patterns with grade and ER and HER2 status but a signature for lymph node status could not be identified; the authors thus concluded that while there may be a biological propensity to metastasize, the influence of time and stochastic processes on tumor metastasis may preclude the identification of a signature of lymph node metastasis (Lu et al. 2008). In a second study, evaluating microarray data from 151 lymph node negative and 144 lymph node positive primary tumors, Weigelt et al. did not identify gene expression differences between tumor types. The authors then applied the lymph node metastasis signature described by Huang et al. to their own external data set and achieved a classification accuracy of only 50%, implying that the signature developed by Huang, using a small sample set and limiting analysis to patients with  $\geq 10$  positive lymph nodes, is not an effective predictor of nodal metastasis (Weigelt et al. 2005). In our study, despite the use of different statistical approaches, genetic signatures of lymph node metastasis that could effectively classify tumors by lymph node status could not be developed. Together with those studies by Lu et al. and Weigelt et al., our data suggest that a molecular signature discriminating patients with lymph node metastasis from those without cannot be developed.

A number of reasons may explain the discrepancy between those groups that have reported molecular signatures of lymph node metastasis and those that have failed to find gene expression differences. For example, study design may affect the ability to detect critical molecular alterations. Most studies identifying a signature of lymph node metastasis relied on small (<40 samples) sample sizes. In addition, multiple models were developed using patients with extremely discordant (negative lymph node status

compared to  $\geq 10$  positive lymph nodes) phenotypes (Bertucci et al. 2000;Huang et al. 2003); these models, therefore, may not apply to the majority of patients who have an intermediate number of positive lymph nodes (Lu et al. 2008). Finally, validation of these signatures on independent sample sets has, to our knowledge, not been reported, and to date, while molecular portraits are used to determine tumor grade, subtype and prognosis, no clinical assay is available to determine lymph node status.

In addition to methodological concerns, lack of a signature of lymph node metastasis may be attributable to biological properties of primary breast tumors, such as the nature and number of cells within a primary tumor with metastatic potential. Injection of melanoma cells into mice demonstrated that tumor cells vary widely in their ability to produce metastases, and cells with metastatic potential are rare within the primary tumor (Fidler and Kripke 1977;Poste and Fidler 1980). This view was challenged by the development of gene expression signatures, such as the 70-gene poor prognosis signature and a molecular signature of metastasis developed from solid tumors (Ramaswamy et al. 2003;van't Veer et al. 2002); because these signatures were derived from bulk tumors, the authors concluded that the majority of cells in the primary tumor have the ability to metastasize. In fact, the ability to predict which tumors will metastasize based on gene signatures derived from primary tumors does not preclude the presence of small subpopulations of cells with full metastatic potential found in localized regions throughout the primary tumor (Fidler and Kripke 2003;Welch 2004). For example, comparison of gene expression patterns between cell line populations that have high compared to low metastatic potential to bone revealed that only a small fraction of cells demonstrated the full bone metastasis signature (Kang et al. 2003). More recently,

the sequencing of a basal-like primary breast tumor and corresponding brain metastasis revealed a significant enrichment of 20 mutations in the metastasis compared to the primary tumor, suggesting that metastases arise from a minority population of cells within the primary tumor (Ding et al. 2010). If these models in which few cells within the primary tumor have full metastatic capacity are correct, genetic signatures from these rare cells will be masked by the majority of tumor cells which do not have full metastatic capacity.

Molecular heterogeneity within tumor subtypes may also preclude the identification of a single signature of metastasis. Breast tumors can be classified by their intrinsic subtypes, including luminal A, luminal B, HER2 positive and basal-like subtype, based on different patterns of gene expression (Perou 2000). These subtypes have been associated with differences in relapse-free and overall survival with the basal-like and HER2 positive subtypes having the shortest survival times (Sorlie et al. 2001). Not only do intrinsic subtypes have different prognoses, but recent studies have shown that each subtype has preferential sites of metastasis: bone was the predominant site of relapse in luminal and HER2 positive tumors but was infrequent in basal-like tumors. In contrast, basal-like tumors had frequent relapse in brain, lung and distant lymph nodes (Kennecke et al. 2010; Smid et al. 2008). Although formal tumor subtyping was not performed, a recent study in lymph node negative primary breast tumors found non-overlapping signatures of metastasis for ER positive and ER negative tumors, suggesting that there are different molecular mechanisms associated with metastasis depending on tumor biology (Yu et al. 2007). Although in our study subtypes did not group cleanly within each cluster

(node negative and node positive), the clustering of node positive triple negative samples suggests that molecular signatures of metastasis may be subtype-specific.

The ability to metastasize may be influenced by not only the tumor cells, but also the microenvironment, both local and distant. Dissemination of tumor cells from the primary site is one of the earliest steps of metastasis; successful invasion and migration of tumor cells requires a number of changes in the breast microenvironment including epithelial-mesenchymal transition (EMT), degradation of the extracellular matrix and angiogenesis. Distant tissue may be subjected to premetastatic niche conditioning, undergoing changes such as recruitment of bone-marrow derived cells that form a favorable environment for tumor cells to grow. Finally, the last stages of metastasis require tumor cells to successfully reach the secondary site, escape senescence and survive and proliferate within a foreign environment (Nguyen and Massagué 2007; Bidard et al. 2008). Given the importance of the microenvironment, molecular characterization of the tumor component alone may not be sufficient in predicting metastatic behavior as a tumor with an aggressive profile may be growing within a non-permissive microenvironment and vice versa. In fact, many signatures of poor prognosis or metastasis include the expression of stromal genes. Thus, consideration of only the tumor epithelial component may fail to capture the full metastatic potential of a primary tumor.

Finally, the ability to metastasize may depend not on biologic features of the primary tumor but on inherent host susceptibility. Out-crossing of a highly metastatic transgenic mouse to a variety of inbred mouse strains resulted in significant variability in

the propensity to metastasize; since each animal received the metastatic transgene, the differences in metastatic capacity have been attributed to genetic background (Hunter 2005). Linkage studies identified candidate metastasis modifier genes in mouse, including Sip1 (Pletcher et al. 2004). Follow-up studies in humans confirmed that SIPA1 is a metastasis susceptibility gene (Crawford et al. 2006;Hsieh et al. 2009). Thus, the ability to successfully metastasize may, at least in part, reflect a systemic, rather than tumor-driven, proclivity.

## **5. CONCLUSIONS**

Despite the use of multiple statistical models, a reliable molecular signature discriminating metastatic from non-metastatic tumors could not be identified. Reasons for this may include both biological properties of the tumor, such as heterogeneity of intrinsic subtypes, paucity of cells within the primary tumor with full metastatic capacity, and influence of the microenvironment, both local and distant, as well as inherited susceptibility to metastasize. Given the current lack of a single, well-validated, molecular assay to discriminate primary tumors with from those without metastatic potential, alternate hypotheses, such as whether different tumor subtypes have unique signatures of metastasis, must be pursued. Although multiple signatures of breast cancer metastasis may have to be developed, these molecular signatures are needed to effectively discriminate patients with and without the propensity to metastasize, thus sparing those at low risk from unnecessary surgical evaluation, while identifying those with favorable pathologic characteristics, who are, nevertheless, likely to develop metastatic disease.

## **ACKNOWLEDGEMENTS**

We thank Dr. Darrell Ellsworth for helpful and critical review of this manuscript. This research was supported by a grant from the United States Department of Defense (Military Molecular Medicine Initiative MDA W81XWH-05-2-0075, Protocol 01-20006). The opinion and assertions contained herein are the private views of the authors and are not to be construed as official or as representing the views of the Department of the Army or the Department of Defense.

## REFERENCES

1. Abba MC, Sun H, Hawkins KA, Drake JA, Hu Y et al. 2007. Breast cancer molecular signatures as determined by SAGE: correlation with lymph node status. *Mol.Cancer Res.* 5, 881-889.
2. American Cancer Society. Breast cancer facts and figures: 2009-2010. 2009. Atlanta, GA, American Cancer Society.
3. Bertucci F, Houlgatte R, Benziane A, Granjeaud S, Adelaide J, et al. 2000. Gene expression profiling of primary breast carcinomas using arrays of candidate genes. *Hum.Mol.Genet.* 9, 2981-2991.
4. Bidard F-C, Pierga J-Y, Vincent-Salomon A and Poupon M-F, 2008. A "class action" against the microenvironment: do cancer cells cooperate in metastasis? *Cancer Metastasis Rev.* 27, 5-10.
5. Carey LA, Perou CM, Livasy CA, Dressler LG, Cowan D, et al. 2006. Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. *J.A.M.A.* 295, 2492-2502.
6. Crawford NP, Ziogas A, Peel DJ, Hess J, Anton-Culver H and Hunter KW 2006. Germline polymorphisms in SIPA1 are associated with metastasis and other indicators of poor prognosis in breast cancer. *Breast Cancer Res.* 8, R16.
7. Ding L, Ellis MJ, Li S, Larson DE, Chen K et al., 2010. Genome remodeling in a basal-like breast cancer metastasis and xenograft. *Nature* 464, 999-1005.
8. Eifel P, Axelson JA, Costa J, Crowley J, Curran WJ.Jr. et al., 2001. National Institutes of Health Consensus Development Conference Statement: adjuvant therapy for breast cancer, November 1-3, 2000. *J.Natl.Cancer Inst.* 93, 979-989.
9. Fidler IJ and Kripke M 1977. Metastasis results from preexisting variant cells within a malignant tumor. *Science* 197, 893-895.
10. Fidler IJ and Kripke ML 2003. Genomic analysis of primary tumors does not address the prevalence of metastatic cells in the population. *Nature Genet.* 34, 23.
11. Fisher B, Redmond C, Fisher ER, Baum M, Wolmark N et al., 1985. Ten-year results of a randomized clinical trial comparing radical mastectomy and total mastectomy with or without radiation. *N.Engl.J.Med.* 312, 674-681.
12. Hsieh FC, Look MP, Sieuwerts AM, Foekens JA and Hunter KW, 2009. Distinct inherited metastasis susceptibility exists for different breast cancer subtypes: a prognosis study. *Breast Cancer Res.* 11, R75.

13. Huang E, Cheng SH, Dressman H, Pittman J, Tsou MH, et al., 2003. Gene expression predictors of breast cancer outcomes. *Lancet* 361, 1590-1596.
14. Hunter KW 2005. The intersection of inheritance and metastasis: the role and implication of germline polymorphism in tumor dissemination. *Cell Cycle* 4, 1719-1721.
15. Kang Y, Siegel PM, Shu W, Drobnjak M, Kakonen SM, et al., 2003. A multigenic program mediating breast cancer metastasis to bone. *Cancer Cell* 3, 537-549.
16. Kennecke H, Yerushalmi R, Woods R, Cheang MC, Voduc Det al., 2010. Metastatic behavior of breast cancer subtypes. *J.Clin.Oncol.* 28, 3271-3277.
17. Kluger HM, Lev DC, Kluger Y, McCarthy MM, Kiriakova G, et al., 2005. Using a xenograft model of human breast cancer metastasis to find genes associated with clinically aggressive disease. *Cancer Res.* 65, 5578-5587.
18. Lu X, Lu X, Wang ZC, Iglehart JD, Zhang W, et al., 2008. Predicting features of breast cancer with gene expression patterns. *Breast Cancer Res* 108, 191-201.
19. Michaelson JS, Silverstein M, Sgroi D, Cheongsiatmoy JA, Taghian A, et al., 2003. The effect of tumor size and lymph node status on breast carcinoma lethality. *Cancer* 98, 2133-2143.
20. National Cancer Institute. SEER cancer statistics review, 1973-1991. NIH Publ. No. 94-2789. 1994. Bethesda, MD, NIH.
21. Newman EA and Newman LA, 2007. Lymphatic mapping techniques and sentinel lymph node biopsy in breast cancer. *Surg.Clin.North Am.* 87, 353-364.
22. Nguyen DX and Massagué J, 2007. Genetic determinants of cancer metastasis. *Nature Rev.Genet.* 8, 341-352.
23. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, et al., 2000. Molecular portraits of human breast tumours. *Nature* 406, 747-752.
24. Pletcher MT, McClurg P, Batalov S, Su AI, Barnes SW, et al., 2004. Use of a dense single nucleotide polymorphism map for in silico mapping in the mouse. *PLoS Biol.* 2, e393.
25. Poste G, and Fidler IJ, 1980. The pathogenesis of cancer metastasis. *Nature* 283, 139-146.
26. Ramaswamy S, Ross KN, Lander ES and Golub TR, 2003. A molecular signature of metastasis in primary solid tumors. *Nat.Genet.* 33, 49-54.

27. Rosen PP, Groshen S, Kinne DW and Norton L, 1993. Factors influencing prognosis in node-negative breast carcinoma: analysis of 767 T1N0M0/T2N0M0 patients with long-term follow-up. *J.Clin.Oncol.* 11, 2090-2100.
28. Smid M, Wang Y, Zhang Y, Sieuwerts AM, Yu J et al., 2008, Subtypes of breast cancer show preferential site of relapse. *Cancer Res.* 68, 3108-3114.
29. Sorlie T, Perou CM, Tibshirania R, Aas T, Geisler S, et al., 2001. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl. Acad. Sci. USA* 98, 10869-10874.
30. Sotiriou C, Wirapati P, Loi S, Harris A, Fox S, et al., 2006. Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. *J.Natl.Cancer Inst.* 98, 262-272.
31. van't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, et al., 2002. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 415, 530-536.
32. van de Vijer MJ, He YD, van't Veer LJ, Dai H, Hart AA, et al., 2002. A gene-expression signature as a predictor of survival in breast cancer. *N.Engl.J.Med.* 347, 1999-2009.
33. Weigelt B, Wessels LFA, Bosma AJ, Nuyten DSA, He YD, et al., 2005. No common denominator for breast cancer lymph node metastasis. *Br.J.Cancer* 93, 924-932.
34. Welch DR 2004. Microarrays bring new insights into understanding of breast cancer metastasis to bone. *Breast Cancer Res.* 6, 61-64.
35. Wilke LG, McCall LM, Posther KE, Whitworth PW, Reintgen DS, 2006. Surgical complications associated with sentinel lymph node biopsy: results from a prospective international cooperative group trial. *Ann.Surg.Oncol.* 13, 491-500.
36. Yu JX, Sieuwerts AM, Zhang Y, Martens JWM, Smid M, et al., 2007. Pathway analysis of gene signatures predicting metastasis of node-negative primary breast cancer. *BMC Cancer* 7, 182.

## **FIGURE LEGEND**

Figure 1. Dendrogram of 54 samples in the training set used to generate the 17-gene signature. Blue and red boxes represent node positive and node negative primary breast tumors, respectively.

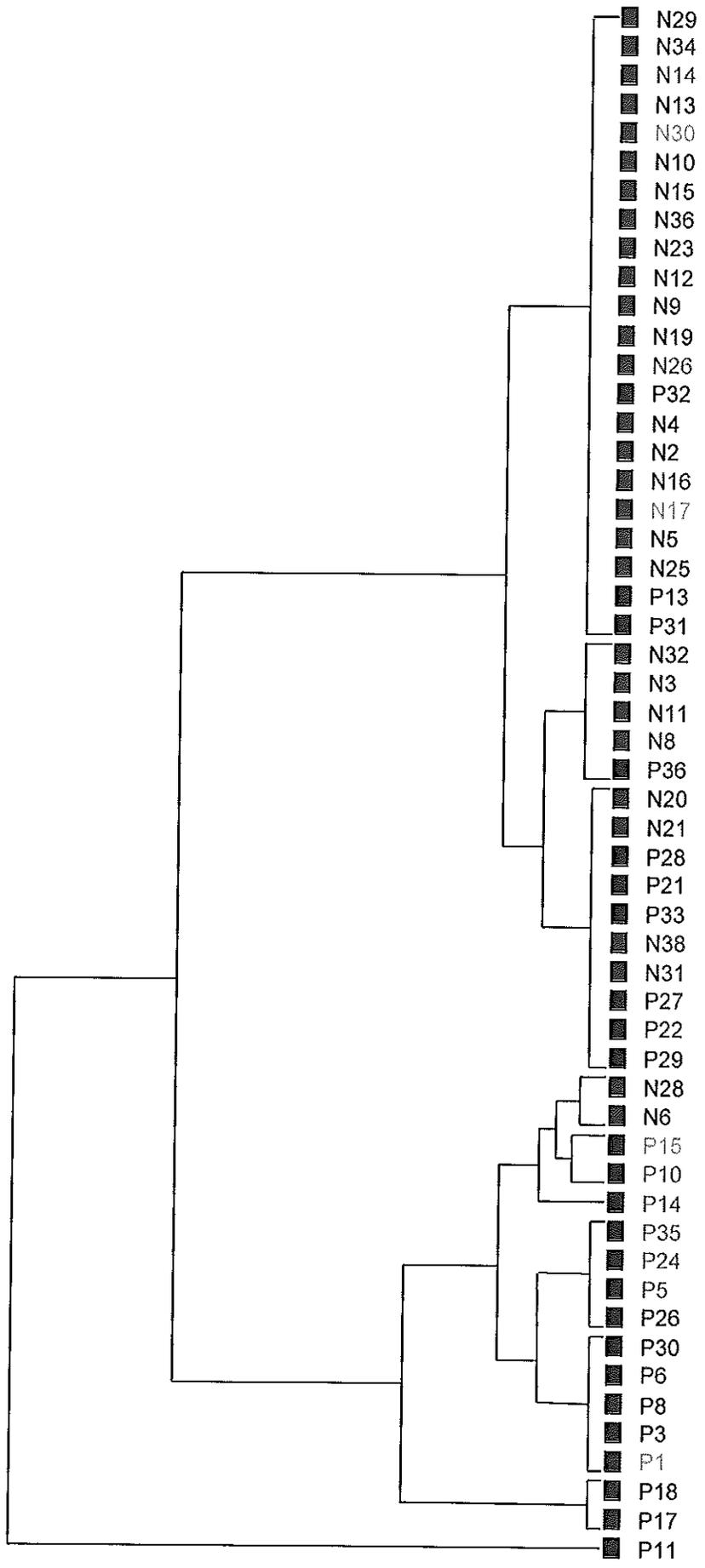
Table 1. Clinical and pathological features of 74 invasive breast tumor specimens used in microarray analysis

	Node Negative (n=39)	Node Positive (n=35)	P Node- vs. Node+
<b>Age</b>			
<50 years	39%	39%	NS
≥50 years	61%	61%	
<b>Histology</b>			
IDCA	95%	89%	NS
ILCA	5%	9%	
Mixed	0%	2%	
<b>Grade</b>			
Well-differentiated	28%	9%	NS
Moderately-differentiated	28%	40%	
Poorly-differentiated	44%	51%	
<b>Hormone Receptor Status<sup>a</sup></b>			
ER+/PR+	49%	50%	NS
ER+/PR-	25%	14%	
ER-/PR-	26%	36%	
<b>HER2 Status</b>			
Positive	18%	22%	NS
Negative	82%	78%	
<b>Tumor Size</b>			
T1	67%	34%	P<0.01
T2	33%	57%	
T3	0%	9%	

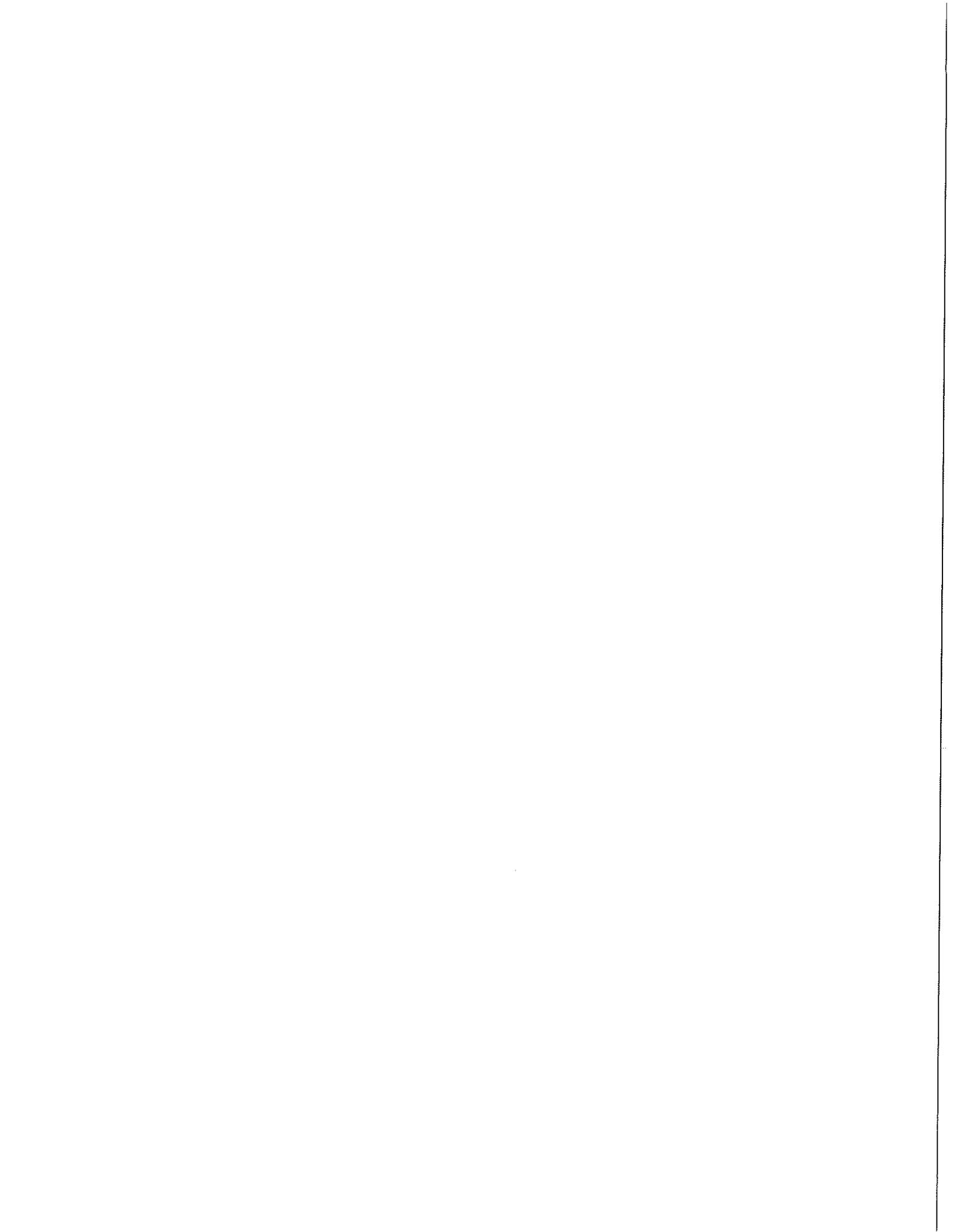
<sup>a</sup>No cases of ER-/PR+ were identified in this group of tumors

Table 2. Top ten rank-ordered genes after ANOVA analysis of 20 node negative and 20 node positive primary breast tumors. qRT-PCR was performed on 74 samples, including 56 out-of-sample tumor specimens

Probe ID	Gene Symbol	Fold-change	P-value Microarray	P-value qRT-PCR
212025_s_at	FLII	9.13	0.00006	0.23
214545_s_at	PROSC	8.52	0.00053	0.12
221885_at	DENND2A	4.82	0.00021	0.77
210764_s_at	CYR61	4.23	0.00048	0.60
218215_s_at	NR1H2	6.60	0.00087	0.81
202024_at	ASNA1	3.43	0.00017	0.40
211529_x_at	HLA-G	3.68	0.00043	0.24
65133_i_at	ZNHIT4	4.46	0.00083	0.90
210196_s_at	PSG1	7.05	0.00128	0.61
208729_x_at	HLA-B	3.94	0.00065	0.13



- Luminal A
- Luminal B
- HER2+
- Triple Negative



## **Association of Clinicopathologic Characteristics with IHC-based Breast Cancer Subtypes**

Jasmine Saini, MS, Windber Research Institute  
Al Kovatich, MS, MDR Global Systems,  
Xing Li, PhD, Windber Research Institute  
Prema Rapuri, PhD, Windber Research Institute  
Anthony Bekash, MD, Windber Research Institute  
Jeffrey A Hooke, MD, Clinical Breast Care Project, Walter Reed Army Medical Center  
Richard J Mural, PhD, Windber Research Institute  
Craig D. Shriver, MD, Clinical Breast Care Project, Walter Reed Army Medical Center  
Hai Hu \*, PhD, Windber Research Institute

\*Corresponding author:

Hai Hu, PhD  
Windber Research Institute  
620 7<sup>th</sup> St.  
Windber, PA 15963  
(814)361-6903 (Office)  
(814)467-6334 (Fax)  
[h.hu@wriwindber.org](mailto:h.hu@wriwindber.org)

## ABSTRACT:

**Introduction** Invasive breast cancers have been classified by gene expression into the following major subclasses – luminal A, luminal B, basal-like, normal-breast-like and HER2+. Recently, there have been reports to characterize these subclasses of cancer based on immunohistochemistry (IHC) assays of estrogen receptor (ER), progesterone receptor (PR) and human epidermal growth factor receptor 2 (HER2). In this study we attempt to understand the association between IHC based breast cancer subtypes and clinicopathologic characteristics, benign breast diseases (BBDs), *in situ* disease and ethnicity specific clinicopathologic properties.

**Methods** The study consists of a total of 482 biopsy confirmed invasive breast cancer patients classified into breast cancer subtypes, based on IHC results ER, PR and HER2. Clinical characteristics, BBDs, *in situ* diseases were studied for association with the subtypes. The data were analyzed using Pearson's chi-square test. A secondary analysis was done with Caucasian and African American ethnicities to attain association between ethnicity specific clinicopathologic properties and breast cancer subtypes.

**Results** A total of 255 (53.0%) luminal A (ER+, PR+, HER2-), 108 (22.4%) luminal B (ER+, PR+/-, HER2+/-), 38 (7.9%) HER2+ (ER-, PR-, HER2+) and 81 (16.8%) triple negative (ER-, PR-, HER2-). Luminal subtypes were mostly old postmenopausal women. High tumor grade, larger tumor size, positive lymph node status and higher AJCC stage were seen more frequently among the triple negative and HER2+ subtypes. Luminal A subtype was highly associated with BBDs. HER2+ subtype had the highest occurrence of ductal carcinoma *In Situ* (p=0.026) while luminal A subtype is more frequent with lobular carcinoma *In Situ* (p=0.0002). Caucasian luminal group was significantly different from the rest of the groups when comparing tumor grade, tumor stage, tumor size, menopausal status and prior use of hormone replacement therapy.

**Conclusions** This study had a unique population of patients where majority of them had equal access to healthcare and we confirmed the reported differential association of subclasses with T-stage, tumor grade, lymph node status, menopausal status and age. In addition, we found that several BBDs and *in situ* disease differentially associate with subtypes.

## **1. INTRODUCTION:**

Breast cancer is the most commonly diagnosed invasive cancer among women and is the second leading cause of cancer mortality in United States. In 2009, it is estimated that 192,370 women will be diagnosed with and 40,170 women will die of breast cancer [1]. Breast cancer has its distinctive molecular and clinical heterogeneity. The causes of this heterogeneity could be due to patient's genetic makeup, lifestyle differences and nutritional/environmental influences. Based on gene expression profiles of breast cancer and normal breast samples, Perou et al. [2],[3] and Sorlie et al. [4] classified breast carcinomas into at least five intrinsic subtypes with distinct clinical outcomes—luminal A, luminal B, HER2+, basal-like and normal-breast-like.

These intrinsic subtypes have been consistently reproduced across a variety of microarray platforms [5],[6] and also have been validated in various different datasets [7, 8]. In the North Carolina Breast Cancer Study, Immunohistochemistry (IHC) surrogate markers have been developed for the first time that can be applied to paraffin embedded tissue for identifying intrinsic breast cancer subtypes and have been verified against gene expression profiles [9]. In the last 2-3 years, the IHC-based classification into breast cancer subtypes as a surrogate has been in a number of studies to analyze the associations between the breast cancer subtypes and clinicopathologic factors, survival and treatment prediction [5, 10-13].

Luminal A and luminal B subtypes express ER and/or PR and are associated with good prognosis and low risk of recurrence. HER2 subtype patients over-express HER2 but not ER and PR. The expression profile of triple negative patients lacks all 3 receptors – ER, PR and HER2. Triple negative and HER2+ tumors show aggressive disease with

large tumor size (>2cm), high tumor grade, lymph node involvement and have high risk of recurrence [11]. Also, triple negative tumors occur mostly among younger African American women when compared with postmenopausal African American or other ethnic population [9].

Racial disparity between Caucasian and African American women has been reported in the incidence and mortality rates from breast cancer. White women have an overall higher incidence of breast cancer than African American women; however African American women have the higher mortality rates and worse survival prognosis for breast cancer compared to other ethnic groups. The poorer outcome for African American women has been attributed to several causes including differences in tumor biology which contributes to the tumor aggressiveness and resistance to the treatment [9, 14, 15], access to quality healthcare resulting in advanced stages of breast cancer in case of African Americans [16], differences in quality of care after diagnosis [17-19] and the socioeconomic factors influencing treatment options [20, 21].

The Clinical Breast Care Project (CBCP) is a US Congressionally mandated breast cancer program funded by the Department of Defense. The clinical facilities for this program are provided mainly at Walter Reed Army medical Center (WRAMC) where equal-access health care is provided to military personnel and dependents. Clinical data is collected using 2 questionnaires - a Core Questionnaire and a Pathology Checklist, and a rigorous quality assurance process is applied.

In the present study, using the CBCP population with largely equal access to healthcare and data collected with similar protocols, we examined the distinction among the breast cancer subtypes classified based on IHC surrogate markers and the risk factors associated with these different subtypes. We also studied the association of ethnicities (African American and Caucasian - the two major ethnic groups in our population) and breast cancer subtypes in this population where the impact of healthcare related socioeconomic factors between the ethnicities are minimized.

## **2. MATERIALS AND METHODS:**

### **2.1 CBCP population:**

All the patients in this study are from CBCP program. CBCP is congressionally mandated military-civilian collaboration, mainly between WRAMC in Washington, D.C, and the Windber Research Institute in Windber, PA. It is an ongoing project with active subject enrollment. The inclusion criteria for the CBCP program includes adults over the age of 18 yrs, mentally competent and willing to provide informed consent, presented to breast centers/clinics with the evidence of possible breast disease or for elective reductive mammoplasty. Normal subjects are also enrolled. HIPPA complaint IRB-approved protocols are used to collect samples of breast and metastatic tissues and blood for the study. At the time of data analysis for this publication more than 4000 patients were recruited into the CBCP.

### **2.2 Data collection - Demographic, lifestyle, health and reproductive data:**

The enrolled subjects were interviewed by research nurses in person to complete a comprehensive life history questionnaire - Core Questionnaire; consisting of 427 data elements covering demographics, health history (i.e. family cancer/other disease

history, reproductive history, hormone and menstrual history, patient's prior treatments/biopsies, etc.) and lifestyle practices (i.e. smoking, alcohol consumption, exercise frequency, etc.). The ethnicity of the subjects was determined by self-identification. The patients can be categorized into following categories based upon their menopausal status - premenopausal, post menopausal, status post hysterectomy (ovaries not removed) and surgically menopausal (both ovaries removed). For this analysis, we only used premenopausal and postmenopausal subjects while studying association of menopausal status with the breast cancer subtypes.

### **2.3 Tumor pathology data:**

For the CBCP subjects requiring a biopsy, a Pathology Checklist is completed which records information regarding the presence of 131 breast disease diagnoses, disease specific biomarkers like ER, PR and HER2 status, tumor grade, tumor size, pathological stage, lymph node metastases, etc. totaling 372 data fields. The pathological stage of the tumor was determined based on the American Joint Commission on Cancer Staging Manual, 6<sup>th</sup> edition, 2002 [22]. The primary T-stage definitions were followed as defined by the American Joint Commission on Cancer staging Manual, 6<sup>th</sup> edition, 2002 [22]. The grade of invasive tumor was determined according to the Nottingham criteria [23].

### **2.4 Immunohistochemistry (IHC)**

Routinely prepared formalin-fixed paraffin-embedded blocks were used for IHC staining which was performed at MDR global systems, LLC Windber, PA. Formalin-fixed paraffin embedded sections were cut at 4 micrometers and placed on positively charged slides. Antibodies were optimized by determining appropriate dilution, incubation time and antigen recovery. Briefly, monoclonal antibodies against ER (clone 1D5/Mouse; Dako),

PR (clone PGR636/Mouse; Dako) and HER2 (clone Poly/Rabbit) were used for IHC staining. The slides were dried at 59°C. The ER and PR slides were placed in Subx solution prior to deparaffination and Dako PT Module was used for antigen recovery. It was followed by washing the slides thoroughly in distilled water before placing them in TBS buffer. The HER2 slides were deparaffinized through a series of xylenes and alcohols and the antigen was recovered in 95°C water bath in Epitope Retrieval Solution. The immunostaining was performed on the Dako Autostainer wherein the primary antibodies were applied and appropriate detection system was then dispensed onto the slides. "DAB" hydrogen peroxide was used for color visualization. Upon completion of staining, all slides were counterstained with hematoxylin. Immunoassaying was conducted according to established protocols and positive and negative controls were included in all staining runs.

IHC stained slides were evaluated for ER, PR expression and HER2 protein over-expression. If >5% tumor cells showed nuclear staining for ER or PR, the tumor was considered ER or PR positive, otherwise the tumor was ER or PR negative. HER2 was scored using the recommended CAP and ASCO guidelines. IHC assay for HER2 was labeled negative for scores 0 or 1+ (absence or <10% partial membranous staining), intermediate for scores 2+ (10-30% weak to moderate membranous staining) or positive for scores 3+ (>30% strong complete membranous staining). Any samples with an intermediate IHC score (2+) for HER2 were further analyzed with the FISH test using Path Vysion HER2-DNA probe kit, and the final HER2 status was positive if the gene was amplified otherwise negative. For histological reasons 102 samples had only FISH test information for HER2 marker to determine the overall HER2 status. Our prior findings showed 95% concordance between the IHC and FISH test results when the

samples are negative (FISH score <1.8) and 70% concordance when the samples are positive (FISH score >2.2).

## **2.5 Sample selection for the study**

Biopsy confirmed invasive female breast cancer patients with known ER, PR and HER2 status were selected. Patients with ER- and PR+ status are a very rare population and were removed from the study (total of 2 patients – Table 1). There are two schools of thought for the definition of luminal B in the literature, some authors suggest luminal B subtype is always HER2 positive [9, 24] while others suggest it could be HER2 negative also [25]. Our definition of luminal B group therefore included ER+ PR- HER2- women and ER+ PR+/- HER2+ women.

The Pathology Checklist records 68 BBDs categorized as inflammatory and reactive lesions, miscellaneous benign lesions, papillary lesions, hyperplasia and columnar cell lesions and biphasic tumors, they were included in the analysis along with ductal or lobular *In Situ* cancers findings. All of these data fields were graded as either being present or absent. Of these 68 BBDs, 24 were observed to co-occur in the invasive study population with a frequency greater than 1% (>5 patients) and were used for analysis. All these benign and *In Situ* findings were observed with the invasive carcinoma from the patient but could be present on different paraffin slides at the time of patient diagnosis.

All of the 482 invasive breast cancer cases were then sub-classified into subtypes based on the IHC results of ER, PR, and HER2 status as detailed in Table1. In these breast cancer subtypes, differences with regard to clinicopathologic characteristics and risk factors were assessed in the entire study group. A secondary analysis was

*Confidential Material – Breast Cancer Subtypes Manuscript*  
performed on the two major ethnic groups of the population, Caucasian and African American, for a better understanding of ethnicity-specific association of clinicopathologic properties with the breast cancer subtypes.

## **2.6 Statistical Analysis**

Statistical analysis was performed using R [26]. Differences between breast cancer subtypes with regard to clinicopathologic characteristics and risk factors were examined using one-way ANOVA for continuous variables using  $p=0.05$  as the threshold for statistical significance, followed by a t-test to identify the significantly contributing groups. For categorical values, Pearson's Chi-square test was used with  $p=0.05$  as the threshold for statistical significance. This was followed by examining the residues from each group and the one with the highest absolute residual value was determined as the group contributing most to the significance. Comparison of the clinicopathologic factors according to breast cancer subtypes in the two ethnicities, Caucasian and African American was also performed using one-way ANOVA for continuous variables and by Pearson's Chi-square test for the categorical variables.

## **3. RESULTS**

The final analysis was performed on the dataset of 482 subjects. In this population, 350 (72.6%) were Caucasian women, 91 (18.9%) were African American, and 41(8.5%) belonged to other ethnicities. Of these 482 subjects, 255 (53%) were luminal A, 108 (22.4%) were luminal B, 81 (16.8%) were triple negative and 38 (7.9%) were HER2+ subtype.

### **3.1 Comparison of clinicopathologic factors among the breast cancer subtypes**

The differences with respect to clinicopathologic factors between the four major IHC-based subtypes of breast cancer are presented in Table 2. The molecular subtypes were significantly associated with the patient's age at diagnosis ( $p=0.03$ ). Women with HER2+ breast cancer subtype (mean $\pm$ SD = 53 $\pm$ 10.36 years) were youngest at diagnosis, followed by women with triple negative (mean $\pm$ SD = 56 $\pm$ 12.10 years), luminal B (mean $\pm$ SD = 58 $\pm$ 13.32 years) and luminal A (mean $\pm$ SD = 60 $\pm$ 12.51 years). Luminal A subtype was diagnosed in patients as young as 27 years and as old as 93 years old (Figure 1). The results of the t-test analysis showed that the luminal A group was significantly older at diagnosis than the triple negative group ( $p=0.025$ ) and HER2+ group ( $p=0.0058$ ) and luminal B group subjects tended to be marginally older at diagnosis than HER2+ group women ( $p=0.068$ ).

Higher tumor grade, greater tumor size, higher AJCC stage and positive lymph node status were highly associated with triple negative and HER2+ subtypes. Significant differences among the breast cancer subtypes were observed based on histological grade ( $p=1.92E-38$ ). Triple negative (85.9%) and HER2+ (88.6%) subtypes had the highest prevalence of poorly differentiated tumors (Grade 3). Luminal A and luminal B subtypes had more frequently well/moderately differentiated tumors (Grade 2 and Grade 1, respectively). Luminal A (75.3%) and luminal B (63.9%) subtypes had mostly patients with small tumor sizes (T1- tumor  $\leq$  2 cm ) in comparison to HER2+ (58.3%) and triple negative (51.3%) groups ( $p=1.75E-06$ ). Higher number of women in the HER2+ (11.1%) subtype had tumor T-stage 4. Luminal A (62.0%) and luminal B (51.9%) patients were mostly in AJCC stage I when compared to HER2+ (39.4%) and triple negative patients (32.9%) ( $p=0.000047$ ). The HER2+ (45.9%) and triple negative

(42.7%) subtypes had higher number of subjects with positive lymph node status compared to that of luminal B (33.3%) and luminal A (25.7%) subtypes ( $p=0.0081$ ).

Both HER2+ (84.2%) and triple negative (84.0%) groups had higher frequency of infiltrating ductal carcinoma than luminal A (68.2%) and luminal B (70.4%) groups ( $p=0.015$ ). HER2+ (13.15%) group also had the highest frequency of micro invasive ductal carcinoma when compared with other groups (triple negative – 7.4%, luminal B - 2.8%, luminal A – 1.2%,  $p=0.0004$ ). On the other hand, luminal A (18.4%) and luminal B (17.6%) groups had higher frequency of infiltrating lobular carcinoma in comparison with HER2+ (2.6%) and triple negative (2.5%) groups ( $p=0.00045$ ).

Luminal A (65.3%) and luminal B (59.1%) subtypes were mostly postmenopausal in comparison to HER2+ (45%) and triple negative (50%) subtypes which had both pre and post menopausal women in almost equal numbers ( $p=0.04$ ). The distribution of ethnicity varied among the subtypes ( $p = 0.0029$ ). Luminal A (84.9%) and HER2+ (85.3%) subtypes were dominantly among the Caucasian Americans whereas African Americans were present relatively high in triple negative (33%) subtype and luminal B (26.3%) subtypes. Luminal B (33.3%) and luminal A (25.3%) subtypes had higher number of subjects with positive history of breast cancer in primary blood family member compared to women in triple negative (20%) and HER2+ (13.2%) subtypes ( $p=0.0492$ ). Patients with a history of hormonal replacement therapy (HRT) use were mostly luminal A (47.0%) subtype while patients with no history of HRT use mostly belonged to luminal B (69.4%) subtype ( $p=0.0043$ ). Women who breastfed were less frequent among the triple negative (40.5%) subtype but were mostly in the HER2+ (67.6%) subtype ( $p=0.035$ ).

No significant difference was observed among the subtypes in the age of menarche, although the triple negative population had a mean age of menarche of 13.11 years which was slightly older than the rest of the population. There were also no significant differences observed in the BMI of the premenopausal and postmenopausal women among the 4 breast cancer subtypes. Mean menopausal age, pregnancy (ever/never), birth control pill use (yes/no), history of diabetes, hypertension, stroke or coronary artery disease, were not significantly different among the four subtypes of breast cancer (data not shown).

### **3.2 Comparison of BBD pathology among the breast cancer subtypes:**

The data on the synchronous presence of benign breast diseases with the invasive tumor sample in a patient is shown in Table 3. Most of the 24 BBDs co-occurred among the luminal A subtype of breast cancer however only 3 statistically significant ones are listed. Luminal A group (25.5%) had higher counts of atypical ductal hyperplasia, followed by luminal B (17.6%), HER2+ (2.6%) and triple negative (7.4%) groups ( $p=0.0001$ ). Luminal A group (71.0%) also had more frequent fibrocystic changes in comparison to the other subtypes (luminal B – 57.4%, HER2+ – 52.6%, triple negative – 58%,  $p=0.0126$ ). Changes of intraductal hyperplasia (mild) were again observed mostly among luminal A (9.8%) group (luminal B – 1.9%, HER2+ – 0.0%, triple negative – 7.4%,  $p=0.0144$ ).

### **3.3 Comparison of In situ pathology among the breast cancer subtypes:**

As shown in Table 3, HER2+ subtype group (81.6%) had highest number of ductal carcinoma *In Situ* (DCIS) followed by luminal B (74.1%) and luminal A groups (63.1%); triple negative group had the least (60.5%) ( $p=0.026$ ). The distribution of lobular carcinoma *in situ* (LCIS) was also different among the 4 breast cancer subtypes, the luminal A (26.3%) and luminal B (22.2%) groups had more cases with lobular carcinoma *in situ* in comparison with HER2+ (7.9%) and triple negative (6.2%) groups ( $p=0.0002$ ).

### **3.4 Comparison of clinicopathologic factors between the breast cancer subtypes in relation to ethnicity:**

A secondary analysis was performed by separating the study subjects based on ethnicity to compare the clinicopathologic factors in the Caucasian ( $n = 350$ ) and African American subjects ( $n = 91$ ) with the breast cancer subtypes and the results are shown in Table 4. Other ethnic groups were removed from this analysis as the numbers were too low to do the analysis. Both the Caucasian and African American populations were mostly positive for ER and have HER2 negative status. On the other hand, African American subjects (56.0%) were mostly PR negative while Caucasian subjects (62.2%) were mostly PR positive ( $p=0.002$ ) (Supplement Table 1).

In our first analysis of the association of the clinicopathologic factors with the breast cancer subtypes including the entire dataset of 482 subjects (presented above), we found that luminal A and luminal B groups showed similar trends while triple negative and HER2+ groups showed similar trends. Therefore, for studying the association of ethnicity and breast cancer subtypes with respect to clinicopathologic factors, we combined luminal A and B groups into luminal group and triple negative and HER2+ groups into non-luminal group. We then compared the clinicopathologic factors between

the luminal and non-luminal groups in the two ethnicities. Table 4 shows only significant findings.

Non-luminal African American population was youngest (mean±SD = 54±12.3 years), followed by non-luminal Caucasian population (mean±SD = 55±11.3 years), luminal African American population (mean±SD = 58±14 years) and the oldest being luminal Caucasian population (mean±SD = 60±12 years) (p = 0.0009). Luminal subtypes of both ethnicities had higher percentage of Grade 1 and Grade 2 tumors while non-luminal subtypes had higher count of Grade 3 tumors (African American women = 89.7%, Caucasian = 85.3%, p = 1.05E-33). Women belonging to the Caucasian luminal group had highest number of subjects with T-stage 1 (73.3%, p = 0.00089) and AJCC stage1 breast cancer (59.6%, p = 4.71E-05). On the other hand, African American in both luminal group (6.6%) and non-luminal group (6.7%) had higher percent of women with AJCC stage IV (p = 4.71E-05) breast cancer compared to other groups. Positive lymph node status was high among non-luminal Caucasian women (48.6%) (African American luminal = 30.5%, African American non-luminal = 28.1%, Caucasian luminal = 34.5%; p =0.0103). Infiltrating ductal carcinoma was mostly found among non-luminal population, irrespective of the ethnicity (African American = 90.0%, Caucasian = 80.0%, p=0.0206). Interestingly, microinvasive ductal carcinoma was least likely to be found in Caucasian luminal population (0.7%) and was almost equally found among both of the non-luminal groups (African American luminal = 4.9%, African American non-luminal = 10.0%, Caucasian non-luminal = 8.8%; p =0.00065). However, Infiltrating lobular carcinoma (African American = 18.0%, Caucasian = 17.8%; p=0.0012) and infiltrating carcinoma (NOS) (African American = 11.5%, Caucasian = 7.4%; p=0.033) were frequent in both African American and Caucasian luminal groups.

Postmenopausal status was predominantly observed in the Caucasian luminal group (68.6%,  $p = 0.00055$ ) and also, higher number of women in this group had a history of HRT use (45.5%,  $p = 0.026$ ). Family history of breast cancer was least prevalent in non-luminal Caucasian (11.3%  $p = 0.0114$ ). Higher number of women in the Caucasian luminal group (54.9%) had a history of breast feeding compared to other groups ( $p=0.038$ ). Overall, Caucasian luminal group was significantly different from the rest of the groups when comparing tumor grade, tumor stage, tumor size, menopausal status and prior use of hormone replacement therapy.

Comparing the benign breast diseases pathology and *in situ* pathology among the luminal and non-luminal groups of the Caucasian and African American populations (Table 5), it was observed that the luminal groups were different from non-luminal groups but both the ethnicities showed similar trends under their respective subtypes. Luminal African American and Caucasian women in comparison to non-luminal had higher number of subjects with diagnosis of atypical ductal hyperplasia (African American- 21.3% vs. 3.3%; Caucasian = 23.0% vs. 7.5%  $p=0.00197$ ), apocrine metaplasia (African American- 52.5% vs. 33.3%; Caucasian = 44.4 vs. 30.0%  $p=0.0287$ ). Benign findings of columnar cell change were mostly synchronously present among non-luminal African American (50.0%  $p=0.0269$ ). *In situ* carcinomas of DCIS type were not found to be significant but LCIS type were highly frequent among the luminal African American (26.2%) and luminal Caucasian (25.2%) women ( $p=7.84E-06$ ) compared to other groups.

#### **4. DISCUSSION**

In the present study, we observed that breast cancer subtypes defined by surrogate IHC markers are significantly associated with clinical and pathological factors, with triple negative and HER2+ subtypes demonstrating unfavorable clinical and pathological prognostic features. We also report for the first time differential association of benign breast diseases and in situ carcinomas (present in the same tumor tissue) with breast cancer subtypes. Majority of our observations are in agreement with other studies which looked at similar associations [9, 11, 24, 27-33]. Overall, our results suggest that women belonging to triple negative and HER2+ subtypes demonstrated higher tumor grade, greater tumor size, higher AJCC stage and positive lymph node status. Our analyses on the breast cancer subtypes and ethnicity revealed that Caucasian luminal group women were significantly different from the rest of the groups with respect to tumor grade, tumor stage, tumor size etc. and had overall association with better prognostic factors.

One of the most important factors contributing to the disparities among the African American and Caucasian is attributed to the unequal access to health care based on the patient's socioeconomic status [34]. The delay in receiving health care could lead to higher mortality rates and higher tumor grade and stage which are frequently observed among the African American population. The availability of the same quality health care to all patients irrespective of their ethnic background in the present study minimizes the contribution of health care disparities. It must also be noted that easy access to health care provides an opportunity for early screening and detection of breast cancer. Due to this opportunity of early detection, we have a higher number of African American population and the mean age of our 482 subjects is lower than the SEER database [35]. Additional strengths of the present study are the use of same standardized methods to

collect the data from the study subjects; pathology reports of >70% subjects being generated by one pathologist and IHC marker information being provided from one laboratory. The above unique aspects of study population enabled us to study the association of breast cancer subtypes and clinicopathologic characteristics in Caucasian and African American women devoid of any major socioeconomic confounding factors.

In the present study, age was significantly associated with breast cancer subtypes with HER2+ group being the youngest. Most of the published studies however reported triple negative population to be the youngest [9, 24, 30-32, 36]. Contrary to this, Bhargava et al [25] classifying invasive tumors into 6 categories found luminal B population to be the youngest. In a population from Africa, Huo et al [28] found no significant differences in the mean age between the breast cancer subtypes. In a cohort of 257 patients with invasive ductal carcinoma NOS, Munoz et al [37] reported HER2+ population to be the youngest and triple negative to be the oldest among the breast cancer subtypes. Thus, study population probably influences mean age at diagnosis among the breast cancer subtypes. In the present study, luminal A and luminal B population were older postmenopausal women while HER2+ and triple negative women were significantly younger premenopausal patients. Triple negative subtype predominantly consists of African American women, which is in agreement with earlier reports [9, 29, 33]. Luminal A and luminal B subtypes also had lower tumor grade, small tumor size, and lower AJCC stage and most of the patients were also lymph node negative. Several studies from different countries and in different ethnic populations have examined the risk factor profiles of breast cancer subtypes [9, 12, 28, 32] and our results are consistent with these observations. It has been consistently reported that basal-like and Her2 subtypes

have an unfavorable prognostic features [4, 9, 30, 32] similar to our results. The ethnicity-based secondary analysis revealed that Caucasian and African American luminal groups had lower tumor grade, smaller tumor size, early tumor size and higher frequency of women with negative lymph node status compared to non-luminal subtypes. Consistent with these observations, Ihemelandu et al studying exclusively the African American women observed that in both premenopausal and postmenopausal women, basal-like and HER2+ subtype were associated with aggressive features while luminal A and B subtypes were associated with better prognostic features [24, 38].

In the present study, the type of cancer by histology was also significantly associated with the breast cancer subtypes, with lobular carcinoma mostly in luminal A group and infiltrating and microinvasive ductal carcinoma mostly in HER2+ and triple negative women. When examined according to ethnicity we again found infiltrating lobular carcinoma and infiltrating carcinoma NOS to be higher in luminal Caucasian and African American women and microinvasive and infiltrating ductal carcinoma to be mostly in nonluminal cases. A similar association has been reported by Spitale et al [32] Yang et al [12] and Carey et al [9] who found higher incidence of lobular carcinomas in luminal A group, while Spitale et al [32] reported higher incidence of ductal carcinoma in HER2+ women.

In the present study, luminal B group had higher frequency of women with family history of breast cancer; higher number of women in luminal A group had history of HRT use and breast feeding was protective for triple negative group. When examined according to ethnicity, luminal Caucasian women were highest prior HRT users; lowest number of women in the nonluminal Caucasian group had family history of breast cancer and

breastfeeding was protective for African American women irrespective of breast cancer subtype. Kwan et al [29] also reported that triple negative women were less likely to breastfeed and luminal A group women to have higher HRT use. Regarding the association with family history, contradictory results have been published. Yang et al [12] and Haffty et al [36] reported a strong association between family history and triple negative women compared to other breast cancer subtypes, while Millikan et al [39] reported no association of family history with any of the subtypes. Phipps et al [31] reported higher percent of luminal patients with positive first-degree family history of breast cancer.

BBDs in women encompass a heterogeneous group of lesions and few of these have been considered to have premalignant potential. Majority of published studies addressed the relative risk for breast cancer from BBDs using subjects diagnosed with BBD during the breast biopsy performed for diagnostic purpose [40-44]. Age and previous family history of breast cancer were reported to be significantly associated with BBDs and women with certain BBDs were reported to have an increased risk of developing future breast cancer. The risks associated with BBD were reported not to be different across ethnicities [45]. In the present study, interestingly we observed co-occurrence of some BBDs in the same tumor tissue and the percentage of women with BBD co-occurrence was higher in the luminal A breast cancer subtype compared to others. Incidentally, women of this group were of highest age and had marginally higher number of subjects with positive family history compared to women of triple negative and HER2+ subtypes. In concurrence with these observations, when we examined the data according to ethnicity, BBDs co-occurrence was again higher in luminal groups of both Caucasian and African American women. There have not been any studies in the

*Confidential Material – Breast Cancer Subtypes Manuscript*

literature which reported the co-occurrence of BBDs with the invasive lesions. However, Worsham et al observed that higher number of older women present with atypia [44] and women with higher incidence of BBDs appear to have a family history of breast cancer in a first degree relative [41, 46]. Thus, it is likely that luminal group of women with advanced age tend to accumulate benign lesions which in due course may lead to malignancy lending support to the hypothesis that breast cancer develops through a linear multistep progressive accumulation of destabilization events. In fact, Hartmann et al reported occurrence of breast cancer during follow up in the same breast in which the benign lesions were detected especially in women with atypia, suggesting that malignant precursors can occur within the benign lesions [41]. Genetic factors could also play a role in the increase in breast cancer risk for women with BBDs. Jorgensen et al found that allelic variants in the DNA repair genes were significantly associated with BBDs, more so in women with family history of breast cancer, again suggesting that BBDs could be an intermediate phenotype in the breast carcinogenesis pathway [47]. Evidence from CGH and LOH studies also suggested that pre-invasive lesions of the breast harbor chromosomal abnormalities at the same loci known to be altered in invasive breast carcinoma again supporting the multistep model for breast carcinogenesis (transition from normal epithelium to invasive via non-atypical and atypical hyperplasia and in situ carcinoma) [48].

In situ carcinomas have also been considered as intermediate in the linear progression hypothesis of breast cancer. DCIS and LCIS are thought to be precursors for IDC and ILC respectively. In this study, we found that higher frequency of women of both luminal A and luminal B subtypes of breast cancer had LCIS co-existence with invasive cancer while higher frequency of women in HER2+ subtype had DCIS co-occurrence with

invasive cancer. When examined according to ethnicity, the co-occurrence of DCIS was not different among the groups; however the co-occurrence of LCIS was significantly higher in luminal groups of Caucasian and African American compared to non-luminal Caucasian and African American women. Spitale et al also reported synchronous presence of an in situ component with the invasive lesion, the prevalence being highest in HER2+/neu subtype and low in basal cell-like (BCI) cases. They proposed that BCI cancers may arise anew while HER2+ cancers may arise through progression of precursor lesions [32]. In agreement of this, Allred et al reported overexpression of HER2+/neu in 56% of pure DCIS and it plays an important role in initiation of ductal carcinomas [49]. None of the other studies examining the clinicopathologic characteristics in relation to breast cancer subtypes looked into this aspect.

Another view that has been proposed is the existence of distinct genetic pathways of tumor evolution which underlie the pathogenesis of different breast tumor subtypes and that they should be studied as distinct entities [50-52]. Using genome wide array based comparative genomic hybridization, Bergamaschi et al and Chin et al reported higher number of gains/losses association with basal like tumor subtype while high-level DNA amplifications were more frequent in luminal B subtype [50, 51]. Similar observations were made with respect to IHC subtypes of familial breast cancer [52].

To explain the different co-occurrences of benign and in situ with the invasive cancer in the present study, the view proposed by Melchor et al would be most suitable. In a recent review, Melchor and Benitez proposed a combination of cancer stem cell (CSC) hypothesis and clonal selection model for tumor development in breast cancer subtypes, CSC hypothesis for the carcinogenesis processes and the clonal selection

*Confidential Material – Breast Cancer Subtypes Manuscript*  
model for the tumor development process [53]. They suggest that carcinogenic events in an ER negative stem cell /progenitor cell would give rise to basal-like, ERBB2 or luminal B breast cancer subtypes, an ER-positive PC would produce a luminal A phenotype supporting a hierarchical breast carcinogenesis model, whereas crucial genomic imbalances are clonally selected during the tumor development. Additional studies are however needed to better understand the breast carcinogenesis processes in each of the breast cancer subtypes.

Some of the limitations of our study include lack of treatment and survival data and limitations due to patient number. However, important treatment and 5 year survival information has just started to become available for majority of the participants and protocols for quality control, data extraction and storage are being perfected. We plan to incorporate this information along with other distinguishing biomarkers like p53 and ki67 and wish to conduct additional analysis on these patients.

In conclusion, we demonstrated a significant association between breast cancer subtypes classified based on surrogate IHC markers and clinicopathologic factors, with triple negative and HER2+ phenotypes associated with unfavorable prognosis factors and luminal A and luminal B phenotypes associated with favorable prognosis factors irrespective of ethnicity and most of the BBDs being associated with the luminal groups. IHC surrogate marker breast cancer sub-classification can give valuable information for prognostic and predictive purposes.

## **LIST OF ABBREVIATIONS**

IHC: Immunohistochemistry

CBCP: Clinical Breast Cancer Project

WRAMC: Walter Reed Army Medical Center

HRT: Hormone Replacement Therapy

HER2: Human Epidermal Growth Factor Receptor 2

DCIS: Ductal Carcinoma *In Situ*

LCIS: Lobular Carcinoma *In Situ*

BMI: Body Mass Index

BCI: Basal Cell-Like

CSC: Cancer Stem Cell

AA: African American

CA: Caucasian

## **COMPETING INTERESTS**

The authors declare that they have no competing interests.

## **AUTHORS CONTRIBUTIONS**

JS – performed database queries and subject selection, led the data analysis, result interpretation, and manuscript preparation.

AK - co-initiated and co-guided this research project with HH, and led the IHC assays.

XL - contributed to data analysis as a senior data analyst.

PR - contributed to the interpretation of the data and manuscript preparation.

AB - initiated the ethnicity-specific data analysis, and provided clinical knowledge to the understanding of the data.

JAH - reviewed the pathology slides and completed the Pathology Checklist for breast cancer cases, and provided breast cancer pathology domain knowledge.

RJM - provided guidance to data analysis and interpretation

CDS - PI of the CBCP; provided clinical guidance and helped in interpreting the results

HH - initiated the research project with AK, formed and managed the project team, and supervised the data analysis, result interpretation, and manuscript preparation.

## **ACKNOWLEDGEMENTS**

We thank the patients who participated in this study. This work was supported by the Clinical Breast Care Project with funds from the US Department of Defense through Henry Jackson Foundation for the Advancement of Military Medicine, Rockville, MD. The views expressed in this manuscript are those of the authors and do not reflect the official policy of the Department of the Army, Department of Defense, or the government of the United States.

## **REFERENCE:**

1. SEER, [http://seer.cancer.gov/csr/1975\\_2006/](http://seer.cancer.gov/csr/1975_2006/).
2. Perou, C.M., et al., *Distinctive gene expression patterns in human mammary epithelial cells and breast cancers*. Proc Natl Acad Sci U S A, 1999. **96**(16): p. 9212-7.
3. Perou, C.M., et al., *Molecular portraits of human breast tumours*. Nature, 2000. **406**(6797): p. 747-52.
4. Sorlie, T., et al., *Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications*. Proc Natl Acad Sci U S A, 2001. **98**(19): p. 10869-74.
5. Sorlie, T., et al., *Distinct molecular mechanisms underlying clinically relevant subtypes of breast cancer: gene expression analyses across three different platforms*. BMC Genomics, 2006. **7**: p. 127.

6. Hu, Z., et al., *The molecular portraits of breast tumors are conserved across microarray platforms*. BMC Genomics, 2006. **7**: p. 96.
7. Sorlie, T., et al., *Repeated observation of breast tumor subtypes in independent gene expression data sets*. Proc Natl Acad Sci U S A, 2003. **100**(14): p. 8418-23.
8. Calza, S., et al., *Intrinsic molecular signature of breast cancer in a population-based cohort of 412 patients*. Breast Cancer Res, 2006. **8**(4): p. R34.
9. Carey, L.A., et al., *Race, Breast Cancer Subtypes, and Survival in the Carolina Breast Cancer Study*. JAMA, 2006. **295**(21): p. 2492-2502.
10. Cheang, M.C., et al., *Ki67 index, HER2 status, and prognosis of patients with luminal B breast cancer*. J Natl Cancer Inst, 2009. **101**(10): p. 736-50.
11. Tamimi, R.M., et al., *Comparison of molecular phenotypes of ductal carcinoma in situ and invasive breast cancer*. Breast Cancer Res, 2008. **10**(4): p. R67.
12. Yang, X.R., et al., *Differences in risk factors for breast cancer molecular subtypes in a population-based study*. Cancer Epidemiol Biomarkers Prev, 2007. **16**(3): p. 439-43.
13. Conforti, R., et al., *Breast cancer molecular subclassification and estrogen receptor expression to predict efficacy of adjuvant anthracyclines-based chemotherapy: a biomarker study from two randomized trials*. Ann Oncol, 2007. **18**(9): p. 1477-83.
14. Curtis, E., et al., *Racial and ethnic differences in breast cancer survival: how much is explained by screening, tumor severity, biology, treatment, comorbidities, and demographics?* Cancer, 2008. **112**(1): p. 171-80.
15. Amend, K., D. Hicks, and C.B. Ambrosone, *Breast cancer in African-American women: differences in tumor biology from European-American women*. Cancer Res, 2006. **66**(17): p. 8327-30.
16. Li, C.I., K.E. Malone, and J.R. Daling, *Differences in breast cancer stage, treatment, and survival by race and ethnicity*. Arch Intern Med, 2003. **163**(1): p. 49-56.
17. Ballard-Barbash, R., et al., *Factors associated with surgical and radiation therapy for early stage breast cancer in older women*. J Natl Cancer Inst, 1996. **88**(11): p. 716-26.
18. Freedman, R.A. and E.P. Winer, *Reducing disparities in breast cancer care: a daunting but essential responsibility*. J Natl Cancer Inst, 2008. **100**(23): p. 1661-3.
19. Bickell, N.A., et al., *Missed opportunities: racial disparities in adjuvant breast cancer treatment*. J Clin Oncol, 2006. **24**(9): p. 1357-62.
20. Mandelblatt, J.S., et al., *Variations in breast carcinoma treatment in older medicare beneficiaries: is it black or white*. Cancer, 2002. **95**(7): p. 1401-14.
21. Bradley, C.J., C.W. Given, and C. Roberts, *Race, socioeconomic status, and breast cancer treatment and survival*. J Natl Cancer Inst, 2002. **94**(7): p. 490-6.
22. Greene, F.L., et al., *AJCC cancer Staging Manual (6th Edition)*. New York: Springer-Verlag 2002.
23. Elston, C.W. and I.O. Ellis, *Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up*. Histopathology, 1991. **19**(5): p. 403-10.
24. Ihemelandu, C.U., et al., *Molecular breast cancer subtypes in premenopausal and postmenopausal African-American women: age-specific prevalence and survival*. J Surg Res, 2007. **143**(1): p. 109-18.
25. Bhargava, R. and D.J. Dabbs, *Luminal B breast tumors are not HER2 positive*. Breast Cancer Res, 2008. **10**(5): p. 404; author reply 405.
26. R, <http://www.R-project.org>.
27. Bhargava, R., et al., *Prevalence, morphologic features and proliferation indices of breast carcinoma molecular classes using immunohistochemical surrogate markers*. Int J Clin Exp Pathol, 2009. **2**(5): p. 444-55.

28. Huo, D., et al., *Population differences in breast cancer: survey in indigenous African women reveals over-representation of triple-negative breast cancer*. J Clin Oncol, 2009. **27**(27): p. 4515-21.
29. Kwan, M.L., et al., *Epidemiology of breast cancer subtypes in two prospective cohort studies of breast cancer survivors*. Breast Cancer Res, 2009. **11**(3): p. R31.
30. Onitilo, A.A., et al., *Breast cancer subtypes based on ER/PR and Her2 expression: comparison of clinicopathologic features and survival*. Clin Med Res, 2009. **7**(1-2): p. 4-13.
31. Phipps, A.I., et al., *Body size and risk of luminal, HER2-overexpressing, and triple-negative breast cancer in postmenopausal women*. Cancer Epidemiol Biomarkers Prev, 2008. **17**(8): p. 2078-86.
32. Spitale, A., et al., *Breast cancer classification according to immunohistochemical markers: clinicopathologic features and short-term survival analysis in a population-based study from the South of Switzerland*. Ann Oncol, 2009. **20**(4): p. 628-35.
33. Trivers, K.F., et al., *The epidemiology of triple-negative breast cancer, including race*. Cancer Causes Control, 2009. **20**(7): p. 1071-82.
34. Smith-Bindman, R., et al., *Does utilization of screening mammography explain racial and ethnic differences in breast cancer?* Ann Intern Med, 2006. **144**(8): p. 541-53.
35. Zhu, K., et al., *Cancer incidence in the U.S. military population: comparison with rates from the SEER program*. Cancer Epidemiol Biomarkers Prev, 2009. **18**(6): p. 1740-5.
36. Haffty, B.G., et al., *Locoregional relapse and distant metastasis in conservatively managed triple negative early-stage breast cancer*. J Clin Oncol, 2006. **24**(36): p. 5652-7.
37. Munoz, M., et al., *Prognostic significance of molecular classification of breast invasive ductal carcinoma*. Arch Gynecol Obstet, 2009. **280**(1): p. 43-8.
38. Ihemelandu, C.U., et al., *Molecular breast cancer subtypes in premenopausal African-American women, tumor biologic factors and clinical outcome*. Ann Surg Oncol, 2007. **14**(10): p. 2994-3003.
39. Millikan, R.C., et al., *Epidemiology of basal-like breast cancer*. Breast Cancer Res Treat, 2008. **109**(1): p. 123-39.
40. Bertelsen, L., et al., *Benign breast disease among first-degree relatives of young breast cancer patients*. Am J Epidemiol, 2008. **168**(3): p. 261-7.
41. Hartmann, L.C., et al., *Benign breast disease and the risk of breast cancer*. N Engl J Med, 2005. **353**(3): p. 229-37.
42. Tamimi, R.M., et al., *Benign breast disease, recent alcohol consumption, and risk of breast cancer: a nested case-control study*. Breast Cancer Res, 2005. **7**(4): p. R555-62.
43. Wang, J., et al., *Lower-category benign breast disease and the risk of invasive breast cancer*. J Natl Cancer Inst, 2004. **96**(8): p. 616-20.
44. Worsham, M.J., et al., *Risk factors for breast cancer from benign breast disease in a diverse population*. Breast Cancer Res Treat, 2009. **118**(1): p. 1-7.
45. Worsham, M.J., et al., *Breast cancer incidence in a cohort of women with benign breast disease from a multiethnic, primary health care population*. Breast J, 2007. **13**(2): p. 115-21.
46. Webb, P.M., et al., *Family history of breast cancer, age and benign breast disease*. Int J Cancer, 2002. **100**(3): p. 375-8.
47. Jorgensen, T.J., et al., *DNA repair gene variants associated with benign breast disease in high cancer risk women*. Cancer Epidemiol Biomarkers Prev, 2009. **18**(1): p. 346-50.
48. Reis-Filho, J.S. and S.R. Lakhani, *The diagnosis and management of pre-invasive breast disease: genetic alterations in pre-invasive lesions*. Breast Cancer Res, 2003. **5**(6): p. 313-9.

49. Allred, D.C., et al., *Overexpression of HER-2/neu and its relationship with other prognostic factors change during the progression of in situ to invasive breast cancer*. *Hum Pathol*, 1992. **23**(9): p. 974-9.
50. Bergamaschi, A., et al., *Distinct patterns of DNA copy number alteration are associated with different clinicopathological features and gene-expression subtypes of breast cancer*. *Genes Chromosomes Cancer*, 2006. **45**(11): p. 1033-40.
51. Chin, K., et al., *Genomic and transcriptional aberrations linked to breast cancer pathophysiologies*. *Cancer Cell*, 2006. **10**(6): p. 529-41.
52. Melchor, L., et al., *Distinct genomic aberration patterns are found in familial breast cancer associated with different immunohistochemical subtypes*. *Oncogene*, 2008. **27**(22): p. 3165-75.
53. Melchor, L. and J. Benitez, *An integrative hypothesis about the origin and development of sporadic and familial breast cancer subtypes*. *Carcinogenesis*, 2008. **29**(8): p. 1475-82.

**Table1: Detailing of the breast cancer subtype classification**

<b>ER</b>	<b>PR</b>	<b>HER2</b>	<b>Subtype</b>
Neg	Neg	Neg	Triple Neg
Neg	Neg	Pos	HER2+
Pos	Neg	Pos	Luminal B
Pos	Pos	Pos	Luminal B
Pos	Neg	Neg	Luminal B
Pos	Pos	Neg	Luminal A
Neg	Pos	Pos	None in our study
Neg	Pos	Neg	Removed from study

Table 2: Distribution of clinicopathological factors among the four subtypes

	All cases n=482	Luminal A n=255 (52.9%)	Luminal B n=108 (22.4%)	Her2 n=38 (7.9%)	Triple Negative n=81 (16.8%)	p
Age						
Mean±SD(years)	58±12.65	60±12.51	58±13.32	53±10.36	56±12.1	0.003
BMI						
Mean±SD	27.89±5.91	27.51±6.08	26.58±4.77	27.13±5.96	28±4.78	0.335
Grade, n (%)						
1	153 (32.3%)	117 (46.2%)	31 (28.7%)	1 (2.8%)	4 (5.1%)	1.92E-38
2	154 (32.5%)	104 (41.1%)	40 (37%)	3 (8.6%)	7 (9.0%)	
3	167 (35.2%)		37 (34.3%)	31 (88.6%)	67 (85.9%)	
T-stage, n (%)						
T1	322 (67.5%)	192 (75.3%)	69 (63.9%)	21 (58.3%)	40 (51.3%)	1.75E-06
T2	127 (26.6%)	54 (21.2%)	31 (28.7%)	9 (25%)	33 (42.3%)	
T3	21 (4.4%)	8 (3.1%)	7 (6.5%)	2 (5.6%)	4 (5.1%)	
T4	7 (1.5%)	1 (0.4%)	1 (0.9%)	4 (11.1%)	1 (1.3%)	
AJCC stage, n (%)						
Stage I	255 (53.1%)	158 (62%)	56 (51.9%)	15 (39.5%)	26 (32.9%)	0.000047
Stage II	161 (33.5%)	77 (30.2%)	34 (31.5%)	12 (31.6%)	38 (48.1%)	
Stage III	50 (10.4%)	15 (5.9%)	13 (12.0%)	10 (26.3%)	12 (15.2%)	
Stage IV	14 (2.9%)	5 (2%)	5 (4.6%)	1 (2.6%)	3 (3.8%)	
Lymph node status, n (%)						
Positive	147 (31.8%)	63 (25.7%)	35(33.3%)	17 (45.9%)	32 (42.7%)	0.0081
Negative	315 (68.2%)	182 (74.3%)	70 (66.7%)	20 (54.1%)	43 (57.3%)	
Tumor characteristics						
Infiltrating ductal carcinoma, n (%)						
Yes	350 (72.6%)	174 (68.2%)	76 (70.4%)	32 (84.2%)	68 (84.0%)	0.0145
No	132 (27.3%)	81 (31.8%)	32 (29.6%)	6 (15.8%)	13 (16.0%)	
Microinvasive ductal carcinoma, n (%)						
Yes	17 (3.5%)	3 (1.2%)	3 (2.8%)	5 (13.2%)	6 (7.4%)	0.0004
No	465 (96.5%)	252 (98.8%)	105 (97.2%)	33 (86.8%)	75 (92.6%)	
Infiltrating Lobular, n (%)						
Yes	69 (14.3%)	47 (18.4%)	19 (17.6%)	1 (2.6%)	2 (2.5%)	0.0004
No	413 (85.7%)	208 (81.6%)	89 (82.4%)	37 (97.4%)	79 (97.5%)	
Menopausal status, n (%)						
Premenopausal	158 (40.2%)	73 (34.8%)	36 (40.9%)	18 (54.5%)	31 (50.0%)	0.0480
Postmenopausal	235 (59.8%)	137 (65.2%)	52 (59.1%)	15 (45.5%)	31 (50.0%)	
Ethnicity, n (%)						
White	350 (79.4%)	197 (84.9%)	73 (73.7%)	29 (85.3%)	51 (67.1%)	0.0029
African American	91 (20.6%)	35 (15.1%)	26 (26.3%)	5 (14.7%)	25 (32.9%)	
Family history of breast cancer, n (%)						
Yes	121 (25.3%)	64 (25.3%)	36 (33.3%)	5 (13.2%)	16 (20.0%)	0.0492
No	358 (74.7%)	189 (74.7%)	72 (66.7%)	33 (86.8%)	64 (80.0%)	
Prior HRT use, n (%)						
Ever	188 (39.3%)	118 (47%)	33 (30.6%)	11 (28.9%)	26 (32.1%)	0.00425
Never	290 (60.7%)	133 (53%)	75 (69.4%)	27 (71.1%)	55 (67.9%)	
History of breast feeding, n (%)						
Ever	228 (50.4%)	130 (53.1%)	45 (45.5%)	23 (67.6%)	30 (40.5%)	0.0353
Never	224 (49.6%)	115 (46.9%)	54 (54.5%)	11 (32.4%)	44 (59.5%)	

BMI, Body Mass Index; HRT, Hormone Replacement Therapy; p-values for only age and BMI were obtained from ANOVA; p-values for rest of the characteristics were obtained for chi-square.

**Table 3: Co-occurrence of BBDs and tumor pathologies in breast cancer subtypes**

	All cases n=482	Luminal A n=255 (52.9%)	Luminal B n=108 (22.4%)	Her2 n=38 (7.9%)	Triple negative n=81 (16.8%)	p
<b>Benign breast diseases</b>						
Atypical ductal hyperplasia, n (%)						
Yes	91 (18.9%)	65 (25.5%)	19 (17.6%)	1 (2.6%)	6 (7.4%)	0.0001
No	391 (81.1%)	190 (74.5%)	89 (82.4%)	37 (97.4%)	75 (92.6%)	
Fibrocystic changes, n (%)						
Yes	310 (64.3%)	181 (71.0%)	62 (57.4%)	20 (52.6%)	47 (58.0%)	0.0126
No	172 (35.7%)	74 (29.0%)	46 (42.6%)	18 (47.4%)	34 (42.0%)	
Intraductal hyperplasia (mild), n (%)						
Yes	33 (6.8%)	25 (9.8%)	2 (1.9%)	0 (0.0%)	6 (7.4%)	0.0144
No	449 (93.2%)	230 (90.2%)	106 (98.1%)	38 (100.0%)	75 (92.6%)	
<b>In situ carcinomas</b>						
DCIS, n (%)						
Yes	321 (66.6%)	161 (63.1%)	80 (74.1%)	31 (81.6%)	49 (60.5%)	0.0258
No	161 (33.4%)	94 (36.9%)	28 (25.9%)	7 (18.4%)	32 (39.5%)	
LCIS, n (%)						
Yes	99 (20.5%)	67 (26.3%)	24 (22.2%)	3 (7.9%)	5 (6.2%)	0.0002
No	383 (79.5%)	188 (73.7%)	84 (77.8%)	35 (92.1%)	76 (93.8%)	

DCIS, Ductal Carcinoma In Situ; LCIS, Lobular Carcinoma In Situ; p-values are from chi-square test.

**Table 4: Association of clinicopathologic characteristics with ethnically distinct luminal and non-luminal breast cancer subtypes**

	All cases n=441	Luminal AA n=61 (13.80%)	Luminal CA n=270 (61.20%)	Non-luminal AA n=30 (6.80%)	Non-luminal CA n=80 (18.10%)	p
Age						
Mean±SD(years)	58.36±12.6	58±14	60±12	54±12.3	55±11.3	0.000902
BMI						
Mean±SD	27.52±6	28±5	27±6	29±5.6	28±5.1	0.364
Tumor grade, n (%)						
Grade 1	143 (33.0%)	17 (27.9%)	121 (45.1%)	1 (3.4%)	4 (5.3%)	1.05E-33
Grade 2	139 (32.1%)	27 (44.3%)	103 (38.4%)	2 (6.9%)	7 (9.3%)	
Grade 3	151 (34.9%)	17 (27.9%)	44 (16.4%)	26 (89.7%)	64 (85.3%)	
T-stage, n (%)						
T1	296 (67.7%)	42 (68.9%)	198 (73.3%)	15 (51.7%)	41 (53.2%)	0.000894
T2	117 (26.8%)	17 (27.9%)	61 (22.6%)	13 (44.8%)	26 (33.8%)	
T3	17 (3.9%)	1 (1.6%)	10 (3.7%)	1 (3.4%)	5 (6.5%)	
T4	7 (1.6%)	1 (1.6%)	1 (0.4%)	0 (0.0%)	5 (6.5%)	
AJCC stage, n (%)						
I	233 (53.1%)	35 (57.4%)	161 (59.6%)	10 (33.3%)	27 (34.6%)	4.71E-05
II	146 (33.3%)	17 (27.9%)	84 (31.1%)	15 (50.0%)	30 (38.5%)	
III	46 (10.5%)	5 (8.2%)	19 (7.0%)	3 (10.0%)	19 (24.4%)	
IV	14 (3.2%)	4 (6.6%)	6 (2.2%)	2 (6.7%)	2 (2.6%)	
Lymph node status, n (%)						
Positive	137 (32.5%)	18 (30.5%)	73 (28.1%)	10 (34.5%)	36 (48.6%)	0.0103
Negative	285 (67.5%)	41 (69.5%)	187 (71.9%)	19 (65.5%)	38 (51.4%)	
Tumor characteristics						
Infiltrating ductal carcinoma, n (%)						
Yes	318 (72.1%)	44 (72.1%)	183 (67.8%)	27 (90.0%)	64 (80.0%)	0.0206
No	123 (27.9%)	17 (27.9%)	87 (32.2%)	3 (10.0%)	16 (20.0%)	
Microinvasive ductal carcinoma, n (%)						
Yes	15 (3.4%)	3 (4.9%)	2 (0.7%)	3 (10.0%)	7 (8.8%)	0.00065
No	426 (96.6%)	58 (95.1%)	268 (99.3%)	27 (90.0%)	73 (91.3%)	
Infiltrating lobular, n (%)						
Yes	62 (14.1%)	11 (18.0%)	48 (17.8%)	0 (0.0%)	3 (3.8%)	0.00123
No	379 (85.9%)	50 (82.0%)	222 (82.2%)	30 (100.0%)	77 (96.3%)	
Infiltrating carcinoma NOS, n (%)						
Yes	28 (6.3%)	7 (11.5%)	20 (7.4%)	0 (0.0%)	1 (1.3%)	0.033
No	413 (93.7%)	54 (88.5%)	250 (92.6%)	30 (100.0%)	79 (98.8%)	
Menopausal status, n (%)						
Premenopausal	143 (39.8%)	26 (53.1%)	70 (31.4%)	10 (52.6%)	37 (54.4%)	0.0005474
Postmenopausal	216 (60.2%)	23 (46.9%)	153 (68.6%)	9 (47.4%)	31 (45.6%)	
Family history of breast cancer, n (%)						
Yes	110 (25.1%)	20 (32.8%)	72 (26.9%)	9 (31.0%)	9 (11.3%)	0.011388
No	328 (74.9%)	41 (67.2%)	196 (73.1%)	20 (69.0%)	71 (88.8%)	
Prior HRT use, n (%)						
Yes	175 (40.0%)	19 (32.2%)	122 (45.5%)	11 (36.7%)	23 (28.8%)	0.026103
No	262 (60.0%)	40 (67.8%)	146 (54.5%)	19 (63.3%)	57 (71.3%)	
History of breastfeeding, n (%)						
Yes	209 (50.6%)	20 (35.1%)	141 (54.9%)	11 (40.7%)	37 (51.4%)	0.038011
No	204 (49.4%)	37 (64.9%)	116 (45.1%)	16 (59.3%)	35 (48.6%)	

NOS – Not otherwise specified; HRT, Hormone Replacement Therapy; p-value for Age and BMI were obtained using ANOVA; other p-values are from chi-square test.

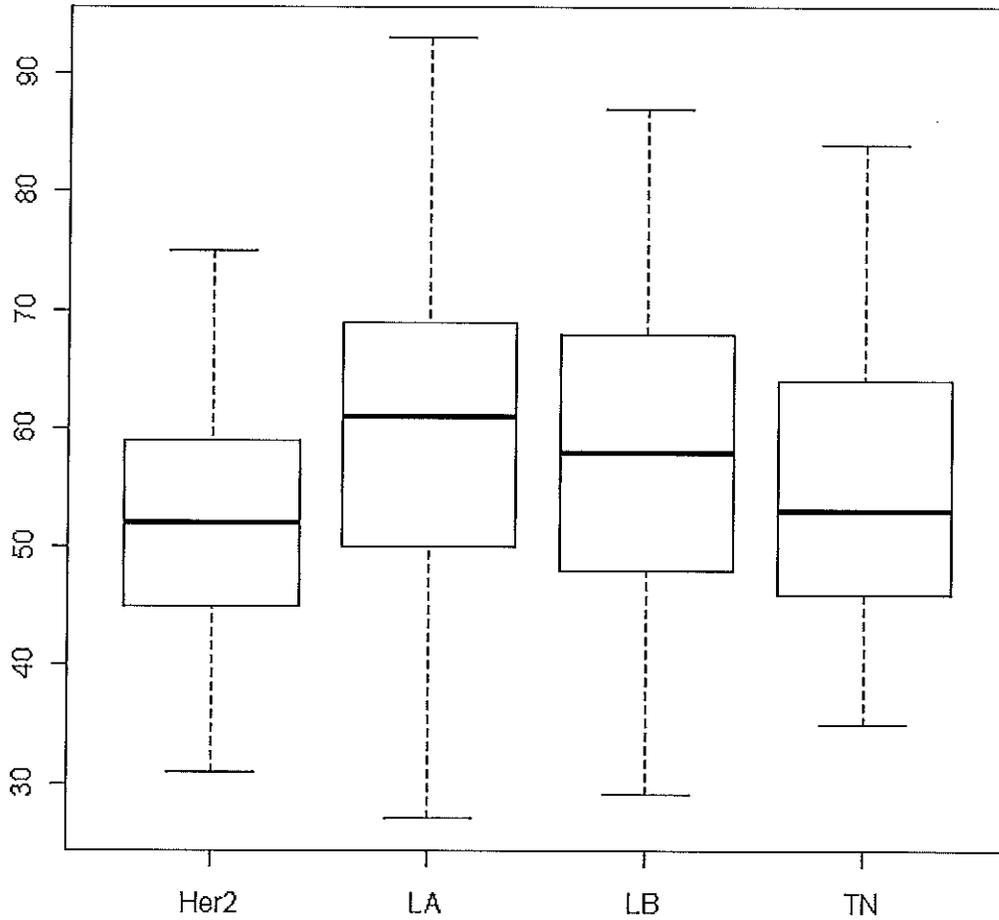
**Table 5: Co-occurrence of BBD and tumor pathologies in the ethnically distinct luminal and non-luminal subtypes**

	All Cases n = 441	Luminal AA n=61 (13.80%)	Luminal CA n=270 (61.20%)	Non-luminal AA n=30 (6.80%)	Non-luminal CA n=80 (18.10%)	p
<b>Benign breast diseases</b>						
Atypical ductal hyperplasia, n (%)						
Yes	82 (18.6%)	13 (21.3%)	62 (23.0%)	1 (3.3%)	6 (7.5%)	0.00197
No	359 (81.4%)	48 (78.7%)	208 (77.0%)	29 (96.7%)	74 (92.5%)	
Apocrine metaplasia, n (%)						
Yes	186 (42.2%)	32 (52.5%)	120 (44.4%)	10 (33.3%)	24 (30.0%)	0.0287
No	255 (57.8%)	29 (47.5%)	150 (55.6%)	20 (66.7%)	56 (70.0%)	
Columnar cell change, n (%)						
Yes	133 (30.2%)	21 (34.4%)	80 (29.6%)	15 (50.0%)	17 (21.3%)	0.0269
No	308 (69.8%)	40 (65.6%)	190 (70.4%)	15 (50.0%)	63 (78.8%)	
<b>In Situ carcinomas</b>						
DCIS, n (%)						
Yes	292 (66.2%)	41 (67.2%)	177 (65.6%)	19 (63.3%)	55 (68.8%)	0.94
No	149 (33.8%)	20 (32.8%)	93 (34.4%)	11 (36.7%)	25 (31.3%)	
LCIS, n (%)						
Yes	91 (20.6%)	16 (26.2%)	68 (25.2%)	3 (10.0%)	4 (5.0%)	7.84E-06
No	350 (79.4%)	45 (73.8%)	202 (74.8%)	27 (90.0%)	76 (95.0%)	

DCIS, Ductal Carcinoma In Situ; LCIS, Lobular Carcinoma In Situ; p-values are from chi-square test.

**Supplement Table 1**

	<b>AA</b> <b>n=91 (20.6%)</b>	<b>CA</b> <b>n=350 (79.4%)</b>	<b>p</b>
ER status, n (%)			
Negative	30 (33.0%)	80 (22.9%)	0.0643557
Positive	61 (67.0%)	270 (77.1%)	
PR status, n (%)			
Negative	51 (56.0%)	132 (37.7%)	0.0023494
Positive	40 (44.0%)	218 (62.3%)	
HER2 status, n (%)			
Negative	79 (86.8%)	294 (84.0%)	0.6177142
Positive	12 (13.2%)	56 (16.0%)	



**Figure1:** Subject age distribution is shown above with their invasive breast cancer subtypes. Luminals have a higher mean age in comparison with HER2+ and triple negative groups. Subsequent pairwise t-test indicated that luminal A is significantly older than triple negative ( $p=0.025$ ) and HER2+ ( $p=0.0058$ ), and there is a trend that luminal B is older than HER2 ( $p=0.068$ ).

Evaluation of the association of the ABO blood group SNP rs505922 with breast cancer phenotypes

Seth Rummel<sup>1</sup>, Craig D. Shriver<sup>2</sup>, Rachel E. Ellsworth<sup>3</sup>

From the Clinical Breast Care Project of the <sup>1</sup>Windber Research Institute, 620 Seventh St, Windber, PA 15963; <sup>2</sup>Walter Reed Army Medical Center, 6900 Georgia Avenue NW, Washington DC, 20307; <sup>3</sup>Henry M. Jackson Foundation for the Advancement of Military Medicine, 620 Seventh Street, Windber, PA 15963

Corresponding Author:

Rachel E. Ellsworth, PhD

Clinical Breast Care Project

620 Seventh Street

Windber, PA 15963

FAX: 814 467-6334

Phone: 814 467-9844

email: [r.ellsworth@wriwindber.org](mailto:r.ellsworth@wriwindber.org)

This research was supported by a grant from the United States Department of Defense (Military Molecular Medicine Initiative MDA W81XWH-05-2-0075, Protocol 01-20006).

## **ABSTRACT**

**Purpose:** To date, evaluation of the association of the ABO blood group and breast cancer has yielded mixed results. SNP rs505922, located within the first intron of the ABO gene, has been associated with the adenocarcinoma subtype of pancreatic cancer. To determine whether the ABO blood group is associated with risk of breast cancer, rs505922 was genotyped in 629 Caucasian women with invasive breast cancer, representing a variety of clinical and pathological tumor types.

**Methods:** Genomic DNA was isolated from blood clots. TaqMan SNP assay C\_2253769\_10 was used to determine the genotypes for rs505922. Statistical analysis was performed using chi-square analysis using a P-value <0.05 to define significance.

**Results:** Genotypes were generated for 100% of the 629 patients in this study. Allele and genotype frequencies did not vary significantly for age at diagnosis, tumor stage, size or grade, hormone, HER2 or lymph node status, intrinsic subtype, tumor type or patient outcome.

**Conclusions:** Allele frequencies for rs505922 did not differ between women with breast cancer and published HapMap frequencies. Further stratification into different tumor phenotypes also failed to reveal an association between rs505922 and any tumor characteristics. Together, these data suggest that the minor allele of rs505922 and the resulting non-O blood types are not associated with increased risk or less favorable tumor characteristics or prognosis in breast cancer.

*Keywords:* ABO blood-group system, breast cancer, phenotypes

## 1. INTRODUCTION

The first association between the ABO blood group and cancer risk was made in 1953 in patients with stomach cancer, where blood group A is associated with increased risk of stomach cancer and blood group O confers a protective advantage in individuals living in England <sup>(1)</sup>. This increased risk of gastric cancer in patients with blood group A has been confirmed as recently as 2010 <sup>(2)</sup>. In oral cancer, loss of the A and B antigens is associated with increased cell migration, in non-small cell lung cancer, the A blood group is associated with improved survival, and in ovarian cancer the B antigen is associated with increased risk <sup>(3-6)</sup>.

A number of studies have investigated the role of the ABO blood group in breast cancer risk and pathology. For example, two publications from the 1950s found no association between ABO blood group and risk of breast cancer in either English or American patient populations <sup>(5;7)</sup>. In more recent studies, while the A antigen has been associated with increased risk of developing invasive ductal carcinoma in 166 Greek women, in a group of 565 Turkish women, no association was seen between breast cancer risk and any ABO blood types <sup>(8;9)</sup>. These differences in risk assessment between not just breast but other tumor types and the ABO blood group may be attributed to differences in ABO frequencies between populations, the importance of using comparable case and control groups, and methods for determining ABO blood types.

In 2009, an agnostic approach was used to identify genes associated with risk of developing pancreatic cancer further supported the association between the ABO blood group and cancer. Genotyping of 558,542 single nucleotide polymorphisms (SNPs) in 1,896 patients with pancreatic cancer and 1,939 controls revealed a significant

association between SNP rs505922, located within the first intron of the ABO glycosyltransferase (ABO) gene <sup>(10)</sup>. The T allele of rs505922, which was found to be protective, is in linkage disequilibrium with the single base pair deletion that encodes the O antigen, suggesting that the A and B blood antigens may be associated with higher risk of pancreatic cancer. In a follow up study, the role of the ABO blood group was re-evaluated in 15,359 patients treated at the European Institute of Oncology with a variety of cancer types. An association between the ABO blood group was detected only for pancreatic cancer <sup>(11)</sup>. Importantly, the protective effect of the O allele was not associated with all types of pancreatic cancer but was limited to the adenocarcinoma subtype.

To determine whether that ABO blood group is associated with risk of breast cancer, SNP rs505922 was genotyped in 629 Caucasian women with invasive breast cancer diagnosed between 2001 and 2010. Data were then evaluated by a number of pathological characteristics including age at diagnosis, hormonal and HER2 status, intrinsic subtype, tumor grade, stage and size, and clinical outcome.

## **2. PATIENTS AND METHODS**

### *2.1. Study Population*

The Clinical Breast Care Project (CBCP) database was queried to identify all Caucasian women with invasive breast cancer diagnosed between 2001 and 2009. Blood samples were collected with approval from the Walter Reed Army Medical Center Human Use Committee and Institutional Review Board (IRB). All subjects enrolled in the CBCP voluntarily agreed to participate and gave written informed consent.

Diagnosis of every specimen was performed by a single, dedicated breast pathologist from hematoxylin and eosin (H&E) stained slides; staging was performed using guidelines defined by the *AJCC Cancer Staging Manual* sixth edition <sup>(12)</sup> and grade assigned using the Nottingham Histologic Score <sup>(13;14)</sup>. ER and PR status were determined by immunohistochemical (IHC) analysis at a clinical laboratory (MDR Global, Windber, PA) and HER2 status was assayed using the PathVysion® HER-2 DNA Probe kit (Abbott Laboratories, Abbott Park, IL) according to manufacturer's protocols.

## 2.2. SNP Genotyping

Genomic DNA was isolated from blood clots using the Gentra Clotspin and Puregene DNA purification kits (Qiagen Inc., Valencia, CA). Ten ng of DNA were amplified in each PCR reaction and each sample was run in duplicate. PCR was performed using the TaqMan SNP assay C\_2253769\_10 (Applied Biosystems, Foster City, CA), representing SNP rs505922 and run on a 7000 Sequence Detection System (Applied Biosystems, Foster City, CA). Genotypes (CC, CT or TT) were determined using the ABI PRISM® 7000 Sequence Detection System software (Applied Biosystems, Foster City, CA).

## 2.3 Statistical Analysis

Differences in genotype or allele frequencies between different clinicopathological characteristics were calculated using chi-square analysis using *rxc* Contingency Tables ([www.physics.csbsju.edu/stats/contingency\\_NROW\\_NCOLUMN\\_form.html](http://www.physics.csbsju.edu/stats/contingency_NROW_NCOLUMN_form.html)). Significance was determined using  $P < 0.05$ .

## 3. RESULTS

### *3.1 Clinicopathological characteristics of the patient population*

The average age at diagnosis for the 629 Caucasian women with invasive breast cancer was 58.94 years (range 28-97 years) and the majority (74%) was post-menopausal at diagnosis. Most of the tumors were invasive ductal carcinomas (IDCA) (70%), ER positive (78%), HER2 negative (82%) and of the luminal A (ER+/HER2-) subtype (74%). Tumors were divided between well-differentiated (33%), moderately-differentiated (38%) and poorly-differentiated (29%). Most of the tumors were early stage with 86% being stage I-II and 67% having negative lymph node status. Seventeen percent of the patients had died of disease.

### *3.2 Frequency of rs505922 in women with invasive breast cancer*

Of the 629 samples assayed in this study, six (1%) produced discordant genotypes between duplicate samples during the original run; these samples were run a second time resulting in concordant genotypes between duplicate samples, resulting in a genotyping accuracy of 100%, with each sample clustering into one of three genotypes. Allele frequencies were 0.65 for the T allele and 0.35 for the C allele. Chi-square analysis revealed that the genotype frequencies were in Hardy-Weinberg Equilibrium ( $P=0.961$ ). Neither genotype nor allele frequencies differed significantly from those reported for Caucasians in the CSHL-HAPMAP panel ( $P=0.809$  and  $P=0.654$ , respectively).

### *3.3 Evaluation of rs505922 genotype and allele frequencies with clinicopathological characteristics*

Allele and genotype frequencies were compared for age at diagnosis, tumor stage, size and grade, hormone, HER2 and lymph node status, intrinsic subtype, tumor type and patient status. No significant differences were seen by genotype (Table 1) or allele (data

not shown). To determine whether the minor C allele was associated with any variable, CC and CT genotypes were combined. As with genotype and allele frequencies, no significant differences were detected.

#### **4. DISCUSSION**

The first suggestion of an association between ABO blood group antigens and malignancy was made almost 100 years ago, yet the role of the ABO blood group in cancer risk and prognosis remains controversial. Results from the pancreatic GWAS study in 2009 renewed interest in the ABO blood group as the T allele of SNP rs505922, located within the first intron of the ABO gene locus was found to confer a protective advantage against pancreatic cancer. The T allele has been found to be in linkage disequilibrium with the single base pair deletion responsible for the O blood group antigen, leading the authors to suggest that people with non-O blood groups have increased risk of developing pancreatic cancer <sup>(10)</sup>.

In addition to the association with defining blood group antigens, variability of SNPs within the ABO region has been associated with circulating levels of tumor necrosis factor alpha (TNF $\alpha$ ), soluble intracellular adhesion molecule 1 (sICAM-1) and alkaline phosphatase (15-17). Because these tumor-promoting factors may also provide an environment favoring breast tumor development, variation of rs505922 may contribute to increased risk of breast cancer development. Comparison of allele frequencies in our cohort of 629 Caucasian women with invasive breast cancer, however, did not differ significantly from Caucasians in the HapMap project, suggesting that the minor allele of rs505922 is not associated with increased risk of breast malignancy.

In the follow-up paper by Iodice et al., which found no association between ABO blood group and breast cancer risk ( $P=0.60$ ) in 7,208 women treated at the European Institute of Oncology (IEO), the importance of tumor heterogeneity was highlighted <sup>(11)</sup>. While the GWAS study found an association between rs505922 and pancreatic cancer overall, evaluation of ABO blood group antigens with subtypes of pancreatic cancer in the IEO revealed that the non-O blood type was associated with the exocrine, but not the endocrine, form of pancreatic cancer. Breast cancer is marked by tremendous heterogeneity in terms of age at diagnosis, tumor pathologies, and molecular characteristics. Thus, the ABO blood group may be associated with increased risk or prognosis of certain subtypes of breast cancer. Neither allele nor genotype frequencies differed significantly by age at diagnosis, tumor stage, grade and size, hormone, HER2 and lymph node status, intrinsic subtype or prognosis, suggesting that unlike pancreatic cancer, the ABO blood group is not associated with either overall risk of or certain subtypes of breast cancer.

Our results are in agreement with those from Iodice et al. derived largely from an Italian patient population as well as those from Dede et al. who found no association in 565 Turkish women <sup>(9;11)</sup>. In contrast, Stamatakis et al. found an association between the A antigen and IDCA and/or poor prognosis in population of Greek women <sup>(8)</sup>. Differences between the Greek study and the other studies that failed to find an association may reflect the small sample size used in the Greek study, which was based on 166 women with breast cancer, which is much smaller than the 565 Turkish, 629 Caucasian American, or 7,208 Italian women included in the other studies.

Although this study does not support a role for the ABO gene in breast cancer incidence or prognosis, this study was limited to individuals with self-reported “white” ancestry. The frequency of the ABO blood group varies globally, with a predominance of the O allele in the Kikuyu of Kenya, Malaysians and Navajos, and of the A allele in Hawaiians, Portuguese and Swiss (<http://www.bloodbook.com/world-abo.html>). Thus, differences in allele frequency of the ABO blood group must be considered when evaluating the role of ABO in disease. For example, the A antigen, which was found to be associated with increased breast cancer risk in a Greek population but not in a Turkish population, is more frequent in Greeks (0.42) than in Turks (0.34), which may influence the association between the A blood group and disease status, although the O group frequency is similar between the two populations (0.40 and 0.43 in Greeks and Turks, respectively).

The variability of the minor allele frequency of SNP rs505922 also varies between populations: in a pilot study from the 1,000 Genomes Project (18), the frequency of rs505922 was 0.32 in YRI (Sub-Saharan Africa), 0.352 in CHB+JPT (China and Japan) and 0.417 in CEU (Caucasians from Utah). Although patients from ten European countries and Shanghai were included in the pancreatic GWAS study, the majority of patients were from a number of cohort studies from the United States. Similar results were achieved when all patients were included in the analysis and when only those of European ancestry were included<sup>(10)</sup>. In our study, all samples were collected from self-described white women living in Washington DC or rural Pennsylvania. Given the marked admixture in the United States, these 629 women likely represent a heterogeneous mixture of European ancestries. Thus, an association that may be detected

in the more heterogeneous Greek population may be masked by the genetic contributions from other European cultures represented in our American patient population.

In conclusion, the minor allele frequency of SNP rs505922, which is in linkage disequilibrium with the non-O ABO blood types, did not differ between Caucasian American women with breast cancer and published HapMap frequencies. Further stratification into different tumor phenotypes also failed to reveal an association between rs505922 and any tumor characteristics. Together, these data suggest that unlike pancreatic cancer, where the minor rs505922 allele was associated with increased risk, especially of exocrine tumors, rs505922 and the resulting non-O blood types, are not associated with increased risk or less favorable tumor characteristics or prognosis in breast cancer.

#### **CONFLICT OF INTEREST STATEMENT**

None declared.

## **ACKNOWLEDGEMENTS**

We thank Brenda Deyarmin for her assistance collecting clinical information for the patients in this study. The opinion and assertions contained herein are the private views of the authors and are not to be construed as official or as representing the views of the Department of the Army or the Department of Defense. This research was supported by a grant from the United States Department of Defense (Military Molecular Medicine Initiative MDA W81XWH-05-2-0075, Protocol 01-20006). The Department of Defense was not involved in the conceptualization or execution of this project.

## REFERENCES

1. Aird I, Bentall HH, Roberts JA. A relationship between cancer of stomach and the ABO blood groups. *Br Med J* 1953;1:799-801.
2. Edgren G, Hjalgrim H, Rostgaard K, Norda R, Wikman A, Melbye M et al. Risk of gastric cancer and peptic ulcers in relation to ABO blood type: a cohort study. *Am J Epidemiol* 2010;172:1280-1285.
3. Gates MA, Wolpin BM, Cramer DW, Hankinson SE, Tworoger SS. ABO blood group and incidence of epithelial ovarian cancer. *Int J Cancer* 2011;128:482-86.
4. Lee JS, Ro JY, Sahin AA, Hong WK, Brown BW, Mountain CF et al. Expression of blood-group antigen A--a favorable prognostic factor in non-small-cell lung cancer. *N Engl J Med* 1991;324:1084-90.
5. Aird I, Bentall HH, Mehigan JA, Roberts JA. The blood groups in relation to peptic ulceration and carcinoma of colon, rectum, breast, and bronchus; an association between the ABO groups and peptic ulceration. *Br Med J* 1954;2:315-21.
6. Dabelsteen E, Gao S. ABO blood-group antigens in oral cancer. *J Dent Res* 2005;84:21-28.
7. Goldenberg IS, Hayes MA. Breast carcinoma and ABO blood groups. *Cancer* 1958;11:973-74.
8. Stamatakis M, Kontzoglou K, Safioleas P, Safioleas C, Manti C, Safioleas M. Breast cancer incidence in Greek women in relation to ABO blood groups and Rh factor. *Int Semin Surg Oncol* 2009;6:14.
9. Dede DS, Aksoy S, Dizdar O, Cerci P, Gullu I, Ozisik Y et al. Blood ABO groups and risk of breast cancer. *Med Oncol* 2010;27:1433.
10. Amundadottir L, Kraft P, Stolzenberg-Solomon RZ, Fuchs CS, Petersen GM, Arslan AA et al. Genome-wide association study identifies variants in the ABO locus associated with susceptibility to pancreatic cancer. *Nat Genet* 2009;41:986-90.
11. Iodice S, Maisonneuve P, Botteri E, Sandri MT, Lowenfels AB. ABO blood group and cancer. *Eur J Cancer* 2010;46:3345-50.
12. American Joint Committee on Cancer. *AJCC Cancer Staging Manual 6th ed.* New York, NY: Springer; 2002.
13. Bloom HJ, Richardson WW. Histological grading and prognosis in breast cancer. *Br J Cancer* 1957;11:359-77.

14. Elston CW, Ellis IO. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow up. *Histopathology* 1991;19:403-10.
15. Yuan X, Waterworth D, Perry JR, Lim N, Song K, Chambers JC et al. Population-based genome-wide association studies reveal six loci influencing plasma levels of liver enzymes. *Am J Hum Genet* 2008;83:520-528.
16. Pare G, Chasman DI, Kellogg M, Zee RY, Rifai N, Badola S et al. Novel association of ABO histo-blood group antigen with soluble ICAM-1: results of a genome-wide association study of 6,578 women. *PLoS Genet* 2008;4:e1000118.
17. Melzer D, Perry JR, Hernandez D, Corsi AM, Stevens K, Rafferty I et al. A genome-wide association study identifies protein quantitative trait loci (pQTLs). *PLoS Genet* 2008;4:e1000072.
18. 1000 Genomes Project Consortium, Durbin RM, Abecasis GR, Altschuler DL, Auton A, Brooks LD et al. A map of human genome variation from population-scale sequencing. *Nature* 2010;467:1061-73.

Table 1. Genotype frequencies by clinicopathological variable. Data was not available for each variable for every patient thus "N" varies by characteristic.

N=629	CC	CT	TT	P-value
Age				P=0.393
<40 years	4 (12%)	11 (32%)	19 (56%)	
40-49 years	12 (9%)	59 (46%)	57 (45%)	
≥50 years	62 (13%)	211 (45%)	194 (42%)	
Tumor Stage				P=0.143
I	49 (15%)	147 (44%)	137 (41%)	
II	18 (9%)	98 (48%)	87 (43%)	
III	11 (15%)	28 (40%)	32 (45%)	
IV	0 (0%)	5 (33%)	10 (67%)	
Tumor Size				P=0.238
T1	56 (14%)	185 (45%)	171 (41%)	
T2	11 (8%)	66 (45%)	69 (47%)	
T3	7 (18%)	17 (45%)	14 (37%)	
Grade				P=0.852
1	25 (13%)	83 (42%)	88 (45%)	
2	30 (13%)	104 (46%)	95 (41%)	
3	18 (10%)	77 (44%)	79 (46%)	
Hormone Status <sup>a</sup>				P=0.583
ER+PR+	54 (14%)	165 (43%)	162 (43%)	
ER+PR-	10 (11%)	46 (49%)	37 (40%)	
ER-PR-	12 (10%)	57 (45%)	56 (45%)	
HER2 Status				P=0.788
Positive	9 (10%)	38 (44%)	40 (46%)	
Negative	60 (13%)	201 (43%)	203 (44%)	
Intrinsic Subtype <sup>b</sup>				P=0.748
Luminal A	54 (14%)	169 (44%)	165 (42%)	
Luminal B	4 (8%)	21 (45%)	22 (47%)	
HER2 enriched	5 (12%)	17 (43%)	18 (45%)	
Triple negative	6 (8%)	32 (42%)	38 (50%)	
Tumor Type				P=0.196
IDCA	41 (10%)	191 (46%)	185 (44%)	
ILCA	18 (19%)	41 (44%)	35 (37%)	
Mixed ILCA/IDCA	4 (13%)	12 (40%)	14 (47%)	
Other <sup>c</sup>	8 (18%)	17 (39%)	19 (43%)	
Lymph Node Status <sup>d</sup>				P=0.546
Negative	51 (12%)	181 (44%)	183 (44%)	
Positive	16 (11%)	74 (48%)	62 (41%)	
Patient Status				P=0.857
Disease-free <sup>e</sup>	23 (13%)	85 (49%)	66 (38%)	
Recurrence	5 (13%)	19 (49%)	15 (38%)	
Dead of Disease	4 (13%)	12 (39%)	15 (48%)	

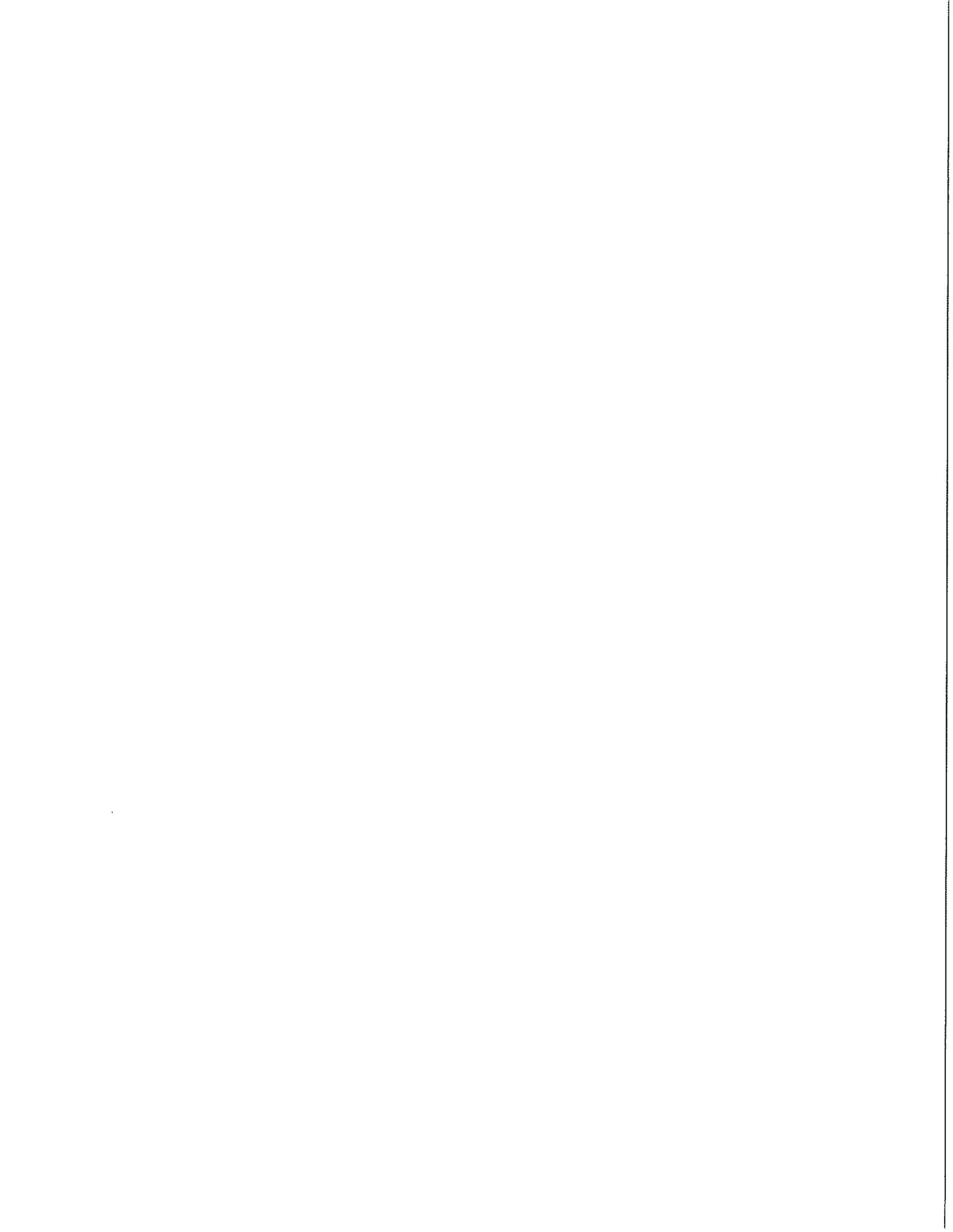
<sup>a</sup>Only a single patient had confirmed ER-PR+ status, thus this rare tumor type was not included in the analysis

<sup>b</sup>Intrinsic subtypes were defined as luminal A = ER and or PR+/HER2-; luminal B = ER and or PR+/HER2+; HER2-enriched = ER and PR-/HER2+ and triple negative = ER, PR and HER2-

<sup>c</sup>Other included histological types including tubular, medullary, apocrine, and mucinous carcinomas

<sup>d</sup>Patients with isolated tumor cells (n=33) were not included in this analysis

<sup>e</sup>Only patients who have been disease-free  $\geq 5$  years were included in this analysis



# **An effective strategy for comprehensive proteome characterization of OCT-embedded breast tumor specimens**

Tao Liu,<sup>1,#</sup> V.S. Kumar Kolli,<sup>2,#</sup> Brianne O. Petritis,<sup>1</sup> Athena A. Schepmoes,<sup>1</sup> Ronald J. Moore,<sup>1</sup> Matthew E. Monroe,<sup>1</sup> Luke Weaver,<sup>2</sup> Brenda Deyarmin,<sup>2</sup> Jennifer Kane,<sup>2</sup> Richard Katenhusen,<sup>2</sup> David Kirchner,<sup>2</sup> Richard J. Mural,<sup>2</sup> Craig Shriver,<sup>3</sup> Richard D. Smith,<sup>1</sup> Karin D. Rodland,<sup>1</sup> and David G. Camp, II<sup>1\*</sup>

<sup>1</sup>Biological Sciences Division, Pacific Northwest National Laboratory, Richland, Washington 99352

<sup>2</sup>Windber Research Institute, Windber, Pennsylvania 15963

<sup>3</sup>General Surgery Services, Walter Reed Army Medical Center, Washington DC 20307

**Running title:** Proteomic analysis of OCT-embedded breast tumor tissue

**Keywords:** quantitative proteomics • breast cancer • tissue • OCT • LMD • LCM • LC-MS • AMT

<sup>#</sup>These authors contributed equally to this work.

<sup>\*</sup>To whom correspondence should be addressed.

David G. Camp, II  
Pacific Northwest National Laboratory  
P.O. Box 999, MSIN: K8-98  
Richland, WA 99352  
Phone: (509) 371-6586  
Fax: (509) 371-6564  
Email: dave.camp@pnl.gov

## Abstract

Optimal cutting temperature (OCT) compound-embedded tissue specimens are reliable resources for morphological, immunohistochemical and molecular examinations. However, severe OCT polymer contamination presents a formidable challenge when analyzing these specimens directly using MS-based proteomics. We have developed a strategy for effective protein extraction and digestion of the OCT-embedded breast tumor specimens and subsequent comprehensive, reproducible, and high-throughput quantitative proteome characterization. Three different “one-tube” protocols were first evaluated for their protein extraction and digestion efficiency, and reproducibility. SCX liquid chromatography was used to remove the polymers while fractionating the peptides from the tryptic digest of prepared OCT sections, followed by LC-MS/MS analysis of each fraction, to create a breast tissue accurate mass and time (AMT) tag database. Breast tissue samples procured by laser microdissection (LMD) from the OCT sections, free from OCT polymer contamination, were then directly analyzed using LC-MS and the AMT tag approach. This effective quantitative proteomics strategy was able to provide >6,000 peptide identifications covering >2,000 proteins, with a protein overlap of >93% and a correlation coefficient of >0.98 for protein abundances between process replicates, through a single LC-MS analysis of each breast tumor LMD sample. It can be widely applied in large-scale biomedical research employing OCT-embedded tissue specimens.

## Introduction

Although cancer is generally considered to be a genetic disease, initiated in the accumulation of multiple mutations, chromosomal rearrangements, and epigenetic changes over time,<sup>1,2</sup> the actual conversion of these changes at the somatic genetic level into the malignant phenotype is a gradual process which must of necessity be executed at the protein level. Alterations in the network of protein-protein interactions induced by changes in abundance or post-translational modification are essential to drive the survival, growth invasion, and metastasis of cancer cells. Recent advancements in high throughput molecular profiling technologies are providing snapshots of the molecular alterations and changes in expression levels exhibited by cancer biopsy tissue specimens, thus contributing to improved understanding of the complex interplay between genetic, transcription, and translational alterations in human cancers.

Molecular profiling technologies are only as informative as the samples analyzed. The ability of laser-based tissue microdissection techniques, e.g., laser capture microdissection (LCM) and laser microbeam microdissection (laser microdissection or LMD for short),<sup>3</sup> to provide controlled samples with a defined cellular composition has increased the utility of molecular profiling studies. Tumor tissues embody a complex mixture of morphologically and phenotypically distinct cell types and this heterogeneous layout of three dimensional tissue structure constrains the interpretation of molecular data from bulk tissue specimens. The isolation of a homogenous population of cells is a prerequisite for understanding the molecular basis of cancer and the inability to isolate pure cell populations has, in the past, confounded the localization of alterations to a specific cell type or to a sub-compartment of the tissue microenvironment. The advent of laser microdissection technique and its ability to directly visualize the cells under study, has allowed for the reduction of tissue heterogeneity and, in some cases, the isolation of specific cell types of interest from heterogeneous tissue.

Interfacing laser microdissection with molecular profiling is revolutionizing molecular pathology and providing new insights into the interactions between malignant epithelial, stromal, and normal tissue sub- compartments. This combination of information is allowing great advances in the understanding of the compartmentalization of the intracellular signaling events of tumor cells in response to the surrounding stroma and the role of this communication in the process of tumor growth and invasion. Archived clinical tissue samples with accurate and complete clinical annotation are improving our understanding of tumor progression and the biology associated with tumor heterogeneity.

Several methods are commonly used to preserve and transport clinical tissue samples, predominantly either flash-freezing or tissue fixation (e.g., formalin, ethanol) and tissue embedding [e.g., paraffin, optimal cutting temperature (OCT) compound] for optimal molecular analysis in disease diagnostics/pathology. While fresh or frozen samples are ideal for comprehensive analysis by molecular profiling technologies such as genomics and MS-based proteomics, the ability to conduct molecular profiling analyses on the more widely available formalin-fixed, paraffin-embedded (FFPE) or OCT-embedded tissue specimens will extend these analyses to the large repositories of archived samples currently available. This is especially promising for retrospective molecular analyses as these samples are typically linked to clinical and clinical outcome data.

Many methods have been developed to extract DNA,<sup>4</sup> RNA<sup>5</sup> or protein<sup>6-8</sup> from FFPE specimens. FFPE specimens are convenient substrates for the application of laser microdissection to procure the small cell populations from the histological sections, and it has been demonstrated that the FFPE specimens can be analyzed by quantitative proteomics at the level of proteins and post-translational modifications (e.g., phosphorylation, N-glycosylation).<sup>9</sup> While laser microdissection samples derived from either fresh-frozen<sup>10,11</sup> or FFPE<sup>9,12,13</sup> tissue specimens can be used for either downstream proteome or genome analysis, there is evidence showing that different degrees of degradation of the

DNA and RNA could occur in FFPE specimens,<sup>14-16</sup> confounding reliable and reproducible gene expression profiling. Without requiring formalin tissue fixation, which could cause modifications of the proteins and nucleic acids in the tissue specimens, OCT-embedding provides a useful alternative to the FFPE method.<sup>17,18</sup> Proteome characterization of OCT-frozen breast tumor samples with laser microdissection has been reported;<sup>19,20</sup> however, the proteome coverage achieved thus far has been typically limited to a few hundred of protein identifications, presumably due to the lack of adequate protocols to extract the protein content and the limited dynamic range and sensitivity of proteomics analytical platforms.

In this study, we compared three different “one-tube” protocols that utilize cleavable surfactants or organic solvents for effective and reproducible protein extraction and digestion of both OCT sections and LMD samples procured from the same normal breast tissue block. We then explored the possibility of using the readily available, but heavily polymer-contaminated OCT section samples for in-depth 2D-LC-MS/MS analysis to establish a tissue protein database for subsequent direct high-throughput LC-MS analysis using the accurate mass and time (AMT) tag approach.<sup>21</sup> Finally LMD samples procured from breast tumor samples, containing 50,000 cells and free from OCT polymer contamination, were analyzed by LC-MS to evaluate the proteome coverage and reproducibility. The results suggest that the new proteomics strategy is highly effective and reproducible, and suitable for large-scale quantitative comparison of the OCT-embedded breast tissue samples for biomedical applications such as biomarker discovery for diagnosis, prognosis, or patient stratification.

## **Materials and Methods**

**Preparation of OCT section and LMD samples.** The normal and cancer tissues that were utilized in this study were selected from the tissue bank at Windber Research Institute (WRI) and were collected from the consented subjects enrolled in HIPAA-Compliant IRB approved clinical protocols at the Clinical Breast Care Project (CBCP) led by the Walter Reed Army Medical Center (WRAMC). All the tissues were

de-identified before they were banked at WRI. The normal tissue was derived from mammoplasty procedures done on women with no history of breast cancer. Cancer tissue was from the surgical resection of invasive breast tumors and both types of tissues were embedded in OCT immediately after surgical excision and frozen. The samples are stored at -180 °C until use.

The OCT-embedded tissue was processed in two ways to generate the protein samples for downstream proteome characterization. To extract the proteins from gross tissue sections, the OCT tissue blocks were mounted onto the chuck of a cryostat CM3050 S (Leica Microsystems, Wetzlar, Germany) and 12  $\mu\text{m}$  sections were cut and the tissue sections were directly used for downstream MS proteomics. To harvest the homogenous population of normal epithelium or epithelial cancerous cells from the OCT tissue blocks, the tissue sections were cut at 8  $\mu\text{m}$  thickness using a cryostat CM3050 S and the sections were transferred to glass slides coated with polyethylene naphthalene membranes (W. Nuhsbaum, Inc., McHenry, IL). The staining protocol consisted of pre-staining steps to fix the sections with 90%, 70%, and 50% ethanol solutions for 30 s each followed by distilled water wash for 1 min. After fixing the sections the Mayer's Hematoxylin (Sigma-Aldrich, St. Louis, MO) was added to the slides and stained for 1 min and then slides were washed with distilled water. The Hematoxylin stained slides were dehydrated in 95% ethanol and stained with Eocin (Richard Allan Scientific, Kalamazoo, MI) for 20 s. The H&E stained slides were visualized on an AS LMD microscope (Leica Microsystems) and the regions of interest from the tissue sections of normal and cancer tissue samples were selected and micro-dissected with the help of a UV laser on the AS LMD. The tissue portions were directly dropped into the Eppendorf tubes for single tube protein recovery methods and three to six tissue sections (8  $\mu\text{m}$ ) were utilized to procure  $\sim 50,000$  cells by LMD depending on the cellularity of the tissue samples.

**Tissue sample pre-treatment.** For proteomics sample preparation, the breast tissue samples (either the OCT sections or the cells procured from the OCT sections using LMD) were first thawed on ice

for 10 min. The OCT section samples were washed briefly with 200  $\mu\text{L}$  of 60% ethanol for three times to remove the OCT polymers on the surface, followed by drying at room temperature for 10 min. The LMD samples were used directly without washing.

**TFE digestion.** Thirty microliters of 50 mM  $\text{NH}_4\text{HCO}_3$ , pH 8 buffer and 30  $\mu\text{L}$  of neat TFE (Sigma-Aldrich) were added to the pre-washed OCT section or LMD sample in a microcentrifuge tube. The tube was wrapped with parafilm and sonicated for 30 s using a Branson Sonifier 250 (Branson Ultrasonics, Danbury, CT) and then set on ice for 30 s; this process was repeated for 3 min. The sample was then incubated at 60  $^\circ\text{C}$  for 2 h with gentle shaking at 300 rpm. Dithiothreitol (DTT) was added for a final concentration of 2 mM and the sample was incubated at 37  $^\circ\text{C}$  for 1 h with gentle shaking. After diluting the sample 5-fold with 50 mM  $\text{NH}_4\text{HCO}_3$ , pH 8 buffer, sequencing grade modified porcine trypsin (Promega, Madison, WI) was added to the sample with 1:50 (w/w) trypsin-to-protein ratio (protein amount in the sample was estimated based on protein assay results of a test TFE tissue lysate with a known amount of cells) and incubated for 3 h at 37  $^\circ\text{C}$  with gentle mixing. The resulting tryptic digest was ultracentrifuged at an  $\text{RCF}_{\text{avg}}$  of 227,000  $\times g$  for 20 min at 4  $^\circ\text{C}$  and the supernatant was transferred to another microcentrifuge tube and concentrated in a Speed-Vac SC 250 Express (Thermo Savant, Holbrook, NY). The sample was resuspended in 25 mM  $\text{NH}_4\text{HCO}_3$  and adjusted to a final concentration of 0.4  $\mu\text{g}/\mu\text{L}$  based on the results of BCA protein assay (Pierce Rockford, IL). Samples were snap frozen and stored at -80  $^\circ\text{C}$  until time for either LC-MS or LC-MS/MS analysis.

**RapiGest digestion.** Sixty microliters of 0.1% RapiGest™ SF Surfactant (Waters, Milford, MA) solution in 50 mM  $\text{NH}_4\text{HCO}_3$  and 2 mM DTT were added to the pre-washed OCT section or LMD sample and the mixture was sonicated as described above. The sample was incubated at 50  $^\circ\text{C}$  for 30 min with gentle shaking, followed by trypsin digestion as described above. After trypsin incubation, 10%

trifluoroacetic acid (TFA) was added to the digested sample for a final TFA concentration of 0.5%. The sample was mixed briefly using a vortex mixer and incubated at 37 °C for 60 min. The resulting RapiGest digest was then ultracentrifuged, concentrated, concentration adjusted, and stored as described in the TFE digestion protocol.

**PPS digestion.** Sixty microliters of 0.1% PPS Silent™ Surfactant (Protein Discovery, Knoxville, TN) solution in 50 mM NH<sub>4</sub>HCO<sub>3</sub> and 2 mM DTT were added to the pre-washed OCT section or LMD sample and the mixture was sonicated as described above. The sample was incubated at 50 °C for 30 min with gentle shaking, followed by trypsin digestion as described above. After trypsin incubation, formic acid was added to the digested sample for a final concentration of 50% (v/v). The sample was mixed briefly using a vortex mixer and incubated at 30 °C for 2 h. The resulting PPS digest was then ultracentrifuged, concentrated, concentration adjusted, and stored as described in the TFE digestion protocol.

**Strong cation exchange (SCX) liquid chromatography separation.** To test the possibility of removing the OCT polymers while fractionating the peptides, two 30-µm thick OCT sections from an OCT-embedded normal breast tissue block were digested using the TFE protocol as described above. SCX separation was performed using 100 µg of the TFE digest on an Agilent 1100 HPLC System (Agilent, Palo Alto, CA). The separation was performed on a PolySulfoethyl A, 200 × 2.1 mm, 5 µm, 300-Å column with a 10 × 2.1 mm guard column (PolyLC, Columbia, MD) at a flow rate of 0.2 mL/min. Mobile phases consisted of 10 mM ammonium formate, pH 3.0, 25% acetonitrile (A) and 500 mM ammonium formate, pH 6.8, 25% acetonitrile (B). Peptides were separated using the following linear gradient: 100% A for 10 min, increasing to 50% B over 50 min and then to 100% B over the next 10 min, and finally, 100% B for the last 10 min. One fraction was collected every minute and the peptide fractions were subsequently pooled into a total of 30 fractions based on elution order. Fractions were dried in a Speed-Vac

concentrator and reconstituted in 30  $\mu\text{L}$  of 25 mM  $\text{NH}_4\text{HCO}_3$ . After the digestion protocol optimization experiments, two 30- $\mu\text{m}$  thick OCT sections from an OCT-embedded breast tumor tissue block were digested by the PPS protocol and fractionated in the same fashion using SCX. The SCX fractions were snap frozen and stored at  $-80^\circ\text{C}$  until time for LC-MS/MS analysis.

**LC-MS/MS and LC-MS analyses.** The HPLC system consisted of an in-house assembled custom configuration of 65-mL Isco Model 65D syringe pumps (Isco, Lincoln, NE), 2-position Valco valves (Valco Instruments, Houston, TX), and a PAL autosampler (Leap Technologies, Carrboro, NC), allowing for fully automated sample analysis across four separate HPLC columns.<sup>22</sup> Reversed-phase capillary HPLC columns were manufactured in-house by slurry packing 3- $\mu\text{m}$  Jupiter  $\text{C}_{18}$  stationary phase (Phenomenex, Torrance, CA) into a 60-cm length of 360  $\mu\text{m}$  o.d. x 75  $\mu\text{m}$  i.d. fused silica capillary tubing (Polymicro Technologies, Phoenix, AZ) that incorporated a 0.5- $\mu\text{m}$  retaining screen in a 1/16" custom laser-bored 75  $\mu\text{m}$  i.d. union (screen and union - Valco Instruments Co., Houston, TX; laser bore - Lenox Laser, Glen Arm, MD). Mobile phase consisted of 0.1% formic acid in water (A) and 0.1% formic acid in acetonitrile (B). The mobile phase was degassed using an in-line Degassex Model DG4400 vacuum degasser (Phenomenex). The HPLC system was equilibrated at 10 kpsi with 100% mobile phase A, and then a mobile phase selection valve was switched 50 min after injection, which created a near-exponential gradient as mobile phase B displaced A in a 2.5 mL active mixer. A 40-cm length of 360  $\mu\text{m}$  o.d. x 15  $\mu\text{m}$  i.d. fused silica tubing was used to split  $\sim 17$   $\mu\text{L}/\text{min}$  of flow before it reached the injection valve (5  $\mu\text{L}$  sample loop). The split flow controlled the gradient speed under conditions of constant pressure operation (10 kpsi). Flow through the capillary HPLC column when equilibrated to 100% mobile phase A was  $\sim 500$  nL/min. The SCX fractions were analyzed by an LTQ mass spectrometer (Thermo Scientific, San Jose, CA) coupled with an in-house developed electrospray ionization (ESI) interface and the ten most intensive peaks in the previous MS spectrum were selected for MS/MS in the linear ion trap.

Other OCT section or LMD samples digested by the three different protocols were analyzed by an LTQ-Orbitrap mass spectrometer (Thermo Scientific). In both LTQ and LTQ-Orbitrap analyses, data was acquired for 100 min, beginning 65 min after sample injection (15 min into gradient); a dynamic exclusion time of 60 s was used to discriminate against previously analyzed ions. Orbitrap spectra (AGC  $1 \times 10^6$ ) were collected from 400-2000 m/z at a resolution of 60,000 followed by data dependent ion trap MS/MS spectra (AGC  $1 \times 10^4$ ) of the six most abundant ions using a collision energy setting of 35%.

**Data analysis.** All MS/MS datasets obtained in the LTQ and LTQ-Orbitrap analyses were searched by SEQUEST (Thermo Scientific) against the human International Protein Index (IPI) database (version 3.41; consisting of 69,731 protein entries; available on line at [www.ebi.ac.uk/IPI](http://www.ebi.ac.uk/IPI)) using no enzyme rules. Tandem MS peaks were generated by extract\_msn.exe, part of the SEQUEST software package (for LTQ data), or DeconMSn<sup>23</sup> (for LTQ-Orbitrap data). No modification of the amino acid residues was employed. The peptide identifications were filtered by stringent criteria established by applying a decoy database search (mass error was used as an additional constraint for filtering the MS/MS data from the LTQ-Orbitrap analyses) to limit the false positive rate to <1% at the peptide level.<sup>24,25</sup>

Quantitative analysis of the breast tissue proteins in a total of 24 LC-MS datasets (3X3 using normal OCT section samples; 3X3 using LMD samples procured from normal OCT sections; 2 using cancer OCT section samples; and 2X2 using LMD samples procured from tumor OCT sections) utilized the PNNL-developed AMT tag approach. An AMT tag database was established by combining peptide and protein identifications obtained from the 2D-LC-MS/MS analyses of normal and cancer OCT section samples. A suite of in-house developed informatics tools were used to process LC-LTQ-Orbitrap MS data and correlate the resulting LC-MS features to the AMT tag database containing the characteristic accurate mass and LC elution time for the peptides identified from breast tissue proteins.<sup>26,27</sup> The initial analysis of raw LC-MS data involved a mass transformation, or deisotoping, step using Decon2LS,<sup>28</sup>

which generated a text file report for each LC-MS dataset including both the monoisotopic masses and the corresponding intensities for all detected species for each mass spectrum. Further downstream data analysis incorporated all of the possible detected peptides into a feature matching and visualization program VIPER<sup>29</sup> to correlate the LC-MS features to the peptide identifications in the AMT tag database and calculate their peak areas. After filtering by mass error and normalized elution time (NET) error, as well as a requirement of each peptide to be observed in at least 4 out of the 24 LC-MS datasets, the false positive rate of the AMT tag peak matching was estimated as <4% using an “11-Da shift” strategy.<sup>30</sup>

The ‘ProteinProphet’ program<sup>31</sup> was then used to define non-redundant protein groups for all peptides identified in both LC-MS/MS and LC-MS analyses as previously described.<sup>32</sup> Briefly, all peptides that passed the filtering criteria were assigned an identical probability score of 1 and entered into the software program (solely for clustering analysis) to generate a final list of nonredundant proteins or protein groups. One protein was randomly selected to represent each corresponding protein group that contains member database entries.

Data normalization and quantification of the relative protein abundances in different samples were performed and visualized using in-house developed software DANTE.<sup>33</sup> Briefly, peptide intensities from the LC-MS analyses were Log2 transformed and normalized using a mean central tendency procedure. Peptide abundances were then “rolled up” to the protein level employing the R-rollup method (based on trends observed at peptide level) implemented in DANTE.

Gene ontology annotation was performed using a software tool STRAP.<sup>34</sup> The final distribution charts were generated using Excel.

## **Results**

### **Protocol optimization for preparing OCT-embedded breast tissue samples for LC-MS analysis.**

Three “one-tube” protein extraction and digestion protocols based on different detergents or organic

solvents, namely PPS, RapiGest, and TFE, were tested for preparing the OCT-embedded breast tissue samples for LC-MS analysis. To compare the performance of the three different protocols, both OCT sections (12  $\mu\text{m}$  thick each) and LMD samples ( $\sim 50,000$  cells each) were prepared from one OCT-embedded normal breast tissue block. Three process replicates were performed for each protocol and sample type, and the resulting peptide samples were analyzed by LC-LTQ-Orbitrap. Peptides and proteins were identified from both LC-MS and MS/MS data using the AMT tag approach and the Sequest algorithm, respectively.

**Figure 1A** shows typical base peak chromatograms of the LC-MS analyses of samples prepared using the three protein extraction and digestion protocols. All the OCT section samples showed overwhelming OCT polymer contamination which heavily suppressed the peptide signals and hence severely compromised their MS detection. One example of the base peak chromatogram of an OCT section sample prepared by the PPS protocol is shown in the bottom panel of **Figure 1A**, and the averaged mass spectrum of the OCT polymers is shown in **Figure 1B**. Under the experimental conditions employed in this study, the OCT polymers were detected as 3+ ion series ranging from 400-700  $m/z$ , 2+ ion series ranging from 400-1100  $m/z$ , and 1+ ion series ranging from 900-2000  $m/z$ . In contrast all the LMD samples prepared by the three protocols showed no obvious sign of OCT polymer contamination and similar peptide elution patterns (**Figure 1A**).

The PPS protocol provided the highest peptide yield. For example, the averaged peptide yields of the PPS, RapiGest, and TFE protocols for the 12- $\mu\text{m}$  OCT sections from the same normal breast tissue block were 85, 70, and 56  $\mu\text{g}$ , respectively. **Table 1** summarizes the peptide and protein identifications obtained from the LC-MS analyses (through the AMT tag approach) of these samples prepared from normal breast tissue. Apparently, the PPS protocol provided the largest numbers of peptide and protein for both OCT sections and LMD samples. The least effective protocol, the TFE protocol, provided only  $\sim 40$ -60% peptide and  $\sim 45$ -65% protein identifications when compared to the PPS protocol. The

comparison of the overall reproducibility of the three protocols, as reflected by the relative standard deviation (RSD) and the averaged overlap of peptide and protein identifications in the three process replicates, was very similar in the LC-MS analyses of the LMD samples; however, the PPS protocol demonstrated the best reproducibility in the LC-MS analyses of the OCT section samples. In addition, LC-MS/MS data from the same analyses displayed the same trends (e.g., PPS provided the best coverage and reproducibility), but in general yielded much less (~30%) peptide and protein identifications than did in the LC-MS analysis (data not shown). Based on these results, the PPS protocol was chosen as the method of choice for the preparation and analysis of the breast tumor samples.

#### **SCX liquid chromatography for simultaneous OCT polymer removal and peptide fractionation.**

In the above analysis of normal breast tissue samples, the samples procured by LMD seemed to be free from the OCT polymer contamination and therefore, were ideal for direct analysis by either LC-MS or LC-MS/MS. However, the accumulation of a sufficient amount of cells by LMD for a 2D-LC-MS/MS analysis for increased proteome coverage could be quite laborious and time consuming. OCT sections can be readily obtained for such 2D analysis, but the severe polymer contamination presented a formidable challenge for direct LC-MS/MS analysis of this type of sample.

Since polyvinyl alcohol and polyethylene glycol (two of the three major components of the OCT compound) are nonionic polymers, we reason that SCX liquid chromatographic separation would remove them from the tryptic peptides that carry positive charge in SCX buffer A (10 mM ammonium formate, pH 3/25% acetonitrile). **Figure 2** shows the UV trace (280 nm) of a SCX separation of an OCT section sample procured from normal breast tissue and digested using the TFE protocol. The large amount of OCT polymers in the sample eluted very early (0-5% B) in the chromatogram as a large and broad peak. Benzalkonium chloride, that has an aromatic ring and consists of ~85% (w/w) of the OCT compound, is probably responsible for most of the absorbance at 280 nm; the other two nonionic

polymers could be eluting at the same time due to their lack of retention under the SCX conditions. Moreover, the OCT polymer contamination did not have any noticeable effect on the fractionation of the tryptic peptides by SCX. LC-MS/MS analysis of the 30 SCX fractions resulted in a total of 6,708 peptide identifications covering 1,994 proteins (1,074 proteins have two or more peptide identifications). The peptide and protein identifications are available in **Supplementary Table 1**.

After the initial optimization of protein extraction and digestion, an OCT section sample procured from breast tumor and digested using the best performing PPS protocol was also fractionated using the same SCX method that removes the OCT polymers. **Figure 3** shows the base peak chromatograms of the LC-MS/MS analyses of the 30 SCX fractions. Clearly, the OCT polymers that once produced dominant signals in LC-MS/MS analysis of the original OCT section sample (see **Figure 3** insert) were no longer detectable under identical LC-MS/MS conditions after the SCX separation. Comparing to the normal OCT section sample, the cancer OCT section sample that went through the same 2D-LC-MS/MS analysis provided substantially better proteome coverage. A total of 19,025 peptides were identified, covering 4557 non-redundant proteins (2852 proteins have two or more peptides). Corresponding peptide and protein identifications are available in **Supplementary Table 2**.

**Single-run LC-MS and LC-MS/MS analysis of breast tumor samples using the PPS protocol.** The best performing PPS protocol for the normal breast tissue was applied to the analysis of breast tumor OCT sections and LMD samples using LC-MS and LC-MS/MS. To test whether increased cell numbers have an effect on the final peptide and protein identifications, two LMD samples containing ~50,000 and 100,000 cells (each with two process replicates) were compared head-to-head; two 12- $\mu$ m thick OCT sections were also prepared from the same breast tumor tissue block for comparison. After LC-LTQ-Orbitrap analysis under the identical conditions, peptide and protein identifications obtained from both single-run LC-MS/MS and LC-MS levels were compared (**Table 2**).

Unlike the analyses of the normal breast tissue samples (**Table 1**), both LC-MS/MS and LC-MS analyses of the breast tumor samples, LMD samples free of OCT polymer contamination performed much better than un-decontaminated LMD samples. For example, in LC-MS/MS analysis the LMD samples provided ~3-fold more peptide identifications than did the OCT section samples (in LC-MS analysis they provided ~1.7-fold more peptide identifications). The sample preparation reproducibility (RSD and overlap in the number of peptide and protein identifications) was also much better when LMD samples were analyzed by both LC-MS/MS and LC-MS, demonstrating a dramatic effect of the OCT polymer removal on significantly increased proteome coverage and much improved reproducibility of the measurements. The severe interference of the OCT polymers on effective LC-MS analysis can be visualized in **Figure 4**: the strong signal from the OCT polymers (diagonal lines indicated by the arrows in **Figure 4A**) simply suppressed almost the entire peptide signal in the OCT polymer elution window and hence, resulted in very limited peptide identifications (**Figure 4B**); in comparison the LMD sample had many more peptide signals (**Figure 4C**) in the same region which led to successful peptide identifications through the AMT tag approach (**Figure 4D**).

As observed in the analyses of normal breast tissue OCT sections and LMD samples, LC-MS analysis of the same types of samples from breast tumor achieved significantly better proteome coverage and reproducibility than that obtained in the LC-MS/MS analysis, attributing to the much lower degree of undersampling in LC-MS. This effect was even more severe for the OCT section samples at both the peptide and protein levels. Moreover, greater than 6,000 peptides covering ~2,000 proteins can be identified from a single LC-MS analysis of the LMD samples (**Table 2**), demonstrating an important advancement in the proteomics analysis of breast tissue specimens.

The 50,000-cell LMD sample and the 100,000-cell LMD sample showed very similar numbers of peptide and protein identifications in both LC-MS/MS and LC-MS analyses. This is because even 50,000 cells were able to yield a sufficient amount of peptides (~20 µg) and the increase in cell number did not

have an effect on the overall peptide yield (i.e., ~40 µg for 100,000 cells) using the PPS protocol, and the loading (2 µg) was the same for these two LMD samples. The PPS protocol was able to robustly extract and digest the proteins in the 50,000-cell and 100,000-cell LMD samples for direct LC-MS analysis.

**Figure 5** shows LC-MS base peak chromatograms of the 50,000- and 100,000-cell LMD samples, and they displayed highly comparable peptide elution and intensity patterns. Comparing to similar analysis of the 50,000-cell LMD samples procured from normal breast tissue (**Figure 1**) and processed by the same PPS protocol, the breast tumor LMD samples showed more distinctive peaks that were more evenly distributed in the base peak chromatograms (**Figure 5**).

Combining the LMD and LC-MS for the analysis of the breast tumor samples resulted in very high overlap in the peptide and protein identifications (see **Table 2**; ~90% and 93%, respectively), representing a significant improvement in qualitative reproducibility compared to that of direct analysis of the OCT section samples applying conventional LC-MS/MS analysis (see **Table 2**; 58% and 74%, respectively). To further evaluate the reproducibility at the protein quantification level, label-free quantification was performed for all the breast tissue samples (both normal and tumor tissue) that were prepared by the PPS protocol and analyzed by LC-MS in process replicates. Both the 50,000- and 100,000-cell breast tumor LMD samples had a very high correlation score ( $R^2 > 0.98$ ) for protein abundances between the replicates (**Figure 6A** and **6B**), again clearly demonstrating the highly reliable nature of such analysis for quantitative proteome characterization of breast tumor specimens. Replicates of the breast tumor OCT section samples (**Figure 6C**) had a lower, though similar, correlation score ( $R^2 \sim 0.89$ ) as that of the replicates of the normal breast tissue LMD samples (**Figure 6D**), reflecting a higher level of variation in the cell type and distribution in those sample types.

## Discussion

Molecular analysis of differences in the DNA, RNA, and protein content of tumor cells compared to their normal counterparts has proven to be a highly useful approach to fundamental advances in our understanding of the processes leading to malignant transformation. In addition, statistically robust changes in gene expression or protein abundance /post-translational modification between normal and cancerous tissue have the potential to serve as biomarkers, either for early diagnosis of cancer, for estimation of prognosis, or for matching of molecular therapies to molecular defects. However, the full potential of molecular analyses cannot be realized until several technical hurdles have been overcome. Chief among these is the difficulty in obtaining sufficiently homogeneous cell populations for valid assessment of molecular markers.

The use of LMD to isolate histologically distinct regions containing a single predominant cell type (e.g., tumor versus adjacent normal, or epithelial versus stromal) has proven highly beneficial for gene expression profiling experiments, particularly as PCR-based methods overcome many of the limitations of small sample size. Proteome analysis of samples isolated by LMD has proven to be much more challenging, in part due to sample size limitations, but also due to the presence of contaminants in tissues preserved with either FFPE or OCT. In this report we compare three alternative “one-tube” protocols, applied to both bulk OCT sections and sub-regions isolated by LMD.

“One-tube” protocols for protein extraction and digestion of small-sized biospecimens have become a popular choice for proteome analysis as these protocols minimize potential sample loss.<sup>35</sup> However, when using “one-tube” protocols without a separate sample clean-up step, it is necessary to avoid the use of strong detergents such as SDS and/or chaotropic agents such as urea or thiourea for protein extraction and denaturation, when it is desired to use LC-MS for analysis, rather than PAGE. Instead, cleavable detergents (such as PPS and RapiGest) that are readily hydrolyzed at low pH into easily removed, soluble by-products with no residual detergent properties are used to generate ESI-compatible peptide preparations for more sensitive LC-MS and LC-MS/MS analysis; simple organic

solvents (e.g., TFE, methanol) are also effective. While these “one-tube” protocols simplify sample processing and minimize sample loss, there remain concerns about protein extraction and digestion efficiency and whether the other cellular by-products (e.g., lipids and carbohydrates) interfere with the peptide-centric proteomics analysis. Our results indicated that the PPS protocol provided the highest peptide yield as well as the largest number of peptide and protein identifications (see **Table 1**) for the OCT-embedded breast tissue samples. However, the effectiveness of the cleavable detergents and organic solvents in terms of protein extraction and digestion can be sample-dependent. For example, the least effective protocol in this study, the TFE protocol, has been working very well in the study of mouse brain samples,<sup>36</sup> implying that the biochemical characteristics of the tissue will have a bearing on the selection of the most effective extraction protocols. In the LMD-processed and PPS-digested breast tissue samples, there are some lower molecular weight species (such as the horizontal lines in **Figure 4C** after scan #12486) that eluted late in the RPLC separation; however, they do not seem to be interfering with the LC-MS analysis through the AMT tag approach (same for the other two protocols, data not shown). The fact that we can reproducibly identify and quantify greater than 6,000 peptides covering ~2,000 proteins in single-run LC-MS analysis of the PPS tissue digests suggest that this “one-tube” protocol is sufficiently effective and robust for most proteomics applications.

One advantage of the proteomics strategy for the breast tissue analysis presented in this study is that readily available OCT sections (using SCX liquid chromatography to remove the polymers while fractionating the peptides) could be used for building an actual tissue AMT tag database. We could then comprehensively and quantitatively analyze the LMD samples using single-run LC-MS analysis for large-scale applications. Previously, analyses of LMD breast tissue samples used AMT tag databases that were established through 2D-LC-MS/MS analyses of cultured cell lines, e.g., human mammary epithelial cells and MCF-7 cells.<sup>10</sup> Although these cell lines represent aspects of normal and malignant breast tissue, they reportedly have a distinct molecular profile compared to primary cells or tissues.<sup>37</sup> As a result,

proteins expressed solely in tumor tissues may be absent from the AMT tag database of cultured cell lines and hence, negatively affect the overall peptide identification efficiency when applying the AMT tag approach. For example, previous LC-MS analysis of LMD breast tissue samples using such cultured cell line databases resulted in ~20% identification rate for the unique LC-MS features;<sup>10</sup> in the current study, we were able to achieve a >50% identification rate at an even lower FDR level (10% vs. 4%) using a tissue AMT tag database that was generated through 2D-LC-MS/MS analysis of samples prepared from exactly the same tissue block. In comparative analysis of different tissues with distinct molecular or phenotypic characteristics, such as metastasis, ER/PR/HER2 status, etc., it is advantageous to establish the AMT tag database for multiple tissue types, even if it is necessary to pool tissue samples.

In this study the breast tumor samples provided much broader proteome coverage than that obtained with the normal breast tissue samples (see **Table 1** and **Table 2** for details). This does not appear to be a function of the amount of protein in the sample, as we loaded similar amount of samples for both LC-MS (2 µg per injection) and 2D-LC-MS/MS (a total of 100 µg for SCX) analyses of the normal and cancer breast tissue specimens. This is also reflected in **Figure 1** and **Figure 5** as the analyses of normal and cancer tissue showed comparable LC-MS peak intensity. The peptide yield for the normal and cancer tissue is comparable, too. For example, the PPS-digested normal and cancer OCT sections (12 µm) yielded 85±8 and 74±7 µg of peptides, respectively. The difference in the extent of proteome coverage observed for normal versus cancer breast tissue may have intrinsic differences in protein content between epithelial and stromal cells (histological images of normal and cancerous breast tissues are available in **Supplementary Figure 1**). A comparison of the GO cellular component terms in the normal tissue and tumor samples showed significant differences in several categories. For instance, compared to their normal tissue counterparts, both tumor LMD and OCT section samples showed consistently higher protein representation in nucleus and endoplasmic reticulum, and lower protein representations in cytoskeleton, plasma membrane, cell surface, and extracellular space (**Figure 7**). To

confirm this, we examined the MS identification of the keratins in the normal tissue and tumor samples, as a specific example of a major cytoskeletal protein. The results showed that indeed keratin proteins have a much higher representation in the normal tissue samples: 1) in the LC-MS analyses of the LMD samples, an increase (normal versus cancer tissue) of 1.8X and 1.6X was observed for the percentage of keratin protein and peptide identifications, respectively; and 2) in the 2D-LC-MS/MS analyses of the OCT section samples, corresponding increase was 2.9X and 5.1X, respectively (see Table 3). Although it was not a head-to-head comparison for the OCT section samples (i.e., the normal tissue sample was digested using the TFE protocol and the tumor sample was digested using PPS protocol), the results showed the same trend in terms of significantly increased keratin proteins in the normal tissue samples. Among the most represented subcellular locations, there was no obvious difference in cytoplasm (~19% of all GO terms), while a significant increase was observed for nucleus (~20% of all GO terms) in the tumor samples, suggesting increased expression of proteins in the nucleus and potentially more heterogeneous populations of the tumor cells. This result is also in alignment with the well-known increase in nucleus to cytoplasm ratio in transformed cells.

To summarize, in this study we developed a highly effective strategy that is suitable for quantitative, comprehensive, and high-throughput proteome characterization of archived OCT-embedded breast tissue specimens for biomedical and clinical applications. This is especially attractive for systems biology study because the OCT-embedded tissue specimens have proven to be reliable resources for immunohistopathological characterization as well as comprehensive molecular examination by multiple “omics” technologies. However, the data presented here by no means represent the best results achievable by this strategy. Many analytical aspects for more sensitive and comprehensive analysis, such as capillary LC columns with smaller inner diameter, longer LC gradient, and faster and more sensitive high-resolution MS instruments, can be employed to further improve the capability of LC-MS coupled to AMT tag databases for analyzing smaller-sized samples and achieving

even broader proteome coverage and dynamic range. In addition, this strategy can be readily implemented for large-scale proteomics analysis of other tissue samples that were preserved using OCT-embedding.

### **Acknowledgement**

We thank the research funds and support received from the US Department of Defense, Henry Jackson Foundation for the Advancement of Military Medicine. Portions of this research were supported by the USAMRMC (W81XWH-05-2-0053) and the NIH National Center for Research Resources for Integrative Biology (RR018522). The experimental work was performed in the Environmental Molecular Sciences Laboratory, a U.S. Department of Energy (DOE) Office of Biological and Environmental Research national scientific user facility on the PNNL campus. PNNL is multiprogram national laboratory operated by Battelle for the DOE under Contract DE-AC05-76RLO 1830.

## References

- (1) Hanahan, D.; Weinberg, R. A., The hallmarks of cancer. *Cell* **2000**, *100*, (1), 57-70.
- (2) Lund, A. H.; van Lohuizen, M., Epigenetics and cancer. *Genes Dev.* **2004**, *18*, (19), 2315-2335.
- (3) Curran, S.; Murray, G. I., An introduction to laser-based tissue microdissection techniques. *Methods Mol. Biol.* **2005**, *293*, 3-8.
- (4) Shi, S. R.; Datar, R.; Liu, C.; Wu, L.; Zhang, Z.; Cote, R. J.; Taylor, C. R., DNA extraction from archival formalin-fixed, paraffin-embedded tissues: heat-induced retrieval in alkaline solution. *Histochem. Cell Biol.* **2004**, *122*, (3), 211-218.
- (5) Finke, J.; Fritzen, R.; Ternes, P.; Lange, W.; Dolken, G., An improved strategy and a useful housekeeping gene for RNA analysis from formalin-fixed, paraffin-embedded tissues by PCR. *BioTechniques* **1993**, *14*, (3), 448-453.
- (6) Crockett, D. K.; Lin, Z. S.; Vaughn, C. P.; Lim, M. S.; Elenitoba-Johnson, K. S. J., Identification of proteins from formalin-fixed paraffin-embedded cells by LC-MS/MS. *Lab. Invest.* **2005**, *85*, (11), 1405-1415.
- (7) Shi, S. R.; Liu, C.; Balgley, B. M.; Lee, C.; Taylor, C. R., Protein extraction from formalin-fixed, paraffin-embedded tissue sections: Quality evaluation by mass spectrometry. *J. Histochem. Cytochem.* **2006**, *54*, (6), 739-743.
- (8) Lemaire, R.; Desmons, A.; Tabet, J. C.; Day, R.; Salzet, M.; Fournier, I., Direct analysis and MALDI imaging of formalin-fixed, paraffin-embedded tissue sections. *J. Proteome Res.* **2007**, *6*, (4), 1295-1305.
- (9) Ostasiewicz, P.; Zielinska, D. F.; Mann, M.; Wisniewski, J. R., Proteome, Phosphoproteome, and N-Glycoproteome Are Quantitatively Preserved in Formalin-Fixed Paraffin-Embedded Tissue and Analyzable by High-Resolution Mass Spectrometry. *J. Proteome Res.* **2010**, Epub ahead of print.
- (10) Umar, A.; Luider, T. M.; Foekens, J. A.; Pasa-Tolic, L., NanoLC-FT-ICR MS improves proteome coverage attainable for similar to 3000 laser-microdissected breast carcinoma cells. *Proteomics* **2007**, *7*, (2), 323-329.
- (11) Umar, A.; Kang, H.; Timmermans, A. M.; Look, M. P.; Meijer-van Gelder, M. E.; den Bakker, M. A.; Jaitly, N.; Martens, J. W. M.; Luider, T. M.; Foekens, J. A.; Pasa-Tolic, L., Identification of a Putative Protein Profile Associated with Tamoxifen Therapy Resistance in Breast Cancer. *Mol. Cell. Proteomics* **2009**, *8*, (6), 1278-1294.
- (12) Hood, B. L.; Darfler, M. M.; Guiel, T. G.; Furusato, B.; Lucas, D. A.; Ringeisen, B. R.; Sesterhenn, I. A.; Conrads, T. P.; Veenstra, T. D.; Krizman, D. B., Proteomic analysis of formalin-fixed prostate cancer tissue. *Mol. Cell. Proteomics* **2005**, *4*, (11), 1741-1753.
- (13) Guo, T.; Wang, W. J.; Rudnick, P. A.; Song, T.; Li, J.; Zhuang, Z. P.; Weil, R. J.; DeVoe, D. L.; Lee, C. S.; Balgley, B. M., Proteome analysis of microdissected formalin-fixed and paraffin-embedded tissue specimens. *J. Histochem. Cytochem.* **2007**, *55*, (7), 763-772.
- (14) Tuefferd, M.; De Bondt, A.; Van Den Wyngaert, I.; Talloen, W.; Verbeke, T.; Carvalho, B.; Clevert, D. A.; Alifano, M.; Raghavan, N.; Amaratunga, D.; Gohlmann, H.; Broet, P.; Camilleri-Broet, S., Genome-wide copy number alterations detection in fresh frozen and matched FFPE samples using SNP 6.0 arrays. *Genes Chromosomes Cancer* **2008**, *47*, (11), 957-964.
- (15) Sanchez-Navarro, I.; Gamez-Pozo, A.; Gonzalez-Baron, M.; Pinto-Marín, Á.; Hardisson, D.; Lopez, R.; Madero, R.; Cejas, P.; Mendiola, M.; Espinosa, E.; Vara, J. A. F., Comparison of gene expression profiling by reverse transcription quantitative PCR between fresh frozen and formalin-fixed, paraffin-embedded breast cancer tissues. *BioTechniques* **2010**, *48*, (5), 389-397.
- (16) von Ahlfen, S.; Missel, A.; Bendrat, K.; Schlumpberger, M., Determinants of RNA Quality from FFPE Samples. *PLoS One* **2007**, *2*, (12), 7.

- (17) Somiari, R. I.; Sullivan, A.; Russell, S.; Somiari, S.; Hu, H.; Jordan, R.; George, A.; Katenhusen, R.; Buchowiecka, A.; Arciero, C.; Brzeski, H.; Hooke, J.; Shriver, C., High-throughput proteomic analysis of human infiltrating ductal carcinoma of the breast. *Proteomics* **2003**, *3*, (10), 1863-1873.
- (18) Lim, J.; Kim, Y.; Lee, W.; Kim, M.; Lee, E. J.; Kang, C. S.; Han, K., Fresh-frozen, optimal cutting temperature (OCT) compound-embedded bone marrow aspirates: a reliable resource for morphological, immunohistochemical and molecular examinations. *Int. J. Lab. Hematol.* **2010**, *32*, (1), E34-E39.
- (19) Zang, L.; Toy, D. P.; Hancock, W. S.; Sgroi, D. C.; Karger, B. L., Proteomic analysis of ductal carcinoma of the breast using laser capture microdissection, LC-MS, and O-16/O-18 isotopic labeling. *J. Proteome Res.* **2004**, *3*, (3), 604-612.
- (20) Johann, D. J.; Rodriguez-Canales, J.; Mukherjee, S.; Prieto, D. A.; Hanson, J. C.; Emmert-Buck, M.; Blonder, J., Approaching Solid Tumor Heterogeneity on a Cellular Basis by Tissue Proteomics Using Laser Capture Microdissection and Biological Mass Spectrometry. *J. Proteome Res.* **2009**, *8*, (5), 2310-2318.
- (21) Smith, R. D.; Anderson, G. A.; Lipton, M. S.; Pasa-Tolic, L.; Shen, Y. F.; Conrads, T. P.; Veenstra, T. D.; Udseth, H. R., An accurate mass tag strategy for quantitative and high-throughput proteome measurements. *Proteomics* **2002**, *2*, (5), 513-523.
- (22) Livesay, E. A.; Tang, K. Q.; Taylor, B. K.; Buschbach, M. A.; Hopkins, D. F.; LaMarche, B. L.; Zhao, R.; Shen, Y. F.; Orton, D. J.; Moore, R. J.; Kelly, R. T.; Udseth, H. R.; Smith, R. D., Fully automated four-column capillary LC-MS system for maximizing throughput in proteomic analyses. *Anal. Chem.* **2008**, *80*, (1), 294-302.
- (23) Mayampurath, A. M.; Jaitly, N.; Purvine, S. O.; Monroe, M. E.; Auberry, K. J.; Adkins, J. N.; Smith, R. D., DeconMSn: a software tool for accurate parent ion monoisotopic mass determination for tandem mass spectra. *Bioinformatics* **2008**, *24*, (7), 1021-3.
- (24) Elias, J. E.; Gygi, S. P., Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Methods* **2007**, *4*, (3), 207-214.
- (25) Qian, W. J.; Liu, T.; Monroe, M. E.; Strittmatter, E. F.; Jacobs, J. M.; Kangas, L. J.; Petritis, K.; Camp, D. G.; Smith, R. D., Probability-based evaluation of peptide and protein identifications from tandem mass spectrometry and SEQUEST analysis: The human proteome. *J. Proteome Res.* **2005**, *4*, (1), 53-62.
- (26) Zimmer, J. S. D.; Monroe, M. E.; Qian, W. J.; Smith, R. D., Advances in proteomics data analysis and display using an accurate mass and time tag approach. *Mass Spectrom. Rev.* **2006**, *25*, (3), 450-482.
- (27) Qian, W. J.; Liu, T.; Petyuk, V. A.; Gritsenko, M. A.; Petritis, B. O.; Polpitiya, A. D.; Kaushal, A.; Xiao, W.; Finnerty, C. C.; Jeschke, M. G.; Jaitly, N.; Monroe, M. E.; Moore, R. J.; Moldawer, L. L.; Davis, R. W.; Tompkins, R. G.; Herndon, D. N.; Camp, D. G.; Smith, R. D.; Program, I. a. t. H. R. t. I. L. S. C. R., Large-Scale Multiplexed Quantitative Discovery Proteomics Enabled by the Use of an O-18-Labeled "Universal" Reference Sample. *J. Proteome Res.* **2009**, *8*, (1), 290-299.
- (28) Jaitly, N.; Mayampurath, A.; Littlefield, K.; Adkins, J. N.; Anderson, G. A.; Smith, R. D., Decon2LS: An open-source software package for automated processing and visualization of high resolution mass spectrometry data. *BMC Bioinformatics* **2009**, *10*, (1), 87-101.
- (29) Monroe, M. E.; Tolic, N.; Jaitly, N.; Shaw, J. L.; Adkins, J. N.; Smith, R. D., VIPER: an advanced software package to support high-throughput LC-MS peptide identification. *Bioinformatics* **2007**, *23*, (15), 2021-3.
- (30) Petyuk, V. A.; Qian, W. J.; Chin, M. H.; Wang, H. X.; Livesay, E. A.; Monroe, M. E.; Adkins, J. N.; Jaitly, N.; Anderson, D. J.; Camp, D. G.; Smith, D. J.; Smith, R. D., Spatial mapping of protein abundances in the mouse brain by voxelation integrated with high-throughput liquid chromatography-mass spectrometry. *Genome Res.* **2007**, *17*, (3), 328-336.

- (31) Nesvizhskii, A. I.; Keller, A.; Kolker, E.; Aebersold, R., A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* **2003**, *75*, (17), 4646-4658.
- (32) Liu, T.; Qian, W. J.; Gritsenko, M. A.; Xiao, W. Z.; Moldawer, L. L.; Kaushal, A.; Monroe, M. E.; Varnum, S. M.; Moore, R. J.; Purvine, S. O.; Maier, R. V.; Davis, R. W.; Tompkins, R. G.; Camp, D. G.; Smith, R. D., High dynamic range characterization of the trauma patient plasma proteome. *Mol. Cell. Proteomics* **2006**, *5*, (10), 1899-1913.
- (33) Polpitiya, A. D.; Qian, W. J.; Jaitly, N.; Petyuk, V. A.; Adkins, J. N.; Camp, D. G.; Anderson, G. A.; Smith, R. D., DAnTE: a statistical tool for quantitative analysis of -omics data. *Bioinformatics* **2008**, *24*, (13), 1556-1558.
- (34) Bhatia, V. N.; Perlman, D. H.; Costello, C. E.; McComb, M. E., Software Tool for Researching Annotations of Proteins: Open-Source Protein Annotation Software with Data Visualization. *Anal. Chem.* **2009**, *81*, (23), 9819-9823.
- (35) Wang, H. X.; Qian, W. J.; Mottaz, H. M.; Clauss, T. R. W.; Anderson, D. J.; Moore, R. J.; Camp, D. G.; Khan, A. H.; Sforza, D. M.; Pallavicini, M.; Smith, D. J.; Smith, R. D., Development and evaluation of a micro- and nanoscale proteomic sample preparation method. *J. Proteome Res.* **2005**, *4*, (6), 2397-2403.
- (36) Wang, H. X.; Qian, W. J.; Chin, M. H.; Petyuk, V. A.; Barry, R. C.; Liu, T.; Gritsenko, M. A.; Mottaz, H. M.; Moore, R. J.; Camp, D. G.; Khan, A. H.; Smith, D. J.; Smith, R. D., Characterization of the mouse brain proteome using global proteomic analysis complemented with cysteinyl-peptide enrichment. *J. Proteome Res.* **2006**, *5*, (2), 361-369.
- (37) Ornstein, D. K.; Gillespie, J. W.; Paweletz, C. P.; Duray, P. H.; Herring, J.; Vocke, C. D.; Topalian, S. L.; Bostwick, D. G.; Linehan, W. M.; Petricoin, E. F.; Emmert-Buck, M. R., Proteomic analysis of laser capture microdissected human prostate cancer and in vitro prostate cell lines. *Electrophoresis* **2000**, *21*, (11), 2235-2242.

**Table 1.** Summary of the peptide and protein identifications obtained from LC-MS analyses of normal breast tissue samples using the three “one-tube” digestion protocols (each with three process replicates). RG: RapiGest™ SF Surfactant.

Sample/ protocol	Peptides			All proteins			Proteins with $\geq 2$ peptides		
	avg. # of IDs	RSD	avg. overlap	avg. # of IDs	RSD	avg. overlap	avg. # of IDs	RSD	avg. overlap
LMD_PPS	2059	6.9%	64.9%	837	6.8%	72.6%	353	8.7%	80.8%
LMD_RG	1383	10.4%	68.6%	661	3.7%	76.9%	278	11.4%	80.3%
LMD_TFE	1181	13.1%	69.6%	534	9.5%	79.0%	228	9.3%	81.1%
OCT_PPS	1859	3.6%	72.8%	836	0.3%	81.2%	342	5.8%	85.0%
OCT_RG	1256	9.3%	66.5%	582	4.9%	75.1%	259	9.0%	83.0%
OCT_TFE	705	14.4%	60.0%	377	18.8%	67.4%	157	10.8%	77.1%

**Table 2.** Summary of the peptide and protein identifications obtained from LC-MS/MS and LC-MS analyses of breast tumor samples using the PPS digestion protocols (each with two process replicates).

Sample/ MS method	Peptides			All proteins			Proteins with >=2 peptides		
	avg. # of IDs	RSD	overlap	avg. # of IDs	RSD	overlap	avg. # of IDs	RSD	overlap
<b>LC-MS/MS<sup>1</sup></b>									
LMD_100K <sup>2</sup>	3313	0.1%	72.6%	1231	0.2%	82.9%	637	3.9%	83.1%
LMD_50K <sup>3</sup>	2917	0.3%	69.5%	1222	5.4%	81.9%	611	2.8%	80.5%
OCT section <sup>4</sup>	1131	20.9%	57.7%	516	19.7%	73.8%	261	19.2%	77.6%
<b>LC-MS<sup>1</sup></b>									
LMD_100K <sup>2</sup>	6373	0.9%	89.6%	1998	1.0%	93.4%	1024	1.7%	93.2%
LMD_50K <sup>3</sup>	6092	2.5%	87.4%	1957	1.7%	92.6%	1020	0.9%	93.3%
OCT section <sup>4</sup>	3551	7.7%	77.4%	1446	5.6%	84.1%	673	5.0%	87.5%

<sup>1</sup>The results were obtained from the same LC-LTQ-Orbitrap analysis

<sup>2</sup>LMD samples with an estimated 100,000 cells

<sup>3</sup>LMD samples with an estimated 50,000 cells

<sup>4</sup>OCT section is 12  $\mu$ m in thickness

**Table 3.** Differential detection of keratins in the normal breast tissue and breast tumor samples.

Sample	Total protein	Keratin	Keratin protein %	Total keratin peptide	Total peptide	Keratin peptide %	Total keratin spectral count	Total spectral count	Keratin spectral count %
cLMD <sup>1</sup>	2265	31	1.37	846	30828	2.74	n/a	n/a	n/a
nLMD <sup>2</sup>	1208	30	2.48	389	8624	4.51	n/a	n/a	n/a
cOCT <sup>3</sup>	4557	23	0.50	134	21301	0.63	340	54138	0.63
nOCT <sup>4</sup>	1994	29	1.45	261	8166	3.20	1231	34715	3.55

<sup>1</sup>Tumor LMD samples (100K and 50K cells) digested by the PPS protocol (two process replicates each); analyzed by single LC-MS.

<sup>2</sup>Normal LMD samples digested by the PPS protocol (three process replicates); analyzed by single LC-MS.

<sup>3</sup>Tumor OCT sample digested by the PPS protocol; analyzed by 2D-LC-MS/MS (30 SCX fractions).

<sup>4</sup>Normal OCT sample digested by the TFE protocol; analyzed by 2D-LC-MS/MS (30 SCX fractions).

**Figure 1.** Base peak chromatograms of LC-MS analyses of LMD and OCT section samples prepared from the same normal breast tissue block using the three different “one-tube” digestion protocols (A) and averaged mass spectrum for the predominant OCT polymers (B). RG: RapiGest surfactant.

**Figure 2.** SCX liquid chromatography separation of a tryptic digest prepared from OCT-embedded normal breast tissue. Simultaneous polymer removal and peptide fractionation were achieved.

**Figure 3.** Base peak chromatograms of LC-MS/MS analyses of a breast tumor OCT section sample digested by the PPS protocol and fractionated by SCX liquid chromatography while having the OCT polymers removed. An OCT section sample procured from the same breast tumor tissue block was analyzed under the same conditions for comparison (insert in the right bottom corner).

**Figure 4.** 2D displays of LC-MS analyses of the OCT section and 50,000-cell LMD samples prepared from the same OCT-embedded breast tumor before (A: OCT section; C: LMD) and after (B: OCT section; D: LMD) peak matching using the AMT tag approach. The predominant OCT polymers in the OCT section sample (the diagonal lines) are indicated by arrows.

**Figure 5.** Base peak chromatograms of LC-MS analyses of LMD samples (estimated 100,000 and 50,000 cells, respectively) prepared from the same OCT-embedded breast tumor using the PPS digestion protocol.

**Figure 6.** Correlation of relative protein abundances in the process replicates of different breast tissue samples prepared by the PPS protocol. Details of the LC-MS analysis and data analysis are described in **Materials and Methods**.

**Figure 7.** Distribution of the GO cellular component terms for the cancer and normal breast tissue samples. Details of the analyses of the four samples are described in the footnote of **Table 3**.

Figure 1.

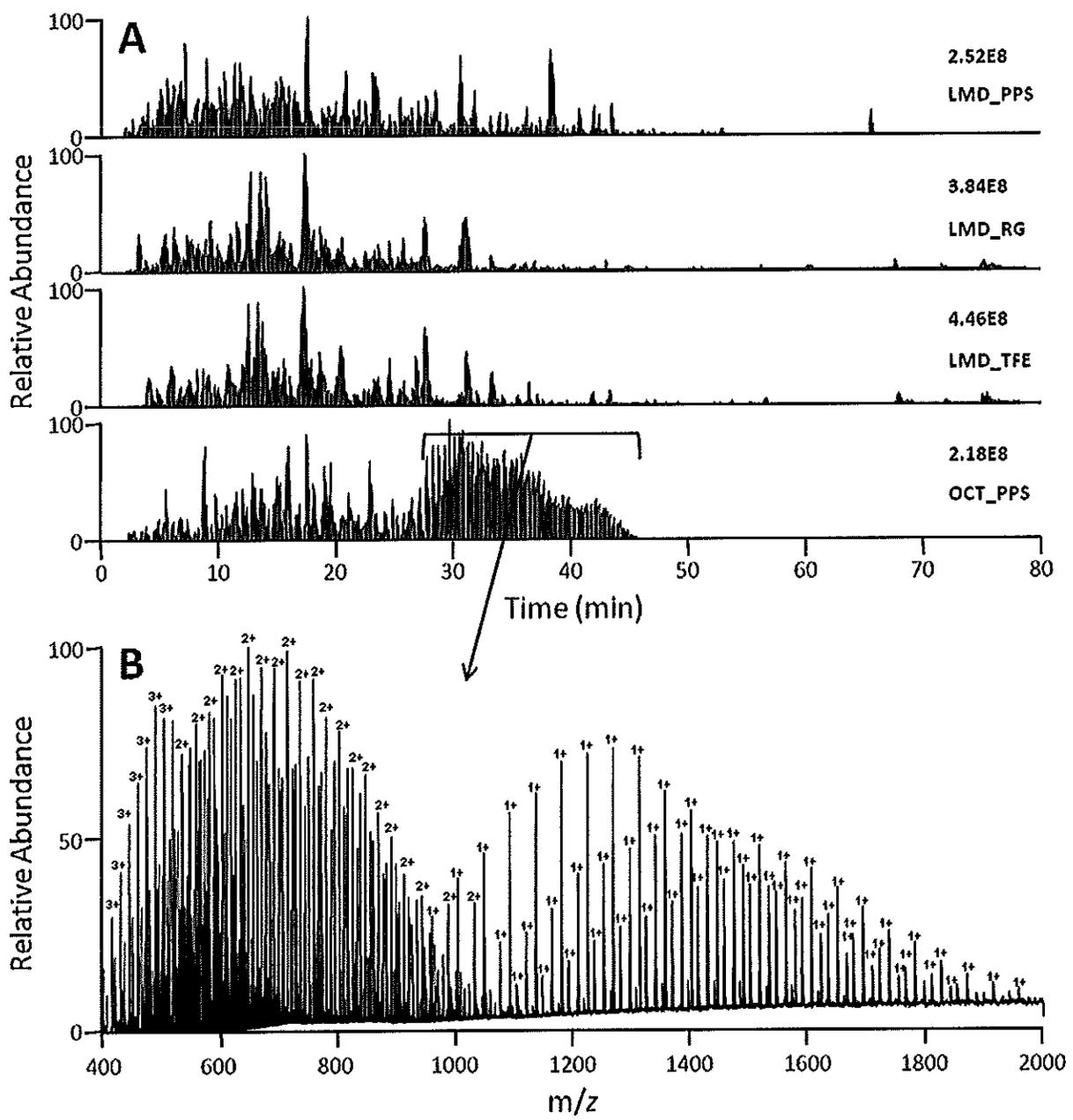


Figure 2.

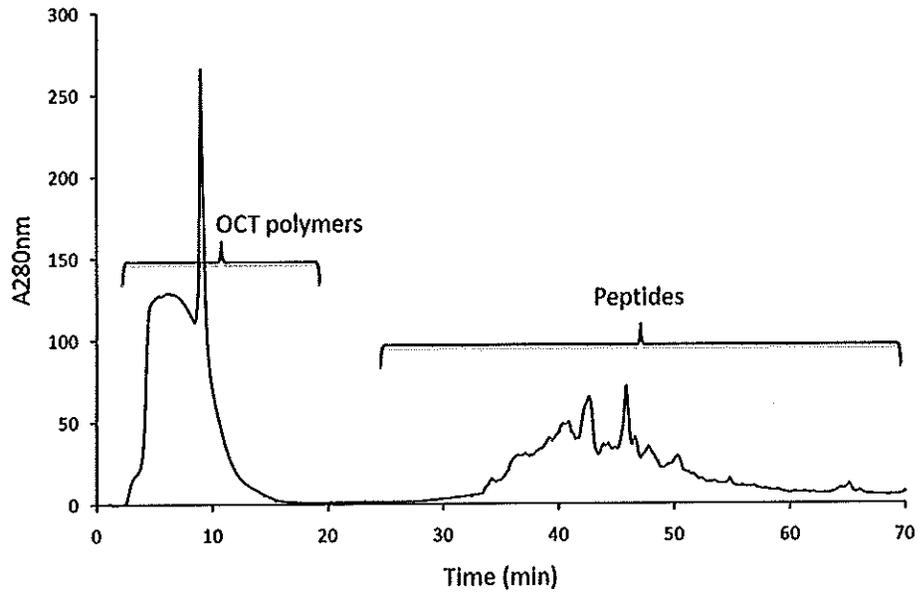


Figure 3.

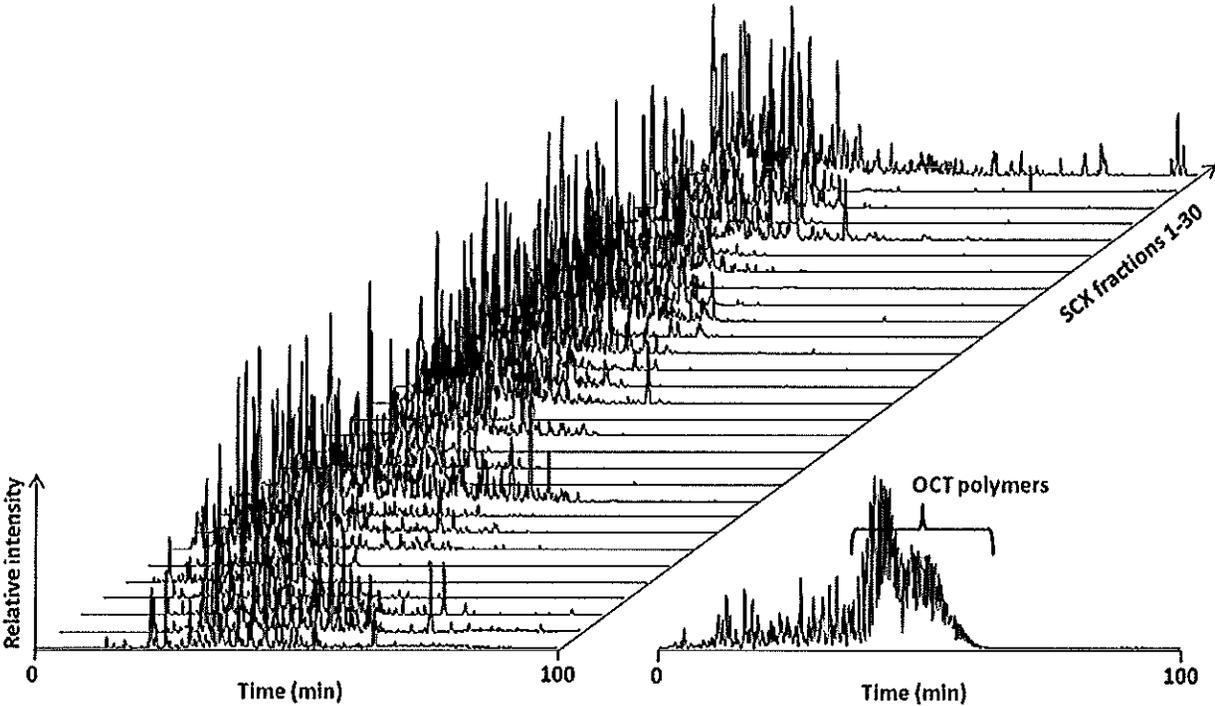


Figure 4.

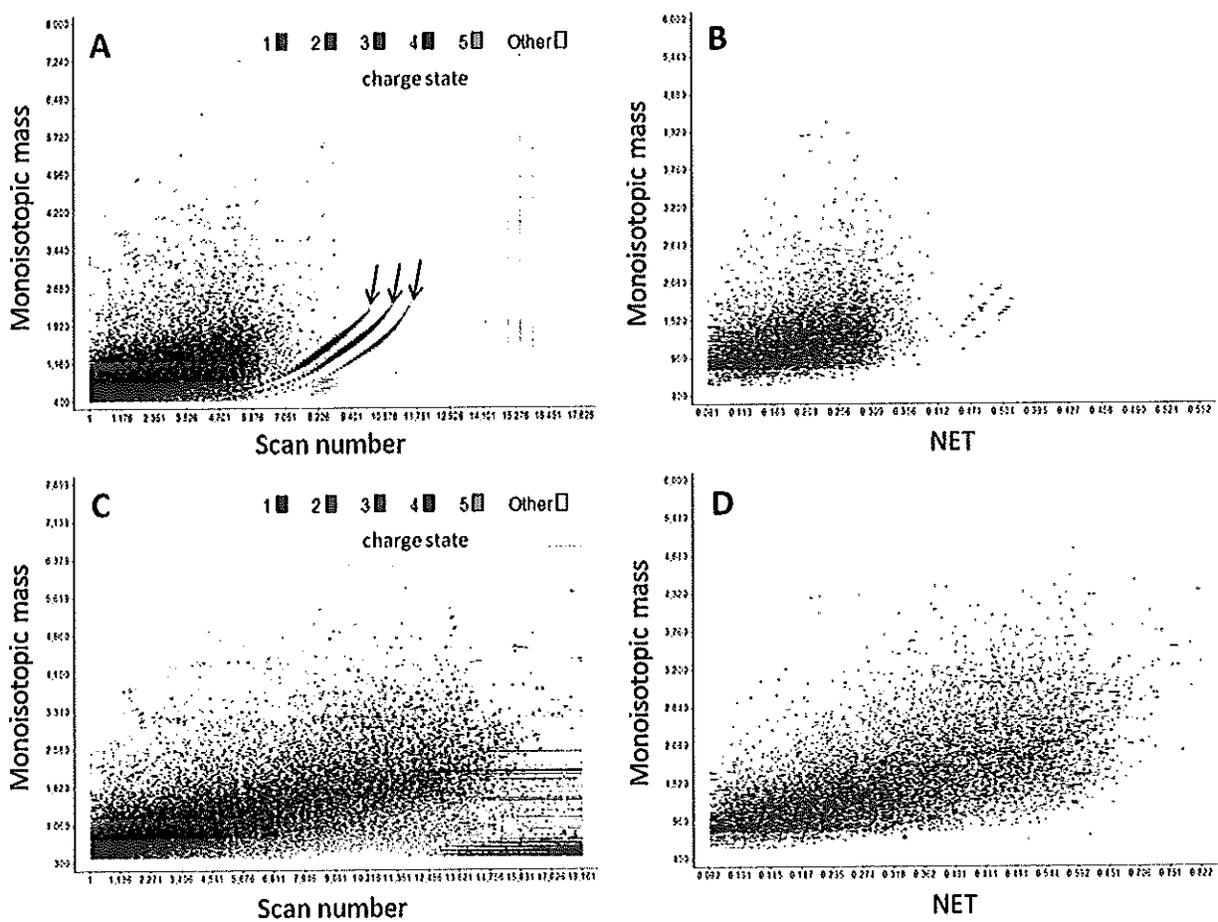


Figure 5.

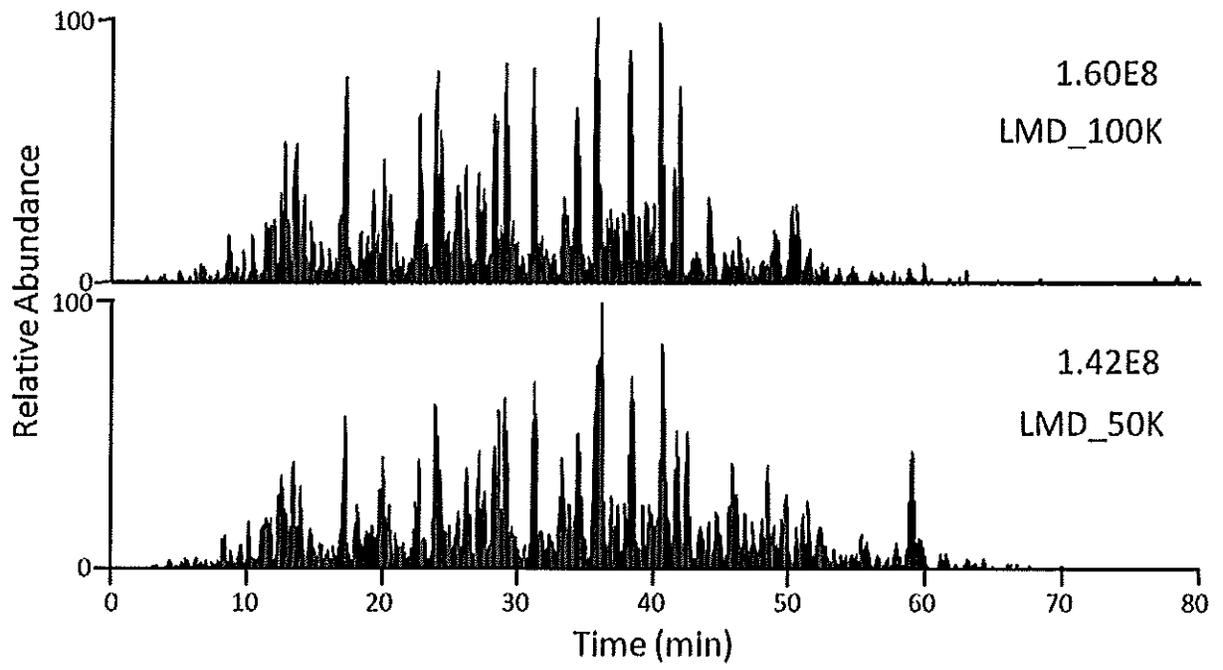


Figure 6.

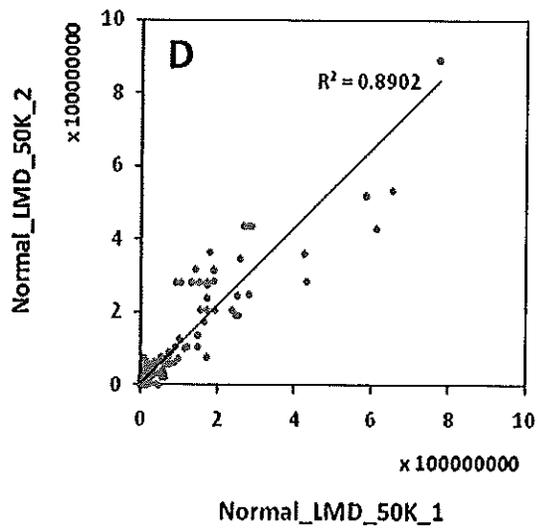
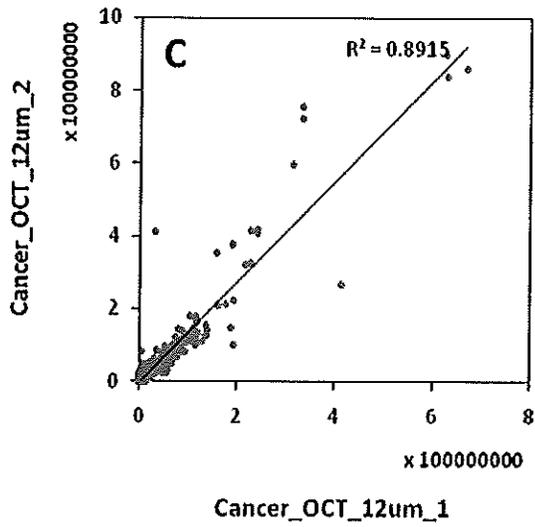
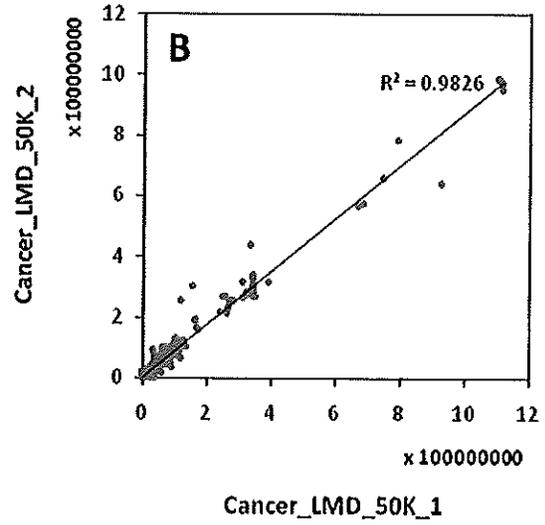
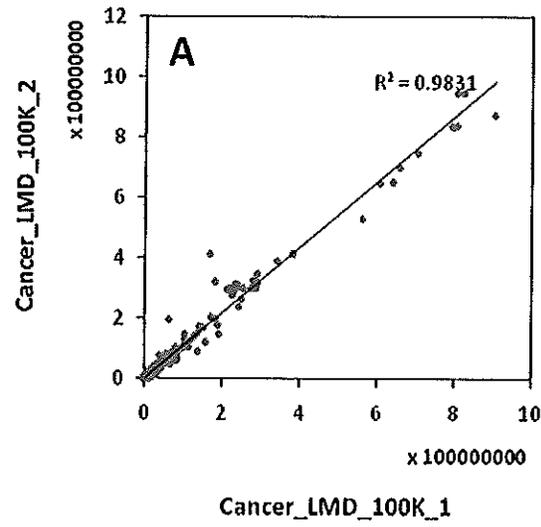
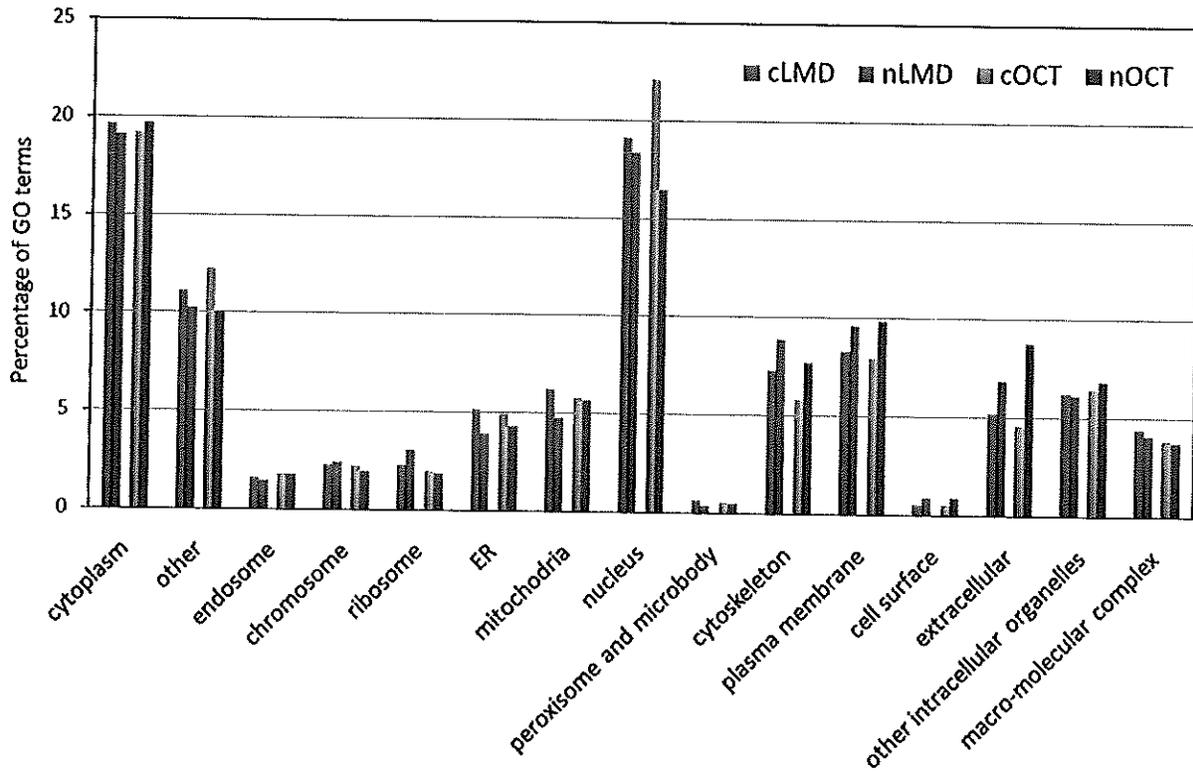
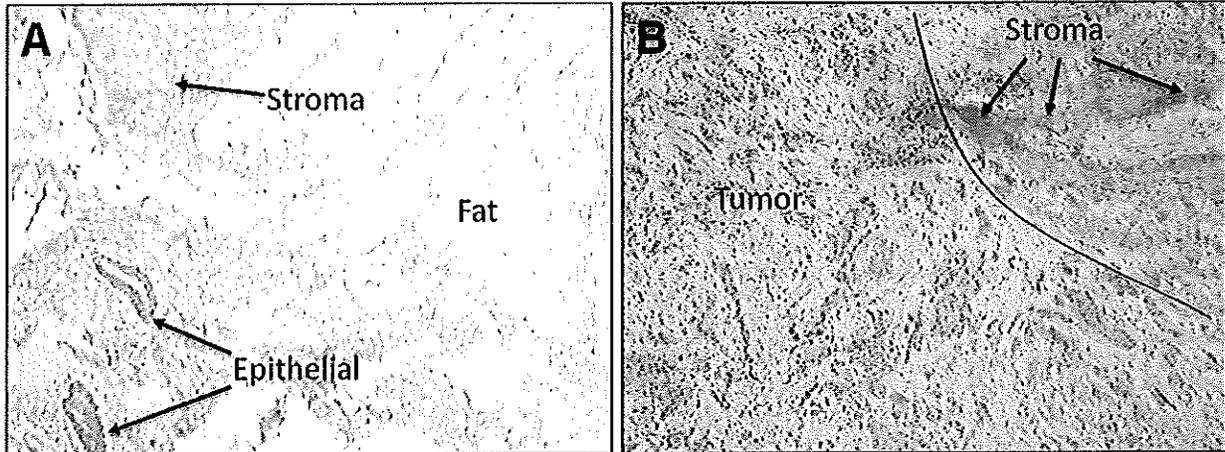
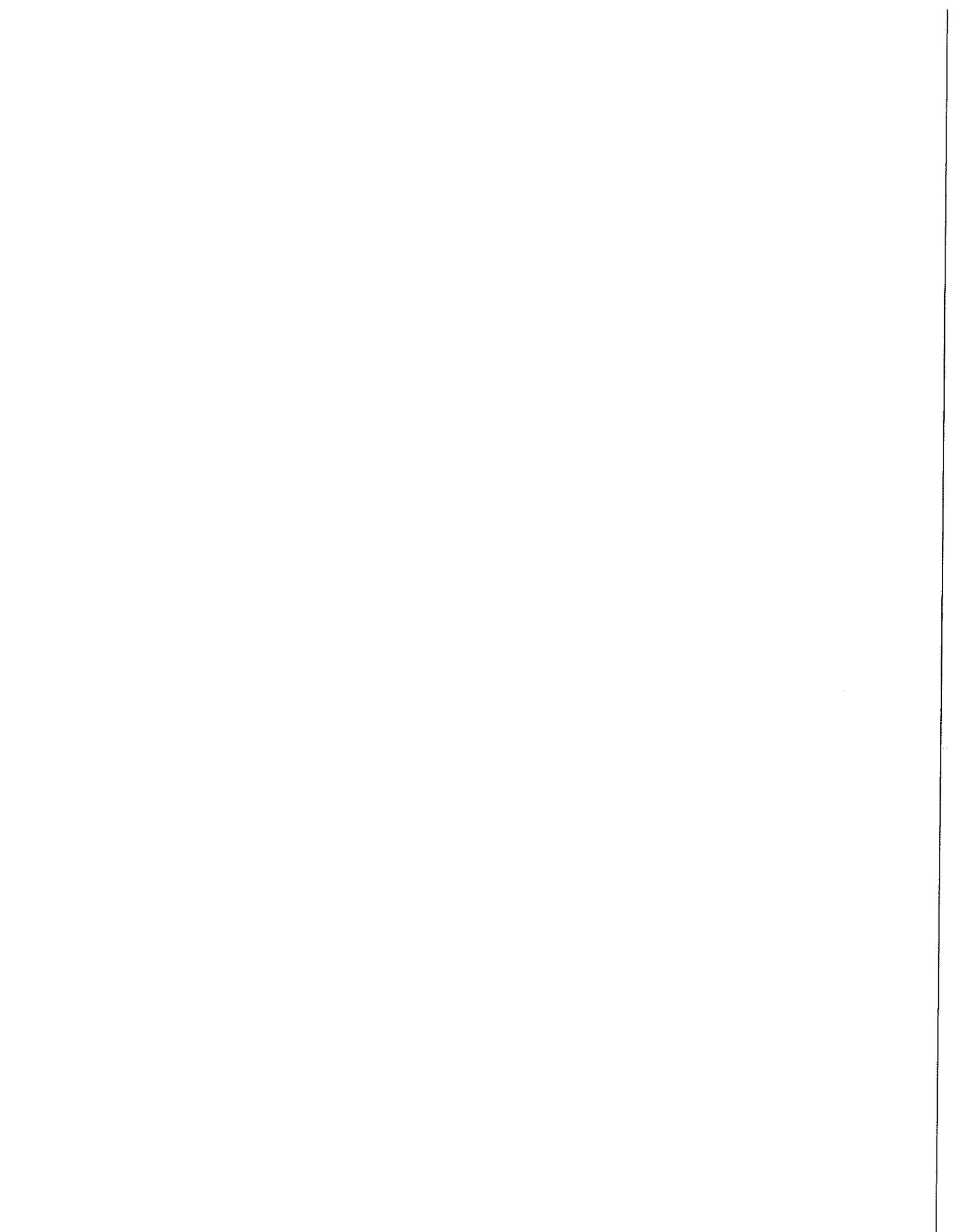


Figure 7.



**Supplementary Figure 1.** Heterogeneity of breast tissue. Panels A and B illustrate the typical heterogeneity observed in histological images of normal (A) and cancerous (B) breast tissue. Epithelial, stromal, and fat components have been labeled. Note the greater cellularity overall of the tumor tissue in Panel B.





## Differential Gene Expression Analysis among Post-menopausal Caucasian Invasive Breast Cancer, Benign and Normal Subjects

Jasmine Saini<sup>1</sup>, Xing Li<sup>1</sup>, Leonid Kvecher<sup>1</sup>, Caroline Larson<sup>1</sup>, Daniel Croft<sup>1</sup>, Yaw-Ching Yang<sup>1</sup>, Jeffrey A. Hooke<sup>2</sup>, Craig D. Shriver<sup>2</sup>, Richard J. Mural<sup>1</sup>, and Hai Hu<sup>1</sup>

<sup>1</sup>Windber Research Institute

<sup>2</sup>Clinical Breast Care Project, Walter Reed Army Medical Center, Washington, DC

**Background:** Breast cancer (BC) is the second leading cause of cancer death among women in United States. There have been studies aiming to develop a blood-based BC detection assay, but most of them were focused on comparing blood samples from BC patients to those from normal subjects. Including benign disease patients in the study is critical in developing a BC detection assay. In addition, BC is a heterogeneous disease with distinct characteristics being associated with ethnicity and menopausal status. Thus we designed a study to analyze gene expression, in peripheral blood samples from invasive BC, benign disease and normal subjects, stratified by menopausal status and ethnicity. Here we report the results from Caucasian postmenopausal women.

**Method:** Subjects were selected from the Clinical Breast Care Project (CBCP). Microarray data using Affymetrix GeneChip Human Genome U133 plus 2.0 arrays, of peripheral blood samples from pathologically confirmed invasive BC (n=17) and benign patients (n=17) were compared to those from normal subjects (n=17). All subjects were postmenopausal Caucasian women, matched for age. Using GenespringGX 11.0 software, the data were normalized, and QC on hybridization and internal controls were performed before filtering out probesets with the lowest 20% signal intensity in each array. Asymptotic t-test with FDR correction was used and significant pathways were obtained.

**Results:** Comparison of gene expression in invasive breast cancer patients (mean $\pm$ SD = 63.7 $\pm$ 7.0 years old) vs. normal subjects (mean $\pm$ SD = 63.2 $\pm$ 7.4 years) identified 1102 significantly different genes, satisfying the thresholds of corrected  $p < 0.05$  and fold change (FC)  $> 1.5$ . Of them, 1003 genes were up regulated and 99 genes being down regulated in BC patients. Comparison of gene expression in benign disease women (mean $\pm$ SD = 61.4 $\pm$ 10.4 years) vs. normal women showed 1320 significantly different genes (corrected  $p < 0.05$ , FC  $> 1.5$ ), with 1121 genes being up regulated and 199 genes being down regulated in benign patients. Of the top 10 genes with highest FC values, 6 genes (MBNL1, FAR1, MDM4, ITGA4, RAB8B and EXOC5) were common in both analysis and were up regulated. NOTCH pathway was up regulated in both benign ( $p=0.038$ ) and invasive ( $p=0.042$ ) groups. Similarly, IL-7 pathway was up regulated in both benign ( $p=0.047$ ) and invasive ( $p=0.030$ ) patients. TCR pathway ( $p=7.36E-04$ )

was up regulated in benign group only. The benign vs. invasive breast cancer subjects did not show significant differential expression.

**Discussion:** Our results provide a list of differentially expressed genes that are mostly up regulated in invasive and benign patients vs. normal subjects. NOTCH pathway is involved in cell-cell communication and angiogenesis. IL7 pathways play an important role in immune system response, and it is also associated with cell proliferation and cell survival signaling. However, TCR pathway is highly significant only in benign patients and stimulation of TCR pathway induces a signaling cascade that ultimately results in activation of induced cell death. This activation could be early body response to prevent cancer development in benign subjects. When more microarray analyses are completed for this study, we hope to obtain a better understanding of the possibility of developing a blood-based BC detection assay.

## **A comparative study of the triple-negative and basal-like breast cancer subtypes**

Chunqing Luo<sup>1</sup>, Hai Hu<sup>1</sup>, Leonid Kvecher<sup>1</sup>, Albert J. Kovatich<sup>2</sup>, Jeffrey A. Hooke<sup>3</sup>, Yaqin Chen<sup>1</sup>, Richard J. Mural<sup>1</sup> and Craig D. Shriver<sup>3</sup>

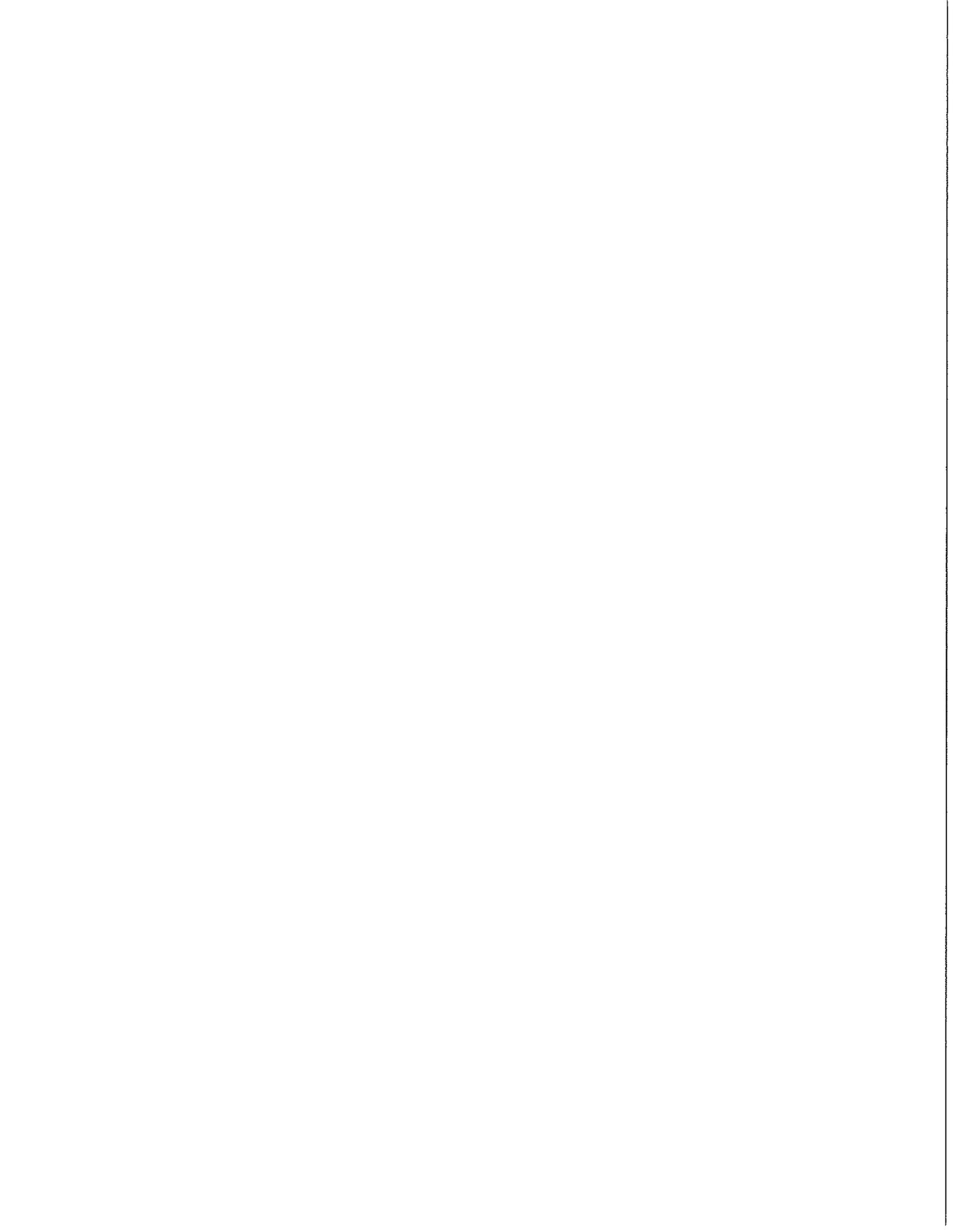
<sup>1</sup>Windber Research Institute, Windber, PA; <sup>2</sup>MDR Global Systems, Windber, PA; <sup>3</sup>Walter Reed National Military Medical Center, Bethesda, MD.

**Background:** Multiple breast cancer subtypes with distinct prognostic differences have been identified based on their molecular characteristics. The Triple-Negative (TN) subtype, identified using clinically-relevant immunohistochemistry (IHC) markers, largely overlaps with the basal-like subtype identified by gene expression profiling. Both have poor prognosis. We performed a preliminary comparative study of the TN/Basal subtypes of breast cancers using molecular data from multiple experimental platforms. A set of breast cancer samples (65) from the Clinical Breast Care Project (CBCP) with comprehensive clinico-pathological and immunophenotypic data were studied, which were also part of The Cancer Genome Atlas (TCGA) breast cancer study.

**Methods:** The 65 breast cancer tissues from the CBCP had IHC results of ER, PR, HER2, Ki67, and p53. The TN subtype was defined as ER-/PR-/HER2-, and the other subtypes were defined according to current practice. Gene expression data of 532 samples (including the 65 CBCP samples) and DNA methylation data of 180 samples (including 25 of the 65 CBCP samples) were downloaded from the TCGA data portal (<http://tcga-data.nci.nih.gov/tcga/tcgaHome2.jsp>). Subtype calls from PAM50 were available from TCGA Breast AWG for 429 of the 532 samples. Clustering analysis was performed using Euclidean distance metric and complete linkage algorithm. Gene expression analysis was performed using Bioconductor “samr” package, for two class unpaired comparison. A p value less than 0.05 and an average fold change of 2 were used as the thresholds for Differentially Expressed Genes (DEGs). The association between gene expression and methylation was calculated using Pearson’s Correlation Test. All statistical analyses were done using R.

**Results:** The 65 CBCP samples were classified into TN (n=12), HER2+ (n=5), LA (n=11), LB (n=31), and undefined (n=6). The 429 TCGA samples were classified as 80 basal-like, 46 Her2+, 187 LA, 109 LB, and 7 normal-like based on gene expression. The IHC-based subclassifications were broadly consistent with the gene-expression based classifications (40/59) with all 12 TN cases called as basal. In analyzing the 65 CBCP samples, unsupervised clustering based on gene expression clustered 9 of the 12 TN into one group, which also included one HER2+ and one LB. Differential gene expression analysis between TN and non-TN yielded 46 DEGs, some of which can well distinguish TN from the other subtypes. Analysis of the 429 TCGA samples (80 basal-like and the remaining 349 samples) yielded 48 DEGs, and 29 of which overlap with the 46 DEGs identified earlier, including ESR1, FOXA1, AGR2, and TFF3. We observed different levels of association between gene expression and methylation for these genes, ranging from the “expected” low methylation/high expression to high methylation/low expression (e.g. TFF3), to completely non-correlated (e.g. ESR1).

**Discussion:** A number of clinically important questions remain unanswered at the molecular level for TN/Basal subtypes. The relationship between the two categories and their further subdivision needs to be addressed. Integrative analysis of data from multiple platforms can enable us to answer these questions and to develop personalized treatment protocols for the molecular subtypes of breast cancers.



## **Development of DNA aptamers that inhibit invasion, migration, and growth of breast cancer cells.**

Rebecca Kopco<sup>1</sup>, Joji Iida<sup>1</sup>, John Lehman<sup>1</sup>, Stella Somiari<sup>1</sup>, Richard Somiari<sup>2</sup>, Richard Mural<sup>1</sup> and Craig Shriver<sup>3</sup>

1: Department of Cell Biology, Windber Research Institute, Windber, PA 15963

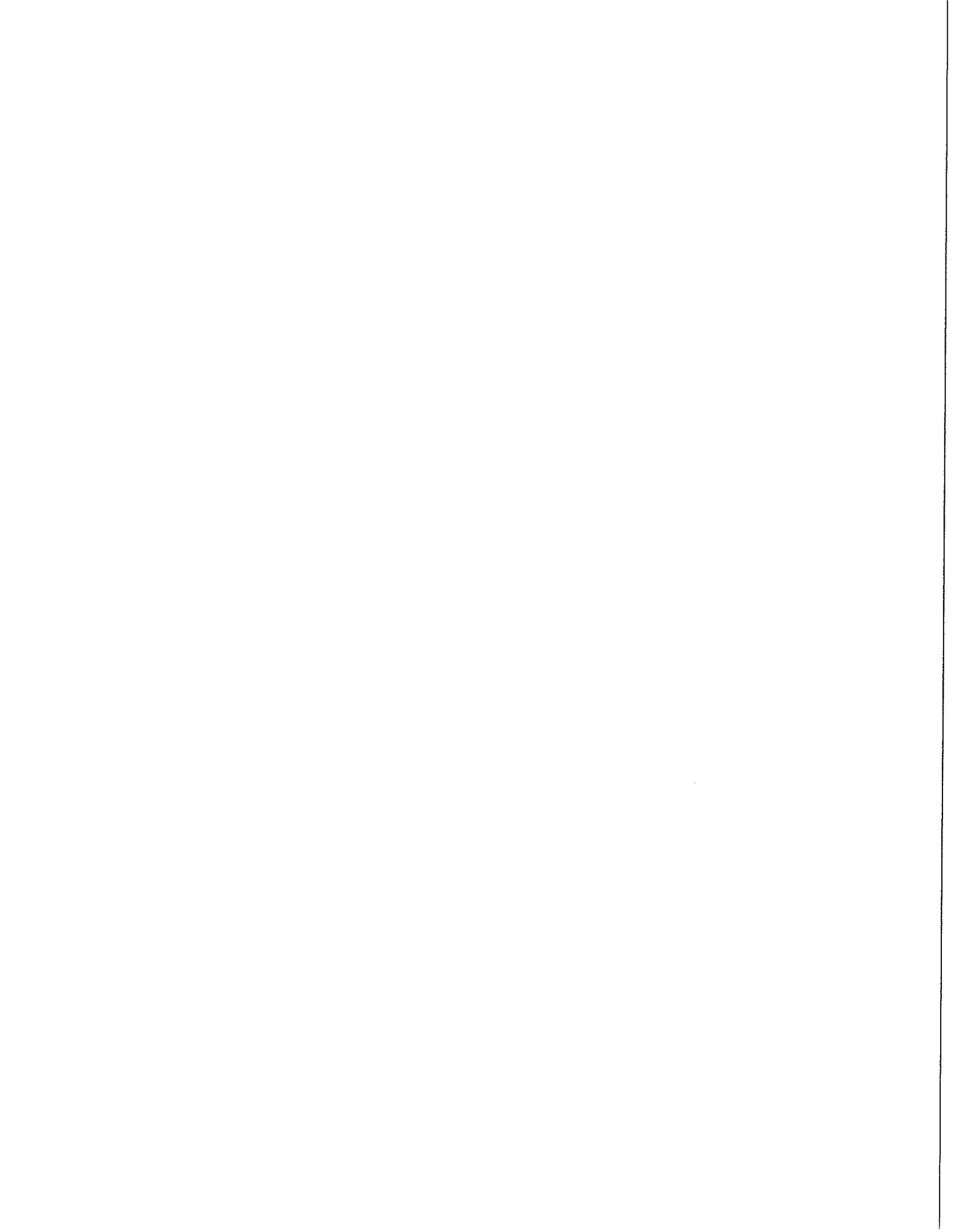
2: ITSI-Biosciences, Johnstown, PA 15901

3: Department of Surgery, Walter-Reed Army Medical Center, Washington, DC 20902

Triple negative (TN) breast cancers are defined by a lack of expression of estrogen, progesterone, and her-2/neu receptors. It is widely recognized that TN breast cancers have a poorer prognosis than other subtypes of breast cancer. Given the lack of effective targeted therapies for TN breast cancer patients, understanding of the mechanisms of growth and invasion of these tumors will provide insight into developing novel approaches to lower the mortality from TN breast cancer.

We utilized three TN breast cancer cell lines (HCC38, HCC1806, HCC1937) as model systems to characterize mechanisms of growth, migration, and invasion. FACS analyses demonstrated that these TN cell lines showed a phenotype of CD44<sup>+</sup>CD24<sup>-/low</sup> epithelial specific antigen (ESA)<sup>+</sup>. An inhibitory antibody against exon v10 of CD44 significantly inhibited three-dimensional (3D) growth of tumor cells. Importantly, this antibody also inhibited alpha2beta1/alpha3beta1 integrins-mediated migration and invasion into matrigel. These results suggest that CD44 harboring exon 10 facilitates progression and metastasis of TN breast cancer cells in vivo by enhancing growth, migration, and invasion. We also demonstrated that this exon interacts with a cell surface tyrosine kinase, EphA2. Thus, we hypothesize that CD44 would form a molecular complex with EphA2 through exon v10 on TN breast cancer cells that plays a key role in facilitating invasion and metastasis.

In order to develop novel reagents to inhibit TN breast cancer invasion and metastasis, we have utilized SELEX (systematic evolution of ligands by exponential enrichment) to isolate DNA aptamers that specifically recognize exon v10 of CD44. DNA aptamers are functional molecules with the appropriate sequence and structure to form a complex with a target molecule. Typically, dissociation constants for these aptamer-target complexes are in the high pico-molar to low nano-molar range, which is comparable to antibodies. Given the fact of the ease of chemical synthesis and modifications of oligonucleotides, DNA aptamers that recognize cancer cells will be easier to develop and have a great impact for the next generation of strategies for cancer detection, diagnosis, and therapy. We describe the development of "function-blocking" DNA aptamers specific for exon v10 of CD44 that inhibits TN breast cancer invasion and metastasis.



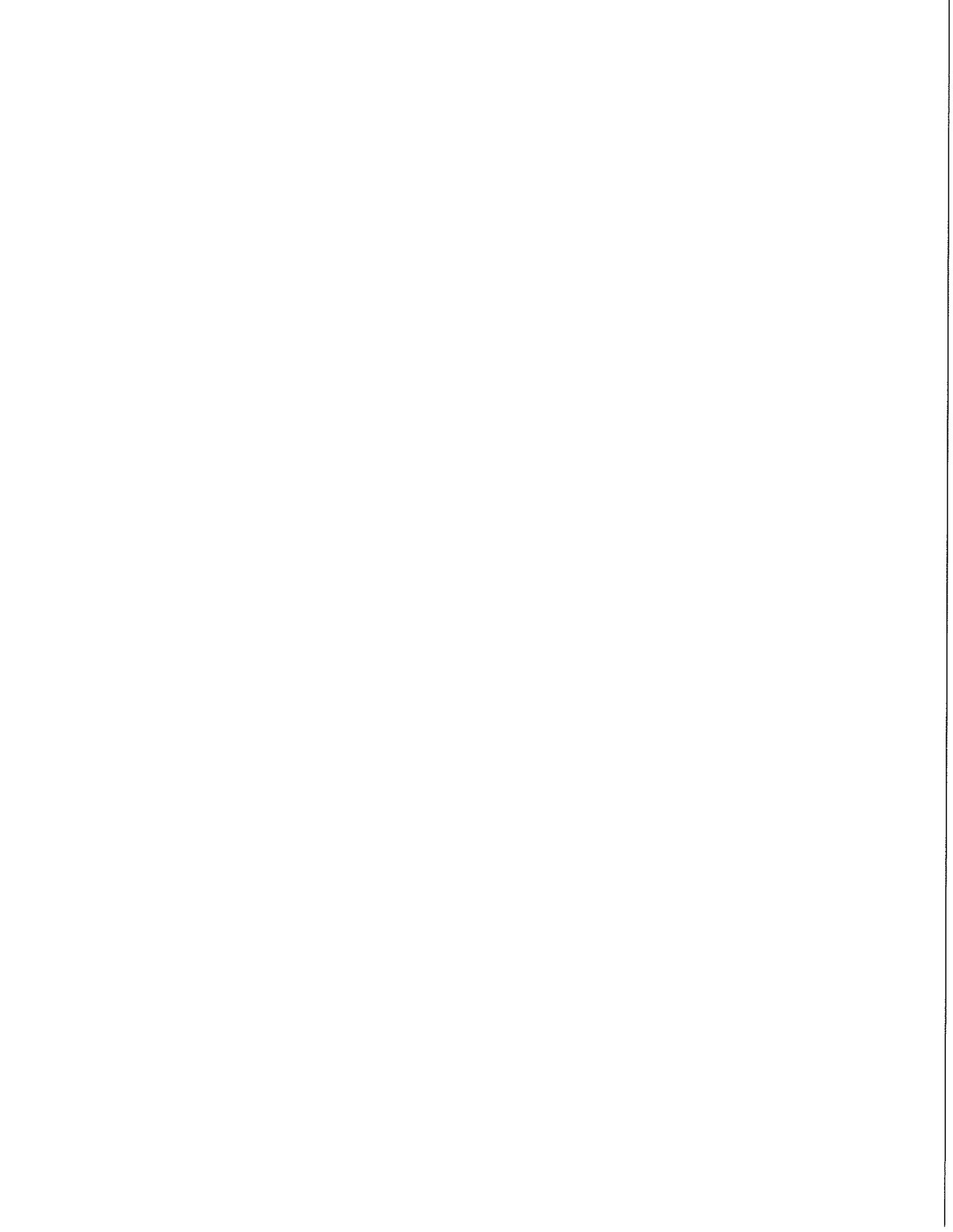
## **The Impact of Matrix Metalloproteinase-1 Promoter 1G/2G Polymorphism on Breast Diseases**

Jing Zhou<sup>1</sup>, Stella Somiari<sup>1</sup>, Susan Lubert<sup>1</sup>, Jasmine Saini, Jeff Hooke, Richard Mural<sup>1</sup>,  
Craig Shriver<sup>1</sup>, Constance Brinckerhoff<sup>2</sup>

<sup>1</sup>Windber Research Institute, Windber, PA

<sup>2</sup>Dartmouth-Hitchcock Medical Center, Lebanon. New Hampshire

Matrix Metalloproteinase-1(MMP-1) is the most ubiquitously expressed interstitial collagenase. Overexpression of MMP-1 has a role in initiating mammary tumorigenesis by both degrading stroma and releasing growth factors and other mitogens for epithelial cells. A single guanine insertion polymorphism in MMP1 promoter region creates a binding site for Ets transcription factor and increases the transcriptional activity of MMP-1 gene. This MMP-1 2G polymorphism has been linked to early onset, increased risk or aggressiveness of several cancers. However, its role in the invasion and metastasis of breast cancer is still controversial. To study the impact of the 2G insertion type polymorphism on breast cancer we analyzed the genotypes of 109 patients (52 invasive breast cancers [IBC], 29 ductal carcinoma in situ [DCIS], 13 atypical ductal hyperplasia [ADH] and 15 benign breast disease). We observed that the 2G/2G genotype frequency was 25%, 28%, 15% and 27% and the 2G allele frequency was 0.43, 0.53, 0.62, and 0.53 for IBC, DCIS, ADH and Benign disease respectively. The 1G/2G genotype frequency was highest within the DCIS (52%), ADH (46%) and Benign disease (53%) groups. Immunohistochemical (IHC) data for MMP-1, HER2, ER/PR and P53 associated with these donors was queried from a data base and this was analyzed with the genotype data. IHC data was scored as either 0 (no expression) or increasing expression of +1, +2 or +3 for MMP-1. We observed that 17% within the IBC group scored 3+ for MMP1 expression while the DCIS, ADH and Benign groups had 15%, 46% and 4% in the 3+ expression range for MMP-1. The expression of MMP-1 was highest (46%; 3+) in the ADH group. Further analyses of HER2, ER/PR and p53 in relation to MMP-1 polymorphism within the invasive breast cancer group showed 2G homozygotes to be higher in the Her-2 positive group (43.5%) than in Her-2 negative group (15.4%) and lower in ER and PR positive groups compared with ER and PR negative groups (21.1% versus 33.3% for ER status and 20.7% versus 28.6% for PR status respectively). The number of 2G homozygotes was also increased in P53 positive group compared with P53 negative group in IBC patients (35% versus 11.8%). We observe that among individuals expressing low levels of MMP1 (1+), 57% have 1G/1G genotype, while among individuals expressing high levels of MMP1 (3+), over 70% are 2G insertion carriers (1G/2G heterozygotes or 2G/2G homozygotes). Therefore, an increasing 2G genotype frequency in IBC group is associated with increasing MMP-1 expression level. The lower 2G allele frequency observed in IBC (0.43) may indicate that the expression of MMP-1 in IBC is controlled not only by this polymorphism but other contributing factors. However, the observation that the ADH group show high MMP-1 expression level (3+) and high 2G allele frequency (0.62) compared to the Benign is consistent with the hypothesis that increased MMP-1 2G polymorphism plays a role in initiation of ADH through up regulation of MMP-1 expression. In conclusion, this pilot study supports the role of MMP-1 2G polymorphism in breast disease onset such as in ADH pathogenesis and it may have a prognostic role in breast disease. These results lay the foundation for a more comprehensive and large scale investigation into the role of MMP-1 polymorphism in breast diseases.



## **Comparison of gene expression profiles of lymph node metastatic and non-metastatic ER+ve breast tumors in pre and post menopausal women.**

Prema B Rapuri<sup>1</sup>, Xing Li<sup>1</sup>, Gerianne Brillhart<sup>1</sup>, Brenda Deyarmin<sup>1</sup>, Leonid Kvecher<sup>1</sup>, Hai Hu<sup>1</sup>, Jeff Hooke<sup>2</sup>, Craig Shriver<sup>2</sup>, Richard Mural<sup>1</sup>

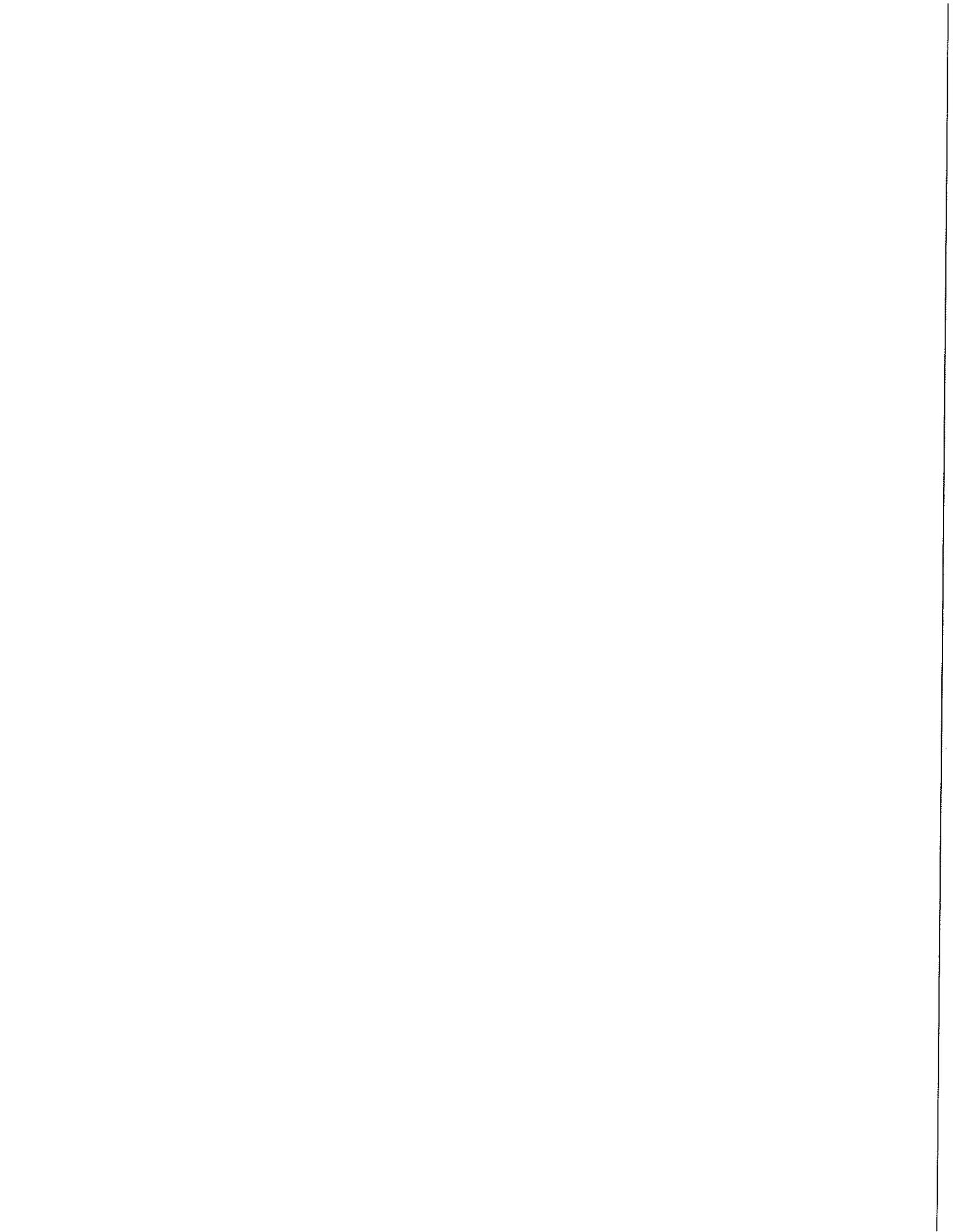
<sup>1</sup>Windber Research Institute, Windber, PA, USA. <sup>2</sup>Walter Reed Army Medical Center, Washington DC, USA.

**Background:** Breast cancer is the most common female cancer in US and is the second leading cause of cancer related death in women. Metastases are the primary cause of cancer morbidity and mortality. Axillary lymph node (LN) status is the most important prognostic factor for breast cancer. The molecular mechanisms that control LN metastasis remains poorly understood. To better understand the various genes and regulatory pathways that drive breast cancer LN metastasis, we compared the gene expression profiles between breast tumors that have metastasized to the LNs and those which have not in the pre and postmenopausal women.

**Material and Methods:** Tumor cells were isolated from the primary tumors (ER+ve) of post menopausal node positive (PMNP; N=12), post menopausal node negative (PMNN; N=12), premenopausal node positive (PRNP; N=12) and premenopausal node negative (PRNN; N=12) women using laser capture microdissection. RNA was isolated using the RNAqueous®-micro kit (Ambion, Austin, TX). Total RNA was converted to Biotin-labelled aRNA using two rounds of amplification with MessageAmp II aRNA amplification kit (Applied Biosystems, Foster City, CA). The aRNA concentration was determined by Nanodrop 1000 and the quality was assessed with a Bioanalyzer. The aRNA was fragmented and hybridized to Human Genome U133 Plus 2.0 GeneChip (Affymetrix, Santa Clara, CA). Microarray raw data were analyzed using a variety of R programming packages for probe density processing, background correction, normalization, quality control/quality assessment, and calculation of gene expression value, etc. To identify differentially expressed genes, Wilcoxon rank sum test with FDR control was performed for pair-wise comparison between different groups. Functional analyses were performed on the identified statistically significant differentially expressed genes to search for the functional categories and pathways in which they are involved and further understand their potential roles in the process of metastasis of breast cancer.

**Results:** Multivariate data mining (hierarchical clustering analysis and principal component analysis, etc) revealed that in postmenopausal women, the node positive and node negative women are well separated, this was not the case in premenopausal women. Further analysis of the PMNN and PMNP groups to identify differentially expressed genes at FDR=0.001 showed that 70 genes were upregulated and 104 genes were downregulated in PMNP vs PMNN groups. Gene function analysis revealed that genes down regulated in the PMNP group compared to PMNN are related to extracellular matrix, cell adhesion, EGF-like pathway, cytoskeleton etc, while the over-expressed genes are related to cell cycle and cell division, condensed chromosome, nuclear lumen, etc.

**Discussion:** The differentially expressed genes between the PMNP and PMNN groups may play a significant role in LN metastasis and may provide prognostic markers as well as potential targets for therapy against early lymphatic metastasis. The probable reason for overlap of gene expression patterns between the PRNP and PRNN groups could be that the complex hormonal milieu in premenopausal women influences the tumorigenesis process differently.



## Ethnicity Difference of Benign Breast Diseases in Breast Cancer and Non-Cancer Patients

Anthony Bekhash<sup>1</sup>, Jasmine Saini<sup>1</sup>, Xing Li<sup>1</sup>, Prema Rapuri<sup>1</sup>, Jeffrey Hooke<sup>2</sup>,  
Albert Kovatich<sup>3</sup>, Richard J. Mural<sup>1</sup>, Craig D. Shriver<sup>2</sup>, Hai Hu<sup>1</sup>

<sup>1</sup>Windber Research Institute, Windber, PA

<sup>2</sup>Walter Reed Army Medical Center, Washington, DC

<sup>3</sup>MDR Global, LLC

**Background:** Breast cancer (BC) is a heterogeneous disease. Ethnicity differences in BC for Caucasian (CA) and African American (AA) women have been reported, but there is no report on ethnicity differences in benign breast diseases (BBDs) although it is known that BBDs may be precursors or risk factors of BCs. In the Clinical Breast Care Project (CBCP), a comprehensive characterization of BBDs is performed on breast biopsies making it possible to conduct a study on the ethnicity difference of BBDs between CA and AA, in the Cancer and Non-Cancer groups respectively.

**Method:** CA and AA patients undergoing a biopsy were selected from CBCP, totaling 1,963. A Pathology Checklist is available, reporting 131 pathologic conditions including 83 BBDs. In addition, a Core Questionnaire covering information such as demographics, medical history, life style, etc. is completed. The Cancer group is composed of 731 CA and 170 AA (including in situ, invasive, and malignant NOS), and the Non-Cancer group is composed of 748 CA and 314 AA. BBDs with a frequency of <1% in each group were removed from the study leaving a total of 26 BBDs analyzed for Cancer group and 25 BBDs for the Non-Cancer group. Pearson's Chi-square test was used to detect statistical significance.

**Results:** The Cancer group showed 6 BBDs significantly associated with the ethnicity; 2 were more frequent in CA, i.e., fat necrosis (5.5% CA vs. 0.6% AA,  $p=0.011$ ), and mild intraductal hyperplasia (6.8% CA vs. 2.4% AA,  $p=0.041$ ). The other 4 BBDs were more frequently observed in AA, which were cysts (56.5% AA vs. 47.2% CA,  $p=0.036$ ), multiple papillomas (15.9% AA vs. 8.3% CA,  $p=0.005$ ), moderate intraductal hyperplasia (MIH) (28.2% AA vs. 18.1% CA,  $p=0.004$ ), and fibroadenomatoid nodule (FN) (11.8% AA vs. 5.3% CA,  $p=0.004$ ). These BBDs were not significantly different in the Non-Cancer group between the two ethnicities.

Six BBDs were significantly associated with the ethnicity in the Non-Cancer group. Four of them were more frequent in CA, including duct ectasia (9.5% CA vs. 3.2% AA,  $p=0.001$ ), microcalcifications (35.3% CA vs. 27.4% AA,  $p=0.015$ ), fibocystic changes (61.0% CA vs. 44.6% AA,  $p=1.30E-06$ ), and sclerosing adenosis (21.7% CA vs. 13.4% AA,  $p=0.002$ ). The other 2 BBDs were more frequent in AA, i.e., sclerosing papilloma (2.6% AA vs. 0.7% CA,  $p=0.025$ ), and

fibroadenoma (29.9% AA vs. 20.9% CA, with  $p=0.002$ ). These BBDs were not significantly different in the Cancer group between the two ethnicities.

**Discussion:** Multiple papillomas and MIH were known moderate risk factors for BC, and in the Cancer group they were more frequently detected in AA than in CA. The significance of the ethnicity difference in the Non-Cancer group is yet to be further explored. It is interesting that not a single BBD was found to be significantly associated with the ethnicity (AA and CA) across the Cancer and the Non-Cancer groups. Thus, the ethnicity difference of BBDs in AA and CA reported here not only suggests possible ethnicity-specific BC risk factors but also generates new hypotheses for future studies.

Genomic instability in the breast microenvironment? A critical evaluation of the evidence

Charvonne Holliday, Seth Rummel, Jeffrey A. Hooke, Craig D. Shriver, Darrell L. Ellsworth, Rachel E. Ellsworth

Charvonne Holliday  
Clinical Breast Care Project  
Windber Research Institute  
620 Seventh Street, Windber, PA 15963  
Phone: (814) 361-6974  
FAX: (814) 467-6334  
Email: [c.holliday@wriwindber.org](mailto:c.holliday@wriwindber.org)

Seth Rummel  
Clinical Breast Care Project  
Windber Research Institute  
620 Seventh Street, Windber, PA 15963  
Phone: (814) 361-6955  
FAX: (814) 467-6334  
Email: [s.rummel@wriwindber.org](mailto:s.rummel@wriwindber.org)

Jeffrey A. Hooke  
Clinical Breast Care Project  
Walter Reed Army Medical Center  
6900 Georgia Avenue, NW  
Washington, DC 20307  
Phone: (202) 782-3459  
FAX: (202) 782-4260  
Email: [jeffrey.hooke1@us.army.mil](mailto:jeffrey.hooke1@us.army.mil)

Craig D. Shriver  
Clinical Breast Care Project  
Walter Reed Army Medical Center  
6900 Georgia Avenue, NW  
Washington, DC 20307  
Phone: (202) 782-0002  
FAX: (202) 782-4260  
Email: [craig.shriver@us.army.mil](mailto:craig.shriver@us.army.mil)

Darrell L. Ellsworth  
Clinical Breast Care Project  
Windber Research Institute  
620 Seventh Street, Windber, PA 15963  
Phone: (814) 361-6911  
FAX: (814) 467-6334  
Email: [d.ellsworth@wriwindber.org](mailto:d.ellsworth@wriwindber.org)

Rachel E. Ellsworth  
Clinical Breast Care Project  
Henry M. Jackson Foundation for the Advancement of Military Medicine  
620 Seventh Street, Windber, PA 15963  
Phone: (814) 361-6926  
FAX: (814) 467-6334  
Email: [r.ellsworth@wriwindber.org](mailto:r.ellsworth@wriwindber.org)

## **SUMMARY**

Tumor-associated stroma plays an active role in breast pathogenesis, with alterations in cell signaling, proliferation and angiogenesis contributing to successful tumorigenesis. Although epigenetic modifications to DNA, as well as changes in RNA and protein expression changes have been identified in tumor-associated stroma, there is considerable debate over the presence of chromosomal alterations, detected as LOH/AI events in histologically-normal stromal cells adjacent to the breast carcinomas. Previous studies have detected LOH/AI at varying distances from the tumor margin and patterns of chromosomal change have been linked to tumor grade and lymph node status. In contrast, recent reports challenge the presence of genomic instability in stromal cells of the breast tumor microenvironment, leading to the speculation that the LOH/AI studies, based almost exclusively on microsatellite-based profiling of formalin-fixed, paraffin-embedded tissues, do not accurately measure the true genomic content of the tumor microenvironment but rather are a reflection of technological artifact. In this review, we present evidence surrounding the debate and critically evaluate arguments – both pro and con- over the presence or absence of genomic instability tumor-associated stroma.

**KEYWORDS**

Breast tumor microenvironment

Microsatellite

Loss of heterozygosity/allelic imbalance

Copy number change

Formalin-fixed paraffin-embedded (FFPE)

Stroma

Breast cancer

## INTRODUCTION

Invasive breast cancer is the most common cancer in women in Western countries and is second only to lung cancer in its lethality. The incidence of breast cancer in the United States has risen by approximately 1.2% per year since 1930 {Bowcock 1999 50 /id}; today, one in eight American women will develop breast cancer during her lifetime. In 2008, it is predicted that 182,460 women will be diagnosed with breast cancer in the United States alone {American Cancer Society 2008 181 /id}. Research aimed at understanding the molecular nature of breast cancer and its progression will have a tremendous impact, not only on the costs associated with breast cancer, but on the nature of diagnosis, treatment and prevention.

Research in the past decade has increased our understanding of how the tumor microenvironment influences tumor development. Breast stroma, which comprises 80% of the normal breast and encompasses fibroblasts, endothelial, smooth muscle, inflammatory and nerve cells, adipocytes, and macromolecules of the extracellular matrix (ECM), serves a supportive role to epithelial cells {Shekhar, Pauley, et al. 2003 633 /id}. A number of physiological changes have been detected in stroma surrounding epithelial neoplasms. For example, changes in fibroblasts in stroma adjacent to breast tumors include altered migratory potential, as well as changes in expression of a number of growth factors {Chiquet-Ehrissmann, Mackie, et al. 1986 634 /id} {Singer, Rasmussen, et al. 1995 635 /id} {Yee, Rosen, et al. 1991 636 /id}; the ability of fibroblasts to produce estrogen suggests that fibroblasts drive epithelial expansion during the early stages of breast cancer {Shekhar, Pauley, et al. 2003 633 /id}. Additional alterations include

changes in cellular polarity, degradation of the ECM and increased angiogenesis {Boudreau & Myers 2003 637 /id} {Shekhar, Pauley, et al. 2003 633 /id}.

Over 10 years ago, chromosomal alterations were detected in morphologically normal TDLUs adjacent to breast carcinomas {Deng, Lu, et al. 1996 97 /id}; a number of additional studies have also detected genomic instability in tumor-associated stroma, furthering the notion that the tumor microenvironment is a dynamic entity that differs from and contributes to tumorigenesis through morphological and molecular alterations {Ellsworth, Ellsworth, et al. 2004 425 /id} {Ellsworth, Ellsworth, et al. 2004 447 /id} {Fukino, Shen, et al. 2004 646 /id} {Fukino, Shen, et al. 2007 645 /id} {Kurose, Hoshaw-Woodard, et al. 2001 44 /id} {Moinfar, Man, et al. 2000 638 /id} {Weber, Shen, et al. 2006 924 /id}. In contrast, recent studies using DNA from fresh frozen tissues to failed to detect chromosomal changes in stromal cells has led some investigators to speculate that the LOH/AI studies do not accurately reflect the chromosomal content of cells of the tumor associated stroma but rather are the result of technical artifacts resulting from the use of microsatellite assays and DNA from FFPE specimens. Here, we present and evaluate data from both sides of the current debate and ponder the question of whether the presence or absence of genomic instability in tumor-associated breast stroma is clinically relevant.

### **Breast Tumor Etiology**

Breast cancer is a complex disease with both environmental and genetic factors contributing to overall risk. Hereditary breast cancer has been associated with mutations in the *BRCA1* and *BRCA2* genes, which control homologous recombination and repair of

somatic damage to DNA and regulate transcription of genes in the DNA damage response pathway {Tan, Marchio, et al. 2008 907 /id}. Efforts to identify genetic risk factors for sporadic cancers have identified five loci, including FGFR2, TNRC9, MAP3K1, LSP1 and a region of chromosome 8q24 {Easton, Pooley, et al. 2007 909 /id}. Additional risk factors for the development of breast cancer include increasing age, family history, reproductive history, environmental exposures and lifestyle choices {Brewster & Bondy 2005 908 /id}.

Breast carcinomas have been well-characterized at the DNA, RNA and protein levels. In addition to heritable mutations in genes such as BRCA1 and BRCA2, tumor-specific mutations in genes such as p53, epigenetic changes and frequent alterations of chromosomes 16q, 17p and 17q have been identified {Miller, Wang, et al. 2003 449 /id} {Petitjean, Achatz, et al. 2007 910 /id} {Agrawal, Murphy, et al. 2007 911 /id}. Identification of patterns of gene expression are now being used clinically to identify patients at risk of recurrence, stratify tumors into histological grade and classify tumors into one of five intrinsic subtypes {Paik, Shak, et al. 2004 540 /id} {Perou, Sorlie, et al. 2000 334 /id} {Sotiriou, Wirapati, et al. 2006 546 /id} {van't Veer, Dai, et al. 2002 335 /id}. Protein assays measuring estrogen and progesterone receptor (ER and PR) and HER2 are now part of routine pathological evaluation while measuring the expression of other proteins such as Ki-67 and TP53 can improve risk assessment of patients with breast cancer {Nicolini, Carpi, et al. 2006 912 /id}.

Defining molecular changes that occur in breast tumors has improved our understanding of tumor development and progression as well as an ability to diagnose and treat women with breast cancer; however, tumorigenesis is a complex process that is

influenced by molecular alterations in tumor epithelial cells and by cell-cell communication with the surrounding stroma. Normal breast stroma is histologically heterogeneous, containing functionally diverse cells such as, which encompasses fibroblasts, endothelial, smooth muscle, inflammatory and nerve cells, adipocytes, and macromolecules of the extracellular matrix (ECM), undergoes a series of phenotypic and functional changes to become activated, providing a permissive environment for the advancement of tumor epithelial cells {Cukierman 2004 919 /id} (Figure 1). For example, in normal stroma, fibroblasts contribute to stability of the ECM, however, once activated these cancer-associated fibroblasts (CAFs) produce growth factors and chemokines, leading to proliferation and progression of tumor epithelial cells {Li, Fan, et al. 2007 921 /id}. Similarly, tumor-associated macrophages (TAMs) facilitate tumor cell migration, angiogenesis and ECM remodeling {Pollard 2008 920 /id}. Finally, increased endothelial cell proliferation and directional migration lead to increased vasculature, providing nutrients and oxygen to tumor epithelial cells {Tlsty & Coussens 2006 922 /id}. Because breast stroma plays such an active and important role in tumorigenesis, increased understanding of the molecular changes that activate tumor stroma is crucial to improving diagnosis and treatment of breast disease.

### **Use of Microsatellite Markers to Identify Chromosomal Alterations**

One of the hallmarks of breast carcinogenesis is global genomic instability, which may include copy number gains or losses of whole or parts of chromosomes. One of the most common methods for measuring chromosomal content is the use of microsatellite markers in loss of heterozygosity (LOH) and allelic imbalance (AI) studies.

Microsatellites, also known as short tandem repeats, are comprised of di-, tri- or tetranucleotides repeated in tandem 5-30 times {Green 2001 109 /id}. Unique DNA sequences flanking each repeat are used to design polymerase chain reaction (PCR) primers to independently assay microsatellites from specific chromosomal regions (Figure 2).

Chromosomal alterations detected using microsatellite markers have been described as LOH or AI events by different authors. Although numerous microsatellite-based studies have used the term LOH to describe chromosomal alterations, LOH is specifically associated with chromosomal deletions, resulting in the conversion of a heterozygous to a homozygous genotype. In contrast, AI describes an imbalance or skewed distribution of one allele, either through amplification or deletion; because microsatellite markers cannot, in fact, distinguish between chromosomal gains and losses, allelic imbalance is the more accurate term to use when performing microsatellite-based studies. AI is often quantified using the formula  $(T1/T2)/(N1/N2)$  where T1 and N1 represent the signal intensities of the less intense alleles and T2 and N2 represent the peak heights of the more intense alleles of tumor and referent samples, respectively {Medintz, Lee, et al. 2000 284 /id}. Currently, no universally accepted criterion is used to define AI; threshold values in the literature vary widely from 20-75% change in tumor allele intensity {Bertheau, Plassa, et al. 2001 101 /id} {Medintz, Lee, et al. 2000 284 /id} {Shen, Yu, et al. 2000 288 /id} {Wang, Huang, et al. 1997 294 /id}.

LOH/AI data has been widely used to characterize breast disease. By 2000, more than 140 publications in the literature described LOH/AI in breast cancer. Meta-analyses of these studies identified 26 unique regions on 16 chromosomes that were frequently

altered in breast carcinomas {Osborne & Hamshere 2000 93 /id}. Similarly, a pooled analysis of 151 studies reporting AI in breast cancer found frequent alterations at ten chromosomal regions, many of which harbored known fragile sites {Miller, Wang, et al. 2003 449 /id}. Using microsatellite data, alterations of a variety of chromosomal regions have been associated with breast tumor characteristics such as histological grade, HER2 status, and lymph node status, providing new insights into tumor development and the complexities underlying breast tumor metastasis {Becker, Ellsworth, et al. 2008 804 /id} {Ellsworth, Ellsworth, et al. 2008 893 /id} {Nagahata, Hirano, et al. 2002 379 /id} {O'Connell, Pekkel, et al. 1998 28 /id} {Roylance, Gorman, et al. 2002 556 /id}.

### **Evidence for Allelic Imbalance/Loss of Heterozygosity in the Breast Tumor**

#### **Microenvironment**

Genomic instability has been well-characterized in breast carcinomas, but until recently, little was known about the frequency of functional importance genomic alterations in non-diseased tissue of the breast. The first high-profile study reporting chromosomal changes in the breast tumor microenvironment used microsatellite markers to assess LOH/AI in DNA isolated from tumor epithelial cells and adjacent normal terminal ductal-lobular units (TDLUs) after manual microdissection. When examining four chromosomal regions showing frequent alterations in breast cancer, 8/30 TDLUs showed chromosomal imbalance, (Table 1); none of the four TDLUs distant to the tumor had detectable LOH/AI {Deng, Lu, et al. 1996 97 /id}. Based on these data, the authors suggest that the definition of molecular margins may be more important than surgical margins in preventing subsequent recurrence.

Another study evaluating molecular changes in 11 manually microdissected areas of breast stroma both close to and distant from *in situ* or invasive carcinomas used 12 microsatellite markers from five chromosomal regions frequently altered in breast cancer {Moinfar, Man, et al. 2000 638 /id}. The reported frequency of LOH/AI ranged from 10-75% in stroma close to disease and 11-57% in stroma at least 15-mm from disease. In contrast, ten reductive mammoplasty samples with no evidence of disease showed no LOH/AI events in either epithelial or stromal cells. Based on these observations, that authors suggested that chromosomal alterations are frequent in stroma not only close to but distant from malignancies, furthermore, molecular changes in stroma may precede those in epithelial cells, possible facilitating tumor invasion.

The development of new technologies for tissue dissection has allowed molecular alterations to be characterized in specific cell types of the heterogeneous breast environment. For example, laser-microdissection (LCM) is useful for isolating relatively pure populations of specific cell types and has become widely used in genomic studies of breast disease. One of the first studies to use LCM to assess LOH/AI in tumor epithelial and nearby stromal cells found higher levels of alterations in tumor epithelial cells compared to stroma, although imbalance at chromosome 2q34 was significantly higher in stromal (61%) compared to epithelial (48%) cells {Kurose, Hoshaw-Woodard, et al. 2001 44 /id}. This same research group subsequently examined tumor epithelial and adjacent stromal fibroblasts from 134 sporadic cancers using a genome-wide panel of 381 microsatellite markers and again reported specific alterations at higher frequencies in stroma compared to epithelial cells. In addition, a number of LOH/AI hotspots where frequent alterations occurred were observed more frequently in stroma than epithelial

cells, leading the authors to conclude that while a limited number of genetic alterations in epithelial cells are sufficient for malignant transformation, it is the large array of hotspots in the stromal cells that influence variability among patients in tumor biology, response to treatment, and outcome {Fukino, Shen, et al. 2004 646 /id}. When these results were compared to patterns of LOH/AI in stromal fibroblasts from BRCA1 and BRCA2 carriers, levels of LOH/AI in the stroma from BRCA1/2 carriers were higher than those from sporadic cases and, in fact, levels of changes were similar in the tumor epithelium (66.2%) and adjacent stroma (59.7%) in hereditary breast cancer patients. The authors thus proposed that BRCA1/2 mutations may lead to genomic instability in the stroma, promoting neoplastic transformation {Weber, Shen, et al. 2006 924 /id}.

An ambitious assessment of LOH/AI in stromal and epithelial cells from 220 primary breast tumors using the same genome-wide panel of 384 markers used in earlier studies examined relationships between clinical parameters and genomic instability in both the epithelial and stromal cells. Significant associations between LOH/AI in stromal fibroblasts on chromosome 11 and tumor grade and chromosomes 1, 2, 5, 18, 20 and 22 and lymph node metastasis were observed; only LOH/AI in epithelial cells on chromosome 14 was associated with progesterone status {Fukino, Shen, et al. 2007 645 /id}.

In our laboratory, we have examined genetic changes in both tumor-adjacent and distant stroma. Tumor epithelial and corresponding adjacent stroma <850  $\mu$ m from the tumor were laser microdissected from 30 formalin-fixed, paraffin-embedded tumor specimens. Using a custom panel of microsatellite markers selected from 26 chromosomal regions frequently altered in breast cancer {Ellsworth, Ellsworth, et al.

2003 290 /id}, we observed discordant patterns of LOH/AI between epithelial and stromal components such that hierarchical clustering samples from the same patient rarely clustered together. In addition, chromosome 11p15 was altered significantly more frequently in stromal compared to epithelial cells {Ellsworth, Ellsworth, et al. 2004 425 /id}. In a second study evaluating LOH/AI data generated from archival breast quadrants from 21 patients who underwent mastectomy, levels of LOH/AI were significantly higher ( $P < 0.05$ ) in the tumor-adjacent (15.4%) compared to distant (3.7%) tissues, suggesting that chromosomal alterations in stroma close to the tumor may promote tumorigenesis {Ellsworth, Ellsworth, et al. 2004 447 /id}.

### **Evidence Against Genetic Changes in the Breast Tumor Microenvironment**

These microsatellite-based AI/LOH studies demonstrating chromosomal alterations in non-diseased breast tumor stroma, both close to and distant from the tumor, suggest a potential role of the microenvironment in breast pathogenesis. While changes in specific chromosomal locations have been associated with different tumor attributes, such as the association between AI/LOH at chromosome 11 and tumor grade {Fukino, Shen, et al. 2007 645 /id}, the underlying genes targeted by these chromosomal changes have not been identified.

To improve understanding of the role different cellular types play in the tumor etiology, Allinen et al. used cell type-specific cell surface markers and magnetic beads to purify epithelial, myoepithelial and endothelial cells, leukocytes, myofibroblasts and fibroblasts from frozen normal, ductal carcinoma *in situ* (DCIS) and invasive breast tissues {Allinen, Beroukhim, et al. 2004 903 /id}. Serial analysis of gene expression

(SAGE) data revealed that while in general, gene expression patterns in cell types from different tissue specimens were highly similar, significant differences in gene expression were detected between normal and diseased states for all cell types. Unlike gene expression changes, however, copy number alterations, assayed using both array comparative genomic hybridization (aCGH) and SNP arrays, were detected only in tumor epithelial cells; no chromosomal alterations were detected in stromal cells from diseased tissues or stromal or epithelial cells from normal tissues.

Because differences between normal and tumor non-epithelial cells are maintained long term, suggesting that heritable genetic alterations are involved in these differences, Hu et al. followed up the study by Allinen et al. by performing methylation-specific digital karyotyping on each stromal cell type {Hu, Yao, et al. 2005 904 /id}. Tumor epithelial and stromal cells were found to be hypomethylated, compared to their normal counterparts. Based on these observations, the authors conclude that epigenetic changes rather than chromosomal alterations, in stromal cells may alter the tumor microenvironment and contribute to breast tumorigenesis {Hu & Polyak 2008 905 /id} {Polyak, Haviv, et al. 2008 906 /id}.

In a recent publication by Qiu et al., copy number analysis was performed on tumor epithelial and stromal cells from breast (n=10) and ovarian (n=25) tumors {Qiu, Hu, et al. 2008 913 /id}. DNA from microdissected epithelial and cancer-associated fibroblasts (CAFs) from fresh frozen specimens, as well as short-term cultures of primary breast CAF were analyzed using Human mapping 500K SNP arrays (Affymetrix, Santa Clara, CA) to identify regions of LOH and/or copy number changes. Although a high frequency of LOH/copy number changes was observed in all of the carcinomas, no

chromosomal alterations were detected in any of the breast or 24/25 of the ovarian CAF samples (Figure 3). Because the SNP data failed to identify LOH on chromosome 11 in CAFs, specifically in regions previously described by microsatellite studies as hotspots for LOH in breast stroma {Fukino, Shen, et al. 2007 645 /id}, a set of five chromosome 11-specific microsatellite markers were used to assess LOH/AI in breast CAFs. None of the 12 CAF samples examined demonstrated LOH by microsatellite analysis, whereas 30% of the matched tumor epithelial cells showed detectable LOH {Qiu, Hu, et al. 2008 913 /id}.

## **Chromosomal Alterations in Breast Stroma: Technical and Biological**

### **Considerations**

#### *Artifacts from FFPE usage and microsatellite analysis*

The inability of recent studies {Allinen, Beroukhir, et al. 2004 903 /id} {Hu, Yao, et al. 2005 904 /id} {Qiu, Hu, et al. 2008 913 /id} to detect copy number changes in breast stromal cells has sparked considerable debate about the validity of LOH/AI results in stroma determined by microsatellite analysis. In particular, some investigators contend that such studies are compromised by technical artifacts, caused by 1) use of FFPE samples, 2) highly multiplexed PCR, 3) non-optimization of microsatellites used on DNA from FFPE samples and 4) lack of replicate PCR reactions {Polyak, Haviv, et al. 2008 906 /id}.

FFPE samples represent valuable specimens for cancer research because pathology laboratories routinely process patient samples, both diseased and non-diseased in paraffin for histological diagnosis, FFPE blocks are typically stored  $\geq 10$  years, and

retrospective outcome data is often available. The quality of DNA in FFPE samples is, however, affected by the fixation and embedding processes as well as long term storage. Formalin causes chemical modification and cross-linking of nucleic acids, while high temperatures used during paraffin-embedding can lead to fragmentation of DNA, resulting in low yields of degraded DNA {Srinivasan, Sedmak, et al. 2002 925 /id}. As a consequence, DNA isolated from FFPE specimens is associated with a higher frequency of mutation artifacts and successful amplification of PCR products >300 bp is difficult {Farrand, Jovanovic, et al. 2002 926 /id}{Williams, Ponten, et al. 1999 927 /id}.

Choice of microsatellites is a critical aspect for accurate assessment of genomic alterations in DNA from FFPE specimens. Given the poor quality and low quantity of DNA typically isolated from FFPE specimens after microdissection, the size PCR products containing the microsatellite markers should be <300 bp. In addition, microsatellite markers should have a high heterozygosity value to increase the likelihood that the markers are informative for assessing LOH/AI. Even when these two criteria are fulfilled, not all microsatellite markers are suitable for use with FFPE material. For example, in developing our panel of microsatellite markers used to assess LOH/AI at chromosomal regions frequently altered in breast cancer, 19 (22%) of the 85 microsatellites originally tested failed to robustly amplify DNA from archival samples {Ellsworth, Ellsworth, et al. 2003 290 /id}. Even with a panel of robust microsatellite markers, PCR amplification was most efficient when markers were assayed individually.

When using microsatellite markers to evaluate genomic instability in DNA from FFPE specimens, a number of validation steps must be used including: 1) repeated genotyping to verify LOH/AI results, and 2) confirmation that LOH/AI is not observed

more frequently for the larger allele {Deng, Lu, et al. 1996 97 /id} {Ellsworth, Ellsworth, et al. 2004 425 /id} {Ellsworth, Ellsworth, et al. 2004 447 /id} {Moinfar, Man, et al. 2000 638 /id}. In addition, assessment of LOH/AI in reductive mammoplasty (RM) samples allows for the determination of background LOH/AI changes. For example, in 10 RM cases with no detectable disease, Moinfar et al. found no LOH/AI events in either the epithelial or stromal compartments, Ellsworth et al. found frequencies of <1% in three RM specimens and 1.5% in histologically normal breast specimens {Ellsworth, Ellsworth, et al. 2004 425 /id} {Ellsworth, Ellsworth, et al. 2005 506 /id} {Moinfar, Man, et al. 2000 638 /id}. These data from samples without disease validate the ability to detect true chromosomal alterations in FFPE specimens: if detection of LOH/AI was a technical artifact, levels of imbalance would be similar between diseased- and non-diseased tissues.

In the study by Kurose et al., each of the 13 PCR assays was repeated for every sample. Although normal epithelial and stromal cells were included in this study, these cell types were microdissected from diseased breasts; disease-free specimens were not included so it is not possible to determine the background levels of LOH/AI detected by this laboratory {Kurose, Hoshaw-Woodard, et al. 2001 44 /id}. In later publications, 372-386 microsatellite markers were evaluated using 72 multiplexed assays, representing 2-8 microsatellites each. Fukino et al. performed a number of experiments as "Validation of veracity of LOH with Multiplex-PCR method in archived materials". Subsets of samples from the original study by Kurose et al., plus additional FFPE samples and four frozen samples corresponding to the original archival specimens were evaluated for LOH/AI with either the original 13 or additional nearby microsatellite markers. In addition, a second technology, real-time LOH analysis, was used {Fukino, Shen, et al. 2004 646

/id}{Kurose, Hoshaw-Woodard, et al. 2001 44 /id}. Although these experiments do validate the results from the Kurose et al. study, it is not clear how these experiments demonstrate the ability to accurately detect LOH/AI events when microsatellites are multiplexed. Given our finding that 22% of markers tested were not suitable for genotyping DNA from laser-microdissected, FFPE samples, it is possible that ~80/380 microsatellites would not provide reliable results as single reactions, much less without the added complication of performing multiplex PCR on low-quality DNA. For example, of the 102,108 possible PCR reactions in the Fukino et al. study, 43,839 informative genotypes were generated {Fukino, Shen, et al. 2004 646 /id}. It would be helpful to know of the 57% genotypes that were not informative, how many were homozygous (and thus not informative) versus how many PCR reactions failed. It is possible that highly multiplexed PCR reactions may be a source of artifact in LOH/AI studies.

Despite the potential artifacts, microsatellite analysis of DNA from FFPE samples has been used successfully for many years in cancer research. Genetic alterations identified by LOH/AI analysis using DNA from FFPE primary breast tumors, including alterations on chromosomes 6q, 8p, 8q, 11q, 13q, 16q, 17p, 17q, and 18q, have also been identified in fresh frozen samples using CGH {Osborne & Hamshere 2000 93 /id}{Tirkkonen, Tanner, et al. 1998 870 /id}. Characterization of DNA from fresh frozen primary breast tumors using CGH revealed that low-grade disease is associated with an overall low frequency of alterations and frequent loss of chromosome 16q while high-grade tumors have frequent genetic alterations including those on chromosomes 8p, 8q, 11q, 13q 17q {Roylance, Gorman, et al. 1999 557 /id}; similar results were later found

using LOH/AI analysis of 185 FFPE primary breast tumors {Ellsworth, Hooke, et al. 2008 577 /id}.

The successful use of LOH/AI data from FFPE samples counters the argument made by Polyak et al. that the published findings in stroma reflect a technological artifact, rather than true genomic instability {Polyak, Haviv, et al. 2008 906 /id}. Should the LOH/AI findings reported by the Eng laboratory, among others, reflect poor quality of DNA rather than true chromosomal alterations, LOH/AI generated in a similar fashion from tumor epithelial cells should be equally affected as factors associated with formalin fixation, paraffin-embedding and long-term storage would affect the entire tissue specimen. As DNA from FFPE breast specimens is now being successfully used to successfully identify copy number changes by CGH and SNP approaches {Joosse, van Beers, et al. 2007 928 /id} {Tuefferd, De Bondt, et al. 2008 929 /id} {Jacobs, Thompson, et al. 2007 930 /id}, other issues must be considered to explain differences between LOH/AI assessments of stroma DNA from FFPE samples and molecular data collected from frozen stromal tissues.

#### *Defining tumor-associated stroma and field effects*

The definition of tumor-associated or tumor-adjacent stroma is highly variable (Figure 4). Most of the LOH/AI studies reported bands of stromal cells ranging from <850  $\mu\text{m}$  from the tumor to <0.5 cm distant were assessed, but not all publications specified distance from the tumor, using definitions such as “close” or “immediately adjacent” (Table 1). When defining genomic alterations, distance from the tumor is critical. Given the symbiotic relationship between tumor and microenvironment, stromal cells closest to the tumor epithelial cells may be most likely to incur critical genomic

alterations. In support for a relationship between levels of genomic instability and distance from the tumor, Heaphy et al. assayed telomere content and allelic imbalance in tumor-associated, histologically normal tissues located either 1 cm or 5 cm from the tumor and found decreasing genomic instability with increasing distance from the tumor {Heaphy, Bisoffi, et al. 2006 937 /id}.

Relationships between chromosomal changes and distance from the tumor may be explained by the field cancerization model, proposed in 1953 by Slaughter et al. to describe the presence of histologically abnormal tissues surrounding carcinomas and subsequent local recurrence of oral cancer {Slaughter, Southwick, et al. 1953 631 /id}. Briefly, a stem cell is thought to accumulate the initiating genetic mutations, which are passed on to daughter cells. This patch of cells escapes the normal mechanisms of cell cycle control and expands, becoming a field of altered cells. As the field grows, additional genetic hits accumulate throughout the field, resulting in a variety of subclones, now developing independently from one another. These subclones have the ability to develop into different types of lesions {Braakhuis, Tabor, et al. 2003 632 /id}. Thus, stromal tissues closest to the tumor may be more likely to show genomic changes than more distant tissues.

Due to the heterogeneous histology of the tumor microenvironment, methods by which stromal cells are collected may also influence the ability to detect chromosomal alterations. While manual microdissection allows for the selection of cell types, the frequency of cross-contamination is higher than when using laser microdissection, which has been shown to have a two-fold increase in the ability to detect chromosomal alterations {Bertheau, Plassa, et al. 2001 101 /id}. Given the importance of obtaining

pure cell populations, copy number changes identified after cell-type specific separation should provide the most accurate molecular profiles in stromal cells. Although this approach failed to detect chromosomal changes in various stromal cell populations {Allinen, Beroukhi, et al. 2004 903 /id}, the cell-enrichment procedure used by Allinen et al. may have included normal stromal cells intervening within the tumor stroma and/or selected out those stromal cells that harbor genetic alterations, which would mask the ability to detect copy number changes {Fukino, Shen, et al. 2004 646 /id}. In addition, the cells used by Allinen et al. were collected after tissue dissociation, thus destroying the spatial relationship to the tumor; by mixing cell types from different distances from the tumor, genetic changes may be masked.

#### *Detection of chromosomal alterations*

The type of technology used to may also influence the ability to detect genomic changes. CGH has an advantage over LOH/AI studies in the ability to identify a chromosomal gain or loss as well as the magnitude of copy number change and can provide a global overview of the genome in from a single assay. CGH cannot, however, detect LOH events without a corresponding copy number loss, thus neither deletion followed by reduplication of the remaining allele nor mitotic recombination would be detected using CGH {O'Connell 2003 931 /id}. In addition, the resolution of CGH is limited, thus small chromosomal alterations may not be detected {Fukino, Shen, et al. 2004 646 /id}. Evaluation of genetic changes in breast specimens using Representational Oligonucleotide Microarray Array (ROMA), a high-resolution microarray technology, has shown that while many alterations were simple, involving broad segments of duplication and deletion, 30% were complex, including a "sawtooth" pattern of changes,

characterized by narrow and frequently alternating segments of chromosomal gains and losses {Hicks, Krasnitz, et al. 2006 932 /id}. Microsatellite markers, which generally measure DNA fragments  $\leq 200$  bp, may therefore, be able to detect small changes not detectable by CGH.

High density arrays of single nucleotide polymorphisms (SNPs) recently used to identify copy number changes, may provide an optimal methodology for accurately identifying chromosomal changes in the tumor microenvironment. Commercial arrays are available with hundreds of thousands of SNPs, providing genome-wide coverage using one to two microarrays. Because SNP arrays provide a genotype in conjunction with a signal intensity for each probe, both copy number changes and LOH can be detected {Gardina, Lo, et al. 2008 933 /id}. Using these powerful assays, Qiu et al. were unable to identify copy number changes and LOH in tumor-associated fibroblasts. It must be noted that Qiu et al., despite not detecting LOH events in their fibroblasts using SNP arrays, re-assessed their breast fibroblast samples with those microsatellite markers on chromosome 11 noted by the Eng group to be highly altered in breast stroma; no LOH/AI was detected {Fukino, Shen, et al. 2004 646 /id} {Fukino, Shen, et al. 2007 645 /id} {Qiu, Hu, et al. 2008 913 /id}.

Identification of genetic instability is highly dependent on the parameters used to define chromosomal alterations. As noted in Table 1, a number of thresholds have been used to define LOH/AI, ranging from only 30% to at least 75% of cells having a detectable LOH/AI event {Deng, Lu, et al. 1996 97 /id} {Moinfar, Man, et al. 2000 638 /id}. Although there has been considerable variation among studies in criteria used to define LOH/AI, making it difficult to compare results, studies using even the most

stringent criteria have detected chromosomal alterations in breast stroma. A similar situation exists for arguments array-based SNP data where a number of different programs are available for analysis, and within each platform, different parameters can be used to control stringency for identifying copy number changes. Thus, when using genome-wide SNP arrays, it is important to optimize parameters for defining copy number changes to discriminate true sporadic changes from normal genome copy number variations {Gorringer & Campbell 2008 934 /id}.

#### *Validation of chromosomal alterations*

Regardless of the platform and methods of data analysis used to define copy number changes, results should be validated with a second, independent platform. Methods for validation include fluorescent in situ hybridization, southern blot analysis, and quantitative-PCR. Allinen et al. validated their CGH results using SNP arrays, confirming the genomic changes in tumor epithelial but not stromal cells {Allinen, Beroukhim, et al. 2004 903 /id}. To our knowledge, validation studies using alternate technologies have not been performed for the LOH/AI results generated from FFPE samples. Although there has been speculation about the causative gene(s) in breast tumor-associated stroma, such as the putative role of the ataxia telangiectasia mutated (ATM) gene on chromosome 11q in the initiation of genomic stability in stroma {Fukino, Shen, et al. 2007 645 /id}, the specific genetic lesion underlying these chromosomal changes have not been identified {Allinen, Beroukhim, et al. 2004 903 /id}.

### **Clinical and Biological Ramifications of Genomic Instability in Breast Tumor Stroma**

Tumor-associated stroma is known to be different than normal stroma at the phenotypic and molecular levels. Reactive stroma is easily recognized by its desmoplastic appearance, epigenetic modifications and gene expression profiles differ between tumor-associated and normal stroma, and growth factors are differentially expressed in tumor-associated compared to normal stroma {Allinen, Beroukhi, et al. 2004 903 /id} {Hu, Yao, et al. 2005 904 /id} {Shekhar, Pauley, et al. 2003 633 /id}. These observations have led to the tissue organization field theory, which suggests that carcinogenesis is not driven by genetic mutations. Under this model, the default state of cells is one of proliferation, not quiescence, and careful tissue organization is needed to inhibit proliferation. Loss of tissue architecture in the microenvironment, rather than changes in DNA, drives tumor formation {Soto & Sonnenschein 2004 935 /id}.

Genomic instability in tumor-associated stromal cells may impact both clinical care for patients with breast cancer and our understanding of breast tumor etiology. For example, stroma adjacent to breast carcinomas from patients treated with epirubicin-cyclophosphamide (FEC-100) chemotherapy demonstrated increased and persistent LOH post-treatment, especially those patients who were non-responsive to treatment {Oudin, Bonnetain, et al. 2007 936 /id}. Associations between genomic instability in stroma and response to treatment suggest that patients with genomically-stable stroma may be more likely to benefit and less likely to develop drug resistance than those patients with genetically-modified stroma.

Likewise, the presence or absence of chromosomal changes in stroma may also have important ramifications for defining surgical margins. Li et al. examined the presences of LOH on chromosome 3p in normal TDLUs adjacent to early-stage breast

carcinomas and found a 3.9-5.2 fold increased risk of recurrence in those patients with LOH {Li, Moore, et al. 2002 938 /id}. Heaphy et al. found a high frequency of telomere shortening and/or allelic imbalance in a substantial number of cells of tumor-adjacent, histologically normal tissues {Heaphy, Bisoffi, et al. 2006 937 /id}. Together these data suggest that although morphologically normal, residual tumor-associated stromal tissue harboring genetic alterations may be associated with increased risk of recurrence.

In addition to these clinical ramifications, determining levels and patterns of genomic instability in the tumor microenvironment may improve our understanding of breast tumor etiology. In our study, chromosomal alterations were detected tumor epithelial and stromal components from *in situ* carcinomas, and at levels similar those seen in the epithelial and stromal components of invasive tumors {Ellsworth, Ellsworth, et al. 2004 425 /id}. Continued investigation of chromosomal changes in the stroma of earlier lesions, including atypical hyperplasias and benign disease, may provide additional information about the development genomic instability arises, and may provide insights into tumor development and progression {Ellsworth, Ellsworth, et al. 2004 939 /id}

## **Conclusions**

Alterations in cell signaling, angiogenesis, cell migration and motility and proliferation establish the tumor microenvironment as a dynamic entity contributing to breast tumorigenesis. Epigenetic changes to DNA as well as changes in RNA and protein expression have been detected in breast stroma; however, whether cells comprising the breast tumor-associated stroma demonstrate genomic instability in the form of LOH/AI or copy number changes is currently the subject of considerable debate. Whether tumor-

associated stroma demonstrates chromosomal alterations may impact the successful treatment of breast cancer and our understanding of pathogenesis. Therefore, it is critical for future studies to accurately define, isolate and assay tumor-associated stromal components to effectively assess genomic stability in the breast tumor microenvironment.

### **Expert commentary**

Since the initial report of chromosomal alterations in morphologically normal TDLUs adjacent to breast carcinomas over 10 years ago, a number of publications using microsatellite data have advanced the study of genomic instability in tumor-associated stroma. Combined evidence from multiple laboratories supports the notion that the tumor microenvironment is a dynamic entity that contributes to tumorigenesis through morphological and molecular alterations. In fact, some investigators have suggested that chromosomal alterations in the microenvironment are responsible for inter-patient variability and are associated with various prognostic factors such as tumor grade and lymph node status. In stark contrast, recent studies using CGH and SNP arrays to detect chromosomal changes in stromal cells collected from fresh frozen tissues have led some investigators to speculate that the LOH/AI studies do not accurately reflect the chromosomal content of cells of the tumor-associated stroma but rather are the result of technical artifacts resulting from microsatellite assays and DNA from FFPE specimens.

We believe that care must be taken when evaluating genomic data generated from both FFPE and fresh frozen stromal specimens as results from each methodology can be affected by a number of variables. In our opinion, the most important obstacle that must be overcome for accurate assessment of genomic instability in breast stroma is a

consensus definition of what constitutes tumor-associated stroma. As noted in Figure 4, distance from the tumor margin used to define “adjacent” stroma varies widely in the literature, from less than 1 mm to 0.5 cm. Stromal cells immediately adjacent to an invasive breast carcinoma may harbor different genomic alterations than fibroblasts or other stromal cells more distant from the tumor. Only when agreement is reached over how to define tumor-associated stroma can accurate genomic assessment be performed.

The current debate over whether genomic instability in the breast tumor microenvironment is marked by genomic instability is reflective of the recent evolution of cancer research. Over the last decade, techniques for assessing allelic imbalance have evolved from single PCR-based assays of microsatellite markers to genomic scale arrays that simultaneously detect copy number and LOH events at hundreds of thousands of SNPs in a single assay. Collection of high-quality research grade tissue specimens may supplement, or even ultimately supplant, FFPE specimens in the molecular assessment of tumor biology. Accurate evaluation of the genomic integrity of tumor-associated stroma has important biological as well as clinical ramifications, impacting the definition of molecular versus surgical margins and implicating the stroma as a more efficacious, less likely to develop drug resistance, target for novel therapeutics. Improved understanding of the tumor microenvironment is, therefore, critical to making meaningful advances in health care for women.

### **Five-year view**

Technology for detecting chromosomal alterations has improved dramatically over the last decade, highlighted by advancements in global profiling of the genome and

ability to detect both copy number changes and LOH using large-scale SNP arrays. Recent refinements allow lower-quality DNA from FFPE tissue specimens to be used successfully on array-based platforms for rapid, global assessment of chromosomal content. In conjunction with these technological improvements, the importance of collecting and preserving fresh frozen specimens for research purposes has led to increased efforts to generate high-quality, research-grade tissue banks. We believe that once a consensus definition of what constitutes tumor-associated stroma is developed, these improved technologies and collections of frozen and matched FFPE specimens will allow the current controversy regarding whether or not the breast tumor microenvironment is characterized by genomic instability to be definitively answered.

### **Key Issues**

- The tumor microenvironment is an active, rather than passive, participant in breast tumor pathogenesis
- Histological characterization and molecular studies have identified morphological, epigenetic and protein changes associated with processes such as angiogenesis, cell signaling and cellular proliferation
- Although LOH/AI studies have detected chromosomal alterations in tumor-associated stroma, these studies assayed genetic changes in FFPE specimens which often yield small quantities of poor quality DNA
- Recent experiments utilizing global array-based technologies and DNA from fresh frozen tissues did not identify LOH or copy number changes in stromal cell

populations, creating debate about the true existence of chromosomal alterations in the breast stroma

- Genomic assessment of both FFPE and fresh frozen samples can be affected by a number of variables including parameters and thresholds used to define copy number changes, underestimation of copy number by using mixed cell populations, validation of results, and most importantly, a consensus definition of true tumor-associated stroma
- Determination of the presence or absence of chromosomal alterations in tumor-associated stroma may have a significant impact on the ability to effectively treat and prevent recurrence of breast cancer

## References

Papers of special note have been highlighted as:

\*of interest

\*\*of considerable interest

## Figure Legends

Figure 1. Contribution of various stromal components to tumorigenesis. Depicted in this figure are endothelial, immune and fibroblast cells, adipocytes and extracellular matrix, which can influence tumorigenesis through angiogenesis, cell signaling, proliferation, ECM remodeling and cell migration.

Figure 2. Design and analysis of a microsatellite marker. PCR primers are designed from unique DNA sequences flanking D18S64 on chromosome 18q21. PCR amplification results in ~million-fold increase in copies of D18S64, with resulting fragments sized between 188-208 bp, depending on the number of (CA) repeats on each chromosome. Electrophoresis of the PCR products discriminates heterozygous genotypes from two patients (left and right panels). Left panel: evidence of LOH - genomic (blood) DNA is heterozygous with alleles at 190 and 198 bp (top), while tumor DNA from the same patient was homozygous with a single allele at 198 bp (bottom). The smaller allele (190) has been completely lost in the tumor. Right panel: evidence of AI - the normal genotype is represented by a larger peak at 192 bp and a smaller peak at 198 bp; the pattern is reversed in the tumor. The normalized allele ratio for this tumor specimen is 0.27 (<0.35) indicating an AI event.

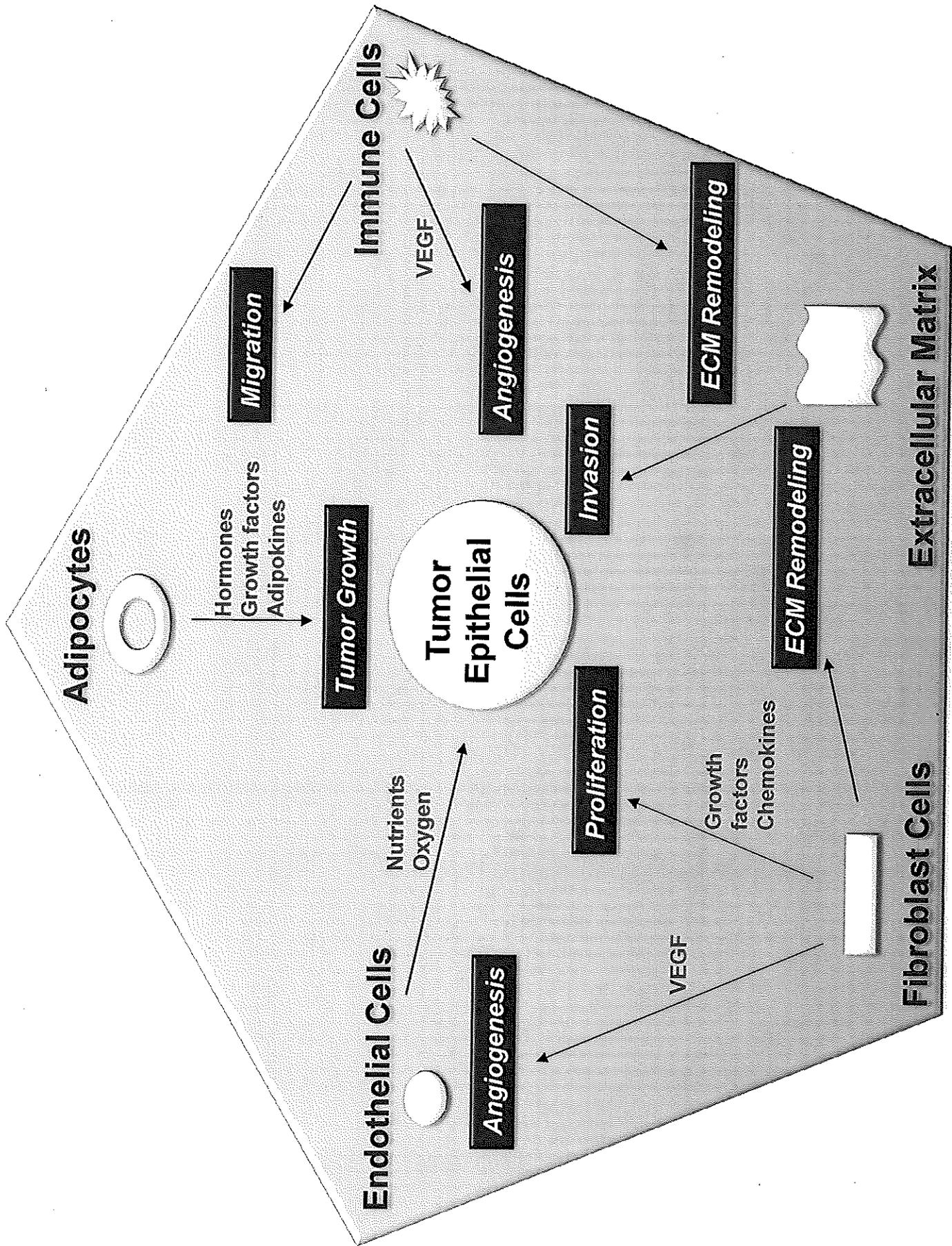
Figure 3. Copy number and LOH events in tumor epithelial and fibroblasts from a medullary breast carcinoma {Qiu, Hu, et al. 2008 913 /id}. Significant changes in copy number are apparent in the epithelial cells (left) while no genetic changes were detected

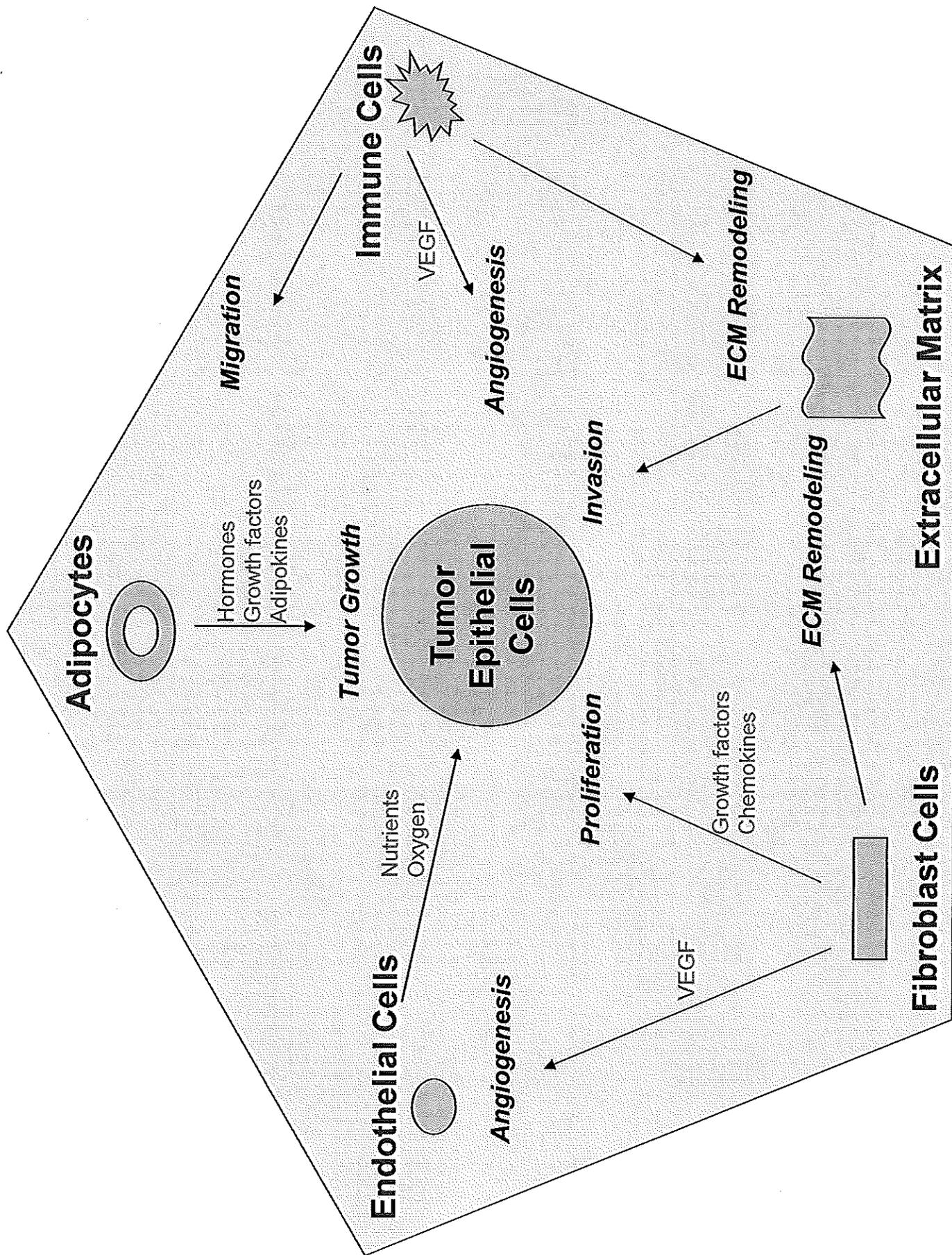
in DNA from fibroblast cells (right). Figure reprinted by permission from Macmillan Publishers Ltd: Nature Genetics, 40(5): 650-655, 2008, <http://www.nature.com/ng/index.html>

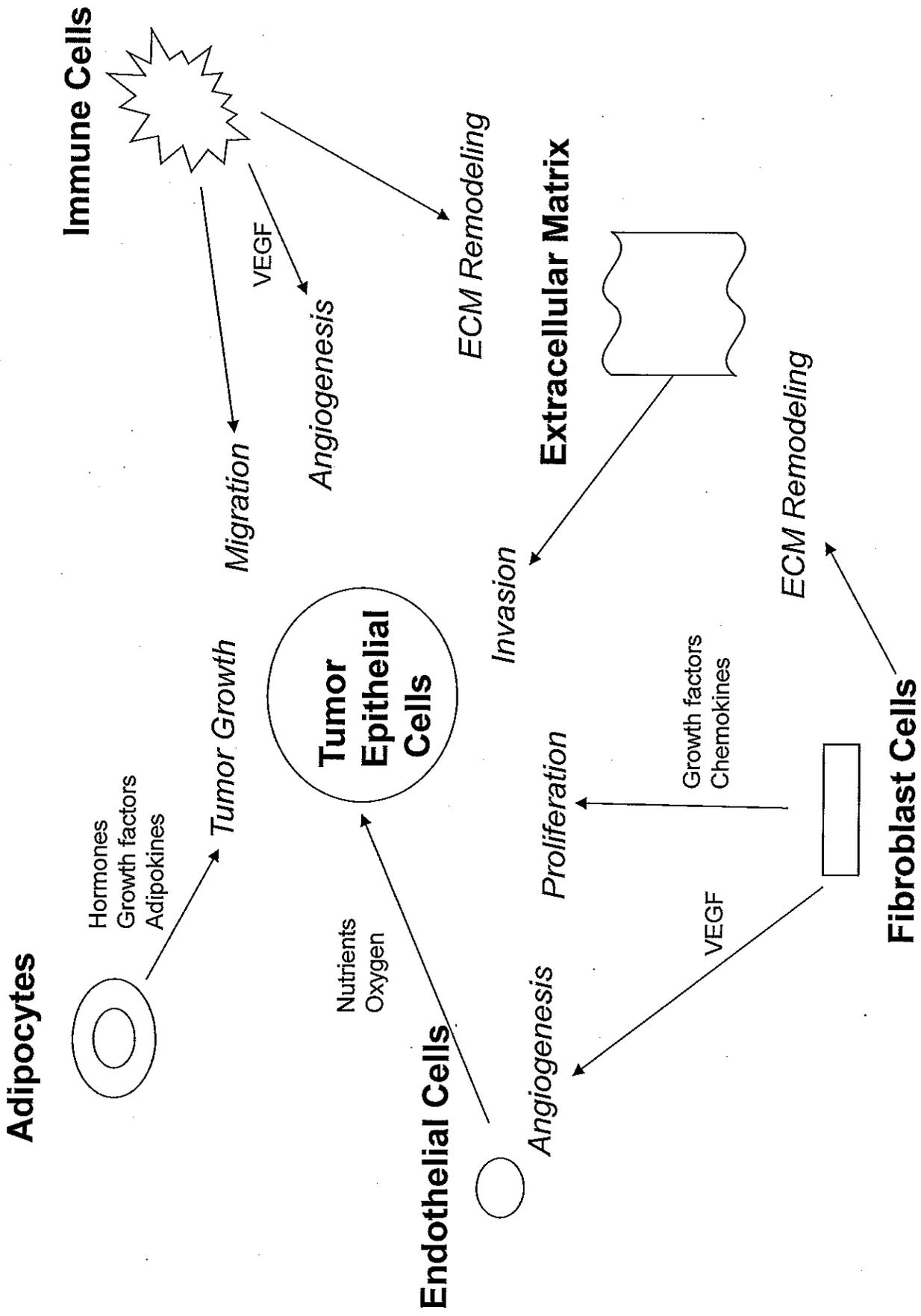
Figure 4. Variability in the definition of tumor-associated stroma. Genomic studies have assessed LOH/AI in tumor-adjacent stroma at various distances from the tumor {Ellsworth, Ellsworth, et al. 2004 939 /id} {Fukino, Shen, et al. 2007 645 /id} {Moinfar, Man, et al. 2000 638 /id} {Weber, Shen, et al. 2006 924 /id}. Of note, the distance of <5cm from the tumor margin utilize by Qiu et al. {Qiu, Hu, et al. 2008 913 /id} is larger than the field of this tumor.

Table 1. Summary of LOH/AI results in stromal cells from archival breast specimens using microsatellite markers.

Cell type(s) examined	Number of samples	Method of isolation	Distance from tumor	LOH/AI criterion	Regions of LOH/AI	Reference
TDLU	30	Manual microdissection	Same section	$\leq 0.70$	3p24	{Deng, Lu, et al. 1996 /id}
Stroma	11	Manual microdissection	Close	$\leq 0.25$	3p14.2, 16q23.1-q24, 17q24,	{Moinfar, Man, et al. 2000 638 /id}
Stroma	11	Manual microdissection	$\leq 15$ mm	$\leq 0.25$	16q23.1-q24, 17q24	{Moinfar, Man, et al. 2000 638 /id}
Stroma	41	LCM	Immediately adjacent	$\leq 0.53$	2q34	{Kurose, Hoshaw-Woodard, et al. 2001 44 /id}
Fibroblasts	134	LCM	Adjacent	$\leq 0.53$	1p35, 6p25, 8q13	{Fukino, Shen, et al. 2004 646 /id}
BRCA1/2 Fibroblasts	51	LCM	$< 0.5$ cm	$\leq 0.66$	1p36, 1p21, 1q42, 7p14, 10p11, 18q21, 18q23, 19q13, 20p13, 20q13	{Weber, Shen, et al. 2006 924 /id}
Fibroblasts	220	LCM	$< 5$ mm	$\leq 0.66$	1, 2, 5, 11, 18, 20, 22	{Fukino, Shen, et al. 2007 645 /id}
Stroma	30	LCM	$< 0.85$ mm	$\leq 0.35$	11p15	{Ellsworth, Ellsworth, et al. 2004 425 /id}
Stroma – distant	21	Whole tissue or LCM	Separate blocks	$\leq 0.35$	8p21-p22	{Ellsworth, Ellsworth, et al. 2004 447 /id}

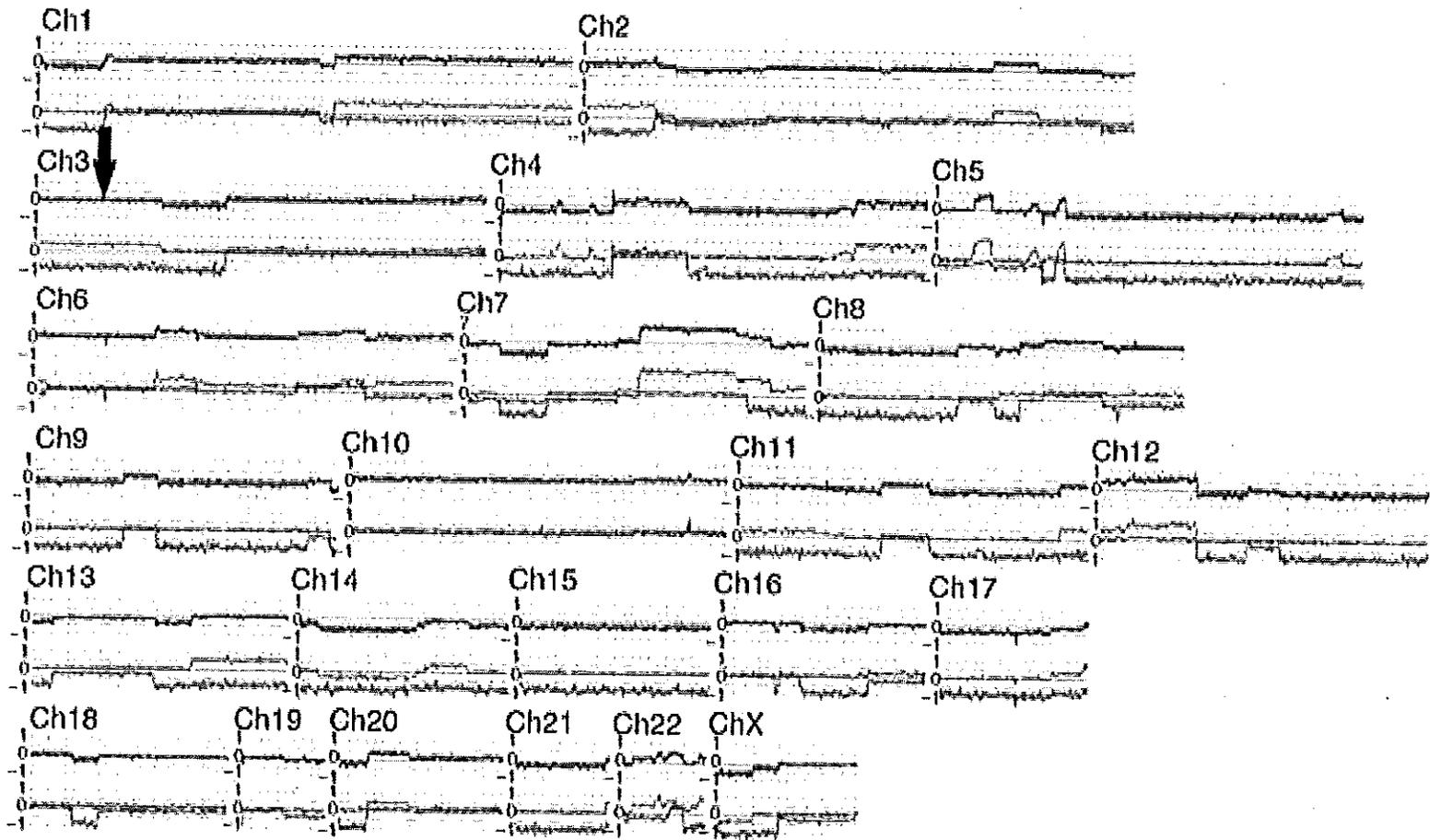




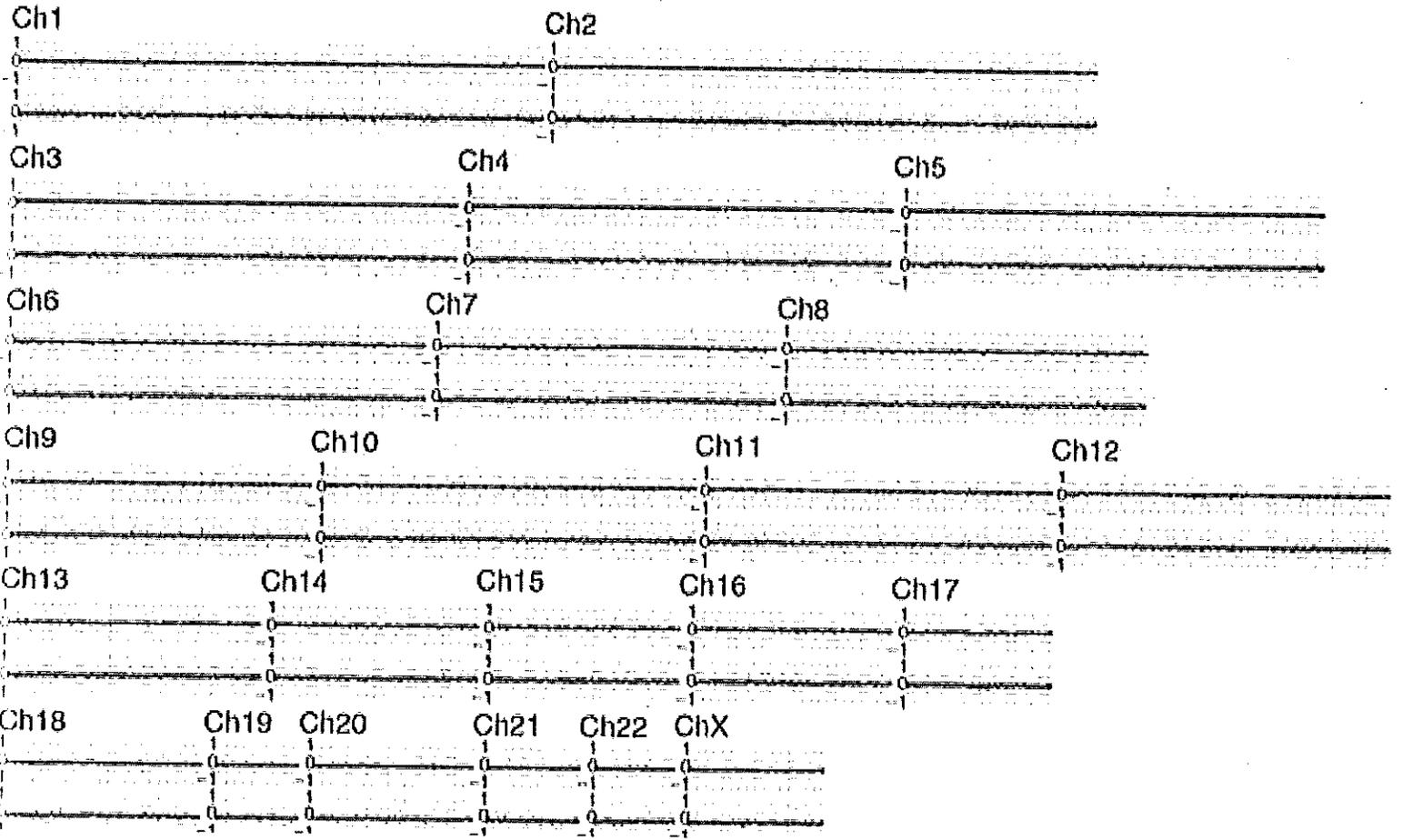


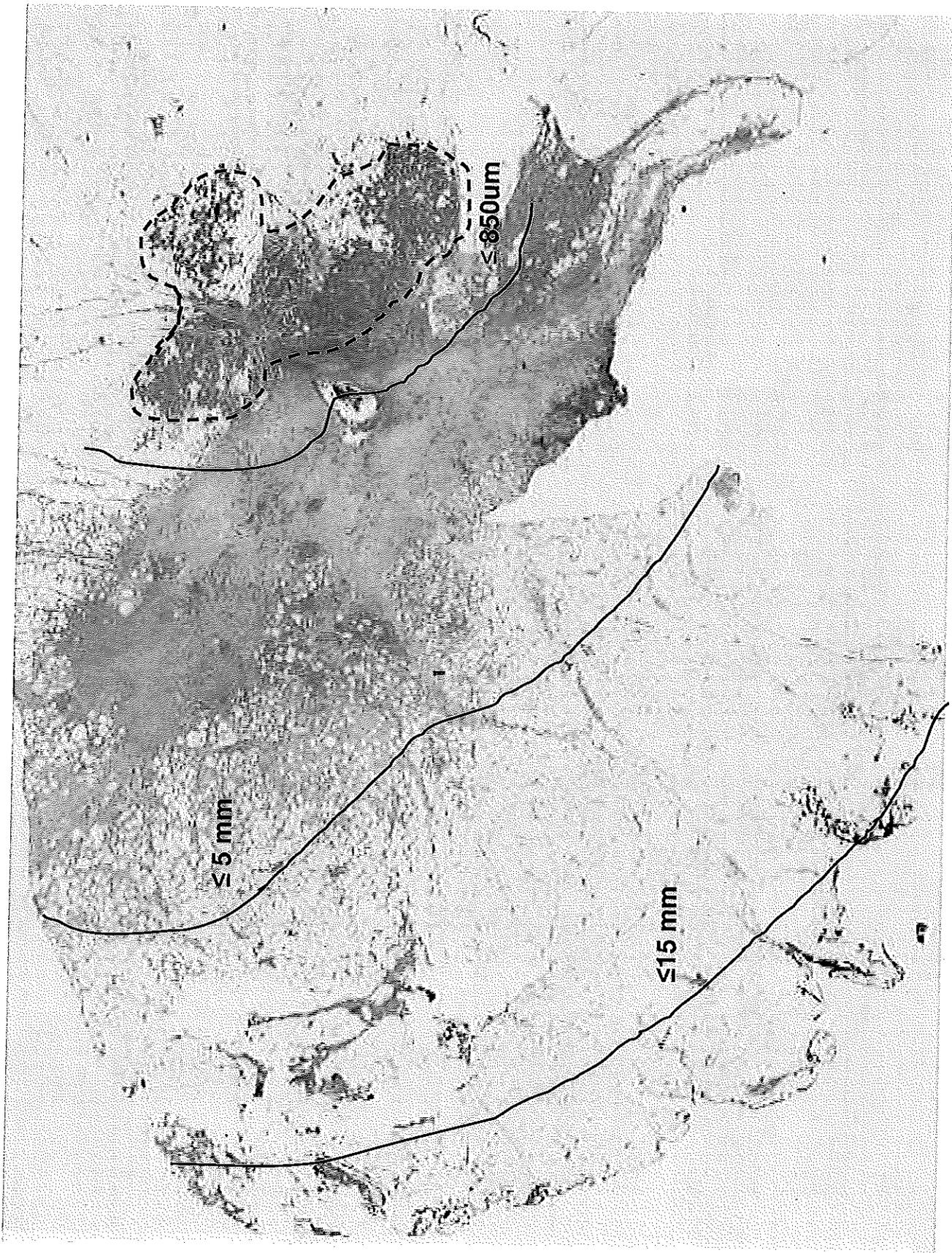


**a** P5077T



**b** P5077F

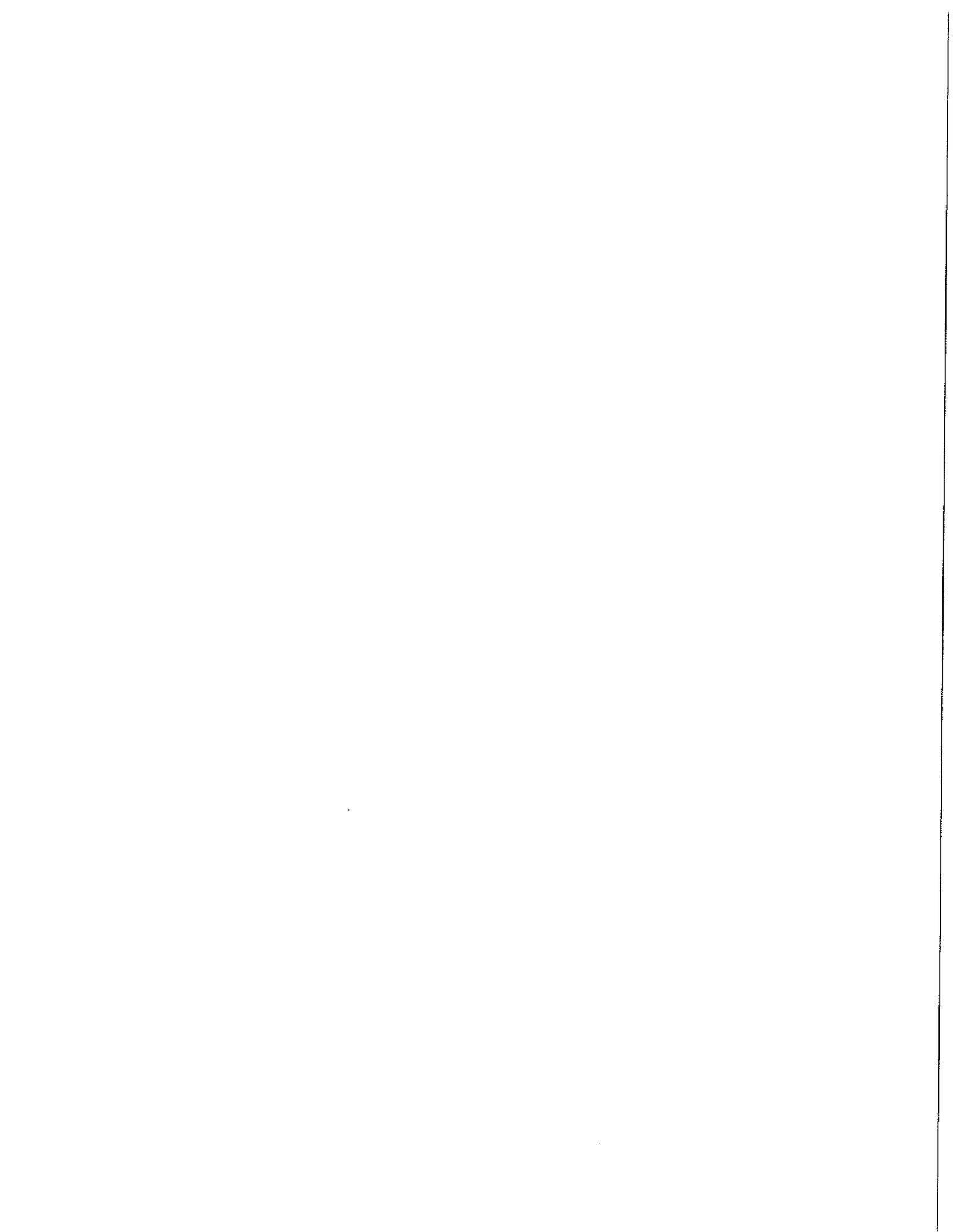




$\leq 850 \mu\text{m}$

$\leq 5 \text{ mm}$

$\leq 15 \text{ mm}$



## Genetic discrimination of aggressive from indolent DCIS

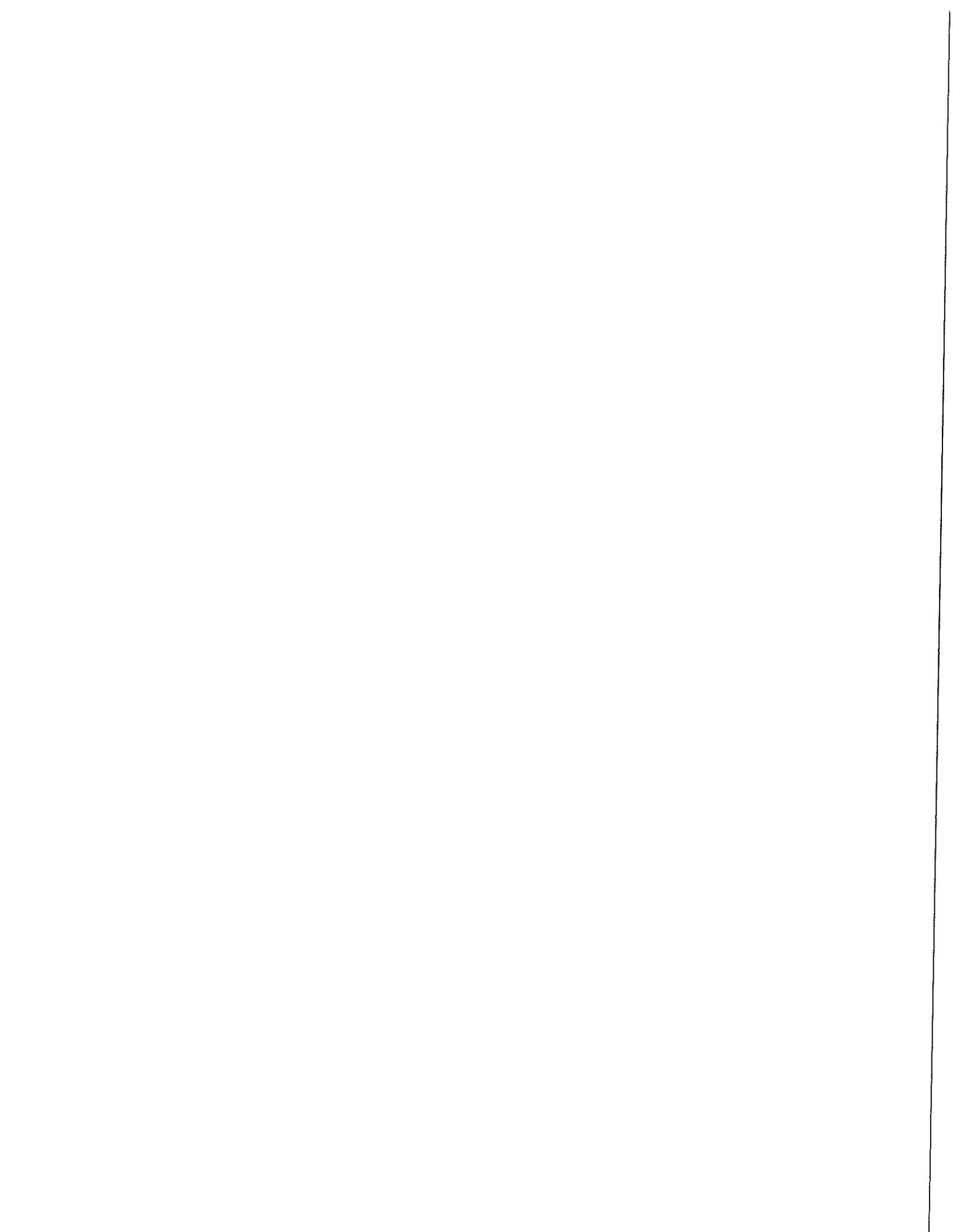
Rachel Ellsworth, Brenda Deyarmin, Heather Patney, Craig Shriver, Kevin Ellison, Derek Thornton, Lorraine Tafra, Zandra Cheng, Martin Rosman

**Background:** Treatment options for DCIS vary widely and include wide-margin surgical excision with or without radiation and/or chemopreventive therapy, or mastectomy. Intuitively, more aggressive treatment options should lead to improved survival rates, however, studies have shown no difference in breast cancer mortality between women treated with wide excision only versus those with excision plus radiation and treatments can be associated with side effects that make follow-up mammography more difficult. To avoid over-treating women with indolent disease, while intensively treating women with aggressive disease, new molecular tools must be developed to supplement pathological information to classify DCIS lesions and predict clinical outcome.

**Methods:** Formalin-fixed paraffin-embedded (FFPE) pure DCIS biopsy specimens were collected from the pathology archives of the Anne Arundel Medical Center, including those with poor prognosis (n=7) and those with  $\geq 5$ -year disease-free survival (n=10). RNA was isolated after laser-microdissection of pure tumor cells and hybridized to Breast Cancer DSA microarrays (Almac Diagnostics). Differentially expressed genes were identified using 2-way ANOVA with Support Vector Machine used to identify genes effective at discriminating good from poor prognosis DCIS. Pathway analysis was performed using MetaCore (GeneGeo).

**Results:** 328 genes were found to be differentially expressed between good and poor prognosis specimens. Further analysis found that a signature of 70 of these 328 genes was effective at discriminating good from poor prognosis DCIS and included genes such as MEF2C, PTK2 and ZBTB2. Pathway analysis revealed that genes involved in cytoskeleton modeling, apoptosis and survival, DNA damage repair and cell adhesion are expressed at lower levels in poor prognosis DCIS while those involved in cell cycle, immune response and cell proliferation are expressed at higher levels.

**Conclusions:** While studies have attempted to identify molecular profiles associated with aggressive DCIS by comparing DCIS co-occurring with invasive disease to pure DCIS, to our knowledge, this is the first study that identified molecular signatures of prognosis in pure DCIS. Although many of the 70 genes found to differ between good and poor prognosis DCIS have not been previously associated with breast cancer or have unknown function, MEF2C and PTK2 have been implicated in invasion and migration, while ZBTB2 is a master regulator of p53 and stimulates cellular proliferation. These data demonstrate aggressive DCIS do differ from indolent DCIS at the genetic level and that these differences may be useful in developing molecular tools to classify DCIS lesions and guide appropriate treatment.



## Genetic signature discriminating metastatic from non-metastatic small tumors

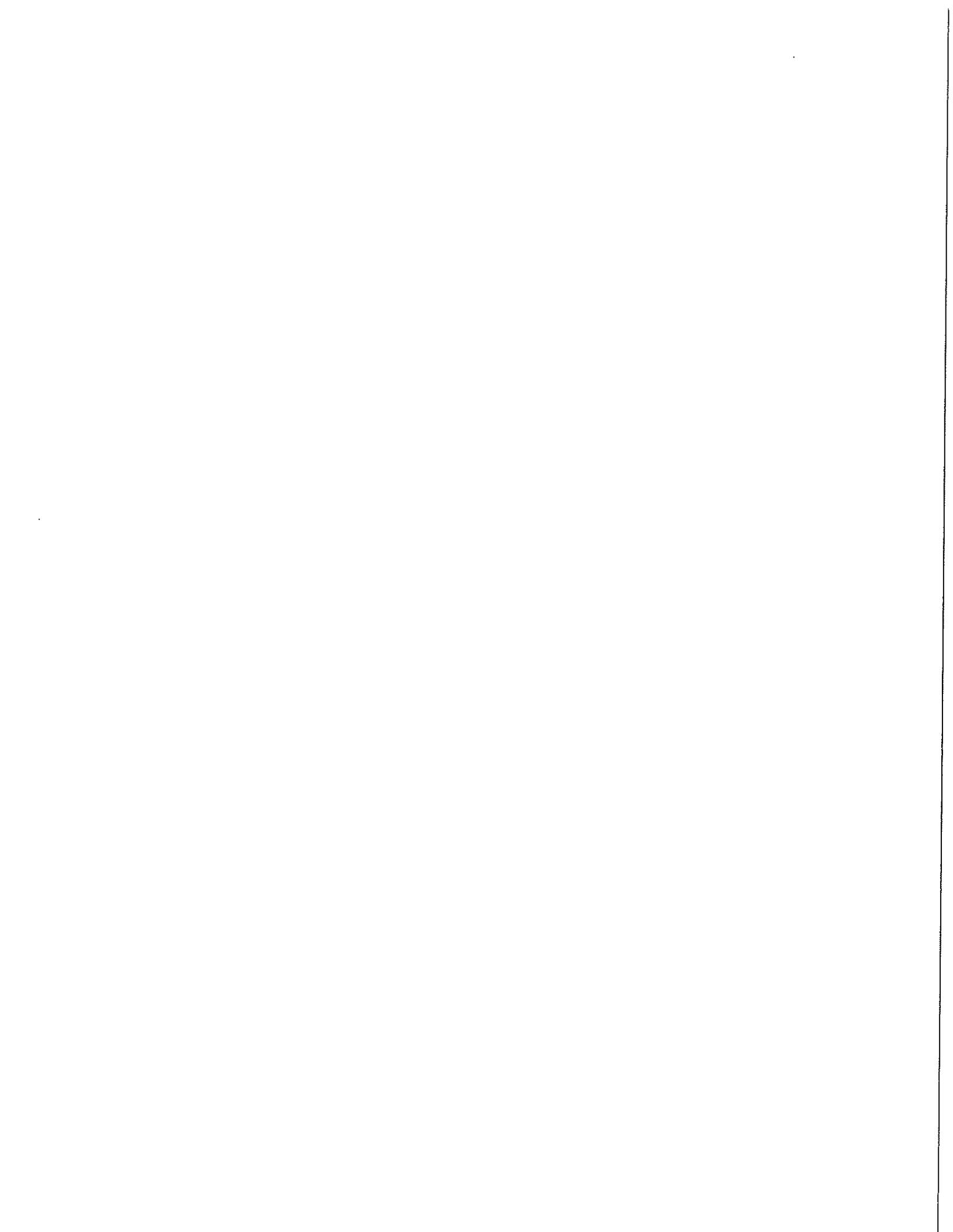
Rachel E. Ellsworth, Allyson L. Valente, Lori A. Field, Jennifer L. Kane, Brad Love, Lorraine Tafra, Craig D Shriver

**Background:** The widespread use of mammographic screening resulted in increased diagnosis of small ( $\leq 2$  cm; T1) tumors. Small tumors are associated with better prognosis, including a lower likelihood of developing metastasis, than larger tumors. Although this lower propensity to metastasize suggests that less aggressive treatments may be warranted in patients with T1 tumors, a subset of patients with small tumors (10-20%) will be diagnosed with lymph node metastasis.

**Methods:** Frozen breast specimens were collected from women with T1 tumors and either negative (n=29) or positive (n=15) lymph node status. RNA was isolated from pure tumor cell populations after laser microdissection. Gene expression data was generated using HG U133A 2.0 arrays (Affymetrix). Differential expression was determined using Mann-Whitney testing using a P-value  $< 0.001$  to define significance. Results for ESR1 were validated by immunohistochemistry.

**Results:** Tumor characteristics did not differ significantly between groups in terms of age at diagnosis, grade, HER2 or PR status; however, tumors from patients with positive lymph nodes (47%) were significantly ( $P < 0.05$ ) more frequently ER negative compared to node negative (14%) patients. Gene expression analysis revealed 17 genes that were differentially expressed between node negative and node positive tumors: 6 with higher expression in node positive, including AURKA, and 11 with higher expression in node negative patients, including ESR1 and EPHX2. Of note, ESR1 was expressed at  $>4X$  higher levels in tumors without metastasis, in agreement with IHC findings.

**Conclusions:** Small metastatic tumors differ in gene expression from those without metastasis. EPHX2 has been implicated as a metastasis suppressor while AURKA has been implicated as a metastasis promoter. These results suggest that small tumors have different propensities to metastasize and the genetic signature may serve as a new molecular tool to discriminate metastatic and non-metastatic small tumors, allowing appropriate treatment and risk assessment to be performed.



## Chromosomal alterations associated with poor outcome in breast cancer patients

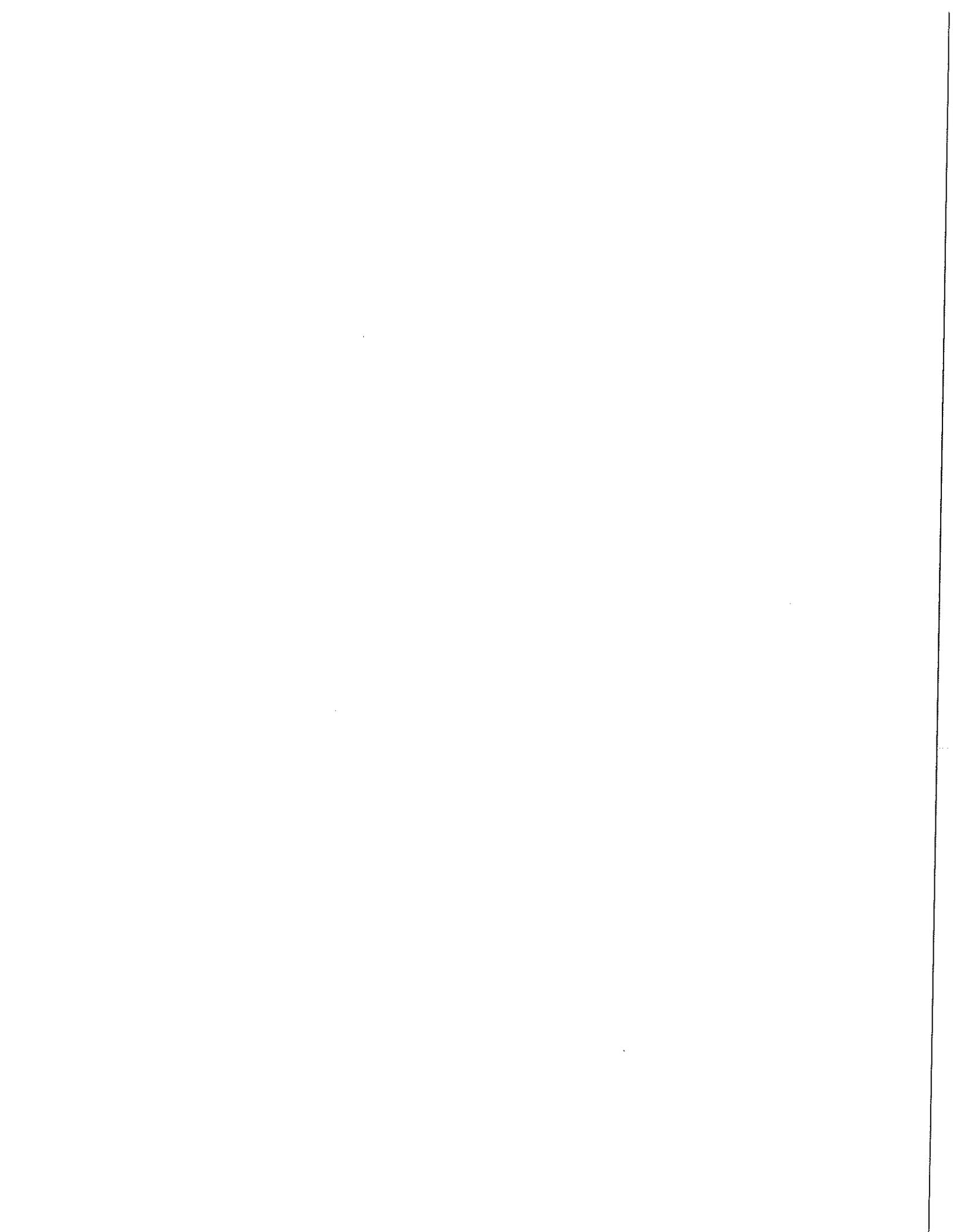
Rachel E Ellsworth, Jamie Leigh Fantacone-Campbell, Allyson Valente, Craig Shriver

**Background:** Despite improvements in detection and treatment, breast cancer remains a disease associated with significant mortality; in 2009, >40,000 women died from breast cancer. Pathologic evaluation is imperfect at predicting which patients will have aggressive disease as patients with otherwise similar pathologies may have very different clinical outcomes. In contrast, identification of molecular alterations associated with disease outcome may serve as accurate and non-subjective measures used to assess risk and prognosis as well as provide new molecular targets for improved treatment.

**Methods:** Between 2002 and 2008, allelic imbalance (AI) data was generated for >300 invasive breast tumors using DNA isolated after laser microdissection of pure tumor cells. A chart review for all patients was completed in December 2008 to identify patient status, as well as date of death or last contact. Levels and patterns of AI were compared between patients dead of disease (DOD) and those disease-free for at least 5 years using Student t-test and chi-square analysis with a significance value of  $P < 0.05$ .

**Results:** Within the 7 year period, 27 patients had died of disease and 99 had  $\geq 5$ -year disease-free survival (DFS); patients with less than 5-year DFS or who died of other causes were not included in further analyses. Levels of AI were significantly higher in tumors from DOD patients (26.6%) compared to DFS patients (20.7%). In addition, AI was found significantly more frequently in DOD tumors at chromosomes 8p22-p21 and 22q12.3 (33% and 28%) compared to DFS patients (12% and 15%); chromosome 16q22-q24 was altered significantly less frequently in DOD (10%) compared to DFS tumors (35%).

**Conclusion:** Tumors associated with poor compared to favorable outcome differ at the molecular level. Deletion of chromosome 8p22 including the downregulation of candidate genes PSD3, LPL and EPHX2 genes has been associated with poor outcome while silencing of the LARGE gene on chromosome 22q has been associated with defects in cell adhesion in highly metastatic cell lines. Alterations in either of these regions may promote metastatic spread. In contrast, loss of 16q appears to have a protective role, with DFS patients demonstrating significantly higher alteration of chromosome 16q compared to DOD patients. Research suggests that loss of the GINS2 gene, which is involved in the initiation of DNA replication, may protect against metastasis. These chromosomal regions and candidate genes may be effective targets for identifying patients with poor prognosis or serve as new targets for therapies to prevent or halt the metastatic process.



## Role of chromosome 16q in breast tumorigenesis

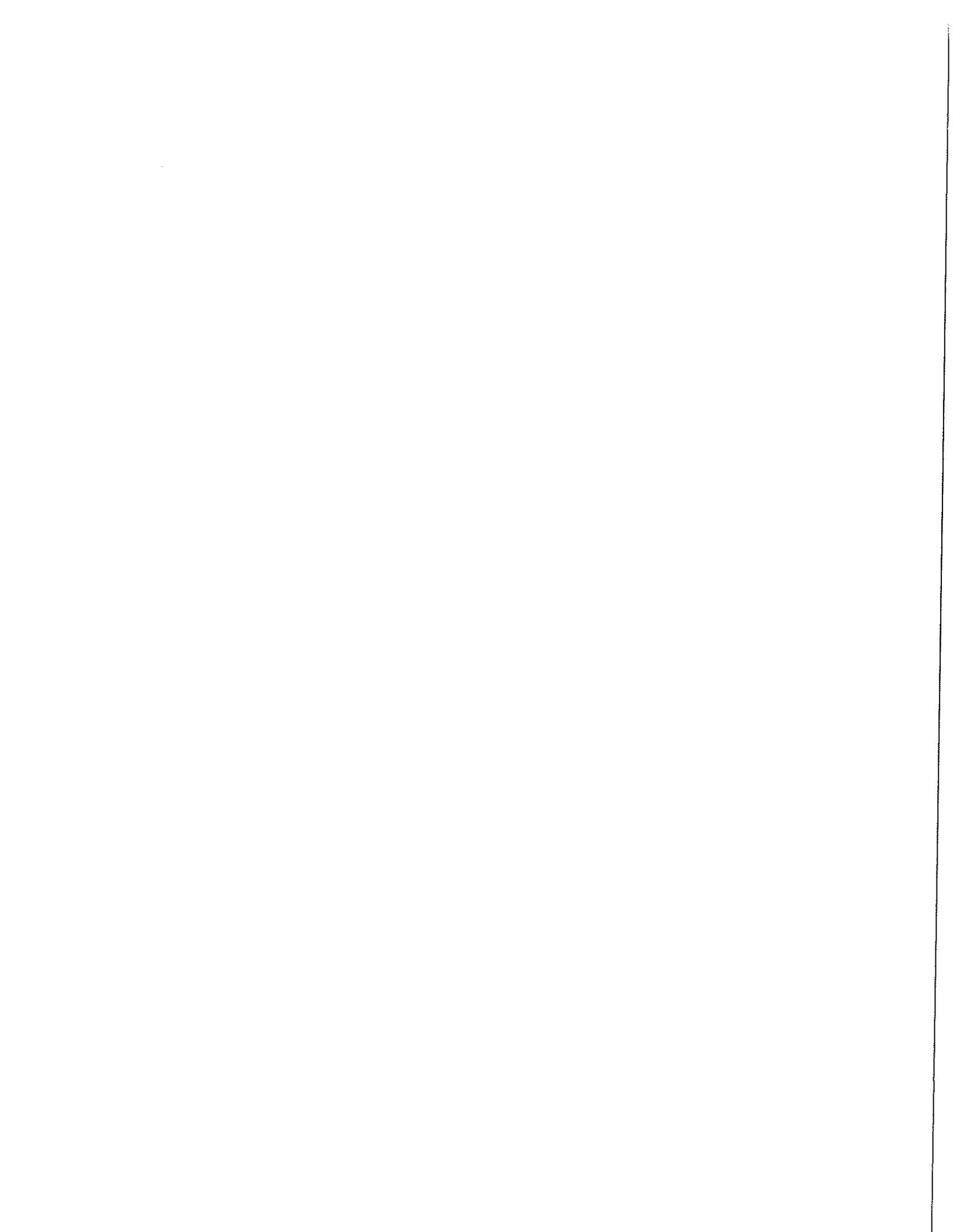
Rachel E. Ellsworth, Heather L. Patney, Jennifer A. Kane, Jeffrey A. Hooke, Craig D. Shriver

**Background:** The mechanism by which high-grade tumors develop is controversial. Studies based largely on observations that the frequency of alterations at chromosome 16q was significantly higher in low-grade tumors suggests that low-grade and high-grade tumors represent separate genetic diseases, while other studies suggest that loss of chromosome 16q is an early and common event in tumorigenesis, and high-grade carcinomas evolve from low-grade tumors. Levels and patterns of chromosomal alterations on chromosome 16q were investigated in low- and high-grade tumors to determine the etiology of high-grade breast tumors.

**Methods:** Tumors were diagnosed and characterized by a single, dedicated breast pathologist. DNA was isolated after laser microdissection from low- (n= 53) and high-grade (n=68) tumors. Genotype data were generated using Human Mapping 250K Sty arrays (Affymetrix). Copy number alterations were detected using Genotyping Console 3.0.2 (Affymetrix) and statistical differences calculated using Fisher's exact test with a cut-off of  $P < 0.05$ .

**Results:** Alteration of chromosome 16q was common with 89% of low-grade and 54% of high-grade tumors ( $P < 0.000005$ ) with a detectable amplification or deletion, although amplifications were more common in high-grade (29%) compared to low-grade (9%) tumors. Deletion of the entire arm of 16q was rare, with only a single high-grade tumor having loss from the centromere to the telomere. Low-grade tumors demonstrated significantly greater ( $P < 0.001$ ) amounts of loss of chromosome 16q (mean = 578 kb, range 0-4,204 kb) compared to high-grade tumors (mean = 95kb, 0-4,225 kb). The most common regions of loss were found at 16q13, 16q21, 16q22.1, 16q23.1 and 16q24.1, all found deleted at significantly higher ( $P < 0.000005$ ) levels in low-grade tumors. No chromosomal regions were altered more frequently in high- compared to low-grade tumors.

**Conclusions:** Under the model of sequential progression, chromosomal alterations in low-grade tumors should be maintained in high-grade tumors; however, data generated here demonstrates that regions of chromosome 16q frequently lost in low-grade tumors are diploid in high-grade tumors. These data, therefore, support a model by which low-grade and high-grade diseases develop along separate genetic pathways, with alterations of chromosome 16q serving as the critical genetic determinant between histological grades.



## Loss of Concordance in Immunohistochemical Markers Between Core Needle Biopsy and the Surgical Specimen in Heterogeneous Breast Tumors

Lauren Greer, Martin Rosman, W. Charles Mylander, Janet Wareham, J. Leigh Campbell, Al Kovatich, Jeffrey Hooke, Joan Woodward, Wen Liang, Robert Buras, Craig Shriver, Lorraine Tafra

Breast tumor heterogeneity, defined as isolated areas with different histology (type and/or grade) or molecular subtypes within a single tumor or among multicentric/multifocal tumors may be present in as high as 10% of patients with breast cancer. Tumor markers (ER, PR, HER2, Ki-67) from core needle biopsies (CNB) of the primary tumor may not be representative of the entire surgical specimen (SS) in heterogeneous tumors. Our aim is to investigate the level of concordance between tumor markers performed on the CNB and the heterogeneous areas of the surgical specimens (SS) in these patients.

A retrospective review of a prospective study was performed on women diagnosed with breast cancer by CNB from January 2009 to June 2011. Subjects treated with neoadjuvant therapy were excluded. Tumor type, grade, ER, PR, HER2, Ki67 expression by IHC were analyzed in the CNB and SS. Subjects with tumor heterogeneity were identified. Contingency tables and agreement modelling were performed to generate Kappa values for heterogeneous patients.

### Results

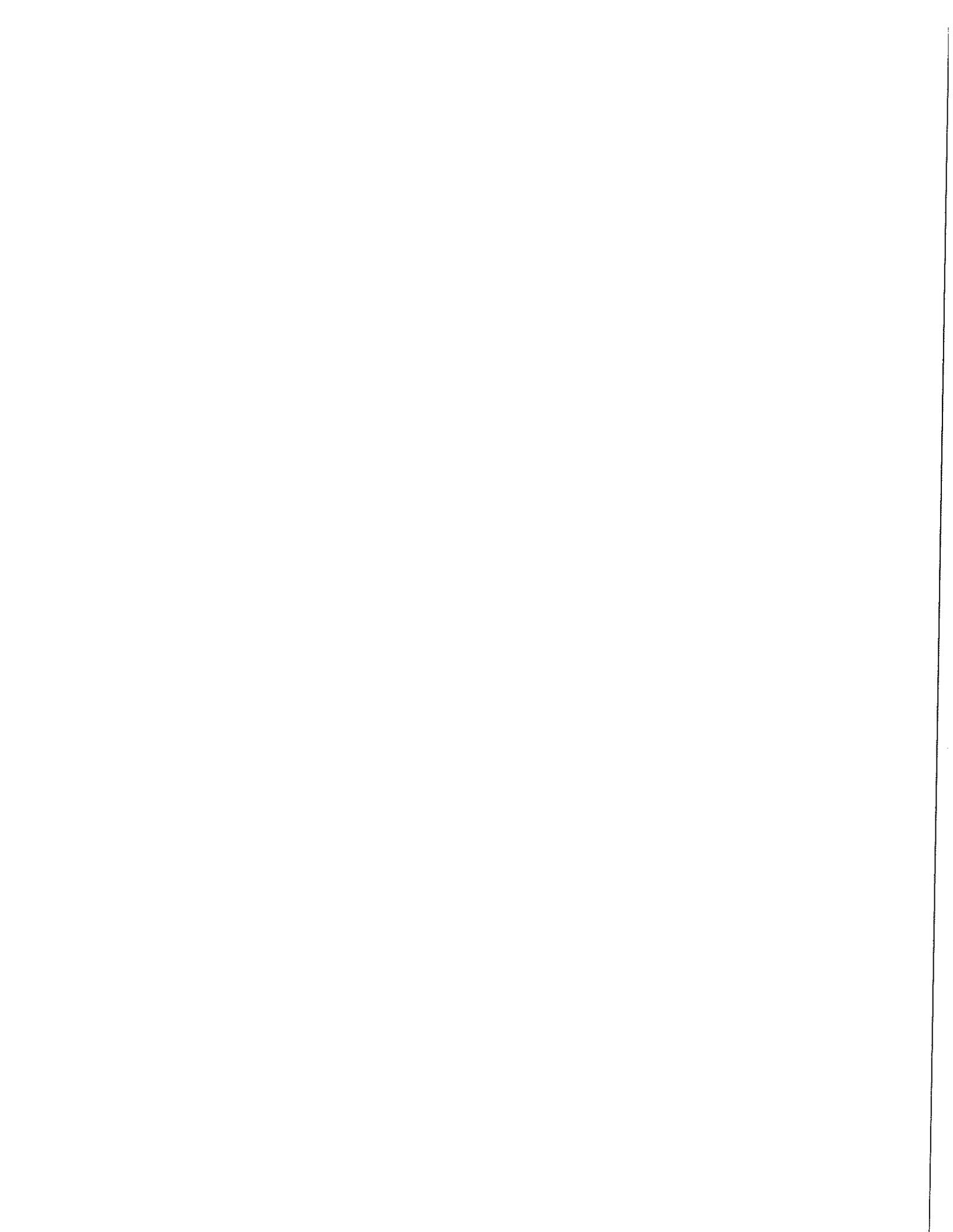
Tumor heterogeneity by H&E and IHC staining was identified in 10 out of \_\_ subjects, either within a single tumor (n=\_) or in different tumors in the same breast (n=). Multiple IHC testing was performed on these different areas of the SS (24 samples analyzed). Tumor sizes ranged from 2mm to 6cm. One patient was noted to have 4 distinct molecular subtypes within the same tumor. There was substantial agreement between the CNB and SS for ER%; moderate agreement for tumor type, PR%, and HER2; and fair agreement for Ki67% (see Chart). Three subjects (30%) had areas of their tumors that stained positive for HER2, which was not detected in their CNB. Three subjects had areas in their SS that were negative for ER and PR, but stained positively in their CNB. No cases were seen where ER or PR were negative on CNB but became positive in the SS.

Blomarker	# of Agreements In total samples	Kappa (κ)	Lands-Koch Agreement Grade
Tumor Type	18/23 (78%)	0.45	Moderate
Histologic Grade	16/22 (73%)	0.51	Moderate
ER%	22/24 (92%)	0.63	Substantial
PR%	20/24 (83%)	0.43	Moderate
HER2	20/24(83%)	0.43	Moderate
Ki67 %	14/22 (64%)	0. 3	Fair

## **Conclusions**

Tumor heterogeneity may be present within a single breast tumor or in separate concurrent tumors within the same breast. The accuracy of single tumor markers by IHC performed on the CNB of subjects with heterogeneity is not always representative of the biology of the whole tumor(s). These data highlights the need for further IHC testing on the SS of these subjects to better tailor targeted therapy.

"The views expressed in this abstract are those of the authors and do not reflect the official policy of the Department of the Army (DOA), Department of Defense (DOD), or US Government."



## DCU: A Data Correction Utility to Correct Clinicopathologic Data in a Data Warehouse

Leonid Kvecher<sup>1</sup>, MS, Weiqiang Wu<sup>1</sup>, MS, Joni Kohr<sup>1</sup>, Craig D. Shriver<sup>2</sup>, MD,  
Richard J. Mural<sup>1</sup>, PhD, and Hai Hu<sup>1</sup>, PhD

<sup>1</sup>Windber Research Institute, Windber, PA

<sup>2</sup>Walter Reed Army Medical Center, Washington, DC

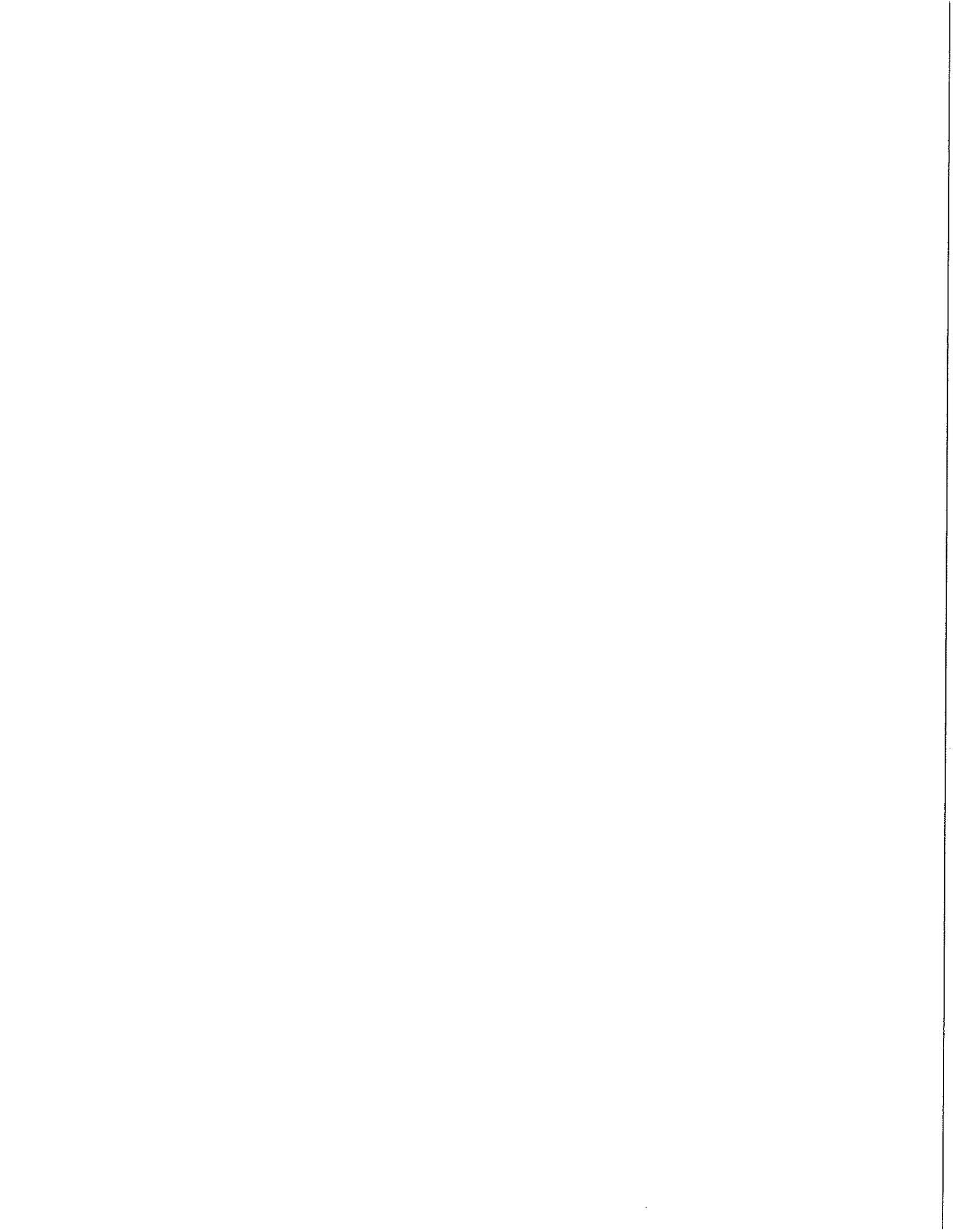
**Abstract:** In the Clinical Breast Care Project (CBCP), clinicopathologic data are collected using hard copy questionnaires, entered into the data tracking system, and loaded to our Data Warehouse for Translational Research (DW4TR). Using Quality Assurance measures, discrepancies in the data are identified and corrected. There is a need to perform systematic data corrections and updates in the DW4TR, and this poster presents a Data Correction Utility (DCU) to serve this purpose.

### Description

**Background:** The Clinical Breast Care Project (CBCP), is a US Congressionally mandated, multi-site translational research program initiated in 2000, with a long term goal of improving the care of patients with breast cancer. In the CBCP, a broad range of clinicopathologic data is collected from participants using multiple hard-copy questionnaires with a total of 799 data fields. The data are entered into a clinical Laboratory Information Management System (LIMS), and loaded into our Data Warehouse for Translational Research (DW4TR) that integrates clinicopathologic data and experimental data. Using Quality Assurance (QA) measures, errors are identified and corrected during the data collection and data entry steps. However, there is a need to correct or update the data individually or systematically in the DW4TR, for example when the cancer staging criteria and biomarker threshold standards changed. Given the restrictions embedded in the LIMS, such changes can only be made individually and manually which is not desired. Therefore we have developed a Data Correction Utility (DCU) to update clinicopathologic data in the DW4TR.

**Methods and Results:** The InforSense Platform was used to develop the DCU in the form of an analytical workflow to correct the data in the Oracle database of the DW4TR. The input to the utility is a flat file with 5 columns, composed of "Subject ID", "Questionnaire ID", "Question ID", "Old Value", and "New Value". The flat file can be manually compiled, or automatically generated using other QA utilities. The DCU first analyzes the information in the first 4 columns to confirm the "Old Value" and prompts the user to resolve any problems identified. Then, one of the three operations of "Insert", "Delete", and "Update" can be performed depending on the nature of the "Old Value" and the "New Value". A graphical user interface was developed so that authorized users, including nurses and the clinical data QA staff, can access the DCU within the network or remotely via Virtual Private Network. As one application example, the DCU eliminated the need for manual individual correction of hundreds of data fields in our LIMS when the standard thresholds for ER+ and PR+ for breast cancer tissues were changed.

**Discussion:** The DCU has enabled authorized clinical and research staff to make corrections of the clinicopathologic data to the DW4TR. This utility answered the needs for data correction when such corrections are difficult or too cumbersome to perform in the LIMS. The principle of the DCU is applicable to other clinical data warehouse systems, especially when there is no LIMS to serve as the data source, or when a batch data correction is needed.



**Impact of lifestyle factors on quality of life and prognosis in breast cancer survivors in the United States**

Rachel E. Ellsworth, Allyson L. Valente, Craig D. Shriver, Barry Bittman, Darrell L. Ellsworth

Rachel E. Ellsworth  
Clinical Breast Care Project  
Henry M. Jackson Foundation for the Advancement of Military Medicine  
620 Seventh Street  
Windber, PA 15963  
Phone: 814-361-6926  
FAX: 814-467-6334  
Email: r.ellsworth@wriwindber.org

Allyson L. Valente  
Clinical Breast Care Project  
Windber Research Institute  
620 Seventh Street  
Windber, PA 15963  
Phone: 814-361-6976  
FAX: 814-467-6334  
Email: a.valente@wriwindber.org

Craig D. Shriver  
Clinical Breast Care Project  
Walter Reed National Military Medical Center  
8901 Rockville Pike, Bldg 19, 3rd floor  
Bethesda, MD 20889  
Phone: 301-295-6333  
Fax: 301-295-9076  
Email: craig.shriver@med.navy.mil

Barry Bittman  
Mind-Body Wellness Center  
18201 Conneaut Lake Road  
Meadville, PA 16335  
Phone 814-333-5061  
FAX: 814-333-5067  
Email: bbittman@mmchs.org

Darrell L. Ellsworth  
Clinical Breast Care Project

Windber Research Institute  
620 Seventh Street  
Windber, PA 15963  
Phone: 814-361-6911  
FAX: 814-467-6334  
Email: [d.ellsworth@wriwindber.org](mailto:d.ellsworth@wriwindber.org)

## **Summary**

Advances in diagnostic screening and adjuvant therapy have dramatically increased the number of breast cancer survivors in the United States, who face numerous issues of physical and mental health, social support, quality of life, and economics. Women living with breast cancer are increasingly interested in lifestyle modification to decrease risk of recurrence and mortality while increasing physical and emotional well-being. Although a healthy diet, frequent physical activity, and stress reduction may be beneficial for decreasing breast cancer risk, studies examining the effects of lifestyle on clinical outcomes such as survival and prognosis have been inconclusive. With the number of breast cancer survivors predicted to increase to 3.4 million by 2015, it is important to develop more effective treatment paradigms that overcome barriers to behavioral modification to improve clinical outcomes and survivorship in breast cancer patients.

**Running Title**

Lifestyle modification in breast cancer survivors

**Keywords:** Breast cancer, lifestyle modification, survivors, diet, physical activity, stress management, quality of life

## **Introduction**

Breast cancer is the most frequently occurring cancer in women and is second only to lung cancer in lethality. Over the last fifty years, lifetime risk of breast cancer has increased such that today one in eight women are expected to develop breast cancer. Projections indicate that more than 230,000 new breast cancers will be diagnosed, and ~40,000 women will die from breast cancer, in the United States (US) in 2011 [1]. Mortality rates increased from 1970 to 1990, but in recent years, death rates have decreased by 28% [2]. This notable decline in mortality, likely attributable in large part to improvements in early detection and adjuvant therapy [3], has produced more than 2.5 million breast cancer survivors in the US today [201].

Breast cancer survivors face numerous issues that can degrade quality-of-life (QOL), including detrimental changes in social support, psychological health, personal finance, and physical activity [4]. Survivors may suffer from anxiety, depression, fear of recurrence, impaired body image, and increased risk of co-morbid conditions [5-7]. With the number of breast cancer survivors in the US predicted to increase to 3.4 million by 2015 [8], it is imperative to develop interventions that will improve QOL in breast cancer survivors.

Studies of behavioral modification in breast cancer survivors are providing new information about how lifestyle factors such as diet and related obesity, physical activity, and stress affect survivorship, as well as knowledge to develop new, effective intervention programs for survivors. In this article, we outline current scientific evidence demonstrating the effects of diet, physical activity, and stress management on QOL in breast cancer survivors, and discuss barriers to widespread implementation of lifestyle modification strategies.

## **The economic impact of breast cancer**

The economic impact on cancer survivors is often overlooked, but has great potential to negatively impact QOL. The National Cancer Institute (NCI) estimates that direct costs for care of women with breast cancer in the US was \$16.5 billion in 2010 [202]. Average lifetime costs for new personalized treatments such as trastuzumab (Herceptin®) for HER2 positive breast cancer are \$44,923 for early breast cancer patients and \$47,728 for women with metastatic disease [9]. Medical out-of-pocket costs can exceed \$1,000 per month during active treatment, and are ~\$500 per month at one year post-diagnosis, coupled with non-medical out-of-pocket costs typically ranging from \$200-\$500 per month [10]. For many patients living with breast cancer, the monetary burden associated with treatment and survival can accentuate stress, precipitate unhealthy behaviors, and impact QOL.

### **Lifestyle modification and human health**

Lifestyle choices and behavioral modifications can significantly impact overall health, as well as susceptibility to cancer, cardiovascular disease, and diabetes. While improvements in hygiene and sanitation have led to a decrease in infectious diseases, the transition from primarily plant-based diets to modern (Western) diets has been associated with an increased incidence of hypertension, heart disease, diabetes, and obesity [11]. Observations that women native to countries with low rates of breast cancer who migrate to the US and adopt a Western lifestyle develop a higher risk of breast cancer suggest that behavior and environmental factors also influence breast cancer etiology [12].

A variety of epidemiological study designs have been used to examine environmental factors and behaviors that contribute to risk of developing breast cancer or long-term prognosis (Table 1). Observational studies performed prospectively or retrospectively usually identify associations between behaviors and outcomes of interest [13]. Observational trials can assess the effectiveness of treatments in diverse patient populations, identify differences among therapeutic options, and examine long-term

effectiveness [14], but may be susceptible to bias and confounding, which limit their ability to measure causality or compare treatments. Conversely, randomized controlled trials (RCTs) are considered the gold standard for measuring the efficacy of interventions on particular outcomes, such as disease recurrence or prognosis. In an RCT, patients with similar characteristics are randomly assigned to a group receiving the intervention or to a control group who receive a placebo or standard treatment. At the conclusion of the trial, a significant difference between the intervention and control groups in the outcome of interest can be attributed to the intervention and not to other factors [15]. Much of the research presented in this paper was derived from observational studies and RCTs measuring the effects of diet, physical activity, and stress reduction on prognosis and QOL in breast cancer survivors.

## **Diet and breast cancer prognosis**

### ***Dietary fat, fiber, fruit, and vegetable intake***

The majority of dietary studies examining the impact of body weight on breast cancer survival observed that obesity at diagnosis is associated with poor prognosis [16]. Similarly, weight gain after diagnosis is common and is associated with mortality, disease recurrence, and development of co-morbid conditions including diabetes and cardiovascular disease [17]. In patients with breast cancer, each five-kilogram gain in weight has been associated with a 13% increase in risk of breast cancer mortality, a 12% increase in all-cause mortality, and a 19% increase in cardiovascular mortality [18]. Given the prediction that 41% of adults in the US will be obese by 2015 [19], improved understanding of the relationship between obesity and prognosis is crucial to developing effective strategies to improve survival and QOL.

The Nurses Health Study (NHS), one of the largest prospective cohort studies in the US, was established in 1976 to follow 121,700 female nurses over time through periodic health assessment

questionnaires. Contrary to earlier studies suggesting that a low-fat diet is associated with improved prognosis [20], food frequency data from 1,982 NHS participants with invasive breast cancer collected over 10 years showed no association between fat intake and mortality [21] (Table 2). A second NHS investigation of 2,619 women with dietary data collected before, and at least one year after, a diagnosis of invasive breast cancer compared women who consumed a “prudent” diet, high in fruits, vegetables, whole-grains, fiber, and protein to women following a “Western” diet, characterized by high amounts of refined grains, processed meat, high-fat dairy, and low amounts of protein and fiber. Although women consuming a Western diet before and after breast cancer diagnosis had a higher risk of all-cause mortality, diet did not impact breast cancer prognosis [22]. A third NHS study evaluating change in body mass index (BMI) in 5,204 women found that weight gain after breast cancer diagnosis was associated with less favorable prognosis, including higher rates of recurrence and mortality [23]. Therefore, weight gain, independent of individual dietary components, may be associated with poor prognosis in breast cancer survivors.

A second large, prospective cohort study, the Life After Cancer Epidemiology (LACE) Study, enrolled 2,321 women from a California-based health-maintenance organization (HMO) and the Utah cancer registry who were diagnosed with breast cancer between 1997 and 2000. Dietary intake of more than 100 foods, beverages, nutritional supplements, and medicinal herbs was assessed by food frequency questionnaires. Preliminary data suggest that dietary patterns of women with breast cancer were similar to those in women free from breast cancer in the general population [24]. Similar to the NHS study, a Western diet was associated with increased risk of all-cause mortality, but was not associated with breast cancer mortality [25].

In the Health, Eating, Activity, and Lifestyle (HEAL) Study, a significant number of African American and Hispanic participants were recruited from California, Washington, and New Mexico. In

this study, women gained an average of 1.7 kg of body weight within three years of breast cancer diagnosis. Overall, 68% of women gained weight and 74% increased their body fat [26]. Women (n=670) with early stage breast cancer, who consumed a better-quality diet (low in calories, added sugar, alcohol, and saturated fat) had a 60% lower risk of all-cause mortality and an 88% lower risk of breast cancer-related mortality [27]. A quality diet also was associated with increased mental and physical QOL, as well as decreased levels of circulating inflammatory markers [28,29].

In contrast to the observational studies described above, the Women's Intervention Nutrition Study (WINS) is a randomized trial measuring the effects of a low-fat diet on relapse-free survival. Women in the intervention group (n=975) undergoing standard treatment for stage I or stage II breast cancer followed a low-fat dietary intervention, while controls (n=1,462) received minimal nutrition counseling. The intervention group showed a significant decrease in fat intake and maintained lower fat consumption for more than five years, while patients in the control group showed no significant change in fat intake. After 60 months of follow-up, relapse-free survival was 24% higher in the intervention group [30]. Although this study focused solely on reducing dietary fat, participants in the intervention arm also increased fruit consumption and lost significantly more weight than controls.

In the Women's Healthy Eating and Living (WHEL) Study, women with stage I–IIIA breast cancer were randomized to an intervention arm (n=1,537) or control arm (n=1,551). The intervention group consumed a high-fruit and vegetable diet, with only 15%-20% of calories from fat, while controls were advised to follow the 5 A Day for Better Health Program diet recommended by the NCI. Nutritional intake for both groups was similar at baseline, but after four years, the intervention group increased vegetable consumption by 65%, fruit consumption by 25%, fiber by 30%, and decreased calories from fat by 13% compared to the control group. Despite improvements in nutrition, breast

cancer event-free survival and mortality did not differ significantly between the intervention and control groups [31].

These studies suggest that a low-fat, high-fiber diet may have only marginal effects on breast cancer mortality. Given the link between poor prognosis and obesity at diagnosis, as well as weight gain after diagnosis, positive effects on survival observed in the WINS may be attributable to incidental weight loss, rather than specific dietary components. Although diet may have no direct effect on breast cancer mortality, healthy eating has been shown to reduce the risk of developing other chronic diseases, such as cardiovascular disease and diabetes. Thus, the American Cancer Society (ACS) recommends a “heart healthy” diet similar to that recommended for cardiovascular health for breast cancer survivors.

### *Alcohol*

In a large meta-analysis that included 58,515 women with invasive breast cancer and 95,067 controls from 53 studies, alcohol consumption was associated with ~7% increased risk of developing breast cancer for every 10 g (unit or drink) of alcohol consumed on a daily basis [32]. Relationships between alcohol consumption and breast cancer survival, however, have not been well defined. One study of 1,268 women  $\leq 45$  years of age with invasive breast cancer assessed the effect of pre-diagnostic alcohol consumption on mortality and determined that women who consumed alcohol during the five-year period prior to diagnosis had a decreased risk of breast-cancer-related mortality compared to non-drinkers [33]. Similarly, the population-based Studies of Epidemiology and Risk Factors in Cancer Heredity (SEARCH) Study of 4,560 breast cancer patients detected a 2% reduction in mortality risk for every unit of alcohol consumed per week [34]. In contrast, the LACE Study detected a 1.3-fold increase in breast cancer recurrence and a 1.5-fold increase in mortality among overweight and obese women who had 3-4 alcoholic drinks per week [35]. Other studies have shown

no relationship between alcohol consumption and survival [36,37]. Given the protective effects of certain types of alcohol on cardiovascular disease, a personal decision on alcohol consumption should be balanced between breast cancer risk and cardiovascular health.

### *Nutraceuticals and functional foods*

Functional foods provide health benefits beyond nutrition. For example, yogurt is a functional food containing live microorganism cultures that do not add nutritive value, but inhibit pathogens and toxin-producing bacteria. In contrast, “nutraceuticals” represent hybrids between traditional nutrients and pharmaceutical compounds, usually available as pills or powders containing bioactive components that provide health benefits [38]. As 75%-87% of breast cancer survivors take dietary supplements to improve survival [39], it is important to understand how functional foods and nutraceuticals affect breast cancer prognosis.

### *Vitamins*

Nearly 50% of adults in the US use vitamin supplements, even though research has provided no convincing evidence of true beneficial effects [40]. At present, the US Preventive Services Task Force and the National Institutes of Health Office of Dietary Supplements and Office of Medical Applications do not recommend use of multivitamins for the prevention of chronic disease [41,42]. Furthermore, two large studies examining the impact of multivitamins on breast cancer risk, the Women’s Health Study (WHS) and Women’s Health Initiative (WHI), found that multivitamin use did not decrease risk of developing breast cancer or influence overall mortality [43,44].

Despite the negative results described above, other studies have observed positive associations between vitamin supplements and improved prognosis in breast cancer survivors. Vitamin A is known

to play a role in inflammation, cell growth, and tissue differentiation, and thus may influence recurrence and metastasis [45]. One study showed that premenopausal survivors diagnosed with node-positive breast cancer and low levels of retinol binding protein, the carrier for vitamin A in plasma, had early disease recurrence [46]. In a separate study of postmenopausal women, low vitamin A levels were associated with poor outcomes, such as distant metastasis and increased mortality [47].

Thiamine (B1), riboflavin (B2), niacin (B3), and folic acid (B9) belong to the water soluble B-complex group of vitamins. Certain B-complex vitamins promote cell growth and division, and thus have been the subject of research examining possible relationships between B vitamins and breast cancer. In a study of 516 postmenopausal women with breast cancer from Orange County, California, women with the highest folate intake had lower mortality risk. Nutrients derived from diet, rather than supplements, were believed to provide the survival advantage [48]. In the Long Island Breast Cancer Study Project (LIBCSP) (n=1,508 women), vitamins B1 and B3 were associated with a 46% and a 39% lower risk, respectively, of all-cause mortality. Vitamin B1 intake was associated with decreased breast cancer mortality [49].

Vitamin C (ascorbic acid) functions as an antioxidant and is important to healthy immune function. In the Orange County study highlighted above, vitamin C was significantly associated with decreased all-cause mortality, but not with breast cancer mortality [48]. In the Shanghai Breast Cancer Survivor Study (SBCSS) of 4,877 women with breast cancer, vitamin C intake was associated with a 44% decrease in all-cause mortality and a 38% decrease in breast cancer recurrence. Similar but less convincing results were observed for vitamin E [50]. An evaluation of the effects of vitamin E and its derivatives on breast cancer cell lines suggests that antioxidants such as vitamin E may prevent disease recurrence by inducing apoptosis in breast cancer cells [51].

Vitamin D is produced naturally by the body during exposure to sunlight, but may be consumed through fortified foods or vitamin supplements. In addition to bone health, vitamin D affects cell proliferation, cell differentiation, and apoptosis. Regional differences in sunlight exposure in the US have been suggested to influence breast cancer mortality — women with greater sunlight exposure had a lower risk of fatal breast cancer, and length of daylight at the time of diagnosis has been linked to breast cancer prognosis [52,53]. Although one group found significantly higher levels of circulating 25-hydroxyvitamin D in women with early stage disease compared to patients with locally advanced or metastatic breast cancer [54], other studies suggest that vitamin D consumption following treatment may have no survival advantage [55].

Nearly 75% of breast cancer patients report taking multivitamins to bolster their immune system and make treatment more effective. Although the results outlined above suggest that vitamin supplements may have beneficial effects on prognosis in breast cancer survivors, the link between vitamin supplements and breast cancer is complex and not well studied. Current recommendations from the ACS suggest that survivors should consider using multivitamins with 100% of the Daily Recommended Values to ensure proper nutrition.

### *Green tea*

Consumption of green tea has long been customary in Asia, but the purported health-promoting benefits of green tea have been recognized only recently in the US. Green tea contains a number of active ingredients including polyphenols, which may confer anti-mutagenic, anti-diabetic, anti-bacterial, anti-inflammatory, or hypo-cholesterolemic advantages [56]. In studies evaluating the effects of green tea consumption in 427 Japanese breast cancer survivors with stage I or stage II disease, recurrence was 16.7% in women who consumed  $\geq 5$  cups of green tea per day, compared to 24.3% in

those consuming <5 cups per day. This protective effect was not observed in patients with stage III breast cancer [57]. Likewise, in a second study of 1,160 Japanese women, consuming 3 or more cups of green tea each day was protective against recurrence in women with stage I disease, but showed no effect in more advanced cases [58]. A meta-analysis of these datasets showed that increased green tea consumption (more than 3 cups per day) was inversely associated with breast cancer recurrence [59]. Little information exists on the relationship between green tea consumption and survival in breast cancer patients; however, as no protective effects from green tea have been observed in advanced disease, the effects of green tea on overall breast cancer survival may be minimal.

### *Soy*

A variety of health benefits have been attributed to consumption of soy-based foods, primarily due to soybean isoflavones, such as genistein and diadzein. Major sources of isoflavones in the American diet are derived from soy-containing products such as oil, flour, soy-milk, and meat substitutes. The chemical structure of isoflavones may enable these compounds to interact with human estrogen receptors, causing effects in the body similar to those induced by estrogen. Experimental studies in cell lines and animal models suggest that genistein has the potential to stimulate cell proliferation, negatively influence prognosis, and impair the effectiveness of tamoxifen therapy in estrogen-receptor-positive (ER+) breast carcinomas [60,61].

Given the results of experimental studies outlined above, which suggest that soy constituents can be estrogenic and potentially risk enhancing, an important question remains: Do the estrogenic effects of soy influence prognosis and survival in humans? In population-based studies in the US, women from the LIBCSP diagnosed with a first primary invasive breast carcinoma (n=1,210) showed reduced hazard ratios for all-cause mortality in the highest quintile of dietary intake compared with the

lowest quintile for flavones, isoflavones, and anthocyanidins. A decrease in breast cancer mortality was detected only in postmenopausal patients [62]. Other studies of diverse populations also observed significant inverse associations between soy consumption and breast cancer recurrence in postmenopausal women [63,64].

The majority of studies examining the relationship between soy consumption and breast cancer prognosis have been performed in Asian populations where long-term soy intake is much higher than in Western countries. Available data suggest that dietary intake of soy-based foods may confer a protective advantage for postmenopausal women with breast cancer. Additional health benefits have been attributed to consumption of soy-based foods, such as lower cholesterol and improved bone health; however, intake of isoflavones at high concentrations typically found in dietary supplements may be detrimental. Until the issue is fully resolved, physicians should allow patients receiving hormonal therapy, or who have ER+ breast cancer, to consume moderate amounts of soy as part of a balanced diet, and may wish to recommend avoiding soy supplements.

As outlined above, most studies examining diet and survival in breast cancer patients were observational, population-based studies, rather than randomized trials. Although a prudent diet was not associated with breast cancer survival, a healthy diet has been shown to have beneficial effects on overall survival in conditions such as diabetes and heart disease, which are frequently seen in breast cancer patients. In addition, a sensible diet can result in significant weight loss, which has been associated with improved prognosis. Thus, patients should be encouraged to adopt a healthy diet that may include modest alcohol consumption, multivitamins, soy, and green tea.

### **Physical activity**

Physical activity is as important as diet for achieving optimal weight and maintaining a healthy lifestyle. In 2008, the US Department of Health and Human Services issued guidelines for physical activity, which included at least 150 minutes of moderate, or 75 minutes of intense, aerobic exercise per week, as well as strength-training at least twice a week to achieve significant health benefits [203]. Strenuous and recreational physical activities have been shown to confer numerous health benefits, including lower risk of developing breast cancer [65-67].

Treatment regimens consisting of surgery with or without chemotherapy and/or radiation have been associated with a number of significant side effects including fatigue, nausea, decreased cardiovascular health, and emotional distress. A number of randomized trials have determined that patients undergoing treatment who engaged in regular exercise had improved health and QOL compared to those who did not exercise [68-72]. Despite these positive findings, less than 50% of patients undergoing treatment, especially chemotherapy, engage in regular physical activity. As demonstrated in the HEAL study, physical activity decreased by approximately two hours per week after diagnosis compared to activity levels one year before diagnosis [73]. Similarly, Belgian women (n=267) with invasive breast cancer showed a significant decrease in physical activity during the first month post-diagnosis, and did not resume normal activity levels during the first year [74].

Risk of developing detrimental side effects, including cardiovascular disease, type 2 diabetes, fatigue, lymphedema, psychological distress, and poor QOL often persists in breast cancer survivors after treatment has been completed. Two small studies initially suggested that exercise may have beneficial effects in post-treatment survivors. In a cross-over study of 24 patients evaluating the effects of physical activity on psychological health, mild-to-moderate intensity exercise for 30 minutes, four days per week for 10 weeks, with or without positive reinforcement, was associated with lower depression and anxiety [75]. In a similar study, postmenopausal breast cancer survivors (n=53) who

performed moderate aerobic exercise three times per week for 15 weeks showed significant improvement in cardiopulmonary function and QOL [76].

Large cohort studies have clearly demonstrated the importance of physical activity in breast cancer patients (Table 3). Data from the NHS from 2,987 women with stage I-III breast cancer showed that any amount of exercise above 3-MET hours per week (the equivalent of walking at a moderate pace of 2-3 miles per hour for one hour) had a beneficial effect on survival. In patients with the highest activity, breast cancer recurrence and mortality were 26%-40% lower than in women with the lowest activity [77]. In the LACE trial, higher self-reported physical activity, including occupational, household, transportation-related, and recreational activity, was significantly associated with lower all-cause mortality in 1,970 breast cancer survivors [Sternfeld, 2009 1318 /id]. The Collaborative Women's Longevity Study (CWLS), which examined 4,482 women diagnosed with breast cancer between 1988 and 2001, suggested that women who engaged in >2.8 MET hours per week had a 35%-49% lower risk of breast cancer-related mortality compared to those who were physically inactive, even after adjusting for confounding factors such as disease stage, treatment modality, and age at diagnosis. Moderate activity appeared to provide the greatest benefit, with vigorous activity providing no additional survival advantage [79]. Likewise, women in the HEAL study who were physically active in the year immediately prior to diagnosis or two years post-diagnosis showed significantly lower breast cancer mortality than women who were inactive. Women who decreased their activity after diagnosis had a four-fold increase in mortality, while those who increased their activity had a 45% lower risk of death [80].

Most research to date strongly supports the link between physical activity and improved breast cancer prognosis and overall survival, but several issues, such as (1) the optimal form of physical activity, (2) the frequency and duration of activity, and (3) an individualized versus a standardized

physical activity regimen remain unresolved. Most studies indicate that moderate-intensity aerobic exercise provides the most benefit in terms of prognosis. The ACS currently recommends at least 30 minutes of physical activity five days a week [81]. In addition, a personalized activity plan may be more effective than a standardized regimen because an individual plan can be customized for different time periods from pre-diagnosis through cancer treatment based on individual needs and abilities. Perhaps the most important factor affecting breast cancer survivors is long-term compliance with any exercise plan. In order to derive maximum benefit, activity must become integrated into each patient's daily routine to ultimately be successful.

### ***Strength training and lymphedema***

Lymphedema results from the accumulation of excess fluid in the lymphatic system after breast cancer surgery and/or radiation therapy and may cause pain, swelling, numbness, stiffness, and fatigue in the affected arm [82]. Approximately 30% of breast cancer survivors suffer from lymphedema, which can adversely affect physical and psychosocial well being [83]. Although often incurable, treatments to lessen the physical effects of lymphedema such as manual drainage and limb compression have been developed.

Risk factors for lymphedema include the number of axillary lymph nodes surgically removed or damaged by radiation, obesity at diagnosis or weight gain after treatment, or activities that irritate or inflame the lymphatic system. To decrease risk of lymphedema, a number of breast cancer advocacy groups such as the ACS, Komen for the Cure, and the National Lymphedema Network have recommended limited use of the affected arm. Although these recommendations were designed as risk reduction strategies, limited use may make daily activities more challenging and actually increase risk for injury [84].

Two recent randomized clinical trials challenged the widespread recommendation that lymphedema patients should limit use of the affected/at-risk arm. The Physical Activity and Lymphedema (PAL) Trial randomized 141 breast cancer survivors with lymphedema and 154 survivors at-risk for lymphedema to either a non-exercising control group or to an intervention group who performed progressive weight training twice a week for one year. The intervention group showed greater improvements in self-reported symptoms and greater overall strength [85]. Importantly, progressive weight lifting did not exacerbate lymphedema. In at-risk patients who had five or more lymph nodes surgically removed, weight training had a noticeable effect on outcome — 22% of controls developed symptoms of lymphedema compared to only 7% of patients who participated in weight training [86]. These observations have revolutionized current concepts of lymphedema treatment and suggest that progressive strength training is in fact beneficial to patients.

### **Stress management**

Psychological distress associated with a diagnosis of breast cancer can evoke an array of emotional experiences, ranging from sadness and despair to anxiety and depression [87]. Although emotional symptoms may dissipate, an appreciable number of survivors continue to experience anxiety and depression for years after diagnosis [88]. Psychological stress may influence cancer risk and survival by altering endocrine responses, circadian rhythms, and immune functions (Figure 1).

Breast cancer survivors who have suffered stressful life events, such as divorce or death of a husband or child, had a significantly higher risk of relapse compared to survivors without such stressors [89]. Similarly, women with strong social ties to relatives, friends, and neighbors had significantly lower risk of breast cancer mortality compared to women lacking a social network [90]. In 2,835 women diagnosed with breast cancer, those who felt socially isolated (no close relationships to friends,

relatives, or adult children) before diagnosis had a 66% increased risk of all-cause mortality and a two-fold increase in breast cancer mortality [91]. Therefore, it is important for breast cancer patients to practice effective stress management techniques, develop a social network, and participate in group support sessions to improve prognosis and QOL.

### *Group support*

Breast cancer patients often experience social isolation during months of treatment and recovery, which precipitate feelings of loneliness, low self esteem, body image issues, and negatively impact their QOL [92]. Two types of group support have been evaluated for their ability to improve QOL and survival: (1) cognitive-behavioral therapy, which uses conversation and active participation to resolve dysfunctional emotions and behaviors in individual or group settings; and (2) psycho-educational therapy, which provides cancer education to patients, teaches coping mechanisms, and offers support and resources to improve psychosocial functioning [93]. Some studies suggest that these interventions have a positive effect on cancer patients for certain symptoms such as distress, depression, anxiety, fatigue, and QOL [94-97], while others conclude that psychosocial interventions are ineffective or limited in producing meaningful emotional benefit [98, {Fors, 2011 1415 /id}]. Limited evidence suggests that group support led by a trained psychologist focusing on strategies to reduce stress, improve mood, and maintain adherence to treatment can successfully lower risk of breast cancer recurrence and mortality [100].

Despite the inconclusive nature of current research, psychosocial interventions are rapidly becoming an important component of cancer care. For health care professionals to provide evidence-based guidance to their cancer patients, high quality well-controlled trials with pre-defined outcome measures are needed to determine the precise benefits of psychosocial interventions.

## *Yoga*

Yoga is a mind-body therapy that originated in India more than 5,000 years ago to reduce stress and promote overall physical and mental well-being. The practice of yoga relies on physical postures, focused breathing, and meditation to stretch muscles, control breathing, and minimize stress through visualization techniques and guided imagery. Yoga is believed to offer numerous psychological and physical health benefits to cancer patients [101], including disease prevention, improved cardiovascular health, weight loss, and stress reduction. Cancer survivors, in particular, tend to use complementary and alternative medicine (CAM) approaches [102] to alleviate common symptoms including fatigue, pain, and insomnia [103], which may persist long after treatment has been completed [104-106].

Initial scientific research has shown that yoga confers a variety of benefits to breast cancer patients. A 12-week yoga intervention significantly improved fatigue scores, as well as physical function, mood, and QOL in a small sample of breast cancer survivors [107]. Likewise, a small pilot study of menopausal women who completed a 10-week yoga program reported significant post-treatment improvement in sleep disturbance, sleep efficiency, and quality of sleep attained [108]. Yoga has been shown to improve psychosocial functioning in breast cancer survivors more than four years post-diagnosis [109], as well as those undergoing active clinical treatment [110].

Chronic pain and other sequelae such as paraesthesia (burning or tingling sensation of the skin), allodynia (pain caused by light touch), and lymphedema are significantly more frequent in breast cancer survivors compared to women in the general population [111-113]. Yoga has been shown to alleviate various forms of chronic pain experienced by many breast cancer survivors, including lower back pain [114], pain associated with fibromyalgia [115], and rheumatoid arthritis [116].

Chronic stress is another common ailment that can negatively impact QOL and survival in women with breast cancer. As an increasing number of women turn to CAM to improve their health after breast cancer diagnosis, it is critical to develop effective methods for managing stress. To date, group support has shown mixed overall results in patients suffering from anxiety and depression, but seems to improve survival in patients with the highest levels of anxiety and depression. Yoga has shown promising results in relieving psychological distress and decreasing fatigue, and may be effective in treating chronic pain associated with side effects of treatment. Most studies to date have been too small to adequately define treatment effects; thus, large-scale studies are urgently needed to quantify potential benefits of group support, yoga, and other measures of stress reduction in breast cancer patients.

### **Lifestyle interventions in patients with metastatic breast cancer**

The majority of intervention studies considering the effects of lifestyle factors on breast cancer prognosis often focus on QOL issues, and may suffer from design limitations that restrict their ability to adequately address survival or recurrence in diverse populations of breast cancer patients [117]. Epidemiologic studies often focus on women with early stage breast cancer and do not usually enroll large numbers of patients with late-stage, metastatic disease. Compared to women with localized disease who have a 98% five-year survival rate, women with stage IV breast cancer have five-year survival rates of only 23%. Patients living with metastatic disease have different health concerns than women free from metastasis. For example, women with stage IV breast cancer may need bone strengthening therapy, psychological counseling to deal with social isolation, nutritional counseling to help with appetite and weight loss, treatment for fatigue, and pain management. Intervention studies must be tailored to the specific needs of patients with metastatic disease, recognizing that diet, exercise,

and stress reduction techniques used in other breast cancer survivors may be less desirable or inappropriate.

Despite limited research on patients with metastatic disease, lifestyle interventions have shown benefit in women with stage IV breast cancer. Psychosocial interventions indicate that women randomized to weekly supportive-expressive group therapy demonstrate improved mood and perception of pain [118], as well as a significantly greater decline in mood disturbance and traumatic stress symptoms compared to controls [119]. Likewise, women randomized to receive cognitive therapy for depression had significantly lower levels of depressive symptoms, anxiety, fatigue, and insomnia following treatment compared to those assigned as wait-list controls [120]. Additional studies have shown that other lifestyle interventions, such as yoga-based palliative care, and integrative treatment with nutrition, fitness, and mind-spirit instruction [121], improve QOL and provide survival benefit.

## **Conclusions**

A multitude of studies investigating the impact of lifestyle modification on breast cancer survivors has produced highly variable and contradictory results. Large-scale dietary studies indicate that specific dietary components may not affect breast cancer prognosis, but achieving and/or maintaining a healthy body weight through diet provides health benefits to breast cancer survivors. Studies of physical activity in breast cancer patients have definitively linked physical activity to improved prognosis and survival, and suggest that women with, or at risk for, lymphedema should engage in progressive strength training to increase mobility and decrease risk of developing symptoms. Studies evaluating stress reduction and the effects of group therapy on QOL have been inconclusive, but many studies evaluated small groups of women and lacked adequate statistical power.

Nevertheless, stress reduction techniques may benefit women with the highest levels of anxiety and depression, and yoga in particular, appears to provide numerous health benefits that improve prognosis. These findings illustrate the need for continued, high-quality research into the effects of lifestyle interventions in breast cancer patients in order to provide breast cancer survivors with state-of-the-art knowledge and optimum methods for improving long-term prognosis and QOL.

### **Expert Commentary**

As the number of breast cancer survivors increases, interest in lifestyle intervention and CAM approaches continues to grow. Although the main focus for many breast cancer survivors is prevention of disease recurrence, survivors are at increased risk for developing chronic conditions such as osteoporosis, diabetes, and cardiovascular disease, as well as non-cancer-related mortality [122]. Behavioral modification may provide breast cancer patients with numerous health benefits, both during and after active treatment. In addition to potential health benefits, lifestyle modification may provide patients with feelings of control and self-determination because they become active participants in managing their own health.

Between September 2006 and February 2007, the President's Cancer Panel examined current evidence regarding the effects of lifestyle behaviors on cancer risk, as well as ways to reduce the national cancer burden by promoting healthy lifestyles [204]. The panel recommended that (1) health care providers coordinate and integrate education and prevention messages related to diet, nutrition, physical activity, and tobacco use with other chronic diseases to promote common risk reduction strategies, and (2) individuals assume personal responsibility for learning about cancer risk and make healthy lifestyle choices. Research has shown that breast cancer patients who receive healthy lifestyle recommendations, such as a recommendation to exercise, from their oncologist actually participate in

these behaviors significantly more than patients following standard treatment [123]; however, only about one-fourth of patients report actually receiving dietary or physical activity recommendations from their primary care physicians [124]. Health care providers must become more involved in recommending (and monitoring implementation of) healthy lifestyle behaviors for their cancer patients. Similarly, breast cancer survivors must be proactive in learning about options for healthier lifestyles and integrating appropriate behaviors into their daily routine.

Despite the large number of studies investigating the effects of lifestyle modification on breast cancer survivorship, few definitive conclusions have emerged. Many lifestyle behaviors may individually affect prognosis, but it is unlikely that a single comprehensive “lifestyle modification package” is appropriate for all survivors. Patients should consider general “common-sense” recommendations: (1) maintain a healthy weight by consuming a diet rich in plant-based foods and limited alcohol; (2) be physically active; (3) avoid smoking or second-hand smoke; and (4) minimize stress. Survivors should strive to incorporate as many healthy behaviors as possible, given their individual physical, emotional, and financial constraints.

### **Five-Year View**

Women with a history of breast cancer constitute the largest group of female cancer survivors, accounting for ~22% of all cancer survivors [205]. As the number of breast cancer survivors continues to grow, it is crucial to identify modifiable factors associated with breast cancer recurrence and survival and to fully understand optimum methods for risk factor reduction. This information will likely come from a variety of sources, but should include:

- Well-designed, high-quality randomized trials and observational studies with adequate statistical power — an example is the Pathways Study, a prospective study of breast cancer

survivorship, which is expecting to enroll 3,000 patients within two months of diagnosis to reduce the effect of survival-bias [125]. Recruiting women with late-stage disease, who frequently have different physical and emotional needs than early-stage patients, will provide much needed information for patients with metastasis.

- Comprehensive clinical programs with research components, such as the Clinical Breast Care Project, a breast cancer research program based at Walter Reed National Military Medical Center, which includes a state-of-the-art comprehensive breast center and research capabilities to examine environmental factors associated with prognosis [126].
- Survivors from diverse ethnic and socioeconomic backgrounds and culturally appropriate intervention programs [127], which would examine important ethnic differences in clinical care and cultural factors influencing behavior modification [128].

Specific issues of highest priority that need to be resolved over the next five years are:

- Who would derive the most benefit from, and thus should participate in, behavioral modification programs? Women of advanced age with significant co-morbidities or metastatic disease may not derive meaningful benefit from lifestyle changes, may be physically unable to participate in certain activity routines, and may be reluctant to adopt new dietary habits.
- Can an optimum combination of lifestyle choices can be determined on an individual patient basis that provides survival and QOL benefits that are superior to a standardized approach?
- What is the optimum way to overcome barriers to implementing and maintaining healthy lifestyle behaviors? Research has shown that (1) age [129] and ethnic [130] differences exist in the likelihood that survivors will adopt healthy lifestyle behaviors, and (2) patients may

not fully adhere over the longer-term to self-directed interventions and may miss potential health benefits [131]. Healthy lifestyle programs need to be developed for a diverse population of breast cancer survivors that have wider adaptability than those presently available.

- Evolution of primary care: Medical providers should not discount CAM approaches, given that a majority of patients believe that these alternative therapies improve their QOL. Physicians should be powerful catalysts for promoting behavior change in their patients, because they are optimally positioned to deliver guidance regarding health promotion.
- Do interactions between molecular factors and behaviors influence breast cancer recurrence and survival? Identifying genetic variation that may enhance or inhibit responses to lifestyle intervention would add an additional level of personalization to patient care.

Finally, integrating healthy behaviors into the daily lives of breast cancer patients faces the ongoing challenge of accessibility to health resources and cost. At present, the cost of treatment for metastatic breast cancer is \$128,556 (95% CI=\$118,409-\$137,644) per patient, and out-of-pocket expenses can reach \$1,500 per month [10]. Additional costs for incorporating healthy lifestyle components that use providers who charge for services [132], such as gym memberships, yoga classes, exercise physiologists, and dietician consultations may be prohibitive to many breast cancer survivors.

The Patient Protection and Affordable Care Act and the Health Care and Education Reconciliation Act were passed in March 2010 [206]. This long overdue overhaul of the American healthcare system is based on a “triple aim” philosophy (improved health in the general population, better patient experiences, and lower per-capita cost) that eventually may translate into improved care

for breast cancer patients. Under this new legislation, prevention will assume a greater role in health care [133,134], and evidence-based intervention programs may become covered services.

### **Key Issues**

- Due mainly to advances in early detection and treatment, there are an estimated 2.5 million breast cancer survivors in the US alone, and this number is projected to increase to 3.4 million within the next four years. Behavioral modification has great potential to improve prognosis and QOL in breast cancer survivors.
- Achieving or maintaining a healthy weight after diagnosis is associated with improved prognosis. A healthy diet may not affect breast cancer prognosis, but has been associated with increased overall survival.
- Limited research suggests that certain types of alcohol, vitamin supplements, soy, and green tea may have some beneficial properties for breast cancer survivors; however, large-scale studies are needed to provide conclusive results and specific recommendations for consumption.
- Moderate physical activity can ameliorate side effects during treatment and improve breast cancer prognosis and overall survival after treatment.
- Contrary to the long-held belief that patients affected with, or at-risk for, lymphedema should limit use of the affected/at-risk arm, progressive supervised strength training can decrease symptoms and improve outcomes.
- Stress reduction techniques such as group support and yoga have shown promising results in improving outcomes and QOL, but further research is essential for quantifying potential benefits in breast cancer patients.

- As more effective healthy lifestyle interventions evolve, important issues such as who will derive the most benefit, when lifestyle interventions should be initiated, how and where programs should be facilitated, and who will pay for healthy lifestyle programs must be resolved in order for lifestyle modification to have a meaningful impact on the lives of breast cancer survivors.

## References

- 1 American Cancer Society. *Cancer facts & figures 2011*. American Cancer Society, Atlanta, GA, USA (2011).
- 2 Jemal A, Ward E, Thun M. Declining death rates reflect progress against cancer. *Plos ONE* 5, e9584 (2010).
- 3 Berry DA, Cronin KA, Plevritis SK *et al*. Effect of screening and adjuvant therapy on mortality from breast cancer. *N. Engl. J. Med.* 353, 1784-1792 (2005).
- 4 Morgan MA. Cancer survivorship: history, quality-of-life issues, and the evolving multidisciplinary approach to implementation of cancer survivorship care plans. *Oncol. Nurs. Forum* 36, 429-436 (2009).
- 5 Stuart K, Brennan M, French J, Houssami N, Boyages J. Life after breast cancer. *Aust. Fam. Physician* 35, 219-224 (2006).
- 6 Rozenberg S, Antoine C, Carly B, Pastijn A, Liebens F. Improving quality of life after breast cancer: prevention of other diseases. *Menopause Int.* 13, 71-74 (2007).
- 7 Montazeri A. Health-related quality of life in breast cancer patients: a bibliographic review of the literature from 1974-2007. *J. Exp. Clin. Cancer Res.* 27, 32 (2008).
- 8 De Angelis R, Tavilla A, Verdecchia A *et al*. Breast cancer survivors in the United States: geographic variability and time trends, 2005-2015. *Cancer* 115, 1954-1966 (2009).
- 9 Garrison LP, Jr., Veenstra DL. The economic value of innovative treatments over the product life cycle: the case of targeted trastuzumab therapy for breast cancer. *Value Health* 12, 1118-1123 (2009).
- 10 Pisu M, Azuero A, McNees P, Burkhardt J, Benz R, Meneses K. The out of pocket cost of breast cancer survivors: a review. *J. Cancer Surviv.* 4, 202-209 (2010).

- 11 Cordain L, Eaton SB, Sebastian A *et al.* Origins and evolution of the Western diet: health implications for the 21st century. *Am. J. Clin. Nutr.* 81, 341-354 (2005).
- 12 Nelson NJ. Migrant studies aid the search for factors linked to breast cancer risk. *J. Natl. Cancer Inst.* 98, 436-438 (2006).
- 13 Yang W, Zilov A, Soewondo P, Bech OM, Sekkal F, Home PD. Observational studies: going beyond the boundaries of randomized controlled trials. *Diabetes Res. Clin. Pract.* 88 Suppl 1, S3-S9 (2010).
- 14 Silverman SL. From randomized controlled trials to observational studies. *Am. J. Med.* 122, 114-120 (2009).
- 15 Stolberg HO, Norman G, Trop I. Randomized controlled trials. *Am. J. Roentgenol.* 183, 1539-1544 (2004).
- 16 Majed B, Moreau T, Senouci K, Salmon RJ, Fourquet A, Asselain B. Is obesity an independent prognosis factor in woman breast cancer? *Breast Cancer Res. Treat.* 111, 329-342 (2008).
- 17 Demark-Wahnefried W, Winer EP, Rimer BK. Why women gain weight with adjuvant chemotherapy for breast cancer. *J. Clin. Oncol.* 11, 1418-1429 (1993).
- 18 Nichols HB, Trentham-Dietz A, Egan KM *et al.* Body mass index before and after breast cancer diagnosis: associations with all-cause, breast cancer, and cardiovascular disease mortality. *Cancer Epidemiol. Biomarkers Prev.* 18, 1403-1409 (2009).
- 19 Wang Y, Beydoun MA. The obesity epidemic in the United States — gender, age, socioeconomic, racial/ethnic, and geographic characteristics: a systematic review and meta-regression analysis. *Epidemiol. Rev.* 29, 6-28 (2007).
- 20 Jain M, Miller AB, To T. Premorbid diet and the prognosis of women with breast cancer. *J. Natl. Cancer Inst.* 86, 1390-1397 (1994).
- 21 Holmes MD, Stampfer MJ, Colditz GA, Rosner B, Hunter DJ, Willet WC. Dietary factors and the survival of women with breast carcinoma. *Cancer* 86, 826-835 (1999).
- 22 Kroenke CH, Fung TT, Hu FB, Holmes MD. Dietary patterns and survival after breast cancer diagnosis. *J. Clin. Oncol.* 23, 9295-9303 (2005).
- 23 Kroenke CH, Chen WY, Rosner B, Holmes MD. Weight, weight gain, and survival after breast cancer diagnosis. *J. Clin. Oncol.* 23, 1370-1378 (2005).
- 24 Caan B, Sternfeld B, Gunderson E, Coates A, Quesenberry C, Slattery ML. Life After Cancer Epidemiology (LACE) Study: a cohort of early stage breast cancer survivors (United States). *Cancer Causes Control* 16, 545-556 (2005).

- 25 Kwan ML, Weltzien E, Kushi LH, Castillo A, Slattery ML, Caan BJ. Dietary patterns and breast cancer recurrence and survival among women with early-stage breast cancer. *J. Clin. Oncol.* 27, 919-926 (2009).
- 26 Irwin ML, McTiernan A, Baumgartner RN *et al.* Changes in body fat and weight after a breast cancer diagnosis: influence of demographic, prognostic, and lifestyle factors. *J. Clin. Oncol.* 23, 774-782 (2005).
- 27 George SM, Irwin ML, Smith AW *et al.* Postdiagnosis diet quality, the combination of diet quality and recreational physical activity, and prognosis after early-stage breast cancer. *Cancer Causes Control* 22, 589-598 (2011).
- 28 Wayne SJ, Baumgartner K, Baumgartner RN, Bernstein L, Bowen DJ, Ballard-Barbash R. Diet quality is directly associated with quality of life in breast cancer survivors. *Breast Cancer Res. Treat.* 96, 227-232 (2006).
- 29 George SM, Neuhauser ML, Mayne ST *et al.* Postdiagnosis diet quality is inversely related to a biomarker of inflammation among breast cancer survivors. *Cancer Epidemiol. Biomarkers Prev.* 19, 2220-2228 (2010).
- 30 Chlebowski RT, Blackburn GL, Thomson CA *et al.* Dietary fat reduction and breast cancer outcome: interim efficacy results from the Women's Intervention Nutrition Study. *J. Natl. Cancer Inst.* 98, 1767-1776 (2006).
- \*\* **Observed significant improvement in prognosis of breast cancer patients on a low-fat diet.**
- 31 Pierce JP, Natarajan L, Caan BJ *et al.* Influence of a diet very high in vegetables, fruit, and fiber and low in fat on prognosis following treatment for breast cancer: the Women's Healthy Eating and Living (WHEL) randomized trial. *JAMA* 298, 289-298 (2007).
- \*\* **Detected no association between a low-fat, high fruit and vegetable diet and outcomes in breast cancer patients.**
- 32 Hamajima N, Hirose K, Tajima K *et al.* Alcohol, tobacco and breast cancer — collaborative reanalysis of individual data from 53 epidemiological studies, including 58,515 women with breast cancer and 95,067 women without the disease. *Br. J. Cancer* 87, 1234-1245 (2002).
- 33 Reding KW, Daling JR, Doody DR, O'Brien CA, Porter PL, Malone KE. Effect of prediagnostic alcohol consumption on survival after breast cancer in young women. *Cancer Epidemiol. Biomarkers Prev.* 17, 1988-1996 (2008).
- 34 Barnett GC, Shah M, Redman K, Easton DF, Ponder BA, Pharoah PD. Risk factors for the incidence of breast cancer: do they affect survival from the disease? *J. Clin. Oncol.* 26, 3310-3316 (2008).

- 35 Kwan ML, Kushi LH, Weltzien E *et al.* Alcohol consumption and breast cancer recurrence and survival among women with early-stage breast cancer: the life after cancer epidemiology study. *J. Clin. Oncol.* 28, 4410-4416 (2010).
- 36 Franceschi S, Dal Maso L, Zucchetto A, Talamini R, Prospective Analysis of Case-control studies on Environmental factors and health (PACE) study group. Alcohol consumption and survival after breast cancer. *Cancer Epidemiol. Biomarkers Prev.* 18, 1011-1012 (2009).
- 37 Flatt SW, Thomson CA, Gold EB *et al.* Low to moderate alcohol intake is not associated with increased mortality after breast cancer. *Cancer Epidemiol. Biomarkers Prev.* 19, 681-688 (2010).
- 38 Espin JC, Garcia-Conesa MT, Tomas-Barberan FA. Nutraceuticals: facts and fiction. *Phytochemistry* 68, 2986-3008 (2007).
- 39 Velicer CM, Ulrich CM. Vitamin and mineral supplement use among US adults after cancer diagnosis. A systematic review. *J. Clin. Oncol.* 26, 665-673 (2008).
- 40 Nesheim MC. What is the research base for the use of dietary supplements? *Public Health Nutr.* 2, 35-38 (1999).
- 41 Morris CD, Carson S. Routine vitamin supplementation to prevent cardiovascular disease: a summary of the evidence for the U.S. Preventive Services Task Force. *Ann. Intern. Med.* 139, 56-70 (2003).
- 42 NIH State-of-the-Science Panel. National Institutes of Health State-of-the-Science Conference Statement: multivitamin/mineral supplements and chronic disease prevention. *Ann. Intern. Med.* 145, 364-371 (2006).
- 43 Ishitani K, Lin J, Manson JE, Buring JE, Zhang SM. A prospective study of multivitamin supplement use and risk of breast cancer. *Am. J. Epidemiol.* 167, 1197-1206 (2008).
- 44 Neuhaus ML, Wassertheil-Smoller S, Thomson C *et al.* Multivitamin use and risk of cancer and cardiovascular disease in the Women's Health Initiative cohorts. *Arch. Intern. Med.* 169, 294-304 (2009).
- 45 Sporn MB, Roberts AB, Goodman DS. *The Retinoids: Biology, Chemistry, and Medicine.* 2nd ed. Raven Press, New York, NY, USA (1994).
- 46 Mehta RR, Hart G, Beattie CW, Das Gupta TK. Significance of plasma retinol binding protein levels in recurrence of breast tumors in women. *Oncology* 44, 350-355 (1987).
- 47 Formelli F, Meneghini E, Cavadini E *et al.* Plasma retinol and prognosis of postmenopausal breast cancer patients. *Cancer Epidemiol. Biomarkers Prev.* 18, 42-48 (2009).

- 48 McEligot AJ, Largent J, Ziogas A, Peel D, Anton-Culver H. Dietary fat, fiber, vegetable, and micronutrients are associated with overall survival in postmenopausal women diagnosed with breast cancer. *Nutr. Cancer* 55, 132-140 (2006).
- 49 Xu X, Gammon MD, Wetmur JG *et al.* B-vitamin intake, one-carbon metabolism and survival among a population-based study of women with breast cancer. *Cancer Epidemiol. Biomarkers Prev.* 17, 2109-2116 (2008).
- 50 Nechuta S, Lu W, Chen Z *et al.* Vitamin supplement use during breast cancer treatment and survival: a prospective cohort study. *Cancer Epidemiol. Biomarkers Prev.* 20, 262-271 (2011).
- 51 Kline K, Yu W, Sanders BG. Vitamin E and breast cancer. *J. Nutr.* 134, 3458S-3462S (2004).
- 52 Garland FC, Garland CF, Gorham ED, Young JF. Geographic variation in breast cancer mortality in the United States: a hypothesis involving exposure to solar radiation. *Prev. Med.* 19, 614-622 (1990).
- 53 Robsahm TE, Tretli S, Dahlback A, Moan J. Vitamin D3 from sunlight may improve the prognosis of breast-, colon- and prostate cancer (Norway). *Cancer Causes Control* 15, 149-158 (2004).
- 54 Palmieri C, MacGregor T, Girgis S, Vigushin D. Serum 25-hydroxyvitamin D levels in early and advanced breast cancer. *J. Clin. Pathol.* 59, 1334-1336 (2006).
- 55 Jacobs ET, Thomson CA, Flatt SW *et al.* Vitamin D and breast cancer recurrence in the Women's Healthy Eating and Living (WHEL) Study. *Am. J. Clin. Nutr.* 93, 108-117 (2011).
- 56 Cabrera C, Artacho R, Giménez R. Beneficial effects of green tea — a review. *J. Am. Coll. Nutr.* 25, 79-99 (2006).
- 57 Nakachi K, Suemasu K, Suga K, Takeo T, Imai K, Higashi Y. Influence of drinking green tea on breast cancer malignancy among Japanese patients. *Jpn. J. Cancer Res.* 89, 254-261 (1998).
- 58 Inoue M, Tajima K, Mizutani M *et al.* Regular consumption of green tea and the risk of breast cancer recurrence: follow-up study from the Hospital-based Epidemiologic Research Program at Aichi Cancer Center (HERPACC), Japan. *Cancer Lett.* 167, 175-182 (2001).
- 59 Ogunleye AA, Xue F, Michels KB. Green tea consumption and breast cancer risk or recurrence: a meta-analysis. *Breast Cancer Res. Treat.* 119, 477-484 (2010).
- 60 Allred CD, Allred KF, Ju YH, Virant SM, Helferich WG. Soy diets containing varying amounts of genistein stimulate growth of estrogen-dependent (MCF-7) tumors in a dose-dependent manner. *Cancer Res.* 61, 5045-5050 (2001).
- 61 Ju YH, Doerge DR, Allred KF, Allred CD, Helferich WG. Dietary genistein negates the inhibitory effect of tamoxifen on growth of estrogen-dependent human breast cancer (MCF-7) cells implanted in athymic mice. *Cancer Res.* 62, 2474-2477 (2002).

- 62 Fink BN, Steck SE, Wolff MS *et al.* Dietary flavonoid intake and breast cancer survival among women on Long Island. *Cancer Epidemiol. Biomarkers Prev.* 16, 2285-2292 (2007).
- 63 Guha N, Kwan ML, Quesenberry CP, Weltzien EK, Castillo AL, Caan BJ. Soy isoflavones and risk of cancer recurrence in a cohort of breast cancer survivors: the Life After Cancer Epidemiology study. *Breast Cancer Res. Treat.* 118, 395-405 (2009).
- 64 Kang X, Zhang Q, Wang S, Huang X, Jin S. Effect of soy isoflavones on breast cancer recurrence and death for patients receiving adjuvant endocrine therapy. *CMAJ* 182, 1857-1862 (2010).
- 65 McTiernan A, Kooperberg C, White E *et al.* Recreational physical activity and the risk of breast cancer in postmenopausal women: the Women's Health Initiative Cohort Study. *JAMA* 290, 1331-1336 (2003).
- 66 Friedenreich CM, Cust AE. Physical activity and breast cancer risk: impact of timing, type and dose of activity and population subgroup effects. *Br. J. Sports Med.* 42, 636-647 (2008).
- 67 George SM, Irwin ML, Matthews CE *et al.* Beyond recreational physical activity: examining occupational and household activity, transportation activity, and sedentary behavior in relation to postmenopausal breast cancer risk. *Am. J. Public Health* 100, 2288-2295 (2010).
- 68 Mock V, Burke MB, Sheehan P *et al.* A nursing rehabilitation program for women with breast cancer receiving adjuvant chemotherapy. *Oncol. Nurs. Forum* 21, 899-907 (1994).
- 69 Dimeo FC, Stieglitz RD, Novelli-Fischer U, Fetscher S, Keul J. Effects of physical activity on the fatigue and psychologic status of cancer patients during chemotherapy. *Cancer* 85, 2273-2277 (1999).
- 70 Segal R, Evans W, Johnson D *et al.* Structured exercise improves physical functioning in women with stages I and II breast cancer: results of a randomized controlled trial. *J. Clin. Oncol.* 19, 657-665 (2001).
- 71 Campbell A, Mutrie N, White F, McGuire F, Kearney N. A pilot study of a supervised group exercise programme as a rehabilitation treatment for women with breast cancer receiving adjuvant treatment. *Eur. J. Oncol. Nurs.* 9, 56-63 (2005).
- 72 Hsieh CC, Sprod LK, Hydock DS, Carter SD, Hayward R, Schneider CM. Effects of a supervised exercise intervention on recovery from treatment regimens in breast cancer survivors. *Oncol. Nurs. Forum* 35, 909-915 (2008).
- 73 Irwin ML, Crumley D, McTiernan A *et al.* Physical activity levels before and after a diagnosis of breast carcinoma: the Health, Eating, Activity, and Lifestyle (HEAL) study. *Cancer* 97, 1746-1757 (2003).
- 74 Devoogdt N, Van Kampen M, Geraerts I *et al.* Physical activity levels after treatment for breast cancer: one-year follow-up. *Breast Cancer Res. Treat.* 123, 417-425 (2010).

- 75 Segar ML, Katch VL, Roth RS *et al.* The effect of aerobic exercise on self-esteem and depressive and anxiety symptoms among breast cancer survivors. *Oncol. Nurs. Forum* 25, 107-113 (1998).
- 76 Courneya KS, Mackey JR, Bell GJ, Jones LW, Field CJ, Fairey AS. Randomized controlled trial of exercise training in postmenopausal breast cancer survivors: cardiopulmonary and quality of life outcomes. *J. Clin. Oncol.* 21, 1660-1668 (2003).
- 77 Holmes MD, Chen WY, Feskanich D, Kroenke CH, Colditz GA. Physical activity and survival after breast cancer diagnosis. *JAMA* 293, 2479-2486 (2005).
- 78 Sternfeld B, Weltzien E, Quesenberry CP Jr *et al.* Physical activity and risk of recurrence and mortality in breast cancer survivors: findings from the LACE study. *Cancer Epidemiol. Biomarkers Prev.* 18, 87-95 (2009).
- 79 Holick CN, Newcomb PA, Trentham-Dietz A *et al.* Physical activity and survival after diagnosis of invasive breast cancer. *Cancer Epidemiol. Biomarkers Prev.* 17, 379-386 (2008).
- 80 Irwin ML, Smith AW, McTiernan A *et al.* Influence of pre- and post-diagnosis physical activity on mortality in breast cancer survivors: the Health, Eating, Activity, and Lifestyle study. *J. Clin. Oncol.* 26, 3958-3964 (2008).
- \* **Description of healthy benefits of exercise before and after breast cancer diagnosis.**
- 81 Byers T, Nestle M, McTiernan A *et al.* American Cancer Society guidelines on nutrition and physical activity for cancer prevention: Reducing the risk of cancer with healthy food choices and physical activity. *CA. Cancer J. Clin.* 52, 92-119 (2002).
- 82 Fu MR, Rosedale M. Breast cancer survivors' experiences of lymphedema-related symptoms. *J. Pain Symptom. Manage.* 38, 849-859 (2009).
- 83 Norman SA, Localio AR, Kallan MJ *et al.* Risk factors for lymphedema after breast cancer treatment. *Cancer Epidemiol. Biomarkers Prev.* 19, 2734-2746 (2010).
- 84 Schmitz KH. Balancing lymphedema risk: exercise versus deconditioning for breast cancer survivors. *Exerc. Sport Sci. Rev.* 38, 17-24 (2010).
- 85 Schmitz KH, Troxel AB, Cheville A *et al.* Physical Activity and Lymphedema (the PAL trial): assessing the safety of progressive strength training in breast cancer survivors. *Contemp. Clin. Trials* 30, 233-245 (2009).
- \*\* **Sentinel study evaluating progressive, supervised weight training in women with lymphedema; led to paradigm shift in activity recommendations for lymphedema patients.**
- 86 Schmitz KH, Ahmed RL, Troxel AB *et al.* Weight lifting for women at risk for breast cancer-related lymphedema: a randomized trial. *JAMA* 304, 2699-2705 (2010).

- \*\* Sentinel study evaluating progressive, supervised weight training in women at risk for lymphedema; led to paradigm shift in treatment for patients who have had axillary lymph node removal, but have not developed lymphedema.**
- 87 Zabora J, BrintzenhofeSzoc K, Curbow B, Hooker C, Piantadosi S. The prevalence of psychological distress by cancer site. *Psychooncology* 10, 19-28 (2001).
- 88 Spiegel D. Psychosocial aspects of breast cancer treatment. *Semin. Oncol.* 24, S1-36-S1-47 (1997).
- 89 Ramirez AJ, Craig TK, Watson JP, Fentiman IS, North WR, Rubens RD. Stress and relapse of breast cancer. *BMJ* 298, 291-293 (1989).
- 90 Waxler-Morrison N, Hislop TG, Mears B, Kan L. Effects of social relationships on survival for women with breast cancer: a prospective study. *Soc. Sci. Med.* 33, 177-183 (1991).
- 91 Kroenke CH, Kubzansky LD, Schernhammer ES, Holmes MD, Kawachi I. Social networks, social support, and survival after breast cancer diagnosis. *J. Clin. Oncol.* 24, 1105-1111 (2006).
- \* Large study demonstrating detrimental effects of social isolation in breast cancer survivors.**
- 92 Rosedale M. Survivor loneliness of women following breast cancer. *Oncol. Nurs. Forum* 36, 175-183 (2009).
- 93 Jacobsen PB, Jim HS. Psychosocial interventions for anxiety and depression in adult cancer patients: achievements and challenges. *CA. Cancer J. Clin.* 58, 214-230 (2008).
- 94 Osborn RL, Demoncada AC, Feuerstein M. Psychosocial interventions for depression, anxiety, and quality of life in cancer survivors: meta-analyses. *Int. J. Psychiatry Med.* 36, 13-34 (2006).
- 95 Tatrow K, Montgomery GH. Cognitive behavioral therapy techniques for distress and pain in breast cancer patients: a meta-analysis. *J. Behav. Med.* 29, 17-27 (2006).
- 96 Zimmermann T, Heinrichs N, Baucom DH. "Does one size fit all?" moderators in psychosocial interventions for breast cancer patients: a meta-analysis. *Ann. Behav. Med.* 34, 225-239 (2007).
- 97 Duijts SF, Faber MM, Oldenburg HS, van BM, Aaronson NK. Effectiveness of behavioral techniques and physical exercise on psychosocial functioning and health-related quality of life in breast cancer patients and survivors — a meta-analysis. *Psychooncology* 20, 115-126 (2011).
- 98 Coyne JC, Stefanek M, Palmer SC. Psychotherapy and survival in cancer: the conflict between hope and evidence. *Psychol. Bull.* 133, 367-394 (2007).
- 99 Fors EA, Bertheussen GF, Thune I *et al.* Psychosocial interventions as part of breast cancer rehabilitation programs? Results from a systematic review. *Psychooncology* 20, 909-918 (2011).

- 100 Andersen BL, Yang HC, Farrar WB *et al.* Psychologic intervention improves survival for breast cancer patients: a randomized clinical trial. *Cancer* 113, 3450-3458 (2008).
- 101 Lin KY, Hu YT, Chang KJ, Lin HF, Tsauo JY. Effects of yoga on psychological health, quality of life, and physical health of patients with cancer: a meta-analysis. *Evid. Based Complement. Alternat. Med.* 2011, 659876 (2011).
- 102 Mao JJ, Palmer CS, Healy KE, Desai K, Amsterdam J. Complementary and alternative medicine use among cancer survivors: a population-based study. *J. Cancer Surviv.* 5, 8-17 (2011).
- 103 Lee ES, Lee MK, Kim SH *et al.* Health-related quality of life in survivors with breast cancer 1 year after diagnosis compared with the general population: a prospective cohort study. *Ann. Surg.* 253, 101-108 (2011).
- 104 Bower JE, Ganz PA, Desmond KA *et al.* Fatigue in long-term breast carcinoma survivors: a longitudinal investigation. *Cancer* 106, 751-758 (2006).
- 105 Minton O, Stone P. How common is fatigue in disease-free breast cancer survivors? A systematic review of the literature. *Breast Cancer Res. Treat.* 112, 5-13 (2008).
- 106 Otte JL, Carpenter JS, Russell KM, Bigatti S, Champion VL. Prevalence, severity, and correlates of sleep-wake disturbances in long-term breast cancer survivors. *J. Pain Symptom Manage.* 39, 535-547 (2010).
- 107 Bower JE, Garet D, Sternlieb B. Yoga for persistent fatigue in breast cancer survivors: results of a pilot study. *Evid. Based Complement. Alternat. Med.* 2011, 623168 (2011).
- 108 Booth-LaForce C, Thurston RC, Taylor MR. A pilot study of a Hatha yoga treatment for menopausal symptoms. *Maturitas* 57, 286-295 (2007).
- 109 Culos-Reed SN, Carlson LE, Daroux LM, Hatley-Aldous S. A pilot study of yoga for breast cancer survivors: physical and psychological benefits. *Psychooncology* 15, 891-897 (2006).
- 110 Vadiraja HS, Raghavendra RM, Nagarathna R *et al.* Effects of a yoga program on cortisol rhythm and mood states in early breast cancer patients undergoing adjuvant radiotherapy: a randomized controlled trial. *Integr. Cancer Ther.* 8, 37-46 (2009).
- 111 Gulluoglu BM, Cingi A, Cakir T, Gercek A, Barlas A, Eti Z. Factors related to post-treatment chronic pain in breast cancer survivors: the interference of pain with life functions. *Int. J. Fertil. Womens Med.* 51, 75-82 (2006).
- 112 Mao JJ, Armstrong K, Bowman MA, Xie SX, Kadakia R, Farrar JT. Symptom burden among cancer survivors: impact of age and comorbidity. *J. Am. Board Fam. Med.* 20, 434-443 (2007).
- 113 Peuckmann V, Ekholm O, Rasmussen NK *et al.* Chronic pain and other sequelae in long-term breast cancer survivors: nationwide survey in Denmark. *Eur. J. Pain* 13, 478-485 (2009).

- 114 Williams K, Abildso C, Steinberg L *et al.* Evaluation of the effectiveness and efficacy of Iyengar yoga therapy on chronic low back pain. *Spine (Phila Pa 1976)* 34, 2066-2076 (2009).
- 115 Carson JW, Carson KM, Porter LS, Keefe FJ, Shaw H, Miller JM. Yoga for women with metastatic breast cancer: results from a pilot study. *J. Pain Symptom Manage.* 33, 331-341 (2007).
- 116 Bosch PR, Traustadottir T, Howard P, Matt KS. Functional and physiological effects of yoga in women with rheumatoid arthritis: a pilot study. *Altern. Ther. Health Med.* 15, 24-31 (2009).
- 117 Kushi LH, Kwan ML, Lee MM, Ambrosone CB. Lifestyle factors and survival in women with breast cancer. *J. Nutr.* 137, 236S-242S (2007).
- 118 Goodwin PJ, Leszcz M, Ennis M *et al.* The effect of group psychosocial support on survival in metastatic breast cancer. *N. Engl. J. Med.* 345, 1719-1726 (2001).
- 119 Classen C, Butler LD, Koopman C *et al.* Supportive-expressive group therapy and distress in patients with metastatic breast cancer: a randomized clinical intervention trial. *Arch. Gen. Psychiatry* 58, 494-501 (2001).
- 120 Savard J, Simard S, Giguere I *et al.* Randomized clinical trial on cognitive therapy for depression in women with metastatic breast cancer: psychological and immunological effects. *Palliat. Support. Care* 4, 219-237 (2006).
- 121 Block KI, Gyllenhaal C, Tripathy D *et al.* Survival impact of integrative cancer care in advanced metastatic breast cancer. *Breast J.* 15, 357-366 (2009).
- \* **Describes beneficial effects of an integrative approach encompassing diet, exercise, and stress reduction in patients with advanced metastatic disease.**
- 122 Demark-Wahnefried W, Aziz NM, Rowland JH, Pinto BM. Riding the crest of the teachable moment: promoting long-term health after the diagnosis of cancer. *J. Clin. Oncol.* 23, 5814-5830 (2005).
- 123 Jones LW, Courneya KS, Fairey AS, Mackey JR. Effects of an oncologist's recommendation to exercise on self-reported exercise behavior in newly diagnosed breast cancer survivors: a single-blind, randomized controlled trial. *Ann. Behav. Med.* 28, 105-113 (2004).
- 124 Sabatino SA, Coates RJ, Uhler RJ, Pollack LA, Alley LG, Zauderer LJ. Provider counseling about health behaviors among cancer survivors in the United States. *J. Clin. Oncol.* 25, 2100-2106 (2007).
- 125 Kwan ML, Ambrosone CB, Lee MM *et al.* The Pathways Study: a prospective study of breast cancer survivorship within Kaiser Permanente Northern California. *Cancer Causes Control* 19, 1065-1076 (2008).

- 126 Ellsworth RE, Zhu K, Bronfman L, Gutchell V, Hooke JA, Shriver CD. The Clinical Breast Care Project: an important resource in investigating environmental and genetic contributions to breast cancer in African American women. *Cell Tissue Bank*. 9, 109-120 (2008).
- 127 Kelley MA. Culturally appropriate breast health educational intervention program for African-American women. *J. Natl. Black Nurses Assoc.* 15, 36-47 (2004).
- 128 Ashing-Giwa KT, Padilla G, Tejero J *et al.* Understanding the breast cancer experience of women: a qualitative study of African American, Asian American, Latina and Caucasian cancer survivors. *Psychooncology* 13, 408-428 (2004).
- 129 Patterson RE, Neuhouser ML, Hedderson MM, Schwartz SM, Standish LJ, Bowen DJ. Changes in diet, physical activity, and supplement use among adults diagnosed with cancer. *J. Am. Diet. Assoc.* 103, 323-328 (2003).
- 130 Demark-Wahnefried W, Peterson B, McBride C, Lipkus I, Clipp E. Current health behaviors and readiness to pursue life-style changes among men and women diagnosed with early stage prostate and breast carcinomas. *Cancer* 88, 674-684 (2000).
- 131 Pinto BM, Rabin C, Papandonatos GD, Frierson GM, Trunzo JJ, Marcus BH. Maintenance of effects of a home-based physical activity program among breast cancer survivors. *Support. Care Cancer* 16, 1279-1289 (2008).
- 132 Patterson RE, Neuhouser ML, Hedderson MM *et al.* Types of alternative medicine used by patients with breast, colon, or prostate cancer: predictors, motives, and costs. *J. Altern. Complement. Med.* 8, 477-485 (2002).
- 133 Schwartz K, Claxton G. The patient protection and affordable care act: how will it affect private health insurance for cancer patients? *Cancer J.* 16, 572-576 (2010).
- 134 Albright HW, Moreno M, Feeley TW *et al.* The implications of the 2010 Patient Protection and Affordable Care Act and the Health Care and Education Reconciliation Act on cancer care delivery. *Cancer* 117, 1564-1574 (2011).

### **Websites**

- 201 American Cancer Society. Learn About Cancer: Breast Cancer  
[www.cancer.org/cancer/breastcancer/detailedguide/breast-cancer-key-statistics](http://www.cancer.org/cancer/breastcancer/detailedguide/breast-cancer-key-statistics)
- 202 National Cancer Institute. Science Serving People: The Cost of Cancer  
[www.cancer.gov/aboutnci/servingpeople/cancer-statistics/costofcancer](http://www.cancer.gov/aboutnci/servingpeople/cancer-statistics/costofcancer)

- 203 US Department of Health and Human Services: Physical Activity Guidelines for Americans  
[www.health.gov/paguidelines/guidelines/default.aspx](http://www.health.gov/paguidelines/guidelines/default.aspx)
- 204 U.S. Department of Health and Human Services, National Institutes of Health, National Cancer Institute. President's Cancer Panel, 2006-2007 Annual Report. Promoting Healthy Lifestyles: Policy, Program, and Personal Recommendations for Reducing Cancer Risk  
<http://deainfo.nci.nih.gov/advisory/pcp/annualReports/pcp07rpt/pcp07rpt.pdf>
- 205 National Cancer Institute. Cancer Survivorship Research: Cancer Control and Population Sciences  
[www.cancercontrol.cancer.gov/ocs/](http://www.cancercontrol.cancer.gov/ocs/)
- 206 HealthCare.gov. Preventive Services Covered under the Affordable Care Act  
[www.healthcare.gov/law/about/provisions/services/lists.html](http://www.healthcare.gov/law/about/provisions/services/lists.html)

## **Acknowledgements**

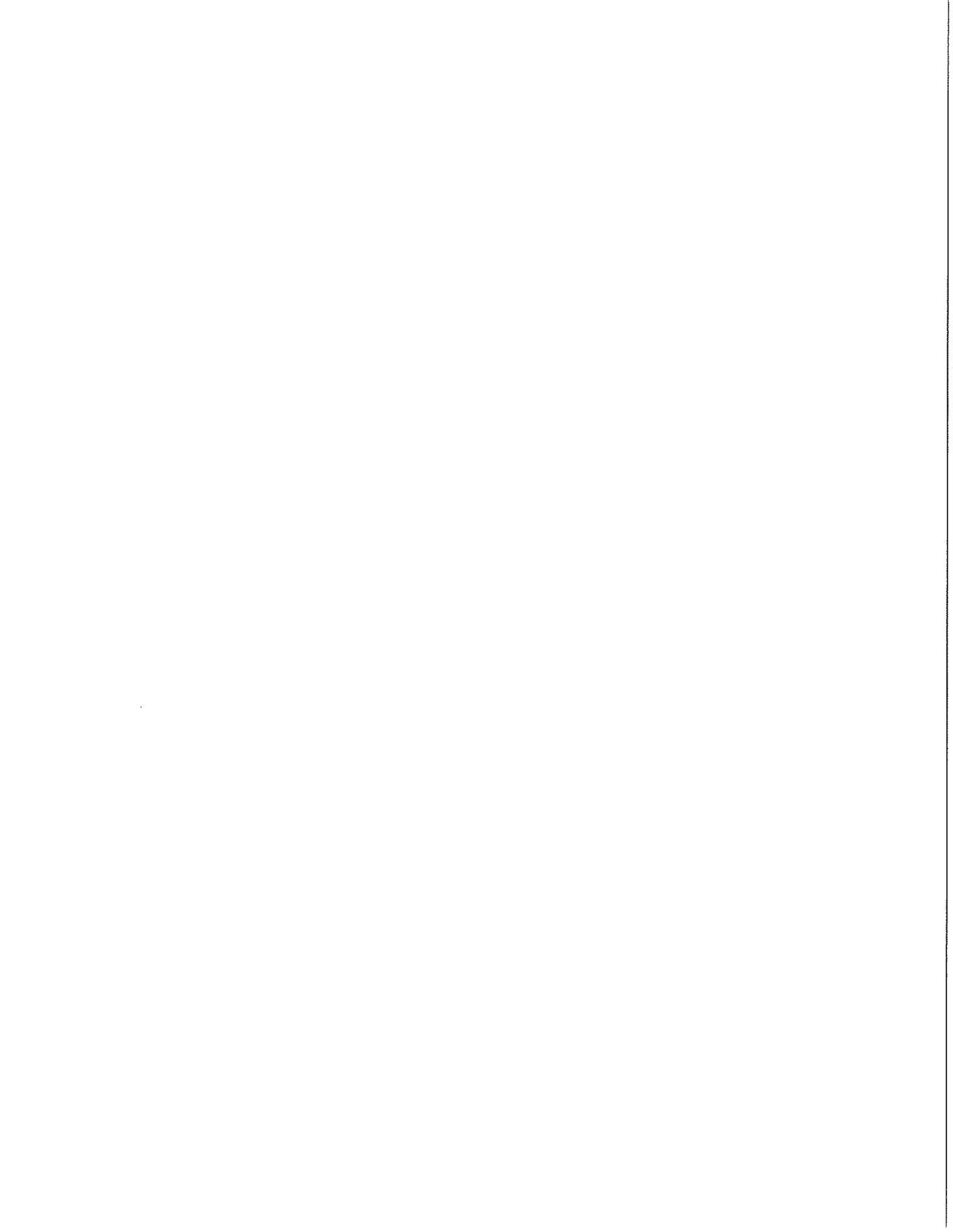
This work was supported by the United States Department of Defense (Military Molecular Medicine Initiative MDA W81XWH-05-2-0075, protocol #01-20006) and was performed under the auspices of the Clinical Breast Care Project, a joint effort of many investigators and staff members whose contributions are gratefully acknowledged. The opinion and assertions contained herein are the private views of the authors and are not to be construed as official or as representing the views of the Department of the Army or the Department of Defense.

## **Figure Legend**

### **Figure 1. Mechanisms by which physiological stress may impact tumorigenesis.**

Stress resulting from a diagnosis of breast cancer may lead to secretion of hormones from the central nervous system, which alters the immune response and promotes angiogenesis, cellular proliferation, and invasion, thereby creating a microenvironment favorable to tumor growth and progression.

Reprinted by permission from Nature Publishing Group: *Nat. Rev. Cancer* 6, 240-248 (2006).



## Identification of Gene Expression Profiles Associated with Different Types of Breast Adipose and Their Relationship to Tumorigenesis

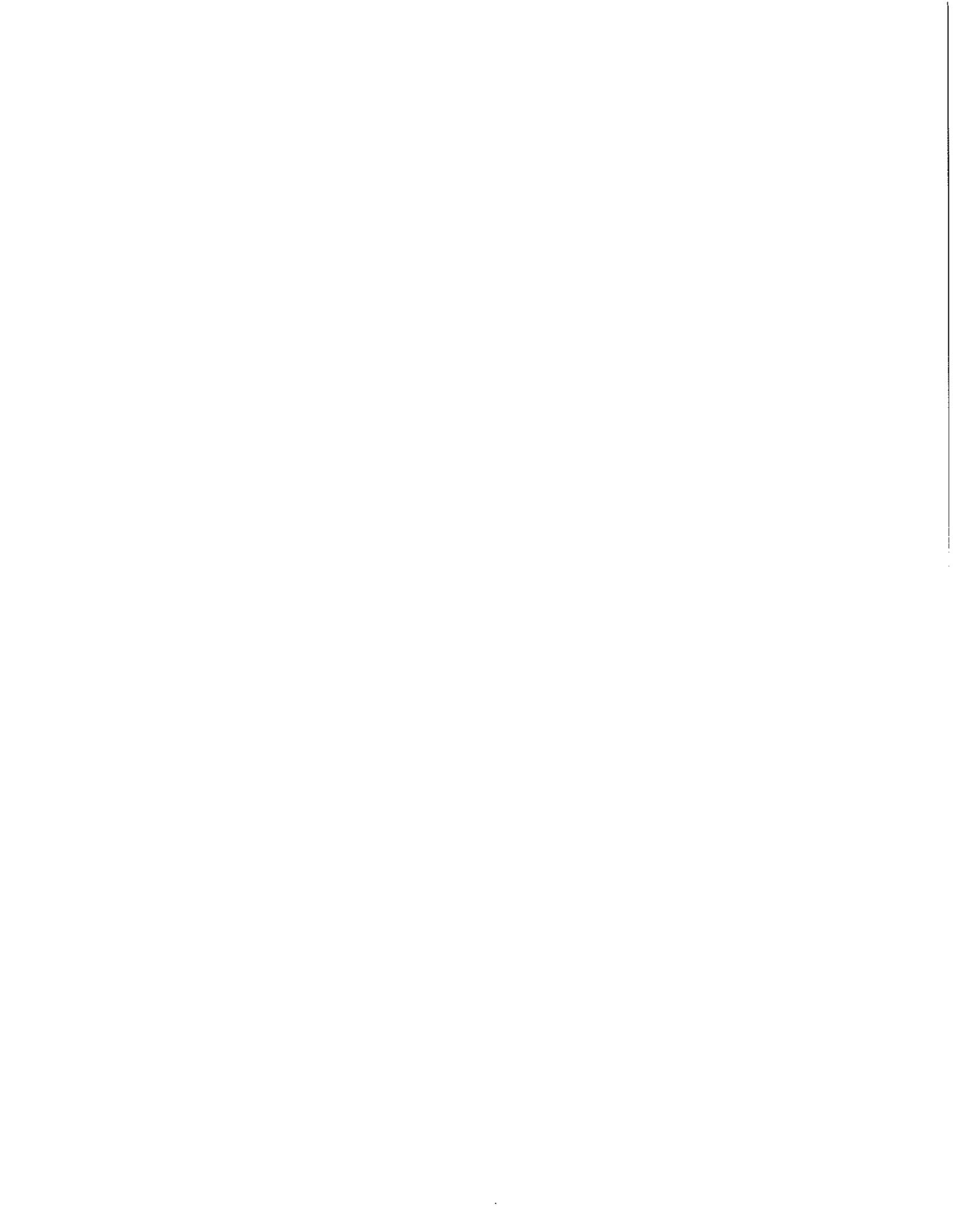
Lori Field, Brenda Deyarmin, Ryan van Laar, Craig Shriver, Rachel Ellsworth

**Background:** Research over the past decade has shown the importance of the stroma in tumorigenesis, however, having long been thought to function only as an inert energy storage depot, the role of adipose tissue in tumorigenesis has been largely ignored. Improved understanding of the role of adipose in tumorigenesis is crucial given the increasing rates of obesity and the use of autologous fat transfer in breast reconstruction.

**Methods:** Adipose, adjacent to and distant from invasive breast tumors, was laser microdissected from 20 post-menopausal women, and from 20 post-menopausal women with non-malignant breast disease. Gene expression data were generated using U133 2.0 microarrays. After quality control and visualization steps, the data were analyzed to identify significant patterns of differential expression between adipose classes, at the individual gene and molecular pathway level.

**Results:** Pathway analysis revealed that immune response differs between non-malignant, distant and tumor adjacent adipose; this response is seen as a gradient with the largest response closest to the tumor. Gene expression differed significantly in adipose from invasive compared to non-malignant breasts with FCGR2A, FOLR2, LGMN, MARCO and NLRP3 expressed at significantly higher levels and HLA-DQB1 and HLA-DQA1 at significantly lower levels in adipose from invasive breasts. Within the invasive breasts, MMP9, PLA2G7, RRM2 and SPP1 were expressed at >3-fold higher levels in adjacent compared to distant adipose.

**Conclusions:** Gene expression levels differ in breast adipose, depending on presence of or proximity to tumor cells. Adipose adjacent to the tumor demonstrated the largest immune response; this response may reflect a reaction to surgical insult from the original biopsy; however, response to surgical injury has been associated with increased ability to metastasize. In addition, within breasts with invasive breast cancer, genes involved in cellular proliferation, degradation of the extracellular matrix and angiogenesis were expressed at higher levels in adjacent compared to distant adipose. Together, these data suggest that adipose is not an inert component of the breast microenvironment but plays an active role in tumorigenesis.



Genomic instability of the breast tumor microenvironment: stromal alterations or technological artifact?

Seth Rummel, Allyson L. Valente, Jennifer L. Kane, Darrell L. Ellsworth, Craig D. Shriver, Rachel E. Ellsworth

**Introduction:** The breast tumor microenvironment plays an active role in tumorigenesis. Molecular alterations, including epigenetic modifications to DNA, and changes in RNA and protein expression have been identified in tumor-associated stroma; however, there is considerable debate as to whether the stroma is characterized by genomic instability or whether detection of chromosomal alterations in the breast stroma is a reflection of technological artifact rather than the true genomic content of the tumor microenvironment.

**Methods:** Surgically-removed breast stroma specimens from 112 women undergoing reductive mammoplasty (n=7), prophylactic mastectomy (N=6) or mastectomy for a diagnosis of breast disease (n=99) were frozen in optimal cutting temperature medium. Allelic imbalance analysis was performed in 484 stromal specimens from 98 women using a panel of 52 microsatellite markers; SNP data was generated from a subset of 86 stromal specimens using 250K SNP arrays (Affymetrix). Copy number alterations were identified using Partek Genomics Suite.

**Results:** AI was not detected in 93% (444/484) of stroma specimens. When compared to AI generated from 77 FFPE stroma specimens, 32 (42%) of which harbored at least one detectable AI event, the frequency of AI in the FFPE specimens (4.62%) was

significantly higher ( $P < 0.0001$ ) than that found in frozen specimens (0.45%). Of the stroma specimens assayed using SNP arrays 95% (82/86) had no detectable alterations and the 11 copy number changes were small and not shared between specimens.

**Conclusions:** The data presented here support a model in which the tumor microenvironment is genetically stable. The direct comparison of copy number alterations between FFPE and frozen research-grade specimens using identical methodologies suggests that past reports of significant AI/LOH in breast stroma, both adjacent to and distant from the tumor, reflects artifact in the archival specimens caused by formalin-fixation, paraffin-embedding and tissue storage.

Evaluation of the association of the ABO blood group SNP rs505922 with breast cancer phenotypes.

Seth Rummel, Craig D. Shriver, Rachel E. Ellsworth

**Purpose:** To date, evaluation of the association of the ABO blood group and breast cancer has yielded mixed results. SNP rs505922, located within the first intron of the ABO gene, has been associated with the adenocarcinoma subtype of pancreatic cancer. To determine whether the ABO blood group is associated with risk of breast cancer, rs505922 was genotyped in 629 Caucasian women with invasive breast cancer, representing a variety of clinical and pathological tumor types.

**Methods:** Genomic DNA was isolated from blood clots. TaqMan SNP assay C\_2253769\_10 was used to determine the genotypes for rs505922. Statistical analysis was performed using chi-square analysis using a P-value <0.05 to define significance.

**Results:** Genotypes were generated for 100% of the 629 patients in this study. Allele and genotype frequencies did not vary significantly for age at diagnosis, tumor stage, size or grade, hormone, HER2 or lymph node status, intrinsic subtype, tumor type or patient outcome.

**Conclusions:** Allele frequencies for rs505922 did not differ between women with breast cancer and published HapMap frequencies. Further stratification into different tumor phenotypes also failed to reveal an association between rs505922 and any tumor characteristics. Together, these data suggest that the minor allele of rs505922 and the

resulting non-O blood types, are not associated with increased risk or less favorable tumor characteristics or prognosis in breast cancer.



## **A comparative study of the triple-negative and basal-like breast cancer subtypes**

Chunqing Luo<sup>1</sup>, Hai Hu<sup>1</sup>, Leonid Kvecher<sup>1</sup>, Albert J. Kovatich<sup>2</sup>, Jeffrey A. Hooke<sup>3</sup>, Yaqin Chen<sup>1</sup>, Richard J. Mural<sup>1</sup> and Craig D. Shriver<sup>3</sup>

<sup>1</sup>Windber Research Institute, Windber, PA; <sup>2</sup>MDR Global Systems, Windber, PA; <sup>3</sup>Walter Reed National Military Medical Center, Bethesda, MD.

**Background:** Multiple breast cancer subtypes with distinct prognostic differences have been identified based on their molecular characteristics. The Triple-Negative (TN) subtype, identified using clinically-relevant immunohistochemistry (IHC) markers, largely overlaps with the basal-like subtype identified by gene expression profiling. Both have poor prognosis. We performed a preliminary comparative study of the TN/Basal subtypes of breast cancers using molecular data from multiple experimental platforms. A set of breast cancer samples (65) from the Clinical Breast Care Project (CBCP) with comprehensive clinico-pathological and immunophenotypic data were studied, which were also part of The Cancer Genome Atlas (TCGA) breast cancer study.

**Methods:** The 65 breast cancer tissues from the CBCP had IHC results of ER, PR, HER2, Ki67, and p53. The TN subtype was defined as ER-/PR-/HER2-, and the other subtypes were defined according to current practice. Gene expression data of 532 samples (including the 65 CBCP samples) and DNA methylation data of 180 samples (including 25 of the 65 CBCP samples) were downloaded from the TCGA data portal (<http://tcga-data.nci.nih.gov/tcga/tcgaHome2.jsp>). Subtype calls from PAM50 were available from TCGA Breast AWG for 429 of the 532 samples. Clustering analysis was performed using Euclidean distance metric and complete linkage algorithm. Gene expression analysis was performed using Bioconductor “samr” package, for two class unpaired comparison. A p value less than 0.05 and an average fold change of 2 were used as the thresholds for Differentially Expressed Genes (DEGs). The association between gene expression and methylation was calculated using Pearson’s Correlation Test. All statistical analyses were done using R.

**Results:** The 65 CBCP samples were classified into TN (n=12), HER2+ (n=5), LA (n=11), LB (n=31), and undefined (n=6). The 429 TCGA samples were classified as 80 basal-like, 46 Her2+, 187 LA, 109 LB, and 7 normal-like based on gene expression. The IHC-based subclassifications were broadly consistent with the gene-expression based classifications (40/59) with all 12 TN cases called as basal. In analyzing the 65 CBCP samples, unsupervised clustering based on gene expression clustered 9 of the 12 TN into one group, which also included one HER2+ and one LB. Differential gene expression analysis between TN and non-TN yielded 46 DEGs, some of which can well distinguish TN from the other subtypes. Analysis of the 429 TCGA samples (80 basal-like and the remaining 349 samples) yielded 48 DEGs, and 29 of which overlap with the 46 DEGs identified earlier, including ESR1, FOXA1, AGR2, and TFF3. We observed different levels of association between gene expression and methylation for these genes, ranging from the “expected” low methylation/high expression to high methylation/low expression (e.g. TFF3), to completely non-correlated (e.g. ESR1).

**Discussion:** A number of clinically important questions remain unanswered at the molecular level for TN/Basal subtypes. The relationship between the two categories and their further subdivision needs to be addressed. Integrative analysis of data from multiple platforms can enable us to answer these questions and to develop personalized treatment protocols for the molecular subtypes of breast cancers.

Evaluation of the association of the ABO blood group SNP rs505922 with breast cancer phenotypes

Seth Rummel<sup>1</sup>, Craig D. Shriver<sup>2</sup>, Rachel E. Ellsworth<sup>3</sup>

From the Clinical Breast Care Project of the <sup>1</sup>Windber Research Institute, 620 Seventh St, Windber, PA 15963; <sup>2</sup>Walter Reed Army Medical Center, 6900 Georgia Avenue NW, Washington DC, 20307; <sup>3</sup>Henry M. Jackson Foundation for the Advancement of Military Medicine, 620 Seventh Street, Windber, PA 15963

Corresponding Author:

Rachel E. Ellsworth, PhD

Clinical Breast Care Project

620 Seventh Street

Windber, PA 15963

FAX: 814 467-6334

Phone: 814 467-9844

email: [r.ellsworth@wriwindber.org](mailto:r.ellsworth@wriwindber.org)

This research was supported by a grant from the United States Department of Defense (Military Molecular Medicine Initiative MDA W81XWH-05-2-0075, Protocol 01-20006).

## **ABSTRACT**

**Purpose:** To date, evaluation of the association of the ABO blood group and breast cancer has yielded mixed results. SNP rs505922, located within the first intron of the ABO gene, has been associated with the adenocarcinoma subtype of pancreatic cancer. To determine whether the ABO blood group is associated with risk of breast cancer, rs505922 was genotyped in 629 Caucasian women with invasive breast cancer, representing a variety of clinical and pathological tumor types.

**Methods:** Genomic DNA was isolated from blood clots. TaqMan SNP assay C\_2253769\_10 was used to determine the genotypes for rs505922. Statistical analysis was performed using chi-square analysis using a P-value <0.05 to define significance.

**Results:** Genotypes were generated for 100% of the 629 patients in this study. Allele and genotype frequencies did not vary significantly for age at diagnosis, tumor stage, size or grade, hormone, HER2 or lymph node status, intrinsic subtype, tumor type or patient outcome.

**Conclusions:** Allele frequencies for rs505922 did not differ between women with breast cancer and published HapMap frequencies. Further stratification into different tumor phenotypes also failed to reveal an association between rs505922 and any tumor characteristics. Together, these data suggest that the minor allele of rs505922 and the resulting non-O blood types are not associated with increased risk or less favorable tumor characteristics or prognosis in breast cancer.

*Keywords:* ABO blood-group system, breast cancer, phenotypes

## 1. INTRODUCTION

The first association between the ABO blood group and cancer risk was made in 1953 in patients with stomach cancer, where blood group A is associated with increased risk of stomach cancer and blood group O confers a protective advantage in individuals living in England <sup>(1)</sup>. This increased risk of gastric cancer in patients with blood group A has been confirmed as recently as 2010 <sup>(2)</sup>. In oral cancer, loss of the A and B antigens is associated with increased cell migration, in non-small cell lung cancer, the A blood group is associated with improved survival, and in ovarian cancer the B antigen is associated with increased risk <sup>(3-6)</sup>.

A number of studies have investigated the role of the ABO blood group in breast cancer risk and pathology. For example, two publications from the 1950s found no association between ABO blood group and risk of breast cancer in either English or American patient populations <sup>(5;7)</sup>. In more recent studies, while the A antigen has been associated with increased risk of developing invasive ductal carcinoma in 166 Greek women, in a group of 565 Turkish women, no association was seen between breast cancer risk and any ABO blood types <sup>(8;9)</sup>. These differences in risk assessment between not just breast but other tumor types and the ABO blood group may be attributed to differences in ABO frequencies between populations, the importance of using comparable case and control groups, and methods for determining ABO blood types.

In 2009, an agnostic approach was used to identify genes associated with risk of developing pancreatic cancer further supported the association between the ABO blood group and cancer. Genotyping of 558,542 single nucleotide polymorphisms (SNPs) in 1,896 patients with pancreatic cancer and 1,939 controls revealed a significant

association between SNP rs505922, located within the first intron of the ABO glycosyltransferase (ABO) gene<sup>(10)</sup>. The T allele of rs505922, which was found to be protective, is in linkage disequilibrium with the single base pair deletion that encodes the O antigen, suggesting that the A and B blood antigens may be associated with higher risk of pancreatic cancer. In a follow up study, the role of the ABO blood group was re-evaluated in 15,359 patients treated at the European Institute of Oncology with a variety of cancer types. As association between the ABO blood group was detected only for pancreatic cancer<sup>(11)</sup>. Importantly, the protective effect of the O allele was not associated with all types of pancreatic cancer but was limited to the adenocarcinoma subtype.

To determine whether that ABO blood group is associated with risk of breast cancer, SNP rs505922 was genotyped in 629 Caucasian women with invasive breast cancer diagnosed between 2001 and 2010. Data were then evaluated by a number of pathological characteristics including age at diagnosis, hormonal and HER2 status, intrinsic subtype, tumor grade, stage and size, and clinical outcome.

## **2. PATIENTS AND METHODS**

### *2.1. Study Population*

The Clinical Breast Care Project (CBCP) database was queried to identify all Caucasian women with invasive breast cancer diagnosed between 2001 and 2009. Blood samples were collected with approval from the Walter Reed Army Medical Center Human Use Committee and Institutional Review Board (IRB). All subjects enrolled in the CBCP voluntarily agreed to participate and gave written informed consent.

Diagnosis of every specimen was performed by a single, dedicated breast pathologist from hematoxylin and eosin (H&E) stained slides; staging was performed using guidelines defined by the *AJCC Cancer Staging Manual* sixth edition <sup>(12)</sup> and grade assigned using the Nottingham Histologic Score <sup>(13;14)</sup>. ER and PR status were determined by immunohistochemical (IHC) analysis at a clinical laboratory (MDR Global, Windber, PA) and HER2 status was assayed using the PathVysion® HER-2 DNA Probe kit (Abbott Laboratories, Abbott Park, IL) according to manufacturer's protocols.

## 2.2. *SNP Genotyping*

Genomic DNA was isolated from blood clots using the Genra Clotspin and Puregene DNA purification kits (Qiagen Inc., Valencia, CA). Ten ng of DNA were amplified in each PCR reaction and each sample was run in duplicate. PCR was performed using the TaqMan SNP assay C\_2253769\_10 (Applied Biosystems, Foster City, CA), representing SNP rs505922 and run on a 7000 Sequence Detection System (Applied Biosystems, Foster City, CA). Genotypes (CC, CT or TT) were determined using the ABI PRISM® 7000 Sequence Detection System software (Applied Biosystems, Foster City, CA).

## 2.3 *Statistical Analysis*

Differences in genotype or allele frequencies between different clinicopathological characteristics were calculated using chi-square analysis using *rxc* Contingency Tables ([www.physics.csbsju.edu/stats/contingency\\_NROW\\_NCOLUMN\\_form.html](http://www.physics.csbsju.edu/stats/contingency_NROW_NCOLUMN_form.html)). Significance was determined using  $P < 0.05$ .

## 3. RESULTS

### *3.1 Clinicopathological characteristics of the patient population*

The average age at diagnosis for the 629 Caucasian women with invasive breast cancer was 58.94 years (range 28-97 years) and the majority (74%) was post-menopausal at diagnosis. Most of the tumors were invasive ductal carcinomas (IDCA) (70%), ER positive (78%), HER2 negative (82%) and of the luminal A (ER+/HER2-) subtype (74%). Tumors were divided between well-differentiated (33%), moderately-differentiated (38%) and poorly-differentiated (29%). Most of the tumors were early stage with 86% being stage I-II and 67% having negative lymph node status. Seventeen percent of the patients had died of disease.

### *3.2 Frequency of rs505922 in women with invasive breast cancer*

Of the 629 samples assayed in this study, six (1%) produced discordant genotypes between duplicate samples during the original run; these samples were run a second time resulting in concordant genotypes between duplicate samples, resulting in a genotyping accuracy of 100%, with each sample clustering into one of three genotypes. Allele frequencies were 0.65 for the T allele and 0.35 for the C allele. Chi-square analysis revealed that the genotype frequencies were in Hardy-Weinberg Equilibrium ( $P=0.961$ ). Neither genotype nor allele frequencies differed significantly from those reported for Caucasians in the CSHL-HAPMAP panel ( $P=0.809$  and  $P=0.654$ , respectively).

### *3.3 Evaluation of rs505922 genotype and allele frequencies with clinicopathological characteristics*

Allele and genotype frequencies were compared for age at diagnosis, tumor stage, size and grade, hormone, HER2 and lymph node status, intrinsic subtype, tumor type and patient status. No significant differences were seen by genotype (Table 1) or allele (data

not shown). To determine whether the minor C allele was associated with any variable, CC and CT genotypes were combined. As with genotype and allele frequencies, no significant differences were detected.

#### **4. DISCUSSION**

The first suggestion of an association between ABO blood group antigens and malignancy was made almost 100 years ago, yet the role of the ABO blood group in cancer risk and prognosis remains controversial. Results from the pancreatic GWAS study in 2009 renewed interest in the ABO blood group as the T allele of SNP rs505922, located within the first intron of the ABO gene locus was found to confer a protective advantage against pancreatic cancer. The T allele has been found to be in linkage disequilibrium with the single base pair deletion responsible for the O blood group antigen, leading the authors to suggest that people with non-O blood groups have increased risk of developing pancreatic cancer <sup>(10)</sup>.

In addition to the association with defining blood group antigens, variability of SNPs within the ABO region has been associated with circulating levels of tumor necrosis factor alpha (TNF $\alpha$ ), soluble intracellular adhesion molecule 1 (sICAM-1) and alkaline phosphatase (15-17). Because these tumor-promoting factors may also provide an environment favoring breast tumor development, variation of rs505922 may contribute to increased risk of breast cancer development. Comparison of allele frequencies in our cohort of 629 Caucasian women with invasive breast cancer, however, did not differ significantly from Caucasians in the HapMap project, suggesting that the minor allele of rs505922 is not associated with increased risk of breast malignancy.

In the follow-up paper by Iodice et al., which found no association between ABO blood group and breast cancer risk ( $P=0.60$ ) in 7,208 women treated at the European Institute of Oncology (IEO), the importance of tumor heterogeneity was highlighted <sup>(11)</sup>. While the GWAS study found an association between rs505922 and pancreatic cancer overall, evaluation of ABO blood group antigens with subtypes of pancreatic cancer in the IEO revealed that the non-O blood type was associated with the exocrine, but not the endocrine, form of pancreatic cancer. Breast cancer is marked by tremendous heterogeneity in terms of age at diagnosis, tumor pathologies, and molecular characteristics. Thus, the ABO blood group may be associated with increased risk or prognosis of certain subtypes of breast cancer. Neither allele nor genotype frequencies differed significantly by age at diagnosis, tumor stage, grade and size, hormone, HER2 and lymph node status, intrinsic subtype or prognosis, suggesting that unlike pancreatic cancer, the ABO blood group is not associated with either overall risk of or certain subtypes of breast cancer.

Our results are in agreement with those from Iodice et al. derived largely from an Italian patient population as well as those from Dede et al. who found no association in 565 Turkish women <sup>(9;11)</sup>. In contrast, Stamatakos et al. found an association between the A antigen and IDCA and/or poor prognosis in population of Greek women <sup>(8)</sup>. Differences between the Greek study and the other studies that failed to find an association may reflect the small sample size used in the Greek study, which was based on 166 women with breast cancer, which is much smaller than the 565 Turkish, 629 Caucasian American, or 7,208 Italian women included in the other studies.

Although this study does not support a role for the ABO gene in breast cancer incidence or prognosis, this study was limited to individuals with self-reported “white” ancestry. The frequency of the ABO blood group varies globally, with a predominance of the O allele in the Kikuyu of Kenya, Malaysians and Navajos, and of the A allele in Hawaiians, Portuguese and Swiss (<http://www.bloodbook.com/world-abo.html>). Thus, differences in allele frequency of the ABO blood group must be considered when evaluating the role of ABO in disease. For example, the A antigen, which was found to be associated with increased breast cancer risk in a Greek population but not in a Turkish population, is more frequent in Greeks (0.42) than in Turks (0.34), which may influence the association between the A blood group and disease status, although the O group frequency is similar between the two populations (0.40 and 0.43 in Greeks and Turks, respectively).

The variability of the minor allele frequency of SNP rs505922 also varies between populations: in a pilot study from the 1,000 Genomes Project (18), the frequency of rs505922 was 0.32 in YRI (Sub-Saharan Africa), 0.352 in CHB+JPT (China and Japan) and 0.417 in CEU (Caucasians from Utah). Although patients from ten European countries and Shanghai were included in the pancreatic GWAS study, the majority of patients were from a number of cohort studies from the United States. Similar results were achieved when all patients were included in the analysis and when only those of European ancestry were included<sup>(10)</sup>. In our study, all samples were collected from self-described white women living in Washington DC or rural Pennsylvania. Given the marked admixture in the United States, these 629 women likely represent a heterogeneous mixture of European ancestries. Thus, an association that may be detected

in the more heterogeneous Greek population may be masked by the genetic contributions from other European cultures represented in our American patient population.

In conclusion, the minor allele frequency of SNP rs505922, which is in linkage disequilibrium with the non-O ABO blood types, did not differ between Caucasian American women with breast cancer and published HapMap frequencies. Further stratification into different tumor phenotypes also failed to reveal an association between rs505922 and any tumor characteristics. Together, these data suggest that unlike pancreatic cancer, where the minor rs505922 allele was associated with increased risk, especially of exocrine tumors, rs505922 and the resulting non-O blood types, are not associated with increased risk or less favorable tumor characteristics or prognosis in breast cancer.

#### **CONFLICT OF INTEREST STATEMENT**

None declared.

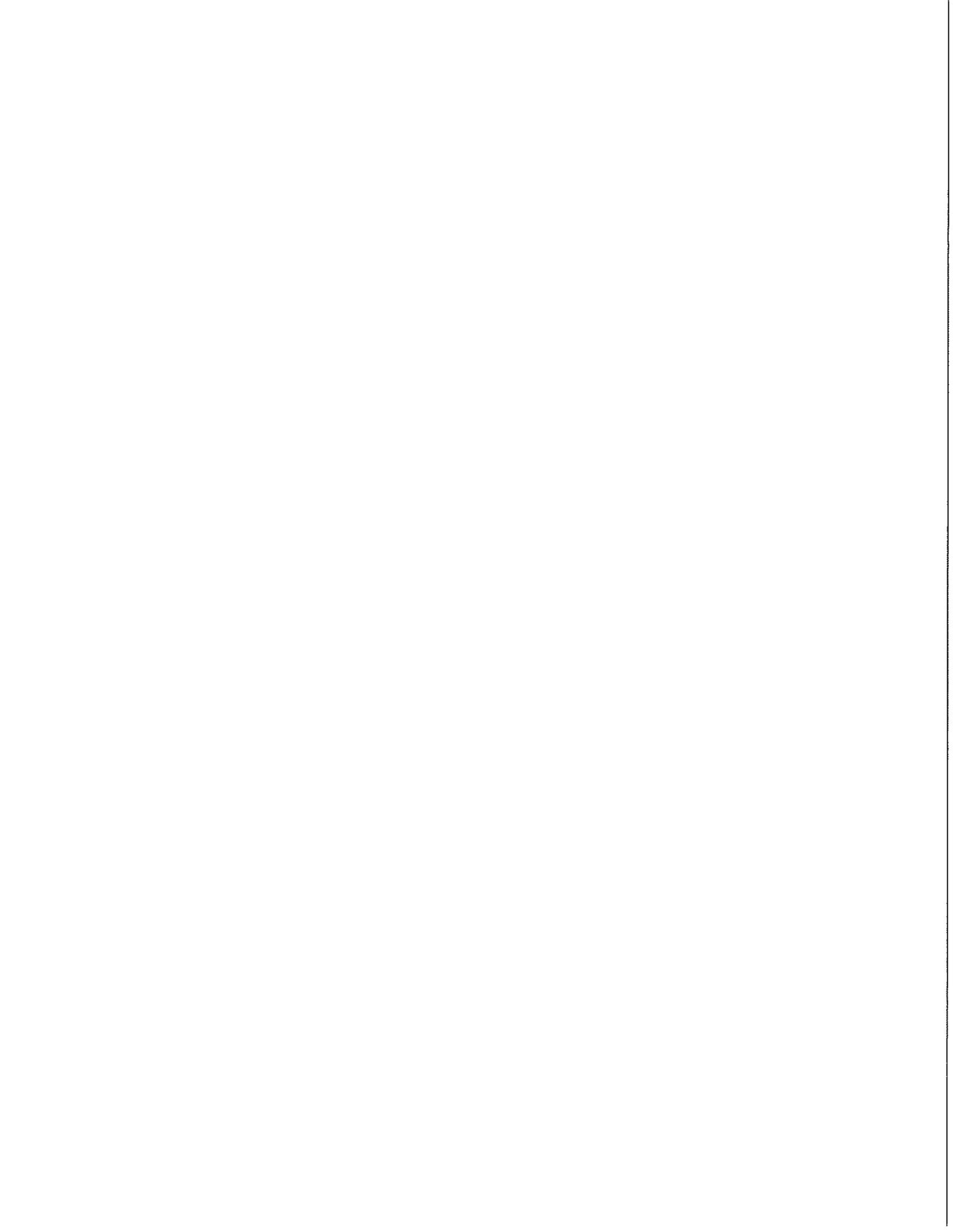
## **ACKNOWLEDGEMENTS**

We thank Brenda Deyarmin for her assistance collecting clinical information for the patients in this study. The opinion and assertions contained herein are the private views of the authors and are not to be construed as official or as representing the views of the Department of the Army or the Department of Defense. This research was supported by a grant from the United States Department of Defense (Military Molecular Medicine Initiative MDA W81XWH-05-2-0075, Protocol 01-20006). The Department of Defense was not involved in the conceptualization or execution of this project.

## REFERENCES

1. Aird I, Bentall HH, Roberts JA. A relationship between cancer of stomach and the ABO blood groups. *Br Med J* 1953;1:799-801.
2. Edgren G, Hjalgrim H, Rostgaard K, Norda R, Wikman A, Melbye M et al. Risk of gastric cancer and peptic ulcers in relation to ABO blood type: a cohort study. *Am J Epidemiol* 2010;172:1280-1285.
3. Gates MA, Wolpin BM, Cramer DW, Hankinson SE, Tworoger SS. ABO blood group and incidence of epithelial ovarian cancer. *Int J Cancer* 2011;128:482-86.
4. Lee JS, Ro JY, Sahin AA, Hong WK, Brown BW, Mountain CF et al. Expression of blood-group antigen A--a favorable prognostic factor in non-small-cell lung cancer. *N Engl J Med* 1991;324:1084-90.
5. Aird I, Bentall HH, Mehigan JA, Roberts JA. The blood groups in relation to peptic ulceration and carcinoma of colon, rectum, breast, and bronchus; an association between the ABO groups and peptic ulceration. *Br Med J* 1954;2:315-21.
6. Dabelsteen E, Gao S. ABO blood-group antigens in oral cancer. *J Dent Res* 2005;84:21-28.
7. Goldenberg IS, Hayes MA. Breast carcinoma and ABO blood groups. *Cancer* 1958;11:973-74.
8. Stamatakis M, Kontzoglou K, Safioleas P, Safioleas C, Manti C, Safioleas M. Breast cancer incidence in Greek women in relation to ABO blood groups and Rh factor. *Int Semin Surg Oncol* 2009;6:14.
9. Dede DS, Aksoy S, Dizdar O, Cerci P, Gullu I, Ozisik Y et al. Blood ABO groups and risk of breast cancer. *Med Oncol* 2010;27:1433.
10. Amundadottir L, Kraft P, Stolzenberg-Solomon RZ, Fuchs CS, Petersen GM, Arslan AA et al. Genome-wide association study identifies variants in the ABO locus associated with susceptibility to pancreatic cancer. *Nat Genet* 2009;41:986-90.
11. Iodice S, Maisonneuve P, Botteri E, Sandri MT, Lowenfels AB. ABO blood group and cancer. *Eur J Cancer* 2010;46:3345-50.
12. American Joint Committee on Cancer. *AJCC Cancer Staging Manual 6th ed.* New York, NY: Springer; 2002.
13. Bloom HJ, Richardson WW. Histological grading and prognosis in breast cancer. *Br J Cancer* 1957;11:359-77.

14. Elston CW, Ellis IO. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow up. *Histopathology* 1991;19:403-10.
15. Yuan X, Waterworth D, Perry JR, Lim N, Song K, Chambers JC et al. Population-based genome-wide association studies reveal six loci influencing plasma levels of liver enzymes. *Am J Hum Genet* 2008;83:520-528.
16. Pare G, Chasman DI, Kellogg M, Zee RY, Rifai N, Badola S et al. Novel association of ABO histo-blood group antigen with soluble ICAM-1: results of a genome-wide association study of 6,578 women. *PLoS Genet* 2008;4:e1000118.
17. Melzer D, Perry JR, Hernandez D, Corsi AM, Stevens K, Rafferty I et al. A genome-wide association study identifies protein quantitative trait loci (pQTLs). *PLoS Genet* 2008;4:e1000072.
18. 1000 Genomes Project Consortium, Durbin RM, Abecasis GR, Altschuler DL, Auton A, Brooks LD et al. A map of human genome variation from population-scale sequencing. *Nature* 2010;467:1061-73.



## The effect of HER2 expression on luminal A breast tumors

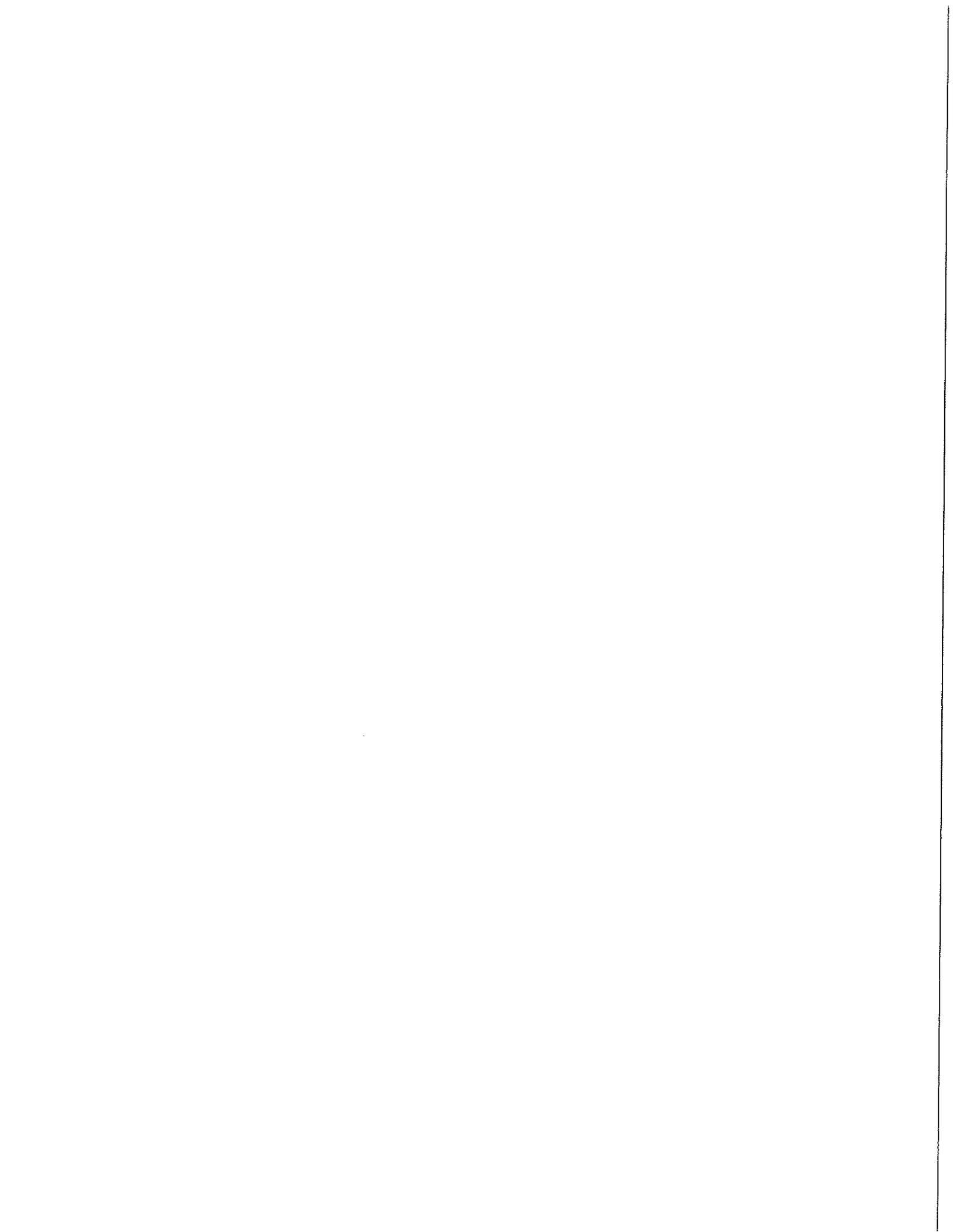
Rachel E. Ellsworth, Christopher Gallagher, Brenda Deyarmin, Jeffrey Hooke, Craig Shriver

**Background:** Luminal A (LumA) tumors are the most common subtype of breast cancer and have been associated with favorable prognosis. Clinically, LumA tumors can be defined using the immunohistochemistry (IHC) based markers, with LumA tumors defined as ER and or PR+/HER2-. Although tumors can be considered LumA if they have >1% expression of ER, negative HER2 status ranges from 0+ (no expression) to 2+ (weak to moderate expression in >10% of cells) without concurrent genomic amplification. It is not clear how moderate levels of HER2 expression influence tumor biology in LumA tumors.

**Methods:** The Clinical Breast Care Project database was queried to identify all hormone receptor positive (ER/PR staining >1%) cases with HER2 status defined as 0+ (n=150) or 2+/not amplified (n=128). Clinicopathological characteristics were compared between groups using chi-square analysis. A subset of tumors (n=25 0+, n = 25 2+) with frozen tumor specimens available were subjected to laser microdissection and hybridized to U133 2.0 microarrays (Affymetrix). Gene expression data was analyzed using Partek Genomics Suite.

**Results:** Average age at diagnosis did not differ between patients with no (58 years) or moderate (59 years) HER2 expression. Patient groups did not differ by ethnicity, tumor stage, grade or size, or lymph node status. Although breast cancer mortality was low, all patients who died of disease (n=6) had moderate levels of HER2 expression ( $P<0.05$ ). At the molecular level, 20 probes from 18 genes were differentially expressed using a false detection rate  $P<0.05$ , >2x-fold change. Six genes, including ESRRG, had significantly higher expression levels and twelve, including MAPT, had significantly lower expression in tumors with moderate HER2 expression.

**Discussion:** LumA tumors with moderate expression levels of HER2 are dissimilar from those with no expression of HER2. Many of the genes down-regulated in HER2 expressing LumA tumors are involved in inhibition of growth and proliferation, apoptosis, and improved survival. In addition, increased expression of ESRGG, as seen in LumA tumors with moderate expression of HER2 has been associated with tamoxifen resistance. Together, these molecular differences suggest that LumA tumors with HER2 expression have more aggressive characteristics than those without HER2 expression and may contribute to the significantly higher mortality rate seen in patients with LumA tumors expressing HER2.



## Molecular alterations associated with breast cancer mortality

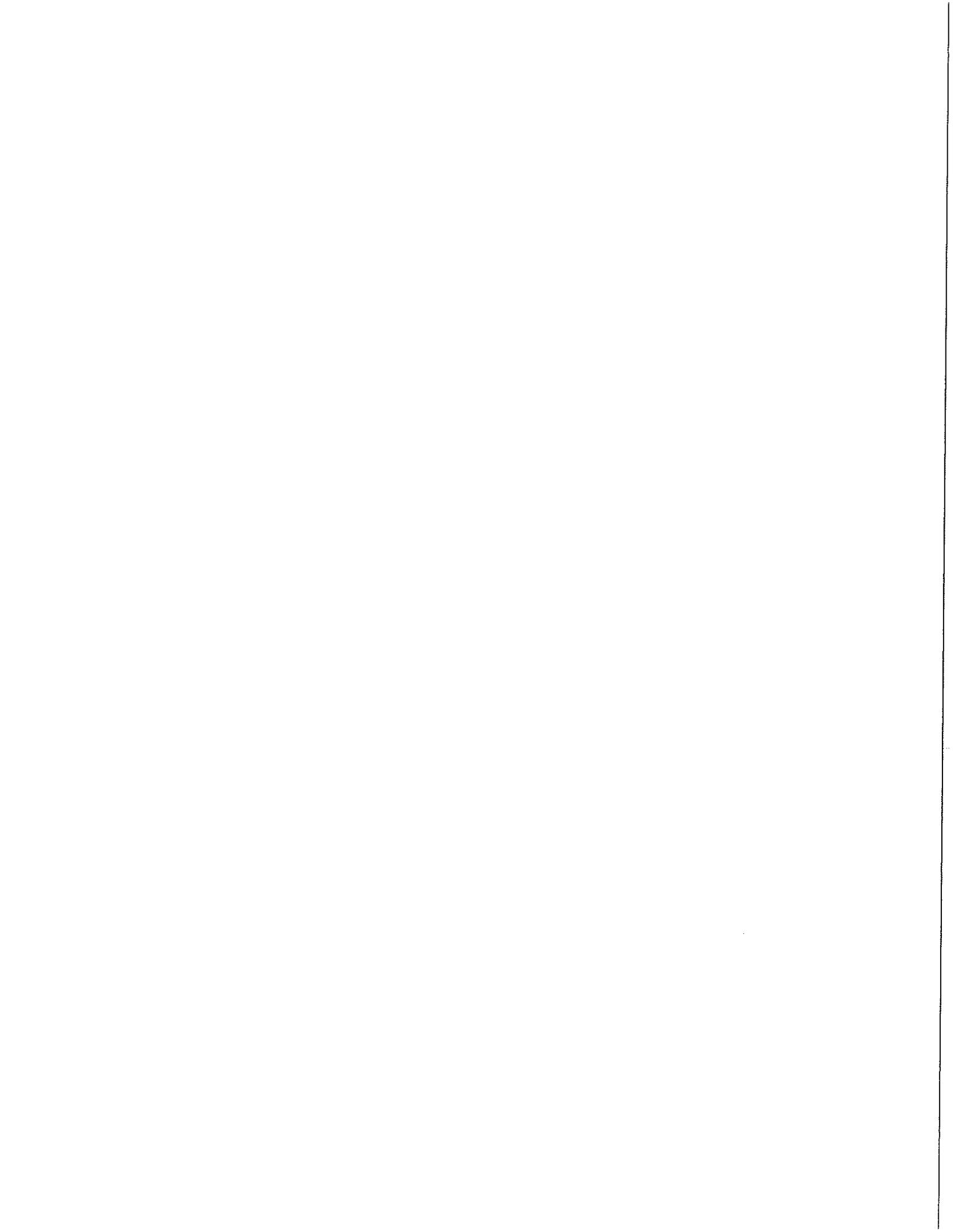
Rachel E Ellsworth, J. Leigh Campbell, Laura Voegtly, Craig D. Shriver

**Background:** Breast cancer is a highly heterogeneous disease and patients with otherwise similar pathologies treated with similar treatment regimens may have very different clinical outcomes. Identification of molecular alterations associated with disease outcome may serve as more accurate and non-subjective measures used to assess risk and prognosis as well as provide new molecular targets for improved treatment.

**Methods:** A bank of allelic imbalance (AI) data representing 26 regions known to be frequently altered in breast cancer has been generated for >300 invasive breast tumors. A chart review was completed for patients diagnosed 2001 - 2005. Levels and patterns of AI were compared between patients who were dead of disease within 5 years of diagnosis (DOD) and those with long-term ( $\geq 5$  years) survival (LTS) using Student t-test and chi-square analysis with a significance value of  $P < 0.05$ . Survival data was analyzed using GraphPad.

**Results:** Levels of AI were significantly higher in tumors from the 24 DOD patients (26.9%) compared to the 70 LTS patients (20.9%). AI was found significantly more frequently in DOD tumors at chromosomes 7q31, 8p22-p21, 18q21 and 22q12.3 (26%, 45%, 33% and 38%, respectively) compared to LTS patients (4%, 14%, 13% and 13%). In contrast, chromosome 16q22-q24 was altered significantly less frequently in DOD (10%) compared to LTS tumors (40%). Survival analysis supported these findings for all but chromosome 22q12.3.

**Conclusion:** Tumors associated with poor outcome are genetically more advanced. Alterations at chromosome 7q31, 8p22, 18q21 and 22q13 may be associated with increased risk of mortality, while loss of 16q appears to have a protective role. Research suggests that loss of the GINS2 gene (16q24), which is involved in the initiation of DNA replication, may protect against metastasis. These chromosomal regions and candidate genes may be effective targets for identifying patients with poor prognosis or serve as new targets for therapies to prevent or halt the metastatic process.



CELLULAR AND MOLECULAR BIOLOGY

## Analysis of Matrix Metalloproteinase-1 Gene Polymorphisms and Expression in Benign and Malignant Breast Tumors

Jing Zhou,<sup>1</sup> Constance Brinckerhoff,<sup>2</sup> Susan Lubert,<sup>1</sup> Kui Yang,<sup>3</sup> Jasmine Saini,<sup>1</sup> Jeffrey Hooke,<sup>4</sup> Richard Mural,<sup>1</sup> Craig Shriver,<sup>4</sup> and Stella Somiari<sup>1</sup>

Clinical Breast Care Project, Windber Research Institute, Windber, Pennsylvania, USA,<sup>1</sup> Dartmouth-Hitchcock Medical Center, Lebanon, New Hampshire, USA,<sup>2</sup> Colorado Foundation for Medical Care, Englewood, Colorado, USA,<sup>3</sup> Clinical Breast Care Project, Walter Reed Army Medical Center, Washington, District of Columbia, USA<sup>4</sup>

A guanine insertion polymorphism in matrix metalloproteinase-1 promoter (MMP-1 2G) is linked to early onset and aggressiveness in cancer. We determined the role of MMP-1 2G on MMP-1 expression and breast cancer severity in patients with breast diseases. We observed no significant difference in genotype distribution among different disease groups. However, MMP-1 expression was significantly higher in atypical ductal hyperplasia than in benign breast disease and in invasive breast cancer compared to *in situ* breast cancer. MMP-1 2G insertion polymorphism in the invasive group also correlated significantly with the expression of MMP-1 and breast cancer prognostic markers HER2 and P53.

**Keywords:** Breast cancers; Detection/diagnosis; Cancer biomarkers

### INTRODUCTION

Matrix metalloproteinases (MMPs) are a large family of calcium-dependent, zinc-containing proteinases. They actively participate in the remodeling of extracellular matrix (ECM) by degrading certain ECM components and promoting cell proliferation, migration, differentiation, apoptosis, and angiogenesis (1–5). MMPs are not only involved in physical processes such as growth, wound healing, and fibrosis but also play an important role in tumor invasion and progression (6–16). Among all the identified MMPs, MMP-1 is the most ubiquitously expressed interstitial collagenase. It is also one of the few MMPs that can cleave collagen I, and III, the major components of ECM, which must be degraded for tumor invasion and metastasis to occur (17). In the literature, MMP-1 has been implicated in both early detection and prognosis of breast cancer. Increased expression of MMP-1 has been reported to be significantly associated with pre-cancerous status (18). Overexpression of MMP-1 mRNA has been correlated with a high frequency of recurrence and fatal outcome (19).

A number of mechanisms can impact the level and the activity of MMPs, among which transcriptional regulation plays one of the most important role. It has been reported that a single guanine insertion at –1607 bp in the promoter region of MMP-1 creates a binding site (5'-GGAT-3') for transcription factor ETS, which can cooperate with the adjacent AP-1 binding site to induce high-level expression of MMP-1 through enhancing the activity of the basal transcriptional machinery, but the 1G allele (5'-GAT-3') does not affect the transcriptional activity of MMP-1 (20, 21). MMP-1 2G polymorphism has been associated with lymph node metastasis in breast cancer (22, 23) and has also been linked to the pathogenesis of other cancers, for example, to the early onset of lung cancer (24), increased risk of nasopharyngeal carcinoma, oral squamous cell carcinoma, and colorectal cancer (25–27), aggressiveness of head and neck cancer (28, 29), development and progression of endometrial carcinoma (30), and negative prognosis in patients with ovarian cancer (31).

In this study, we sequenced the genotype of 132 patients categorized as invasive, *in situ* breast cancer, atypical hyperplasia, or benign breast disease and analyzed the tissue expression of MMP-1 in these diseases to determine the impact of MMP-1 2G insertion polymorphism and MMP-1 expression on the occurrence and severity of breast cancer. We further explored the possible relationship of MMP-1 expression and MMP-1 polymorphism with other known clinical markers of breast cancer such as HER2 and P53.

### MATERIALS AND METHODS

#### Subject selection

A total of 180 patients consisting of 136 breast cancer, 15 atypical ductal hyperplasia (ADH), and 29 benign breast disease were recruited in this study. The breast cancer patients comprised 71 invasive breast cancers (IBCs) and 65 ductal/lobular carcinoma *in situ* (DCIS/LCIS). The benign patients were mainly those exhibiting fibrocystic changes

Correspondence to: Stella Somiari, PhD, Clinical Breast Care Project, Windber Research Institute, 620 7th Street, Windber, PA 15963, USA.  
email: s.somiari@wriwindber.org

(stromal fibrosis, cysts, apocrine metaplasia, sclerosing adenosis, and intraductal hyperplasia). Other benign diagnoses included microcalcifications, adenomatous change, columnar cell change, and intraductal papillomas. All participants were fully consented patients recruited under the Institutional Review Board approved protocols of the Clinical Breast Care Project (CBCP). They were female, 18 years of age and above, who were attending the breast clinics of the CBCP at the Walter Reed Army Medical Center (WRAMC) in Washington, DC, and the Joyce Murtha Breast Care Center (JMBCC)/Windber Medical Center (WMC) in Windber, PA. The median ages for IBC, DCIS/LCIS, ADH, and benign groups were 62 years (range: 36–86), 55 years (range: 25–84), 49 years (19–74), and 45 years (range: 26–71), respectively. Characteristics of the patients with breast cancer are summarized in Table 1. Sequencing data for MMP-1 polymorphism were obtained for 132 patients, including 52 IBCs, 50 DCIS/LCIS, 13 ADH, and 17 benign breast diseases. Immunohistochemistry data for MMP-1 expression were determined in 165 patients, which included 65 IBCs, 64 DCIS/LCIS, 11 ADH, and 25 benign breast diseases.

### Genotyping

Blood clot was obtained after processing peripheral blood into serum. Blood DNA was extracted from blood clots using the QIAamp blood DNA commercial kit (Qiagen). The DNA was then utilized for polymerase chain reaction (PCR) to amplify *MMP-1* gene as earlier described (21). Briefly, PCR reaction was carried out in a total volume of

50  $\mu$ L with 100 ng genomic DNA, 0.2 mM dNTP, 2 mM  $MgSO_4$ , 50 nM each of the two primers, and 1 unit of platinum Taq DNA high-fidelity polymerase (Invitrogen). The primer sequences for amplifying MMP-1 were: forward primer 5'-TTGCCAGATGGGACAGTGTATGAG-3' and reverse primer 5'-ACATTAAATTGTCTTGGGTACTGGT-3'. PCR started with 30 s incubation at 94°C followed by 10 cycles of 30 s incubation at 94°C, 30 s incubation at 52°C, and 2 min incubation at 68°C; 30 cycles of 30 s incubation at 94°C, 30 s incubation at 55°C, and 1 min 30 s at 68°C with 4 min incubation at 68°C. The amplification was verified on 1.2% agarose gel.

PCR products were cleaned up with Exonuclease I and Shrimp Alkaline Phosphatase (Affymetrics) to remove primers. Cycle sequencing reactions were performed with a nested primer (5'-AGTGTCTTTGGTCTCTGC-3') and purified by ethanol precipitation (including 7.5 M ammonium acetate). DNA pellets were then dried, resuspended in MegaBACE Loading Solution, and sequenced using the MegaBACE 1000 (GE Healthcare). Data were analyzed with Sequence Analysis v4.0 software (GE Healthcare).

### Immunohistochemistry (IHC)

Formalin-fixed, paraffin-embedded sections of the selected cases were retrieved from the tissue repository to determine the expression level of MMP-1 by IHC. The pathological diagnosis of each case was confirmed by the CBCP pathologist using a freshly prepared hematoxylin and eosin slide. For all confirmed cases, seven slides of 4-micron sections were cut and subjected to IHC analysis for MMP-1. Antibody optimization was performed to establish the required optimal conditions. For the determination of MMP-1 expression, we utilized a polyclonal rabbit antibody at 1:80 dilutions (LabVision, Catalog #RB-1536-P). The antibody was derived against a synthetic peptide from the middle region of the human MMP-1 and can recognize both the 52 kDa (unglycosylated)/57 kDa (glycosylated) species of the pro-form and 42 kDa (unglycosylated)/47 kDa (glycosylated) active form of MMP-1 (32, 33). Detailed information on antibody specificity is as earlier described (32, 33).

All IHC was performed at MDR Global Inc, Windber, PA, using standardized procedures and protocols. Briefly, each section was placed on positively charged slides, which were then dried at 59°C, before placing in SubX solution prior to deparaffinization and antigen recovery using Dako PT Module. The nonenzymatic antigen recovery step was performed in a patented microwave oven and the slides were washed thoroughly in distilled water before placing in TBS buffer. Immunostaining was performed on the Dako Autostainer with the application of primary antibody followed by washing the slides again with TBS buffer. The appropriate detection reagent was then dispensed onto the slides followed by another wash in TBS buffer and diaminobenzidine "DAB"/hydrogen peroxide for color visualization. All the slides were counterstained with hematoxylin. Preanalytical processes such as antigen retrieval and dilution determination were performed on a known positive control, in this case, placenta (syncytiotrophoblast layer). Placenta provided

Table 1. Clinicopathological characteristics of patients with breast cancer

Parameters		n = 136 (%)
Age (years)	<59	58 (43)
	≥59	61 (45)
Ethnicity	AA	29 (21)
	White	77 (57)
	Other	13 (10)
Menopausal status	Premenopausal	36 (26)
	Postmenopausal	82 (60)
Tumor stage	0	64 (47)
	I	22 (16)
	II	26 (19)
	III	9 (7)
	IV	8 (6)
Histological grade	G1	14 (10)
	G2	34 (25)
	G3	29 (21)
Tumor size	T1	29 (21)
	T2	32 (24)
	T3	6 (4)
Nodal status	Negative	97 (71)
	Positive	32 (24)
ER	Negative	24 (18)
	Positive	88 (65)
PR	Negative	48 (35)
	Positive	64 (47)
HER2	Negative	40 (29)
	Positive	30 (22)
P53	Negative	29 (21)
	Positive	28 (21)

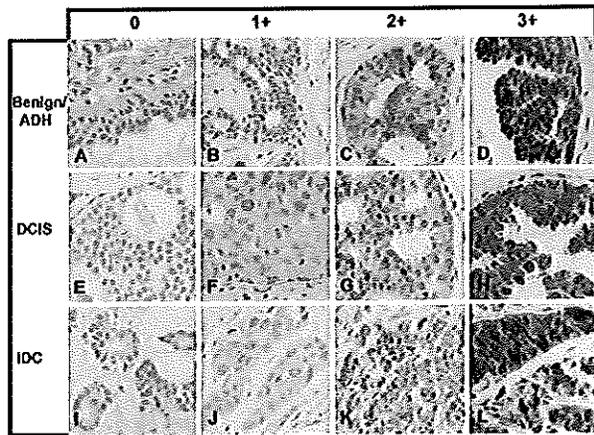


Figure 1. Immunostaining of MMP-1 expression in benign breast disease, atypical ductal hyperplasia, and malignant breast tumor. MMP-1 expression in breast tissues was visualized by immunohistochemical staining with a specific rabbit antihuman polyclonal antibody. Staining was predominantly in the cytoplasm of normal and neoplastic cells. Examples of MMP-1 staining in (A–B) benign breast disease; (C–D) atypical ductal hyperplasia (ADH); (E–H) ductal carcinoma *in situ* (DCIS); and (I–L) invasive ductal carcinoma (IDC). Magnification: 200 $\times$  for all pictures.

internal negative control as well (blood vessels and lymphatics) and these elements were evaluated in each run. The immunohistochemistry result for MMP-1 was assessed according to MMP-1 staining intensity in tumor cells as follows: 0, no staining; 1+, mild staining; 2+, moderate staining; and 3+, intense staining (Figure 1). Available IHC data for ER, PR, HER2, and P53 were queried from the clinical database of the CBCP. Evaluation of the expression of ER, PR, and P53 was based on the percentage of nuclear staining of the tumor cells (for ER/PR: 0–5% negative,  $\geq$ 5% positive; for P53: 0–50% negative,  $\geq$ 50% positive). HER2 expression was assessed according to ASCO/CAP guideline and was based on the intensity and percentage of the membrane staining of tumor cells: 0/1+, negative; 2+, reflex to FISH; and 3+, positive (34). Detailed information on the antibodies used to detect the protein markers is provided in Supplementary Table 1.

#### Statistical analysis

Statistical analyses were carried out using SAS 9.1 for Windows (SAS Institute, Cary, NC). The association between MMP-1 expression/polymorphism and disease status was determined using chi-square or Fisher's exact test when the

numbers of cases were small. Jonckheere–Terpstra test for trend was used to assess the correlation between MMP-1 genotype and expression level. Jonckheere–Terpstra test is a nonparametric statistical test, which examines the ordered differences of MMP-1 expression level among polymorphism classes. Odds ratios (ORs) and 95% confidence intervals (CIs) were calculated by logistic regression. In all cases,  $p < .05$  was required for significance.

## RESULTS

### MMP-1 expression in subjects

MMP-1 was primarily detected in the cytoplasm of normal and neoplastic cells (Figure 1). In the ADH group, 91% (10 of 11) of the patients expressed MMP-1 at high levels (2+/3+), while in the benign group, only 48% (12 of 25) had high levels of expression of MMP-1. The expression of MMP-1 was significantly associated with the disease status ( $p = .016$ ; Table 2). Compared to subjects expressing low levels (0/1+) of MMP-1, the ORs for subjects expressing high levels (2+/3+) of MMP-1 to develop ADH increased 10.83 times (OR = 10.83; 95% CI, 1.2–97.8; Table 2). When the expression of MMP-1 in IBC was compared with *in situ* breast cancer (DCIS/LCIS), we found that 58% of the IBC patients had high MMP-1 expression, which was significantly higher than that of the DCIS/LCIS group ( $p = .014$ ; Table 2). Patients expressing high levels of MMP-1 had a 2.20-fold increased risk of developing IBC (OR = 2.20; 95% CI, 1.09–4.44; Table 2).

### The distribution of MMP-1 1G/2G genotype in subjects

The distribution of 1G/2G genotypes in the 132 selected patients was according to the Hardy–Weinberg equilibrium ( $\chi^2 = 0.26$ ;  $p = 0.88$ ). There was no significant difference in the distribution of the MMP-1 1G/2G genotypes between benign and ADH groups or between IBC and DCIS/LCIS groups ( $p > .05$ ; Table 3). Allele frequencies between these groups were also not statistically significant ( $p > .05$ ; Table 3).

### Correlation of MMP-1 expression and MMP-1 genotypes

We observed a significant association between the expression of MMP-1 and MMP-1 genotypes in the IBC group ( $p = .032$ ; Table 4). In this group, 57% of the patients expressing low MMP-1 had the 1G/1G homozygote genotype while 80% of those expressing high MMP-1 (3+) were either 1G/2G carriers (40%) or homozygous for 2G (40%). These data are

Table 2. The expression of MMP-1 in subjects with breast diseases

	MMP1 expression		OR	95% CI	p
	Low (0/1+) n (%)	High (2+/3+) n (%)			
Benign (n = 25)	13 (52)	12 (48)	1		
ADH (n = 11)	1 (9)	10 (91)	10.83	1.20–97.80	.016 <sup>a</sup>
DCIS/LCIS (n = 64)	39 (61)	25 (39)	1		
IBC (n = 65)	27 (42)	38 (58)	2.2	1.09–4.44	.014 <sup>b</sup>

<sup>a</sup>The expression of MMP-1 in ADH is significantly higher than in benign breast disease ( $p < .05$ , Fisher's exact test).

<sup>b</sup>The expression of MMP-1 in invasive breast cancer is significantly higher than in *in situ* breast cancer ( $p < .05$ ,  $\chi^2$  test).

ADH, atypical ductal hyperplasia; DCIS, ductal carcinoma *in situ*; LCIS, lobular carcinoma *in situ*; IBC-invasive breast cancer.

Table 3. MMP-1 genotype distribution and allele frequencies in patients with breast diseases

Genotype	Benign (n = 17)		ADH (n = 13)		DCIS/LCIS (n = 50)		IBC (n = 52)	
	n	f	n	f	n	f	n	f
1G/1G	3	0.18	5	0.39	10	0.20	19	0.36
1G/2G	10	0.59	6	0.46	29	0.58	18	0.35
2G/2G	4	0.24	2	0.15	11	0.22	15	0.29
p	.068 <sup>a</sup>				.051 <sup>b</sup>			
Allele								
1G	16	0.47	16	0.62	49	0.49	56	0.54
2G	18	0.53	10	0.38	51	0.51	48	0.46
p	.197 <sup>a</sup>				.489 <sup>b</sup>			

<sup>a</sup>The difference in genotype distribution or allele frequencies between benign breast disease and ADH groups is not statistically significant ( $p > .05$ , Fisher's exact test).

<sup>b</sup>The difference in genotype distribution or allele frequencies between *in situ* and invasive breast cancer groups is not statistically significant ( $p > .05$ ,  $\chi^2$  test).

ADH, atypical ductal hyperplasia; DCIS, ductal carcinoma *in situ*; LCIS, lobular carcinoma *in situ*; IBC, invasive breast cancer.

displayed graphically in Figure 2(A). The 2G allele frequency for IBC patients expressing low MMP-1 was 0.29 (for 1+) and 0.44 (for 2+), while IBC patients expressing high MMP-1 (3+) had a 2G allele frequency of 0.6 (Figure 2(B)). Thus, in the IBC patients, the expression of MMP-1 was positively correlated with the 2G allele. This association between MMP-1 2G insertion polymorphism and MMP-1 expression was not observed in the DCIS/LCIS group ( $p = .23$ ; Table 4).

#### Correlation of MMP-1 expression with clinicopathological parameters

Correlation of MMP-1 expression with clinicopathological parameters in patients with breast cancer is summarized in Table 5. We did not observe a significant correlation of MMP-1 expression with age, ethnicity, menopausal status, tumor stage, tumor size, lymph node metastasis, or ER, PR, and P53 status. However, we observed a significant correlation of MMP-1 expression with HER2 status ( $p = .0039$ ). In the HER2- group, 71% (27 of 38) of patients had high MMP-1 expression, while in HER2+ group, only 35% (9 of 26) of the patients expressed high levels of MMP-1 (Table 5). We also found marginal correlation of MMP-1 expression with tumor histological grade ( $p = .0496$ ; Table 5).

Table 4. Correlation between MMP-1 expression and genotype in patients with invasive and *in situ* breast cancer

Genotype	MMP-1 expression					
	IBC (n = 41)			DCIS/LCIS (n = 33)		
	1+ (n = 14)	2+ (n = 17)	3+ (n = 10)	1+ (n = 12)	2+ (n = 12)	3+ (n = 9)
	n (%)	n (%)	n (%)	n (%)	n (%)	n (%)
1G/1G	8 (57)	6 (35)	2 (20)	2 (17)	3 (25)	0 (0)
1G/2G	4 (29)	7 (41)	4 (40)	7 (58)	5 (42)	6 (67)
2G/2G	2 (14)	4 (24)	4 (40)	3 (25)	4 (33)	3 (33)
p	.032			.230		

Jonckheere-Terpstra test.

DCIS, ductal carcinoma *in situ*; LCIS, lobular carcinoma *in situ*; IBC, invasive breast cancer.

#### Correlation of MMP-1 genotype with clinicopathological parameters

We analyzed the data to determine if the MMP-1 2G insertion polymorphism had any association with other clinicopathological parameters in patients with breast cancer. No significant correlation between genotype distribution of MMP-1 with age, ethnicity, menopausal status, lymph node metastasis, tumor stage, histograde, tumor size, and ER/PR status was observed ( $p > .05$ ; Table 6). However, we observed a significant difference in the genotype distribution of the MMP-1 between the HER2+ and HER2- group in IBC patients (Table 6). Notably, the percentage of 1G/1G homozygote was similar in the two groups, but the percentage of 1G/2G heterozygotes was significantly decreased and the percentage of the 2G/2G homozygotes was significantly increased in HER2+ group compared to the HER2- group ( $p = .0401$ ; Table 6). Patients carrying the 2G/2G genotype had a 7.5-fold (95% CI, 1.48-37.9;  $p = .032$ ; Table 6) increased odds of expressing HER2 compared to the 1G/2G carriers. However, the odds for 1G/1G carriers expressing HER2 was only increased by 2.78-fold, which was not statistically significant ( $p = .29$ ).

We also observed a significant association between MMP-1 genotype distribution and the expression of P53 in IBC patients ( $p = .0399$ ; Table 6). Compared with 1G/2G carriers, the odds for the 2G/2G carriers to express P53 increased by 5.88-fold (95% CI, 5.26-33.3;  $p = .049$ ), while the odds for the 1G/1G carriers to express P53 increased by 5.26-fold (95% CI, 1.18-25;  $p = .038$ ).

#### DISCUSSION

##### Expression of MMP-1 on breast diseases development

We determined the expression of MMP-1 in patients with benign breast disease, ADH (a precancerous condition that poses an increased risk of breast cancer development), DCIS/LCIS (*in situ*), and IBC. We observed significantly increased expression of MMP-1 in ADH patients compared to benign breast disease ( $p = .016$ ) and a dramatic increase in the expression of MMP-1 in IBC compared to DCIS/LCIS ( $p = .014$ ). Interestingly, we also found that the expression of MMP-1 in the *in situ* foci of IBC tumors is significantly higher than that of pure DCIS specimens ( $p = .035$ , data not shown) but not significantly different from that of the invasive foci of IBC specimens ( $p = .595$ , data not shown).

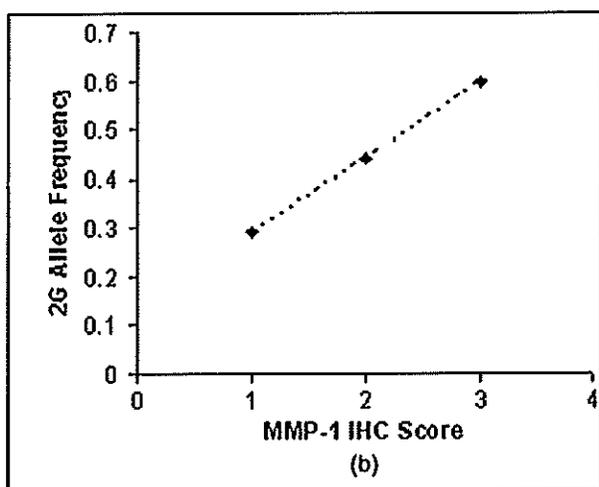
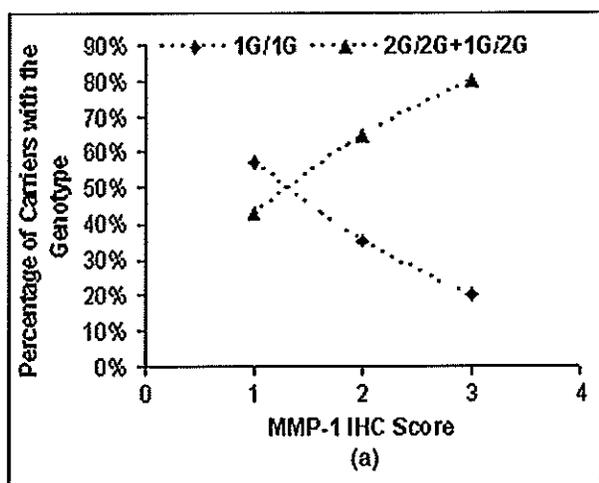


Figure 2 (A). MMP-1 2G insertion polymorphism is correlated with MMP-1 expression in invasive breast cancer (IBC). Subjects with IBC were grouped according to the expression level of MMP-1 (1+, 2+, and 3+). The percentage of homozygotes (1G/1G and 2G/2G) and heterozygotes (1G/2G) was then plotted against MMP-1 expression level based on the immunohistochemical score of 1+, 2+, and 3+. As MMP-1 expression increased, the percentage of 1G/1G homozygotes decreased, while the percentage of 2G carriers (1G/2G heterozygotes) and homozygotes (2G/2G) increased. Results show a relationship between 2G insertion and MMP-1 expression in IBC patients. (B) The MMP-1 2G allele frequency correlates with MMP-1 expression level. Allele frequency of subjects in the IBC group was plotted against the expression level of MMP-1 (1+, 2+, and 3+) as obtained from immunohistochemistry readings. The 2G allele frequency increased with increasing expression of MMP-1.

These observations indicate an association of MMP-1 expression with breast disease status/severity and are consistent with earlier studies that support (a) the role of MMP-1 as a candidate precancerous marker for early detection of breast cancer (18) and (b) the notion that overexpression of MMP-1 is important for the invasion and progression of breast cancer (19–21).

MMP-1 has been reported as a useful marker for the identification of ADH patients without a history of cancer who would likely develop breast cancer in the future (17). Our

limited data on follow-up history of these ADH patients prevent us from further discussion on the role of MMP-1 in identifying the ADH patients with a potential to develop breast cancer.

#### The role of MMP-1 2G polymorphism on MMP-1 expression and breast cancer severity

The reported MMP-1 genotype frequencies in healthy Caucasian women are 0.28, 0.54, and 0.18 for 1G/1G, 1G/2G, and 2G/2G, respectively (35), and this is not significantly different from the genotype frequency we observed in IBC, DCIS/LCIS, ADH, or benign breast disease group ( $p > .05$ ). The distribution of the MMP-1 genotype in our studied groups was also not significantly different ( $p > .05$ ), suggesting that MMP-1 polymorphism may not contribute to the development of breast cancer. This is consistent with the result from a recent meta-analysis study, which failed to show any significant association between MMP-1 2G polymorphism and breast cancer susceptibility (36).

However, we observed an association between the 2G genotype and MMP-1 expression in the IBC patient group ( $p = .032$ ; Table 4) but not in the DCIS/LCIS group ( $p > .05$ ; Table 4). This correlation was also not observed in ADH or benign group. These observations indicate that MMP-1 2G polymorphism may contribute significantly to the increased expression of MMP-1 in IBC but not in the *in situ* breast cancer or the less severe breast disease conditions, and may thus correlate with breast disease status or breast cancer severity. However, our limited sample size prevents us from making any strong conclusions. A larger sample size study is required to establish the full potential of MMP-1 polymorphism in breast cancer progression.

It is worth noting that while the 2G polymorphism may contribute to increased expression of MMP-1 in IBC, other factors may contribute more significantly to increased expression of MMP-1 in ADH group. It has been reported that MMP-1 expression can be increased by many stimuli, including growth factors, inflammatory interleukins, heat, and UV-light, which can stimulate mitogen-activated protein kinases (MAPKs) and activate MMP-1 transcription (37–40).

#### Correlation of MMP-1 2G polymorphism with other clinical markers of breast cancer

Higher 2G allele frequencies have been reported in breast cancer patients with lymph node metastasis compared with patients without metastasis, indicating the role of MMP-1 2G polymorphism as a potential prognostic marker for breast cancer (22, 23, 41). In our study, although we observed increased 2G/2G genotype frequency (35%) in the lymph node positive group compared with lymph node negative group (19%), the difference did not reach statistical significance ( $p = .098$ ).

However, we did find a correlation between MMP-1 genotype with other breast cancer markers, such as HER2 and P53. There was a significant association between the MMP-1 2G polymorphism and HER2 expression in IBC patients ( $p = .0401$ ; Table 6). Forty-four percent (10 of 23) of patients expressing HER2 had the homozygous 2G/2G genotype

Table 5. Correlation of MMP-1 expression with clinicopathological parameters in patients with breast cancer

Parameters	MMP-1 expression		Total n	p
	Low (0/1+) n (%)	High (2+/3+) n (%)		
Age (years)				
≤59	36 (65)	19 (35)	55	.0545
>59	26 (47)	29 (53)	55	
Ethnicity				
African-American	14 (58)	10 (42)	24	.7142
White	40 (54)	34 (46)	74	
Menopausal				
Premenopausal	19 (59)	13 (41)	32	.6437
Postmenopausal	42 (55)	35 (45)	77	
Stage				
0	40 (65)	22 (35)	61	.057
I	8 (38)	13 (62)	21	
II	9 (39)	14 (61)	23	
III/IV	9 (64)	5 (36)	14	
Histological grade				
G1	7 (54)	6 (46)	13	.0496
G2	8 (25)	24 (75)	32	
G3	13 (54)	11 (46)	24	
Tumor size				
≤2 cm	10 (37)	17 (63)	27	.3177
>2cm	16 (50)	16 (50)	32	
Lymph node status				
Negative	49 (54)	41 (46)	90	.9355
Positive	15 (54)	13 (46)	28	
ER status				
Negative	8 (40)	12 (60)	20	.1959
Positive	46 (56)	36 (44)	82	
PR status				
Negative	23 (52)	21 (48)	44	.8328
Positive	31 (54)	26 (46)	57	
HER2/neu protein				
Negative	11 (29)	27 (71)	38	.0039
Positive	17 (65)	9 (35)	26	
P53 status				
Negative	9 (39)	14 (61)	23	.7435
Positive	9 (35)	17 (65)	26	

p values < 0.05 are required for statistical.

compared to 15% of the patients (4 of 26) not expressing HER2 (Table 6). The distribution curve of the genotype shifted to a higher frequency of 2G homozygotes in HER2+ group. In breast cancer, the amplification of the Her2/neu oncogene is a well-established marker, which is associated with relapse and short survival (42, 43). HER2 overexpression has also been correlated with poor outcome (44–46). The association of the MMP-1 2G insertion polymorphism with this clinical biomarker again indicates the potential role of this polymorphism in predicting breast cancer severity.

Studies have shown that both MMP-1 2G polymorphism and HER2 overexpression can upregulate MMP-1 expression via different molecular mechanisms (20, 21, 47). The 2G single nucleotide polymorphism of MMP-1 affects the basal transcription of MMP-1 through regulating ETS-1 and AP-1 (20), while HER2 can further increase MMP-1 expression through activating ER81, another member of ETS family (47). In the current study, we did not observe a positive correlation between HER2 and MMP-1 expression in the patients with breast cancer. On the contrary, the MMP-1 expression was higher in HER2-negative group compared with HER2-positive group (Table 5). This could be due to un-

known mechanisms, for example, a negative feedback effect from increased MMP-1 expression on HER2 transcription or a result of the overall effect of different regulators on MMP-1 expression. We did, however, observe a correlation between the 2G insertion polymorphism and MMP-1 expression in the IBC patients only ( $p = .032$ ) and not in the *in situ* group, which supports the notion that MMP-1 2G polymorphism is associated with disease severity in breast cancer.

Overexpression of P53 is also associated with breast cancer progression. Both P53 mutation and accumulation have been associated with poor prognosis (48–51). Thor et al. found significant correlation of P53 accumulation with short metastasis-free and overall survival in both sporadic and familial breast cancers (51). Wild-type P53 can downregulate MMP-1 expression by disrupting the communication between AP-1 and basal transcriptional complex in human foreskin fibroblast cells (52, 53). However, it has been reported that the repression effect of P53 on MMP-1 expression is lost in osteogenic sarcoma cells overexpressing mutant P53 (52). More detailed study on the effect of the accumulated or mutant P53 on MMP-1 expression in breast cancer will be required to gain a better understanding of the

Table 6. Correlation of MMP-1 genotype and clinicopathological characteristics in patients with breast cancer

Parameters	Genotype			Total n	p
	1G/1G n (%)	1G/2G n (%)	2G/2G n (%)		
Age (years)					
≤55	13 (33)	18 (45)	10 (25)	40	.8624
>55	15 (36)	16 (38)	11 (26)	42	
Ethnicity					
African-American	7 (37)	9 (47)	3 (16)	19	.6340
White	20 (33)	24 (39)	17 (28)	61	
Menopausal					
Premenopausal	9 (32)	13 (46)	6 (21)	28	.7592
Postmenopausal	18 (33)	21 (39)	15 (28)	54	
Stage					
0	10 (23)	25 (57)	9 (20)	44	.2323
I	7 (39)	6 (33)	5 (28)	18	
II	6 (38)	7 (44)	3 (19)	16	
III/IV	4 (36)	2 (18)	5 (45)	11	
Histological grade					
G1	6 (60)	3 (30)	1 (10)	10	.5673
G2	6 (29)	8 (38)	7 (33)	21	
G3	8 (38)	7 (33)	6 (29)	21	
Tumor size					
<2 cm	8 (36)	8 (36)	6 (27)	22	.9759
>2cm	9 (38)	8 (33)	7 (29)	24	
Lymph node status					
Negative	20 (29)	35 (51)	13 (19)	68	.0982
Positive	8 (40)	5 (25)	7 (35)	20	
ER status					
Negative	7 (47)	3 (20)	5 (33)	15	.1050
Positive	19 (31)	30 (49)	12 (20)	61	
PR status					
Negative	12 (38)	11 (34)	9 (28)	32	.1217
Positive	14 (32)	22 (50)	8 (18)	44	
HER2/neu protein					
Negative	10 (38)	12 (46)	4 (15)	26	.0401
Positive	9 (39)	4 (17)	10 (43)	23	
P53 status					
Negative	5 (24)	13 (62)	3 (14)	21	.0399
Positive	10 (45)	5 (23)	7 (32)	22	

p values < 0.05 are required for statistical.

interrelationship between P53 and MMP-1. So far, we did not observe a significant association between MMP-1 and P53 expression in our study (Table 5). As we have mentioned previously that MMP-1 level can be affected by many factors (37–40), our result indicated that P53 may be a minor contributor compared to other regulators in deterring MMP-1 levels in breast cancer. However, we observed a significant difference in the distribution of MMP-1 genotypes between P53+ and P53– patients ( $p = .0399$ ; Table 6). There was a significant decrease in the number of heterozygotes (1G/2G) and an increase in the number of homozygotes (1G/1G and 2G/2G) in P53+ patients compared to P53– patients ( $p = .010$ ). The overall effect of this difference on MMP-1 expression in P53+/- patients may need further investigation. Although the observed pattern of MMP-1 genotype distribution in P53 expressing tumors varied from that observed in the HER2 expressing tumors, these observations both point to some level/degree of connectivity between the MMP-1 polymorphism and breast markers associated with disease severity.

In this study, we provide evidence that overexpression of MMP-1 is associated with both increased risk and local invasion in breast cancer. While the MMP-1 2G insertion polymorphism may not contribute to either event, other regulators must be involved in upregulating MMP-1 levels in the different steps leading to breast cancer onset and invasion. However, the MMP-1 2G polymorphism is associated with breast cancer severity, which is supported by its correlation with MMP-1 expression and the expression status of HER2 and P53 in IBC patients. Overexpression of MMP-1 induced by this polymorphism together with other regulators may predict an adverse outcome of breast cancer. More comprehensive studies are required to fully understand the mechanism of the interaction of this polymorphism with other molecular markers, and to explore the impact of this polymorphism alone or with other clinical markers on disease-free or overall survival in breast cancer. Such studies will elucidate the impact of MMP-1 2G polymorphism on breast cancer development and progression.

## CONCLUSION

To study the impact of MMP-1 1G/2G polymorphism on breast cancer development, we genotyped 132 patients with benign breast disease, ADH, DCIS/LCIS, and IBC and analyzed our results with expression data of the following breast cancer markers, MMP-1, ER, PR, HER2, and P53. We found that overexpression of MMP-1 is associated with breast cancer onset and invasion. We report here, for the first time, that MMP-1 2G insertion polymorphism is significantly associated with HER2 and P53 expression. HER2 overexpression, P53 accumulation, or the coexistence of these two, are known molecular markers in breast cancer progression (42, 44, 46, 51, 54). The association of MMP-1 2G insertion polymorphism with these two biomarkers in IBC indicates a role of MMP-1 2G polymorphism in predicting breast cancer severity and may provide new molecular tools for subclassification of IBC.

## ACKNOWLEDGMENT

This research was supported by a grant from the United States Department of Defense (Military Molecular Medicine Initiative MDA W81XWHH-05-2-0075, Protocol 01-20006), and by NIH-AR-26599 and NIH-CA-77267 to CB. The opinion and assertions contained herein are the private views of the authors and are not to be construed as official or as representing the views of the Department of the Army or the Department of Defense.

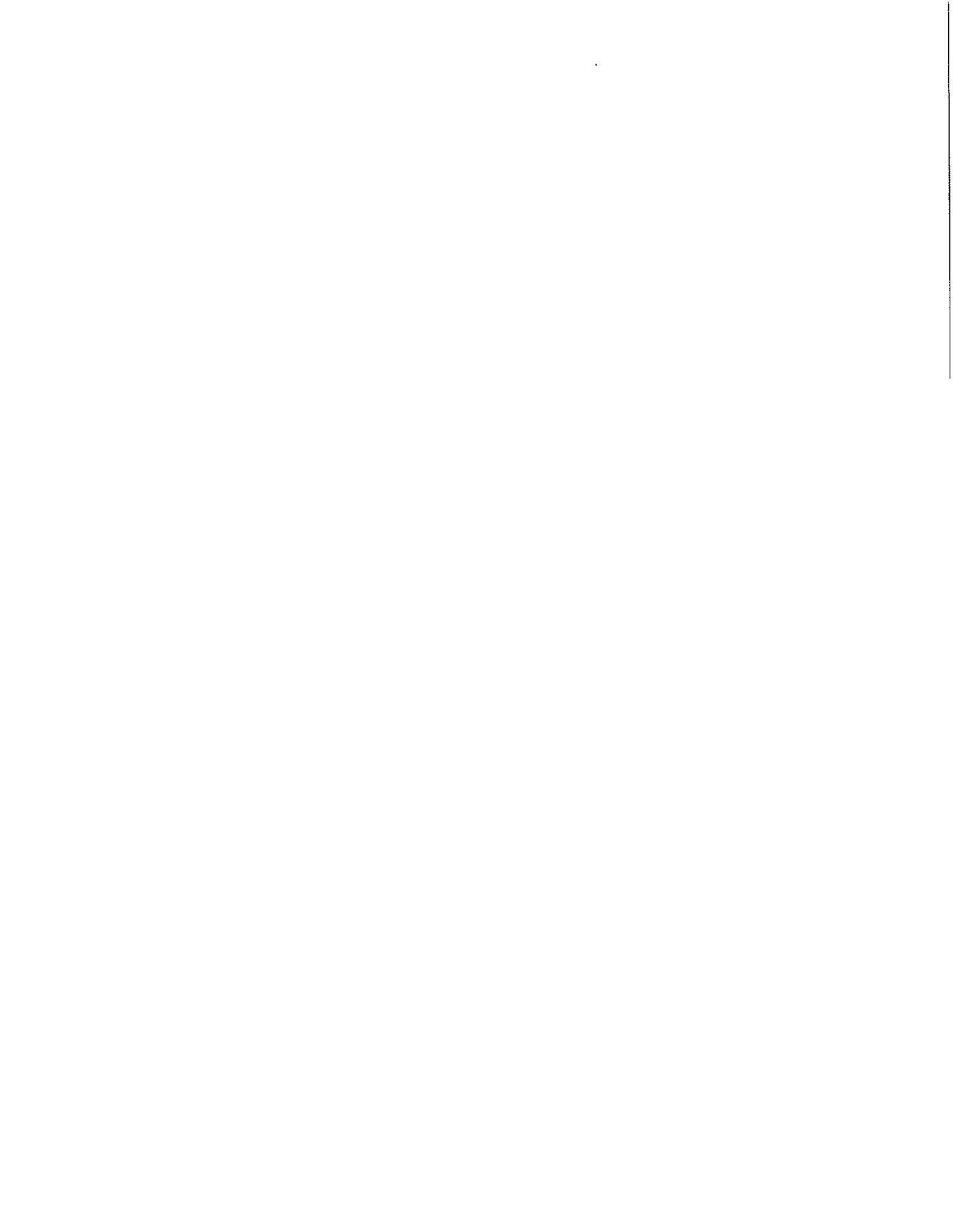
## DECLARATION OF INTEREST

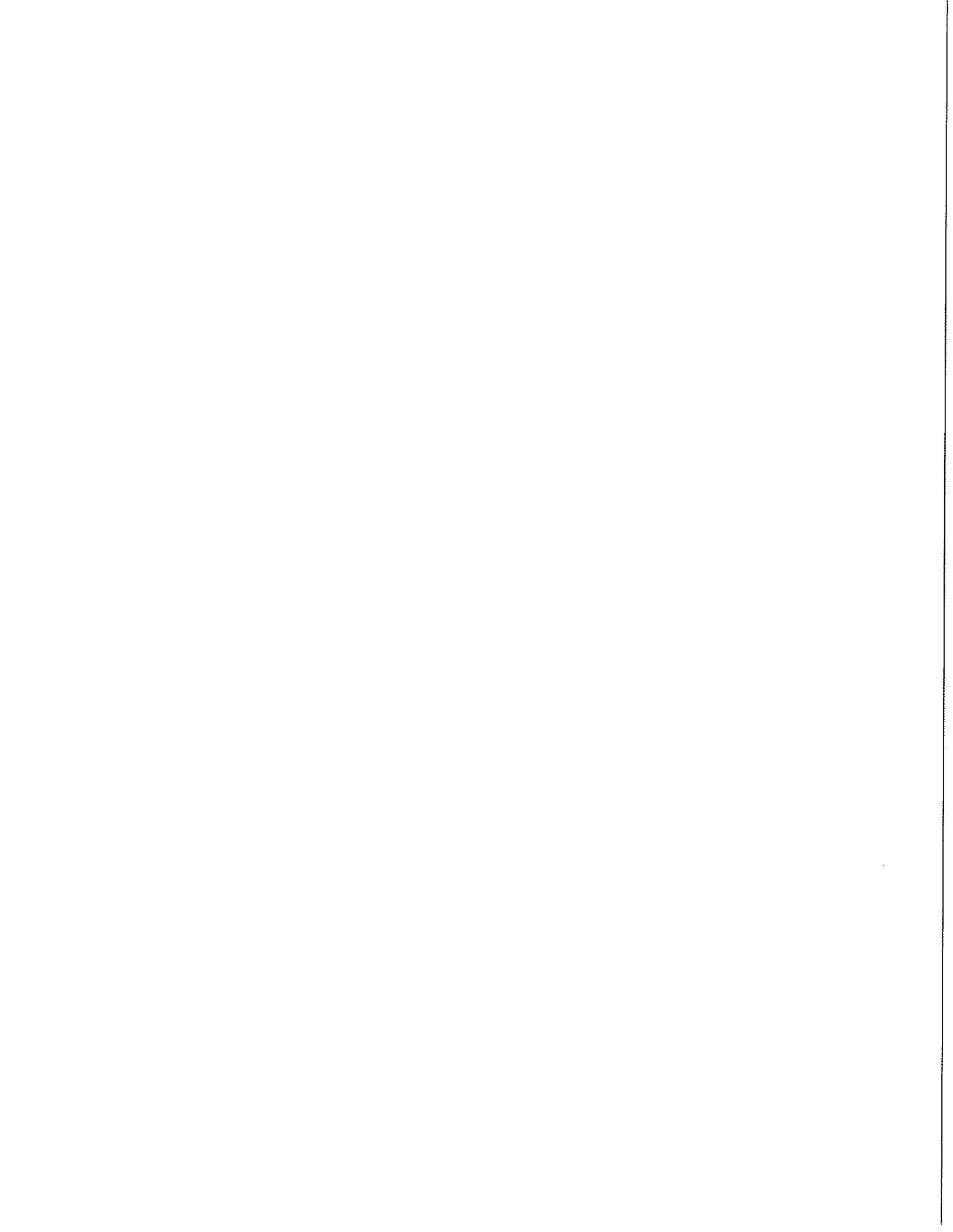
The authors report no declarations of interest. The authors alone are responsible for the content and writing of the paper.

## REFERENCES

- Menon B, Singh M, Singh K. Matrix metalloproteinases mediate beta-adrenergic receptor-stimulated apoptosis in adult rat ventricular myocytes. *Am J Physiol Cell Physiol* 2005;289:C168–C176.
- McCawley LJ, Wright J, LaFleur BJ, Crawford HC, Matrisian LM. Keratinocyte expression of MMP3 enhances differentiation and prevents tumor establishment. *Am J Pathol* 2008;173:1528–1539.
- Wang W, Pan H, Murray K, Jefferson BS, Li Y. Matrix metalloproteinase-1 promotes muscle cell migration and differentiation. *Am J Pathol* 2009;174:541–549.
- Zeng ZZ, Yao H, Staszewski ED, Rockwood KF, Markwart SM, Fay KS, Spalding AC, Livant DL. Alpha(5)beta(1) integrin ligand PHSRN induces invasion and alpha(5) mRNA in endothelial cells to stimulate angiogenesis. *Transl Oncol* 2009;2:8–20.
- Shan B, Morris CA, Zhuo Y, Shelby BD, Levy DR, Lasky JA. Activation of proMMP-2 and Src by HHV8 vGPCR in human pulmonary arterial endothelial cells. *J Mol Cell Cardiol* 2007;42:517–525.
- Gonzalez Rodriguez I, Rivas del Fresno M, Gil Ugarteburu R, Gonzalez Arriaga P, Lopez Cima F, Fernandez Samoano A, Muramendiarez Fernandez V, Fernandez Garcia I, Perez-Carral JR, Tardon A. Expression of matrix metalloproteinase-9 in prostate cancer. Preliminary experience. *Arch Esp Urol* 2010;63:119–124.
- Szarvas T, Becker M, vom Dorp F, Gethmann C, Totsch M, Bankfalvi A, Schmid KW, Romics I, Rubben H, Ergun S. Matrix metalloproteinase-7 as a marker of metastasis and predictor of poor survival in bladder cancer. *Cancer Sci* 2010;101:1300–1308.
- Yi YC, Chou PT, Chen LY, Kuo WH, Ho ES, Han CP, Yang SF. Matrix metalloproteinase-7 (MMP-7) polymorphism is a risk factor for endometrial cancer susceptibility. *Clin Chem Lab Med* 2010;48:337–344.
- Kim YW, Ko YT, Kim NK, Chung HC, Min BS, Lee KY, Park JP, Kim H. A comparative study of protein expression in primary colorectal cancer and synchronous hepatic metastases: the significance of matrix metalloproteinase-1 expression as a predictor of liver metastasis. *Scand J Gastroenterol* 2010;45:217–225.
- Garg P, Sarma D, Jeppsson S, Patel NR, Gewirtz AT, Merlin D, Sitaraman SV. Matrix metalloproteinase-9 functions as a tumor suppressor in colitis-associated cancer. *Cancer Res* 2010;70:792–801.
- Li M, Xiao T, Zhang Y, Feng L, Lin D, Liu Y, Mao Y, Guo S, Han N, Di X, Zhang K, Cheng S, Gao Y. Prognostic significance of matrix metalloproteinase-1 levels in peripheral plasma and tumour tissues of lung cancer patients. *Lung Cancer* 2010;69:341–347.
- Luukkaa H, Klemi P, Hirsimaki P, Vahlberg T, Kivisaari A, Kahari VM, Grenman R. Matrix metalloproteinase (MMP)-7 in salivary gland cancer. *Acta Oncol* 2010;49:85–90.
- Luukkaa H, Klemi P, Leivo I, Makitie AA, Irish J, Gilbert R, Perez-Ordóñez B, Hirsimaki P, Vahlberg T, Kivisaari A, Kahari VM, Grenman R. Expression of matrix metalloproteinase-1, -7, -9, -13, Ki-67, and HER-2 in epithelial-myoepithelial salivary gland cancer. *Head Neck* 2010;32:1019–1027.
- Yeh YC, Sheu BS, Cheng HC, Wang YL, Yang HB, Wu JJ. Elevated serum matrix metalloproteinase-3 and -7 in *H. pylori*-related gastric cancer can be biomarkers correlating with a poor survival. *Dig Dis Sci* 2010;55:1649–1657.
- Scherf DB, Dally H, Muller P, Werle-Schneider G, Jager B, Edler I, Tuengerthal S, Fischer JR, Drings P, Bartsch H, Risch A. Single nucleotide polymorphisms in matrix metalloproteinase genes and lung cancer chemotherapy response and prognosis. *Eur Respir J* 2010;35:381–390.
- Zohny SF, Fayed ST. Clinical utility of circulating matrix metalloproteinase-7 (MMP-7), CC chemokine ligand 18 (CCL18) and CC chemokine ligand 11 (CCL11) as markers for diagnosis of epithelial ovarian cancer. *Med Oncol* 2010 Dec;27(4):1246–1253. Epub 2009 Nov 24.
- Brinckerhoff CE, Matrisian LM. Matrix metalloproteinases: a tail of a frog that became a prince. *Nat Rev Mol Cell Biol* 2002;3:207–214.
- Poola I, DeWitty RL, Marshalleck JJ, Bhatnagar R, Abraham J, Leffall LD. Identification of MMP-1 as a putative breast cancer predictive marker by global gene expression analysis. *Nat Med* 2005;11:481–483.
- Cheng S, Tada M, Hida Y, Asano T, Kuramae T, Takemoto N, Hamada J, Miyamoto M, Hirano S, Kondo S, Moriuchi T. High MMP-1 mRNA expression is a risk factor for disease-free and overall survivals in patients with invasive breast carcinoma. *J Surg Res* 2008;146:104–109.
- Tower GB, Coon CI, Brinckerhoff CE. The 2G single nucleotide polymorphism (SNP) in the MMP-1 promoter contributes to high levels of MMP-1 transcription in MCF-7/ADR breast cancer cells. *Breast Cancer Res Treat* 2003;82:75–82.
- Rutter JL, Mitchell TI, Buttice G, Meyers J, Gusella JF, Ozelius LJ, Brinckerhoff CE. A single nucleotide polymorphism in the matrix metalloproteinase-1 promoter creates an Ets binding site and augments transcription. *Cancer Res* 1998;58:5321–5325.
- Przybylowska K, Zielinska J, Zadrozny M, Krawczyk T, Kulig A, Wozniak P, Rykala J, Kolacinska A, Morawiec Z, Drzewoski J, Blasiak J. An association between the matrix metalloproteinase 1 promoter gene polymorphism and lymph node metastasis in breast cancer. *J Exp Clin Cancer Res* 2004;23:121–125.
- Hughes S, Agbaje O, Bowen RL, Holliday DL, Shaw JA, Duffy S, Jones JL. Matrix metalloproteinase single-nucleotide polymorphisms and haplotypes predict breast cancer progression. *Clin Cancer Res* 2007;13:6673–6680.
- McColgan P, Sharma P. Polymorphisms of matrix metalloproteinases 1, 2, 3 and 9 and susceptibility to lung, breast

- and colorectal cancer in over 30,000 subjects. *Int J Cancer* 2009;125:1473–1478.
25. Kondo S, Wakisaka N, Schell MJ, Horikawa T, Sheen TS, Sato H, Furukawa M, Pagano JS, Yoshizaki T. Epstein-Barr virus latent membrane protein 1 induces the matrix metalloproteinase-1 promoter via an Ets binding site formed by a single nucleotide polymorphism: enhanced susceptibility to nasopharyngeal carcinoma. *Int J Cancer* 2005;115:368–376.
  26. Ghilardi G, Biondi ML, Mangoni J, Leviti S, DeMonti M, Guagnellini E, Scorza R. Matrix metalloproteinase-1 promoter polymorphism 1G/2G is correlated with colorectal cancer invasiveness. *Clin Cancer Res* 2001;7:2344–2346.
  27. Vairaktaris E, Yapijakis C, Derka S, Serefolglou Z, Vassiliou S, Nkenke E, Ragos V, Vylliotis A, Spyridonidou S, Tsigris C, Yannopoulos A, Tesseromatis C, Neukam FW, Patsouris E. Association of matrix metalloproteinase-1 (-1607 1G/2G) polymorphism with increased risk for oral squamous cell carcinoma. *Anticancer Res* 2007;27:459–464.
  28. O-charoenrat P, Leksrisakul P, Sangruchi S. A functional polymorphism in the matrix metalloproteinase-1 gene promoter is associated with susceptibility and aggressiveness of head and neck cancer. *Int J Cancer* 2006;118:2548–2553.
  29. Hashimoto T, Uchida K, Okayama N, Imate Y, Suehiro Y, Hamanaka Y, Ueyama Y, Shinozaki F, Yamashita H, Hinoda Y. Association of matrix metalloproteinase (MMP)-1 promoter polymorphism with head and neck squamous cell carcinoma. *Cancer Lett* 2004;211:19–24.
  30. Nishioka Y, Kobayashi K, Sagae S, Ishioka S, Nishikawa A, Matsushima M, Kanamori Y, Minaguchi T, Nakamura Y, Tokino T, Kudo R. A single nucleotide polymorphism in the matrix metalloproteinase-1 promoter in endometrial carcinomas. *Jpn J Cancer Res* 2000;91:612–615.
  31. Six L, Grimm C, Leodolter S, Tempfer C, Zeillinger R, Sliutz G, Speiser P, Reinthaller A, Hefler LA. A polymorphism in the matrix metalloproteinase-1 gene promoter is associated with the prognosis of patients with ovarian cancer. *Gynecol Oncol* 2006;100:506–510.
  32. Birkedal-Hansen B, Moore WG, Taylor RE, Bhowan AS, Birkedal-Hansen H. Monoclonal antibodies to human fibroblast collagenase. Inhibition of enzymatic activity, affinity purification of the enzyme, and evidence for clustering of epitopes in the NH<sub>2</sub>-terminal end of the activated enzyme. *Biochemistry* 1988;27:6751–6758.
  33. Windsor LJ, Birkedal-Hansen H, Birkedal-Hansen B, Engler JA. An internal cysteine plays a role in the maintenance of the latency of human fibroblast collagenase. *Biochemistry* 1991;30:641–647.
  34. Wolff AC, Hammond MB, Schwartz JN, Hagerty KL, Allred DC, Cote RJ, Dowsett M, Fitzgibbons PL, Hanna WM, Langer A, McShane LM, Paik S, Pegram MD, Perez EA, Press MF, Rhodes A, Sturgeon C, Taube SE, Tubbs R, Vance GH, van de Vijver M, Wheeler TM, Hayes DF. American Society of Clinical Oncology/College of American Pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer. *Arch Pathol Lab Med* 2007;131:18–43.
  35. Walch K, Nagele F, Zeillinger R, Vytiska-Binstorfer E, Huber JC, Hefler LA. A polymorphism in the matrix metalloproteinase-1 gene promoter is associated with the presence of polycystic ovary syndrome in Caucasian women. *Fertil Steril* 2005;83:1565–1567.
  36. Zhou P, Du LF, Lv GQ, Yu XM, Gu YL, Li JP, Zhang C. Current evidence on the relationship between four polymorphisms in the matrix metalloproteinases (MMP) gene and breast cancer risk: a meta-analysis. *Breast Cancer Res Treat* 2011;127:813–818.
  37. Johansson N, Ala-aho R, Uitto V, Grenman R, Fusenig NE, Lopez-Otin C, Kahari VM. Expression of collagenase-3 (MMP-13) and collagenase-1 (MMP-1) by transformed keratinocytes is dependent on the activity of p38 mitogen-activated protein kinase. *J Cell Sci* 2000;113(Pt 2):227–235.
  38. Reunanen N, Westermarck J, Hakkinen L, Holmstrom TH, Elo I, Eriksson JE, Kahari VM. Enhancement of fibroblast collagenase (matrix metalloproteinase-1) gene expression by ceramide is mediated by extracellular signal-regulated and stress-activated protein kinase pathways. *J Biol Chem* 1998;273:5137–5145.
  39. Park CH, Lee MJ, Ahn J, Kim S, Kim HH, Kim KH, Eun HC, Chung JH. Heat shock-induced matrix metalloproteinase (MMP)-1 and MMP-3 are mediated through ERK and JNK activation and via an autocrine interleukin-6 loop. *J Invest Dermatol* 2004;123:1012–1019.
  40. Di Girolamo N, Coroneo MT, Wakefield D. UVB-elicited induction of MMP-1 expression in human ocular surface epithelial cells is mediated through the ERK1/2 MAPK-dependent pathway. *Invest Ophthalmol Vis Sci* 2003;44:4705–4714.
  41. Przybylowska K, Kluczna A, Zadrozny M, Krawczyk T, Kulig A, Rykala J, Kolacinska A, Morawiec Z, Drzewoski J, Blasiak J. Polymorphisms of the promoter regions of matrix metalloproteinases genes MMP-1 and MMP-9 in breast cancer. *Breast Cancer Res Treat* 2006;95:65–72.
  42. Slamon DJ, Clark GM, Wong SG, Levin WJ, Ullrich A, McGuire WL. Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science* 1987;235:177–182.
  43. Ross JS, Fletcher JA, Linette GP, Stec J, Clark E, Ayers M, Symmans WF, Puzstai L, Bloom KJ. The Her-2/neu gene and protein in breast cancer 2003: biomarker and target of therapy. *Oncologist* 2003;8:307–325.
  44. Mitchell MS, Press MF. The role of immunohistochemistry and fluorescence in situ hybridization for HER2/neu in assessing the prognosis of breast cancer. *Semin Oncol* 1999;26:108–116.
  45. Chen XS, Ma CD, Wu JY, Yang WT, Lu HF, Wu J, Lu JS, Shao ZM, Shen ZZ, Shen KW. Molecular subtype approximated by quantitative estrogen receptor, progesterone receptor and Her2 can predict the prognosis of breast cancer. *Tumori* 2010;96:103–110.
  46. Choi YH, Ahn JH, Kim SB, Jung KH, Gong GY, Kim MJ, Son BH, Ahn SH, Kim WK. Tissue microarray-based study of patients with lymph node-negative breast cancer shows that HER2/neu overexpression is an important predictive marker of poor prognosis. *Ann Oncol* 2009;20:1337–1343.
  47. Bosc DG, Goueli BS, Janknecht R. HER2/Neu-mediated activation of the ETS transcription factor ER81 and its target gene MMP-1. *Oncogene* 2001;20:6215–6224.
  48. Yang CH, Cristofanilli M. The role of p53 mutations as a prognostic factor and therapeutic target in inflammatory breast cancer. *Future Oncol* 2006;2:247–255.
  49. Colozza M, Cardoso F, Sotiriou C, Larsimont D, Piccart MJ. Bringing molecular prognosis and prediction to the clinic. *Clin Breast Cancer* 2005;6:61–76.
  50. Coradini D, Daidone MG. Biomolecular prognostic factors in breast cancer. *Curr Opin Obstet Gynecol* 2004;16:49–55.
  51. Thor AD, Moore DH II, Edgerton SM, Kawasaki ES, Reihnsaus E, Lynch HT, Marcus JN, Schwartz L, Chen LC, Mayall BH, et al. Accumulation of p53 tumor suppressor gene protein: an independent marker of prognosis in breast cancers. *J Natl Cancer Inst* 1992;84:845–855.
  52. Sun Y, Wenger L, Rutter JL, Brinckerhoff CE, Cheung HS. p53 down-regulates human matrix metalloproteinase-1 (Collagenase-1) gene expression. *J Biol Chem* 1999;274:11535–11540.
  53. Sun Y, Zeng XR, Wenger L, Firestein GS, Cheung HS. P53 down-regulates matrix metalloproteinase-1 by targeting the communications between AP-1 and the basal transcription complex. *J Cell Biochem* 2004;92:258–269.
  54. Yamashita H, Nishio M, Toyama T, Sugiura H, Zhang Z, Kobayashi S, Iwase H. Coexistence of HER2 over-expression and p53 protein accumulation is a strong prognostic molecular marker in breast cancer. *Breast Cancer Res* 2004;6:R24–R30.





## **Congruence between patterns of microRNA expression and histologic grading of invasive breast carcinomas**

Ellsworth DL, Croft DT, Jr, Field LA, Deyarmin B, Kane J, Ellsworth RE, Hooke JA, Shriver CD. Windber Research Institute, Windber, PA; Henry M. Jackson Foundation for the Advancement of Military Medicine, Rockville, MD; Walter Reed Army Medical Center, Washington, DC.

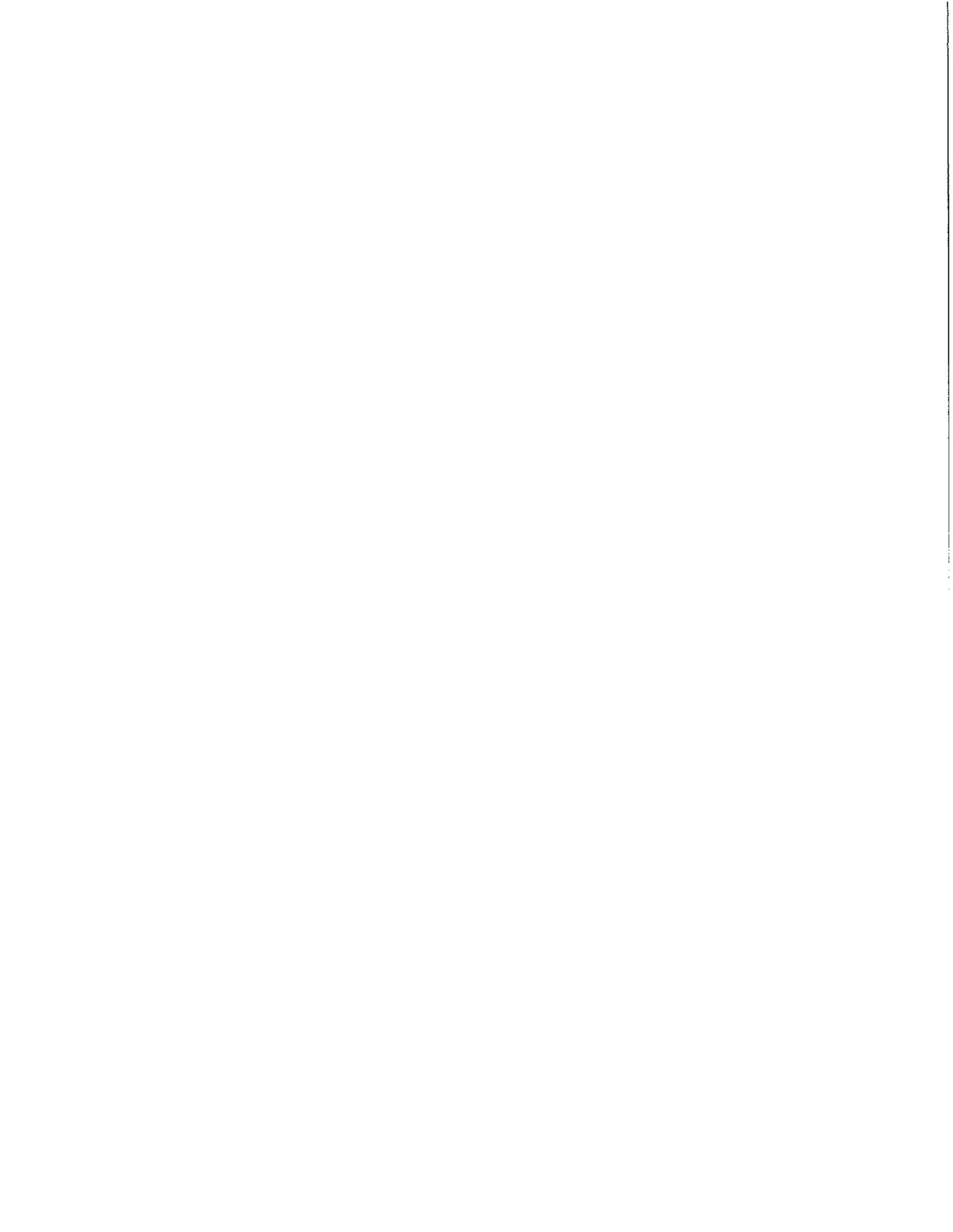
**Background:** Histologic grading may be used as an indicator of prognosis in breast cancer; patients with low-grade carcinomas have ~85% ten-year survival compared to just 45% survival in patients with high-grade disease. Although useful for risk stratification, assigning nuclear grade is subjective, and a large proportion of carcinomas are classified as intermediate-grade with uncertain prognosis, thus limiting clinical utility. MicroRNAs (miRNAs) regulate gene expression and serve an important role in breast cancer development. In this study we examined miRNA expression profiles in low-grade and high-grade breast carcinomas to determine if miRNA expression is associated with pathological classifications of tumor grade.

**Methods:** Breast tumors were obtained from 69 patients enrolled in the Clinical Breast Care Project. Samples were partitioned into low-grade (n=30) or high-grade (n=39) categories using the Nottingham Histologic Score. Following laser microdissection of frozen tissue sections, miRNA was isolated from pure populations of breast tumor cells and hybridized to Affymetrix GeneChip® miRNA arrays containing over 800 human miRNA probes. Expression profiles were analyzed with Partek Genomics Suite using the miRNA Expression Module.

**Results:** We identified 30 unique miRNAs that showed differential expression at a False Discovery Rate (FDR)  $p < 0.05$  between low-grade and high-grade breast carcinomas. Gene targets for these miRNAs function primarily in metabolism and cell communication. Expression of hsa-miR-18a and hsa-miR-572 was significantly different between histologic grades at an FDR  $p < 1 \times 10^{-8}$  and hierarchical clustering based on these miRNAs correctly classified 97% (29/30) of low-grade and 90% (35/39) of high-grade tumors. miR-18a has been shown to inhibit ER signaling and promote cellular differentiation, while the role of miR-572 in breast carcinogenesis is not well known.

**Conclusions:** Dysregulation of miRNAs may accompany changes in cellular morphology typically used in histologic classification of breast carcinomas. Patterns of miRNA expression may improve reproducibility and clinical utility of tumor grading and may prove useful for prediction of recurrence and survival for patients with intermediate-grade carcinomas.

<http://www.call4abstracts.com/sabcs/>



## **Fibroadenomatoid changes are more prevalent in middle-aged women and have a positive association with invasive breast cancer**

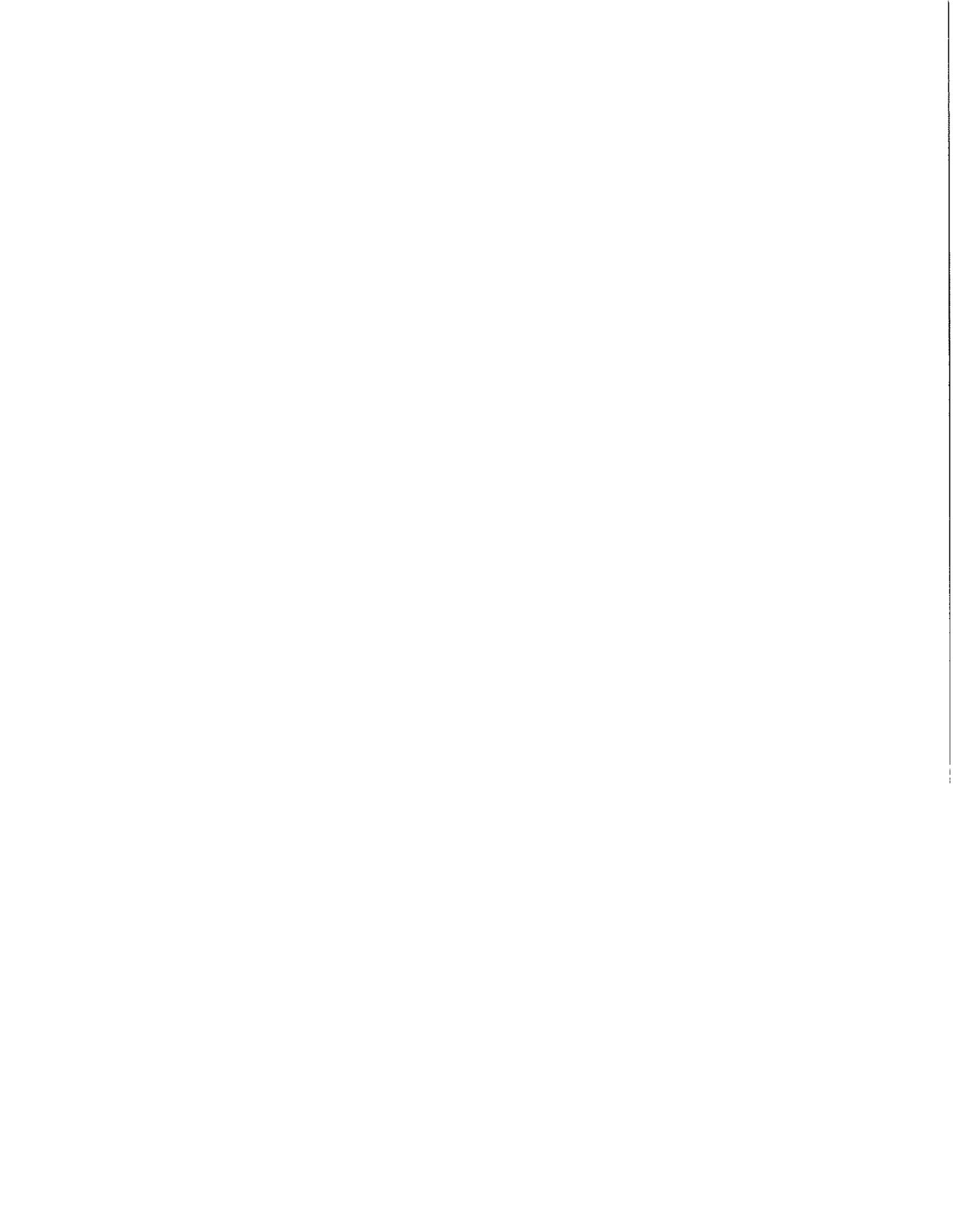
Yaqin Chen, Anthony Bekhash, Albert J. Kovatich, Jeffrey A. Hooke, Leonid Kvecher, Edith P. Mitchell, Hallgeir Rui, Richard J. Mural, Craig D. Shriver, and Hai Hu

**Background:** The role of benign breast diseases (BBDs) in the development of invasive breast cancers (IBCs) has been studied for many years. Some BBDs have been studied comprehensively (e.g., fibrocystic changes (FCC)) while less is known about other BBDs (e.g., fibroadenomatoid changes (FAC)). FAC has been considered by some researchers as a precursor of fibroadenoma (FA). Conclusions from different studies vary, partially due to different interpretation methods and diagnostic criteria when multiple hospitals and pathologists were involved. In this study, we used subjects in the Clinical Breast Care Project (CBCP) from a military medical center where pathology slides were reviewed by a single breast pathologist to study FAC, FA, and FCC in comparison to the published literature.

**Methods:** Subjects were enrolled in the study following IRB-approved, HIPAA-compliant protocols. All the clinicopathologic data are available from the CBCP data warehouse (DW4TR). In the CBCP, FCC is composed of 4 components: stromal fibrosis, cysts, apocrine metaplasia, and sclerosing adenosis. Two modeling studies were performed. i) For the BBDs and IBC association study, two groups of subjects were identified: 1136 subjects diagnosed with “Benign” or “Atypical” diseases, and 619 cases diagnosed with IBCs. A logistic regression model was developed for the prediction of IBCs by the 3 BBDs and 2 well-established risk factors (RF): age (younger,  $\leq 40$ ; middle-aged, 41-60; older,  $> 60$ ) and race (Caucasian, African American, Asian, and other). ii) For the RF association study with the BBDs, 6 additional RFs reported to be associated with these BBDs were identified from the literature: current use of oral contraceptives, number of live births, education, body mass index, hormonal replacement therapy, and IBC family history. These 8 RFs were used to develop a logistic regression model for each of the BBDs. The analyses were performed in SAS.

**Results:** In the first study, age and race were confirmed as RFs for IBCs. FAC was positively associated with IBC (OR=3.04, 95%CI=2.06 to 4.50). FA was negatively associated with IBC, and the level of the association was stronger in women without FCC (OR=0.15, 95%CI=0.08 to 0.28), compared to women with FCC (OR=0.40, 95%CI=0.24 to 0.65). FCC was not significantly associated with IBC. Results from the second study indicated that, age was significantly associated with FAC ( $p=0.015$ ), specifically the middle-aged women were more likely to have FAC compared to younger women (OR=2.03, 95%CI=1.23 to 3.34), while the older women were at a non-significantly increased risk. Trends of association with FAC were also noted for the number of live birth ( $p=0.095$ ), ethnicity ( $p=0.096$ ), and current oral contraceptive pill use ( $p=0.077$ ). The FCC model results were in general consistent with the literature, and we also confirmed that age was negatively associated with the diagnosis of FA.

**Discussion:** Our study was consistent with FCC findings in the literature. We observed that FAC was positively associated with IBC, whereas FA was negatively associated. Also, FAC occurred more often in middle-aged women while FAs occurrence was higher in younger women. Our results suggest that FAC and FA may be two different diseases.



## Fibrocystic changes have different age-dependent patterns in benign, in situ, and invasive breast cancer patients

Anthony Bekhash, Albert J. Kovatich, Yaqin Chen, Jeffrey A. Hooke, Leonid Kvecher, Richard J. Mural, Craig D. Shriver, and Hai Hu

**Background:** Fibrocystic changes (FCCs) have been known to develop in the aging process. Breast cancers in younger and elderly women are known to have different characteristics. Less is known about alterations in the microenvironment for disease development in middle-aged women who undergo menopausal changes. The Walter Reed Army Medical Center, through the Clinical Breast Care Project, has enrolled over 2000 subjects undergoing a biopsy; all the pathology was reviewed by a single pathologist. These subjects provide an opportunity to study the age-dependent pattern of FCC across benign, in situ (IS) and invasive breast carcinoma (IBC) patient populations.

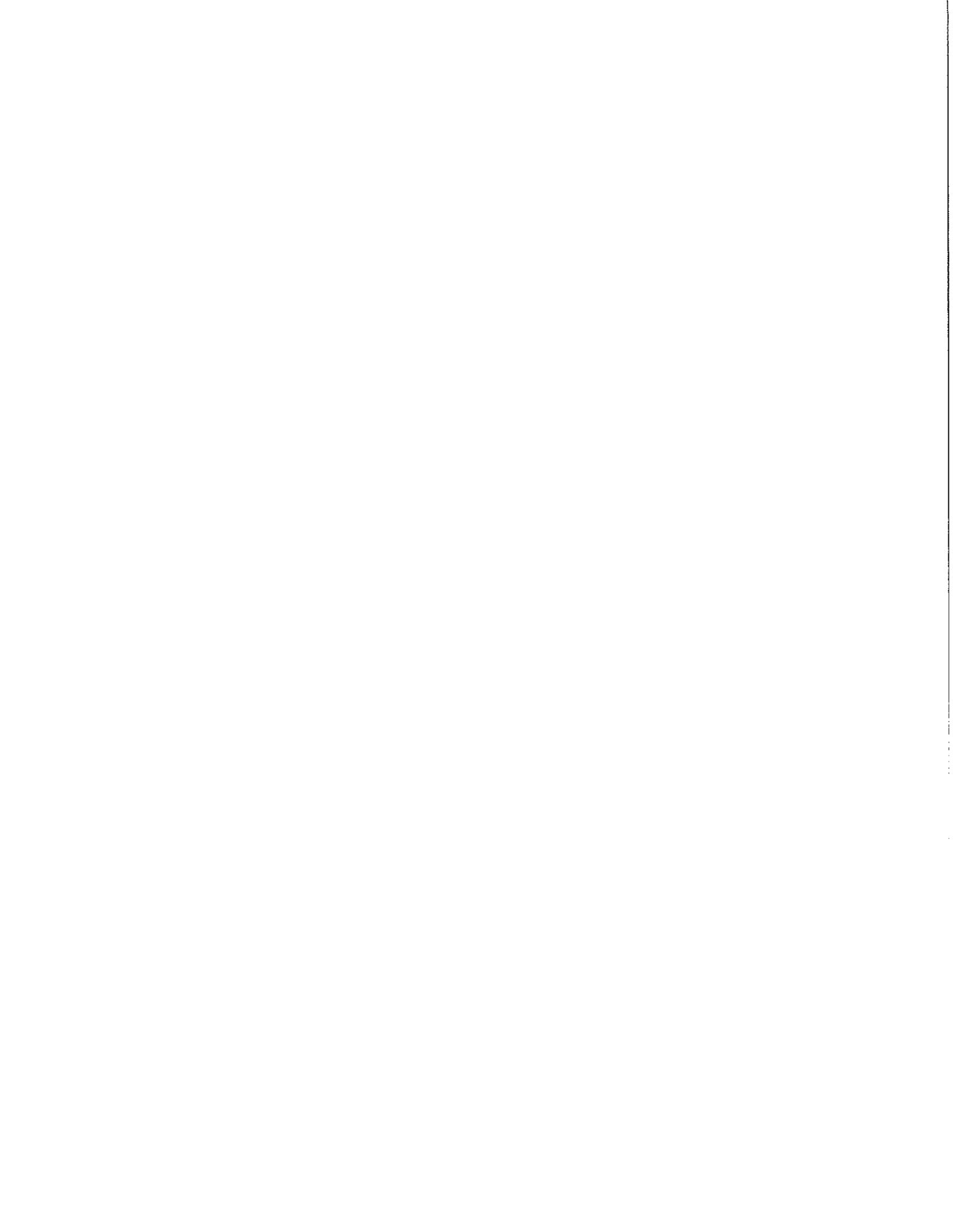
**Methods:** Subjects were enrolled following IRB-approved protocols. A total of 2060 subjects were identified for this study, including 1226 benign/atypical, 193 IS, and 641 IBC patients. Patients were divided into three age groups:  $\leq 45$  years, 46-65 years, and  $\geq 66$  years. Chi-Square test in the SAS was used for statistical analysis across the age groups for the occurrence of FCCs. For IS and IBC patients, additional analysis was performed on high grade (grade 3) and non-high grade (grades 1 and 2 combined) cancers. For IS, grade analysis was performed on ductal carcinoma (DCIS).

**Results:** As shown in Table 1, FCC occurrence rate is significantly lower in younger patients in the benign population. In the IBC population, this rate is significantly higher in middle-aged patients. This higher rate is significant in the non-high grade patients but disappears in the high grade patients. In the IS population, the same trend exists that the FCC rate is higher in middle-aged patients. This trend is retained in non-high grade DCIS but again lost in high grade DCIS.

**Table 1. Occurrence rate of fibrocystic changes in different patient populations across the three age groups**

Pathology/Group	$\leq 45$ y % (n)	46-65 y % (n)	$\geq 66$ y % (n)	p
Benign	39.5%, (676)	53.5%, (413)	52.6%, (137)	<0.0001
Invasive	56.1%, (132)	66%, (309)	55.5%, (200)	0.028
Non-high grade	54%, (76)	73.2%, (194)	61.4%, (145)	0.005
High grade	63.8%, (30)	56.8%, (95)	45.2%, (42)	0.201
In situ	72.7%, (33)	79.2%, (101)	64.4%, (59)	0.122
Non-high grade	66.7%, (15)	82.6%, (46)	66.7%, (36)	0.201
High grade	72.7%, (11)	66.7%, (21)	60%, (15)	0.792

**Discussion:** The observation that FCC increases with age in the benign population, but is bell-shaped in IBC and IS populations, supports a hypothesis that physiological changes in the middle-aged women may alter the microenvironment for disease development. The observation that the bell-shaped pattern is due to non-high grade DCIS but not due to high-grade DCIS suggests that changes in the microenvironment contribute to breast disease development.





**Role for  $\beta$ -catenin/TCF signaling pathways in regulation of migration and growth of triple negative breast cancer cells.**

**Joji Iida<sup>1a</sup>, John R. Lehman<sup>1</sup>, Rebecca Clancy<sup>1</sup>, Stella Somiari<sup>1</sup>, Rachel E. Ellsworth<sup>1</sup>,  
Richard J. Mural<sup>1</sup>, and Craig D. Shriver<sup>2</sup>**

**1. Windber Research Institute, Windber, PA 15963.**

**2. Department of Surgery, Walter-Reed Army Medical Center, Washington, DC20902**

**<sup>a</sup>To whom correspondence should be addressed. E-mail: [g.iida@wriwindber.org](mailto:g.iida@wriwindber.org)**

**Running title: wnt signaling pathway in triple-negative breast cancer**

**Abbreviations used in this study**

Triple negative: TN

Matrix metalloproteinase: MMP

Extracellular matrix: ECM

## **Abstract**

There is substantial evidence indicating the key role of wnt signaling pathways in breast cancer progression and metastasis. Previous studies reported that a synthetic inhibitor for  $\beta$ -catenin/TCF pathway, FH535, inhibited growth of cancer cells but not normal fibroblast cells. In this study, we demonstrated that FH535 significantly inhibited  $\alpha2\beta1/\alpha3\beta1$  integrin-mediated cell migration of triple negative (TN) breast cancer cells (HCC38). Further studies suggest that treatment of cells with FH535 inhibited tyrosine phosphorylation of paxillin at Tyr<sup>118</sup>, but not of focal adhesion kinase (FAK) at Tyr<sup>397</sup>. Interestingly, treatment of HCC38 cells with FH535 diminished expression of c-Abl. Given that c-Abl is one of tyrosine kinases that phosphorylates paxillin, our results suggest that the c-Abl-paxillin axis would play a key role in facilitating tumor migration to type I collagen. We also demonstrated that FH535 inhibited the production of metastatic-related genes including MMP-2, MMP-9, and TrkB. Finally, we demonstrated that FH535 was a potent growth inhibitor for HCC38 cells but not for T47D (luminal B type) when cultured in three dimensional (3D) type I collagen gel. Thus, FH535 inhibited multiple pathways that enhance malignant tumor cell phenotypes, suggesting that targeting  $\beta$ -catenin/TCF signaling pathway would be beneficial strategies for treatment of TN breast cancer phenotype and possibly other cancers in which this signaling pathway is activated.

## **Introduction**

Breast cancer is one of the leading causes of death in women today in Western countries. Although the mortality rate of this cancer has been decreasing as a result of the development of early detection methods such as mammography as well as adjuvant therapy with tamoxifen and chemotherapeutic drugs, metastatic breast cancer continues to be the most fatal condition for women. Triple negative (TN) breast cancers are defined by a lack of expression of estrogen, progesterone, and HER-2/neu receptors. It is widely recognized that TN breast cancers have a poorer prognosis than any other subtypes of breast cancer. Given the lack of effective targeted therapies for TN breast cancer patients, a better understanding of the mechanisms of growth and invasion in the tissue microenvironment will provide valuable insight into developing novel approaches to lower the mortality associated with TN breast cancer.

To successfully metastasize, a cancer cell has to migrate into surrounding tissue architecture, detach from the primary tumor, migrate into surrounding tissue architecture, and intravasate into a blood or lymphatic vessel, arrest in a secondary site, extravasate into the secondary tissues and start growing. Integrins are a family of cell adhesion receptors expressed on the cell surface that mediate cellular interaction with extracellular matrix (ECM) components [1]. There is substantial evidence suggesting a pivotal role for integrin-mediated cell adhesion and subsequent signaling pathways, including tyrosine kinases, for facilitating tumor migration, invasion and growth in the tissue architecture [2]. Thus, further characterizing mechanisms of integrin-mediated cell adhesion and signaling pathways would provide insight for not only understanding the processes of tumor progression but also generating therapeutic strategies for cancer patients.

There are several lines of evidence indicating that inappropriate activation of wnt signaling pathway in cancer cells leads to a translocation of  $\beta$ -catenin to the nucleus to stimulate wnt-pathway responsive genes including matrix degrading enzymes, which are important for tumor invasion and metastasis [3]. FH535 is a synthetic inhibitor of the TCF/ $\beta$ -catenin pathway through its inhibition of the complex formation containing  $\beta$ -catenin and Tcf/Lef in the nucleus, which leads to decreased gene transcription activity for genes such as Tcf-4 [4]. Importantly, FH535 has been shown to inhibit growth of colon cancer cells [4], suggesting that the wnt-signaling pathway would play a key role for facilitating tumor growth and metastasis.

In this study, we used FH535 as a tool for evaluating  $\beta$ -catenin/TCF pathway in promoting growth and migration of TN breast cancer cells. Our results suggest that this pathway is important for promoting growth, migration, and metastasis-related gene expressions (i.e. MMP-2, MMP-9, and TrkB). The  $\beta$ -catenin/TCF signaling pathway plays a pivotal role in promoting TN breast cancer progression and metastasis. This pathway therefore may be a promising therapeutic target for TN phenotypes of breast cancer as well as other breast cancer phenotypes.

## **Materials and Methods**

**Cell lines:** HCC38 and T47D cells were purchased from ATCC (Manassas, VA, USA) and maintained in RPMI1640 containing 10% FBS.

**Antibodies and chemicals:** Poly-HEMA, FH535, and MOPC-21 (IgG<sub>1</sub>) were purchased from Sigma-Aldrich (St. Louis, MO, USA). Rat tail type I collagen, anti- $\beta$ 1 (clone P4C10 IgG<sub>1</sub>), anti- $\alpha$ 2 (clone P1E6 IgG<sub>1</sub>), anti- $\alpha$ 3 (clone P1B5 IgG<sub>1</sub>) integrin, anti-actin, and anti-Ki67 antibodies were purchased from Millipore (Billerica, MA, USA). Anti-phosphopaxillin (Tyr<sup>118</sup>), anti-paxillin, anti-phosphotyrosine, anti-FAK, anti-phosphoFAK (Tyr<sup>397</sup>), and anti-cAbl antibodies were purchased from Cell Signaling Technology (Beverly, MA, USA). HRP-conjugated secondary antibodies were purchased from Jackson Immunoresearch Laboratories (West Grove, PA, USA). WST-1 was purchased from Calbiochem (Gibbstown, NJ, USA). Human MMP MultiAnalyte Profiling Base Kit and Bead Sets Type I collagen solution was purchased from R&D (Minneapolis, MN). Antigen-retrieval solution was purchased from Biogenex (Fremont, CA). Genome U133 Plus 2.0 GeneChip arrays were purchased from Affymetrix (Affymetrix, Santa Clara, CA). Other chemicals were purchased from Sigma-Aldrich.

**Growth Assays:** Cells were harvested with PBS containing 5 mM EDTA and washed three times with RPMI1640-10%FBS. Cells were then resuspended in the same media at a concentration of  $10^5$  cells/ml. For two dimensional growth assays, cells ( $10^4$  cells/100  $\mu$ l/well) were incubated in the presence or absence of FH535 (0.2-20  $\mu$ M) For three dimensional growth assays, 96-well plates were coated with 100  $\mu$ g/ml poly-HEMA for inhibiting attachment of cells as described previously [5,6].

An aliquot of 100  $\mu$ l was added into each well and incubated overnight. Then, an aliquot of 100  $\mu$ l of RPMI-10%FBS containing FH535 at various concentrations (0.2-20  $\mu$ M) was added as described in each well. The plates were incubated overnight and cell growth was measured using colorimetric assay system with WST-1. Each experiment was performed in quadruplets and the data were shown as mean  $\pm$  S.D. of absorbance at 450 nm.

Collagen gel culture: Cell growth was studied in three-dimensional collagen gel as described previously [7]. Briefly, purified type I collagen gels were prepared by adding 10X RPMI, FBS (final concentration of 10%) and neutralized by adding 1N NaOH to give 2.5 mg/ml of type I collagen. Immediately, cells were added at a concentration of  $5 \times 10^5$  cells/ml and the collagen/cell mixture was added to wells, and warmed at 37°C for completing gelling for 4-5 hours. When gels had formed, culture media (RMPI1640 containing 10% FBS) was added and cultured for eight days by changing media every three days. FH535 was added in both gel and medium throughout the experiments at a final concentration of 1  $\mu$ M.

Morphological studies: After eight days of incubation, cell/collagen matrix was fixed by immersion in PBS-10% formalin for 4-5 hours and embedded in paraffin. The paraffin section was serially cut at 5  $\mu$ m and mounted on slide glasses. Every seven sections, cell/collagen matrices were stained with hematoxylin and eosin to ensure the presence of cells in the serial sections. The sections were deparaffinized and treated with antigen-retrieval solution according to the manufacturer's instructions. Tissue sections were stained with various antibodies and visualized by DAB staining with counter staining by DAPI to localize cells. The results were shown by the (mean  $\pm$  standard deviation) of %

of positive cells (DAB-positive) out of the total cells (DAPI-positive) from three independent experiments.

Migration assays: Tumor cell migration was characterized using Transwell systems (Costar, 8.0µm pore size) as described previously [6]. Briefly, cells were harvested with 5 mM EDTA, washed, and resuspended in serum free RPMI1640 media at a concentration of  $5 \times 10^5$  cells/ml. An aliquot of 100 µl of the cell suspension was added into the upper chamber of Transwell system. Type I collagen was dissolved in serum-free RPMI1640 media at a concentration of 3 µg/ml and added into the lower chamber of the system. Antibodies (5µg/ml) or FH535 (0.01 to 1 µM) were added in both upper and lower chambers. Migration assays were performed at 37°C for 4-5 hours. The inserts were removed and then fixed in 10% formalin in PBS for five minutes and stained with 0.5% crystal violet for five minutes. The inserts were washed and the upper surface of the membranes was wiped with a cotton swab to remove non-migratory cells. Migrated cells were counted at randomly selected three fields. Each experiment was performed in triplicate. The results were shown as mean cell numbers +/- S.D. per mm<sup>2</sup>.

Western blotting: Cells were cultured in 6 well plates ( $7 \times 10^5$  cells/well/1ml) coated with 3 µg/ml of type I collagen for 4 hours in the presence or absence of FH535 (1 µM). Plates were washed with PBS and lysed in SDS-sample buffer. After fragmenting DNA by a sonicator, proteins were separated on SDS-PAGE and transferred onto Immobilon-P membrane as we described previously [6]. The membranes were blotted with primary antibodies followed by HRP-conjugated secondary antibodies. The proteins were detected with enhanced chemiluminescence (ECL) reactions. For serial

immunodetection by different antibodies, membranes were stripped and ensured that there was no carry over signals from the last detection by ECL.

Gelatin zymography: The visualizations of MMP-2 and MMP-9 in the serum free conditioned media of HCC38 cells were performed using gelatin zymography as previously described [6]. Briefly, cells ( $7 \times 10^5$  cells per well) were cultured in the growth media in six-well plates overnight. Cells were washed with warmed serum free RPMI1640 five times and then incubated in the presence or absence of FH535 (1  $\mu$ M) in serum free RPMI1640 overnight. Conditioned media (10 $\mu$ l) was mixed with 10 $\mu$ l of SDS sample buffer (6% SDS, 0.16 M Tris, 30%glycerol, and 0.2% bromophenol blue) and then separated on SDS-PAGE containing 1 mg/ml gelatin. The gels were washed two times with 2.5% Triton-X100 and then incubated in buffer containing 20 mM Tris-HCl (pH 7.5), 1.25% Triton-X100, 5 mM CaCl<sub>2</sub>, and 1 mM ZnCl<sub>2</sub> overnight at 37°C. The gels were stained with Coomassie blue and de-stained. As a standard, conditioned media prepared from ConA-stimulated HT1080 cells was used to localize pro- (p), intermediate- (i), and active- (a) forms of MMP-2 as well as pro-MMP-9 [6]. Cells were briefly washed with PBS and total RNA was extracted as described below.

Luminex assays: The amount of MMP-2 and MMP-9 secreted in the different cell culture media samples was measured using the xMAP Technology (Luminex). We used the Fluorokine MAP, Human MMP MultiAnalyte Profiling Base Kit and Bead Sets. In order to measure MMP-2, and MMP-9, serum free conditioned media prepared as described in the previous section were diluted 5-fold with Calibrator Diluent RD5-37. To generate a standard curve, the standard solution provided was reconstituted in 0.9 ml of Calibrator Diluent RD5-37. Fifty micro liters of microparticle bead mixtures for MMP-2 and MMP-9 were pipetted into the 96-well micro plates. Fifty  $\mu$ l of the

prepared standard dilutions were added to the wells and 50  $\mu$ l of the cell culture media to be tested were added to the remaining wells. The plates were then incubated for two hours at room temperature on a horizontal orbital microplate shaker (set at 500 +/- 50 rpm) in the dark. After incubation, the plates were washed three times with washing buffer using the vacuum manifold and 50  $\mu$ l of Biotin Antibody Cocktail added to each well and incubated for one hour and washed. Fifty  $\mu$ l of Streptavidin-phycoerythrin conjugate (PE) was added to all the wells and incubated for 30 minutes. After incubation, 100  $\mu$ l of wash buffer was added to each well and incubated for 2 min on the shaker. The level of the analytes in the samples was detected with the Luminex 100 system. The data were presented as ng/ml.

RT-PCR analysis: Total RNA was prepared from cells treated with FH535 as described in the section of gelatin/zymography using RNA extraction kit (Qiagen, Maryland, USA). The same amount of RNA (1 $\mu$ g) was transcribed into DNA using reverse transcriptase with random hexamer primers according to the manufactures' protocols (SuperScript III First strand synthesis kit, Invitrogen, Carlsbad, CA, USA). The sequences of the primers were as follows.

- 1) MMP-2: forward 5'-CAA TAC CTG AAC ACC TT-3'  
reverse 5'- CTG TAT GTG ATC TGG TT-5'
- 2) MMP-9: forward 5'- AGA TTC CAA ACC TTT GAG-3'  
reverse 5'- GGC CTT GGA AGA TGA ATG-3'
- 3) TCF-4: forward 5'- TTC AAA GAC GAC GGC GAA CAG-3'  
reverse 5'- TTG CTG TAC GTG ATA AGA GGC G-3'
- 4) GAPDH: forward 5'- TGA TGA CAT CAA GAA GGT GGT GAA G-3'  
reverse 5'- TCC TTG GAG GCC ATG TGG GCC A-3'

PCR reactions were performed by heating at 94°C (2min) followed by 30 cycles of 94°C (30 sec), 55°C (30 sec), and 72°C (1min), and a final extension at 72°C 7min) with Platinum Taq DNA polymerase (Invitrogen, Carlsbad, CA, USA). The PCR products were separated on 2% agarose gel and visualized using ethidium bromide.

Statistical analysis: Statistical significance was calculated by Student's two-tailed t-test (paired). *P* values less than 0.05 were considered as statistically significant.

## **Results:**

### **FH535 inhibited growth of human triple negative (TN) breast cancer cells (HCC38)**

FH535 has been demonstrated to inhibit growth of various tumor cells including colon, lung, and hepatocellular carcinomas [4]. As a first set of studies, we tested whether FH535 inhibits growth of triple negative breast cancer cells. We evaluated cell growth with two experimental systems: a conventional two-dimensional (2D) growth assay and a three-dimensional (3D) growth assay in which cells could not attach to the wells but grow in suspension. In both these growth assay systems, FH535 significantly inhibited tumor growth (**Figure 1A and 1B**). These results further support the notion of FH535 as a potent growth inhibitor for cancer cells, implicating that  $\beta$ -catenin/TCF pathway may play a key role in facilitating tumor growth.

### **FH535 inhibited $\alpha 2\beta 1/\alpha 3\beta 1$ integrins-mediated HCC38 migration toward type I collagen**

Enhanced migration has been recognized as one of the hallmarks of malignant tumor cell phenotypes. Migration of breast cancer cells into surrounding connective tissue architecture is the important process for establishing metastasis. Type I collagen is a major component of connective tissues and migration toward this protein is implicated as a key step for invading into tissues.

Previous studies demonstrated that tumor migration to type I collagen is mediated by both or either of  $\alpha 2\beta 1$  and  $\alpha 3\beta 1$  integrin [8,9]. Since HCC38 cells express both  $\alpha 2$  and  $\alpha 3$  integrin subunits evaluated by FACS analysis (not shown), we tested inhibitory antibodies against these subunits for their abilities to inhibit HCC38 cell migration to type I collagen. Both anti- $\alpha 2$  and  $\alpha 3$  integrin subunits antibodies as well as anti- $\beta 1$  integrin subunit antibody almost completely inhibited tumor migration

(**Figure 2A**), while isotype-matched control antibody (MOPC-21) did not affect migration of tumor cells (**Figure 2A**). These results indicate that  $\alpha 2\beta 1$  and  $\alpha 3\beta 1$  integrins mediated HCC38 migration to type I collagen.

Under these experimental conditions, we tested FH535, which is a specific inhibitor for  $\beta$ -catenin/TCF signaling pathway, for its ability to regulate HCC38 migration to type I collagen. In addition to the previous studies demonstrating that FH535 selectively inhibits growth of colon and lung cancer cells [4], our results indicated that this compound was a significant inhibitor of tumor cell migration even at 0.1  $\mu\text{M}$  (**Figure 2B**), consistent with the previous studies [10]. These results suggest that wnt signaling pathways regulate  $\alpha 2\beta 1$  and  $\alpha 3\beta 1$  integrins-mediated tumor migration.

Previous studies demonstrated that tyrosine kinase pathways stimulated through ligation of integrins to their specific ligands play a key role in facilitating cell migration [11,12,13]. In order to test whether wnt signaling pathways were involved in tyrosine kinase pathways in HCC38 cells, cells were treated with FH535 (1  $\mu\text{M}$ ) on type I collagen-coated substrate for 4 hours. Total cell lysates were separated on SDS-PAGE and transferred on Immobilon-P membranes. Blotting with anti-phosphotyrosine (PY) antibody suggest that phosphorylation of several proteins (see arrows in the **Figure 3A**) was inhibited in the presence of FH535 compared to control cells (**Figure 3A**). Previous studies suggest that tyrosine phosphorylation of Tyr<sup>118</sup> of paxillin is implicated to play a role in tumor migration [14]. Blotting with anti-phosphopaxillin (Try<sup>118</sup>) antibody suggests that the phosphorylation was inhibited by treatment with FH535 (**Figure 3B**). However, the amounts of paxillin were not affected by the treatments (**Figure 3B**). As one of the mechanisms, these results

suggest that FH535 inhibited  $\alpha 2\beta 1/\alpha 3\beta 1$ -integrin mediated migration to type I collagen by regulating tyrosine kinase pathways involving phosphorylation of Tyr<sup>118</sup> of paxillin. FH535 did not inhibit cell adhesion to type I collagen (not shown), suggesting that initial attachment of cells mediated by  $\alpha 2\beta 1/\alpha 3\beta 1$ -integrin would not be targets of FH535. Indeed, treatment of cells with FH535 did not alter phosphorylation of Tyr<sup>397</sup> of focal adhesion kinase (FAK) under these experimental conditions (**Figure 3C**). In contrast, treatment of HCC38 cells with FH535 diminished the protein of c-Abl (**Figure 3D**). Previous studies suggest that c-Abl is one of the tyrosine kinases that phosphorylate paxillin [15]. Thus, these results suggest that  $\beta$ -catenin/TCF pathway would regulate the expression or degradation of c-Abl that facilitate tumor migration by phosphorylating paxillin.

#### Down stream target molecules of FH535 in HCC38 cells.

There is substantial evidence demonstrating pro-metastatic gene products including MMP-1, MMP-2, and MMP-9 play key roles in tumor invasion and metastasis such as breast cancer, melanoma, and colon cancer [16,17,18]. In order to ask the question whether  $\beta$ -catenin/TCF signaling pathway was involved in MMP-2 and MMP-9 production from HCC38 cells, cells were cultured in the presence of FH535 in serum-free media overnight and the conditioned media were analyzed for expression of MMP-2 and MMP-9. When cells were treated with FH535, the production of both MMP-2 and MMP-9 was almost completely suppressed as evaluated on gelatin-zymography (**Figure 4A**). We, then, quantified proteins levels of MMP-2 and MMP-9 by utilizing antibody-based assay systems (Luminex) which enable us to quantitate the amount of the proteins with a higher sensitive manner compared to gelatin-zymography. Treatment of HCC 38 cells with FH535, the expression of both MMP-2 (**Figure 4B**) and MMP-9 (**Figure 4C**) were significantly inhibited. These results further

support a key role of  $\beta$ -catenin/TCF pathway in regulating the productions of these pro-metastatic gene products [10].

We then tested whether MMP-2 and MMP-9 were downstream targets of FH535 treatment by reverse transcriptase (RT)-PCR analysis. When cells were treated with FH535 at a concentration of 1  $\mu$ M, the expression of mRNA for MMP-2 and MMP-9 were greatly inhibited (**Figure 4D**), suggesting that FH535 regulates the production of MMP-2 and MMP-9 by inhibiting transcription of these genes. Previous studies showed that treatment of colon cancer cells (HCT116) with FH535 cells inhibited the transcription of TCF-4 [4], and this notion was also partially supported in HCC38 cells by demonstrating a decreased expression of TCF-4 in cells treated with FH535 (**Figure 4D**).

We also demonstrated that TrkB expression was inhibited by treatment of cells with FH535 at a concentration of 1  $\mu$ M (Figure 5). Previous studies suggest that Trk2 plays a key role in regulating epithelial-mesenchymal transition (EMT) in breast cancer. These demonstrated that activated TrkB caused highly invasive and metastatic tumor cells and importantly, this kinase also induce epithelial-mesenchymal transition (EMT) [19]. Taken, together, our results demonstrated that treatment of TN cancer cells with FH535 inhibited gene expression profiles of metastasis-related genes such as MMP-2, MMP-9 and TrkB, which implicate a key role of  $\beta$ -catenin/TCF pathway in facilitating tumor invasion and metastasis by enhancing matrix degradation and regulating cell morphology.

FH535 regulated growth of HCC38 in 3D type I collagen culture.

We further evaluated FH535 in regulating 3D collagen gel culture systems as described in materials and methods. Although the *in vivo* tissue architecture consists of ECM proteins, stromal fibroblasts, and soluble growth factors, recent studies suggest that ECM could approximate the tissues and provide models for various tumor cell growth including breast cancer cells [20,21,22]. As a model system to evaluate breast cancer cells proliferation and invasion, type I collagen has been used as a matrix for supporting tissue architectures [23,24]. We, therefore, tested if FH535 inhibited proliferation of breast cancer cells by evaluating Ki67 expressions (**Figure 6**). When HCC38 cells were cultured for eight days in the presence of FH535 at a concentration of 1  $\mu$ M, cell proliferation was significantly inhibited compared to control cells (**Figure 6A**). Interestingly, proliferation of other phenotype of breast cancer cells, T47D (ER<sup>+</sup>, PR<sup>+</sup>, her2/neu<sup>+</sup>), was not inhibited by FH535 under our experimental conditions (**Figure 6B**), suggesting that  $\beta$ -catenin/TCF pathway may regulate tumor proliferation in a cell-type specific manner.

## **Discussion**

There is increasing evidence demonstrating that wnt signaling pathways play key roles not only in the physiological but also in the disease processes [25,26,27]. In the normal breast, expressions of wnt-signal components are tightly regulated, thus dictating its role in balancing stem cell-renewal, maintenance, and differentiation during embryonic and postnatal development [25]. Thus, it is not surprising that dysregulation of the expression and biological activities of wnt- signaling components would be one of the key steps in the progression of breast cancer. Indeed, activation of wnt signaling is implicated as an important feature of triple-negative (TN)/basal-like breast cancer and is predictive of worse overall survival [28]. In this study, we demonstrated that a  $\beta$ -catenin/TCF inhibitor, FH535, is a potent inhibitor for growth and migration of TN breast cancer cells. Our results not only suggest the importance of  $\beta$ -catenin/TCF signaling in breast cancer progression but also provide additional insights of targeting this signaling pathway for the therapy of breast cancer such as TN phenotype.

Consistent with the previous results reported by Handell et al. [4], our results suggest that FH535 was a potent growth inhibitor for TN phenotype of breast cancer cells when cultured as conventional two-dimensional (2D) conditions. We also demonstrated that tumor growth by incubating cells as a suspension, in which cells do not attach to ECM, was also sensitive to FH535. During metastatic progression, tumor cells are in a state lacking attachment to ECM and susceptible to apoptosis. Upon detachment, normal epithelial cells undergo apoptosis, however, oncogenic transformation of cells acquire properties to become resistance to this process. Given the fact that  $\beta$ -catenin/TCF pathway is activated in many cancer types including breast, colon, and prostate, it is possible to target this pathway as a therapeutic strategies would provide effective treatments.

Enhanced migration to ECM proteins is one of the key steps for establishing metastasis. We demonstrated that FH535 significantly inhibited  $\alpha 2\beta 1/\alpha 3\beta 1$ -integrin mediated tumor migration, suggesting that  $\beta$ -catenin/TCF pathway would regulate integrin-mediated signaling pathways including tyrosine kinase pathways. We demonstrated that FH535 inhibited phosphorylation of Tyr<sup>118</sup> of paxillin without affecting phosphorylation of Tyr<sup>397</sup> of FAK. These results suggest that  $\beta$ -catenin/TCF pathway would regulate downstream signaling of integrin ligation for facilitating tumor migration. Previous studies suggest that phosphorylation of Tyr<sup>38</sup> and Tyr<sup>118</sup> of paxillin occur upon cell adhesion to extracellular matrix (ECM) [29], and FAK [30], Src kinase families [31], and c-Abl [15] mediate these phosphorylations. Lim et. al., recently reported that tyrosine phosphorylation of WW-binding protein 2 (WBP2) plays a key role in mediating breast cancer growth, supporting a notion in which FH535 may have a function to regulate tyrosine kinase pathways [32]. In this study, our results suggest that c-Abl is a downstream target of  $\beta$ -catenin/TCF pathway for promoting tumor migration by phosphorylating paxillin. Recent studies demonstrated that c-Abl plays a key role in regulating growth as well as enhancing drug resistance in breast cancer cells [33,34]. Importantly, small synthetic inhibitor for c-Abl inhibits growth of TN breast cancer cells [35]. Thus, blocking the functions of c-Abl as well as its expression could be a beneficial treatment for TN phenotype of breast cancer.

Production and activation of MMP-2 and MMP-9 have been characterized as key processes in both physiological and pathological processes [36]. In breast cancer progression, there is substantial evidence demonstrating MMP-1, MMP-2 and MMP-9 promotes tumor invasion and metastasis [37,38,39,40]. Recent studies showed that the inhibition of  $\beta$ -catenin by specific siRNA inhibited

MMP-2 and MMP-9 expression in colon cancer cells, inflamed tissue, and mesenchymal stem cells [41,42,43]. Thus, our results further support a notion in which MMP-2 and MMP-9 are downstream target of  $\beta$ -catenin/TCF pathway [10]. By analyzing mRNA expression profiles in control and FH535-treated HCC38 cells, we demonstrated that protein expression of TrkB was inhibited in the presence of FH535 in tumor cells. Although the details of the mechanisms of epithelial-mesenchymal transition (EMT) are not clear yet, previous studies suggest that TrkB kinase plays a key role in the process by regulating downstream target including activation and/or expression of mitogen-activated kinase and transcription factors (i.e. Twist and Snail) [19,44] . We are currently evaluating activation and expression of these downstream molecules for further characterization of mechanisms of EMT in TN breast cancer cells.

Recent studies suggest that three-dimensional (3D) cell culture model emulates a more physiologically relevant microenvironment [21]. For example, down-regulation of  $\beta$ 1 integrin and growth factor signaling pathways in 3D culture systems resulted in reversion of the malignant phenotypes which were demonstrated as a growth arrest and reformation of tissue polarity [45]. Keely and colleagues reported that tumor microenvironment consists of type I collagen matrix regulates breast cancer proliferation and differentiation [24,46,47] . In this study, we demonstrated that inhibition of  $\beta$ -catenin/TCF pathway significantly inhibited HCC38 cells proliferation but not in T47D cells. Although the activation status of  $\beta$ -catenin/TCF pathway in T47D cells is not clear, it is possible that this cell line utilizes different growth mechanisms from HCC38 in type I collagen gel. Indeed, T47D cells grow in the matrix through the activation of GTPase (i.e. Rho) and Rho kinase (ROCK) [24]. Recent studies suggest that wnt signaling pathways are activated in TN phenotype of breast cancer tissues [48,49], implicating specific signaling pathways would regulate cell-type

specific growth in tissue architecture, thus further characterization of tumor growth in tissue microenvironments are required for establishing a better treatment for breast cancer patients.

In this study, we demonstrated the pivotal role for  $\beta$ -catenin/TCF pathway in enhancing tumor migration and growth by using FH535 as a tool. Our results suggest that the potential cross-talk between  $\beta$ -catenin/TCF and tyrosine kinase pathways involving c-Abl-paxillin axis. It is expected that targeting  $\beta$ -catenin/TCF pathway would down-regulate key malignant cellular phenotypes and metastasis-related genes, which would be promising therapeutic strategies for TN phenotype of breast cancer as well as possibly other cancer types which express activated and/or elevated  $\beta$ -catenin/TCF signaling pathway.

**Acknowledgment:**

This research was supported by a grant from the United States Department of Defense (Military Molecular Medicine Initiative MDA W81XWHH-05-2-0075, Protocol 01-20006). The authors thank Dr. Patricia Keely (University of Wisconsin, Madison) for advising and discussing the three dimensional collagen gel assays. We also thank to Brenda Deyarmin and Jennifer Kane for advising histology and immunohistochemistry procedures. The opinion and assertions contained herein are the private views of the authors and are not to be construed as official or as representing the views of the Department of the Army or the Department of Defense.

## References

1. Hynes RO (2009) The extracellular matrix: not just pretty fibrils. *Science* 326: 1216-1219.
2. Cox D, Brennan M, Moran N Integrins as therapeutic targets: lessons and opportunities. *Nat Rev Drug Discov* 9: 804-820.
3. Behrens J (2000) Control of beta-catenin signaling in tumor development. *Ann N Y Acad Sci* 910: 21-33; discussion 33-25.
4. Handeli S, Simon JA (2008) A small-molecule inhibitor of Tcf/beta-catenin signaling down-regulates PPARgamma and PPARdelta activities. *Mol Cancer Ther* 7: 521-529.
5. Drury LJ, Wendt MK, Dwinell MB CXCL12 chemokine expression and secretion regulates colorectal carcinoma cell anoikis through bim-mediated intrinsic apoptosis. *PLoS One* 5: e12895.
6. Iida J, Wilhelmson KL, Price MA, Wilson CM, Pei D, et al. (2004) Membrane type-1 matrix metalloproteinase promotes human melanoma invasion and growth. *J Invest Dermatol* 122: 167-176.
7. Keely PJ, Fong AM, Zutter MM, Santoro SA (1995) Alteration of collagen-dependent adhesion, motility, and morphogenesis by the expression of antisense alpha 2 integrin mRNA in mammary cells. *J Cell Sci* 108 ( Pt 2): 595-607.
8. Carter WG, Wayner EA, Bouchard TS, Kaur P (1990) The role of integrins alpha 2 beta 1 and alpha 3 beta 1 in cell-cell and cell-substrate adhesion of human epidermal cells. *J Cell Biol* 110: 1387-1404.
9. Languino LR, Gehlsen KR, Wayner E, Carter WG, Engvall E, et al. (1989) Endothelial cells use alpha 2 beta 1 integrin as a laminin receptor. *J Cell Biol* 109: 2455-2462.
10. Vaid M, Prasad R, Sun Q, Katiyar SK (2011) Silymarin Targets beta-Catenin Signaling in Blocking Migration/Invasion of Human Melanoma Cells. *PLoS One* 6: e23000.

11. Cunningham-Edmondson AC, Hanks SK (2009) p130Cas substrate domain signaling promotes migration, invasion, and survival of estrogen receptor-negative breast cancer cells. *Breast Cancer (London)* 2009: 39-52.
12. Tavora B, Batista S, Reynolds LE, Jadeja S, Robinson S, et al. Endothelial FAK is required for tumour angiogenesis. *EMBO Mol Med* 2: 516-528.
13. Zhao X, Guan JL Focal adhesion kinase and its signaling pathways in cell migration and angiogenesis. *Adv Drug Deliv Rev*.
14. Azuma K, Tanaka M, Uekita T, Inoue S, Yokota J, et al. (2005) Tyrosine phosphorylation of paxillin affects the metastatic potential of human osteosarcoma. *Oncogene* 24: 4754-4764.
15. Lewis JM, Schwartz MA (1998) Integrins regulate the association and phosphorylation of paxillin by c-Abl. *J Biol Chem* 273: 14225-14230.
16. Bourboulia D, Stetler-Stevenson WG Matrix metalloproteinases (MMPs) and tissue inhibitors of metalloproteinases (TIMPs): Positive and negative regulators in tumor cell adhesion. *Semin Cancer Biol* 20: 161-168.
17. Gialeli C, Theocharis AD, Karamanos NK Roles of matrix metalloproteinases in cancer progression and their pharmacological targeting. *Febs J* 278: 16-27.
18. Radisky ES, Radisky DC Matrix metalloproteinase-induced epithelial-mesenchymal transition in breast cancer. *J Mammary Gland Biol Neoplasia* 15: 201-212.
19. Smit MA, Geiger TR, Song JY, Gitelman I, Peeper DS (2009) A Twist-Snail axis critical for TrkB-induced epithelial-mesenchymal transition-like transformation, anoikis resistance, and metastasis. *Mol Cell Biol* 29: 3722-3737.
20. Bissell MJ (1999) Tumor plasticity allows vasculogenic mimicry, a novel form of angiogenic switch. A rose by any other name? *Am J Pathol* 155: 675-679.

21. Bissell MJ, Labarge MA (2005) Context, tissue plasticity, and cancer: are tumor stem cells also regulated by the microenvironment? *Cancer Cell* 7: 17-23.
22. Bissell MJ, Radisky D (2001) Putting tumours in context. *Nat Rev Cancer* 1: 46-54.
23. Provenzano PP, Inman DR, Eliceiri KW, Knittel JG, Yan L, et al. (2008) Collagen density promotes mammary tumor initiation and progression. *BMC Med* 6: 11.
24. Wozniak MA, Desai R, Solski PA, Der CJ, Keely PJ (2003) ROCK-generated contractility regulates breast epithelial cell differentiation in response to the physical properties of a three-dimensional collagen matrix. *J Cell Biol* 163: 583-595.
25. Micalizzi DS, Farabaugh SM, Ford HL Epithelial-mesenchymal transition in cancer: parallels between normal development and tumor progression. *J Mammary Gland Biol Neoplasia* 15: 117-134.
26. Niehrs C, Boutros M Trafficking, acidification, and growth factor signaling. *Sci Signal* 3: pe26.
27. Tao J, Chen S, Lee B Alteration of Notch signaling in skeletal development and disease. *Ann N Y Acad Sci* 1192: 257-268.
28. Geyer FC, Lacroix-Triki M, Savage K, Arnedos M, Lambros MB, et al. beta-Catenin pathway activation in breast cancer is associated with triple-negative phenotype but not with CTNNB1 mutation. *Mod Pathol*.
29. Burridge K, Turner CE, Romer LH (1992) Tyrosine phosphorylation of paxillin and pp125FAK accompanies cell adhesion to extracellular matrix: a role in cytoskeletal assembly. *J Cell Biol* 119: 893-903.
30. Schaller MD, Parsons JT (1995) pp125FAK-dependent tyrosine phosphorylation of paxillin creates a high-affinity binding site for Crk. *Mol Cell Biol* 15: 2635-2645.
31. Klinghoffer RA, Sachsenmaier C, Cooper JA, Soriano P (1999) Src family kinases are required for integrin but not PDGFR signal transduction. *Embo J* 18: 2459-2471.

32. Lim SK, Orhant-Prioux M, Toy W, Tan KY, Lim YP (2011) Tyrosine phosphorylation of transcriptional coactivator WW-domain binding protein 2 regulates estrogen receptor {alpha} function in breast cancer via the Wnt pathway. *Faseb J* 25: 3004-3018.
33. Srinivasan D, Plattner R (2006) Activation of Abl tyrosine kinases promotes invasion of aggressive breast cancer cells. *Cancer Res* 66: 5648-5655.
34. Zhao H, Ou-Yang F, Chen IF, Hou MF, Yuan SS, et al. (2010) Enhanced resistance to tamoxifen by the c-ABL proto-oncogene in breast cancer. *Neoplasia* 12: 214-223.
35. Finn RS, Dering J, Ginther C, Wilson CA, Glaspy P, et al. (2007) Dasatinib, an orally active small molecule inhibitor of both the src and abl kinases, selectively inhibits growth of basal-type/"triple-negative" breast cancer cell lines growing in vitro. *Breast Cancer Res Treat* 105: 319-326.
36. Kessenbrock K, Plaks V, Werb Z Matrix metalloproteinases: regulators of the tumor microenvironment. *Cell* 141: 52-67.
37. Bohn OL, Nasir I, Brufsky A, Tseng GC, Bhargava R, et al. (2009) Biomarker profile in breast carcinomas presenting with bone metastasis. *Int J Clin Exp Pathol* 3: 139-146.
38. Del Casar JM, Gonzalez-Reyes S, Gonzalez LO, Gonzalez JM, Junquera S, et al. Expression of metalloproteases and their inhibitors in different histological types of breast cancer. *J Cancer Res Clin Oncol* 136: 811-819.
39. Garcia MF, Gonzalez-Reyes S, Gonzalez LO, Junquera S, Berdize N, et al. Comparative study of the expression of metalloproteases and their inhibitors in different localizations within primary tumours and in metastatic lymph nodes of breast cancer. *Int J Exp Pathol* 91: 324-334.
40. Gonzalez LO, Gonzalez-Reyes S, Junquera S, Marin L, Gonzalez L, et al. Expression of metalloproteases and their inhibitors by tumor and stromal cells in ductal carcinoma in situ of the breast and their relationship with microinvasive events. *J Cancer Res Clin Oncol* 136: 1313-1321.

41. Karow M, Popp T, Egea V, Ries C, Jochum M, et al. (2009) Wnt signalling in mouse mesenchymal stem cells: impact on proliferation, invasion and MMP expression. *J Cell Mol Med* 13: 2506-2520.
42. Kim TH, Kim SH, Seo JY, Chung H, Kwak HJ, et al. Blockade of the Wnt/beta-catenin pathway attenuates bleomycin-induced pulmonary fibrosis. *Tohoku J Exp Med* 223: 45-54.
43. Wang XH, Sun X, Meng XW, Lv ZW, Du YJ, et al. beta-catenin siRNA regulation of apoptosis- and angiogenesis-related gene expression in hepatocellular carcinoma cells: potential uses for gene therapy. *Oncol Rep* 24: 1093-1099.
44. Smit MA, Peeper DS Zeb1 is required for TrkB-induced epithelial-mesenchymal transition, anoikis resistance and metastasis. *Oncogene*.
45. Park CC, Zhang H, Pallavicini M, Gray JW, Baehner F, et al. (2006) Beta1 integrin inhibitory antibody induces apoptosis of breast cancer cells, inhibits growth, and distinguishes malignant from normal phenotype in three dimensional cultures and in vivo. *Cancer Res* 66: 1526-1535.
46. Kenny PA, Lee GY, Bissell MJ (2007) Targeting the tumor microenvironment. *Front Biosci* 12: 3468-3474.
47. Provenzano PP, Eliceiri KW, Keely PJ (2009) Shining new light on 3D cell motility and the metastatic process. *Trends Cell Biol* 19: 638-648.
48. King TD, Suto MJ, Li Y (2011) The Wnt/beta-catenin signaling pathway: A potential therapeutic target in the treatment of triple negative breast cancer. *J Cell Biochem*.
49. Geyer FC, Lacroix-Triki M, Savage K, Arnedos M, Lambros MB, et al. (2011) beta-Catenin pathway activation in breast cancer is associated with triple-negative phenotype but not with CTNNB1 mutation. *Mod Pathol* 24: 209-231.

**Figure legends:**

**Figure 1: Effect of FH535 on the growth of triple-negative (TN) breast cancer cells (HCC38).**

(A) Cells were harvested and resuspended in RPMI1640-10%FBS at a concentration of  $10^5$  cells/ml. An aliquot (100  $\mu$ l) of cell suspension was incubated with various antibodies at a concentration of 1FH535 for 24 hours in 96 well plates. Cell growth was measured as an Absorbance at 450 nm of WST-1 as described in manufacture's instruction and expressed as mean  $\pm$  S.D. Experiments were repeated three times. \*  $p < 0.001$ , \*\*  $p < 0.05$  (by Student's two-tailed paired  $t$ -test).

(B) Cells were prepared as described above and incubated in the presence of FH535 (0.01-1  $\mu$ M) on plate coated with poly-HEMA as described in materials and methods. Cell growth was measured as an Absorbance at 450 nm of WST-1 as described in manufacture's instruction and expressed as mean  $\pm$  S.D. Experiments were repeated three times.

\*  $p < 0.001$ , \*\*  $p < 0.05$  (by Student's two-tailed paired  $t$ -test).

**Figure 2: FH535 inhibited migration of HCC38 cells to type I collagen.**

(A) Cells were harvested, washed, and resuspended in RPMI-serum free media at a concentration of  $5 \times 10^5$  cells/ml. Type I collagen was used as a chemoattractant at a concentration of 3  $\mu$ g/ml. Antibodies (5 $\mu$ g/ml) were added in both cell suspension and type I collagen solution and incubated for 4 hours at 37°C. Migrated cells were manually counted and expressed as mean  $\pm$  S.D. Experiments were repeated three times. \*  $p < 0.001$ , \*\*  $p < 0.05$  (by Student's two-tailed paired  $t$ -test)

(B) FH535 (0.01-1 $\mu$ M) were added in both cell suspension and type I collagen solution and incubated for 4 hours at 37°C. Migrated cells were manually counted and expressed as mean +/- S.D. Experiments were repeated three times.

\*  $p < 0.001$ , \*\*  $p < 0.05$  (by Student's two-tailed paired  $t$ -test)

Figure 3: Effect of FH535 on the tyrosine phosphorylation pathways in HCC38 cells on type I collagen.

Cells were harvested and cultured on substrate coated with type I collagen for 4 hours in the presence or absence of FH535 (1  $\mu$ M). Cells were lysed and the same amount of proteins was separated on SDS-PAGE followed by transferred onto Immobilon-P membranes. Membranes were blotted with anti-phosphotyrosine antibody (PY) (A), anti-phosphopaxillin (Tyr<sup>118</sup>) followed by anti-paxillin antibody (B), with anti-phosphoFAK (Tyr<sup>397</sup>) followed by anti-FAK antibody (C), or anti-c-Abl antibody (D). Actin was used as a loading control. Membranes were incubated with HRP-conjugated secondary antibody and signals were detected by ECL.

Figure 4: FH535 inhibited production of MMP-2 and MMP-9 from HCC38 cells.

(A) Cells ( $7 \times 10^5$  cells/well) were cultured overnight in RPMI-1640-10%FBS, washed, and replaced with serum-free RPMI1640 media in the presence or absence of FH535 (1  $\mu$ M) overnight. The conditioned media were analyzed on gelatin/zymography. The localization of pro-MMP-9, pro (p), intermediate (i), and active (a) forms of MMP-2 were shown in the left column using serum free conditioned media prepared from ConA-stimulated HT1080 cells. Conditioned media were analyzed for the expression of MMP-2 (B) and MMP-9 (C) by Luminex . \*  $p < 0.001$  (by Student's two-tailed paired  $t$ -test)

(D) Total RNA was extracted and the same amount of RNA was transcribed to DNA using random hexamer. The DNA was used as a template for PCR analysis for MMP-2, MMP-9, and Tcf-4. GAPDH was used as a control.

Figure 5: Effect of FH535 on the expression of TrkB protein

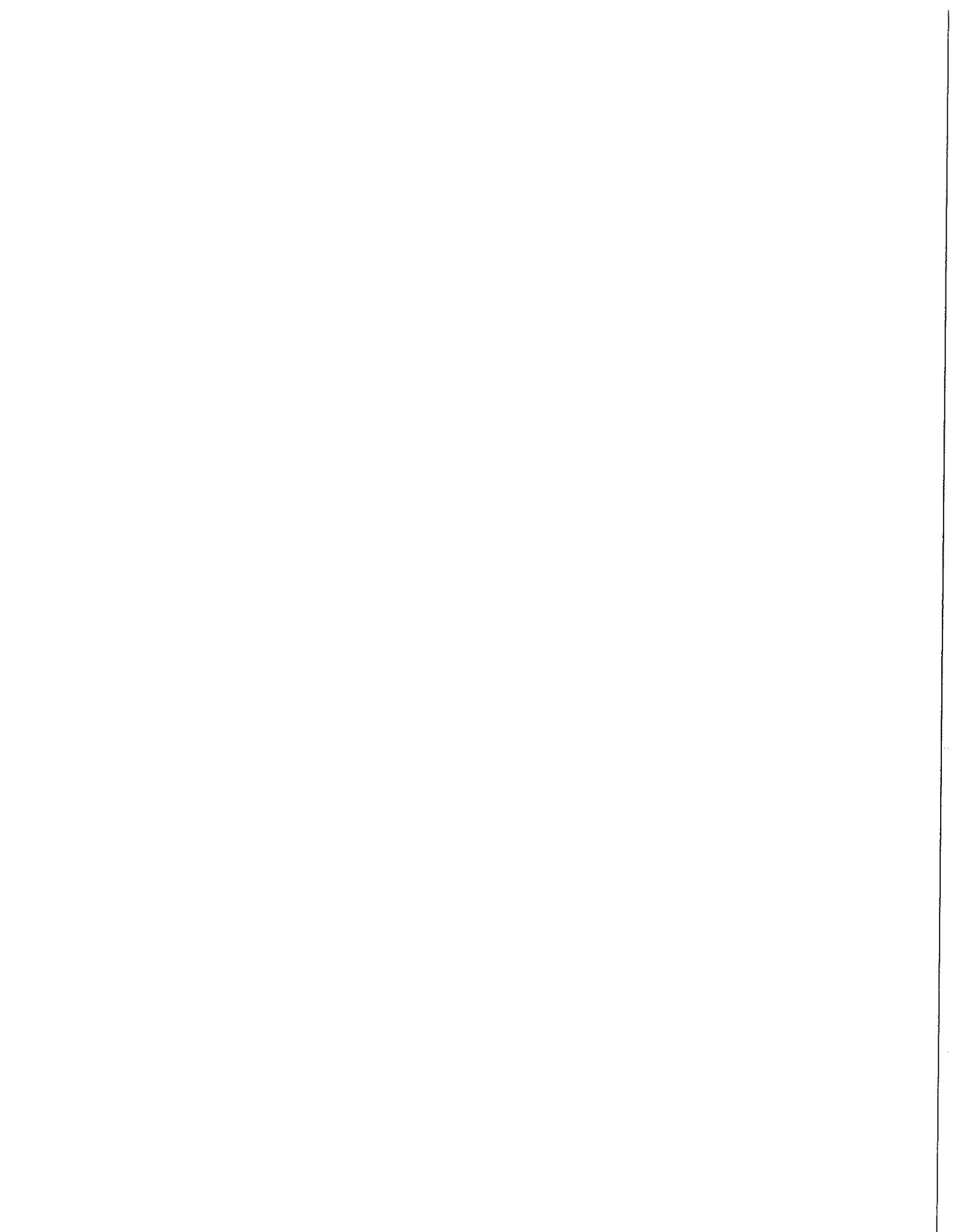
Cells were harvested and cultured on substrate coated with type I collagen for 4 hours in the presence or absence of FH535 (1  $\mu$ M). Cells were lysed and the same amount of proteins was separated on SDS-PAGE followed by transferred onto Immobilon-P membranes. Membranes were blotted with anti-TrkB antibody. Membranes were re-stripped and stained with anti-actin antibody as a loading control.

Figure 6: Proliferation of tumor cells in three-dimensional type I collagen gel in the presence of FH535.

Cells (HCC38 and T47D) were cultured in type I collagen gel as described in material and methods. Cell/gel matrices were fixed and embedded in paraffin. The paraffin section was serially cut at 5  $\mu$ m and mounted on slide glasses. Tissue sections were stained with anti-Ki67 antibody or control IgG followed by HRP-conjugated secondary antibody as visualizing by DAB staining with counter staining by DAPI to localize cells. The results were demonstrated by the (mean  $\pm$  standard deviation) of % of positive cells for DAB from three independent experiments.

\*  $p < 0.001$  (by Student's two-tailed paired  $t$ -test).

n.s: not significance (by Student's two-tailed paired  $t$ -test).



## Does Breast Tumor Heterogeneity Necessitate Further Immunohistochemical Staining on Surgical Specimens?

**Authors:** Lauren T. Greer, M.D., Martin Rosman, M.D., Charles Mylander, Ph.D., Albert J. Kovatich, M.S., Jeffrey Hooke, M.D., Wen Liang, D.O., Robert R. Buras, M.D., FACS, Craig D. Shriver, M.D., FACS, Lorraine Tafra, M.D., FACS

**Introduction:** Prognostic and predictive tumor markers (ER, PR, HER2, Ki-67) in breast cancer are most commonly performed on core needle biopsies (CNB) of the primary tumor. Because treatment recommendations are influenced by these markers, it is imperative to verify strong concordance between tumor markers derived from IHC testing on CNB specimens and the corresponding surgical specimens (SS).

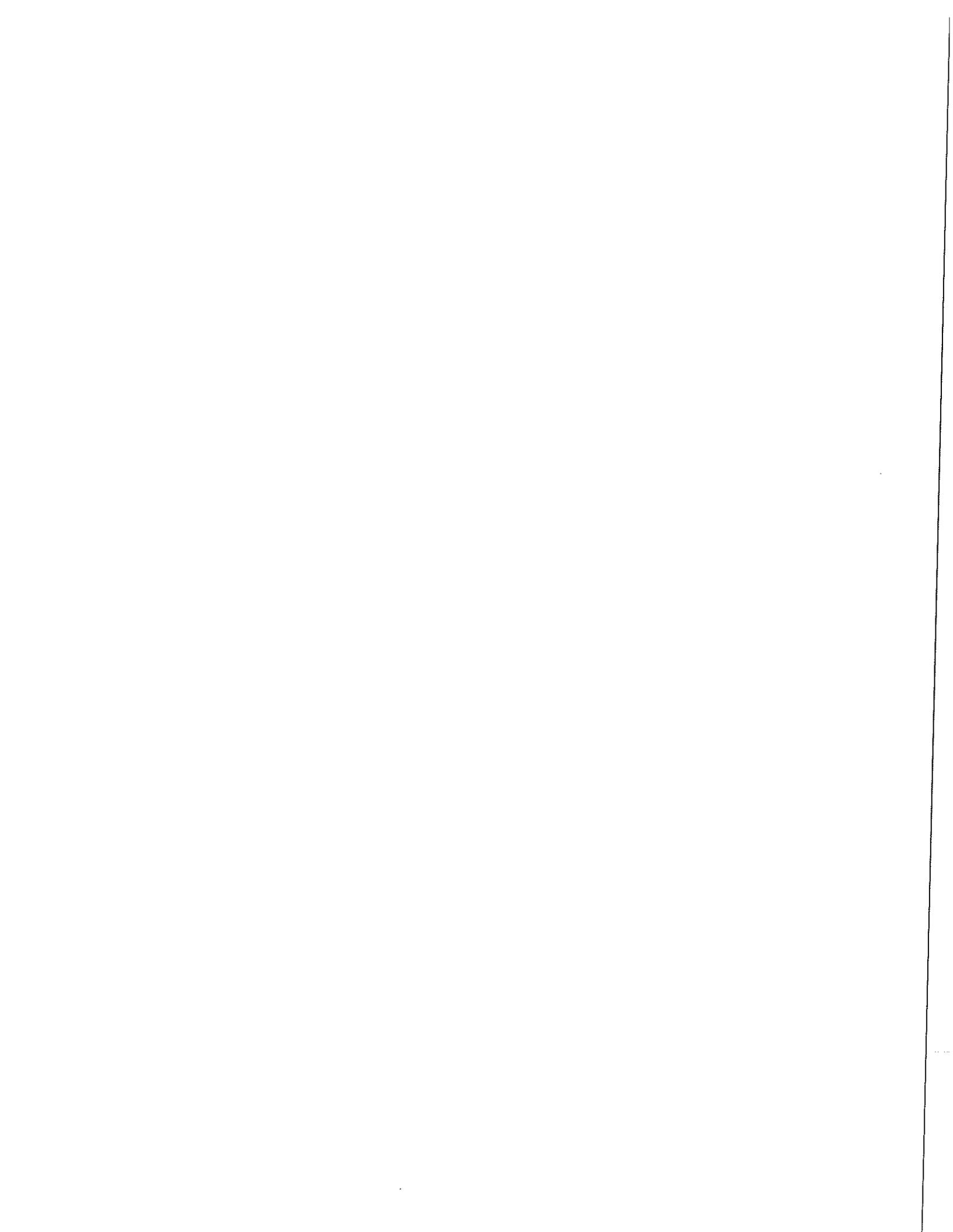
**Methods:** A prospective study was performed on 132 women (151 samples) with breast cancer diagnosed from January 2009 to March 2011. Tumor type, grade, ER, PR, HER2, Ki67 expression by IHC were retrospectively analyzed in the CNB and SS. Contingency tables and agreement modeling were performed to generate Kappa values. ER% and PR% were calculated based on positivity being  $\geq 1\%$ , HER2 was considered positive if CNB showed 2+(FISH positive) or 3+ but on SS was based on 3+ only (no FISH performed on SS).

**Results:** There was substantial agreement between the CNB and SS for tumor type, ER%, PR%, HER2; moderate agreement for grade; and fair agreement for Ki67% (see Chart). The CNB contained higher percentages of ER, PR and Ki67 than the SS. Distinct molecular subtypes (tumor heterogeneity) by H&E and IHC were noted in the SS of 14 subjects (11%). These patients had multiple IHC testing performed on different areas of tumor. In 3 of these subjects (21%), HER2 was found to be positive in an area of the SS but was negative on the CNB. No cases were seen where ER or PR were negative on CNB but became positive in the SS. However, 2 subjects in the heterogeneous group had distinctly different regions in their tumor where one was ER positive and another was ER negative.

**Conclusion:** The heterogeneous distribution of antigens in breast cancer tumors raises concern that the CNB may not adequately represent the true biologic profile in all patients. There is strong concordance for tumor type, ER, and PR between CNB and SS (although a decline was noted from CNB to SS); however, HER2 activity does not appear to be adequately detected on CNB in patients with heterogeneous tumors. This data highlights that the IHC testing on the CNB alone may not be adequate to tailor targeted therapy in all patients.

Blomarker	# Agreements	Kappa ( $\kappa$ )	Landis-Koch Agreement Grade	95% CI
Tumor Type	124/151 (82%)	0.66	Substantial	(0.55, 0.77)
Histologic Grade	84/133 (63%)	0.44	Moderate	(0.31, 0.56)
ER%	142/151 (94%)	0.73	Substantial	(0.56, 0.89)
PR%	134/151 (89%)	0.69	Substantial	(0.55, 0.82)
HER2	128/134 (96%)	0.68	Substantial	(0.43, 0.92)
Ki67 %	52/119 (44%)	0.21	Fair	(0.10, 0.31)

"The views expressed in this abstract are those of the authors and do not reflect the official policy of the Department of the Army (DOA), Department of Defense (DOD), or US Government."



## Molecular drivers of breast tumor differentiation

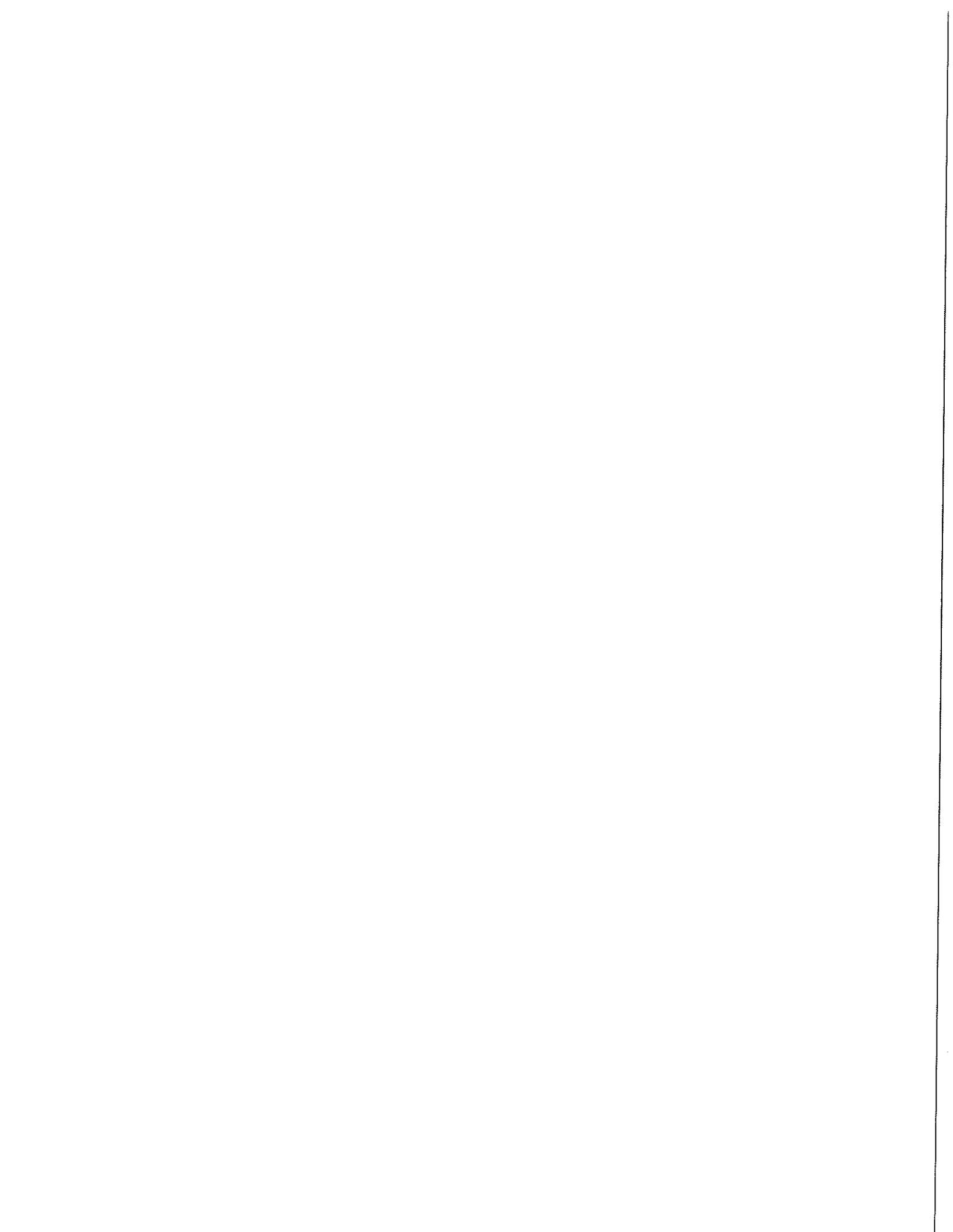
HL Patney, RE Ellsworth, CD Shriver

**Background:** Histological grade classifies breast carcinomas into low- (G1), intermediate- and high-grade (G3). Developmental pathways for high-grade disease remain controversial with both linear models, in which high-grade tumors evolve from low-grade precursors, as well as models that depict G1 and G3 carcinomas as distinct molecular diseases, based largely on a higher frequency of loss of chromosome 16q in low-grade tumors. Here, we utilized global SNP arrays to simultaneously measure copy number changes and LOH at high resolution to determine the evolutionary relationship between low- and high-grade tumors.

**Methods:** The Clinical Breast Care Project database was queried to identify all G1 and G3 tumors with frozen tumor specimens available. Tumor specimens (49 G1 and 77 G3) were subjected to laser microdissection and hybridized to the GeneChip Human Mapping 250K Sty arrays (Affymetrix). Copy number and LOH analysis were performed using Partek Genomics Suite segmentation analysis.

**Results:** The most common alteration in both groups was gain of chromosome 1q (71% G1 and 69% G3), with G1 tumors also had frequent loss of 16q (90%) and G3 tumors having frequent gain of 8q and loss of 8p, 15q and 17p. Loss of chromosome 16q and gain of chromosome 8q were significantly higher ( $P < 0.00001$ ) in G1 and G3 tumors, respectively. The predominant type of alteration of chromosome 16q in G1 tumors (65%) was loss of the entire arm; in G3 tumors, 12% had complete loss, 10% had no alteration and the other samples had complex patterns of small regions (<Mb) or gain or loss. Patterns of alteration on chromosome 16q included while in G3 tumors, nine (12%) had complete loss of 16q, and eight (10%) had no alterations; the majority, however, had complex patterns of change including small (<1 Mb) deletions or amplifications scattered across 16q and complex sawtooth patterns of alternating deletions and amplifications. LOH was not found in diploid or amplified regions.

**Conclusions:** Gain of chromosome 1q is common in low- and high-grade tumors and may be an early event in tumorigenesis. G3 tumors did not demonstrate frequent loss of chromosome 16q, nor was frequent LOH detected in copy-neutral or amplified regions thus suggesting that the complex patterns of gain and loss of chromosome 16q in G3 tumors are not the result of segmental gains subsequent to large-scale loss of 16q. Together, these data suggest that G3 tumors do not evolve from a well-differentiated precursor, but rather the development of G1 tumors is driven by loss of 16q, while G3 tumors are driven by more complex changes including gain of 8q and loss of 8p, 15q and 17p.



## **Identification of blood-based biomarkers for breast cancer detection**

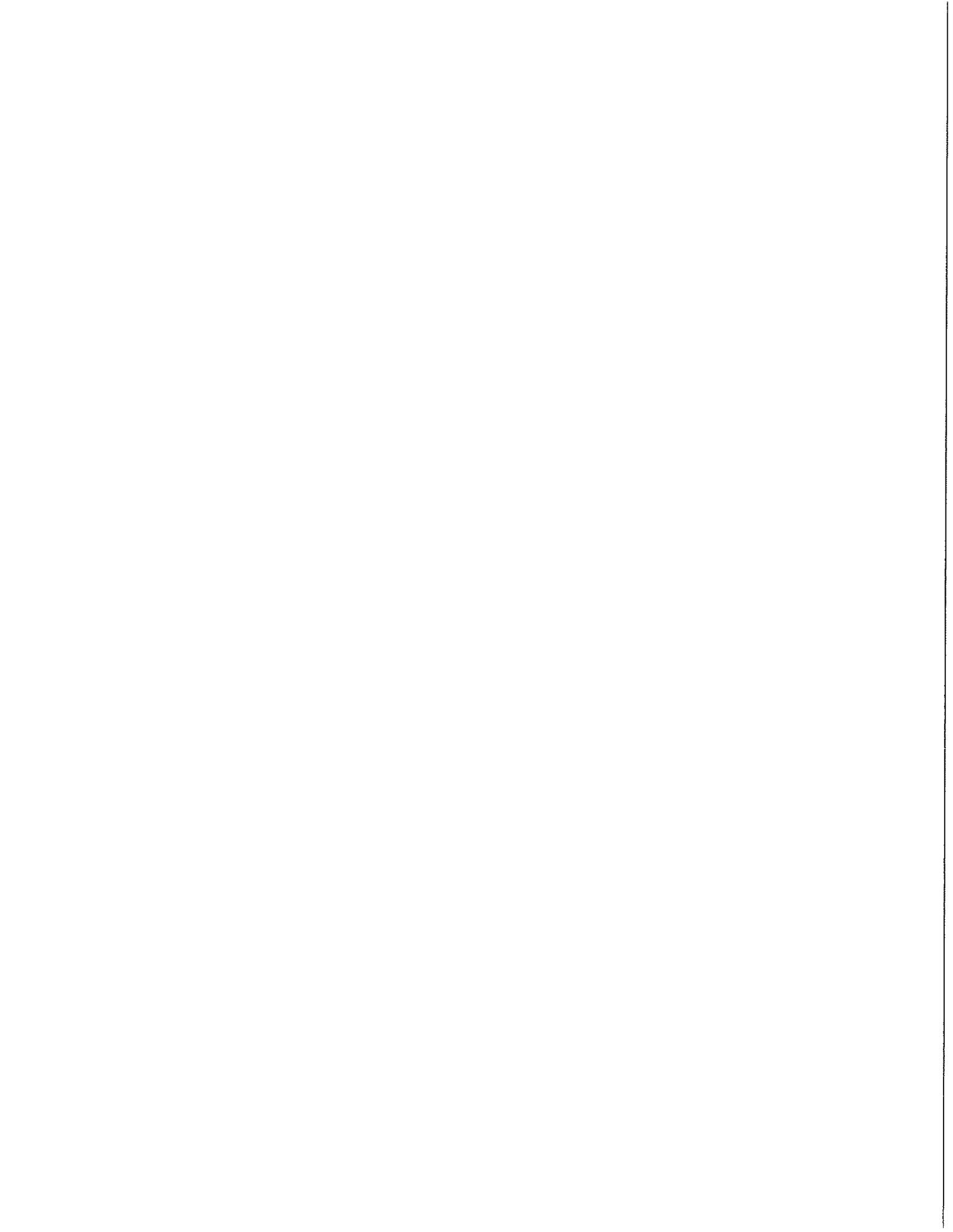
Lori A. Field, Brenda Deyarmin, Rachel E. Ellsworth, and Craig D. Shriver.

**Background:** Current methods for determining whether or not a patient has breast cancer can be costly, time-consuming and require invasive surgical procedures. In addition to being more cost-effective and time-efficient, the development of a reliable blood test that can distinguish between women with and without breast cancer would also ease the anxiety and stress associated with more invasive procedures. Through comparison of whole-blood gene expression profiles, we have identified differences in gene expression between patients with and without breast cancer that may be useful as a potential screening tool for breast cancer.

**Methods:** Whole blood from 50 female controls and 63 women with invasive breast cancer was collected into PAXgene tubes. RNA was isolated using the PAXgene Blood RNA kit (PreAnalytiX) and globin mRNA levels reduced using the GLOBINclear kit (Applied Biosystems). Globin-reduced RNA was amplified with the MessageAmp II aRNA Amplification kit and labeled with biotin-11-UTP (Applied Biosystems). aRNA was fragmented and hybridized to the Affymetrix HG U133A 2.0 expression array. CEL files were analyzed using Partek Genomics Suite and genes differentially expressed between controls and invasive cases were identified using a FDR < 0.05 and fold change greater than 2.

**Results:** Comparison of blood-based gene expression patterns between controls and invasive breast cancer patients identified 55 probes representing 51 genes that were differentially expressed. Twenty genes were more highly expressed in controls while 31 genes were more highly expressed in patients with invasive breast cancer. Hierarchical cluster analysis produced two groups; all of the controls clustered together (100% specificity) while 11 of 63 invasive cases (83% sensitivity) were incorrectly classified. A sub-analysis comparing invasive patients with (n=29) and without (n=34) lymph node metastases could not identify a gene expression signature that could discriminate between these two groups of patients.

**Discussion:** These data suggest that gene expression profiles from whole blood can be used to differentiate between patients with and without breast cancer. Those genes with higher expression in controls function in cytoskeleton organization and assembly, immune response, signal transduction, and protein trafficking. Functions of genes with higher expression in patients with invasive breast cancer include signal transduction, transcription, translation, and ribosomal assembly, regulation of cell growth and DNA repair. Further investigation of these gene expression differences is warranted to determine their utility as a potential blood-screening tool for the detection of breast cancer.



# Loss of Stat5a protein in breast cancer is associated with tumor progression and unfavorable clinical outcomes

Amy R. Peck<sup>1</sup>, Agnieszka K. Witkiewicz<sup>2</sup>, Chengbao Liu<sup>1</sup>, Alexander C. Klimowicz<sup>3</sup>, Ginger A. Stringer<sup>4</sup>, Edward Pequignot<sup>4</sup>, Boris Freydin<sup>4</sup>, Ning Yang<sup>1</sup>, Thai H. Tran<sup>1</sup>, Anne L. Rosenberg<sup>5</sup>, Jeffrey A. Hooke<sup>6</sup>, Albert J. Kovatich<sup>7</sup>, Craig D. Shriver<sup>6</sup>, David L. Rimm<sup>8</sup>, Anthony M. Magliocco<sup>9</sup>, Terry Hyslop<sup>4</sup>, and Hallgeir Rui<sup>1,2</sup>

<sup>1</sup>Department of Cancer Biology, Kimmel Cancer Center, Thomas Jefferson University, Philadelphia, Pennsylvania

<sup>2</sup>Department of Pathology, Thomas Jefferson University, Philadelphia, Pennsylvania

<sup>3</sup>Department of Oncology, Tom Baker Cancer Centre, Alberta Cancer Board, Alberta, Canada

<sup>4</sup>Division of Biostatistics, Thomas Jefferson University, Philadelphia, Pennsylvania

<sup>5</sup>Department of Surgery, Thomas Jefferson University, Philadelphia, Pennsylvania

<sup>6</sup>Walter Reed National Military Medical Center, Bethesda, Maryland

<sup>7</sup>MDR Global Systems, LLC, Windber, Pennsylvania

<sup>8</sup>Department of Pathology, Yale University School of Medicine, New Haven, Connecticut

<sup>9</sup>Department of Pathology, Lee Moffitt Cancer Center & Research Institute, Tampa, Florida

Supported by NIH grants R01-CA101841 and R01-CA118740 (H.R.), Komen for the Cure Promise Grant KG091116 (H.R., A.K.W., C.L., J.A.H., A.J.K., C.D.S., T.H.), DOD CDMRP Predoctoral Fellowship (A.R.P.), and NCI Support Grant 1P30CA56036 to the Kimmel Cancer Center. The Project is funded, in part, under a Commonwealth University Research Enhancement Program grant with the Pennsylvania Department of Health (H.R.). The Department specifically disclaims responsibility for any analyses, interpretations or conclusions. The views expressed in this article are those of the authors and do not reflect the official policy of the Department of the Army (DOA), Department of Defense (DOD), or US Government.

<sup>†</sup>Address correspondence to: Hallgeir Rui, Department of Cancer Biology, Thomas Jefferson University, 233 S. 10<sup>th</sup> St, Philadelphia, PA 19107. [hallgeir.rui@jefferson.edu](mailto:hallgeir.rui@jefferson.edu)

**Running title:** Stat5a and breast cancer prognosis and antiestrogen response.

**Key words:** Stat5a, breast cancer prognosis, antiestrogen therapy outcome

**Word count:** 3,600

**Figures and Tables:** 3 Figures, 3 Tables, 1 Supplementary Table

**Abstract** (248 words)

**Purpose:** In breast cancer, reduced levels of nuclear localized and tyrosine phosphorylated Stat5a/b are associated with unfavorable prognosis and increased risk of antiestrogen therapy failure. Growth and differentiation of normal mammary epithelia require Stat5a but not Stat5b. We therefore investigated the association of Stat5a levels in breast cancer with prognosis and antiestrogen therapy failure.

**Experimental Design:** Stat5a protein levels were quantified *in situ* by automated immunofluorescence analysis in a breast cancer progression material (normal, DCIS, invasive cancer, and lymph node metastases) and by immunofluorescence or DAB-chromogen detection in four archival cohorts of patients with clinical outcome data using univariate and multivariate statistical models, adjusting for tumor grade, tumor size, lymph node status and ER/PR and Her2 status.

**Results:** Protein levels of Stat5a but not Stat5b were reduced in invasive breast cancer tissues and lymph node metastases compared to normal epithelia (Material I, n=180). In two cohorts of therapy-naïve, node-negative breast cancer patients (Material II, n=233; Material III, n=291), low expression of nuclear Stat5a was an independent marker of poor prognosis. Multivariate analysis of two additional patient cohorts treated with antiestrogen monotherapy (Material IV, n=75; Material V, n=97) revealed that low tumor levels of Stat5a in both cohorts were associated with a greater than 4-fold risk of antiestrogen therapy failure.

**Conclusions:** Loss of Stat5a in invasive breast cancer represents a new independent marker of poor prognosis in node-negative, non-adjuvant treated breast cancer patients. Stat5a holds potential to become a useful diagnostic tool to predict tumor response to antiestrogen therapy.

***Translational Relevance:*** (120-150 words; current 139)

Collectively, our observations suggest that low levels of Stat5a are not only a poor prognostic factor associated with cancer progression, but are also associated with antiestrogen therapeutic outcome and may have utility as a predictive marker. There is a great need for better predictive markers for antiestrogen therapy. Approximately 75% of newly diagnosed breast cancer patients have estrogen receptor positive disease and are predicted to respond to antiestrogen therapy, yet close to one third of patients with estrogen receptor-positive breast cancers fail to respond to antiestrogen therapy due to inherent or acquired resistance. Many markers including human epidermal growth factor receptor 2 (HER2) are only weakly predictive of response to antiestrogen therapy. If validated in randomized clinical trial materials, levels of nuclear localized Stat5a may help identify patients with estrogen receptor-positive breast cancer who are at increased risk of antiestrogen therapy failure.

## **Introduction** (4560 words, 50 references)

Signal transducer and activator of transcription-5a (Stat5a) was originally isolated from mammary glands of pregnant sheep and initially termed "Mammary gland factor" (Wakao, Gouilleux *et al.* 1994). Canonical Stat5a signaling in mammary epithelia involves prolactin-induced activation of tyrosine kinase Jak2 (Gouilleux, Wakao *et al.* 1994; Xie, LeBaron *et al.* 2002; Wagner, Krempler *et al.* 2004) which then phosphorylates Stat5a on tyrosine residue Y694. Phosphorylated Stat5a molecules undergo functional dimerization, nuclear translocation, and DNA binding to modulate expression of target genes and promote cell survival, proliferation, and differentiation (Gouilleux, Wakao *et al.* 1994; Liu, Robinson *et al.* 1997; Nevalainen, Xie *et al.* 2002). Stat5a becomes highly phosphorylated and hyperactivated during terminal differentiation of mammary tissue in pregnancy and lactation (Liu, Robinson *et al.* 1997; Nevalainen, Xie *et al.* 2004). Gene knockout studies in mice have revealed that Stat5a, but not the highly homologous Stat5b, is critical for specification in the mammary gland of the alveolar secretory epithelial cell lineage during pregnancy and for terminal differentiation and lactation (Liu, Robinson *et al.* 1997; Udy, Towers *et al.* 1997). Consistent with the key role of Stat5a as a mediator of prolactin receptor-Jak2 signaling and differentiation of the alveolar secretory cell lineage, gene deletion of Stat5a in mice (Liu, Robinson *et al.* 1997) also mimics the disruption of terminal mammary epithelial differentiation and lactation observed in mice deficient in prolactin, prolactin receptors, or Jak2 (Horseman, Zhao *et al.* 1997; Ormandy, Binart *et al.* 1997; Wagner, Krempler *et al.* 2004).

In healthy human and rodent mammary epithelia outside of pregnancy, Stat5a is phosphorylated and transcriptionally active at a basal level (Nevalainen, Xie *et al.* 2002). Whereas normal human breast epithelia are positive for nuclear localized, tyrosine phosphorylated Stat5 (Nuc-pYStat5a/b), a progressive loss of Nuc-pYStat5a/b occurs in invasive and metastatic breast cancer (Nevalainen, Xie *et al.* 2004; Tran, Utama *et al.* 2010; Peck, Witkiewicz *et al.* 2011). Loss of Nuc-pYStat5a/b has been

demonstrated to be a strong and independent marker of poor prognosis in node-negative breast cancer (Nevalainen, Xie *et al.* 2004; Peck, Witkiewicz *et al.* 2011) and was also a strong independent marker associated with antiestrogen therapy failure in primary breast cancer (Peck, Witkiewicz *et al.* 2011). However, the mechanisms underlying the loss of Nuc-pYStat5a/b during breast cancer progression have not been elucidated. One mechanism may involve disruption of Stat5 phosphorylation during breast cancer progression through upregulation of the Jak2 tyrosine phosphatase, PTP1B. We have reported that PTP1B effectively disrupts prolactin-induced Stat5 tyrosine phosphorylation in human breast cancer cell lines, and that levels of PTP1B in invasive human breast cancer specimens were inversely correlated with levels of Nuc-pYStat5a/b (Johnson, Peck *et al.* 2010). However, alternative mechanisms underpinning loss of Nuc-pYStat5a/b in progressing breast cancer are also likely to exist.

Several studies have reported low levels of Stat5 protein in invasive breast cancer, suggesting that reduced protein levels of Stat5a and/or Stat5b may contribute to reduced levels of Nuc-pYStat5a/b. First, analysis of whole cell extracts from invasive breast tumor tissue indicated that only 1 of 63 cases had strong Stat5 DNA binding activity and that tyrosine phosphorylated Stat5 protein was undetectable by immunoblotting in all cases (Widschwendter, Tonko-Geymayer *et al.* 2002). Second, only 34% of 517 invasive breast cancer cases were positive for epithelial cell Stat5 expression detected by immunohistochemistry and nuclear Stat5 was detected in as few as 18 (3%) of cases (Yamashita, Nishio *et al.* 2006). Third, cytoplasmic or nuclear Stat5a was detected in 17% of 30 cases of breast adenocarcinomas compared to nearly 100% positive expression in normal tissue (Bratthauer, Strauss *et al.* 2006). In contrast, our initial study of a breast cancer progression material using a pan-Stat5a/b antibody indicated that malignant breast tumors generally remained positive for expression of Stat5 protein across the progression series, although quantification was not attempted and the antibody used did not discern between Stat5a and Stat5b (Nevalainen, Xie *et al.* 2004). Furthermore, other groups reported high frequencies of nuclear Stat5a expression in breast

adenocarcinomas, ranging from 48% to 74% (Cotarla, Ren *et al.* 2004; Shan, Yu *et al.* 2004). Collectively, these discrepant and incomplete data warrant a more systematic effort to quantify Stat5a or Stat5b protein expression during human breast cancer progression in relation to normal breast tissue.

Considering the importance of Stat5a in normal mammary development, particularly its role in epithelial differentiation, disruption of Stat5a signaling could lead to less differentiated and more aggressive tumors. Several efforts have been made to correlate Stat5 with clinical disease response. Total cellular Stat5 protein expression was reported to be prognostic of patient outcome in an ER-positive subset of tumors but not in the entire ER-positive/negative population, however nuclear localization or phosphorylation status of Stat5 was not determined (Yamashita, Nishio *et al.* 2006). Low levels of nuclear localized and tyrosine phosphorylated Stat5a/b (Nuc-pYStat5a/b), detected using a tyrosine-phosphorylation specific antibody, have been established to be a marker of poor prognosis in multiple independent cohorts of node-negative, therapy-naïve breast cancer patients (Nevalainen, Xie *et al.* 2004; Peck, Witkiewicz *et al.* 2011). In contrast, DNA binding analysis using a phospho-Stat5a/b antibody and whole tumor tissue extracts from a small cohort of patients did not find a significant correlation with disease outcome (Widschwendter, Tonko-Geymayer *et al.* 2002). However, stromal cell components in the tissue extracts may have diluted the Stat5 DNA binding activity from the epithelial tumor compartment, a limitation not present in direct immunohistochemical analyses of tissue sections. Importantly, low levels of Nuc-pYStat5a/b also were associated with increased risk of failure of antiestrogen therapy in node-negative and node-positive breast cancer (Peck, Witkiewicz *et al.* 2011) and undetectable Stat5 protein was associated with poor response to post-relapse antiestrogen therapy (Yamashita, Nishio *et al.* 2006). The prognostic utility of the Stat5a protein or the association of Stat5a expression with response to conventional adjuvant antiestrogen therapy has not been investigated to date.

The present study employed selective antibodies and immunofluorescence labeling on the Automated Quantitative Analysis (AQUA) platform (Camp, Chung *et al.* 2002; McCabe, Dolled-Filhart *et al.* 2005) to quantify levels of Stat5a or Stat5b proteins in a breast cancer progression material. Unexpectedly, the analyses revealed frequent and selective loss of nuclear as well as total Stat5a protein in invasive breast cancer and lymph node metastases, whereas expression of total cellular and nuclear Stat5b protein remained unchanged. Stat5a expression levels were further analyzed in four distinct archival cohorts of breast cancer with clinical outcome data, employing both traditional diaminobenzidine (DAB)-chromogen immunohistochemistry (IHC) with pathologist scoring and immunofluorescence-based quantification on the AQUA platform. We present novel data indicating that loss of nuclear localized Stat5a is prognostic of unfavorable breast cancer outcome in patients who did not receive systemic adjuvant therapy and is furthermore associated with elevated risk of failure of antiestrogen therapy. Thus the previously observed loss of Nuc-pYStat5a/b during breast cancer progression and the prognostic value of Nuc-pYStat5a/b may in part result from loss of Stat5a protein expression.

## Materials and Methods

**Breast Tumor Specimens.** Five independent and deidentified clinical cohorts of breast cancer tissues were evaluated. Tissues were represented as formalin-fixed and paraffin-embedded whole tissue sections or tissue microarrays. The research use of deidentified tissues was approved by the ethics committee of the respective institutions. Clinical information was not available for progression Material I, and patient demographics and clinical features of Materials II, III, and IV are presented in **Table 1**.

Material I was a breast cancer progression tissue array constructed using cutting edge matrix assembly (LeBaron, Crismon *et al.* 2005) and represented 180 unmatched patient specimens, including 40 normal breast tissues, 20 ductal carcinoma *in situ* (DCIS), 100 invasive ductal carcinomas (IDC), and 20 lymph node breast cancer metastases (Tran, Utama *et al.* 2010). Tissues were obtained from Thomas Jefferson University Hospital archives. The array contained 59% estrogen receptor (ER)-positive, 42% progesterone receptor (PR)-positive, and 20% Her2-positive cases as determined by pathologist scoring of standard immunohistochemistry. Stat5a scores, detected by immunofluorescence and quantified by automated quantitative analysis (AQUA), were obtained for 126 cases and Stat5b AQUA scores were obtained for 116 cases. Material II, obtained through the National Cancer Institute's Cooperative Breast Cancer Tissue Resource, was comprised of whole tissue sections from patients with node-negative IDC who did not receive adjuvant systemic therapy (N=233). Nuclear Stat5a (Nuc-Stat5a) scores were calculated through pathologist scoring of traditional DAB-chromogen immunohistochemistry (DAB-IHC) and obtained for 223 tumor specimens. Material III was a breast cancer tissue microarray (0.6mm cores) from Yale University pathology archives, representing 291 node-negative IDC patients who did not receive adjuvant systemic therapy (Dolled-Filhart, Ryden *et al.* 2006). Nuc-Stat5a AQUA scores were obtained for 239 tumors. Material IV was comprised of whole tissue sections from tumors of node-negative IDC patients who received adjuvant hormone monotherapy (N=75). Specimens were obtained through the National Cancer

Institute's Cooperative Breast Cancer Tissue Resource. Nuc-Stat5a DAB-IHC scores were obtained for 73 tumor specimens. Material V was a 0.6 mm tumor tissue core microarray ("Tamoxifen 50/50 array") from a random series of breast cancer patients identified through the Alberta Cancer Registry (Calgary, Alberta, Canada) who received adjuvant hormone monotherapy for up to 60 months and an average of 30 months (Peck, Witkiewicz *et al.* 2011). The array was constructed in triplicate from tumors of 50 patients who died of breast cancer, classified as resistant to antiestrogen therapy, and 50 patients with greater than 5-years follow-up without breast cancer recurrence, classified as good responders. AQUA analysis yielded informative data on Nuc-Stat5a for 71 cases.

***Quantification of Stat5a and Stat5b by automated quantitative analysis (AQUA).*** Rabbit polyclonal anti-sera for detection of Stat5a and Stat5b were generated against unique epitopes of each protein as previously described (Nevalainen, Xie *et al.* 2002). Antigen retrieval was performed using the DAKO PT-module with citric acid buffer (pH 6.0). Immunofluorescent staining was performed on a Dako Autostainer with Stat5a or Stat5b antisera diluted 1:8,000 or 1:4,000, respectively, and coincubated with mouse monoclonal anti-pancytokeratin (clone AE1/AE3, DAKO, 1:100 dilution) to define the tumor mask (Camp, Chung *et al.* 2002) for 20 min. HRP-conjugated anti-rabbit IgG and Alexa-488-conjugated anti-mouse IgG secondary antibodies were added for 30 min, followed by 10 min incubation with Cy5-tyramide (Perkin Elmer, Waltham, MA). Slides were coverslipped using DAPI-containing mounting media. The AQUA/PM2000 platform (HistoRx, New Haven, CT) (Camp, Chung *et al.* 2002) was used to quantify fluorescence-based immunostaining by automatically scanning the slides and capturing fluorescent images in three channels, FITC/Alexa-488 (cytokeratin), Cy5 (Stat5a, Stat5b) or DAPI (nuclei). AQUA scores were objectively derived by calculating the mean signal intensities within the cell nuclei of the epithelial compartment, defined by cytokeratin-positive and DAPI-positive mapping and blinded to clinical data. A significant number of cases were not interpretable by AQUA due to loss of histospots or insufficient staining quality or

tumor sampling (<5% tumor area) required for automated analysis (Gustavson, Bourke-Martin et al. 2009).

**Detection of Stat5a by DAB-chromogen immunohistochemistry.** Paraffin-embedded, formalin-fixed tissues were deparaffinized and rehydrated prior to antigen retrieval using citric acid buffer (pH 6.0), blocking of endogenous peroxidase activity using 0.3% H<sub>2</sub>O<sub>2</sub> for 10 min at room temperature, and reduction of non-specific immunoglobulin binding using normal goat serum for 2 h at room temperature. Stat5a antibody was incubated with tissue slides at a final dilution of 1:8,000 for 30 min at room temperature using a Dako Autostainer Plus (Dako, Carpinteria, CA). Antigen-antibody complexes were detected using biotinylated goat anti-rabbit secondary antibody (Biogenex, San Ramon, CA) followed by streptavidin-horseradish-peroxidase complex, using 3,3-(prime) DAB as chromogen and Mayer hematoxylin as counterstain. Individual breast tumor samples were scored by a pathologist, blinded to clinical outcome, for transcriptionally active Stat5a as defined by detectable nuclear localization of the protein (Nuc-Stat5a).

**Statistical Methods.** One-way ANOVA with Dunnett T3 pairwise post-hoc test assuming unequal variances (SPSS v15.0, SPSS Inc, Chicago, IL) was used to test for differences in Stat5a or Stat5b levels between breast histology groups in progression Material I. For Materials II, IV and V, clinical endpoints for survival analyses were breast cancer-specific survival (CSS) and time-to-recurrence (TTR) of either local or distant disease according to consensus definitions (Punt, Buyse et al. 2007). For Material III only CSS was available. Statistical software R version 2.11.1 (R Development Core Team 2011) was used to determine the optimal cutpoint for low and high Stat5a-expressing tumors as a function of patient survival. Kaplan-Meier survival curves were generated and differences in survival were measured using the log-rank test or adjusted Cox regression when the proportional hazard assumption passed (completed globally using a Wald chi-square test for each cohort and outcome multivariate model) or adjusted Weibull regression (assessment made graphically) when

proportional hazard assumption failed (SAS v9.2; SAS Institute, Cary, NC). When available, variables included in the adjusted models were tumor grade, tumor size, lymph node status, ER/PR expression, Her2 status, and Nuc-Stat5a.

## Results

***Levels of nuclear localized Stat5a are diminished during breast cancer progression.*** Reduced levels of Nuc-pYStat5a/b previously have been demonstrated in more poorly differentiated tumors and aggressive disease when compared to normal epithelia (Cotarla, Ren et al. 2004; Nevalainen, Xie et al. 2004; Bratthauer, Strauss et al. 2006; Peck, Witkiewicz et al. 2011). To determine whether levels of Nuc-Stat5a were correspondingly reduced over breast cancer progression, a breast tissue progression array (Material I) that included normal breast, ductal carcinoma *in situ* (DCIS), invasive ductal carcinomas (IDC), and lymph node metastases, was stained for Stat5a or Stat5b by immunofluorescence labeling and quantified by AQUA (**Figure 1**). Levels of Nuc-Stat5a were markedly reduced during breast cancer progression with significant loss of Nuc-Stat5a expression in IDC ( $p < 0.001$ ) and greatest loss in lymph node metastases ( $p < 0.001$ ; **Figure 1A, 1C**). Interestingly, loss of Nuc-Stat5a over progression occurred in parallel with our previously reported loss of Nuc-pYStat5a/b in the same breast progression tissue array (Peck, Witkiewicz et al. 2011). In contrast, Nuc-Stat5b remained detectable and did not change over progression (**Figure 1B, 1C**). Stat5a and Stat5b differ most within the C-terminal transactivation domain and selective antibodies have been generated to this region of each protein (Nevalainen, Xie et al. 2002). To confirm selectivity of each antibody, Stat5a or Stat5b proteins were immunoprecipitated from T47D breast cancer cell lysates and resolved by gel electrophoresis. Cross-immunoblotting using the Stat5a or Stat5b antibodies demonstrated lack of cross-reactivity, while immunoblotting with a pan-Stat5a/b antibody detected the slightly slower-migrating Stat5a band (94 kDa) in the Stat5a immunoprecipitate and the slightly faster migrating Stat5b band (92 kDa) in the Stat5b immunoprecipitate (**Figure 1D**). These quantitative *in*

*situ* data provide novel information indicating that expression of Nuc-Stat5a, but not Nuc-Stat5b, is lost over breast cancer progression.

***Cellular expression of Stat5a but not Stat5b protein is suppressed over breast cancer progression.*** The absence of Nuc-Stat5a protein in invasive breast cancer and lymph node metastases is consistent with our previous reports showing a loss of Nuc-pYStat5a/b (Nevalainen, Xie *et al.* 2004; Peck, Witkiewicz *et al.* 2011). However, it is unclear to what extent the observed reduction in Nuc-Stat5a is a result of loss of tyrosine phosphorylation and nuclear translocation or an overall suppression of Stat5a protein expression. We have previously shown that levels of the Jak2 phosphatase, PTP1B, were inversely correlated with Nuc-pYStat5a/b expression in human breast cancer and that PTP1B suppressed prolactin-induced Stat5a phosphorylation levels (Johnson, Peck *et al.* 2010), suggesting that the previously reported reduction in Nuc-pYStat5a/b in breast cancer is at least in part due to reduced phosphorylation of Stat5a. To determine whether reduction of Nuc-pYStat5a/b and Nuc-Stat5a levels could also be attributed to an overall suppression of Stat5a protein expression, we quantified the levels of total cellular Stat5a protein within the breast tissue progression array. Interestingly, total cellular levels of Stat5a protein were significantly reduced from normal epithelia to more advanced breast cancer disease (**Figure 1E**), thus paralleling the detected loss of Nuc-Stat5a. In contrast, total cellular Stat5b levels remained constant across the progression material (**Figure 1F**).

Phospho-tyrosine motifs surrounding the conserved amino acid residues Y694 of Stat5a and Y699 of Stat5b are identical and cannot be distinguished by selective phospho-tyrosine Stat5 antibodies. In the absence of selective phospho-tyrosine-Stat5a or phospho-tyrosine-Stat5b antibodies, levels of nuclear Stat5a or nuclear Stat5b can be measured as a proxy for tyrosine-phosphorylated and transcriptionally active protein. Following *ex vivo* prolactin stimulation of viable human breast tissue explants, levels of Nuc-Stat5a were significantly induced in the luminal epithelial cells (**Figure 1G**,

upper panels). Correspondingly, Nuc-pYStat5a/b was highly enriched in the epithelial cell nuclei of the same prolactin-stimulated tissues (**Figure 1G**, lower panels). In contrast, prolactin-inducible levels of Nuc-Stat5b were only weakly detected (data not shown). We conclude that loss of Nuc-Stat5a in invasive breast cancer and metastases is a reflection of loss of cellular expression of Stat5a protein and not simply a result of reduced tyrosine phosphorylation or nuclear exclusion.

***Nuclear localized Stat5a is an independent marker of prognosis in node-negative breast cancer.*** Considering that the observed loss of Nuc-Stat5a corresponds with our previously identified loss of Nuc-pYStat5a/b, we hypothesized that Nuc-Stat5a likewise would predict clinical outcome. We first evaluated whole tumor tissue sections from patients with node-negative breast cancer who did not receive any systemic adjuvant treatment (Material II) for expression of Nuc-Stat5a using DAB-IHC. Survival analyses were performed to evaluate time to recurrence (TTR) and cancer-specific survival (CSS). Low levels of Nuc-Stat5a were associated with increased risk of breast cancer recurrence (TTR; Log-rank  $p=0.003$ ,  $N=223$ , **Figure 2A**; univariate Weibull regression hazard ratio (HR)=2.60 (1.36, 4.96),  $p=0.004$ ,  $N=218$ , **Supplementary Table 1**). Importantly, low expression of Nuc-Stat5a also was associated with poor breast cancer-specific survival (CSS; Log-rank  $p=0.007$ ,  $N=223$ , **Figure 2B**; univariate Weibull regression HR=2.35 (1.20, 4.58),  $p=0.012$ ,  $N=218$ ; **Table 2**). To validate these data, we analyzed levels of Nuc-Stat5a using objective, automated quantitative analysis (AQUA) of immunofluorescence in a tissue microarray from an independent cohort of node-negative breast cancer patients who also did not receive any adjuvant therapy (Material III). Nuc-Stat5a levels, quantified by the AQUA platform, predicted breast cancer survival in Material III (CSS; Log-rank  $p=0.021$ ,  $N=239$ , **Figure 2C**; univariate Weibull regression HR=2.13 (1.01, 4.49),  $p=0.047$ ,  $N=190$ , **Table 2**). TTR data was not available for Material III.

Nuc-Stat5a remained an independent marker of disease prognosis in multivariate analyses in both Materials II and III. When adjusting for standard clinical and pathological markers in Material II,

patients with low Nuc-Stat5a expression had an adjusted 2.5-fold increased risk of disease recurrence (TTR; multivariate Weibull regression HR=2.55 (1.31, 4.98),  $p=0.006$ , N=218, **Supplementary Table 1**) and a 2.3-fold greater risk of dying from breast cancer (CSS; multivariate Weibull regression HR=2.34 (1.16, 4.71),  $p=0.018$ , N=218, **Table 2**). Likewise, in Material III, AQUA quantification of Nuc-Stat5a was an independent prognostic indicator of breast cancer-specific survival as reflected in a 2.2-fold increased risk of death (CSS; multivariate Weibull regression HR=2.15 (1.02, 4.55),  $p=0.045$ , N=190, **Table 2**). We found comparable effect sizes in two separate cohorts and by two analytical approaches, and conclude that Nuc-Stat5a is an independent prognostic marker of outcome in patients with lymph node-negative breast cancer. The effect sizes observed using Nuc-Stat5a as a marker of favorable outcome in these prognostic cohorts were comparable to those observed using Nuc-pYStat5 in these same cohorts (Peck, Witkiewicz et al. 2011).

***Loss of nuclear Stat5a is associated with increased risk of antiestrogen therapy failure.*** To determine if Stat5a would be a significant predictor of patient response to antiestrogen therapy, we analyzed Nuc-Stat5a expression levels by DAB-IHC and pathologist scoring of whole tissue sections from a cohort of node-negative breast cancer patients who received adjuvant antiestrogen therapy but did not receive chemotherapy. The absence of detectable Nuc-Stat5a in tumors from these patients indicated a significantly increased risk of breast cancer-specific death (CSS; Log-rank  $p<0.001$ , N=73, **Figure 3A**; univariate Cox regression HR=6.73 (1.88, 24.05),  $p=0.003$ , N=73). Patients with undetectable Nuc-Stat5a were also at an increased risk of breast cancer recurrence (TTR; Log-rank  $p=0.003$ , N=73, **Figure 3B**; univariate Cox regression HR=5.08 (1.52, 17.01),  $p=0.008$ , N=73). Multivariate analysis of this cohort revealed that Nuc-Stat5a remained an independent marker of patient outcome after adjustment for other tumor parameters both by CSS (multivariate Cox regression HR=4.19 (1.13, 15.48),  $p=0.032$ , N=73) and by TTR (multivariate Cox regression HR=4.27 (1.20, 15.19),  $p=0.025$ , N=73). ER/PR status could not be included in the

multivariate model for Material IV, because only one of the patients with all clinical parameters evaluable for multivariate analysis had an ER/PR-negative tumor.

Analysis of Nuc-Stat5a expression as a function of survival was then performed in an independent tissue microarray of tumors which included both node-negative and node-positive patients who were also treated exclusively with antiestrogen monotherapy (Material V, N=97). Nuc-Stat5a was detected by immunofluorescence and quantified by AQUA. Patients whose tumors expressed low levels of Nuc-Stat5a were at markedly increased risk of failing antiestrogen treatment in this independent material. Risk of death from breast cancer was significantly increased in patients with low levels of Nuc-Stat5a (CSS Log-rank  $p=0.025$ , N=71, **Figure 3C**; univariate Cox regression HR=2.75 (1.33, 5.69),  $p=0.006$ , N=55, **Table 3**). Univariate Cox regression also revealed a significant increased risk of breast cancer recurrence in these patients (TTR univariate Cox regression HR=2.15 (1.06, 4.35),  $p=0.033$ , N=55), although despite a clear separation of the two groups by Kaplan-Meier graphical representation, log-rank analysis did not reach statistical significance (TTR Log-rank  $p=0.11$ , N=71, **Figure 3D**). In multivariate analysis adjusting for tumor size, tumor grade, lymph node status, ER/PR status, and Her2 overexpression, Nuc-Stat5a remained an independent marker of survival in Material V predicting a 5-fold increased risk of breast cancer-related death (CSS; multivariate Cox regression HR=4.95 (1.87, 13.06),  $p=0.001$ , N=55, **Table 3**). Positive node status (CSS; multivariate Cox regression HR=4.72 (1.90, 11.74),  $p<0.001$ , N=55) and tumor size (CSS; multivariate Cox regression tumor size 2-<5cm HR=3.51 (1.26, 9.74),  $p=0.016$  and tumor size  $\geq 5$ cm HR=8.54 (1.92, 37.94),  $p=0.005$ , N=55) were also independent predictors of survival (**Table 3**). While tumors with low Nuc-Stat5a trended towards earlier breast cancer recurrence, TTR did not reach statistical significance in multivariate analysis of Material IV (TTR; multivariate Cox regression HR=2.14 (0.89, 5.12)  $p=0.087$ , N=55). Collectively, based on two independent cohorts of patients treated with antiestrogen therapy alone, low levels were in both cohorts associated with elevated risk of failure of antiestrogen

treatment. These initial studies justify further analysis of Nuc-Stat5a as a potentially clinically useful measurement of response to antiestrogen therapy.

## Discussion

The present study revealed a significant loss of both total cellular and nuclear levels of Stat5a in invasive breast cancer and lymph node metastases compared to levels of this transcription factor in normal breast epithelia and DCIS. In contrast, cellular and nuclear Stat5b protein levels remained unchanged, suggesting divergent expression and involvement of Stat5a and Stat5b during breast cancer progression. Analysis of two independent cohorts totaling more than 500 patients with therapy-naïve, lymph node-negative breast cancer, further revealed that loss of nuclear localized Stat5a (Nuc-Stat5a) was an independent marker of poor prognosis. In both patient cohorts reduced levels of Nuc-Stat5a, as detected by either DAB-IHC or AQUA, were associated with greater than two-fold increased risk of death from breast cancer after adjustment for other clinical parameters in multivariate analyses. In node-negative breast cancer Nuc-Stat5a may therefore be considered a strong prognostic marker (Stearns and Hayes 2002). Importantly, in two additional and distinct cohorts of patients who received adjuvant antiestrogen monotherapy, loss of Nuc-Stat5a was an independent marker associated with greater than four-fold increased risk of death from breast cancer. Low levels of Nuc-Stat5a therefore also hold potential to become a new clinically useful predictive marker of resistance to antiestrogen therapy, provided that this predictive association is validated in tumors from clinical trial cohorts of patients randomized for antiestrogen therapy. Collectively, these novel observations on Stat5a protein levels in breast cancer extend previous immunohistochemical reports documenting prognostic and antiestrogen response-predictive associations of nuclear localized, tyrosine phosphorylated Stat5a/b (Nuc-pYStat5a/b) (Nevalainen, Xie *et al.* 2004; Peck, Witkiewicz *et al.* 2011) or cellular Stat5 (Yamashita, Nishio *et al.* 2006).

Functional roles have previously been proposed for Stat5a in both initiation of breast cancer and during progression of established disease. *In vivo* studies in genetically engineered mice have supported a tumor initiation and cell survival role of Stat5a in the mammary gland (Humphreys and Hennighausen 1999; Iavnilovitch, Groner *et al.* 2002; Ren, Cai *et al.* 2002). However, a separate

model for the role of Stat5a signaling in established human breast cancer has emerged, suggesting that regardless of potential tumor initiating effects, Stat5a maintains established malignant breast tumors in a more differentiated state and thereby may suppress tumor progression (Wagner and Rui 2008). Support for this hypothesis comes from *in vitro* data indicating that activated Stat5a maintains cellular differentiation and suppresses epithelial-to-mesenchymal transition and invasive characteristics of human breast cancer cell lines (Sultan, Xie et al. 2005; Nouhi, Chughtai et al. 2006; Sultan, Brim et al. 2008; Tran, Utama et al. 2010) as well as immunohistochemical observations of positive associations between nuclear Stat5a and more well-differentiated tumors (Cotarla, Ren et al. 2004; Bratthauer, Strauss et al. 2006). We propose that pro-differentiation and invasion-suppressive effects of Stat5a in established human breast cancer provides a biological mechanism underpinning the observed prognostic value of nuclear Stat5a protein levels.

A general limitation of previous studies of Stat5 expression in breast cancer has been the lack of analytic distinction of Stat5a from the highly homologous Stat5b. Stat5a and Stat5b display many similar and redundant functions and structural features but are proteins encoded by separate genes and also have different regulatory features and functions (Kirken, Malabarba et al. 1997; Grimley, Dong et al. 1999; Yamashita, Nevalainen et al. 2001) as well as distinct tissue specific expression patterns (Liu, Robinson et al. 1995; Liu, Robinson et al. 1996; Liu, Robinson et al. 1997; Nevalainen, Xie et al. 2002). In the absence of antibodies selective for either phosphorylated Stat5a (Tyr694) or phosphorylated Stat5b (Tyr699) due to sequence-identity of these motifs, levels of nuclear localized Stat5a or Stat5b protein may be used as a substitute for phosphorylated and transcriptionally active dimers. Thus far, systematic analyses of Stat5a and Stat5b levels have not been assessed in breast cancer, and insight into the expression patterns of Stat5a and Stat5b proteins in breast cancer has remained incomplete due to conflicting data and preliminary analyses. Levels of Nuc-Stat5a or Nuc-pYStat5a/b have been reported in several studies to be reduced in breast cancer compared to normal breast epithelia, which consistently express nuclear localized Stat5 (Nevalainen, Xie et al. 2004;

Bratthauer, Strauss et al. 2006; Peck, Witkiewicz et al. 2011). However one study reported that Stat5a infrequently localized to the nucleus in normal breast epithelia compared to invasive ductal carcinomas (Shan, Yu et al. 2004) but this material was small and may therefore not be representative. Two other studies with apparently conflicting data reported nuclear Stat5 in as few as 3% of invasive breast cancer cases (Yamashita, Nishio et al. 2006), whereas nuclear Stat5a was detected in 74% of invasive breast cancer cases (Cotarla, Ren et al. 2004). Confounding the interpretation of these prior studies are their general lack of comparison of Stat5 levels in breast cancer to normal breast epithelia, their use of different antibodies, and the variable thresholds used for determining positive Stat5 staining. In a previous study based on staining with a pan-Stat5a/b antibody, we observed detectable Stat5a/b protein in all cases of normal breast epithelia and the majority of malignant breast tumors, whereas Nuc-pYStat5a/b protein status was frequently negative in advanced malignant disease (Nevalainen, Xie et al. 2004). However, in that study analyses were not quantitative and the reported detection of continued presence of total Stat5a/b protein is most likely explained by the persistent expression of Stat5b across the progression material as was quantified in the present study (**Figure 1F**). The novel data presented here demonstrate that Stat5a is the primary Stat5 protein lost during progression of breast cancer.

There are several limitations of this study. Our cohorts are confined to relatively small sample sizes, there is some loss of evaluable tissues on the tissue microarrays, and there are some missing or incomplete clinical data that reduces the statistical power of the multivariate analyses. Despite a strong predictive association of Nuc-Stat5a in CSS analyses of Material V of antiestrogen-treated patients, TTR in Material V failed to reach statistical significance by univariate log-rank and multivariate Cox regression models, although univariate Cox regression indicated a modest 2-fold significant increased risk of breast cancer recurrence. This discrepancy could be a result of limited numbers of cases or reflect that data may not be missing at random, although specific patterns within missing data did not emerge within any cohort of patients examined in this study (data not shown).

Retrospective analysis is a further limitation and the predictive analyses should be followed up with studies of randomized prospective trials. In follow-up studies of Stat5a as a biomarker of patient outcome and response to therapy, it will be paramount to ensure proper quantification and standardization of protein levels and achieving consistency in stratification of tumors into low and high expressers. The observed favorable prognosis associated with continued expression of Nuc-Stat5a in invasive breast cancer supports the hypothesis that in established human breast cancer, Stat5a signaling is important for maintaining tumor differentiation and suppressing disease progression (Sultan, Xie *et al.* 2005; Sultan, Brim *et al.* 2008; Wagner and Rui 2008). Further molecular mechanistic studies *in vivo* will be needed to determine conclusively whether casual mechanisms exist by which Stat5a suppresses disease progression or affects tumor responsiveness to antiestrogen therapy.

Nuc-Stat5a may emerge as a useful breast cancer marker of prognosis and antiestrogen responsiveness. Considering that as many as one third of patients with ER-positive breast cancer develop resistance to antiestrogen therapy and relapse (Osborne 1998; Clarke, Leonessa *et al.* 2001; Schiff, Massarweh *et al.* 2004; Prat and Baselga 2008), identification of patients with increased risk of therapy failure for stratification into more aggressive therapies is essential. Many strategies have been proposed to identify patients with therapy-resistant tumors, including PCR-based assays, gene expression profiling, or immunohistochemical detection of biomarkers such as Her2, EGFR, IGF-IR, progesterone receptor, p27, AIB1 (reviewed in (Osborne and Schiff 2011)), IRS-1, caveolin-1, retinoblastoma tumor suppressor, and nuclear localized and tyrosine phosphorylated Stat5a/b (Lehn, Ferno *et al.*; Migliaccio, Wu *et al.*; Bosco, Wang *et al.* 2007; Mercier, Casimiro *et al.* 2008; Witkiewicz, Dasgupta *et al.* 2009; Ertel, Dean *et al.* 2010; Peck, Witkiewicz *et al.* 2011). Nuc-Stat5a represents a new and potentially strong marker for further evaluation as a tool to identify antiestrogen therapy-responsive versus resistant primary breast cancer.

**Acknowledgements.** Whole tissue sections and clinical data for Material II were generously provided by the National Cancer Institute's Cooperative Breast Cancer Tissue Resource. We thank Eva Andersson for expert technical help with immunostaining of sections of Material II and Melanie Gironde for editorial assistance.

**Conflicts of interest.** The following authors or their immediate family members have indicated a financial interest. David L. Rimm, HistoRx (New Haven, CT).

**References (current 52 references; limit 50)**

- Bosco, E. E., Y. Wang, et al. (2007). "The retinoblastoma tumor suppressor modifies the therapeutic response of breast cancer." *J Clin Invest* **117**(1): 218-28.
- Bratthauer, G. L., B. L. Strauss, et al. (2006). "STAT 5a expression in various lesions of the breast." *Virchows Arch* **448**(2): 165-71.
- Camp, R. L., G. G. Chung, et al. (2002). "Automated subcellular localization and quantification of protein expression in tissue microarrays." *Nat Med* **8**(11): 1323-7.
- Clarke, R., F. Leonessa, et al. (2001). "Cellular and molecular pharmacology of antiestrogen action and resistance." *Pharmacol Rev* **53**(1): 25-71.
- Cotarla, I., S. Ren, et al. (2004). "Stat5a is tyrosine phosphorylated and nuclear localized in a high proportion of human breast cancers." *Int J Cancer* **108**(5): 665-71.
- Dolled-Filhart, M., L. Ryden, et al. (2006). "Classification of breast cancer using genetic algorithms and tissue microarrays." *Clin Cancer Res* **12**(21): 6459-68.
- Ertel, A., J. L. Dean, et al. (2010). "RB-pathway disruption in breast cancer: differential association with disease subtypes, disease-specific prognosis and therapeutic response." *Cell Cycle* **9**(20): 4153-63.
- Gouilleux, F., H. Wakao, et al. (1994). "Prolactin induces phosphorylation of Tyr694 of Stat5 (MGF), a prerequisite for DNA binding and induction of transcription." *Embo J* **13**(18): 4361-9.
- Grimley, P. M., F. Dong, et al. (1999). "Stat5a and Stat5b: fraternal twins of signal transduction and transcriptional activation." *Cytokine Growth Factor Rev* **10**(2): 131-57.
- Gustavson, M. D., B. Bourke-Martin, et al. (2009). "Development of an unsupervised pixel-based clustering algorithm for compartmentalization of immunohistochemical expression using Automated QUantitative Analysis." *Appl Immunohistochem Mol Morphol* **17**(4): 329-37.
- Horseman, N. D., W. Zhao, et al. (1997). "Defective mammapoiesis, but normal hematopoiesis, in mice with a targeted disruption of the prolactin gene." *Embo J* **16**(23): 6926-35.
- Humphreys, R. C. and L. Hennighausen (1999). "Signal transducer and activator of transcription 5a influences mammary epithelial cell survival and tumorigenesis." *Cell Growth Differ* **10**(10): 685-94.
- Iavnilovitch, E., B. Groner, et al. (2002). "Overexpression and forced activation of stat5 in mammary gland of transgenic mice promotes cellular proliferation, enhances differentiation, and delays postlactational apoptosis." *Mol Cancer Res* **1**(1): 32-47.
- Johnson, K. J., A. R. Peck, et al. (2010). "PTP1B suppresses prolactin activation of Stat5 in breast cancer cells." *Am J Pathol* **177**(6): 2971-83.
- Kirken, R. A., M. G. Malabarba, et al. (1997). "Prolactin stimulates serine/tyrosine phosphorylation and formation of heterocomplexes of multiple Stat5 isoforms in Nb2 lymphocytes." *J Biol Chem* **272**(22): 14098-103.
- LeBaron, M. J., H. R. Crismon, et al. (2005). "Ultrahigh density microarrays of solid samples." *Nat Methods* **2**(7): 511-3.
- Lehn, S., M. Ferno, et al. "A non-functional retinoblastoma tumor suppressor (RB) pathway in premenopausal breast cancer is associated with resistance to tamoxifen." *Cell Cycle* **10**(6): 956-62.
- Liu, X., G. W. Robinson, et al. (1995). "Cloning and expression of Stat5 and an additional homologue (Stat5b) involved in prolactin signal transduction in mouse mammary tissue." *Proc Natl Acad Sci U S A* **92**(19): 8831-5.
- Liu, X., G. W. Robinson, et al. (1996). "Activation of Stat5a and Stat5b by tyrosine phosphorylation is tightly linked to mammary gland differentiation." *Mol Endocrinol* **10**(12): 1496-506.
- Liu, X., G. W. Robinson, et al. (1997). "Stat5a is mandatory for adult mammary gland development and lactogenesis." *Genes Dev* **11**(2): 179-86.
- McCabe, A., M. Dolled-Filhart, et al. (2005). "Automated quantitative analysis (AQUA) of in situ protein expression, antibody concentration, and prognosis." *J Natl Cancer Inst* **97**(24): 1808-15.
- Mercier, I., M. C. Casimiro, et al. (2008). "Human breast cancer-associated fibroblasts (CAFs) show caveolin-1 downregulation and RB tumor suppressor functional inactivation: Implications for the response to hormonal therapy." *Cancer Biol Ther* **7**(8): 1212-25.

- Migliaccio, I., M. F. Wu, et al. "Nuclear IRS-1 predicts tamoxifen response in patients with early breast cancer." Breast Cancer Res Treat **123**(3): 651-60.
- Nevalainen, M. T., J. Xie, et al. (2002). "Basal activation of transcription factor signal transducer and activator of transcription (Stat5) in nonpregnant mouse and human breast epithelium." Mol Endocrinol **16**(5): 1108-24.
- Nevalainen, M. T., J. Xie, et al. (2004). "Signal transducer and activator of transcription-5 activation and breast cancer prognosis." J Clin Oncol **22**(11): 2053-60.
- Nouhi, Z., N. Chughtai, et al. (2006). "Defining the role of prolactin as an invasion suppressor hormone in breast cancer cells." Cancer Res **66**(3): 1824-32.
- Ormandy, C. J., N. Binart, et al. (1997). "Mammary gland development in prolactin receptor knockout mice." J Mammary Gland Biol Neoplasia **2**(4): 355-64.
- Osborne, C. K. (1998). "Tamoxifen in the treatment of breast cancer." N Engl J Med **339**(22): 1609-18.
- Osborne, C. K. and R. Schiff (2011). "Mechanisms of endocrine resistance in breast cancer." Annu Rev Med **62**: 233-47.
- Peck, A. R., A. K. Witkiewicz, et al. (2011). "Loss of nuclear localized and tyrosine phosphorylated stat5 in breast cancer predicts poor clinical outcome and increased risk of antiestrogen therapy failure." J Clin Oncol **29**(18): 2448-58.
- Prat, A. and J. Baselga (2008). "The role of hormonal therapy in the management of hormonal-receptor-positive breast cancer with co-expression of HER2." Nat Clin Pract Oncol **5**(9): 531-42.
- Punt, C. J., M. Buyse, et al. (2007). "Endpoints in adjuvant treatment trials: a systematic review of the literature in colon cancer and proposed definitions for future trials." J Natl Cancer Inst **99**(13): 998-1003.
- R Development Core Team (2011). *A Language and Environment for Statistical Computing*. Vienna, Austria, R Foundation for Statistical Computing.
- Ren, S., H. R. Cai, et al. (2002). "Loss of Stat5a delays mammary cancer progression in a mouse model." Oncogene **21**(27): 4335-9.
- Schiff, R., S. A. Massarweh, et al. (2004). "Cross-talk between estrogen receptor and growth factor pathways as a molecular target for overcoming endocrine resistance." Clin Cancer Res **10**(1 Pt 2): 331S-6S.
- Shan, L., M. Yu, et al. (2004). "Possible role of Stat5a in rat mammary gland carcinogenesis." Breast Cancer Res Treat **88**(3): 263-72.
- Stearns, V. and D. F. Hayes (2002). *Clinical Evaluation Criteria for Tumor Markers. Tumor markers: physiology, pathobiology, technology, and clinical applications*. E. P. Diamandis, AACCC Press: 19-23.
- Sultan, A. S., H. Brim, et al. (2008). "Co-overexpression of Janus kinase 2 and signal transducer and activator of transcription 5a promotes differentiation of mammary cancer cells through reversal of epithelial-mesenchymal transition." Cancer Sci **99**(2): 272-9.
- Sultan, A. S., J. Xie, et al. (2005). "Stat5 promotes homotypic adhesion and inhibits invasive characteristics of human breast cancer cells." Oncogene **24**(5): 746-60.
- Tran, T. H., F. E. Utama, et al. (2010). "Prolactin inhibits BCL6 expression in breast cancer through a Stat5a-dependent mechanism." Cancer Res **70**(4): 1711-21.
- Udy, G. B., R. P. Towers, et al. (1997). "Requirement of STAT5b for sexual dimorphism of body growth rates and liver gene expression." Proc Natl Acad Sci U S A **94**(14): 7239-44.
- Wagner, K. U., A. Krempler, et al. (2004). "Impaired alveologenesis and maintenance of secretory mammary epithelial cells in Jak2 conditional knockout mice." Mol Cell Biol **24**(12): 5510-20.
- Wagner, K. U. and H. Rui (2008). "Jak2/Stat5 signaling in mammaryogenesis, breast cancer initiation and progression." J Mammary Gland Biol Neoplasia **13**(1): 93-103.
- Wakao, H., F. Gouilleux, et al. (1994). "Mammary gland factor (MGF) is a novel member of the cytokine regulated transcription factor gene family and confers the prolactin response." Embo J **13**(9): 2182-91.
- Widschwendter, A., S. Tonko-Geymayer, et al. (2002). "Prognostic significance of signal transducer and activator of transcription 1 activation in breast cancer." Clin Cancer Res **8**(10): 3065-74.
- Witkiewicz, A. K., A. Dasgupta, et al. (2009). "An absence of stromal caveolin-1 expression predicts early tumor recurrence and poor clinical outcome in human breast cancers." Am J Pathol **174**(6): 2023-34.

- Xie, J., M. J. LeBaron, et al. (2002). "Role of tyrosine kinase Jak2 in prolactin-induced differentiation and growth of mammary epithelial cells." J Biol Chem **277**(16): 14020-30.
- Yamashita, H., M. T. Nevalainen, et al. (2001). "Role of serine phosphorylation of Stat5a in prolactin-stimulated beta-casein gene expression." Mol Cell Endocrinol **183**(1-2): 151-63.
- Yamashita, H., M. Nishio, et al. (2006). "Stat5 expression predicts response to endocrine therapy and improves survival in estrogen receptor-positive breast cancer." Endocr Relat Cancer **13**(3): 885-93.

**Table 1.** Characteristics of Material II, III, IV, and V. ER - estrogen receptor; PR - progesterone receptor; SD - standard deviation.

Variable	Material II Prognosis (N = 233)		Material III Prognosis (N = 291)		Material IV Antiestrogen (N = 75)		Material V Antiestrogen (N = 97)		
	No.	%	No.	%	No.	%	No.	%	
Center	Fox Chase	64	27	-	-	19	25	-	-
	Kaiser	79	34	-	-	35	47	-	-
	Univ of Miami	30	13	-	-	11	15	-	-
	Washington Univ	60	26	-	-	10	13	-	-
Race	Asian	1	0.4	0	0	1	1.3	-	-
	Black	16	7	3	1	6	8	-	-
	White	216	93	287	99	67	1.3	-	-
	Other	0	0	1	0.3	1	89	-	-
Age	<50	42	18	86	30	9	12	6	6
	≥50	191	82	205	70	66	88	91	94
Size (cm)	<2	117	50	114	39	41	55	39	40
	≥2 - <5	107	46	133	46	34	45	43	44
	≥5	9	4	33	11	0	0	11	11
	Missing	0	0	11	4	0	0	4	4
Grade	1	61	26	67	23	17	23	14	14
	2	104	45	137	47	45	60	47	48
	3	68	29	48	16	13	17	32	33
	Missing			39	13			4	4
ER Status	Negative	44	19	97	33	1	1.3	11	11
	Positive	186	80	163	56	72	96	80	82
	Missing	3	1.3	31	11	2	3	6	6
PR Status	Negative	65	28	103	35	9	12	-	-
	Positive	122	52	151	52	56	75	-	-
	Missing	46	20	37	13	10	13	-	-
ER/PR	Negative	44	19	68	23	1	1.3	12	12
	Positive	178	76	192	66	72	96	83	86
	Missing	11	5	31	11	2	3	2	2
HER2 Status	Negative	-	-	217	75	-	-	77	79
	Positive	-	-	33	11	-	-	13	13
	Missing	-	-	41	14	-	-	7	7
Nodal Status	Negative	233	100	291	100	73	100	42	43
	Positive	0	0	0	0	0	0	42	43
	Missing	0	0	0	0	0	0	13	13
Chemotherapy	Untreated	233	100	-	-	75	100		
	Treated	0	0	-	-	0	0		
	Missing	0	0	-	-	0	0		
Hormone therapy	Untreated	233	100	-	-	0	0	5	5
	Treated	0	0	-	-	75	100	92	95
Radiation therapy	Untreated	187	80	-	-	48	64	36	37
	Treated	46	20	-	-	27	36	59	61
	Missing	0	0	-	-	0	0	2	2
p5A Status	Low	30	13	200	69	9	12	24	25
	High	193	83	39	13	64	85	47	48
	Missing	10	4	52	18	2	3	26	27
CSS Events	Number	52	22	107	37	10	13	56	58
Evaluable subjects for multivariate	Number	218	94	190	65	73	97	55	57
Age (yrs)	mean (range)	62.3 (31 - 88)		57.3 (24 - 86)		64.6 (43 - 88)		69.8 (38 - 89)	
Tumor Size	mean (range)	2.1 (0.6 - 7.5)		2.5 (0.4 - 11)		1.9 (0.5 - 4.5)		2.7 (0.4 - 11)	
Follow-up (m)	median (range)	126 (3 - 326)		160 (1 - 425)		117 (10 - 195)		41 (4 - 143)	
p5A Value	mean (range)	32 (0 - 90)		653 (175 - 2099)		34.1 (0 - 80)		1949 (727 - 4626)	
Year Dx	range	1974 - 1990		1953 - 1980		1986 - 1996		1990 - 2000	

**Table 2.** Univariate and multivariate survival analyses of breast cancer specific survival (CSS) in Material II and III. Weibull regression survival analysis was used to evaluate prognostic factors. HR - hazard ratio; CI - confidence interval; ER - estrogen receptor; PR - progesterone receptor.

Material II CSS (N=218)		N	Multivariate Adjusted (Weibull)		Univariate Unadjusted (Weibull)	
Variable *			Hazard Ratio (95% CI)	p-value	Hazard Ratio (95% CI)	p-value
Grade	1	57	1	-	1	-
	2	95	1.01 (0.45 to 2.26)	0.977	1.33 (0.62 to 2.85)	0.472
	3	66	1.59 (0.65 to 3.87)	0.306	1.90 (0.88 to 4.08)	0.102
Size	<2 cm	108	1	-	1	-
	2-<5 cm	101	1.59 (0.82 to 3.08)	0.166	1.88 (1.03 to 3.45)	0.041
	>5 cm	9	2.38 (0.74 to 7.60)	0.145	2.86 (0.94 to 8.76)	0.065
ER/PR Status	Neg	43	1	-	1	-
	Pos	176	1.49 (0.68 to 3.26)	0.321	0.87 (0.45 to 1.70)	0.688
Stat5A	Low (0)	188	2.34 (1.16 to 4.71)	0.018	2.35 (1.20 to 4.58)	0.012
	High (>0)	30	1	-	1	-
* Global test: $\chi^2(4) = 17.18, p = 0.0018$						
Material III CSS (N=190)		N	Multivariate Adjusted (Weibull)		Univariate Unadjusted (Weibull)	
Variable *			Hazard Ratio (95% CI)	p-value	Hazard Ratio (95% CI)	p-value
Grade	1	42	1	-	1	-
	2	110	1.55 (0.84 to 2.84)	0.162	1.42 (0.77 to 2.59)	0.258
	3	38	1.11 (0.50 to 2.47)	0.797	1.10 (0.52 to 2.35)	0.801
Size	<2 cm	71	1	-	1	-
	2-<5 cm	98	2.10 (1.19 to 3.72)	0.011	2.01 (1.14 to 3.55)	0.016
	≥5 cm	21	3.48 (1.61 to 7.54)	0.002	3.09 (1.45 to 6.58)	0.004
ER/PR Status	Neg	41	1	-	1	-
	Pos	149	0.93 (0.49 to 1.75)	0.824	0.91 (0.52 to 1.59)	0.738
Her2 Status	Neg	166	1	-	1	-
	Pos	24	1.02 (0.47 to 2.20)	0.963	0.97 (0.48 to 1.96)	0.939
Stat5A	Low (<881)	158	2.15 (1.02 to 4.55)	0.045	2.13 (1.01 to 4.49)	0.047
	High (≥881)	32	1	-	1	-
* Global test: $\chi^2(5) = 10.05, p = 0.074$						

**Table 3.** Univariate and multivariate Cox regression survival analysis of breast cancer-specific survival (CSS) and time to recurrence (TTR) as a function of Nuc-Stat5a in breast cancer patients treated with antiestrogen monotherapy (Material IV and V). HR - hazard ratio; CI - confidence interval; ER/PR status - estrogen or progesterone receptor positive.

Material IV CSS (N=73) Variable *		N	Multivariate Adjusted (Cox)		Univariate Unadjusted (Cox)	
			Hazard Ratio (95% CI)	p-value	Hazard Ratio (95% CI)	p-value
Grade	1	17	1	-	1	-
	2	43	3.42 (0.42 to 27.87)	0.251	3.74 (0.46 to 30.52)	0.218
	3	13	2.60 (0.23 to 29.30)	0.438	3.32 (0.30 to 36.66)	0.327
Size	<2 cm	40	1	-	1	-
	≥2-<5 cm	33	4.29 (0.86 to 21.38)	0.075	5.79 (1.23 to 27.26)	0.026
Stat5A	Low (0)	9	4.19 (1.13 to 15.48)	0.032	6.73 (1.88 to 24.05)	0.003
	High (>0)	64	1	-	1	-
* Global test: $\chi^2(3) = 2.83, P = 0.42$						
Material IV TTR (N=73) Variable *		N	Multivariate Adjusted (Cox)		Univariate Unadjusted (Cox)	
			Hazard Ratio (95% CI)	p-value	Hazard Ratio (95% CI)	p-value
Grade	1	17	1	-	1	-
	2	43	1.20 (0.30 to 4.81)	0.794	1.27 (0.32 to 5.04)	0.729
	3	13	1.59 (0.31 to 8.22)	0.581	1.87 (0.37 to 9.39)	0.448
Size	<2 cm	40	1	-	1	-
	>2-<5 cm	33	1.63 (0.49 to 5.43)	0.429	2.24 (0.73 to 6.86)	0.158
Stat5A	Low (0)	9	4.27 (1.20 to 15.19)	0.025	5.08 (1.52 to 17.01)	0.008
	High (>0)	64	1	-	1	-
* Global test: $\chi^2(3) = 1.47, P = 0.69$						
Material V CSS (N=55) Variable *		N	Multivariate Adjusted (Cox)		Univariate Unadjusted (Cox)	
			Hazard Ratio (95% CI)	p-value	Hazard Ratio (95% CI)	p-value
Grade	1	8	1	-	1	-
	2	28	3.48 (0.91 to 13.31)	0.068	1.11 (0.36 to 3.41)	0.855
	3	19	2.83 (0.65 to 12.28)	0.164	2.25 (0.73 to 6.91)	0.158
Size	<2 cm	24	1	-	1	-
	2-<5 cm	26	3.51 (1.26 to 9.74)	0.016	4.22 (1.76 to 10.13)	0.001
	≥5 cm	5	8.54 (1.92 to 37.94)	0.005	6.94 (2.01 to 24.00)	0.002
LN Status	Neg	29	1	-	1	-
	Pos	26	4.72 (1.90 to 11.74)	<.001	4.30 (1.98 to 9.33)	<.001
ER/PR Status	Neg	8	1	-	1	-
	Pos	47	0.73 (0.22 to 2.38)	0.597	0.37 (0.16 to 0.86)	0.021
Her2 Status	Neg	48	1	-	1	-
	Pos	7	2.87 (0.92 to 8.96)	0.070	3.07 (1.23 to 7.63)	0.016
Stat5A	Low (<1454)	16	4.95 (1.87 to 13.06)	0.001	2.75 (1.33 to 5.69)	0.006
	High (≥1454)	39	1	-	1	-
* Global test: $\chi^2(6) = 3.19, P = 0.78$						
Material V TTR (N=55) Variable *		N	Multivariate Adjusted (Cox)		Univariate Unadjusted (Cox)	
			Hazard Ratio (95% CI)	p-value	Hazard Ratio (95% CI)	p-value
Grade	1	8	1	-	1	-
	2	28	1.17 (0.37 to 3.72)	0.794	0.87 (0.31 to 2.41)	0.783
	3	19	1.24 (0.33 to 4.68)	0.746	1.98 (0.71 to 5.55)	0.191
Size	<2 cm	24	1	-	1	-
	2-<5 cm	26	3.38 (1.30 to 8.80)	0.013	3.75 (1.67 to 8.43)	0.001
	≥5 cm	5	8.17 (2.26 to 29.50)	0.001	6.91 (2.27 to 21.06)	<.001
LN Status	Neg	29	1	-	1	-
	Pos	26	3.46 (1.57 to 7.64)	0.002	3.42 (1.67 to 7.01)	<.001
ER/PR Status	Neg	8	1	-	1	-
	Pos	47	0.51 (0.17 to 1.48)	0.214	0.30 (0.13 to 0.68)	0.004
Her2 Status	Neg	48	1	-	1	-
	Pos	7	3.35 (1.13 to 9.93)	0.029	3.06 (1.24 to 7.56)	0.015
Stat5A	Low (<1454)	16	2.14 (0.89 to 5.12)	0.087	2.15 (1.06 to 4.35)	0.033
	High (≥1454)	39	1	-	1	-
* Global test: $\chi^2(6) = 4.67, P = 0.59$						

**Supplementary Table 1.** Univariate and multivariate Cox regression survival analyses of time to recurrence (TTR) of breast cancer in Material II. HR - hazard ratio; CI - confidence interval; ER - estrogen receptor; PR - progesterone receptor.

Material II TTR (N=218) Variable *		N	Multivariate Adjusted (Weibull)		Univariate Unadjusted (Weibull)	
			Hazard Ratio (95% CI)	p-value	Hazard Ratio (95% CI)	p-value
Grade	1	57	1	-	1	-
	2	95	1.19 (0.55 to 2.60)	0.655	1.52 (0.72 to 3.21)	0.278
	3	66	1.74 (0.72 to 4.18)	0.215	2.06 (0.96 to 4.42)	0.065
Size	<2 cm	108	1	-	1	-
	2-<5 cm	101	1.56 (0.84 to 2.92)	0.159	1.86 (1.04 to 3.32)	0.036
	>5 cm	9	2.18 (0.70 to 6.78)	0.178	2.73 (0.91 to 8.23)	0.074
ER/PR Status	Neg	43	1	-	1	-
	Pos	175	1.48 (0.70 to 3.14)	0.306	0.86 (0.45 to 1.63)	0.634
Stat5A	Low (0)	30	2.55 (1.31 to 4.98)	0.006	2.60 (1.36 to 4.96)	0.004
	High (>0)	188	1	-	1	-

\* Global test:  $\chi^2(4) = 15.06, p = 0.0046$

## Figure Legends

**Figure 1.** Nuclear localization of Stat5a (Nuc-Stat5a) is lost during breast cancer progression. **A)** Detection of nuclear localized Stat5a by immunofluorescence and quantified by AQUA revealed a significant reduction in Nuc-Stat5a protein in invasive ductal carcinoma (IDC; n=66) and lymph node metastases (n=19) when compared to normal breast tissue (n=23) and ductal carcinoma *in situ* (DCIS; n=18). **B)** Levels of nuclear localized Stat5b (Nuc-Stat5b) remained unchanged between normal (n=24), DCIS (n=12), IDC (n=67) and lymph node metastases (n=13) in the same breast progression array (Material I). **C)** Representative images of Stat5a and Stat5b detected by immunofluorescence in normal human breast tissue and lymph node metastases. **D)** Immunoprecipitation of Stat5a (94 kDa) or Stat5b (92 kDa) protein from T47D breast cancer cells, followed by immunoblotting with the Stat5a, Stat5b, or a pan-Stat5a/b antibody, revealed specificity and lack of cross-reactivity of the Stat5a and Stat5b antibodies. **E-F)** Levels of total cellular Stat5a protein (**E**), but not total cellular Stat5b protein (**F**), were significantly reduced over breast cancer progression (Material I). **G)** Stat5a protein (upper panels) translocated to the nucleus following *ex vivo* prolactin stimulation of human breast tissue explants. Similarly, tyrosine phosphorylated Stat5a/b (pY-Stat5a/b) is induced in the nuclei of the same breast tissue following *ex vivo* stimulation with prolactin (lower panels). Statistical differences (ANOVA, Dunnett T3 post-hoc test) in levels of nuclear and total cellular Stat5a or Stat5b over breast cancer progression in relation to normal histological type are indicated; \*\*\*  $p < 0.001$ , \*  $p < 0.05$ . AQUA – Automated Quantitative Analysis; DCIS - ductal carcinoma *in situ*; IDC - invasive ductal carcinoma; LN Met - lymph node metastasis.

**Figure 2.** Loss of nuclear Stat5a (Nuc-Stat5a) predicts unfavorable breast cancer prognosis. **A-B)** Nuclear localization of Stat5a was detected by standard DAB IHC and pathologist review of whole tissue sections in Material II. Kaplan-Meier analysis indicated that patients with low levels of Nuc-Stat5a had **A)** reduced time-to-recurrence (TTR) of breast cancer and **B)** poor breast cancer-specific survival (CSS). **C)** Immunofluorescence and quantitative AQUA analysis of Nuc-Stat5a revealed by

Kaplan-Meier analysis that loss of Nuc-Stat5a was prognostic of poor breast cancer-specific survival (CSS) in Material III. Censored cases (+) and number of patients per group are indicated.

**Figure 3.** Low levels of nuclear Stat5a (Nuc-Stat5a) predict poor response to antiestrogen therapy. **A-B)** Nuc-Stat5a was detected by DAB-chromogen IHC and pathologist scoring of whole tissue sections from node-negative breast cancer patients treated with antiestrogen monotherapy (Material IV). The lack of nuclear localization of Stat5a predicted **A)** poor breast cancer-specific survival (CSS) and **B)** reduced time-to-recurrence (TTR) of breast cancer. **C-D)** Nuc-Stat5a expression levels were measured by immunofluorescence and quantified by AQUA in node-negative and positive breast cancer patients treated with antiestrogen monotherapy (Material V). Low levels of Nuc-Stat5a was predictive of **C)** poor cancer-specific survival (CSS) and **D)** reduced TTR of breast cancer. Kaplan-Meier plots with censored cases (+) and number of patients per group indicated.

Figure 1.

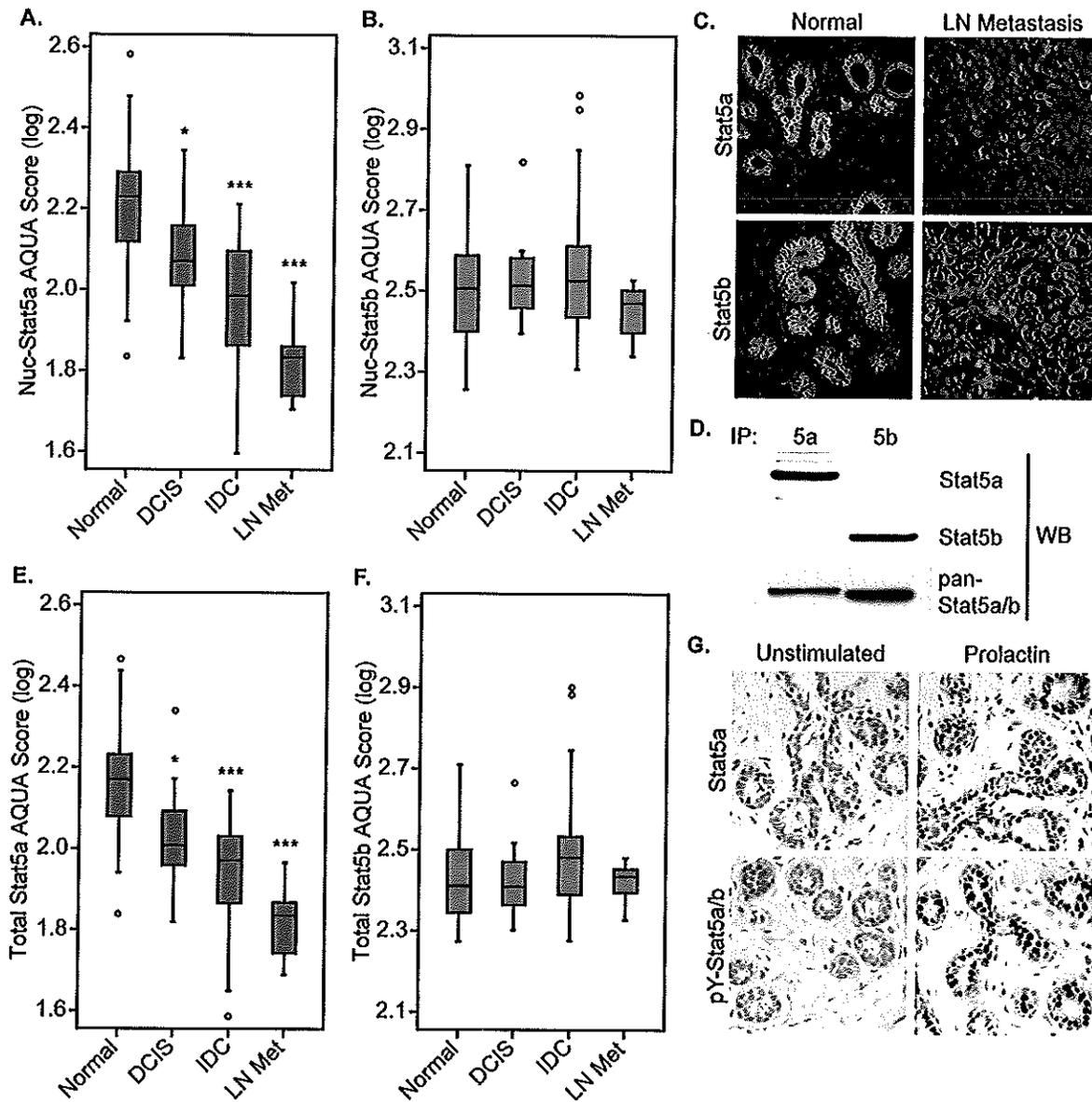


Figure 2.

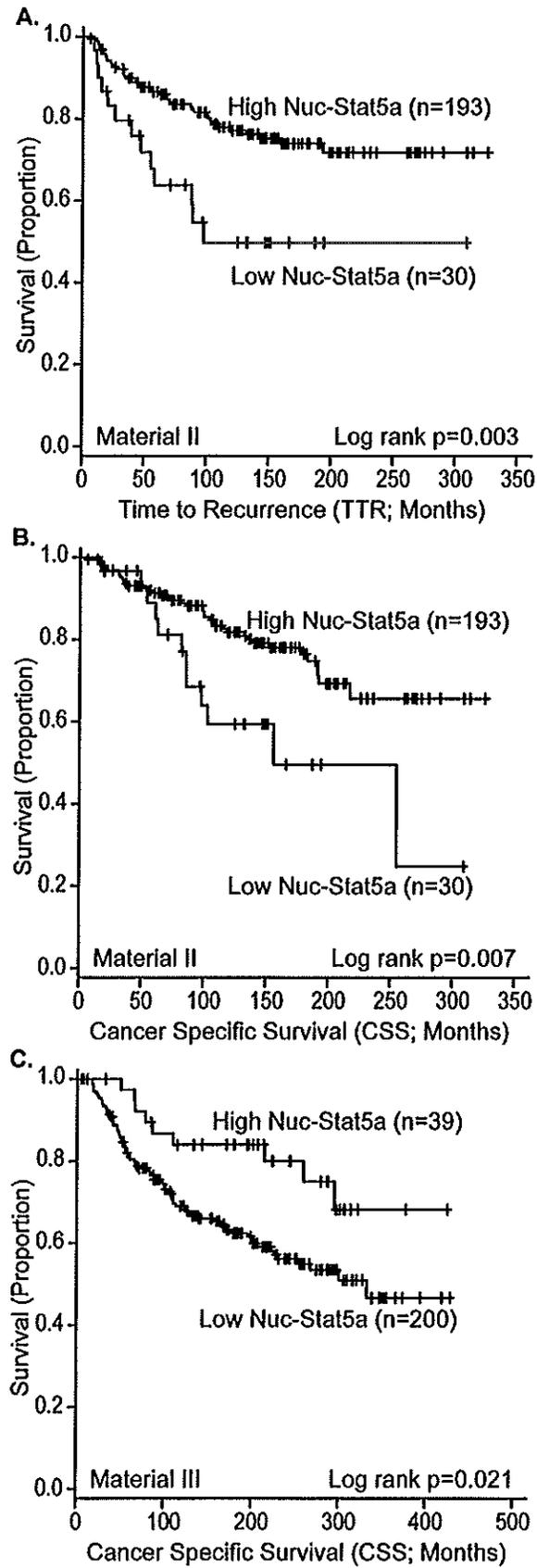
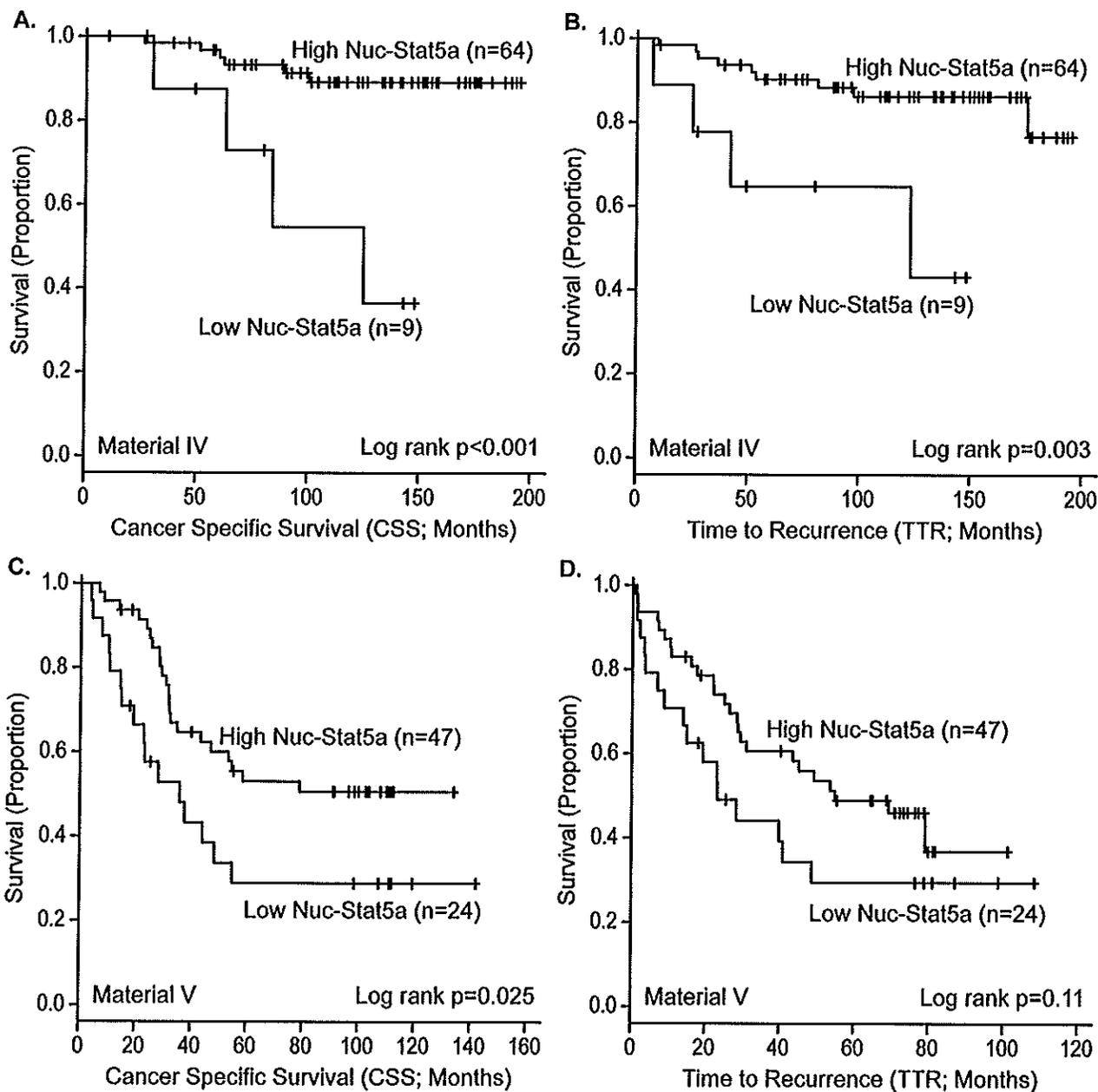
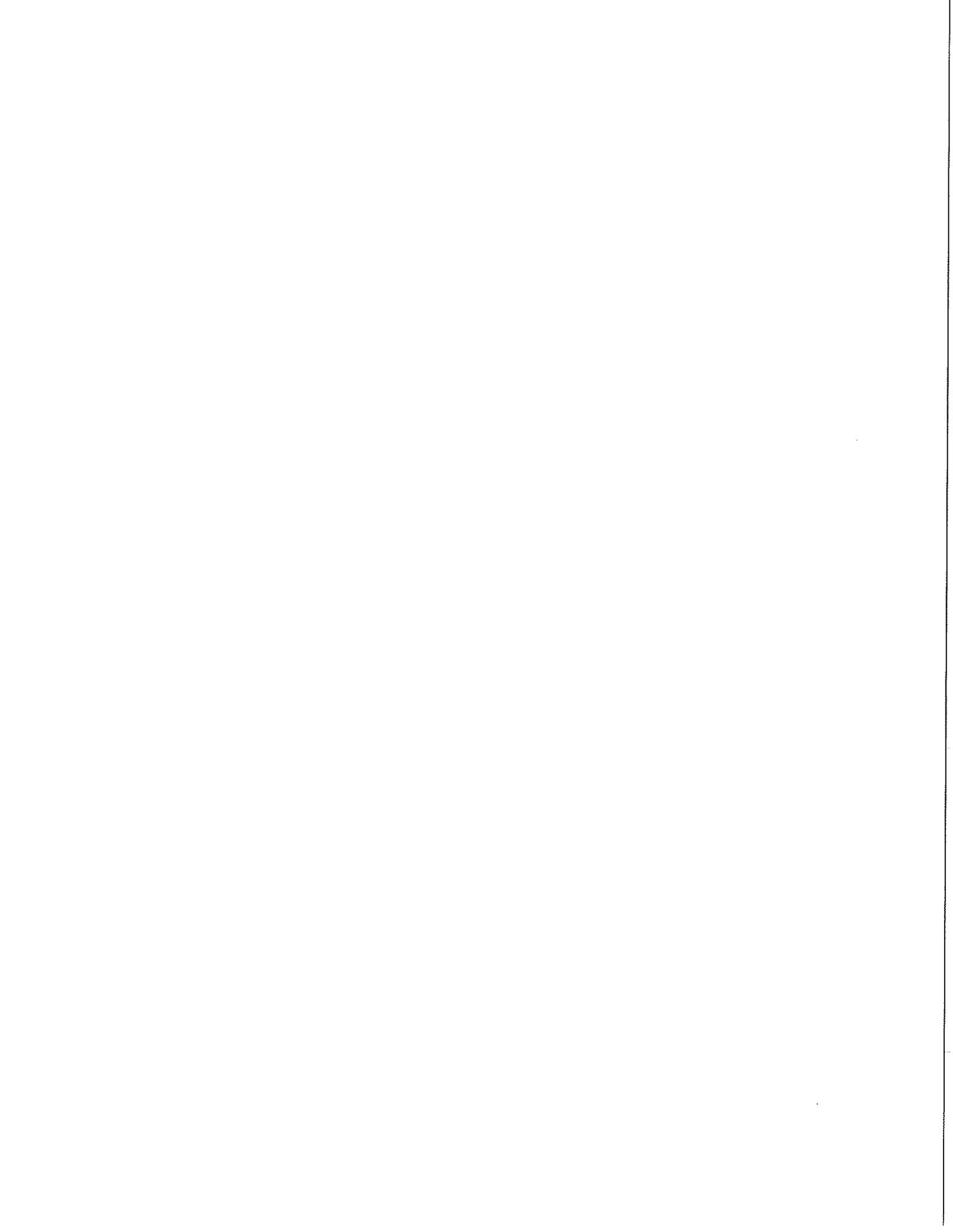


Figure 3.





**Title:**

**Matrix Metalloproteinase-1 2G Insertion Polymorphism is a Prognostic Marker for Predicting Breast Cancer Severity**

Jing Zhou<sup>1</sup>, Constance Brinckerhoff<sup>2</sup>, Susan Lubert<sup>1</sup>, Kui Yang<sup>3</sup>, Jasmine Saini <sup>1</sup>, Jeffrey Hooke<sup>4</sup>, Richard Mural<sup>1</sup>, Craig Shriver<sup>4</sup>, Stella Somiari<sup>1\*</sup>

<sup>1</sup>Clinical Breast Care Project, Windber Research Institute, 620 7th Street, Windber, PA 15963, USA

<sup>2</sup>Dartmouth-Hitchcock Medical Center, Lebanon, NH, USA

<sup>3</sup> Colorado Foundation for Medical Care, 23 Inverness Way East, Suit 100, Englewood, CO 80112-5708, USA

<sup>4</sup>Clinical Breast Care Project, Walter Reed Army Medical Center, 6900 Georgia Avenue, NW Washington, DC 20307, USA

\*Corresponding Author

Email: [s.somiari@wriwindber.org](mailto:s.somiari@wriwindber.org)

Tel: 814-4679844

Fax: 814-4676334

Journal: *Breast Cancer Research and Treatment*

## Abstract

**Purpose.** Matrix Metalloproteinase-1 (MMP-1) is a ubiquitously expressed interstitial collagenase. Over-expression of MMP-1 is associated with initiating mammary tumorigenesis by degrading stroma and releasing growth factors. A single guanine insertion in the MMP-1 promoter region creates the binding site 5'-GGAT-3', for the ETS transcription factor, which increases the expression of MMP-1. MMP-1 2G insertion polymorphism is linked to early onset, increased risk or aggressiveness of several cancers. We studied the effect of this polymorphism on MMP-1 expression and severity of breast cancer.

**Method.** The genotype of 132 patients was sequenced. This consisted of 52 invasive breast cancers, 50 ductal carcinoma in situ/lobular carcinoma in situ, 13 atypical ductal hyperplasia and 17 benign breast disease. Genotype data were analyzed together with immunohistochemical results for MMP-1, ER/PR, HER2, and P53.

**Results.** We found no significant difference in the distribution of MMP-1 genotype among the study groups. However, in the invasive breast cancer group, the 2G insertion polymorphism was significantly correlated with the expression of MMP-1 ( $P = 0.03$ ), HER2 and P53 ( $P = 0.04$  respectively). The expression of MMP-1 was significantly higher in atypical ductal hyperplasia patients compared to those with benign breast disease ( $P = 0.02$ ) and in the invasive cancer group compared to the *in situ* cancer group ( $P = 0.03$ ).

**Conclusions.** The MMP-1 2G insertion polymorphism is associated with breast cancer severity and this may have a biological association with the expression of other clinical prognostic markers. MMP-1 over-expression is associated with increased risk of breast cancer and local invasion.

**Key words:** breast cancer, gene polymorphism, matrix metalloproteinases (MMPs), MMP-1 gene, tumor progression, prognosis

## Introduction

Matrix Metalloproteinases (MMPs) are a large family of calcium-dependent zinc-containing proteinases. They can degrade components of extracellular matrix (ECM) and function in a variety of ways, which include promoting cell proliferation, migration, differentiation, apoptosis and angiogenesis [1-5]. The MMPs are involved in physical processes such as growth, wound healing and fibrosis, and also play a role in tumor invasion and metastasis in almost all human cancers [6-16]. Matrix metalloproteinase-1 (MMP-1) is the most ubiquitously expressed interstitial collagenase. It cleaves collagen I, II, and III, the major components of ECM. MMP-1 has been implicated in early detection and prognosis of breast cancer. Increased expression of MMP-1 has been reported to be significantly associated with precancerous status [17] and over-expression of MMP-1 mRNA has been correlated with high frequency of recurrence and fatal outcome in breast cancer [18].

Among a number of regulatory mechanisms that can impact the level and the activity of MMPs, transcriptional regulation plays one of the most important roles. A single guanine insertion at -1607 bp in the promoter region of MMP-1 creates a binding site (5'-GGAT-3') for transcription factor ETS and enhances the transcriptional activity of MMP-1, while the 1G allele (5'-GAT-3') does not. This ETS binding site can cooperate with the adjacent AP-1 binding site to induce high level expression of MMP-1 through enhancing the activity of the basal transcriptional machinery [19,20].

In breast cancer, association between MMP-1 2G/2G genotype and lymph node metastasis has been reported and MMP-1 1G/2G polymorphism has been proposed as a prognostic marker [21,22]. Also, the MMP-1 2G polymorphism has been linked to early-onset of lung cancer [23], increased risk of nasopharyngeal carcinoma, oral squamous cell carcinoma and colorectal cancer [24-26], aggressiveness of head and neck cancer [27,28], development and progression of endometrial carcinoma [29] and negative prognosis in patients with ovarian cancer [30].

To determine the impact of MMP-1 2G insertion polymorphism on the occurrence and progression of breast cancer, we sequenced the genotype of a total of 132 patients who either had breast cancer or breast disease to see if the MMP-1 1G/2G genotype correlates with the expression of MMP-1 and /or the disease

status of the patient. We further explored the possible relationship of this polymorphism with known clinical markers of breast cancer such as HER2 and P53.

## **Materials and Methods**

### *Subject selection*

All participants were fully consented patients recruited under the Institutional Review Board approved protocols of the Clinical Breast Care Project (CBCP). They were women, 18 years of age and above, who were attending the breast clinics of the CBCP at the Walter Reed Army Medical Center (WRAMC) in Washington DC, USA, and the Joyce Murtha Breast Care Center (JMBCC)/Windber Medical Center (WMC) in Windber, PA, USA. A total of 180 patients were recruited out of which we obtained sequencing data for 132 patients consisting of 52 invasive breast cancer [IBC], 50 ductal/lobular carcinoma in situ [DCIS/LCIS], 13 atypical ductal hyperplasia [ADH] and 17 benign breast disease. The median ages for IBC, DCIS, ADH and benign groups were 62 years (range: 36 - 86), 55 years (range: 25 - 84), 49 years (19 -74) and 45 years (range: 26 - 71) respectively. Immunohistochemistry data for MMP-1 was determined for 165 patients.

### *Genotyping*

Blood DNA was extracted using the QIAamp blood DNA commercial kit (Qiagen). Blood DNA was specifically from blood clots obtained after processing peripheral blood into serum. The DNA was then utilized for PCR as earlier described [20]. Briefly, PCR reaction was carried out in a total volume of 50  $\mu$ l with 100ng genomic DNA, 0.2mM dNTP, 2mM MgSO<sub>4</sub>, 50nM each of the two primers and 1unit of platinum Taq DNA high fidelity polymerase (Invitrogen). The primer sequences for amplifying MMP-1 were: forward primer 5'TTGCCAGATGGGACAGTGTATGAG-3'; reverse primer 5'-ACATTAAATTGTCTTGGGTACTGGT-3'. The PCR reaction started with 30 seconds incubation at 94°C followed by 10 cycles of 30 seconds incubation at 94°C, 30 seconds incubation at 52°C, and 2 minutes

incubation at 68°C; 30 cycles of 30 seconds incubation at 94°C, 30 seconds incubation at 55°C, and 1 minute 30 seconds at 68°C with 4 minutes incubation at 68°C. The amplification was verified on 1.2% agarose gel.

PCR products were cleaned up with Exonuclease I & Shrimp Alkaline Phosphatase enzymes (Affymetrics) to remove primers. Cycle sequencing reactions were performed with a nested primer (5'-AGTGTTCTTTGGTCTCTGC-3') and purified by ethanol precipitation (including 7.5M ammonium acetate). DNA pellets were then dried, resuspended in MegaBACE Loading Solution and sequenced using the MegaBACE 1000 (GE Healthcare). Data was analyzed with Sequence Analysis v4.0 software (GE Healthcare).

### *Immunohistochemistry (IHC)*

Formalin-fixed paraffin embedded sections of the selected cases were retrieved from the tissue repository to determine expression level of MMP-1 by IHC. The pathological diagnosis of each case was confirmed by the CBCP pathologist using a freshly prepared Hematoxylin and Eosin slide. For all confirmed cases, seven slides of 4 microns sections were cut and subjected to IHC analysis for MMP-1. Antibody optimization was performed to establish the required optimal conditions. For the determination of MMP-1 expression, we utilized a polyclonal rabbit antihuman antibody (LabVision) at 1:80 dilution. All IHC was performed at MDR Global Inc., Windber, PA, using standardized procedures and protocols. Briefly, each section was placed on positively charged slides which were then dried at 59°C, before placing in SubX solution prior to deparaffinization and antigen recovery using Dako PT Module. The non-enzymatic antigen recovery step was performed in a patented microwave oven and the slides were washed thoroughly in distilled water before placing in TBS buffer. Immunostaining was performed on the Dako Autostainer with the application of primary antibody followed by washing the slides again with TBS buffer. The appropriate detection reagent was then dispensed onto the slides followed by another wash in TBS buffer and Diaminobenzidine "DAB"/hydrogen peroxide for color visualization. All the slides were counterstained with hematoxylin. Appropriately selected tissue specimens were used as positive controls for each primary antibody. IHC data for MMP-1 was scored as either 0 (no expression) or increasing expression of 1+, 2+ or 3+ (Figure1).

Immunohistochemistry data for ER, PR, HER2, and P53 were queried from the clinical database of the CBCP and subjected to analyses with data on MMP-1 genotyping and expression.

### *Statistical Analysis*

Statistical analyses were carried out using SAS 9.1 for Windows (SAS institute, Cary, NC). The association between MMP-1 expression/polymorphism and disease status was determined using Chi-square or Fisher's exact test when the numbers of the cases were small. The Jonckheere-Terpstra test for trend was used to assess the correlation between MMP-1 genotype and expression level. The Jonckheere-Terpstra test is a nonparametric statistical test, which examines the ordered differences of MMP-1 expression level among polymorphism classes. Odds Ratios (ORs) and 95% Confidence Intervals (CIs) were calculated by logistic regression. In all cases,  $P < 0.05$  was required for significance.

### **Results**

The distribution of 1G/2G genotypes in the 132 selected patients was according to the Hardy-Weinberg equilibrium [20]. In the ADH group, 91% (10 out of 11) of the patients showed high expression of MMP-1, while in the benign group, only 48% (12 out of 25) expressed MMP-1 at a high level. The expression of MMP-1 was significantly associated with the disease status ( $P = 0.02$ ; Table 1). Compared to subjects expressing low levels (0/1+) of MMP-1, the odds for subjects expressing high levels (2+/3+) of MMP-1 to develop ADH increased 10.83 times (OR = 10.83; 95% CI, 1.2-97.8; Table 1). When the expression of MMP-1 in IBC was compared with *in situ* breast cancer (DCIS/LCIS), we found that 58% of the IBC patients had high MMP-1 expression, which was significantly higher than that of the DCIS/LCIS group ( $\chi^2 = 4.79$ , Df = 1,  $P = 0.03$ ; Table 2). Patients expressing high levels of MMP-1 had a 2.20 fold increased risk of developing invasive breast cancer (OR = 2.20; 95% CI, 1.09-4.44; Table 2).

There was no significant difference in the distribution of the MMP-1 1G/2G genotypes between the study groups. Differences observed in allele frequencies between the groups were not statistically significant ( $P >$

0.05) (Data not shown). However, within the IBC group, we observed a significant association between the expression of MMP-1 and the MMP-1 genotypes (Table 3;  $P = 0.03$ ). In this group, 57% of the patients expressing low MMP-1 had the 1G/1G homozygote genotype while 80% of those expressing high MMP-1 (3+) were either 1G/2G carriers or homozygous for 2G as displayed graphically in Figure 2a. Figure 2b shows the direct correlation between increased MMP-1 expression and the 2G allele frequency. This association between the MMP-1 2G insertion polymorphism and MMP-1 expression was not observed in the other study groups. Finally, we analyzed the data to determine if the MMP-1 2G insertion polymorphism had any association with other known breast cancer prognostic markers such as ER/PR, HER2 and P53 in the IBC group. We observed a significant difference in the distribution of the MMP-1 genotypes between the HER2+ and HER2- patients (Table 4). Notably, the percentage of 1G/1G homozygote was similar in these two groups (Table 4); but the percentage of 1G/2G heterozygotes significantly decreased and the percentage of the 2G/2G homozygotes significantly increased in HER2+ group compared to the HER2- group ( $P = 0.04$ ; Table 4). Compared to the 1G/2G carriers, patients carrying the 2G/2G genotype have a 7.5-fold (95%CI, 1.48-37.9;  $P = 0.03$ ; Table 4) increased odds of expressing HER2; however, the odds for 1G/1G carriers to express HER2 was only increased 2.78 fold, which was not statistically significant ( $P = 0.29$ ). We also observed significant association between MMP-1 genotype distribution and the expression of P53 in IBC patients (Table 5). Compared with 1G/2G carriers, the odds for the 2G/2G carriers to express P53 increased 5.88 fold (95%CI, 5.26-33.3;  $P = 0.049$ ) while the odds for the 1G/1G carriers to express P53 increased 5.26 fold (95%CI, 1.18-25;  $P = 0.038$ ; Table 5). There was no association between the polymorphism and the ER/PR status of the IBC patients ( $P > 0.05$ ) (Data not shown).

## Discussion

We determined the expression of MMP-1 in patients with benign breast disease, ADH (a precancerous condition which poses an increased risk of breast cancer development), *in situ* and invasive breast cancers. We observed significantly increased expression of MMP-1 in ADH patients compared to benign breast disease ( $P = 0.02$ ) and a dramatic increase in the expression of MMP-1 in IBC compared with DCIS/LCIS ( $P = 0.03$ ). Interestingly, we also found that the expression of MMP-1 in the *in situ* foci of IBC tumors are also

significantly higher than that of pure DCIS specimens ( $P = 0.035$ , data not shown), but not significantly different from that of the invasive foci of IBC specimens ( $P = 0.595$ , data not shown). These observations indicate an association between MMP-1 expression and disease status in breast cancer and a role for MMP-1 in the development and progression of breast cancer. Our results are consistent with earlier studies that support (i) the role of MMP-1 as a candidate precancerous marker for early detection of breast cancer [17] and (ii) the notion that over-expression of MMP-1 is important for the invasion and progression of breast cancer [18,19,21].

The genotype distribution of the MMP-1 between the different patient groups, was not significantly different ( $P > 0.05$ ), suggesting that the MMP-1 polymorphism may not contribute to the development of breast cancer. However, our study shows an association between the 2G genotype and MMP-1 expression in advanced disease as observed in the IBC patient group ( $P = 0.03$ ; Table 3). The 2G allele frequency for invasive breast cancer patients expressing low MMP-1 was 0.29 (for 1+) and 0.44 (for 2+); while IBC patients expressing high MMP-1 (3+) had a 2G allele frequency of 0.6 (Figure 2b). Thus in the IBC patients, the expression of MMP-1 was positively correlated with the 2G allele. However, we did not find association of MMP-1 genotypes and MMP-1 expression in all the other study groups. While the 2G polymorphism may contribute to increased expression of MMP-1 in IBC, the observation in ADH indicates that other factors may contribute to increased expression of MMP-1 in this subgroup of breast cancer patients. As reported, MMP-1 expression can be increased by many stimuli including growth factors, inflammatory interleukins, heat and UV-light, which can stimulate mitogen-activated protein kinases (MAPK) and activate MMP-1 transcription [31-34].

Higher 2G allele frequencies have been reported in breast cancer patients with lymph node-metastasis compared with patients without metastasis, which highlights this polymorphism as a prognostic marker for breast cancer [21,22,35]. Consistently, we found significant association between the MMP-1 2G polymorphism and HER2 expression in IBC patients (Table 4). Forty four percent (10 out of 23) of patients expressing HER2, had the homozygous 2G/2G genotype compared to 15% of the patients (4 out of 26) not expressing HER2 (Table 4). The distribution curve of the genotype shifted to a higher frequency of 2G homozygotes in HER2+ group. In breast cancer, the amplification of the Her2/neu oncogene is a well established prognostic marker which is associated with relapse and short survival [36,37]; HER2 over-

expression has also been correlated with poor outcome [38-40]. The association of the MMP-1 2G insertion polymorphism with this clinical biomarker further indicates its role in breast cancer prognosis. Studies have shown that both MMP-1 2G polymorphism and HER2 over-expression can up-regulate MMP-1 expression via different molecular mechanisms [19,41]. The 2G single nucleotide polymorphism of MMP-1 affects the basal transcription of MMP-1 through regulating ETS-1 and AP-1[19,20], while HER2 can further increase MMP-1 expression through activating ER81, another member of ETS family [41]. We can therefore speculate a worse prognosis for IBC patients who carry the MMP-1 2G/2G genotype and also over-express HER2, in which the expression of MMP-1 would be dramatically increased leading to lymph vascular metastasis and poor overall survival [18,42]. Although in the current study we did not observe a significant relationship between HER2 and MMP-1 over expression in the IBC group (data not shown), this could be due to the effect of unknown mechanisms, for example, a negative feedback effect from increased MMP-1 expression on HER2 transcription. We did, however, observe a correlation between the 2G insertion polymorphism and MMP-1 expression in the IBC patients only ( $P = 0.03$ ) and not in the other study categories, further supporting the notion that MMP-1 2G polymorphism is associated with disease severity and poor prognosis in breast cancer. We did not specify the localization of the MMP-1 expression (stromal versus tumor cell) and are aware that this could have implications in breast cancer prognosis [43].

Over expression of P53 is also associated with breast cancer progression and is reported as yet another clinical prognostic marker for breast cancer. Both P53 mutation and accumulation have been associated with poor prognosis [44-47]. Thor et al found significant correlation of P53 accumulation with short metastasis-free and overall survival in both sporadic and familial breast cancers [47]. They also found association between accumulation of P53 and P53 mutation [47]. Wild type P53 can down regulate MMP-1 expression by disrupting the communication between AP-1 and basal transcriptional complex in human foreskin fibroblast cells [48,49]. However, the repression effect of P53 on MMP-1 expression is lost in osteogenic sarcoma cells over expressing mutant P53 [48]. It would be interesting to study the effect of accumulated or mutated P53 on MMP-1 expression in breast cancer. So far, we did not observe an association between MMP-1 and P53 expression in our study, but we observed a significant difference in the distribution of MMP-1 genotypes between P53+ and P53- patients ( $P = 0.04$ ; Table 4). There was a significant decrease in the number of heterozygotes (1G/2G) and increased number of homozygotes (1G/1G and 2G/2G) in the

P53+ patients compared to the P53- patients ( $P = 0.02$ ). The overall effect of this difference on MMP-1 expression in P53+/- patients may need further investigation. Although the observed pattern of MMP-1 genotype distribution in P53 expressing tumors varied from that observed in the HER2 expressing tumors, these observations point to some level/degree of connectivity between the MMP-1 polymorphism and other factors associated with breast cancer severity and poor prognosis.

In summary, our results reveal a contribution of the MMP-1 -1607nt single nucleotide polymorphism to breast cancer severity as observed in its association with MMP-1 and HER2 over expression in IBC (Table 3 and 4). Our study also shows that MMP-1 over expression is associated with increased risk (Table 1) and local invasion (Table 2) of breast cancer. Based on our results and the multistep model of breast carcinogenesis, the transition from normal epithelium to invasive breast cancer via non-atypical hyperplasia, atypical hyperplasia and *in situ* carcinoma [50-53], we postulate a model for the impact of MMP-1 2G insertion polymorphism on breast cancer (Figure 3).

The MMP-1 2G polymorphism and over expression of MMP-1 have potential as prognostic markers in breast cancer. More comprehensive studies are required to fully understand the mechanism of the interaction of this polymorphism with other prognostic markers, and to explore the impact of this polymorphism alone or with other prognostic markers on disease free or overall survival in breast cancer. Such studies will elucidate the full potential of MMP-1 2G polymorphism as an independent prognostic marker in breast cancer.

## Conclusion

To study the impact of MMP-1 1G/2G polymorphism on breast cancer development, we genotyped 132 patients with benign breast disease, ADH, DCIS/LCIS and IBC and analyzed our results with expression data of the following breast cancer prognostic markers, MMP-1, ER, PR, HER2 and P53. We found that over expression of MMP-1 is associated with breast cancer onset and disease severity. We report here, for the first time, that MMP-1 2G insertion polymorphism is significantly associated with HER2 and P53 expression. HER2 over-expression, P53 accumulation, or the coexistence of these two, are known molecular markers in breast cancer prognosis [36,38,40,47,54]. The association of MMP-1 2G insertion polymorphism with these

two biomarkers in invasive breast cancer supports the role of MMP-1 2G polymorphism in breast cancer prognosis and may provide new molecular tools for sub classification of invasive breast cancer.

### **Acknowledgement**

This research was supported by a grant from the United States Department of Defense (Military Molecular Medicine Initiative MDA W81XWHH-05-2-0075, Protocol 01-20006), and by NIH-AR-26599 and NIH-CA-77267 to CB. The opinion and assertions contained herein are the private views of the authors and are not to be construed as official or as representing the views of the Department of the Army or the Department of Defense.

## Figure Legends

**Fig.1** Representative immunohistochemical staining of MMP-1 in a patient with invasive breast cancer (ID #100002283). The figure shows the image magnified at 20X and 40X.

**Fig.2a** MMP-1 2G insertion polymorphism is correlated with MMP-1 expression in IBC.

Subjects with IBC were grouped according to the expression level of MMP-1(1+, 2+ and 3+). The percentage of homozygotes (1G/1G and 2G/2G) and heterozygotes (1G/2G) were then plotted against MMP-1 expression level based on the Immunohistochemical score of 1+, 2+ and 3+. As MMP-1 expression increased, the percentage of 1G/1G homozygotes decreased, while the percentage of 2G carriers (1G/2G heterozygotes) and homozygotes (2G/2G) increased. Results show a relationship between 2G insertion and MMP-1 expression in IBC patients.

**Fig.2b** The MMP-1 2G allele frequency correlates with MMP-1 expression level.

Allele frequency of subjects in the IBC group were plotted against the expression level of MMP-1(1+, 2+ and 3+) as obtained from immunohistochemistry readings. The 2G allele frequency increased with increasing expression of MMP-1.

**Fig.3** A model of the impact of MMP-1 2G insertion polymorphism on breast cancer.

Over- expression of MMP-1 contributes to both increased risk and local invasion in breast cancer. While the MMP-1 2G insertion polymorphism may not contribute to either event, other regulators must be involved in up-regulating MMP-1 levels in the different steps leading to breast cancer onset and development. However, the MMP-1 2G polymorphism is significantly associated with disease severity and this is supported by its correlation with MMP-1 expression and the expression status of HER2 and P53 in IBC patients. Over-expression of MMP-1 induced by this polymorphism together with other regulators predicts an adverse

outcome of breast cancer. Thus, MMP-1 2G insertion polymorphism is a potential marker for breast cancer prognosis.

## References

1. Menon B, Singh M, Singh K (2005) Matrix metalloproteinases mediate beta-adrenergic receptor-stimulated apoptosis in adult rat ventricular myocytes. *Am J Physiol Cell Physiol* 289 (1):C168-176. doi:00606.2004 [pii]  
10.1152/ajpcell.00606.2004
2. McCawley LJ, Wright J, LaFleur BJ, Crawford HC, Matrisian LM (2008) Keratinocyte expression of MMP3 enhances differentiation and prevents tumor establishment. *Am J Pathol* 173 (5):1528-1539. doi:ajpath.2008.080132 [pii]  
10.2353/ajpath.2008.080132
3. Wang W, Pan H, Murray K, Jefferson BS, Li Y (2009) Matrix metalloproteinase-1 promotes muscle cell migration and differentiation. *Am J Pathol* 174 (2):541-549. doi:ajpath.2009.080509 [pii]  
10.2353/ajpath.2009.080509
4. Zeng ZZ, Yao H, Staszewski ED, Rockwood KF, Markwart SM, Fay KS, Spalding AC, Livant DL (2009) alpha(5)beta(1) Integrin Ligand PHSRN Induces Invasion and alpha(5) mRNA in Endothelial Cells to Stimulate Angiogenesis. *Transl Oncol* 2 (1):8-20
5. Shan B, Morris CA, Zhuo Y, Shelby BD, Levy DR, Lasky JA (2007) Activation of proMMP-2 and Src by HHV8 vGPCR in human pulmonary arterial endothelial cells. *J Mol Cell Cardiol* 42 (3):517-525. doi:S0022-2828(06)00763-2 [pii]  
10.1016/j.yjmcc.2006.08.004
6. Gonzalez Rodriguez I, Rivas del Fresno M, Gil Ugarteburu R, Gonzalez Arriaga P, Lopez Cima F, Fernandez Samoano A, Muruamendiaraz Fernandez V, Fernandez Garcia I, Perez-Carral JR, Tardon A (2010) Expression of matrix metalloproteinase-9 in prostate cancer. Preliminary experience. *Arch Esp Urol* 63 (2):119-124
7. Szarvas T, Becker M, vom Dorp F, Gethmann C, Totsch M, Bankfalvi A, Schmid KW, Romics I, Rubben H, Ergun S (2010) Matrix metalloproteinase-7 as a marker of metastasis and predictor of poor survival in bladder cancer. *Cancer Sci* 101 (5):1300-1308. doi:CAS1506 [pii]

10.1111/j.1349-7006.2010.01506.x

8. Yi YC, Chou PT, Chen LY, Kuo WH, Ho ES, Han CP, Yang SF (2010) Matrix metalloproteinase-7 (MMP-7) polymorphism is a risk factor for endometrial cancer susceptibility. *Clin Chem Lab Med* 48 (3):337-344. doi:10.1515/CCLM.2010.082
9. Kim YW, Ko YT, Kim NK, Chung HC, Min BS, Lee KY, Park JP, Kim H (2010) A comparative study of protein expression in primary colorectal cancer and synchronous hepatic metastases: the significance of matrix metalloproteinase-1 expression as a predictor of liver metastasis. *Scand J Gastroenterol* 45 (2):217-225. doi:10.3109/00365520903453158
10. Garg P, Sarma D, Jeppsson S, Patel NR, Gewirtz AT, Merlin D, Sitaraman SV (2010) Matrix metalloproteinase-9 functions as a tumor suppressor in colitis-associated cancer. *Cancer Res* 70 (2):792-801. doi:0008-5472.CAN-09-3166 [pii]  
10.1158/0008-5472.CAN-09-3166
11. Li M, Xiao T, Zhang Y, Feng L, Lin D, Liu Y, Mao Y, Guo S, Han N, Di X, Zhang K, Cheng S, Gao Y (2010) Prognostic significance of matrix metalloproteinase-1 levels in peripheral plasma and tumour tissues of lung cancer patients. *Lung Cancer* 69 (3):341-347. doi:S0169-5002(09)00673-4 [pii]  
10.1016/j.lungcan.2009.12.007
12. Luukkaa H, Klemi P, Hirsimaki P, Vahlberg T, Kivisaari A, Kahari VM, Grenman R (2010) Matrix metalloproteinase (MMP)-7 in salivary gland cancer. *Acta Oncol* 49 (1):85-90. doi:10.3109/02841860903287197
13. Luukkaa H, Klemi P, Leivo I, Makitie AA, Irish J, Gilbert R, Perez-Ordóñez B, Hirsimaki P, Vahlberg T, Kivisaari A, Kahari VM, Grenman R (2010) Expression of matrix metalloproteinase-1, -7, -9, -13, Ki-67, and HER-2 in epithelial-myoepithelial salivary gland cancer. *Head Neck* 32 (8):1019-1027. doi:10.1002/hed.21277
14. Yeh YC, Sheu BS, Cheng HC, Wang YL, Yang HB, Wu JJ (2010) Elevated serum matrix metalloproteinase-3 and -7 in *H. pylori*-related gastric cancer can be biomarkers correlating with a poor survival. *Dig Dis Sci* 55 (6):1649-1657. doi:10.1007/s10620-009-0926-x

15. Scherf DB, Dally H, Muller P, Werle-Schneider G, Jager B, Edler L, Tuengerthal S, Fischer JR, Drings P, Bartsch H, Risch A (2010) Single nucleotide polymorphisms in matrix metalloproteinase genes and lung cancer chemotherapy response and prognosis. *Eur Respir J* 35 (2):381-390. doi:09031936.00125608 [pii] 10.1183/09031936.00125608
16. Zohny SF, Fayed ST (2009) Clinical utility of circulating matrix metalloproteinase-7 (MMP-7), CC chemokine ligand 18 (CCL18) and CC chemokine ligand 11 (CCL11) as markers for diagnosis of epithelial ovarian cancer. *Med Oncol*. doi:10.1007/s12032-009-9366-x
17. Poola I, DeWitty RL, Marshalleck JJ, Bhatnagar R, Abraham J, Leffall LD (2005) Identification of MMP-1 as a putative breast cancer predictive marker by global gene expression analysis. *Nat Med* 11 (5):481-483. doi:nm1243 [pii] 10.1038/nm1243
18. Cheng S, Tada M, Hida Y, Asano T, Kuramae T, Takemoto N, Hamada J, Miyamoto M, Hirano S, Kondo S, Moriuchi T (2008) High MMP-1 mRNA expression is a risk factor for disease-free and overall survivals in patients with invasive breast carcinoma. *J Surg Res* 146 (1):104-109. doi:S0022-4804(07)00382-4 [pii] 10.1016/j.jss.2007.05.032
19. Tower GB, Coon CI, Brinckerhoff CE (2003) The 2G single nucleotide polymorphism (SNP) in the MMP-1 promoter contributes to high levels of MMP-1 transcription in MCF-7/ADR breast cancer cells. *Breast Cancer Res Treat* 82 (2):75-82. doi:10.1023/B:BREA.0000003948.14026.7c
20. Rutter JL, Mitchell TI, Buttice G, Meyers J, Gusella JF, Ozelius LJ, Brinckerhoff CE (1998) A single nucleotide polymorphism in the matrix metalloproteinase-1 promoter creates an Ets binding site and augments transcription. *Cancer Res* 58 (23):5321-5325
21. Przybyłowska K, Zielinska J, Zadrozny M, Krawczyk T, Kulig A, Wozniak P, Rykała J, Kolacinska A, Morawiec Z, Drzewoski J, Błasiak J (2004) An association between the matrix metalloproteinase 1 promoter gene polymorphism and lymphnode metastasis in breast cancer. *J Exp Clin Cancer Res* 23 (1):121-125

22. Hughes S, Agbaje O, Bowen RL, Holliday DL, Shaw JA, Duffy S, Jones JL (2007) Matrix metalloproteinase single-nucleotide polymorphisms and haplotypes predict breast cancer progression. *Clin Cancer Res* 13 (22 Pt 1):6673-6680. doi:13/22/6673 [pii]  
10.1158/1078-0432.CCR-07-0884
23. McColgan P, Sharma P (2009) Polymorphisms of matrix metalloproteinases 1, 2, 3 and 9 and susceptibility to lung, breast and colorectal cancer in over 30,000 subjects. *Int J Cancer* 125 (6):1473-1478. doi:10.1002/ijc.24441
24. Kondo S, Wakisaka N, Schell MJ, Horikawa T, Sheen TS, Sato H, Furukawa M, Pagano JS, Yoshizaki T (2005) Epstein-Barr virus latent membrane protein 1 induces the matrix metalloproteinase-1 promoter via an Ets binding site formed by a single nucleotide polymorphism: enhanced susceptibility to nasopharyngeal carcinoma. *Int J Cancer* 115 (3):368-376. doi:10.1002/ijc.20849
25. Ghilardi G, Biondi ML, Mangoni J, Leviti S, DeMonti M, Guagnellini E, Scorza R (2001) Matrix metalloproteinase-1 promoter polymorphism 1G/2G is correlated with colorectal cancer invasiveness. *Clin Cancer Res* 7 (8):2344-2346
26. Vairaktaris E, Yapijakis C, Derka S, Serefoglou Z, Vassiliou S, Nkenke E, Ragos V, Vylliotis A, Spyridonidou S, Tsigris C, Yannopoulos A, Tesseromatis C, Neukam FW, Patsouris E (2007) Association of matrix metalloproteinase-1 (-1607 1G/2G) polymorphism with increased risk for oral squamous cell carcinoma. *Anticancer Res* 27 (1A):459-464
27. P Oc, Leksrisakul P, Sangruchi S (2006) A functional polymorphism in the matrix metalloproteinase-1 gene promoter is associated with susceptibility and aggressiveness of head and neck cancer. *Int J Cancer* 118 (10):2548-2553. doi:10.1002/ijc.21644
28. Hashimoto T, Uchida K, Okayama N, Imate Y, Suehiro Y, Hamanaka Y, Ueyama Y, Shinozaki F, Yamashita H, Hinoda Y (2004) Association of matrix metalloproteinase (MMP)-1 promoter polymorphism with head and neck squamous cell carcinoma. *Cancer Lett* 211 (1):19-24. doi:10.1016/j.canlet.2004.01.032 S0304383504001351 [pii]
29. Nishioka Y, Kobayashi K, Sagae S, Ishioka S, Nishikawa A, Matsushima M, Kanamori Y, Minaguchi T, Nakamura Y, Tokino T, Kudo R (2000) A single nucleotide polymorphism in the matrix metalloproteinase-1 promoter in endometrial carcinomas. *Jpn J Cancer Res* 91 (6):612-615

30. Six L, Grimm C, Leodolter S, Tempfer C, Zeillinger R, Sliutz G, Speiser P, Reinthaller A, Hefler LA (2006) A polymorphism in the matrix metalloproteinase-1 gene promoter is associated with the prognosis of patients with ovarian cancer. *Gynecol Oncol* 100 (3):506-510. doi:S0090-8258(05)00759-6 [pii] 10.1016/j.ygyno.2005.08.049
31. Johansson N, Ala-aho R, Uitto V, Grenman R, Fusenig NE, Lopez-Otin C, Kahari VM (2000) Expression of collagenase-3 (MMP-13) and collagenase-1 (MMP-1) by transformed keratinocytes is dependent on the activity of p38 mitogen-activated protein kinase. *J Cell Sci* 113 Pt 2:227-235
32. Reunanen N, Westermarck J, Hakkinen L, Holmstrom TH, Elo I, Eriksson JE, Kahari VM (1998) Enhancement of fibroblast collagenase (matrix metalloproteinase-1) gene expression by ceramide is mediated by extracellular signal-regulated and stress-activated protein kinase pathways. *J Biol Chem* 273 (9):5137-5145
33. Park CH, Lee MJ, Ahn J, Kim S, Kim HH, Kim KH, Eun HC, Chung JH (2004) Heat shock-induced matrix metalloproteinase (MMP)-1 and MMP-3 are mediated through ERK and JNK activation and via an autocrine interleukin-6 loop. *J Invest Dermatol* 123 (6):1012-1019. doi:JID23487 [pii] 10.1111/j.0022-202X.2004.23487.x
34. Di Girolamo N, Coroneo MT, Wakefield D (2003) UVB-elicited induction of MMP-1 expression in human ocular surface epithelial cells is mediated through the ERK1/2 MAPK-dependent pathway. *Invest Ophthalmol Vis Sci* 44 (11):4705-4714
35. Przybylowska K, Kluczna A, Zadrozny M, Krawczyk T, Kulig A, Rykala J, Kolacinska A, Morawiec Z, Drzewoski J, Blasiak J (2006) Polymorphisms of the promoter regions of matrix metalloproteinases genes MMP-1 and MMP-9 in breast cancer. *Breast Cancer Res Treat* 95 (1):65-72. doi:10.1007/s10549-005-9042-6
36. Slamon DJ, Clark GM, Wong SG, Levin WJ, Ullrich A, McGuire WL (1987) Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science* 235 (4785):177-182
37. Ross JS, Fletcher JA, Linette GP, Stec J, Clark E, Ayers M, Symmans WF, Puzstai L, Bloom KJ (2003) The Her-2/neu gene and protein in breast cancer 2003: biomarker and target of therapy. *Oncologist* 8 (4):307-325

38. Mitchell MS, Press MF (1999) The role of immunohistochemistry and fluorescence in situ hybridization for HER2/neu in assessing the prognosis of breast cancer. *Semin Oncol* 26 (4 Suppl 12):108-116
39. Chen XS, Ma CD, Wu JY, Yang WT, Lu HF, Wu J, Lu JS, Shao ZM, Shen ZZ, Shen KW (2010) Molecular subtype approximated by quantitative estrogen receptor, progesterone receptor and Her2 can predict the prognosis of breast cancer. *Tumori* 96 (1):103-110
40. Choi YH, Ahn JH, Kim SB, Jung KH, Gong GY, Kim MJ, Son BH, Ahn SH, Kim WK (2009) Tissue microarray-based study of patients with lymph node-negative breast cancer shows that HER2/neu overexpression is an important predictive marker of poor prognosis. *Ann Oncol* 20 (8):1337-1343. doi:mdp003 [pii]  
10.1093/annonc/mdp003
41. Bosc DG, Goueli BS, Janknecht R (2001) HER2/Neu-mediated activation of the ETS transcription factor ER81 and its target gene MMP-1. *Oncogene* 20 (43):6215-6224. doi:10.1038/sj.onc.1204820
42. Baker EA, Stephenson TJ, Reed MW, Brown NJ (2002) Expression of proteinases and inhibitors in human breast cancer progression and survival. *Mol Pathol* 55 (5):300-304
43. Eck SM, Cote AL, Winkelman WD, Brinckerhoff CE (2009) CXCR4 and matrix metalloproteinase-1 are elevated in breast carcinoma-associated fibroblasts and in normal mammary fibroblasts exposed to factors secreted by breast cancer cells. *Mol Cancer Res* 7 (7):1033-1044. doi:1541-7786.MCR-09-0015 [pii]  
10.1158/1541-7786.MCR-09-0015
44. Yang CH, Cristofanilli M (2006) The role of p53 mutations as a prognostic factor and therapeutic target in inflammatory breast cancer. *Future Oncol* 2 (2):247-255. doi:10.2217/14796694.2.2.247
45. Colozza M, Cardoso F, Sotiriou C, Larsimont D, Piccart MJ (2005) Bringing molecular prognosis and prediction to the clinic. *Clin Breast Cancer* 6 (1):61-76
46. Coradini D, Daidone MG (2004) Biomolecular prognostic factors in breast cancer. *Curr Opin Obstet Gynecol* 16 (1):49-55. doi:00001703-200402000-00010 [pii]

47. Thor AD, Moore DH, II, Edgerton SM, Kawasaki ES, Reihnsaus E, Lynch HT, Marcus JN, Schwartz L, Chen LC, Mayall BH, et al. (1992) Accumulation of p53 tumor suppressor gene protein: an independent marker of prognosis in breast cancers. *J Natl Cancer Inst* 84 (11):845-855
48. Sun Y, Wenger L, Rutter JL, Brinckerhoff CE, Cheung HS (1999) p53 down-regulates human matrix metalloproteinase-1 (Collagenase-1) gene expression. *J Biol Chem* 274 (17):11535-11540
49. Sun Y, Zeng XR, Wenger L, Firestein GS, Cheung HS (2004) P53 down-regulates matrix metalloproteinase-1 by targeting the communications between AP-1 and the basal transcription complex. *J Cell Biochem* 92 (2):258-269. doi:10.1002/jcb.20044
50. Reis-Filho JS, Lakhani SR (2003) The diagnosis and management of pre-invasive breast disease: genetic alterations in pre-invasive lesions. *Breast Cancer Res* 5 (6):313-319. doi:10.1186/bcr650  
bcr650 [pii]
51. Buerger H, Otterbach F, Simon R, Poremba C, Diallo R, Decker T, Riethdorf L, Brinkschmidt C, Dockhorn-Dworniczak B, Boecker W (1999) Comparative genomic hybridization of ductal carcinoma in situ of the breast-evidence of multiple genetic pathways. *J Pathol* 187 (4):396-402. doi:10.1002/(SICI)1096-9896(199903)187:4<396::AID-PATH286>3.0.CO;2-L [pii]  
10.1002/(SICI)1096-9896(199903)187:4<396::AID-PATH286>3.0.CO;2-L
52. Dupont WD, Page DL (1985) Risk factors for breast cancer in women with proliferative breast disease. *N Engl J Med* 312 (3):146-151. doi:10.1056/NEJM198501173120303
53. Allred DC, Mohsin SK, Fuqua SA (2001) Histological and biological evolution of human premalignant breast disease. *Endocr Relat Cancer* 8 (1):47-61
54. Yamashita H, Nishio M, Toyama T, Sugiura H, Zhang Z, Kobayashi S, Iwase H (2004) Coexistence of HER2 over-expression and p53 protein accumulation is a strong prognostic molecular marker in breast cancer. *Breast Cancer Res* 6 (1):R24-30. doi:10.1186/bcr738  
bcr738 [pii]

*Table 1. The expression of MMP-1 in subjects with benign breast disease versus atypical hyperplasia*

MMP1 expression	Benign (n=25) n (%)	ADH (n=11) n (%)	OR	95%CI	P*
Low (0/1+)	13 (52)	1 (9)	1		
High (2+/3+)	12 (48)	10 (91)	10.83	1.20-97.80	0.02

\* Fisher's exact test, 2- tailed  
ADH-atypical ductal hyperplasia

Table 2. The expression of MMP-1 in subjects with *in situ* versus invasive breast cancer

MMP1 expression	DCIS/LCIS (n=64) n (%)	IBC (n=65) n (%)	OR	95%CI	P*
Low (0/1+)	39 (61)	27 (42)	1		
High (2+/3+)	25 (39)	38 (58)	2.20	1.09-4.44	0.03

\*  $\chi^2$  test, 2-tailed

DCIS-ductal carcinoma *in situ*; LCIS-lobular carcinoma *in situ*; IBC-invasive breast cancer

*Table 3. MMP-1 2G insertion polymorphism is correlated with MMP-1 expression in IBC*

Genotype	MMP-1 expression (n=41)		
	1+ (n=14) n (%)	2+ (n=17) n (%)	3+ (n=10) n (%)
1G/1G	8 (57)	6 (35)	2 (20)
1G/2G	4 (29)	7 (41)	4 (40)
2G/2G	2 (14)	4 (24)	4 (40)

*P = 0.03 (Jonckheere-Terpstra test)*

Table 4. MMP-1 1G/2G polymorphism is correlated with HER2 expression in IBC

Genotype	HER2-(n=26)	HER2+(n=23)	OR	95%CI	P*
	n ( f)	n ( f)			
1G/1G	10(0.39)	9 (0.39)	2.78	0.64-12.06	0.29
1G/2G	12(0.46)	4 (0.17)	1		
2G/2G	4 (0.15)	10 (0.44)	7.50	1.48-37.90	0.03

Overall  $P^* = 0.04$

\* Fisher's exact test, 2- tailed

Table 5. MMP-1 1G/2G polymorphism is correlated with P53 expression in IBC

Genotype	P53-(n=21)		P53+(n=22)		OR	95%CI	P*
	n (f)	n (f)	n (f)	n (f)			
1G/1G	5 (0.24)	10 (0.45)	5.26	1.18-25.00			0.038
1G/2G	13 (0.62)	5 (0.23)	1				
2G/2G	3 (0.14)	7 (0.32)	5.88	5.26-33.30			0.049

Overall P\* = 0.04

\* Fisher's exact test, 2- tailed

Fig.1

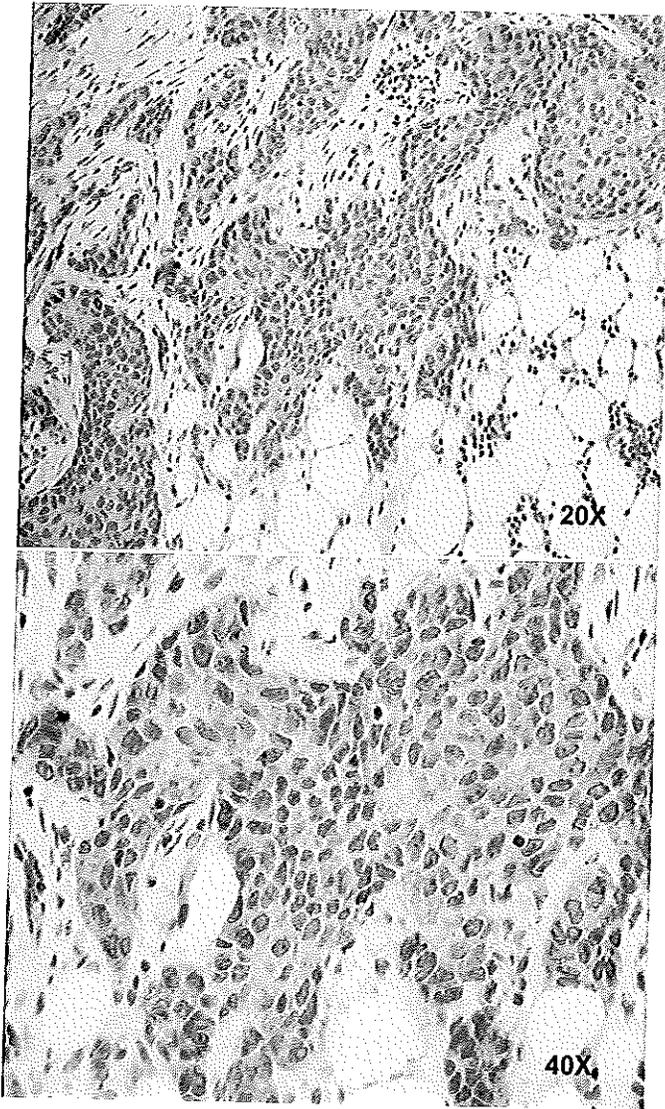


Fig.2a

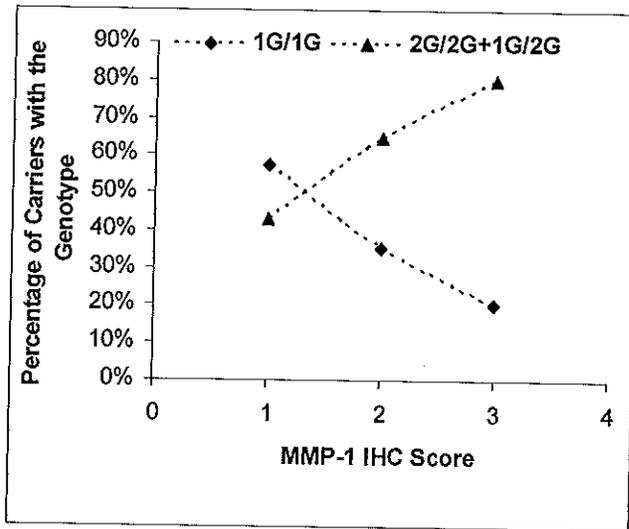
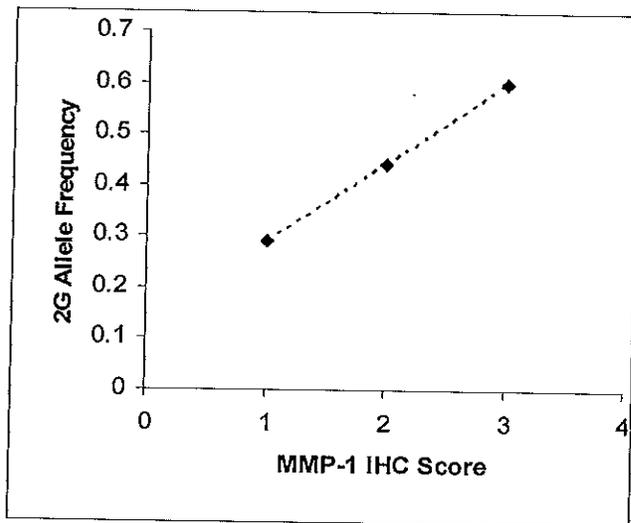
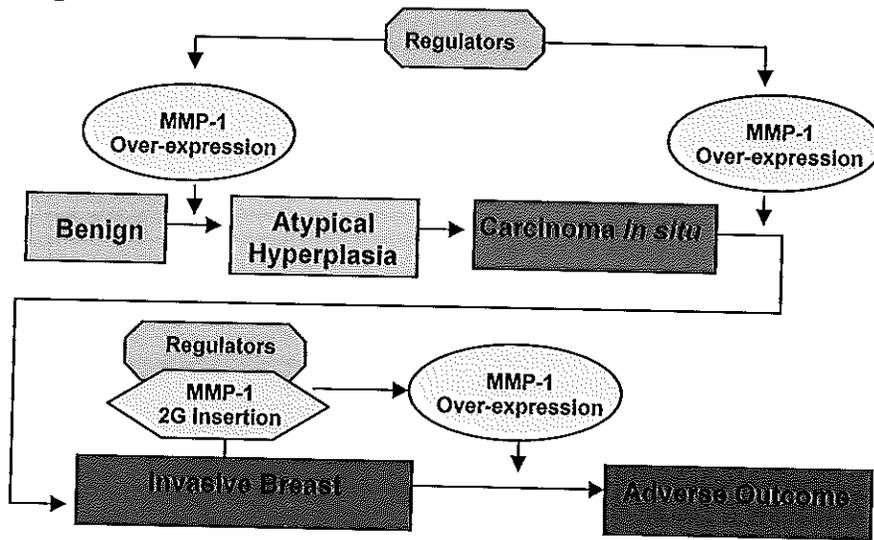


Fig.2b

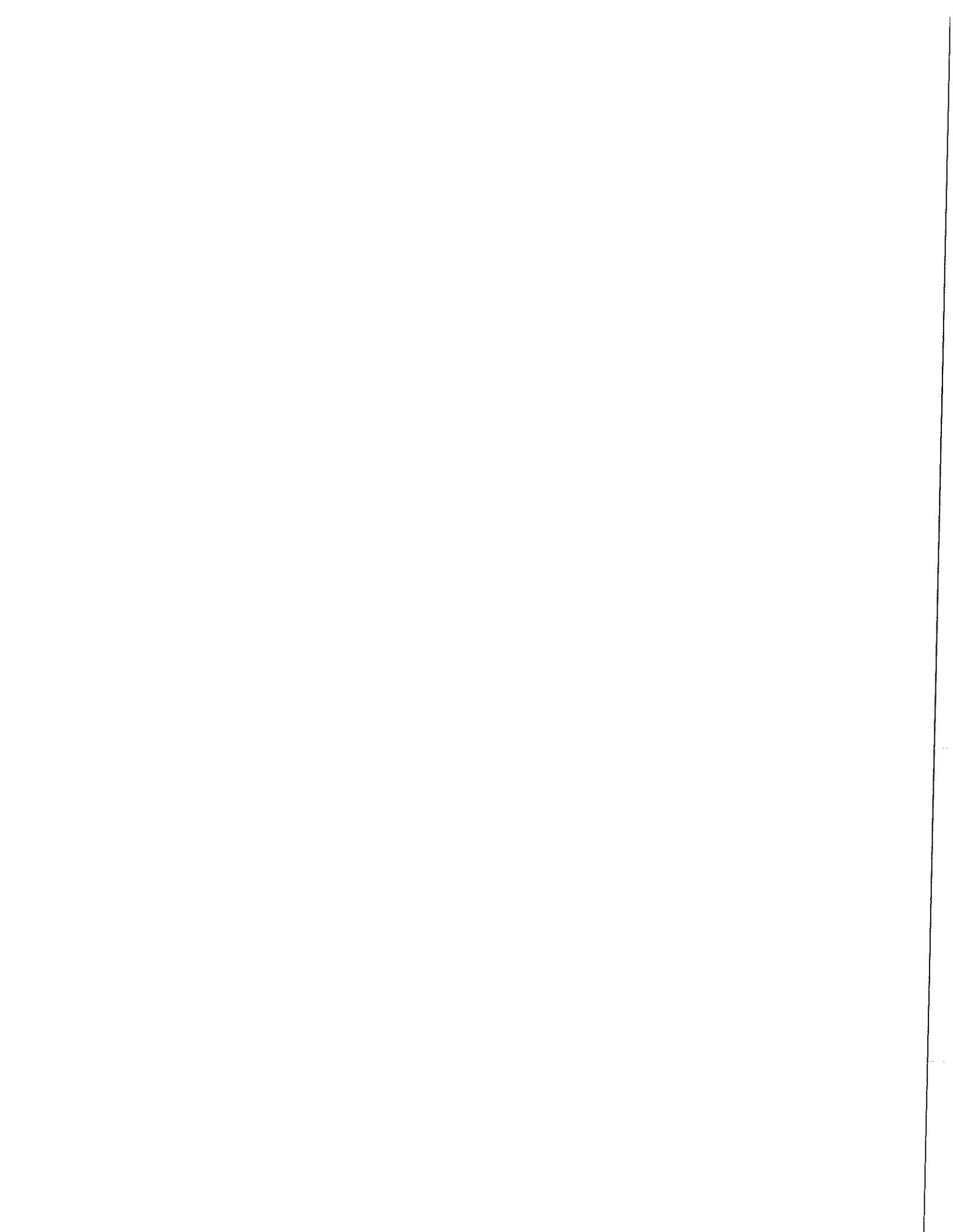


Created with excel

Fig.3



MS office drawing objects



**Title: Are We Army or Navy? Don't Ask, Don't Tell - The Challenge of Merging Two Breast Centers into One**

Primary Author: Lee Bronfman, RN, MA, CCRP, Administrative Director  
Walter Reed Army Medical Center  
Phone: 202-782-0001 Fax: 202-782-4260  
e-mail: [lee.bronfman@na.amedd.army.mil](mailto:lee.bronfman@na.amedd.army.mil)

Mark Friend, Assistant Department Head  
Breast Care Department  
National Naval Medical Center  
Bethesda, Maryland  
e-mail: [mark.friend@med.navy.mil](mailto:mark.friend@med.navy.mil)

**Category I: Office Operations, Administration and Programs Subcategory B. Administration**  
1. Staffing: Selection Process, Justification, etc.

**Introduction:** The 2005 Base Realignment and Closure (BRAC) Act recommended closing 33 major United States military bases and the "realignment" (either enlarging or shrinking) of 29 others. As part of a Base Realignment and Closure announcement on May 13, 2005, the Department of Defense proposed replacing Walter Reed Army Medical Center with a new Walter Reed National Military Medical Center (WRNMMC). The new center would be on the grounds of the National Naval Medical Center in Bethesda, Maryland, seven miles (11 km) from WRAMC's current location in Washington, D.C. The proposal is part of a program to transform medical facilities into joint facilities, with staff including Army, Navy, and Air Force medical personnel. The Secretary of Defense was required to begin implementing the recommendations by September 15, 2007 and must complete implementation no later than September 15, 2011.

**Methods:** Leadership set the tone with COL Craig Shriver named the integrated Chief for Breast Care and CAPT Philip Perdue named the Integrated Chief of General Surgery. Each began seeing breast clinic at the other's facility and frequently met to compare the best (and worst) of each. The Army Clinical Breast Care Project has a three day annual offsite meeting each year and our Navy colleagues were invited to attend. On the last morning of the 2009 meeting, COL Shriver and CAPT Perdue led a brainstorming session to generate ideas on:

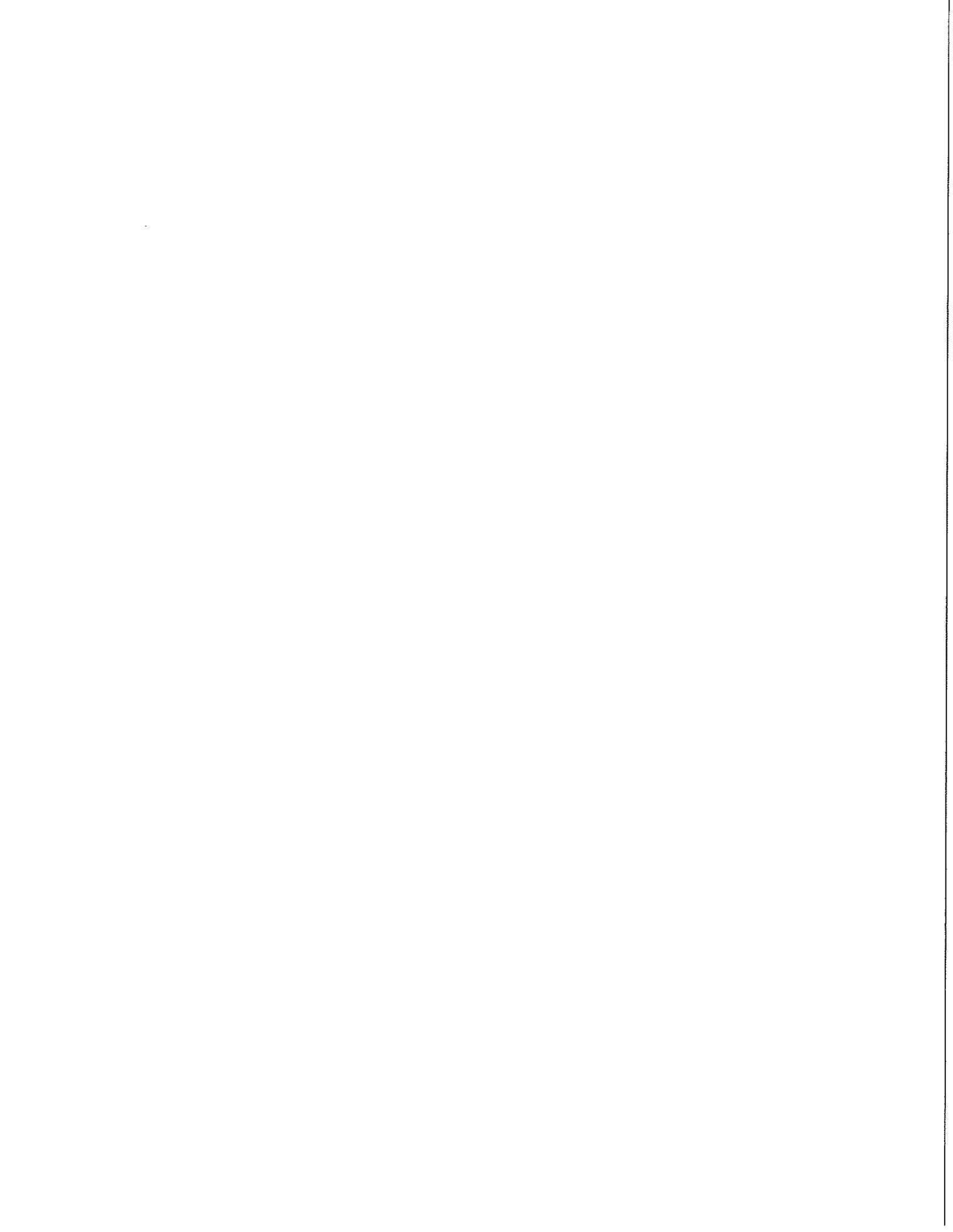
- What do we want our new breast center to be/do/have?
- What resources (people) do we have now to do that? What don't we have?
- Plan/Timeline for Way Forward: Meetings, Site Visits, Working Groups
- Patient/Research Flow at New Center

The administrative and nursing leadership from both the Army and Navy facilities established work groups to address issues such as clinic flow, the multidisciplinary conference, clinic décor and research.

**Results:** Although very much a work in progress, close working relationships have begun between Army and Navy leadership and administrative personnel. Finishing ahead of schedule, the new Breast and Research Center at the new Walter Reed National Military Medical Center is open for business with Navy personnel in place. Walter Reed Army Medical Center, scheduled to close in September of 2011, continues to be open for business as usual. The Navy Breast Center, a comprehensive center with a co-located imaging center has not been heavily involved in research, whereas the Army facility has functioned as a translational research center of excellence since its creation in 2000. Preliminary plans will include staffing the co-located imaging center at the new Breast and Research Center with existing Navy personnel. The clinical

care of patients will initially be provided by the existing Navy staff and the research component which includes consenting of patients, completion of questionnaires, and data QA provided by the Army personnel.

**Conclusions:** The new Breast and Research Center at the new Walter Reed Military Medical Center is in its infancy. Cheryl Boglarsky, Ph.D., Director of Research at Human Synergistics International identifies steps to successful mergers and acquisitions. Step one requires everyone to understand that a merger will transform two organizations into a new, single entity. Step two requires a new vision, and step three involves establishing a culture that can deliver that vision. The fourth step is appropriate leadership that embraces the vision. The new Breast and Research Center at the new Walter Reed National Military Medical Center is located in a Comprehensive Cancer Center whose Implementation Team is being led by the Integrated Chief for Breast Care. The vision of a world class cancer center providing excellence in clinical care and research has been determined and we hope, with proper care and feeding by line managers through executive leadership, the systems, structures, technology and training will be set in motion to achieve this vision and the ideal culture for this new Breast and Research Center.





Display Settings: Abstract



Comput Methods Programs Biomed. 2012 Sep 14. pii: S0169-2607(12)00192-7. doi: 10.1016/j.cmpb.2012.08.010. [Epub ahead of print]

## **QAiT: A quality assurance issue tracking tool to facilitate the improvement of clinical data quality.**

Zhang Y, Sun W, Gutcheon EM, Kvecher L, Kohr J, Bekhash A, Shriver CD, Liebman MN, Mural RJ, Hu H.  
Windber Research Institute, Windber, PA 15963, United States.

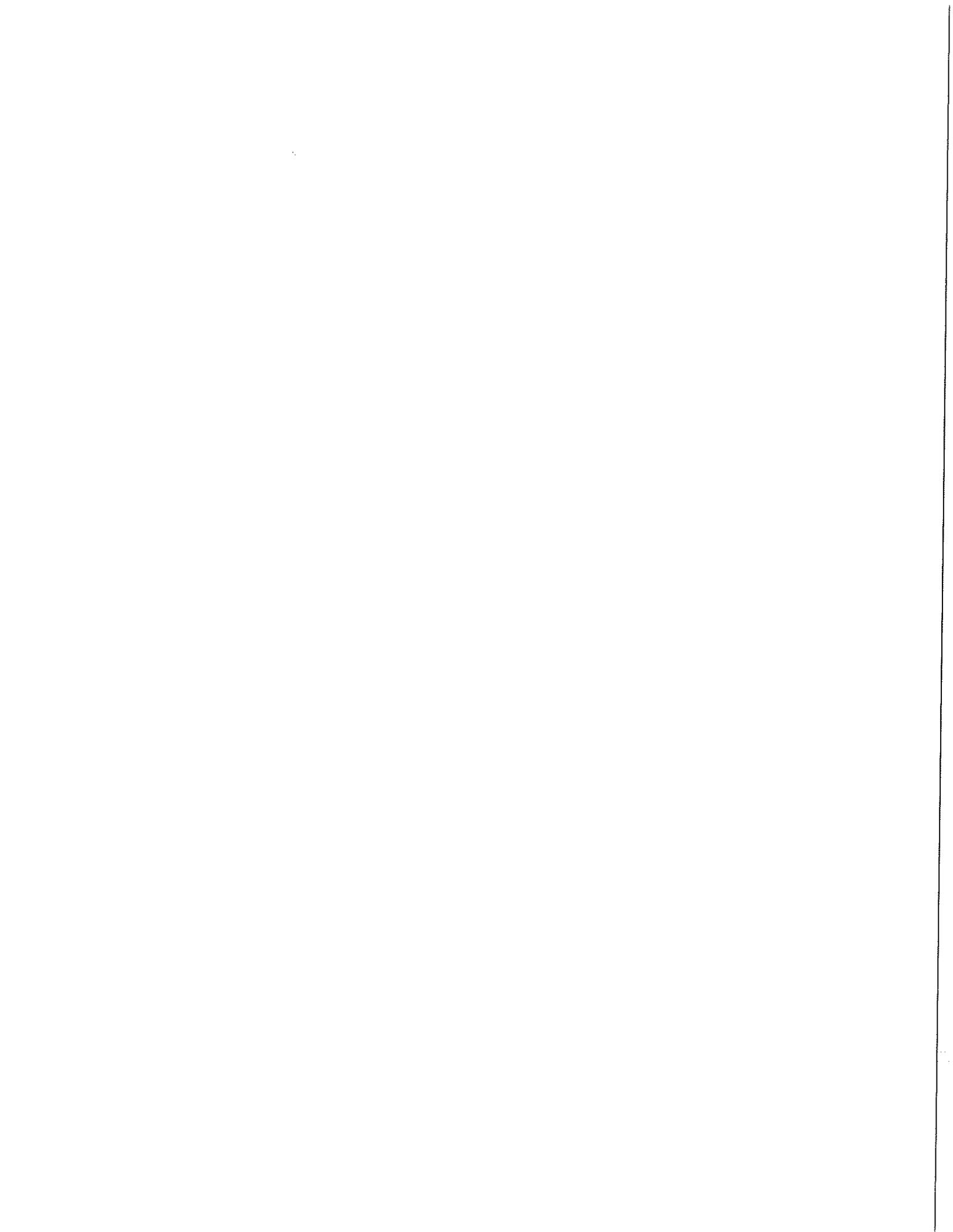
### **Abstract**

In clinical and translational research as well as clinical trial projects, clinical data collection is prone to errors such as missing data, and misinterpretation or inconsistency of the data. A good quality assurance (QA) program can resolve many such errors though this requires efficient communications between the QA staff and data collectors. Managing such communications is critical to resolving QA problems but imposes a major challenge for a project involving multiple clinical and data processing sites. We have developed a QA issue tracking (QAiT) system to support clinical data QA in the Clinical Breast Care Project (CBCP). This web-based application provides centralized management of QA issues with role-based access privileges. It has greatly facilitated the QA process and enhanced the overall quality of the CBCP clinical data. As a stand-alone system, QAiT can supplement any other clinical data management systems and can be adapted to support other projects.

Copyright © 2012 Elsevier Ireland Ltd. All rights reserved.

PMID:22985872[PubMed - as supplied by publisher]

**LinkOut - more resources**





## DW4TR: A Data Warehouse for Translational Research

Hai Hu<sup>a,\*</sup>, Mick Correll<sup>c,1</sup>, Leonid Kvecher<sup>a</sup>, Michelle Osmond<sup>b,2</sup>, Jim Clark<sup>b,2</sup>, Anthony Bekhash<sup>a</sup>, Gwendolyn Schwab<sup>a</sup>, De Gao<sup>d,2</sup>, Jun Gao<sup>d,2</sup>, Vladimir Kubatin<sup>c,2</sup>, Craig D. Shriver<sup>e</sup>, Jeffrey A. Hooke<sup>e</sup>, Larry G. Maxwell<sup>e</sup>, Albert J. Kovatich<sup>f</sup>, Jonathan G. Sheldon<sup>b,3</sup>, Michael N. Liebman<sup>a,4</sup>, Richard J. Mural<sup>a</sup>

<sup>a</sup>Windber Research Institute, Windber, PA, USA

<sup>b</sup>InforSense LLC, London, UK

<sup>c</sup>InforSense Ltd., Boston, MA, USA

<sup>d</sup>InforSense Ltd., Shanghai, China

<sup>e</sup>Walter Reed Army Medical Center, Washington, DC, USA

<sup>f</sup>MDR Global, Windber, PA, USA

### ARTICLE INFO

#### Article history:

Received 29 September 2010

Accepted 4 August 2011

Available online 22 August 2011

#### Keywords:

Data Warehouse

Translational research

Patient-centric data model

User interface

Ontology

### ABSTRACT

The linkage between the clinical and laboratory research domains is a key issue in translational research. Integration of clinicopathologic data alone is a major task given the number of data elements involved. For a translational research environment, it is critical to make these data usable at the point-of-need. Individual systems have been developed to meet the needs of particular projects though the need for a generalizable system has been recognized. Increased use of Electronic Medical Record data in translational research will demand generalizing the system for integrating clinical data to support the study of a broad range of human diseases. To ultimately satisfy these needs, we have developed a system to support multiple translational research projects. This system, the Data Warehouse for Translational Research (DW4TR), is based on a light-weight, patient-centric modularly-structured clinical data model and a specimen-centric molecular data model. The temporal relationships of the data are also part of the model. The data are accessed through an interface composed of an Aggregated Biomedical-Information Browser (ABB) and an Individual Subject Information Viewer (ISIV) which target general users. The system was developed to support a breast cancer translational research program and has been extended to support a gynecological disease program. Further extensions of the DW4TR are underway. We believe that the DW4TR will play an important role in translational research across multiple disease types.

© 2011 Elsevier Inc. All rights reserved.

### 1. Introduction

Translational research requires integration of clinicopathologic data, biospecimen data, and molecular data, across multiple data collection and generation platforms [1–4]. It is also critical to make these data usable at the point-of-need. From the data management perspective, clinicopathologic data and molecular data have distinct features. The former are characterized by a large number of data fields, often with missing values due to failure to collect, non-existence or inaccessibility of the data. Molecular data, on the other hand, are characterized by a limited number of data

fields, large number of records, and ever-evolving data types associated with the development of new technologies. While a translational research project typically collects and generates its own clinicopathologic and molecular data, public domain data are often used as well. Furthermore, temporal information needs to be properly managed and presented.

Many clinical and translational research projects use questionnaires as a clinical data collection instrument, and currently using data from Electronic Medical Records (EMRs) for translational research is gaining momentum as EMR systems replace paper-based medical records. Compared to EMRs, questionnaires are simpler as only questions useful to the specific research program of the disease would be included, thus managing such data does not require a comprehensive system covering all the human diseases. An EMR system cannot be used as a data management system for translational research, due to the transactional nature of the former and the reporting and analysis requirements of the latter. The requirements from the US Health Insurance Portability and Accountability Act of 1996 (HIPAA) for protection of patients' Protected Health

\* Corresponding author. Address: Windber Research Institute, 620 7th St., Windber, PA 15963, USA. Fax: +1 814 467 6334.

E-mail address: [h.hu@wriwindber.org](mailto:h.hu@wriwindber.org) (H. Hu).

<sup>1</sup> Present address: Dana-Farber Cancer Institute, Boston, MA, USA.

<sup>2</sup> Present address: IDBS, Burlington, MA, USA.

<sup>3</sup> Present address: Oracle, Swindon, UK.

<sup>4</sup> Present address: Strategic Medicine, Kennett Square, PA, USA.

## Abbreviations

2D-DIGE	2-Dimensional, Difference In Gel Electrophoresis	IT	Information Technology
ABB	Aggregated Biomedical-Information Browser	LC-MS	Liquid-Chromatography Mass-Spectrometry
aCGH	Array Comparative Genomic Hybridization	LIMS	Laboratory Information Management System
AJAX	Asynchronous JavaScript and XML	MeSH	Medical Subject Heading
AJCC	American Joint Committee on Cancer	MIAME	Minimum Information About a Microarray Experiment
ASCO	American Society of Clinical Oncology	MS	Mass Spectrometry
BioPAX	Biological Pathway Exchange	NBIA	National Biomedical Imaging Archive
BMI	Body-Mass Index	NCI	National Cancer Institute
caBIG <sup>a</sup>	Cancer Biomedical Informatics Grid	NCICB	National Cancer Institute Center for Bioinformatics
caBIO	Cancer Bioinformatics Infrastructure Objects	NIH	National Institutes of Health
caCORE	cancer Common Ontological Reference Environment	OCT	Optical Cutting Temperature medium
caDSR	cancer Data Standards Registry and Repository	OLAP	On-Line Analytical Processing
CAP	College of American Pathologists	MOLAP	Multi-dimensional OLAP
CBCP	Clinical Breast Care Project	ROLAP	Relational OLAP
CDW	Clinical Data Warehouse	OWL	Web Ontology Language
CGEMS	Cancer Genetic Markers of Susceptibility	PHI	Protected Health Information
CTSA	Clinical and Translational Science Award	PI	Principal Investigator
DICOM	Digital Imaging Communications in Medicine	PR	Progesterone Receptor
DNA	Deoxyribonucleic Acid	QA	Quality Assurance
DW	Data Warehouse	RAID	Redundant Array of Independent Disks
DW4TR	Data Warehouse for Translational Research	RAM	Random-Access Memory
EAV	Entity-Attribute-Value	REMBRANDT	Repository of Molecular Brain Neoplasia Data
EMR	Electronic Medical Record	RDBMS	Relational Database Management System
ER	Estrogen Receptor	RIN	RNA Integrity Number
ETL	Extract, Transform, and Load	RNA	Ribonucleic Acid
FF	Flash Frozen	RT-PCR	Real-Time Polymerase Chain Reaction
FFPE	Formalin-Fixed, Paraffin-Embedded	SAS	Statistical Analysis System
FISH	Fluorescence In Situ Hybridization	SNOMED-CT	Systemized Nomenclature of Medicine-Clinical Terms
GB	Giga-Byte	SNP	Single-Nucleotide Polymorphism
GDP	Gynecological Disease Program	SPSS	Statistical Package for the Social Sciences
GO	Gene Ontology	STRIDE	Stanford Translational Research Integrated Database Environment
H&E	Hematoxylin and Eosin	TB	Tera-Byte
HER2	Human Epidermal growth factor Receptor 2	UML	Unified Modeling Language
HIPAA	Health Insurance Portability and Accountability Act of 1996	UMLS	Unified Medical Language System
HL7	Health Level 7	VCDE	Vocabulary and Common Data Elements
I2B2	Informatics for Integrating Biology and the Bedside	VPD	Virtual Private Database
ID	Identification	VPN	Virtual Private Network
IHC	Immunohistochemistry	WRAMC	Walter Reed Army Medical Center
IRB	Institutional Review Board	XML	Extensible Markup Language
ISIV	Individual Subject Information Viewer		

Information (PHI) also differ between the two systems. In addition, translational research may not need all of the data contained in an EMR and it requires data from other sources as well including molecular data. It is important that a data centralization system be developed for translational research capable of inputting clinical data from questionnaires and EMRs for any human disease, and integrating molecular data from different biochemical, genomic and proteomic experimental platforms.

Data warehousing as an important component of biomedical informatics, provides a way for data integration in clinical and translational research [2,5–13]. A Data Warehouse (DW) is an information repository for data analysis and reporting [14,15]. For raw clinical data, an Entity-Attribute-Value (EAV) data model is often used due to their sparseness and dynamic nature [16–18]. Such data models require the development of ontologies to manage the complex relationships between clinical concepts/questions and possible answers. The user interface is often provided by an On-Line Analytical Processing (OLAP) tool, which provides multidimensional data analysis capabilities by aggregating data across hierarchically organized dimensions using either a multidimensional (MOLAP) or relational (ROLAP) model [14,15,19]; the former

relies on pre-computed data cubes whereas the latter directly queries a relational database.

Developing a comprehensive DW system to meet the needs of translational research faces many challenges. Based on our own experience, such challenges include; management of operationally important information, such as HIPAA compliance and PHI, data input from questionnaires with version control, data input from EMRs, biospecimen collection and banking, data de-identification, molecular images, multiple-platforms of molecular data, temporal information, and data ownership, etc. From the system development perspective challenges include; development and application of ontologies including controlled vocabulary and modeling of clinical data, molecular data, image data, and temporal data. In addition, development of graphical user interfaces for different levels of users is important, and data security and accessibility should be properly addressed.

Ontology development is very important to data integration and exchange; for the purpose of this paper we focus on its physical artifacts of terminologies/controlled vocabulary and information model/data model [20]. Many ontologies have been developed, each serving a specific purpose. In the clinical field,

commonly accepted standards/ontologies include; Health Level 7 (HL7) for exchange of clinical messages to support clinical practice and healthcare service [21]; Medical Subject Heading (MeSH) for medical literature indexing which has also proven useful for classification of biomedical entities [22]; Systemized Nomenclature of Medicine—Clinical Terms (SNOMED-CT), as a comprehensive clinical healthcare terminology organized hierarchically [23]; and Digital Imaging Communications in Medicine (DICOM) for medical image annotations [24]. For biological research, Gene Ontology (GO), arguably the most successful biological ontology, is aimed to standardize the representation of gene and gene product attributes across species and databases [25]; Minimum Information About a Microarray Experiment (MIAME) for microarray-based technologies is generally accepted by the community [26]; Biological Pathway Exchange (BioPAX) is for pathway data exchange [27]. NCI Thesaurus provides reference terminology for many NCI and other systems covering vocabulary for clinical care and translational and basic research [28]. Systems have been developed to unify existing ontologies or provide ontology library services, for example the BioPortal by the US National Center for Biomedical Ontology [29], and the Ontology Lookup Service by European Bioinformatics Institute based on the Open Biomedical Ontology format [30]. The Unified Medical Language System (UMLS) developed by the US National Library of Medicine, aims to provide ad hoc linkages across life science ontologies [31]. A number of reviews are available including ones discussing development, implementation, and use of ontologies [20,32–34].

Temporal information is also very important in modeling health and disease development. We consider disease as a process, not a state. Breast cancer, for example, takes many years to develop before a lesion can be detected by mammography [35]. First considered by artificial intelligence researchers in the 1980s, temporal reasoning and temporal data query became a topic of interest in biomedical/medical informatics in the 1990s [36–38]. To enable “controlled language use” of the temporal information in healthcare, a European Prestandard CEN ENV 12381 has been developed [39]. Temporal information needs to be collected, modeled, and presented. Collection of temporal information can be done explicitly through questionnaires or data forms which record historical information, or abstracted through the time-stamps of data collected and analyzed over time. For the latter, temporal-abstraction and temporal-querying systems have been developed which have proven clinically useful in monitoring clinical disorders [40–42]. A data model comprising Parameters, Events, and Constants has been defined for a temporal query language [43,44], where Parameters are basically entities with values bearing a time-stamp, Events are external acts act upon a patient which could be time-stamped or occur over a time interval, and Constants are non-temporal measures. There have been reports of a data management or DW system incorporating temporal information for clinical trials, time-series gene expression microarray experiment, and medical information events in a hospital setting [38,45,46].

The many challenges faced in developing a comprehensive DW system for translational research make this a difficult undertaking. In practice, different systems have been developed focusing on different challenges based on specific needs of supported program(s). For example, I2B2, Informatics for Integrating Biology and the Bedside [47], took a top-down “Enterprise” approach with an open architecture enabling integration of outside modules performing specialized functions that are needed for translational research [48,49]. It address a very important and specialized niche in the clinical informatics space, namely, providing researchers with sufficient access to clinical data to enable study design and cohort selection, while minimizing many of the patient privacy risks and concerns [50–54]. A central conundrum of translational research is that access to detailed clinical data typically requires proper

informed consent and IRB approval; however, in order to design a cohort study or to determine if a particular project is even feasible, a researcher often needs access to aggregate statistics for key clinical attributes of interest at an early stage of the process. I2B2 overcomes these challenges by enabling researchers to design and execute queries against a de-identified data mart that return only aggregate counts; furthermore, results are subtly obfuscated to avoid identification by more sophisticated combinations of queries. Having been deployed for a number of projects across multiple sites, I2B2 is a mature software offering. However, it does not currently provide the level of detailed analysis necessary for clinical and translational research beyond cohort selection and study design [51–54]. STRIDE, The Stanford Translational Research Integrated Database Environment [55], focuses on a Clinical Data Warehouse (CDW) and supports a virtual biospecimen model for accessing several specialized tissue banks [13,56,57]. It also has a Research Database management system supporting multiple logically separate research databases. An EAV model is used for data storage and HL7 Reference Information Model is used for data representation. Its semantic layer consists of a framework supporting multiple terminologies. Some of the Clinical and Translational Science Award (CTSA) centers [61] also have data warehousing efforts.

There are also data warehousing efforts in the cancer Biomedical Informatics Grids (caBIG<sup>®</sup>) [62]. caBIG<sup>®</sup> is an information network enabling members of the cancer community to share data and knowledge based on a core Grid architecture. It was launched in 2003 by the NCI Center for Bioinformatics (NCICB) of the US NIH [63,64]. One of its data warehousing efforts is caBIO, a major component of caCORE (cancer Common Ontological Reference Environment) that was developed before the launch of caBIG<sup>®</sup> [65], enabling access of biomedical annotations from curated data sources in an integrated view. As one of the released products of caBIG<sup>®</sup>, the current version caCORE 3.x includes Enterprise Vocabulary Services for hosting and managing vocabulary, cancer Data Standards Registry and Repository (caDSR) for hosting and managing metadata, and the caCORE Software Development Kit [66]. Another data warehousing approach for caGrid is a semantic web-based data warehouse for creating relationships among caGrid models, accomplished through the transformation of semantically-annotated caBIG<sup>®</sup> Unified Modeling Language (UML) information models into Web Ontology Language (OWL) ontologies that preserve those semantics [67]. Still another caBIG effort, caIntegrator [58], is a framework for data integration currently capable of integrating microarray data in caArray and imaging data from the National Biomedical Imaging Archive (NBIA). It has been applied to several projects, e.g., REMBRANDT [59] and CGEMS [60]. These efforts are best described as development projects using the caIntegrator framework. None of these efforts in caBIG, however, enables users to dynamically interrogate the clinicopathologic data in a multidimensional manner by non-informatics specialists using an intuitive interface.

We began our DW development focused on a questionnaire-based system, but later realized that a patient-centric and expandable data model would be a better solution. We envisioned developing a system to support translational research in general and began by developing a system to support the Clinical Breast Care Project (CBCP) [68] but designed it to be expandable to support additional programs. This approach has been validated by our successful expansion of the system to support a second translational research program, the Gynecological Disease Program (GDP) [69]. It is currently under further expansion to support a third translational research program, a consortium effort headed by Thomas Jefferson University involving five organizations which is characterizing a set of 5000 invasive breast cancer cases. All these projects have multiple platforms for clinical and molecular studies.

Development of our system, the Data Warehouse for Translational Research (DW4TR), focused on three challenges in DW development. The first was the development of a pragmatic data model to satisfy the needs of questionnaire-based clinical data input. Despite the existence of a large number of clinical and biological ontologies, there are often gaps, in supporting a research project, between the available ontologies and the specific project needs [70,71]. Many groups therefore create their own ontologies [72,73]. Given the characteristics of clinical and molecular data, it is more challenging to develop a general and flexible system for clinical data; therefore we have made this our first priority [9]. In our case, both CBCP and GDP utilizes not only data elements covered by existing ontologies, but also detailed and specific data elements that are currently not covered. Comparing what we need to what is available, we concluded that completely adopting one or two existing ontologies for our DW development is not feasible (c.f. Section 5). Therefore, we developed our own framework ontology but referred to existing ones, to satisfy our immediate needs first, planning to map our ontology to existing ontologies at a later time to enable transportability.

The second challenge, we focused on was interface development. A user-friendly interface tailored to the non-informatics specialist is needed to ensure both the utility and adoption of this system. The interface that we developed is composed of an Aggregated Biomedical-Information Browser (ABB) and an Individual Subject Information Viewer (ISIV). ABB is ROLAP-based, thus all the calculations and queries are performed at the time of use rather than relying on a pre-calculated data cube, enabling clinicians and scientists who are not informatics experts to dynamically interrogate any combination of the myriad of data elements stored in the DW4TR. The third challenge we focused on was the management of temporal information. We defined three temporal data types and applied them to data attributes to enable the use of the temporal information in the DW4TR from both interfaces. In addition to focusing on these three challenges, we also addressed a number of other challenges described earlier. For example, in managing de-identified clinical information of CBCP and GDP, we developed solutions to satisfy different requirements for data security and accessibility that is also described.

## 2. Methods

### 2.1. System design and development

Initially developed to support the immediate needs of managing the clinical, biospecimen, and molecular data for the CBCP, our design of the DW4TR to be flexible and extensible enabled subsequent expansion to support GDP. Both programs use questionnaires for collecting clinicopathologic information, making expansion of the light-weight production system straight forward.

#### 2.1.1. The data model

The EAV model is used for the original raw clinical data which are subsequently hosted in the DW in an extensible data model reflecting the ontology, where the relationships among the attributes are presented in a meaningful way to the users through a hierarchical organization. The extensible data model is currently composed of the following components:

**2.1.1.1. Patient-centric clinical data model.** The clinical data model was developed to reflect the physician–patient interaction. The ontology framework was developed by a multi-disciplinary team composed of research physicians, biomedical informaticians, IT developers, surgeons, pathologists, and gynecologists and was based on the standard clinical workflows and clinical experiences.

The development not only referred to existing standards and ontologies mainly MeSH and SNOMED-CT but also to the NCI Thesaurus and caBIG VCDE [22,23,28,74], reflecting other reported methods and efforts [32,33,75]. This ontology is reflected on the physical data model, which is hierarchical and composed of a number of primary modules, e.g., "Diagnostic", which are made up of secondary modules (e.g., "Biopsy") and attributes. A secondary module is further composed of tertiary modules and attributes, and so on. An attribute is a fine-grained object that is composed of the name, the data type, and the temporal characteristics (see below). A simple attribute is often called a data element, e.g., sex. A complex attribute is composed of multiple facets, e.g., exercise is composed of frequency, intensity, duration, and type. This way, each data element of a study is represented in the model either by an attribute or by a sub-module. The sub-modules are structured hierarchically to reflect the ontology.

**2.1.1.2. Specimen-centric molecular study data model.** Molecular (biochemical, genomic and proteomic) analysis are performed on collected specimens using multiple experimental platforms. Every experiment is done using biospecimens, including body fluids, solid tissues, and their derivatives. Thus, a specimen-centric data model is a natural choice for these types of data, which are connected to the clinical data model hierarchy as a series of sub-modules. We initially focused on immunohistochemistry (IHC), Fluorescence In Situ Hybridization (FISH), and gene expression microarray data and adopted existing clinical or molecular data standards and guidelines [26,76,77]. The storage and retrieval of high throughput molecular data poses a number of IT challenges. A result file from a single microarray can be tens of megabytes, and it is not uncommon for a single experiment to make use of hundreds of such arrays. Common approaches for handling this type of data include: organizing output files into a hierarchical file system, loading of raw or partially transformed data into a DW, or a hybrid approach of maintaining certain metadata in a database but referring back to original output files with file pointers. Given the large number of molecular modalities used in our current and future research, no single method was considered suitable, therefore our system was designed to take an "all of the above approach" that makes use of flexible data processing pipelines. With this approach distinct pipelines can be created across or even within assay types and all of the storage methods discussed above can be supported. This approach has enabled us to optimize storage of each assay group individually, as well as provide a mechanism for handling changing technology and legacy data.

**2.1.1.3. Temporal data model.** Critical to human disease studies is the proper representation of temporal information. We define three types of temporal data; (1) Static, which is data with no temporal dimension, e.g., ethnicity. (2) Event, which is associated with a specific time point, e.g., a surgical procedure. (3) Interval, which is associated with a starting and ending time point, e.g., a course of medication. Each attribute in the data model is tagged with a proper temporal status and populated with proper values for temporal information representation and analysis.

**2.1.1.4. Medical image data model.** Medical imaging is an important part of clinical practice and a source of data for various studies. We initially focus on digitized and digital mammograms, which follows the DICOM standards [24] and we de-identified the images using DICOM Anonymizer Pro. The data files associated with these images are large and they impose a challenge in data management, e.g., storing those image files in the database of the DW will drastically impede the system performance. Besides, analyzing images typically requires specialized software. Thus, in the DW4TR, we store the image files on a file server; the file locations are then

stored in the DW and logically linked to the patient. The annotations and a thumb nail of the image are also stored in the database to enable efficient search of relevant features. We intend to apply the same principle to other images.

**2.1.1.5. Relationships between questionnaires and the data model.** The data for both CBCP and GDP studies are collected using questionnaires each having multiple questions. Each question has one or more data elements. The data elements could contain values and temporal information. If a data element contains a value, it is mapped to an attribute(s). Some questions ask about the same event at different time points, and these questions are mapped to the same attribute with corresponding temporal information. There are also different questionnaires collecting the same information, e.g. for different studies, and they have been mapped to the same attributes as well. Finally each attribute is mapped to the data model as described above.

### 2.1.2. User interface development

The requirement to develop an intuitive, user friendly system that could be used directly by clinicians and researchers as opposed to being a specialist tool useful only to the technically savvy, led to our decision to take the dimensional modeling approach. While traditional query interfaces are considered intuitive by IT professionals, we found that the dimensional modeling approach was generally preferred by our target demographic. A number of criteria were considered in designing the physical data structure to support the clinical data warehouse, the first of which was the ability to handle a large number of dimensions. A custom binning capability that is convenient to use, was an important feature requested by the users. As is the case with any longitudinal study, the temporal dimension of data is critical, and therefore proper handling of the temporal dimension was another important concern. Efficient handling of sparsely distributed data and support for multiple terminology sets were also important criteria. During such comprehensive communications with the end users, we concluded that all the desired analyses could be categorized as either an aggregate data analysis or an individualized data analysis. Thus, we designed a ROLAP-based ABB for the former, and an ISIV for the latter.

## 2.2. Implementation

### 2.2.1. Development environment

Two DELL Windows server systems, one for the application and the other for the database, were used. Both servers have two Quad Core Intel® Xeron® processors and 8 GB RAM. The database server hard drive is 1 TB and for the application server 0.5 TB, both are in the RAID 5 configuration. Development is based on the Oracle RDBMS 10g, using the InforSense platform that enables a series of analyses to be performed in an analytical workflow environment with each analysis step represented as a node in the workflow [78]. The user-interface is Java based and web accessible. The production system has the same configuration. The DW4TR also supports a one-server system configuration.

### 2.2.2. Data model and interface implementation

Our physical data model is composed of two distinct elements, an attribute repository, and an attribute ontology. The attribute repository stores data in the EAV model. The attribute ontology is a separate hierarchical data model capturing and providing structure to the relationships among the attributes that have been collected so that the ontology (logical patient-centric or specimen-centric data model) are physically positioned in a hierarchical organization of attributes for presentation in a meaningful way to the users of the system.

In implementing the data model, first the source questionnaires are reviewed and a logical mapping of the data attributes onto the newly designed patient modules is developed. This process requires the definition of the data variable type and other metadata for different data attributes, and often requires splitting or merging data elements in order to accommodate the difference in requirements of capturing the data versus analyzing the data.

Parallel to the development of the physical and logical data model was the development of the end user interfaces. Our system provides users with two distinct applications: the ABB, and the ISIV. These Java tools are entirely web based with zero client side footprint, and do not require any additional browser plug-ins. They provide users with a highly interactive and dynamic experience, making use of Asynchronous JavaScript and XML (AJAX) techniques for asynchronous rendering and updating. Both tools are deployed within the InforSense web portal environment, which provides the services used for data source connection pooling, user authentication and authorization, as well as web session management.

### 2.2.3. Data loading

Multiple sources of data in different format are loaded into the DW4TR, including direct Oracle-to-Oracle load from the Laboratory Information Management System (LIMS) and flat files [12,79]. Data loading is performed through a standard process referred to as Extract, Transform, and Load (ETL) [14,15].

## 2.3. Data sources

The current DW4TR is host to clinical data from CBCP and GDP. Subjects are enrolled into the programs via HIPAA compliant, Institutional Review Board (IRB) approved protocols at multiple participating clinical sites, and these protocols explicitly specified WRI as the data integration center for the programs. Clinical data are collected through multiple questionnaires, and biological specimens are collected, processed, banked, and analyzed using genomic and proteomic experimental technologies.

CBCP and GDP happen to be two distinct types of programs in administrative structure, clinical dataset, and tissue banking. They have different data security and access requirements, commanding distinct solutions in the DW4TR. CBCP is a centralized program using one set of the questionnaires, one LIMS, one Tissue Bank, and one data warehouse (the DW4TR). All the participating organizations use approved master IRB protocols from the Walter Reed Army Medical Center (WRAMC), with minor adaptations to the requirements of the local institution for local IRB approval, which is subsequently approved by the IRB of the US Army Medical Research and Materiel Command that contracted the Henry Jackson Foundation to manage the CBCP. Proper paperwork was done to enable centralized tissue banking and data warehousing at WRI. The IRB approved questionnaires contain certain date information such as date of birth which patients are consented to, thus the CBCP clinical data is a "Limited Data Set" per HIPAA definition. The clinical data and biospecimens are de-identified, and each subject is represented by a CBCP number. The link table is securely maintained at the office of the CBCP site PI which is strictly accessible only to the PI or his/her designee at that site. When clinical data Quality Assurance (QA) problems are identified in any participating institutions, they are all reported to WRAMC and the WRAMC clinical data team coordinates the QA resolving process. As of February 2011, over 5000 subjects have been enrolled in the study using two major questionnaires to collect up to 799 elements of clinicopathology data per subject. More than 43,000 specimens (including aliquots) have been collected and genomic and proteomic experiments have been conducted using these specimens.

GDP, on the other hand, is a consortium program constituting of nine clinical and research organizations with WRAMC serving as the lead institution. Each participating organization follows its own IRB-approved protocol allowing only “Safe Harbor Data Set”, and each site manages and owns its own questionnaire-specific datasets. All the captured data are in a de-identified form, and each subject is represented by a GDP study ID. Each site PI maintains the link key to the identity of the subject enrolled at that site. All the de-identified data are then sent to WRI, and eventually loaded into the DW4TR. Aggregated subject information can be shared across the consortium members, but the detailed individual subject information in the DW4TR is only accessible to the originating clinical site. In GDP, about 500 subjects have been enrolled in the study. A dozen questionnaires were used to collect more than 7600 elements of data focusing on surgical procedures, pathology, family history, psychology, food intake, etc. Currently GDP is consolidating the questionnaires being used reducing the data elements to less than 1400.

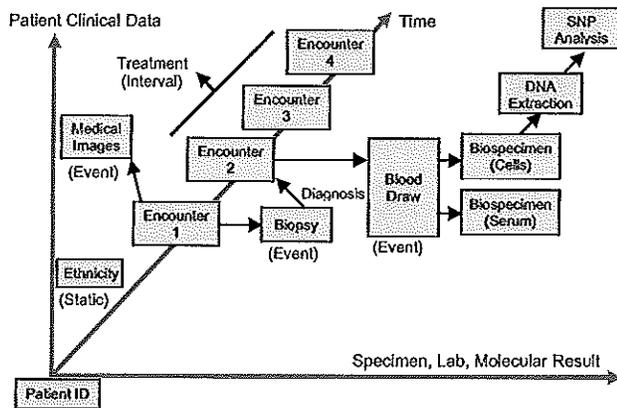


Fig. 1. Illustration in a 3-dimensional diagram of the relationship between clinical data (vertical axis), molecular data (horizontal axis), and temporal information (axis pointing into the page) represented by our data model. The patient is represented by Patient ID at origin. The three temporal data types are illustrated. This patient had clinical encounters at four time points, and “Encounter 1” resulted in medical images and biopsy which led to “Encounter 2”, which resulted in biospecimens that could be used for different types of molecular studies. “Treatment” started from “Encounter 2” and ended at “Encounter 4”.

### 3. Results

#### 3.1. Overview of the DW4TR and the relationships between involved data types

Key to the development of the DW4TR is the understanding of the complex relationships between data collected or generated with multiple platforms. Fig. 1 illustrates our understanding of such relationships in a 3-dimensional patient data space which we can use to represent our data models.

Fig. 2 is a conceptual view of the DW4TR, as a hybrid system, that integrates the data collected or generated by the supported programs, and federates most of other data needed in the studies. Blocks with solid lines are the areas already developed or currently under development, and blocks with dashed lines are areas for future development. The whole DW is composed of a data tier, a middle tier, and an application tier. In the data tier, an extended EAV model is used for the raw clinical data. The middle tier is based on a patient-centric, modularly-structured clinical data model (including a medical image data model), integrating a specimen-centric molecular data model, and a temporal data model. The clinical and molecular data models are designed to be extensible. The application tier is composed of the ABB and the ISIV, and other utilities. Additional applications will be developed to federate and present other data needed but not generated by the supported translational research programs.

#### 3.2. The ontology and the extensible data model

As a bottom-up approach, we developed a pragmatic light-weight ontology using controlled terminology both from existing ontologies and supplemental ones we developed to support the detailed program needs. The relationships between these terminologies in the ontology are reflected on the physical data models described below.

The patient-centric clinical data model was developed to reflect the physician-patient interaction. As illustrated in Fig. 3, the model is composed of six major modules: Medical History, Physical Exam, Diagnostic, Treatment, Outcome, and Scheduling and Consent. Each of these modules is composed of attributes and sub-modules; e.g., the Diagnostic module is composed of Biopsy, Imaging, and

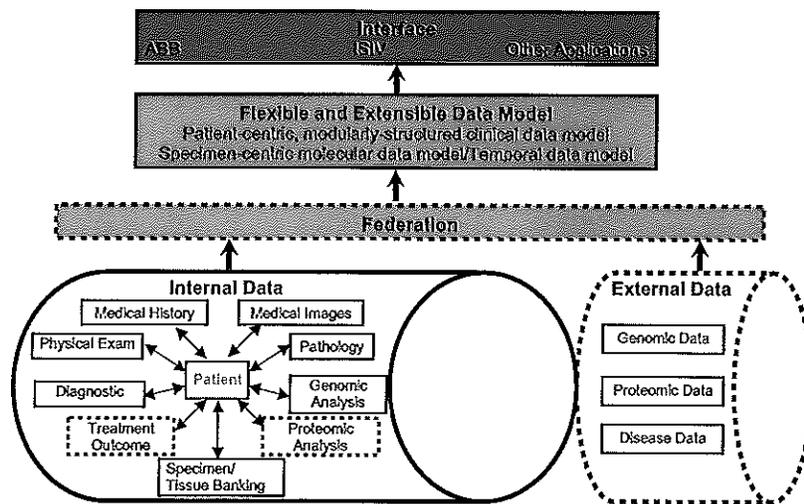


Fig. 2. The structure of the DW4TR including a data tier, a middle tier (blue), and an application tier (green). The data tier is composed of ‘Internal Data’ generated by the supported research project, and the ‘External Data’ not generated by but needed in the research. The “Internal Data” are integrated, and the “External Data” could be in an integrated or federated form but will be federated with the “Internal Data”. Solid line blocks: developed or in development; dashed line blocks: future development. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

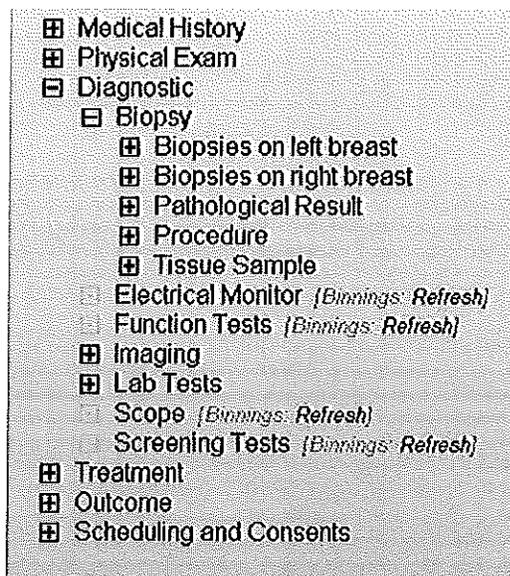


Fig. 3. A screenshot of the ontology structure of the patient-centric, modularly-structured clinical hierarchical data model.

Lab Tests, etc. Each of these sub-modules can then be further composed of its sub-modules and attributes. Many of these sub-modules are disease independent but some are disease specific. This modularly structured data model enables us to adapt the system to support the study of a new disease, by re-using disease independent modules, and developing a limited number of new modules specific to the new disease. The data model has a hierarchical structure. Users with domain knowledge can explore this hierarchical structure to retrieve needed data elements. Users with less specific domain knowledge may not be able to explore this data model as effectively though they can use the keyword search capability of the system to identify, for selection, the data elements containing keywords of interest.

### 3.2.1. The specimen-centric molecular study data model

The specimen-centric molecular study data model is connected to the clinical data model as sub-modules. For example, CBCP breast tissues and lymph nodes in different preservation types (e.g., OCT, FFPE, and FF) are represented by sub-modules of Diagnostic-Biopsy-Tissue Samples, blood samples and their derivatives (e.g., PAXgene tube, plasma, cells, serum, and clots) are represented by sub-modules of Diagnostic-Lab Tests. Genomic, proteomic, and other molecular studies are performed with molecular derivatives of the specimens (e.g. DNA, RNA, and protein) on different experimental platforms, e.g., IHC, FISH, gene expression microarray, SNP microarray, 2D-DIGE, MS, LC-MS, etc. We have completed the modules for the IHC and FISH assay data, and a proof-of-principle has been implemented for gene expression microarrays (c.f. Figs. 4 and 5).

### 3.2.2. Temporal information definition and presentation

All attributes are tagged with one of the three defined temporal properties. All temporal information, including static information, can be viewed in the ISIV (c.f. Fig. 7). Within the ABB users have the ability to define “time filters” for the different columns in the display (c.f. Fig. 6).

### 3.2.3. Example data module definition

Table 1 is a simplified illustration of three defined data modules. Thus, “Ethnic Group” is a simple (single attribute) data

module allowing multiple values for a given subject, and is “Static”. “Alcohol Use” is a complex data module with three facets, and covers an “Interval”. “Surgical Proc(EDURE)” is a module similar to “Alcohol Use” but is an “Event”.

## 3.3. DW4TR interfaces—ABB and ISIV

### 3.3.1. ABB

This interface allows the user to analyze the data by dynamically creating a multidimensional pivot view. The rows are categorical data fields of interest, in a hierarchical structure. The columns can be data elements of either a numerical or categorical nature, in a flat or hierarchical structure. Although the view is a two dimensional table, the user can effectively explore data of theoretically unlimited dimensions by simply building multiple levels of a hierarchical data structure in both the rows and columns, and expanding them. The number of dimensions (levels of hierarchy) is only practically limited by the performance of the system (see Section 3.6). Fig. 4 (Panels A and B) shows screenshots of how an ABB view of five hierarchical levels was created and used to explore one subclass of breast cancer patients among CBCP Caucasian American subjects, who have triple-negative tumors, i.e., ER (estrogen receptor), PR (progesterone receptor), and HER2 (human epidermal growth factor receptor 2) negative.

Using the ABB, simple analyses such as mean, standard deviation, counts and percentages can be performed. Views on subject counts or specimen counts can be created. The results of interest can be printed, or exported to an Excel spreadsheet for additional analysis using other specialized software. A subject set of interest can also be saved as a cohort, for example the triple-negative Caucasian American breast cancer group shown in Fig. 4B, and used in a subsequent study or analyzed with a different application. The cohort can be analyzed using the ISIV described in the next section, or rendered to additional complicated analysis that the InforSense Analytical Workflow platform supports (not shown). The cohort data can also be exported to an Excel spreadsheet for analysis by specialized software outside of the DW4TR.

ABB offers many other features including a unique custom data binning capability. Multiple binning methods can be used, either automatically or manually, to create bins to explore data elements that contain either numerical or categorical values. For different studies, a user may need to use the same information in different ways. As shown in Fig. 4 (Panels C and D), the ethnic group attribute contains more than a dozen different possible values from CBCP and GDP. If the user wants to concentrate on Caucasians, African Americans, and Asian Americans, he/she can create three bins for them, and place the rest in “Other”. A histogram of the binning can then be viewed. Custom data binning can be done with categorical data as well as numerical data, and in our practice viewing histograms helps to adjust the binning strategy to ensure that there are a sufficient number of records available in each bin.

One type of molecular data can be directly accessed through ABB is IHC. Fig. 4B shows data for three molecules (ER, PR and HER2) which are routinely assayed in the CBCP, for which there are established clinical standards for clinical IHC [76,77]. Two other molecules, Ki67 and p53, are also routinely assayed though there are no current clinical standards for these molecules. In CBCP, all five molecules are characterized by protein expression assayed by IHC, and HER2 is further characterized with gene copy number ratio assessed using FISH as needed. Both IHC and FISH data types are supported in the production system. The IHC submodule contains Result, Percentage, Intensity, and Staining Pattern for nuclear proteins (ER and PR) or Result and Score for membrane-bound proteins (HER2). The FISH submodule contains Result and Ratio. Both submodules belong to “Pathologic Results” of “Biopsy” of “Diagnostic” in the data model. Specific business rules are applied in

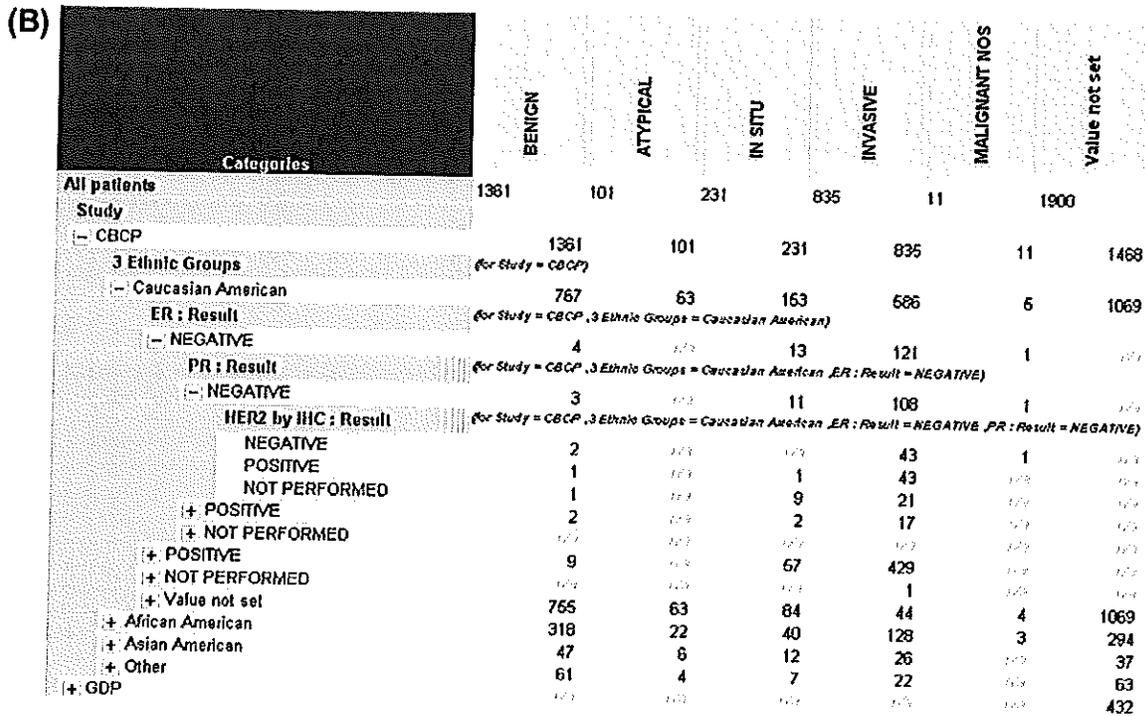
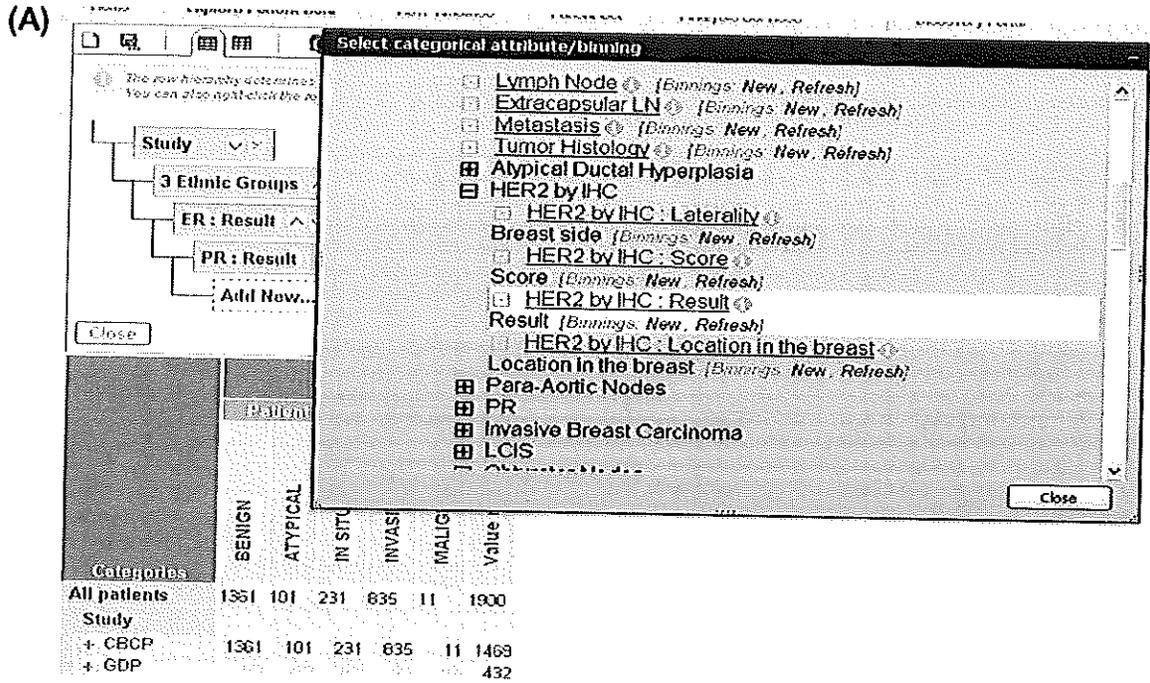


Fig. 4. Screenshots of ABB. (A) View construction. The column is 'Patient Pathology Category', and the rows are, hierarchically, 'Study', '3 Ethnic Group' (binned from 'Ethnic Group' as shown in Panels C and D), 'ER Result', 'PR Result', and the screen was captured when 'HER2 by IHC: Result' was being selected. The inset shows a portion of the data module where the data element resides. (B) View exploration. Rows of five levels were explored, and at the end the numbers are shown for triple negative (ER-, PR- and Her2-) subjects for CBCP Caucasian American in each pathology category. Similar analysis can be done for other ethnic groups. (C) The custom data binning feature. Here a manual binning option is selected. Bins created are 'Caucasian American', 'African American', 'Asian American', and 'Other', with Null defaulted. Available values are in the window of 'Unassigned Values', which can be dragged into the bins on the left to complete the binning. (D) Histogram of the binning result, which can be viewed during custom data binning or at the time of use.

the ETL process, for example the final HER2 expression result is determined by IHC Score (0 or 1+ as negative, and 3+ as positive), but when IHC Score is ambiguous (2+) the Ratio of the FISH assay is used, with  $\leq 1.8$  as negative and  $\geq 2.2$  as positive, and an intermediate ratio is considered as ambiguous.

A proof-of-principle model of microarray-based molecular data is in our test system; this feature is the only feature described in Section 3 that has not yet been applied to the production system. The MIAME standards [26] are followed with some modification, for example, our "sample ID" eliminates the need for subject and

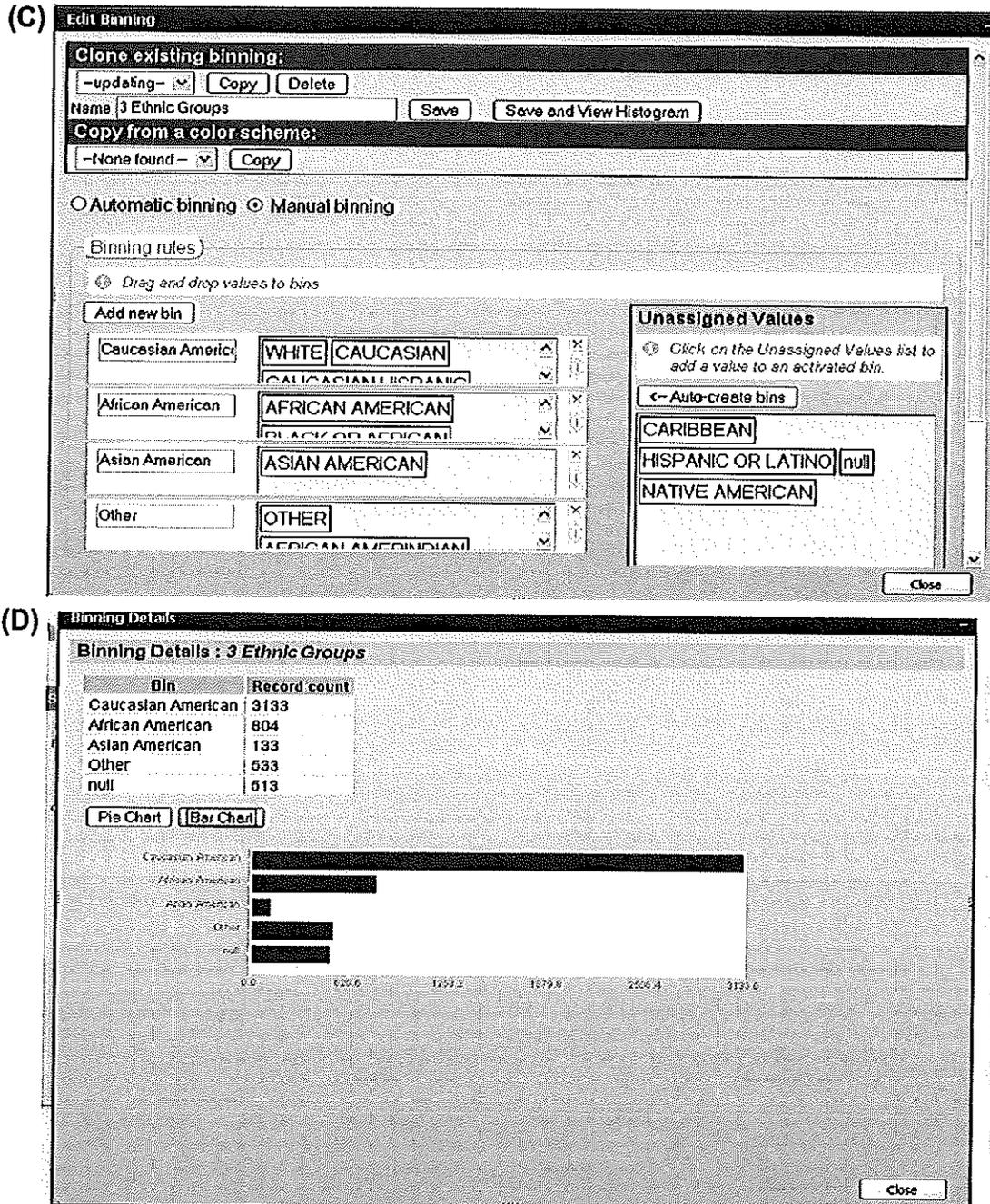


Fig. 4 (continued)

specimen information. The metadata are modeled, including sample ID, bio-molecule information (lab protocol, 260/280 ratio, RIN, etc.), array information (platform, array ID/batch number, etc.), experimental information (project name, PI, lab operator, date, etc.), and results (call rate, 3'/5' ratio, Pass/Fail flag, data file location, etc.). Metadata are accessible via ABB for selecting microarray data of interest in a cohort defined by clinicopathologic characteristics, but we saved the array raw data files in a file server due to their size and the fact that their analysis typically requires specialized software that read in the raw data file directly. Fig. 5 shows an example how a set of microarray data can be identified from previous experiments performed on samples from

a cohort of interest. This information may lead to a secondary use of the microarray data, or help the researcher to design a new experiment.

ABB also contains a utility "Time Filter" to apply temporal information in the analysis of aggregated information. The time filters are used to limit or constrain the data that are returned much like any other type of filter criteria. Time filters can be created based on absolute dates, patient age, or relative to other events. Fig. 6 shows one example how "Alcohol Usage" was analyzed for CBCP subjects when they were at different age ranges. "Alcohol Usage" is an Interval temporal data type with values self-reported by subjects covering different periods of time in patient's life. Note that direct

Categories	Patient Count						Value not set
	ATYPICAL	BENIGN	IN SITU	INVASIVE	MALIGNANT NOS		
All entities	4896	114	1607	301	1286	12	1620
2 Ethnic Groups							
- Caucasian American	3270	69	922	203	914	8	1183
Menopausal Status (for 2 Ethnic Groups = Caucasian American)							
- POST-MENOPAUSAL	1348	30	230	103	522	5	473
Microarray Experiment : Platform (for 2 Ethnic Groups = Caucasian American, Menopausal Status = POST-MENOPAUSAL)							
AFFY GENECHIP U133 PLUS2	138		3		76		61
Value not set	1210	30	227	103	446	5	412
- PRE-MENOPAUSAL	1283	24	484	66	235	3	480
Microarray Experiment : Platform (for 2 Ethnic Groups = Caucasian American, Menopausal Status = PRE-MENOPAUSAL)							
AFFY GENECHIP U133 PLUS2	46		2		19		26
Value not set	1237	24	482	66	216	3	454
+ STATUS POST HYSTERECTOMY	343	13	78	26	97		133
+ SURGICALLY MENOPAUSAL	379	11	86	25	126		140

Fig. 5. Screenshot of ABB showing exploration of available gene expression microarray data on biospecimens of a CBCP cohort of interest. Subjects were Caucasian American, post-menopausal or pre-menopausal, and experiments were performed on an Affymetrix platform. Subjects were mostly invasive breast cancer patients and normal subjects (as represented by 'Value Not Set' in the last column). The numbers from each pathology category do not completely add up to the total patient count on the left-most column since a couple of patients changed the category status with time. Highlighted row of subjects (light blue) can be saved as a cohort for further studies.

custom-binning of subject's age would not be useful to this analysis.

### 3.3.2. ISIV

This interface is for viewing and analysis of detailed subject information, most importantly the temporal-related information. It takes subject IDs as the input either by direct entry, or selecting from a pre-defined cohort. Events and Intervals of interest can be selected, and applied to the whole subject set. Information can be viewed independently for every subject, or aligned across the subject set on Events of interest to enable comparison between subjects. Static information can be displayed as well, so all the information about the subject can be studied. An example screenshot of ISIV is shown in Fig. 7.

Both the ABB and ISIV require minimal training before one can start to use them. The average training time for a clinician or scientist is 15 min, and the user can learn more features by using the system. Interestingly, for a high school student the average training time is only 10 min.

### 3.4. Direct data extraction from the database

The raw data in the DW4TR can be directly queried from the database, as a complete or partial dataset depending on the research need. The data can be presented in either a flat file structure or in Excel spreadsheet, for use with specialized data analysis software such as SAS, SPSS, or R. Such queries have been developed for and applied to both CBCP and GDP data. Direct data extraction requires a database administrator and is not the focus of this paper, thus is only briefly described here.

### 3.5. Extensibility of the DW4TR

The DW4TR was developed to support the CBCP with a vision to support translational research in general. This extensibility was tested and validated by expanding the system to support the GDP. Its flexibility was tested by our successful revision of the model commanded by the consolidation of GDP questionnaires and data elements. A number of disease-independent attributes and submodules developed for the CBCP implementation were re-used for the GDP instance, for example demographics and elements of past medical history. Many other attributes/submodules were modified to satisfy the requirements of GDP, for example alcohol usage. For GDP-specific data elements, additional modules were developed. In addition, the GDP contains a psychology questionnaire and a food questionnaire that are program-specific but not disease-specific. These were developed and structured to the overall patient-centric clinical data model, and could potentially be re-used for future programs. In addition, the GDP has different data security and accessibility requirements, and we have developed additional mechanism for it as detailed in Section 4. With the experience gained in supporting the GDP, we are confident that we will succeed in the current further extension of the system to support the third translational research program for the consortium led by Thomas Jefferson University.

### 3.6. System performance

To evaluate the performance potential of the DW4TR, we conducted a stress test using the CBCP data in a one-server system configuration. The test was done from a user's perspective, by

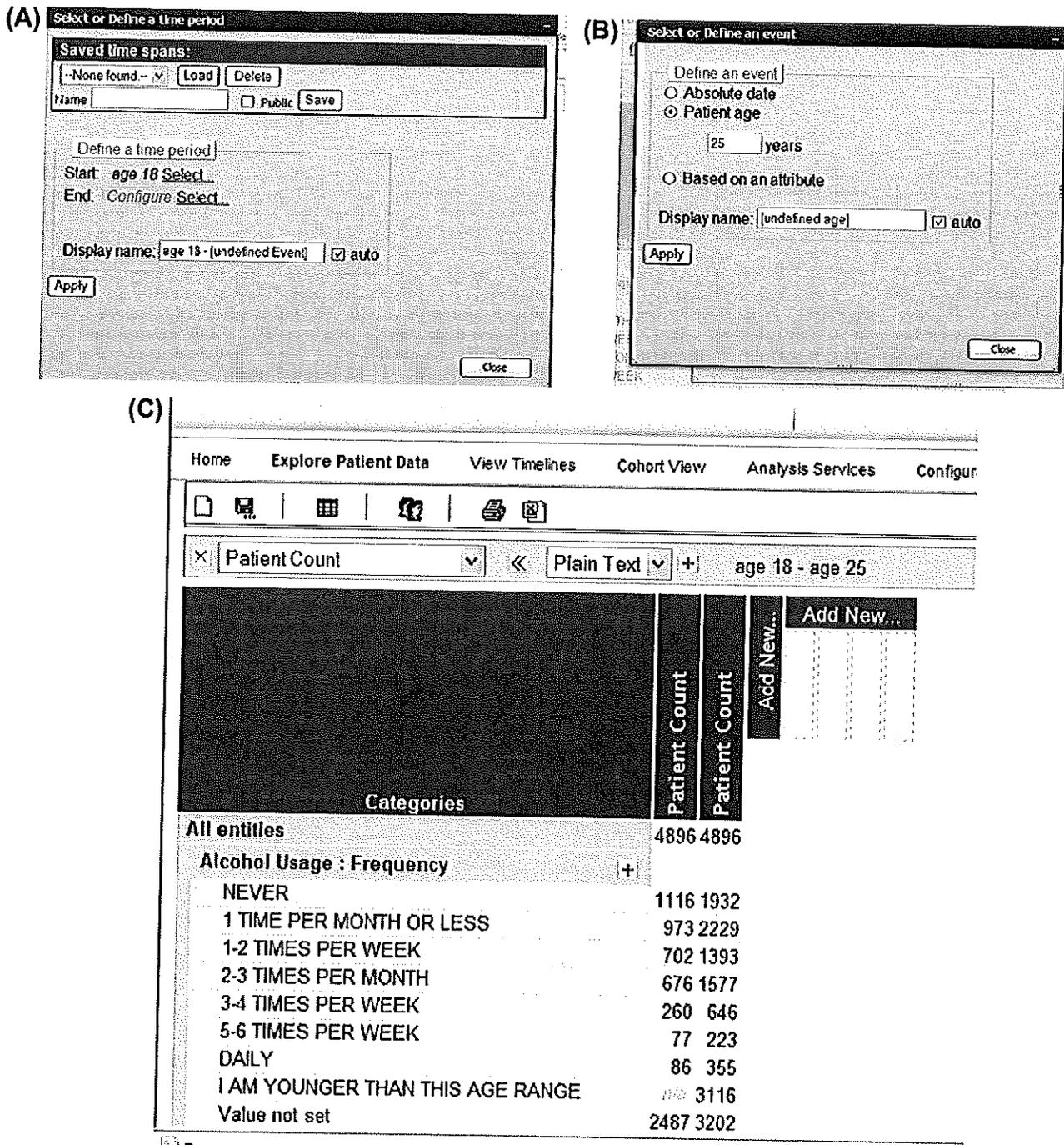


Fig. 6. Use of temporal information in ABB to study data of interests. (A) Screenshot showing the use of the 'Time Filter' to define the starting and ending time for study, with starting time already being configured but not the ending time. (B) Configuring the ending time based on Age, but it can also be based on Absolute Date or an Attribute. (C) Example showing the alcohol usage history across the CBCP subjects. The first column shows the usage frequencies of subjects when they were between 18–25 years of age, and the second column shows the usage frequencies across the whole population. Additional age ranges can be configured to show the corresponding alcohol usage habit.

measuring how long it takes for the data to completely display in a hierarchical view for the first time—note that subsequent display of the same data is much faster. The view is composed of three levels of Pathology Category, Ethnicity, and Body-Mass Index (BMI, with custom-binning) in the rows and record counts in the columns. Starting from the then patient number of 4234, we artificially replicated the records in the database to 8468, 16,936, 33,872, and 67,744 without performing a database tune-up other than Oracle table analysis. It took one second or less, for the level

1 (Pathology Category) and level 2 (Ethnicity) data to be displayed in all these tests. For level 3 (BMI), it took 2, 4, 12, 39, and 134 s respectively to display the data when the number of records doubled stepwise from 4234 to 67,744. Given that the current CBCP annual subject enrollment is about 600, it will take another 20 years before 16,936 subjects are enrolled when 12 s are needed to display 3 levels of data. We expect that the system performance will be further dramatically improved with hardware upgrades and Oracle tuning.

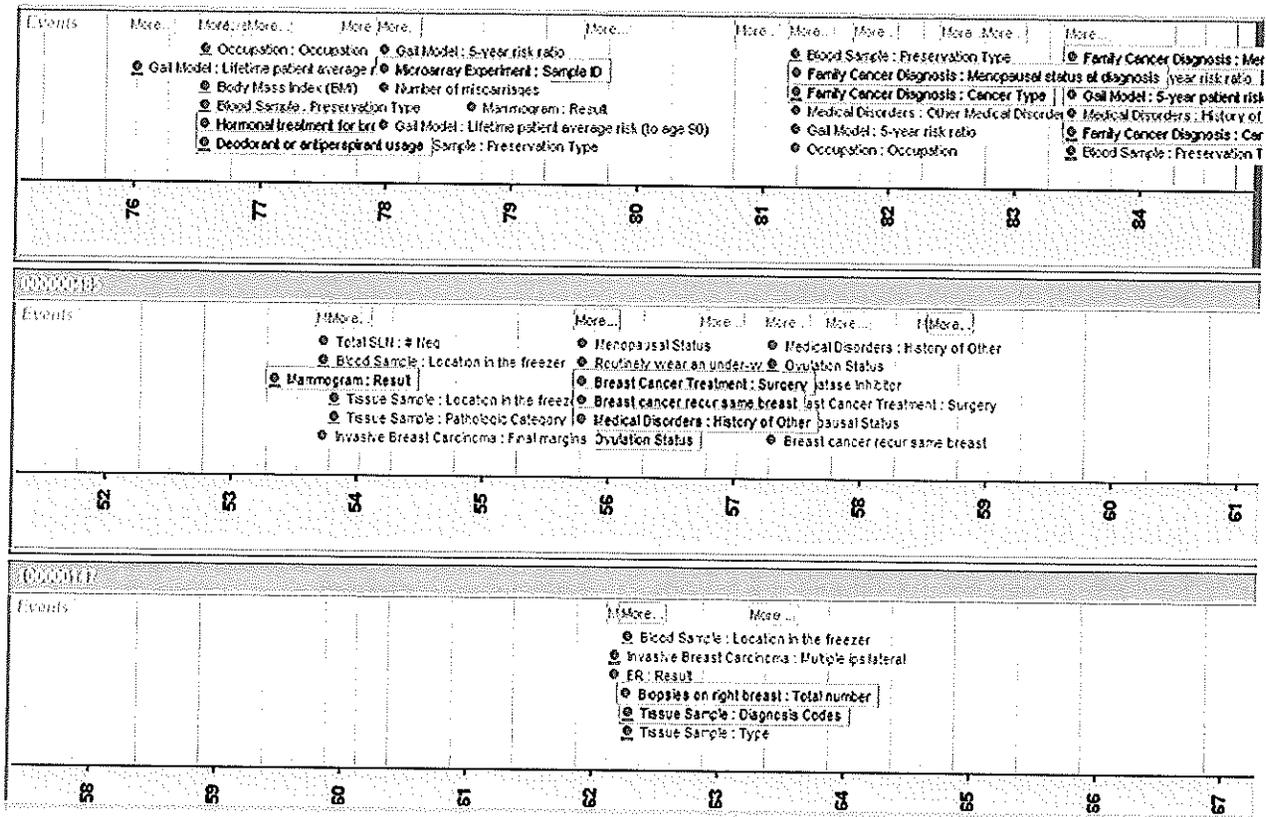


Fig. 7. Screenshot of the ISIV. Temporal information from three subjects is shown. Users can zoom in or out, align the information on a specific event for cross-subject comparison, and explore more detailed information by clicking on the event of interest. The temporal axis shows the age but can be toggled to calendar year.

Table 1  
Example data module definition.

ID	Attribute	Facet	Data type	Units	Multiple values	From date	To date	Temporal information	Parent ID
23	Ethnic Group		VARCHAR2		Yes			Static	
103	Alcohol Use		COMPLEX		No	mm/dd/yyyy	mm/dd/yyyy	Interval	
104	Alcohol Use	Name	VARCHAR2		No	mm/dd/yyyy	mm/dd/yyyy	Interval	103
105	Alcohol Use	Frequency	NUMBER	/day	No	mm/dd/yyyy	mm/dd/yyyy	Interval	103
106	Alcohol Use	Amount	NUMBER	drinks	No	mm/dd/yyyy	mm/dd/yyyy	Interval	103
1541	Surgical Proc		COMPLEX		No	mm/dd/yyyy		Event	
1542	Surgical Proc	Name	VARCHAR2		No	mm/dd/yyyy		Event	1541
1543	Surgical Proc	Code	VARCHAR2		No	mm/dd/yyyy		Event	1541
1544	Surgical Proc	Laterality	VARCHAR2		No	mm/dd/yyyy		Event	1541

3.7. Example use cases

The DW4TR has been playing an important role in supporting translational research across user groups in both WRI and WRAMC. Here we report four example use cases to demonstrate the utility of the system.

3.7.1. The DW4TR enables clinicians to study the integrated data directly

When first exposed to the DW4TR, CBCP clinical users appreciated the simple interface which gave them their first opportunity to directly study the integrated clinicopathologic and biospecimen data. They were happy to see that the results shown on the ABB confirmed their qualitative clinical observations. In one short session of less than 1 h, the program PI identified several lines of evidence which subsequently led to an abstract accepted to the San Antonio Breast Cancer Symposiums [80]. The system has since

been further improved in both the data model structure and the user interface.

3.7.2. The DW4TR enables enhanced cohort and biospecimen selection

Before the DW4TR was developed, cohort and biospecimen selection for any research project was a major task involving multiple data sources including: Core Questionnaires, Pathology Checklists, and IHC assay reports, in the form of an Oracle database, a Microsoft Access database, Excel spreadsheets, and occasionally hard copy reports. Many manual steps were involved, and the procedure was not standardized. The selection of biospecimens was especially challenging when there were temporal restrictions, for example identifying blood samples drawn in a subject cohort before certain kinds of invasive procedures were performed. This requirement alone could take an experienced researcher several days since multiple questionnaires are involved and a number of surgical procedures could be performed on one subject and that there were 15

**Table 2**  
Example classifications of breast neoplasms in MeSH, SNOMED, and DW4TR.

Name	Classifications
MeSH1	3. Diseases (C) → Skin and Connective Tissue Diseases [C17] → Skin Diseases [C17.800] → Breast Diseases [C17.800.090] → Breast Neoplasms [C17.800.090.500]
MeSH2	3. Diseases (C) → Neoplasms [C04] → Neoplasms by Site [C04.588] → Breast Neoplasms [C04.588.180]
SNOMED1	SNOMED CT Concept (SNOMED RT + CTV3) → Clinical finding (finding) → Disease (disorder) → Disorder by body site (disorder) → Disorder of body system (disorder) → Disorder of breast (disorder) → Neoplasm of breast (disorder)
SNOMED2	SNOMED CT Concept (SNOMED RT + CTV3) → Clinical finding (finding) → Finding by site (finding) → Finding of body region (finding) → Finding of trunk structure (finding) → Finding of region of thorax (finding) → Breast finding (finding) → Disorder of breast (disorder) → Neoplasm of breast (disorder)
DW4TR	Medical History → Past Medical History → Major Adult Illnesses → Breast Cancer

different kinds of such procedures. Using the DW4TR interface and additional utilities, the time needed to perform such tasks has now been reduced to minutes. Currently any cohort and biospecimen selection can be done in hours instead of days or weeks.

### 3.7.3. The DW4TR serves as a research environment for risk factor assessment

The DW4TR provides an effective research environment for cancer risk factor analysis. Several studies examining breast cancer risk factors have been done using this system [81–85], and several new studies are underway.

### 3.7.4. The DW4TR facilitates virtual experimental studies

Virtual experimental studies can be performed as integrated information associated with completed research projects is available. Using the information stored in the DW4TR and the file server, we identified the gene expression microarray data from three previous projects, on two types of specimens from three subject groups. From these datasets we were able to create two virtual experiments, and both studies were accepted to leading scientific conferences [86,87].

## 4. Security and accessibility

Data security and access control are important factors in the design of the DW4TR, particularly as our system can be expanded to include multiple research studies. The use of clinical data in translational research is governed by HIPAA compliant IRB approved protocols, and typically cannot be unconditionally shared. Thus we have designed and developed a set of data security and access solutions to satisfy such requirements.

The first layer of security for the system is data de-identification. Prior to being loaded into the DW4TR, PHI is stripped away from the collected data. At this point subjects will have been assigned a unique research subject ID for linking together subject information across multiple sources. The next level of security makes use of the "Virtual Private Database" (VPD) feature of Oracle, which allows for row level access control within the DW. While one of the strengths of the system is the ability to perform cross-study analysis, in many instances it is necessary to restrict data visibility among different users of the system. In our security model, all users are assigned roles which dictate the data they are permitted to access. The VPD system is designed such that every data element in the DW is tagged with a study identifier that is associated with the different user roles, enabling data access control. The third level of security is at the application tier. All access to the DW occurs through a password protected web portal that is served by an application server. The final layer of security is a network firewall and Virtual Private Network (VPN) for which the 168-bit encryption Secure Sockets Layer VPN from Juniper Networks Inc. is deployed. All components of the system, including the DW4TR and the web portal, are protected behind the local network firewall and are only accessible from either within the Institute itself, or by first establishing a VPN connection to the network.

These security and access control measures were developed based on the needs of the supported programs, and can be selectively applied to individual programs. For the two programs currently supported by the DW4TR, CBCP protocols allow data access by all CBCP researchers, and VPN access to the DW4TR is granted with written permission of the CBCP program PI. For GDP, aggregated biomedical information is allowed for use by all the GDP researchers, but the detailed subject information is only allowed to the originating clinical sites unless additional paperwork is completed. Thus, we have designed and developed a role-based site-specific privilege system to enable GDP users to use ABB and the ISIV. DW4TR can also be configured to disable functions and utilities of choice. Currently at the request of the GDP PI, we have disabled the ISIV functions for GDP users so that the GDP data are only available through ABB to GDP users for aggregated information. When a GDP researcher identifies a cohort of interest, additional data use agreements will be completed to request individual subjects' information from the corresponding clinical sites, for approval by the involved sites and the GDP program PI. With a written request from the GDP program PI, WRI will supply the detailed data from the DW4TR to the researcher.

## 5. Discussion

We have taken a bottom-up approach to develop a DW4TR to support translational research programs. The system is extensible, integrating internally generated clinicopathologic and molecular data based on a light-weight ontology, with a patient-centric, modularly structured clinical data model supplemented with a specimen-centric molecular data model. Two interfaces have been developed targeting non-informatician end users, with the ABB supporting aggregated data analysis and ISIV supporting single subject information analysis. The system enables study cohort and biospecimen selection and is capable of handling temporal relationships between clinical data and biospecimen data collected at multiple points in time.

The development of our data model relies on using or creating ontologies for clinical and molecular data types. We will focus this discussion on clinical data ontology. In fact, for molecular data types, the scale of work is much smaller and the existing ontologies are mostly adequate. For example, we found that the MIAME standard met our needs for modeling gene expression microarray metadata. Our analysis indicated that existing ontologies for clinical data cannot be readily adapted to satisfy our specific research needs, due to gaps between existing ontologies and our needs as well as overlaps and inconsistencies between existing standards and ontologies [70,71]. For example, the CBCP Pathology Checklist contains 372 data elements including 131 breast pathology conditions. When the data form was developed and revised, the AJCC (American Joint Committee on Cancer) guidelines and ASCO/CAP (American Society of Clinical Oncology/College of American Pathologists) guidelines were followed; these guidelines, like many other guidelines, have evolved over the years [88–91]. Many data elements present in the CBCP lack standard descriptions, for example

for many benign diseases. Some biomarkers used in the CBCP (Ki67 and p53) do not even have a standard clinical IHC protocol. Therefore, descriptions or definitions of such data elements were developed based on medical textbooks and pathologist's clinical practice for the Pathology Checklist, which we modeled ourselves.

There are classifications in existing standards/ontologies that we find cumbersome or which do not meet our needs [92]. We show as an example in Table 2 the classifications of "Breast Neoplasms" by two existing ontologies. In MeSH, it is classified under "Skin Diseases". This classification (named "MeSH1" in Table 2) is not available in the SNOMED-CT classifications under either Skin Diseases or Connective Tissue Diseases. In MeSH, "Breast Neoplasms" is also classified under "Neoplasms" (in a way we agree with, "MeSH2" in Table 2). However, in the two classifications, "Breast Neoplasms" have different codes, which may cause a problem in the practical applications of data modeling. In SNOMED-CT, the term "Neoplasm of Breast" is used, and there are several classifications leading to it with relatively long paths. We show two of them, "SNOMED1" and "SNOMED2" in Table 2. In our pragmatic approach to support the CBCP, applying such a complete classification is cumbersome. In addition, reconciliation of discrepancies among the available ontologies and our understanding of the problem, for all the data elements in the CBCP, is a major task and not one of our immediate project needs. Therefore, we have developed a classification path for "Breast Cancer", which is more light-weight but adequate to our needs (see "DW4TR" in Table 2).

Finally, some questions in the questionnaires of the two programs we currently support do not comply with existing standards, but we are still required to model them to enable proper storage and presentation of the data. Such problems have been reported to program PIs and were taken into consideration, for example in the consolidation of the GDP questionnaires. In addition, existing ontologies are not static, for example MeSH is currently updated weekly and SNOMED-CT is updated monthly. Thus, after analyzing what is available and what we need to support, we determined that our best approach was to design a light-weight ontology incorporating existing standards and ontologies where possible and to map our classification system to existing ones at a later time. The mapping table(s) will serve as the buffer between the two moving targets which will ensure transportability and un-interrupted use of the system. We will also supplement those standards with new ontologies and common data elements that we define.

The DW4TR currently supports limited number of molecular data types and we will continue to develop data models to support other molecular study platforms, including SNP microarray and array CGH data. We expect that the method developed for gene expression microarray will in general be applicable to these array-based technologies. We will explore support for RT-PCR, tissue microarrays, mass spectrometry, and Next Generation sequencing technologies. These will depend on the needs of the users and programs that we support.

Currently a prototype has been developed to enable access to medical and molecular images through the ISIV, including digital/digitized mammograms and gene expression microarray images. A thumbnail of the image is stored in the database of the DW with searchable annotations. The full-size image is stored in a file server instead of the DW4TR database for performance considerations. We plan to follow the DICOM standard for medical images and will explore how to best annotate molecular images. Additional images we need to support include, pathology H&E slide images, biomarker IHC images, and tissue microarray images. Note that all the medical images have to be de-identified for research purpose.

Currently the DW4TR does not support EMR for data input. We have a stepwise plan to begin using EMR records for CBCP subjects and we are looking at subject de-identification, data extraction, and data model expansion. Although we do not plan to take a

systematic "Enterprise" approach, we will learn from such approaches as used by I2B2 and others. When EMRs are used as an input, a number of functions need to be integrated into the system including subject de-identification and information extraction, for which the I2B2 development teams have made good progresses [50,54,93]. While I2B2 clearly plays an important role in the study design phase including addressing de-identification and natural language processing, it was not designed to be a comprehensive solution that can provide the level of detailed access necessary for later stage research [50–54]. The ABB described here provides the same ability to generate aggregate statistics for population stratification and cohort selection, but by doing this in a web-based pivot table interface it also facilitates trend identification and discovery in a way traditional query interfaces cannot. From a data modeling perspective, the ABB also allows for more sophisticated handling of complex, multi-faceted attributes whereas I2B2 only offers aggregation around the single pivot point of the patient. I2B2 makes good use of ontologies, but has not yet developed an ontology with the level of detail described here [54,94]. Furthermore, by design I2B2 does not enable access to row level clinical data and detailed clinical timeline provided by the ISIV [54,94].

In conclusion, we have developed the DW4TR to support a breast cancer translational research program and then expanded it to support a gynecological disease translational research program. It is based on a light-weight ontology and equipped with an interface capable of handling temporal information, designed for use by non-informatician specialists including clinicians and laboratory scientists. The system supports both *in silico* and *in vitro* studies. With its proven extensibility, we believe that the DW4TR will play an important role in translational research across multiple disease studies.

## Acknowledgments

We thank Mr. Anton Oleynikov, Mr. Faruk Cay, Mr. Raveen Sharma, Dr. Csaba Mihaly, Mr. Eric Babyak, and Dr. Yonghong Zhang, for their help in this project. We thank Ms. Joni Kohr, Dr. Hallgeir Rui, and Mr. John Eberhardt for their help in manuscript preparation and revision. We also thank the anonymous reviewers for their constructive comments that resulted in enhanced quality of this paper. This work was supported by the Clinical Breast Care Project, and the Gynecological Disease Program, with funds from the US Department of Defense through Henry Jackson Foundation for the Advancement of Military Medicine, Rockville, MD. The views expressed in this paper are those of the authors and do not reflect the official policy of the Department of the Army, Department of Defense, or the government of the United States.

## References

- [1] Hu H, Mural RJ, Liebman MN, Hu H, Mural RJ, Liebman MN, et al. Biomedical informatics in translational research. Boston, London: Artech House; 2008.
- [2] Bernstam EV, Hersh WR, Johnson SB, Chute CG, Nguyen H, Sim I, et al. Synergies and distinctions between computational disciplines in biomedical research: perspective from the Clinical and Translational Science Award programs. *Acad Med* 2009;84:964–70.
- [3] Payne PR, Embi PJ, Sen CK. Translational informatics: enabling high-throughput research paradigms. *Physiol Genomics* 2009;39:131–40.
- [4] Sarkar IN. Biomedical informatics and translational medicine. *J Transl Med* 2010;8:22.
- [5] Nadkarni PM, Reeders ST, Zhou J. CECIL: a database for storing and retrieving clinical and molecular information on patients with Alport syndrome. In: *Proc annu symp comput appl med care*; 1993. p. 649–53.
- [6] Hu H, Brzeski H, Hutchins J, Ramaraj M, Qu L, Xiong R, et al. Biomedical informatics: development of a comprehensive data warehouse for clinical and genomic breast cancer research. *Pharmacogenomics* 2004;5:933–41.
- [7] Brammen D, Katzer C, Rohrig R, Weismuller K, Maier M, Hossain H, et al. An integrated data-warehouse-concept for clinical and biological information. *Stud Health Technol Inform* 2005;116:9–14.

- [8] Murphy SN, Mendis ME, Berkowitz DA, Kohane I, Chueh HC. Integration of clinical and genetic data in the i2b2 architecture. In: AMIA annu symp proc; 2006. p. 1040.
- [9] Hu H, Correll M, Osmond M, Gao J, Oleynikov A, Sheldon J, et al. A clinical data warehouse to support translational research. In: 15th annual international conference on Intelligent Systems for Molecular Biology (ISMB), Vienna, Austria; July 21–25, 2007.
- [10] Burgun A, Bodenreider O. Accessing and integrating data and knowledge for biomedical research. *Yearb Med Inform* 2008;91–101.
- [11] Rossille D, Burgun A, Pangault-Lorho C, Fest T. Integrating clinical, gene expression, protein expression and preanalytical data for in silico cancer research. *Stud Health Technol Inform* 2008;136:455–60.
- [12] Hu H. Data centralization. In: Hu H, Mural RJ, Liebman MN, editors. *Biomedical informatics in translational research*; 2008. p. 141–71 [chapter 8].
- [13] Lowe HJ, Ferris TA, Hernandez PM, Weber SC. STRIDE – an integrated standards-based translational research informatics platform. In: AMIA annu symp proc; 2009. p. 391–5.
- [14] Inmon WH, Kelley C. *Developing the data warehouse*. QED Publishing Group; 1993.
- [15] Kimball R, Ross M. *The data warehouse Toolkit*. 2nd ed. John Wiley and Sons; 2002.
- [16] Stead WW, Hammond WE, Straube MJ. A chartless record – is it adequate? *J Med Syst* 1983;7:103–9.
- [17] Niedner CD. The entity-attribute-value data model in radiology informatics. In: *Proceedings of the 10th conference on computer applications in radiology*, Anaheim, CA; 1990.
- [18] Nadkarni PM, Brandt C. Data extraction and ad hoc query of an entity-attribute-value database. *J Am Med Inform Assoc* 1998;5:511–27.
- [19] Codd F, Codd SB, Salley CT. *Providing OLAP (Online Analytical Processing) to user-analysts: an IT mandate*. San Jose: Codd & Date, Inc; 1993.
- [20] Rubin DL, Shah NH, Noy NF. Biomedical ontologies: a functional perspective. *Brief Bioinform* 2008;9:75–90.
- [21] Health Level 7. <<http://www.hl7.org/>>.
- [22] Lowe HJ, Barnett GO. Understanding and using the medical subject headings (MeSH) vocabulary to perform literature searches. *JAMA* 1994;271:1103–8.
- [23] SNOMED Clinical Terms® (SNOMED CT®). <[http://www.nlm.nih.gov/research/umls/Snomed/snomed\\_main.html](http://www.nlm.nih.gov/research/umls/Snomed/snomed_main.html)>.
- [24] DICOM – Digital Imaging Communications in Medicine. <<http://www.medical.nema.org/>>.
- [25] Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The gene ontology consortium. *Nat Genet* 2000;25:25–9.
- [26] Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, et al. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat Genet* 2001;29:365–71.
- [27] BioPAX: Biological Pathways Exchange. <<http://www.biopax.org/>>.
- [28] Sioutos N, de Coronado S, Haber MW, Hartel FW, Shaiu WL, Wright LW. NCI Thesaurus: a semantic model integrating cancer-related clinical and molecular information. *J Biomed Inform* 2007;40:30–43.
- [29] Rubin DL, Lewis SE, Mungall CJ, Misra S, Westerfield M, Ashburner M, et al. National Center for Biomedical Ontology: advancing biomedicine through structured organization of scientific knowledge. *OMICS* 2006;10:185–98.
- [30] Cote RG, Jones P, Apweiler R, Hermjakob H. The Ontology Lookup Service, a lightweight cross-platform tool for controlled vocabulary queries. *BMC Bioinform* 2006;7:97.
- [31] Lindberg DA, Humphreys BL, McCray AT. The unified medical language system. *Methods Inf Med* 1993;32:281–91.
- [32] Bard JB, Rhee SY. Ontologies in biology: design, applications and future challenges. *Nat Rev Genet* 2004;5:213–22.
- [33] Yu AC. Methods in biomedical ontology. *J Biomed Inform* 2006;39:252–66.
- [34] Cimino JJ, Zhu X. The practical impact of ontologies on biomedical informatics. *Yearb Med Inform* 2006;124–35.
- [35] Weedon-Fekjaer H, Lindqvist BH, Vatten LJ, Aalen OO, Trelli S. Breast cancer tumor growth estimated through mammography screening data. *Breast Cancer Res* 2008;10:R41.
- [36] Shoham Y. *Reasoning about change: time and causation from the standpoint of artificial intelligence*. Cambridge, MA, USA: MIT Press; 1988.
- [37] Shahar Y. A framework for knowledge-based temporal abstraction. *Art Intell* 1997;90:79–133.
- [38] Deshpande AM, Brandt C, Nadkarni PM. Temporal query of attribute-value patient data: utilizing the constraints of clinical studies. *Int J Med Inform* 2003;70:59–77.
- [39] Ceusters W, Steurs F, Zanstra P, Van Der Haring E, Rogers J. From a time standard for medical informatics to a controlled language for health. *Int J Med Inform* 1998;48:85–101.
- [40] Das AK, Shahar Y, Tu SW, Musen MA. A temporal-abstraction mediator for protocol-based decision-support systems. In: *Proc annu symp comput appl med care*; 1994. p. 320–4.
- [41] Shahar Y, Das AK, Tu SW, Kraemer FB, Musen MA. Knowledge-based temporal abstraction for diabetic monitoring. In: *Proc annu symp comput appl med care*; 1994. p. 697–701.
- [42] Nguyen JH, Shahar Y, Tu SW, Das AK, Musen MA. A temporal database mediator for protocol-based decision support. In: *Proc AMIA annu fall symp*; 1997. p. 298–302.
- [43] Post AR, Harrison Jr JH. PROTEMPA: a method for specifying and identifying temporal sequences in retrospective data for patient selection. *J Am Med Inform Assoc* 2007;14:674–83.
- [44] Post AR, Sovarel AN, Harrison JH, Jr. Abstraction-based temporal data retrieval for a Clinical Data Repository. In: AMIA annu symp proc; 2007. p. 603–7.
- [45] Chen J, Zhao P, Massaro D, Clerch LB, Almon RR, DuBois DC, et al. The PEPR GeneChip data warehouse, and implementation of a dynamic time series query tool (SGQT) with graphical interface. *Nucleic Acids Res* 2004;32:D578–81.
- [46] Yamamoto Y, Namikawa H, Inamura K. Development of a time-oriented data warehouse based on a medical information event model. *Igaku Butsuri* 2002;22:327–33.
- [47] Informatics for Integrating Biology and the Bedside (I2B2). <<https://www.i2b2.org/>>.
- [48] Mendis M, Wattanasin N, Kuttan R, Pan W, Phillips L, Hackett K, et al. Integration of hive and cell software in the i2b2 architecture. In: AMIA annu symp proc; 2007. p. 1048.
- [49] Mendis M, Phillips LC, Kuttan R, Pan W, Gainer V, Kohane I, et al. Integrating outside modules into the i2b2 architecture. In: AMIA annu symp proc; 2008. p. 1054.
- [50] Uzuner O, Luo Y, Szolovits P. Evaluating the state-of-the-art in automatic identification. *J Am Med Inform Assoc* 2007;14:550–63.
- [51] Barrett N, Weber-Jahnke JH. Applying natural language processing toolkits to electronic health records – an experience report. *Stud Health Technol Inform* 2009;143:441–6.
- [52] Deshmukh VG, Meystre SM, Mitchell JA. Evaluating the informatics for integrating biology and the bedside system for clinical research. *BMC Med Res Methodol* 2009;9:70.
- [53] Meystre SM, Deshmukh VG, Mitchell J. A clinical use case to evaluate the i2b2 Hive: predicting asthma exacerbations. In: AMIA annu symp proc; 2009. p. 442–6.
- [54] Murphy SN, Weber G, Mendis M, Gainer V, Chueh HC, Churchill S, et al. Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). *J Am Med Inform Assoc* 2010;17:124–30.
- [55] The stanford translational research integrated database environment (STRIDE). <<http://clinicalinformatics.stanford.edu/STRIDE/>>.
- [56] Hernandez P, Podchyska T, Weber S, Ferris T, Lowe H. Automated mapping of pharmacy orders from two electronic health record systems to RxNorm within the STRIDE clinical data warehouse. *AMIA Annu Symp Proc* 2009;2009:244–8.
- [57] Lowe HJ, Huang Y, Regula DP. Using a statistical natural language parser augmented with the UMLS specialist lexicon to assign SNOMED CT codes to anatomic sites and pathologic diagnoses in full text pathology reports. In: AMIA annu symp proc; 2009. p. 386–90.
- [58] caIntegrator. <<https://cabig.nci.nih.gov/tools/caIntegrator/>>.
- [59] Repository of Molecular Brain Neoplasia Data (REMBRANDT). <<http://caIntegrator-Info.nci.nih.gov/rembrandt/>>.
- [60] Cancer Genetic Markers of Susceptibility (CGEMS). <<http://cgems.cancer.gov/>>.
- [61] Clinical and Translational Science Award (CTSA) on the website of National Center for Research Resources. <[http://www.ncrr.nih.gov/clinical\\_research\\_resources/clinical\\_and\\_translational\\_science\\_awards/](http://www.ncrr.nih.gov/clinical_research_resources/clinical_and_translational_science_awards/)>.
- [62] The Cancer Biomedical Informatics Grid (caBIG). <<http://cabig.nci.nih.gov/>>.
- [63] von Eschenbach AC, Buetow KH. Cancer informatics vision: caBIGTM. *Cancer Inform* 2006;2:22–4.
- [64] The Cancer Biomedical Informatics Grid (caBIG): infrastructure and applications for a worldwide research community. *Stud Health Technol Inform* 2007;129:330–4.
- [65] Covitz PA, Hartel F, Schaefer C, De Coronado S, Fragoso G, Sahni H, et al. caCORE: a common infrastructure for cancer informatics. *Bioinformatics* 2003;19:2404–12.
- [66] Komatsoufis GA, Warzel DB, Hartel FW, Shanbhag K, Chilukuri R, Fragoso G, et al. caCORE version 3: implementation of a model driven, service-oriented architecture for semantic interoperability. *J Biomed Inform* 2008;41:106–23.
- [67] McCusker JP, Phillips JA, Gonzalez Beltran A, Finkelstein A, Krauthammer M. Semantic web data warehousing for caGrid. *BMC Bioinform* 2009;10(Suppl. 10):S2.
- [68] The Clinical Breast Care Project. <[www.cbcp.info/](http://www.cbcp.info/)>.
- [69] The Gynecological Disease Program. <<http://www.gyndisease.org/index.html>>.
- [70] Rector AL, Zanstra PE, Solomon WD, Rogers JE, Baud R, Ceusters W, et al. Reconciling users' needs and formal requirements: issues in developing a reusable ontology for medicine. *IEEE Trans Inf Technol Biomed* 1998;2:229–42.
- [71] Richesson RL, Krischer J. Data standards in clinical research: gaps, overlaps, challenges and future directions. *J Am Med Inform Assoc* 2007;14:687–96.
- [72] Mohanty SK, Mistry AT, Amin W, Parwani AV, Pople AK, Schmandt L, et al. The development and deployment of Common Data Elements for tissue banks for translational research in cancer – an emerging standard based approach for the Mesothelioma Virtual Tissue Bank. *BMC Cancer* 2008;8:91.
- [73] Min H, Manion EJ, Goralczyk E, Wong YN, Ross E, Beck JR. Integration of prostate cancer clinical data using an ontology. *J Biomed Inform* 2009;42:1035–45.
- [74] caBIG Vocabularies & Common Data Elements (VCDE) Workspace. <<https://cabig.nci.nih.gov/workspaces/VCDE/>>.
- [75] Brandt CA, Cohen DB, Shifman MA, Miller PL, Nadkarni PM, Frawley SJ. Approaches and informatics tools to assist in the integration of similar clinical research questionnaires. *Methods Inf Med* 2004;43:156–62.
- [76] Wolff AC, Hammond ME, Schwartz JN, Hagerty KL, Allred DC, Cote RJ, et al. American Society of Clinical Oncology/College of American Pathologists

- guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer. *Arch Pathol Lab Med* 2007;131:18–43.
- [77] Hammond ME, Hayes DF, Dowsett M, Allred DC, Haggerty KL, Badve S, et al. American Society of Clinical Oncology/College of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer. *Arch Pathol Lab Med* 2010;134:907–22.
- [78] Beaulah SA, Correll MA, Munro RE, Sheldon JG. Addressing informatics challenges in Translational Research with workflow technology. *Drug Discov Today* 2008;13:771–7.
- [79] Hu H, Kvecher L. Data tracking systems. In: Hu H, Mural RJ, Liebman MN, editors. *Biomedical informatics in translational research*; 2008. p. 111–36 [chapter 7].
- [80] Hu H, Zhang Y, Kvecher L, Sun W, Hooke J, Mural RJ, et al. Different characteristics of invasive breast cancers between Caucasian and African American women. In: *The 29th San Antonio breast cancer symposium*, San Antonio, TX; 14–17 December 2006.
- [81] Maskery SM, Hu H, Hooke J, Shriver CD, Liebman MN. A Bayesian derived network of breast pathology co-occurrence. *J Biomed Inform* 2008;41:242–50.
- [82] Bekhash A, Maskery SM, Kvecher L, Correll M, Zhang Y, Hooke J, et al. Clinical breast care project data warehouse as a research environment for breast cancer risk factor studies; submitted for publication.
- [83] Bekhash A, Maskery SM, Kvecher L, Hooke J, Liebman MN, Shriver CD, et al. A pilot study of known or controversial breast cancer risk factors using the Clinical Breast Care Project database as a research environment. In: *The 30th San Antonio breast cancer symposium*, San Antonio, TX; 13–16 December 2007.
- [84] Saini J, Kovatich A, Bekhash A, Hooke J, Mural RJ, Shriver CD, et al. Association of clinicopathologic characteristics with IHC-based breast cancer subtypes. *Cancer Res* 2009;69:635.
- [85] Bekhash A, Saini J, Li X, Rapuri P, Hooke AJ, Mural RJ, et al. Ethnicity difference of benign breast diseases in breast cancer and non-cancer patients. *Cancer Res* 2010;288s–9s.
- [86] Saini J, Li X, Kvecher L, Larson C, Croft D, Yang YC, et al. Differential gene expression analysis among post-menopausal caucasian invasive breast cancer, benign and normal subjects. *Cancer Res* 2010;245s.
- [87] Li X, Rapuri P, Melley J, Brillhart G, Wu W, Kvecher L, et al. Comparative analysis of gene expression profiles in human breast cancer from microarray data using breast tissues and peripheral blood samples. In: *International conference on intelligent systems for molecular biology (ISMB)*, Boston, MA; 11–13 July 2010.
- [88] Edge SB, Byrd DR, Compton CC, Fritz AG, Greene FL, Trotti A. *AJCC cancer staging manual*. 7th ed. New York, NY: Springer-Verlag; 2010 [6th printing].
- [89] Lester SC, Bose S, Chen YY, Connolly JL, de Baca ME, Fitzgibbons PL, et al. Protocol for the examination of specimens from patients with invasive carcinoma of the breast. *Arch Pathol Lab Med* 2009;133:1515–38.
- [90] Lester SC, Bose S, Chen YY, Connolly JL, de Baca ME, Fitzgibbons PL, et al. Protocol for the examination of specimens from patients with ductal carcinoma in situ of the breast. *Arch Pathol Lab Med* 2009;133:15–25.
- [91] Printz C. New AJCC cancer staging manual reflects changes in cancer knowledge. *Cancer* 2010;116:2–3.
- [92] Andrews JE, Richesson RL, Krischer J. Variation of SNOMED CT coding of clinical research concepts among coding experts. *J Am Med Inform Assoc* 2007;14:497–506.
- [93] Childs LC, Enelow R, Simonsen L, Heintzelman NH, Kowalski KM, Taylor RJ. Description of a rule-based system for the i2b2 challenge in natural language processing for clinical data. *J Am Med Inform Assoc* 2009;16:571–5.
- [94] Murphy SN, Mendis M, Hackett K, Kuttan R, Pan W, Phillips LC, et al. Architecture of the open-source clinical research chart from Informatics for Integrating Biology and the Bedside. In: *AMIA annu symp proc*; 2007. p. 548–52.



## DW4TR: A Data Warehouse for Translational Research

Hai Hu<sup>a,\*</sup>, Mick Correll<sup>c,1</sup>, Leonid Kvecher<sup>a</sup>, Michelle Osmond<sup>b,2</sup>, Jim Clark<sup>b,2</sup>, Anthony Bekhash<sup>a</sup>, Gwendolyn Schwab<sup>a</sup>, De Gao<sup>d,2</sup>, Jun Gao<sup>d,2</sup>, Vladimir Kubatin<sup>c,2</sup>, Craig D. Shriver<sup>e</sup>, Jeffrey A. Hooke<sup>e</sup>, Larry G. Maxwell<sup>e</sup>, Albert J. Kovatich<sup>f</sup>, Jonathan G. Sheldon<sup>b,3</sup>, Michael N. Liebman<sup>a,4</sup>, Richard J. Mural<sup>a</sup>

<sup>a</sup>Windber Research Institute, Windber, PA, USA

<sup>b</sup>InforSense LLC, London, UK

<sup>c</sup>InforSense Ltd., Boston, MA, USA

<sup>d</sup>InforSense Ltd., Shanghai, China

<sup>e</sup>Walter Reed Army Medical Center, Washington, DC, USA

<sup>f</sup>MDR Global, Windber, PA, USA

### ARTICLE INFO

#### Article history:

Received 29 September 2010

Accepted 4 August 2011

Available online 22 August 2011

#### Keywords:

Data Warehouse

Translational research

Patient-centric data model

User interface

Ontology

### ABSTRACT

The linkage between the clinical and laboratory research domains is a key issue in translational research. Integration of clinicopathologic data alone is a major task given the number of data elements involved. For a translational research environment, it is critical to make these data usable at the point-of-need. Individual systems have been developed to meet the needs of particular projects though the need for a generalizable system has been recognized. Increased use of Electronic Medical Record data in translational research will demand generalizing the system for integrating clinical data to support the study of a broad range of human diseases. To ultimately satisfy these needs, we have developed a system to support multiple translational research projects. This system, the Data Warehouse for Translational Research (DW4TR), is based on a light-weight, patient-centric modularly-structured clinical data model and a specimen-centric molecular data model. The temporal relationships of the data are also part of the model. The data are accessed through an interface composed of an Aggregated Biomedical-Information Browser (ABB) and an Individual Subject Information Viewer (ISIV) which target general users. The system was developed to support a breast cancer translational research program and has been extended to support a gynecological disease program. Further extensions of the DW4TR are underway. We believe that the DW4TR will play an important role in translational research across multiple disease types.

© 2011 Elsevier Inc. All rights reserved.

### 1. Introduction

Translational research requires integration of clinicopathologic data, biospecimen data, and molecular data, across multiple data collection and generation platforms [1–4]. It is also critical to make these data usable at the point-of-need. From the data management perspective, clinicopathologic data and molecular data have distinct features. The former are characterized by a large number of data fields, often with missing values due to failure to collect, non-existence or inaccessibility of the data. Molecular data, on the other hand, are characterized by a limited number of data

fields, large number of records, and ever-evolving data types associated with the development of new technologies. While a translational research project typically collects and generates its own clinicopathologic and molecular data, public domain data are often used as well. Furthermore, temporal information needs to be properly managed and presented.

Many clinical and translational research projects use questionnaires as a clinical data collection instrument, and currently using data from Electronic Medical Records (EMRs) for translational research is gaining momentum as EMR systems replace paper-based medical records. Compared to EMRs, questionnaires are simpler as only questions useful to the specific research program of the disease would be included, thus managing such data does not require a comprehensive system covering all the human diseases. An EMR system cannot be used as a data management system for translational research, due to the transactional nature of the former and the reporting and analysis requirements of the latter. The requirements from the US Health Insurance Portability and Accountability Act of 1996 (HIPAA) for protection of patients' Protected Health

\* Corresponding author. Address: Windber Research Institute, 620 7th St., Windber, PA 15963, USA. Fax: +1 814 467 6334.

E-mail address: [h.hu@wriwindber.org](mailto:h.hu@wriwindber.org) (H. Hu).

<sup>1</sup> Present address: Dana-Farber Cancer Institute, Boston, MA, USA.

<sup>2</sup> Present address: IDBS, Burlington, MA, USA.

<sup>3</sup> Present address: Oracle, Swindon, UK.

<sup>4</sup> Present address: Strategic Medicine, Kennett Square, PA, USA.

## Abbreviations

2D-DIGE	2-Dimensional, Difference In Gel Electrophoresis	IT	Information Technology
ABB	Aggregated Biomedical-Information Browser	LC-MS	Liquid-Chromatography Mass-Spectrometry
aCGH	Array Comparative Genomic Hybridization	LIMS	Laboratory Information Management System
AJAX	Asynchronous JavaScript and XML	MeSH	Medical Subject Heading
AJCC	American Joint Committee on Cancer	MIAME	Minimum Information About a Microarray Experiment
ASCO	American Society of Clinical Oncology	MS	Mass Spectrometry
BioPAX	Biological Pathway Exchange	NBIA	National Biomedical Imaging Archive
BMI	Body-Mass Index	NCI	National Cancer Institute
caBIG <sup>a</sup>	Cancer Biomedical Informatics Grid	NCICB	National Cancer Institute Center for Bioinformatics
caBIO	Cancer Bioinformatics Infrastructure Objects	NIH	National Institutes of Health
caCORE	cancer Common Ontological Reference Environment	OCT	Optical Cutting Temperature medium
caDSR	cancer Data Standards Registry and Repository	OLAP	On-Line Analytical Processing
CAP	College of American Pathologists	MOLAP	Multi-dimensional OLAP
CBCP	Clinical Breast Care Project	ROLAP	Relational OLAP
CDW	Clinical Data Warehouse	OWL	Web Ontology Language
CGEMS	Cancer Genetic Markers of Susceptibility	PHI	Protected Health Information
CTSA	Clinical and Translational Science Award	PI	Principal Investigator
DICOM	Digital Imaging Communications in Medicine	PR	Progesterone Receptor
DNA	Deoxyribonucleic Acid	QA	Quality Assurance
DW	Data Warehouse	RAID	Redundant Array of Independent Disks
DW4TR	Data Warehouse for Translational Research	RAM	Random-Access Memory
EAV	Entity-Attribute-Value	REMBRANDT	Repository of Molecular Brain Neoplasia Data
EMR	Electronic Medical Record	RDBMS	Relational Database Management System
ER	Estrogen Receptor	RIN	RNA Integrity Number
ETL	Extract, Transform, and Load	RNA	Ribonucleic Acid
FF	Flash Frozen	RT-PCR	Real-Time Polymerase Chain Reaction
FFPE	Formalin-Fixed, Paraffin-Embedded	SAS	Statistical Analysis System
FISH	Fluorescence In Situ Hybridization	SNOMED-CT	Systemized Nomenclature of Medicine-Clinical Terms
GB	Giga-Byte	SNP	Single-Nucleotide Polymorphism
GDP	Gynecological Disease Program	SPSS	Statistical Package for the Social Sciences
GO	Gene Ontology	STRIDE	Stanford Translational Research Integrated Database Environment
H&E	Hematoxylin and Eosin	TB	Tera-Byte
HER2	Human Epidermal growth factor Receptor 2	UML	Unified Modeling Language
HIPAA	Health Insurance Portability and Accountability Act of 1996	UMLS	Unified Medical Language System
HL7	Health Level 7	VCDE	Vocabulary and Common Data Elements
I2B2	Informatics for Integrating Biology and the Bedside	VPD	Virtual Private Database
ID	Identification	VPN	Virtual Private Network
IHC	Immunohistochemistry	WRAMC	Walter Reed Army Medical Center
IRB	Institutional Review Board	XML	Extensible Markup Language
ISIV	Individual Subject Information Viewer		

Information (PHI) also differ between the two systems. In addition, translational research may not need all of the data contained in an EMR and it requires data from other sources as well including molecular data. It is important that a data centralization system be developed for translational research capable of inputting clinical data from questionnaires and EMRs for any human disease, and integrating molecular data from different biochemical, genomic and proteomic experimental platforms.

Data warehousing as an important component of biomedical informatics, provides a way for data integration in clinical and translational research [2,5–13]. A Data Warehouse (DW) is an information repository for data analysis and reporting [14,15]. For raw clinical data, an Entity-Attribute-Value (EAV) data model is often used due to their sparseness and dynamic nature [16–18]. Such data models require the development of ontologies to manage the complex relationships between clinical concepts/questions and possible answers. The user interface is often provided by an On-Line Analytical Processing (OLAP) tool, which provides multidimensional data analysis capabilities by aggregating data across hierarchically organized dimensions using either a multidimensional (MOLAP) or relational (ROLAP) model [14,15,19]; the former

relies on pre-computed data cubes whereas the latter directly queries a relational database.

Developing a comprehensive DW system to meet the needs of translational research faces many challenges. Based on our own experience, such challenges include; management of operationally important information, such as HIPAA compliance and PHI, data input from questionnaires with version control, data input from EMRs, biospecimen collection and banking, data de-identification, molecular images, multiple-platforms of molecular data, temporal information, and data ownership, etc. From the system development perspective challenges include; development and application of ontologies including controlled vocabulary and modeling of clinical data, molecular data, image data, and temporal data. In addition, development of graphical user interfaces for different levels of users is important, and data security and accessibility should be properly addressed.

Ontology development is very important to data integration and exchange; for the purpose of this paper we focus on its physical artifacts of terminologies/controlled vocabulary and information model/data model [20]. Many ontologies have been developed, each serving a specific purpose. In the clinical field,

commonly accepted standards/ontologies include; Health Level 7 (HL7) for exchange of clinical messages to support clinical practice and healthcare service [21]; Medical Subject Heading (MeSH) for medical literature indexing which has also proven useful for classification of biomedical entities [22]; Systemized Nomenclature of Medicine—Clinical Terms (SNOMED-CT), as a comprehensive clinical healthcare terminology organized hierarchically [23]; and Digital Imaging Communications in Medicine (DICOM) for medical image annotations [24]. For biological research, Gene Ontology (GO), arguably the most successful biological ontology, is aimed to standardize the representation of gene and gene product attributes across species and databases [25]; Minimum Information About a Microarray Experiment (MIAME) for microarray-based technologies is generally accepted by the community [26]; Biological Pathway Exchange (BioPAX) is for pathway data exchange [27]. NCI Thesaurus provides reference terminology for many NCI and other systems covering vocabulary for clinical care and translational and basic research [28]. Systems have been developed to unify existing ontologies or provide ontology library services, for example the BioPortal by the US National Center for Biomedical Ontology [29], and the Ontology Lookup Service by European Bioinformatics Institute based on the Open Biomedical Ontology format [30]. The Unified Medical Language System (UMLS) developed by the US National Library of Medicine, aims to provide ad hoc linkages across life science ontologies [31]. A number of reviews are available including ones discussing development, implementation, and use of ontologies [20,32–34].

Temporal information is also very important in modeling health and disease development. We consider disease as a process, not a state. Breast cancer, for example, takes many years to develop before a lesion can be detected by mammography [35]. First considered by artificial intelligence researchers in the 1980s, temporal reasoning and temporal data query became a topic of interest in biomedical/medical informatics in the 1990s [36–38]. To enable “controlled language use” of the temporal information in healthcare, a European Prestandard CEN ENV 12381 has been developed [39]. Temporal information needs to be collected, modeled, and presented. Collection of temporal information can be done explicitly through questionnaires or data forms which record historical information, or abstracted through the time-stamps of data collected and analyzed over time. For the latter, temporal-abstraction and temporal-querying systems have been developed which have proven clinically useful in monitoring clinical disorders [40–42]. A data model comprising Parameters, Events, and Constants has been defined for a temporal query language [43,44], where Parameters are basically entities with values bearing a time-stamp, Events are external acts act upon a patient which could be time-stamped or occur over a time interval, and Constants are non-temporal measures. There have been reports of a data management or DW system incorporating temporal information for clinical trials, time-series gene expression microarray experiment, and medical information events in a hospital setting [38,45,46].

The many challenges faced in developing a comprehensive DW system for translational research make this a difficult undertaking. In practice, different systems have been developed focusing on different challenges based on specific needs of supported program(s). For example, I2B2, Informatics for Integrating Biology and the Bedside [47], took a top-down “Enterprise” approach with an open architecture enabling integration of outside modules performing specialized functions that are needed for translational research [48,49]. It address a very important and specialized niche in the clinical informatics space, namely, providing researchers with sufficient access to clinical data to enable study design and cohort selection, while minimizing many of the patient privacy risks and concerns [50–54]. A central conundrum of translational research is that access to detailed clinical data typically requires proper

informed consent and IRB approval; however, in order to design a cohort study or to determine if a particular project is even feasible, a researcher often needs access to aggregate statistics for key clinical attributes of interest at an early stage of the process. I2B2 overcomes these challenges by enabling researchers to design and execute queries against a de-identified data mart that return only aggregate counts; furthermore, results are subtly obfuscated to avoid identification by more sophisticated combinations of queries. Having been deployed for a number of projects across multiple sites, I2B2 is a mature software offering. However, it does not currently provide the level of detailed analysis necessary for clinical and translational research beyond cohort selection and study design [51–54]. STRIDE, The Stanford Translational Research Integrated Database Environment [55], focuses on a Clinical Data Warehouse (CDW) and supports a virtual biospecimen model for accessing several specialized tissue banks [13,56,57]. It also has a Research Database management system supporting multiple logically separate research databases. An EAV model is used for data storage and HL7 Reference Information Model is used for data representation. Its semantic layer consists of a framework supporting multiple terminologies. Some of the Clinical and Translational Science Award (CTSA) centers [61] also have data warehousing efforts.

There are also data warehousing efforts in the cancer Biomedical Informatics Grids (caBIG<sup>®</sup>) [62]. caBIG<sup>®</sup> is an information network enabling members of the cancer community to share data and knowledge based on a core Grid architecture. It was launched in 2003 by the NCI Center for Bioinformatics (NCICB) of the US NIH [63,64]. One of its data warehousing efforts is caBIO, a major component of caCORE (cancer Common Ontological Reference Environment) that was developed before the launch of caBIG<sup>®</sup> [65], enabling access of biomedical annotations from curated data sources in an integrated view. As one of the released products of caBIG<sup>®</sup>, the current version caCORE 3.x includes Enterprise Vocabulary Services for hosting and managing vocabulary, cancer Data Standards Registry and Repository (caDSR) for hosting and managing metadata, and the caCORE Software Development Kit [66]. Another data warehousing approach for caGrid is a semantic web-based data warehouse for creating relationships among caGrid models, accomplished through the transformation of semantically-annotated caBIG<sup>®</sup> Unified Modeling Language (UML) information models into Web Ontology Language (OWL) ontologies that preserve those semantics [67]. Still another caBIG effort, caIntegrator [58], is a framework for data integration currently capable of integrating microarray data in caArray and imaging data from the National Biomedical Imaging Archive (NBIA). It has been applied to several projects, e.g., REMBRANDT [59] and CGEMS [60]. These efforts are best described as development projects using the caIntegrator framework. None of these efforts in caBIG, however, enables users to dynamically interrogate the clinicopathologic data in a multidimensional manner by non-informatics specialists using an intuitive interface.

We began our DW development focused on a questionnaire-based system, but later realized that a patient-centric and expandable data model would be a better solution. We envisioned developing a system to support translational research in general and began by developing a system to support the Clinical Breast Care Project (CBCP) [68] but designed it to be expandable to support additional programs. This approach has been validated by our successful expansion of the system to support a second translational research program, the Gynecological Disease Program (GDP) [69]. It is currently under further expansion to support a third translational research program, a consortium effort headed by Thomas Jefferson University involving five organizations which is characterizing a set of 5000 invasive breast cancer cases. All these projects have multiple platforms for clinical and molecular studies.

Development of our system, the Data Warehouse for Translational Research (DW4TR), focused on three challenges in DW development. The first was the development of a pragmatic data model to satisfy the needs of questionnaire-based clinical data input. Despite the existence of a large number of clinical and biological ontologies, there are often gaps, in supporting a research project, between the available ontologies and the specific project needs [70,71]. Many groups therefore create their own ontologies [72,73]. Given the characteristics of clinical and molecular data, it is more challenging to develop a general and flexible system for clinical data; therefore we have made this our first priority [9]. In our case, both CBCP and GDP utilizes not only data elements covered by existing ontologies, but also detailed and specific data elements that are currently not covered. Comparing what we need to what is available, we concluded that completely adopting one or two existing ontologies for our DW development is not feasible (c.f. Section 5). Therefore, we developed our own framework ontology but referred to existing ones, to satisfy our immediate needs first, planning to map our ontology to existing ontologies at a later time to enable transportability.

The second challenge, we focused on was interface development. A user-friendly interface tailored to the non-informatics specialist is needed to ensure both the utility and adoption of this system. The interface that we developed is composed of an Aggregated Biomedical-Information Browser (ABB) and an Individual Subject Information Viewer (ISIV). ABB is ROLAP-based, thus all the calculations and queries are performed at the time of use rather than relying on a pre-calculated data cube, enabling clinicians and scientists who are not informatics experts to dynamically interrogate any combination of the myriad of data elements stored in the DW4TR. The third challenge we focused on was the management of temporal information. We defined three temporal data types and applied them to data attributes to enable the use of the temporal information in the DW4TR from both interfaces. In addition to focusing on these three challenges, we also addressed a number of other challenges described earlier. For example, in managing de-identified clinical information of CBCP and GDP, we developed solutions to satisfy different requirements for data security and accessibility that is also described.

## 2. Methods

### 2.1. System design and development

Initially developed to support the immediate needs of managing the clinical, biospecimen, and molecular data for the CBCP, our design of the DW4TR to be flexible and extensible enabled subsequent expansion to support GDP. Both programs use questionnaires for collecting clinicopathologic information, making expansion of the light-weight production system straight forward.

#### 2.1.1. The data model

The EAV model is used for the original raw clinical data which are subsequently hosted in the DW in an extensible data model reflecting the ontology, where the relationships among the attributes are presented in a meaningful way to the users through a hierarchical organization. The extensible data model is currently composed of the following components:

**2.1.1.1. Patient-centric clinical data model.** The clinical data model was developed to reflect the physician-patient interaction. The ontology framework was developed by a multi-disciplinary team composed of research physicians, biomedical informaticians, IT developers, surgeons, pathologists, and gynecologists and was based on the standard clinical workflows and clinical experiences.

The development not only referred to existing standards and ontologies mainly MeSH and SNOMED-CT but also to the NCI Thesaurus and caBIG VCDE [22,23,28,74], reflecting other reported methods and efforts [32,33,75]. This ontology is reflected on the physical data model, which is hierarchical and composed of a number of primary modules, e.g., "Diagnostic", which are made up of secondary modules (e.g., "Biopsy") and attributes. A secondary module is further composed of tertiary modules and attributes, and so on. An attribute is a fine-grained object that is composed of the name, the data type, and the temporal characteristics (see below). A simple attribute is often called a data element, e.g., sex. A complex attribute is composed of multiple facets, e.g., exercise is composed of frequency, intensity, duration, and type. This way, each data element of a study is represented in the model either by an attribute or by a sub-module. The sub-modules are structured hierarchically to reflect the ontology.

**2.1.1.2. Specimen-centric molecular study data model.** Molecular (biochemical, genomic and proteomic) analysis are performed on collected specimens using multiple experimental platforms. Every experiment is done using biospecimens, including body fluids, solid tissues, and their derivatives. Thus, a specimen-centric data model is a natural choice for these types of data, which are connected to the clinical data model hierarchy as a series of sub-modules. We initially focused on immunohistochemistry (IHC), Fluorescence In Situ Hybridization (FISH), and gene expression microarray data and adopted existing clinical or molecular data standards and guidelines [26,76,77]. The storage and retrieval of high throughput molecular data poses a number of IT challenges. A result file from a single microarray can be tens of megabytes, and it is not uncommon for a single experiment to make use of hundreds of such arrays. Common approaches for handling this type of data include: organizing output files into a hierarchical file system, loading of raw or partially transformed data into a DW, or a hybrid approach of maintaining certain metadata in a database but referring back to original output files with file pointers. Given the large number of molecular modalities used in our current and future research, no single method was considered suitable, therefore our system was designed to take an "all of the above approach" that makes use of flexible data processing pipelines. With this approach distinct pipelines can be created across or even within assay types and all of the storage methods discussed above can be supported. This approach has enabled us to optimize storage of each assay group individually, as well as provide a mechanism for handling changing technology and legacy data.

**2.1.1.3. Temporal data model.** Critical to human disease studies is the proper representation of temporal information. We define three types of temporal data: (1) Static, which is data with no temporal dimension, e.g., ethnicity. (2) Event, which is associated with a specific time point, e.g., a surgical procedure. (3) Interval, which is associated with a starting and ending time point, e.g., a course of medication. Each attribute in the data model is tagged with a proper temporal status and populated with proper values for temporal information representation and analysis.

**2.1.1.4. Medical image data model.** Medical imaging is an important part of clinical practice and a source of data for various studies. We initially focus on digitized and digital mammograms, which follows the DICOM standards [24] and we de-identified the images using DICOM Anonymizer Pro. The data files associated with these images are large and they impose a challenge in data management, e.g., storing those image files in the database of the DW will drastically impede the system performance. Besides, analyzing images typically requires specialized software. Thus, in the DW4TR, we store the image files on a file server; the file locations are then

stored in the DW and logically linked to the patient. The annotations and a thumb nail of the image are also stored in the database to enable efficient search of relevant features. We intend to apply the same principle to other images.

**2.1.1.5. Relationships between questionnaires and the data model.** The data for both CBCP and GDP studies are collected using questionnaires each having multiple questions. Each question has one or more data elements. The data elements could contain values and temporal information. If a data element contains a value, it is mapped to an attribute(s). Some questions ask about the same event at different time points, and these questions are mapped to the same attribute with corresponding temporal information. There are also different questionnaires collecting the same information, e.g. for different studies, and they have been mapped to the same attributes as well. Finally each attribute is mapped to the data model as described above.

### 2.1.2. User interface development

The requirement to develop an intuitive, user friendly system that could be used directly by clinicians and researchers as opposed to being a specialist tool useful only to the technically savvy, led to our decision to take the dimensional modeling approach. While traditional query interfaces are considered intuitive by IT professionals, we found that the dimensional modeling approach was generally preferred by our target demographic. A number of criteria were considered in designing the physical data structure to support the clinical data warehouse, the first of which was the ability to handle a large number of dimensions. A custom binning capability that is convenient to use, was an important feature requested by the users. As is the case with any longitudinal study, the temporal dimension of data is critical, and therefore proper handling of the temporal dimension was another important concern. Efficient handling of sparsely distributed data and support for multiple terminology sets were also important criteria. During such comprehensive communications with the end users, we concluded that all the desired analyses could be categorized as either an aggregate data analysis or an individualized data analysis. Thus, we designed a ROLAP-based ABB for the former, and an ISIV for the latter.

## 2.2. Implementation

### 2.2.1. Development environment

Two DELL Windows server systems, one for the application and the other for the database, were used. Both servers have two Quad Core Intel® Xeon® processors and 8 GB RAM. The database server hard drive is 1 TB and for the application server 0.5 TB, both are in the RAID 5 configuration. Development is based on the Oracle RDBMS 10g, using the InforSense platform that enables a series of analyses to be performed in an analytical workflow environment with each analysis step represented as a node in the workflow [78]. The user-interface is Java based and web accessible. The production system has the same configuration. The DW4TR also supports a one-server system configuration.

### 2.2.2. Data model and interface implementation

Our physical data model is composed of two distinct elements, an attribute repository, and an attribute ontology. The attribute repository stores data in the EAV model. The attribute ontology is a separate hierarchical data model capturing and providing structure to the relationships among the attributes that have been collected so that the ontology (logical patient-centric or specimen-centric data model) are physically positioned in a hierarchical organization of attributes for presentation in a meaningful way to the users of the system.

In implementing the data model, first the source questionnaires are reviewed and a logical mapping of the data attributes onto the newly designed patient modules is developed. This process requires the definition of the data variable type and other metadata for different data attributes, and often requires splitting or merging data elements in order to accommodate the difference in requirements of capturing the data versus analyzing the data.

Parallel to the development of the physical and logical data model was the development of the end user interfaces. Our system provides users with two distinct applications: the ABB, and the ISIV. These Java tools are entirely web based with zero client side foot print, and do not require any additional browser plug-ins. They provide users with a highly interactive and dynamic experience, making use of Asynchronous JavaScript and XML (AJAX) techniques for asynchronous rendering and updating. Both tools are deployed within the InforSense web portal environment, which provides the services used for data source connection pooling, user authentication and authorization, as well as web session management.

### 2.2.3. Data loading

Multiple sources of data in different format are loaded into the DW4TR, including direct Oracle-to-Oracle load from the Laboratory Information Management System (LIMS) and flat files [12,79]. Data loading is performed through a standard process referred to as Extract, Transform, and Load (ETL) [14,15].

## 2.3. Data sources

The current DW4TR is host to clinical data from CBCP and GDP. Subjects are enrolled into the programs via HIPAA compliant, Institutional Review Board (IRB) approved protocols at multiple participating clinical sites, and these protocols explicitly specified WRI as the data integration center for the programs. Clinical data are collected through multiple questionnaires, and biological specimens are collected, processed, banked, and analyzed using genomic and proteomic experimental technologies.

CBCP and GDP happen to be two distinct types of programs in administrative structure, clinical dataset, and tissue banking. They have different data security and access requirements, commanding distinct solutions in the DW4TR. CBCP is a centralized program using one set of the questionnaires, one LIMS, one Tissue Bank, and one data warehouse (the DW4TR). All the participating organizations use approved master IRB protocols from the Walter Reed Army Medical Center (WRAMC), with minor adaptations to the requirements of the local institution for local IRB approval, which is subsequently approved by the IRB of the US Army Medical Research and Materiel Command that contracted the Henry Jackson Foundation to manage the CBCP. Proper paperwork was done to enable centralized tissue banking and data warehousing at WRI. The IRB approved questionnaires contain certain date information such as date of birth which patients are consented to, thus the CBCP clinical data is a "Limited Data Set" per HIPAA definition. The clinical data and biospecimens are de-identified, and each subject is represented by a CBCP number. The link table is securely maintained at the office of the CBCP site PI which is strictly accessible only to the PI or his/her designee at that site. When clinical data Quality Assurance (QA) problems are identified in any participating institutions, they are all reported to WRAMC and the WRAMC clinical data team coordinates the QA resolving process. As of February 2011, over 5000 subjects have been enrolled in the study using two major questionnaires to collect up to 799 elements of clinicopathology data per subject. More than 43,000 specimens (including aliquots) have been collected and genomic and proteomic experiments have been conducted using these specimens.

GDP, on the other hand, is a consortium program constituting of nine clinical and research organizations with WRAMC serving as the lead institution. Each participating organization follows its own IRB-approved protocol allowing only “Safe Harbor Data Set”, and each site manages and owns its own questionnaire-specific datasets. All the captured data are in a de-identified form, and each subject is represented by a GDP study ID. Each site PI maintains the link key to the identity of the subject enrolled at that site. All the de-identified data are then sent to WRI, and eventually loaded into the DW4TR. Aggregated subject information can be shared across the consortium members, but the detailed individual subject information in the DW4TR is only accessible to the originating clinical site. In GDP, about 500 subjects have been enrolled in the study. A dozen questionnaires were used to collect more than 7600 elements of data focusing on surgical procedures, pathology, family history, psychology, food intake, etc. Currently GDP is consolidating the questionnaires being used reducing the data elements to less than 1400.

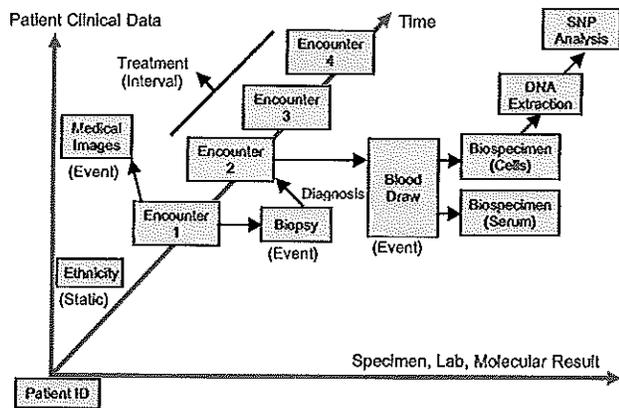


Fig. 1. Illustration in a 3-dimensional diagram of the relationship between clinical data (vertical axis), molecular data (horizontal axis), and temporal information (axis pointing into the page) represented by our data model. The patient is represented by Patient ID at origin. The three temporal data types are illustrated. This patient had clinical encounters at four time points, and “Encounter 1” resulted in medical images and biopsy which led to “Encounter 2”, which resulted in biospecimens that could be used for different types of molecular studies. “Treatment” started from “Encounter 2” and ended at “Encounter 4”.

### 3. Results

#### 3.1. Overview of the DW4TR and the relationships between involved data types

Key to the development of the DW4TR is the understanding of the complex relationships between data collected or generated with multiple platforms. Fig. 1 illustrates our understanding of such relationships in a 3-dimensional patient data space which we can use to represent our data models.

Fig. 2 is a conceptual view of the DW4TR, as a hybrid system, that integrates the data collected or generated by the supported programs, and federates most of other data needed in the studies. Blocks with solid lines are the areas already developed or currently under development, and blocks with dashed lines are areas for future development. The whole DW is composed of a data tier, a middle tier, and an application tier. In the data tier, an extended EAV model is used for the raw clinical data. The middle tier is based on a patient-centric, modularly-structured clinical data model (including a medical image data model), integrating a specimen-centric molecular data model, and a temporal data model. The clinical and molecular data models are designed to be extensible. The application tier is composed of the ABB and the ISIV, and other utilities. Additional applications will be developed to federate and present other data needed but not generated by the supported translational research programs.

#### 3.2. The ontology and the extensible data model

As a bottom-up approach, we developed a pragmatic light-weight ontology using controlled terminology both from existing ontologies and supplemental ones we developed to support the detailed program needs. The relationships between these terminologies in the ontology are reflected on the physical data models described below.

The patient-centric clinical data model was developed to reflect the physician-patient interaction. As illustrated in Fig. 3, the model is composed of six major modules: Medical History, Physical Exam, Diagnostic, Treatment, Outcome, and Scheduling and Consent. Each of these modules is composed of attributes and sub-modules; e.g., the Diagnostic module is composed of Biopsy, Imaging, and

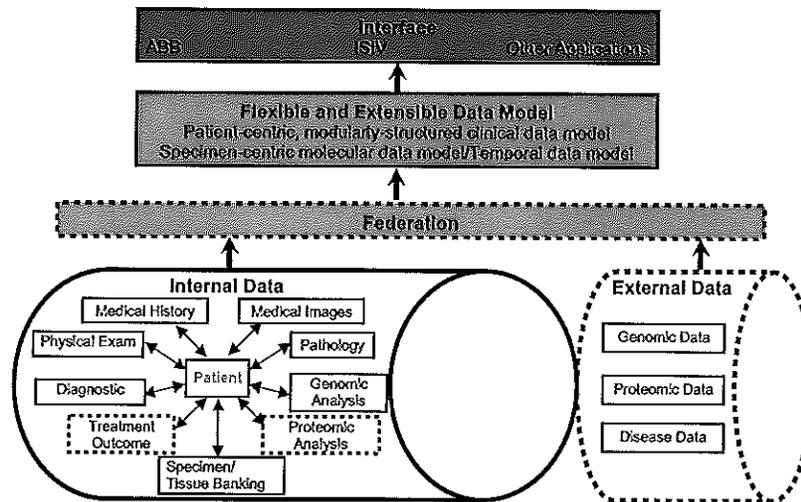


Fig. 2. The structure of the DW4TR including a data tier, a middle tier (blue), and an application tier (green). The data tier is composed of ‘Internal Data’ generated by the supported research project, and the ‘External Data’ not generated by but needed in the research. The “Internal Data” are integrated, and the “External Data” could be in an integrated or federated form but will be federated with the “Internal Data”. Solid line blocks: developed or in development; dashed line blocks: future development. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

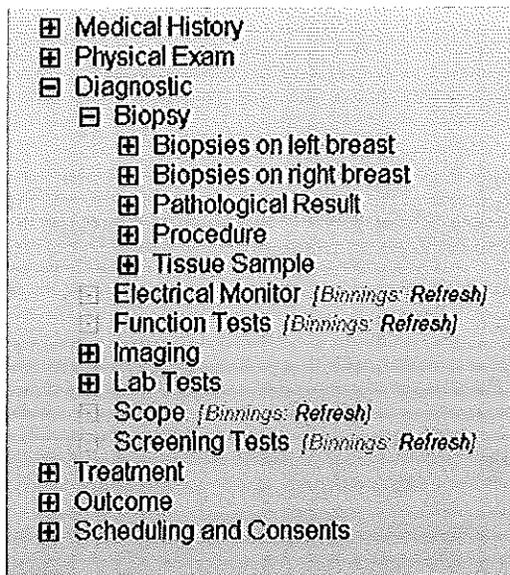


Fig. 3. A screenshot of the ontology structure of the patient-centric, modularly-structured clinical hierarchical data model.

Lab Tests, etc. Each of these sub-modules can then be further composed of its sub-modules and attributes. Many of these sub-modules are disease independent but some are disease specific. This modularly structured data model enables us to adapt the system to support the study of a new disease, by re-using disease independent modules, and developing a limited number of new modules specific to the new disease. The data model has a hierarchical structure. Users with domain knowledge can explore this hierarchical structure to retrieve needed data elements. Users with less specific domain knowledge may not be able to explore this data model as effectively though they can use the keyword search capability of the system to identify, for selection, the data elements containing keywords of interest.

### 3.2.1. The specimen-centric molecular study data model

The specimen-centric molecular study data model is connected to the clinical data model as sub-modules. For example, CBCP breast tissues and lymph nodes in different preservation types (e.g., OCT, FFPE, and FF) are represented by sub-modules of Diagnostic-Biopsy-Tissue Samples, blood samples and their derivatives (e.g., PAXgene tube, plasma, cells, serum, and clots) are represented by sub-modules of Diagnostic-Lab Tests. Genomic, proteomic, and other molecular studies are performed with molecular derivatives of the specimens (e.g. DNA, RNA, and protein) on different experimental platforms, e.g., IHC, FISH, gene expression microarray, SNP microarray, 2D-DIGE, MS, LC-MS, etc. We have completed the modules for the IHC and FISH assay data, and a proof-of-principle has been implemented for gene expression microarrays (c.f. Figs. 4 and 5).

### 3.2.2. Temporal information definition and presentation

All attributes are tagged with one of the three defined temporal properties. All temporal information, including static information, can be viewed in the ISIV (c.f. Fig. 7). Within the ABB users have the ability to define “time filters” for the different columns in the display (c.f. Fig. 6).

### 3.2.3. Example data module definition

Table 1 is a simplified illustration of three defined data modules. Thus, “Ethnic Group” is a simple (single attribute) data

module allowing multiple values for a given subject, and is “Static”. “Alcohol Use” is a complex data module with three facets, and covers an “Interval”. “Surgical Proc(EDURE)” is a module similar to “Alcohol Use” but is an “Event”.

## 3.3. DW4TR Interfaces—ABB and ISIV

### 3.3.1. ABB

This interface allows the user to analyze the data by dynamically creating a multidimensional pivot view. The rows are categorical data fields of interest, in a hierarchical structure. The columns can be data elements of either a numerical or categorical nature, in a flat or hierarchical structure. Although the view is a two dimensional table, the user can effectively explore data of theoretically unlimited dimensions by simply building multiple levels of a hierarchical data structure in both the rows and columns, and expanding them. The number of dimensions (levels of hierarchy) is only practically limited by the performance of the system (see Section 3.6). Fig. 4 (Panels A and B) shows screenshots of how an ABB view of five hierarchical levels was created and used to explore one subclass of breast cancer patients among CBCP Caucasian American subjects, who have triple-negative tumors, i.e., ER (estrogen receptor), PR (progesterone receptor), and HER2 (human epidermal growth factor receptor 2) negative.

Using the ABB, simple analyses such as mean, standard deviation, counts and percentages can be performed. Views on subject counts or specimen counts can be created. The results of interest can be printed, or exported to an Excel spreadsheet for additional analysis using other specialized software. A subject set of interest can also be saved as a cohort, for example the triple-negative Caucasian American breast cancer group shown in Fig. 4B, and used in a subsequent study or analyzed with a different application. The cohort can be analyzed using the ISIV described in the next section, or rendered to additional complicated analysis that the InforSense Analytical Workflow platform supports (not shown). The cohort data can also be exported to an Excel spreadsheet for analysis by specialized software outside of the DW4TR.

ABB offers many other features including a unique custom data binning capability. Multiple binning methods can be used, either automatically or manually, to create bins to explore data elements that contain either numerical or categorical values. For different studies, a user may need to use the same information in different ways. As shown in Fig. 4 (Panels C and D), the ethnic group attribute contains more than a dozen different possible values from CBCP and GDP. If the user wants to concentrate on Caucasians, African Americans, and Asian Americans, he/she can create three bins for them, and place the rest in “Other”. A histogram of the binning can then be viewed. Custom data binning can be done with categorical data as well as numerical data, and in our practice viewing histograms helps to adjust the binning strategy to ensure that there are a sufficient number of records available in each bin.

One type of molecular data can be directly accessed through ABB is IHC. Fig. 4B shows data for three molecules (ER, PR and HER2) which are routinely assayed in the CBCP, for which there are established clinical standards for clinical IHC [76,77]. Two other molecules, Ki67 and p53, are also routinely assayed though there are no current clinical standards for these molecules. In CBCP, all five molecules are characterized by protein expression assayed by IHC, and HER2 is further characterized with gene copy number ratio assessed using FISH as needed. Both IHC and FISH data types are supported in the production system. The IHC submodule contains Result, Percentage, Intensity, and Staining Pattern for nuclear proteins (ER and PR) or Result and Score for membrane-bound proteins (HER2). The FISH submodule contains Result and Ratio. Both submodules belong to “Pathologic Results” of “Biopsy” of “Diagnostic” in the data model. Specific business rules are applied in

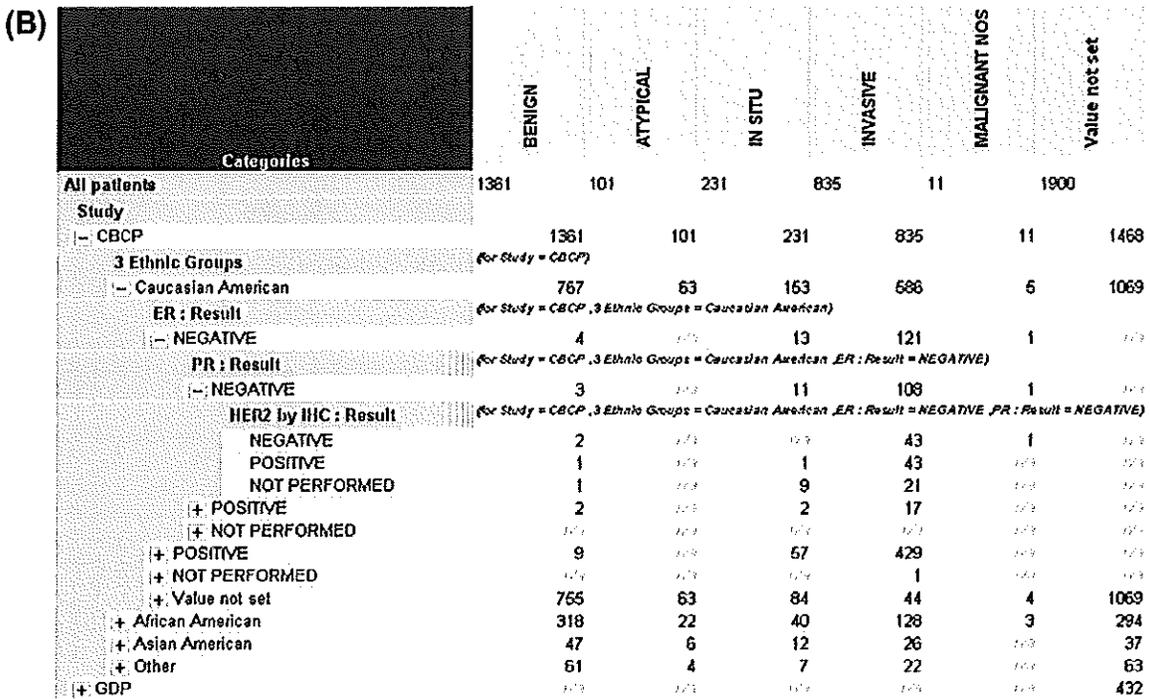
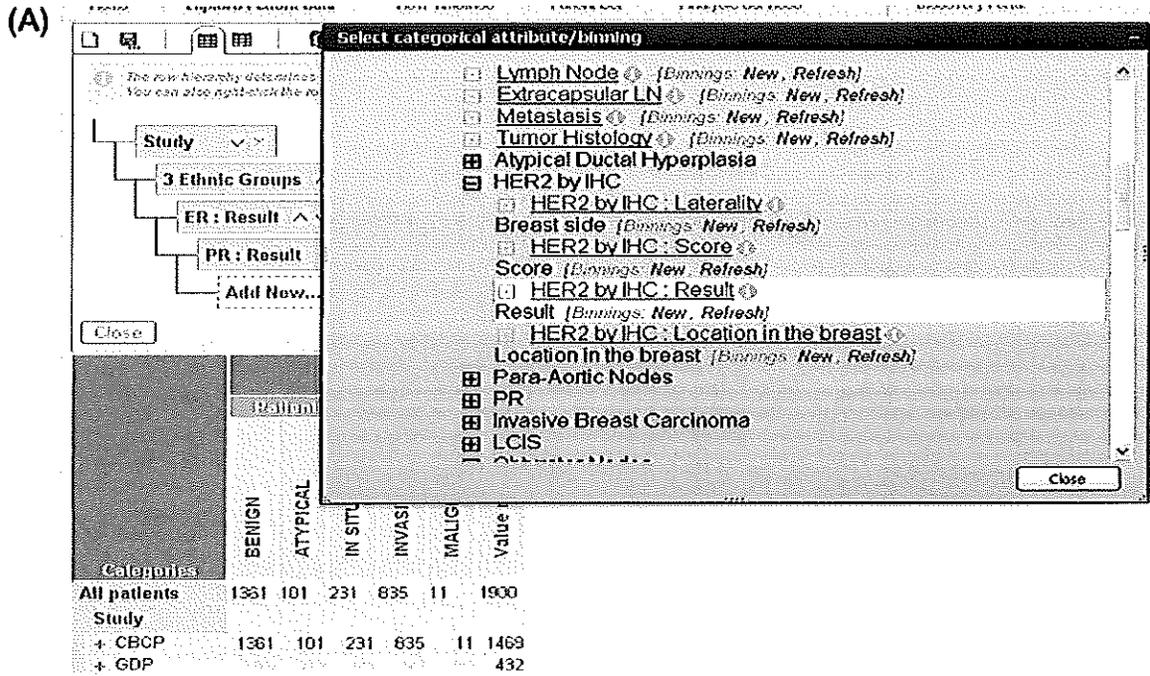


Fig. 4. Screenshots of ABB. (A) View construction. The column is 'Patient Pathology Category', and the rows are, hierarchically, 'Study', '3 Ethnic Group' (binned from 'Ethnic Group' as shown in Panels C and D), 'ER Result', 'PR Result', and the screen was captured when 'HER2 by IHC: Result' was being selected. The inset shows a portion of the data module where the data element resides. (B) View exploration. Rows of five levels were explored, and at the end the numbers are shown for triple negative (ER-, PR- and Her2-) subjects for CBCP Caucasian American in each pathology category. Similar analysis can be done for other ethnic groups. (C) The custom data binning feature. Here a manual binning option is selected. Bins created are 'Caucasian American', 'African American', 'Asian American', and 'Other', with Null defaulted. Available values are in the window of 'Unassigned Values', which can be dragged into the bins on the left to complete the binning. (D) Histogram of the binning result, which can be viewed during custom data binning or at the time of use.

the ETL process, for example the final HER2 expression result is determined by IHC Score (0 or 1+ as negative, and 3+ as positive), but when IHC Score is ambiguous (2+) the Ratio of the FISH assay is used, with  $\leq 1.8$  as negative and  $\geq 2.2$  as positive, and an intermediate ratio is considered as ambiguous.

A proof-of-principle model of microarray-based molecular data is in our test system; this feature is the only feature described in Section 3 that has not yet been applied to the production system. The MIAME standards [26] are followed with some modification, for example, our "sample ID" eliminates the need for subject and

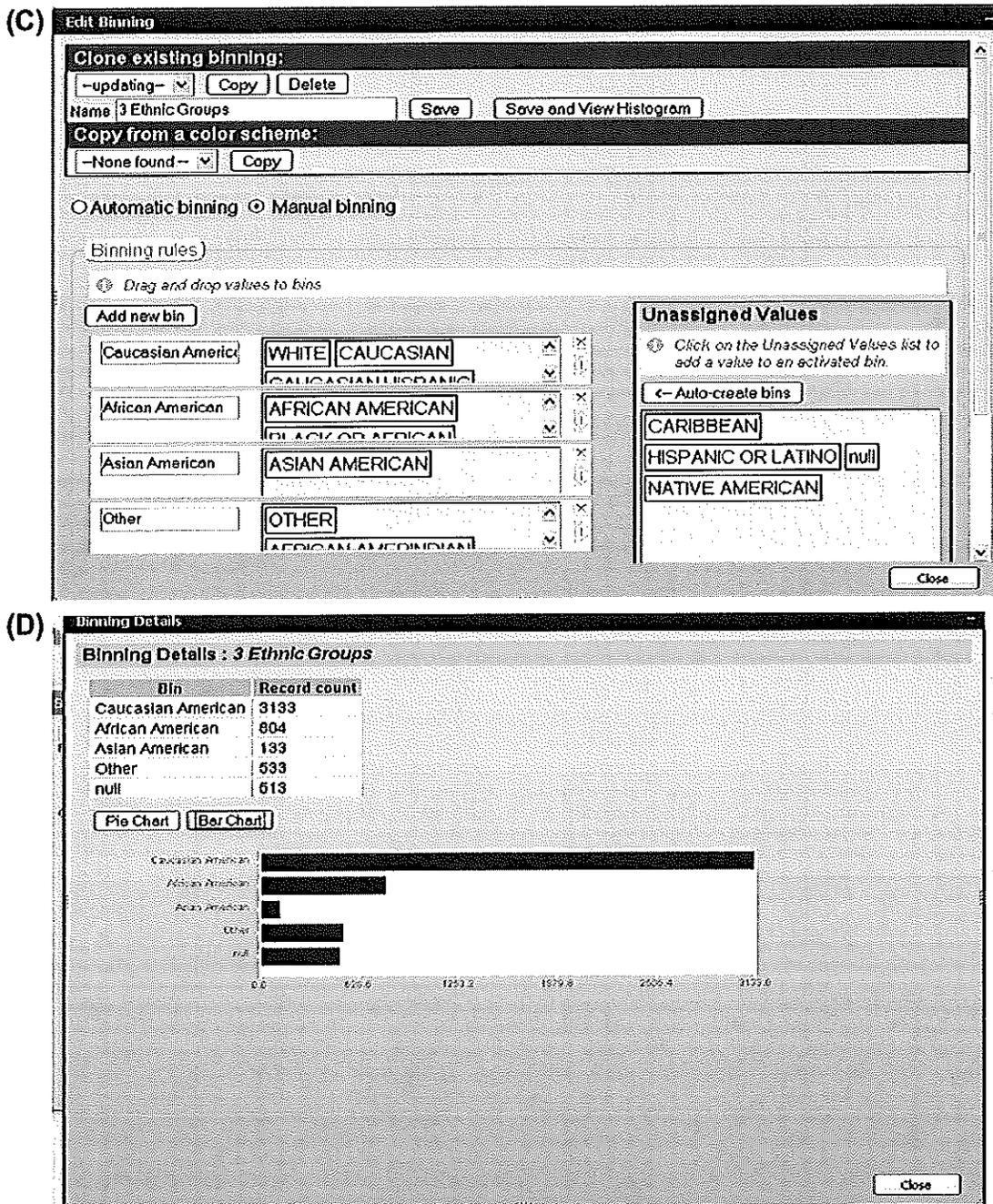


Fig. 4 (continued)

specimen information. The metadata are modeled, including sample ID, bio-molecule information (lab protocol, 260/280 ratio, RIN, etc.), array information (platform, array ID/batch number, etc.), experimental information (project name, PI, lab operator, date, etc.), and results (call rate, 3'/5' ratio, Pass/Fail flag, data file location, etc.). Metadata are accessible via ABB for selecting microarray data of interest in a cohort defined by clinicopathologic characteristics, but we saved the array raw data files in a file server due to their size and the fact that their analysis typically requires specialized software that read in the raw data file directly. Fig. 5 shows an example how a set of microarray data can be identified from previous experiments performed on samples from

a cohort of interest. This information may lead to a secondary use of the microarray data, or help the researcher to design a new experiment.

ABB also contains a utility "Time Filter" to apply temporal information in the analysis of aggregated information. The time filters are used to limit or constrain the data that are returned much like any other type of filter criteria. Time filters can be created based on absolute dates, patient age, or relative to other events. Fig. 6 shows one example how "Alcohol Usage" was analyzed for CBCP subjects when they were at different age ranges. "Alcohol Usage" is an Interval temporal data type with values self-reported by subjects covering different periods of time in patient's life. Note that direct

Categories	Patient Count						
	ATYPICAL	BENIGN	IN SITU	INVASIVE	MALIGNANT NOS	Value not set	
All entities	4896	114	1607	301	1286	12	1620
2 Ethnic Groups							
- Caucasian American	3270	69	922	203	914	8	1183
Menopausal Status	(for 2 Ethnic Groups = Caucasian American)						
- POST-MENOPAUSAL	1348	30	230	103	522	5	473
Microarray Experiment : Platform	(for 2 Ethnic Groups = Caucasian American ,Menopausal Status = POST-MENOPAUSAL)						
AFFY GENECHIP U133 PLUS2	138		3		76		61
Value not set	1210	30	227	103	446	5	412
- PRE-MENOPAUSAL	1283	24	484	66	235	3	480
Microarray Experiment : Platform	(for 2 Ethnic Groups = Caucasian American ,Menopausal Status = PRE-MENOPAUSAL)						
AFFY GENECHIP U133 PLUS2	46		2		19		26
Value not set	1237	24	482	66	216	3	454
+ STATUS POST HYSTERECTOMY	343	13	78	26	97		133
+ SURGICALLY MENOPAUSAL	379	11	86	25	126		140

Fig. 5. Screenshot of ABB showing exploration of available gene expression microarray data on biospecimens of a CBCP cohort of interest. Subjects were Caucasian American, post-menopausal or pre-menopausal, and experiments were performed on an Affymetrix platform. Subjects were mostly invasive breast cancer patients and normal subjects (as represented by 'Value Not Set' in the last column). The numbers from each pathology category do not completely add up to the total patient count on the left-most column since a couple of patients changed the category status with time. Highlighted row of subjects (light blue) can be saved as a cohort for further studies.

custom-binning of subject's age would not be useful to this analysis.

### 3.3.2. ISIV

This interface is for viewing and analysis of detailed subject information, most importantly the temporal-related information. It takes subject IDs as the input either by direct entry, or selecting from a pre-defined cohort. Events and Intervals of interest can be selected, and applied to the whole subject set. Information can be viewed independently for every subject, or aligned across the subject set on Events of interest to enable comparison between subjects. Static information can be displayed as well, so all the information about the subject can be studied. An example screenshot of ISIV is shown in Fig. 7.

Both the ABB and ISIV require minimal training before one can start to use them. The average training time for a clinician or scientist is 15 min, and the user can learn more features by using the system. Interestingly, for a high school student the average training time is only 10 min.

### 3.4. Direct data extraction from the database

The raw data in the DW4TR can be directly queried from the database, as a complete or partial dataset depending on the research need. The data can be presented in either a flat file structure or in Excel spreadsheet, for use with specialized data analysis software such as SAS, SPSS, or R. Such queries have been developed for and applied to both CBCP and GDP data. Direct data extraction requires a database administrator and is not the focus of this paper, thus is only briefly described here.

### 3.5. Extensibility of the DW4TR

The DW4TR was developed to support the CBCP with a vision to support translational research in general. This extensibility was tested and validated by expanding the system to support the GDP. Its flexibility was tested by our successful revision of the model commanded by the consolidation of GDP questionnaires and data elements. A number of disease-independent attributes and submodules developed for the CBCP implementation were re-used for the GDP instance, for example demographics and elements of past medical history. Many other attributes/submodules were modified to satisfy the requirements of GDP, for example alcohol usage. For GDP-specific data elements, additional modules were developed. In addition, the GDP contains a psychology questionnaire and a food questionnaire that are program-specific but not disease-specific. These were developed and structured to the overall patient-centric clinical data model, and could potentially be re-used for future programs. In addition, the GDP has different data security and accessibility requirements, and we have developed additional mechanism for it as detailed in Section 4. With the experience gained in supporting the GDP, we are confident that we will succeed in the current further extension of the system to support the third translational research program for the consortium led by Thomas Jefferson University.

### 3.6. System performance

To evaluate the performance potential of the DW4TR, we conducted a stress test using the CBCP data in a one-server system configuration. The test was done from a user's perspective, by

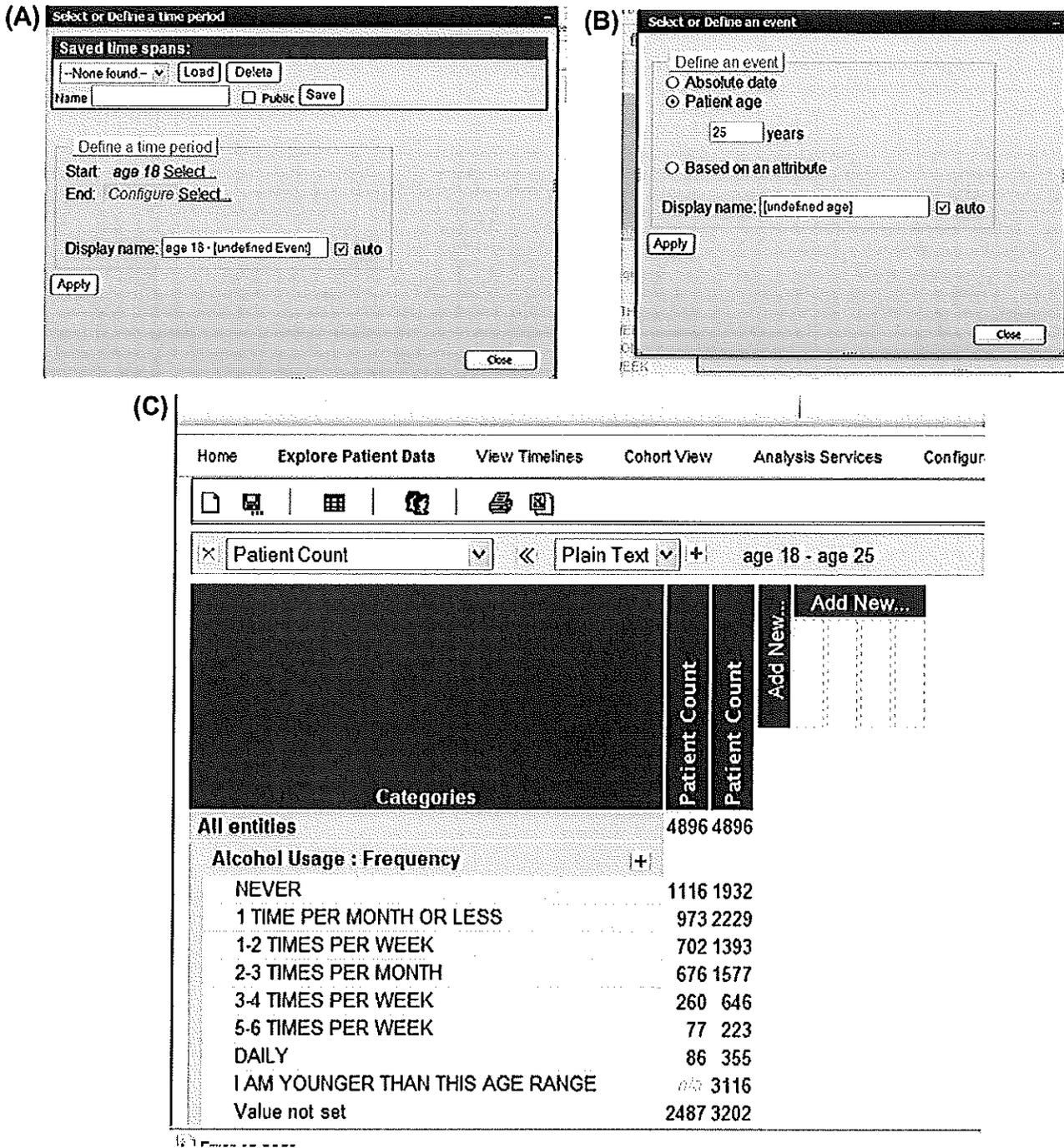


Fig. 6. Use of temporal information in ABB to study data of interests. (A) Screenshot showing the use of the 'Time Filter' to define the starting and ending time for study, with starting time already being configured but not the ending time. (B) Configuring the ending time based on Age, but it can also be based on Absolute Date or an Attribute. (C) Example showing the alcohol usage history across the CBCP subjects. The first column shows the usage frequencies of subjects when they were between 18–25 years of age, and the second column shows the usage frequencies across the whole population. Additional age ranges can be configured to show the corresponding alcohol usage habit.

measuring how long it takes for the data to completely display in a hierarchical view for the first time—note that subsequent display of the same data is much faster. The view is composed of three levels of Pathology Category, Ethnicity, and Body-Mass Index (BMI, with custom-binning) in the rows and record counts in the columns. Starting from the then patient number of 4234, we artificially replicated the records in the database to 8468, 16,936, 33,872, and 67,744 without performing a database tune-up other than Oracle table analysis. It took one second or less, for the level

1 (Pathology Category) and level 2 (Ethnicity) data to be displayed in all these tests. For level 3 (BMI), it took 2, 4, 12, 39, and 134 s respectively to display the data when the number of records doubled stepwise from 4234 to 67,744. Given that the current CBCP annual subject enrollment is about 600, it will take another 20 years before 16,936 subjects are enrolled when 12 s are needed to display 3 levels of data. We expect that the system performance will be further dramatically improved with hardware upgrades and Oracle tuning.

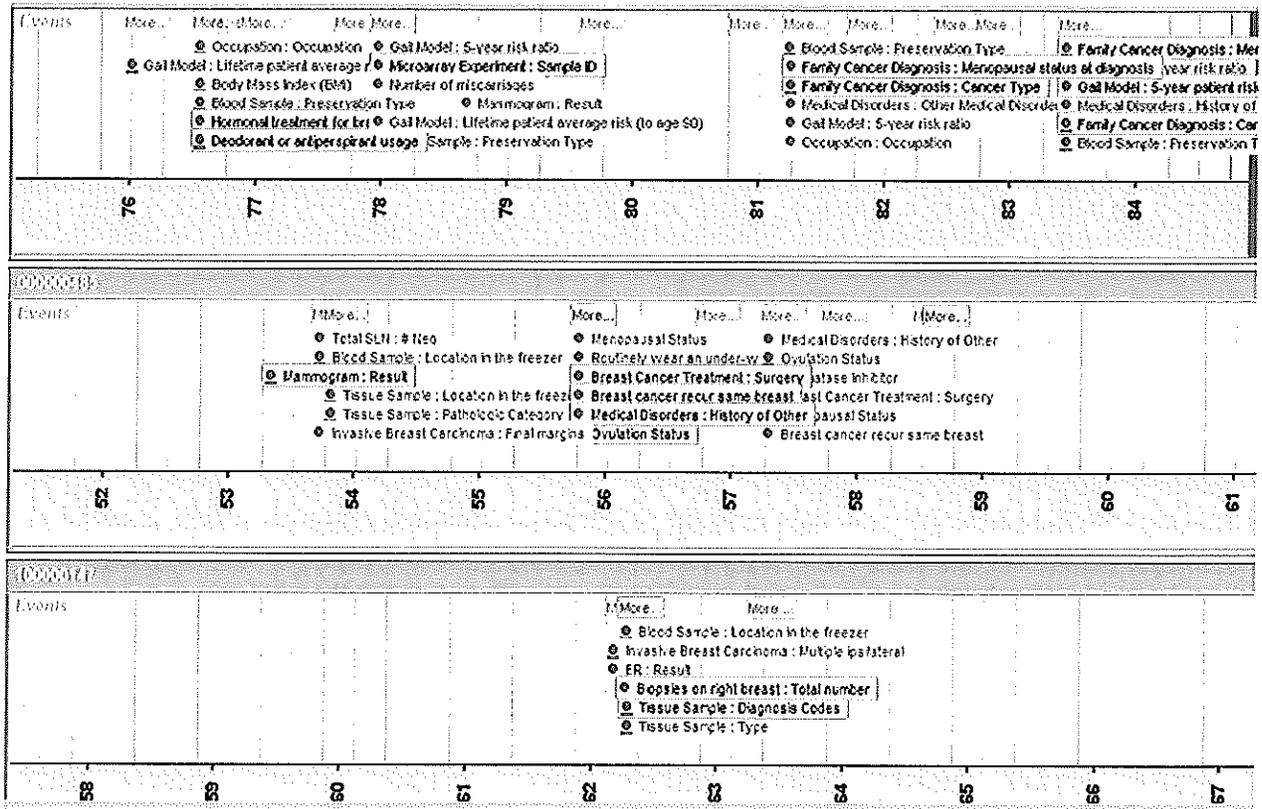


Fig. 7. Screenshot of the ISIV. Temporal information from three subjects is shown. Users can zoom in or out, align the information on a specific event for cross-subject comparison, and explore more detailed information by clicking on the event of interest. The temporal axis shows the age but can be toggled to calendar year.

Table 1  
Example data module definition.

ID	Attribute	Facet	Data type	Units	Multiple values	From date	To date	Temporal information	Parent ID
23	Ethnic Group		VARCHAR2		Yes			Static	
103	Alcohol Use		COMPLEX		No	mm/dd/yyyy	mm/dd/yyyy	Interval	
104	Alcohol Use	Name	VARCHAR2		No	mm/dd/yyyy	mm/dd/yyyy	Interval	103
105	Alcohol Use	Frequency	NUMBER	/day	No	mm/dd/yyyy	mm/dd/yyyy	Interval	103
106	Alcohol Use	Amount	NUMBER	drinks	No	mm/dd/yyyy	mm/dd/yyyy	Interval	103
1541	Surgical Proc		COMPLEX		No	mm/dd/yyyy		Event	
1542	Surgical Proc	Name	VARCHAR2		No	mm/dd/yyyy		Event	1541
1543	Surgical Proc	Code	VARCHAR2		No	mm/dd/yyyy		Event	1541
1544	Surgical Proc	Laterality	VARCHAR2		No	mm/dd/yyyy		Event	1541

3.7. Example use cases

The DW4TR has been playing an important role in supporting translational research across user groups in both WRI and WRAMC. Here we report four example use cases to demonstrate the utility of the system.

3.7.1. The DW4TR enables clinicians to study the integrated data directly

When first exposed to the DW4TR, CBCP clinical users appreciated the simple interface which gave them their first opportunity to directly study the integrated clinicopathologic and biospecimen data. They were happy to see that the results shown on the ABB confirmed their qualitative clinical observations. In one short session of less than 1 h, the program PI identified several lines of evidence which subsequently led to an abstract accepted to the San Antonio Breast Cancer Symposiums [80]. The system has since

been further improved in both the data model structure and the user interface.

3.7.2. The DW4TR enables enhanced cohort and biospecimen selection

Before the DW4TR was developed, cohort and biospecimen selection for any research project was a major task involving multiple data sources including: Core Questionnaires, Pathology Checklists, and IHC assay reports, in the form of an Oracle database, a Microsoft Access database, Excel spreadsheets, and occasionally hard copy reports. Many manual steps were involved, and the procedure was not standardized. The selection of biospecimens was especially challenging when there were temporal restrictions, for example identifying blood samples drawn in a subject cohort before certain kinds of invasive procedures were performed. This requirement alone could take an experienced researcher several days since multiple questionnaires are involved and a number of surgical procedures could be performed on one subject and that there were 15

**Table 2**  
Example classifications of breast neoplasms in MeSH, SNOMED, and DW4TR.

Name	Classifications
MeSH1	3. Diseases (C) → Skin and Connective Tissue Diseases [C17] → Skin Diseases [C17.800] → Breast Diseases [C17.800.090] → Breast Neoplasms [C17.800.090.500]
MeSH2	3. Diseases (C) → Neoplasms [C04] → Neoplasms by Site [C04.588] → Breast Neoplasms [C04.588.180]
SNOMED1	SNOMED CT Concept (SNOMED RT + CTV3) → Clinical finding (finding) → Disease (disorder) → Disorder by body site (disorder) → Disorder of body system (disorder) → Disorder of breast (disorder) → Neoplasm of breast (disorder)
SNOMED2	SNOMED CT Concept (SNOMED RT + CTV3) → Clinical finding (finding) → Finding by site (finding) → Finding of body region (finding) → Finding of trunk structure (finding) → Finding of region of thorax (finding) → Breast finding (finding) → Disorder of breast (disorder) → Neoplasm of breast (disorder)
DW4TR	Medical History → Past Medical History → Major Adult Illnesses → Breast Cancer

different kinds of such procedures. Using the DW4TR interface and additional utilities, the time needed to perform such tasks has now been reduced to minutes. Currently any cohort and biospecimen selection can be done in hours instead of days or weeks.

### 3.7.3. The DW4TR serves as a research environment for risk factor assessment

The DW4TR provides an effective research environment for cancer risk factor analysis. Several studies examining breast cancer risk factors have been done using this system [81–85], and several new studies are underway.

### 3.7.4. The DW4TR facilitates virtual experimental studies

Virtual experimental studies can be performed as integrated information associated with completed research projects is available. Using the information stored in the DW4TR and the file server, we identified the gene expression microarray data from three previous projects, on two types of specimens from three subject groups. From these datasets we were able to create two virtual experiments, and both studies were accepted to leading scientific conferences [86,87].

## 4. Security and accessibility

Data security and access control are important factors in the design of the DW4TR, particularly as our system can be expanded to include multiple research studies. The use of clinical data in translational research is governed by HIPAA compliant IRB approved protocols, and typically cannot be unconditionally shared. Thus we have designed and developed a set of data security and access solutions to satisfy such requirements.

The first layer of security for the system is data de-identification. Prior to being loaded into the DW4TR, PHI is stripped away from the collected data. At this point subjects will have been assigned a unique research subject ID for linking together subject information across multiple sources. The next level of security makes use of the “Virtual Private Database” (VPD) feature of Oracle, which allows for row level access control within the DW. While one of the strengths of the system is the ability to perform cross-study analysis, in many instances it is necessary to restrict data visibility among different users of the system. In our security model, all users are assigned roles which dictate the data they are permitted to access. The VPD system is designed such that every data element in the DW is tagged with a study identifier that is associated with the different user roles, enabling data access control. The third level of security is at the application tier. All access to the DW occurs through a password protected web portal that is served by an application server. The final layer of security is a network firewall and Virtual Private Network (VPN) for which the 168-bit encryption Secure Sockets Layer VPN from Juniper Networks Inc. is deployed. All components of the system, including the DW4TR and the web portal, are protected behind the local network firewall and are only accessible from either within the Institute itself, or by first establishing a VPN connection to the network.

These security and access control measures were developed based on the needs of the supported programs, and can be selectively applied to individual programs. For the two programs currently supported by the DW4TR, CBCP protocols allow data access by all CBCP researchers, and VPN access to the DW4TR is granted with written permission of the CBCP program PI. For GDP, aggregated biomedical information is allowed for use by all the GDP researchers, but the detailed subject information is only allowed to the originating clinical sites unless additional paperwork is completed. Thus, we have designed and developed a role-based site-specific privilege system to enable GDP users to use ABB and the ISIV. DW4TR can also be configured to disable functions and utilities of choice. Currently at the request of the GDP PI, we have disabled the ISIV functions for GDP users so that the GDP data are only available through ABB to GDP users for aggregated information. When a GDP researcher identifies a cohort of interest, additional data use agreements will be completed to request individual subjects' information from the corresponding clinical sites, for approval by the involved sites and the GDP program PI. With a written request from the GDP program PI, WRI will supply the detailed data from the DW4TR to the researcher.

## 5. Discussion

We have taken a bottom-up approach to develop a DW4TR to support translational research programs. The system is extensible, integrating internally generated clinicopathologic and molecular data based on a light-weight ontology, with a patient-centric, modularly structured clinical data model supplemented with a specimen-centric molecular data model. Two interfaces have been developed targeting non-informatician end users, with the ABB supporting aggregated data analysis and ISIV supporting single subject information analysis. The system enables study cohort and biospecimen selection and is capable of handling temporal relationships between clinical data and biospecimen data collected at multiple points in time.

The development of our data model relies on using or creating ontologies for clinical and molecular data types. We will focus this discussion on clinical data ontology. In fact, for molecular data types, the scale of work is much smaller and the existing ontologies are mostly adequate. For example, we found that the MIAME standard met our needs for modeling gene expression microarray metadata. Our analysis indicated that existing ontologies for clinical data cannot be readily adapted to satisfy our specific research needs, due to gaps between existing ontologies and our needs as well as overlaps and inconsistencies between existing standards and ontologies [70,71]. For example, the CBCP Pathology Checklist contains 372 data elements including 131 breast pathology conditions. When the data form was developed and revised, the AJCC (American Joint Committee on Cancer) guidelines and ASCO/CAP (American Society of Clinical Oncology/College of American Pathologists) guidelines were followed; these guidelines, like many other guidelines, have evolved over the years [88–91]. Many data elements present in the CBCP lack standard descriptions, for example

for many benign diseases. Some biomarkers used in the CBCP (Ki67 and p53) do not even have a standard clinical IHC protocol. Therefore, descriptions or definitions of such data elements were developed based on medical textbooks and pathologist's clinical practice for the Pathology Checklist, which we modeled ourselves.

There are classifications in existing standards/ontologies that we find cumbersome or which do not meet our needs [92]. We show as an example in Table 2 the classifications of "Breast Neoplasms" by two existing ontologies. In MeSH, it is classified under "Skin Diseases". This classification (named "MeSH1" in Table 2) is not available in the SNOMED-CT classifications under either Skin Diseases or Connective Tissue Diseases. In MeSH, "Breast Neoplasms" is also classified under "Neoplasms" (in a way we agree with, "MeSH2" in Table 2). However, in the two classifications, "Breast Neoplasms" have different codes, which may cause a problem in the practical applications of data modeling. In SNOMED-CT, the term "Neoplasm of Breast" is used, and there are several classifications leading to it with relatively long paths. We show two of them, "SNOMED1" and "SNOMED2" in Table 2. In our pragmatic approach to support the CBCP, applying such a complete classification is cumbersome. In addition, reconciliation of discrepancies among the available ontologies and our understanding of the problem, for all the data elements in the CBCP, is a major task and not one of our immediate project needs. Therefore, we have developed a classification path for "Breast Cancer", which is more light-weight but adequate to our needs (see "DW4TR" in Table 2).

Finally, some questions in the questionnaires of the two programs we currently support do not comply with existing standards, but we are still required to model them to enable proper storage and presentation of the data. Such problems have been reported to program PIs and were taken into consideration, for example in the consolidation of the GDP questionnaires. In addition, existing ontologies are not static, for example MeSH is currently updated weekly and SNOMED-CT is updated monthly. Thus, after analyzing what is available and what we need to support, we determined that our best approach was to design a light-weight ontology incorporating existing standards and ontologies where possible and to map our classification system to existing ones at a later time. The mapping table(s) will serve as the buffer between the two moving targets which will ensure transportability and un-interrupted use of the system. We will also supplement those standards with new ontologies and common data elements that we define.

The DW4TR currently supports limited number of molecular data types and we will continue to develop data models to support other molecular study platforms, including SNP microarray and array CGH data. We expect that the method developed for gene expression microarray will in general be applicable to these array-based technologies. We will explore support for RT-PCR, tissue microarrays, mass spectrometry, and Next Generation sequencing technologies. These will depend on the needs of the users and programs that we support.

Currently a prototype has been developed to enable access to medical and molecular images through the ISIV, including digital/digitized mammograms and gene expression microarray images. A thumbnail of the image is stored in the database of the DW with searchable annotations. The full-size image is stored in a file server instead of the DW4TR database for performance considerations. We plan to follow the DICOM standard for medical images and will explore how to best annotate molecular images. Additional images we need to support include, pathology H&E slide images, biomarker IHC images, and tissue microarray images. Note that all the medical images have to be de-identified for research purpose.

Currently the DW4TR does not support EMR for data input. We have a stepwise plan to begin using EMR records for CBCP subjects and we are looking at subject de-identification, data extraction, and data model expansion. Although we do not plan to take a

systematic "Enterprise" approach, we will learn from such approaches as used by I2B2 and others. When EMRs are used as an input, a number of functions need to be integrated into the system including subject de-identification and information extraction, for which the I2B2 development teams have made good progresses [50,54,93]. While I2B2 clearly plays an important role in the study design phase including addressing de-identification and natural language processing, it was not designed to be a comprehensive solution that can provide the level of detailed access necessary for later stage research [50–54]. The ABB described here provides the same ability to generate aggregate statistics for population stratification and cohort selection, but by doing this in a web-based pivot table interface it also facilitates trend identification and discovery in a way traditional query interfaces cannot. From a data modeling perspective, the ABB also allows for more sophisticated handling of complex, multi-faceted attributes whereas I2B2 only offers aggregation around the single pivot point of the patient. I2B2 makes good use of ontologies, but has not yet developed an ontology with the level of detail described here [54,94]. Furthermore, by design I2B2 does not enable access to row level clinical data and detailed clinical timeline provided by the ISIV [54,94].

In conclusion, we have developed the DW4TR to support a breast cancer translational research program and then expanded it to support a gynecological disease translational research program. It is based on a light-weight ontology and equipped with an interface capable of handling temporal information, designed for use by non-informatician specialists including clinicians and laboratory scientists. The system supports both *in silico* and *in vitro* studies. With its proven extensibility, we believe that the DW4TR will play an important role in translational research across multiple disease studies.

#### Acknowledgments

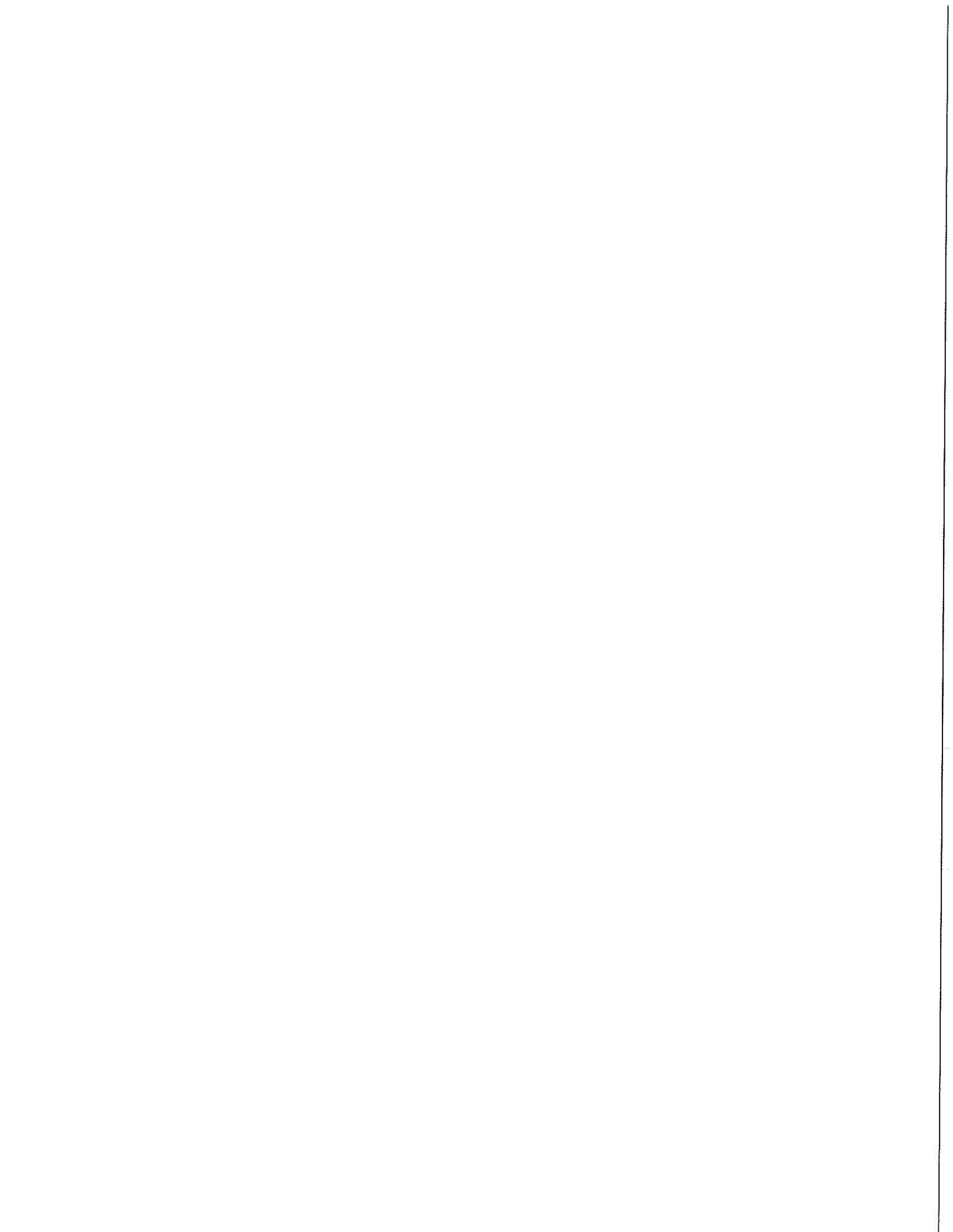
We thank Mr. Anton Oleynikov, Mr. Faruk Cay, Mr. Raveen Sharma, Dr. Csaba Mihaly, Mr. Eric Babyak, and Dr. Yonghong Zhang, for their help in this project. We thank Ms. Joni Kohr, Dr. Hallgeir Rui, and Mr. John Eberhardt for their help in manuscript preparation and revision. We also thank the anonymous reviewers for their constructive comments that resulted in enhanced quality of this paper. This work was supported by the Clinical Breast Care Project, and the Gynecological Disease Program, with funds from the US Department of Defense through Henry Jackson Foundation for the Advancement of Military Medicine, Rockville, MD. The views expressed in this paper are those of the authors and do not reflect the official policy of the Department of the Army, Department of Defense, or the government of the United States.

#### References

- [1] Hu H, Mural RJ, Liebman MN, Hu H, Mural RJ, Liebman MN, et al. Biomedical informatics in translational research. Boston, London: Artech House; 2008.
- [2] Bernstam EV, Hersh WR, Johnson SB, Chute CG, Nguyen H, Sim I, et al. Synergies and distinctions between computational disciplines in biomedical research: perspective from the Clinical and Translational Science Award programs. *Acad Med* 2009;84:964–70.
- [3] Payne PR, Embi PJ, Sen CK. Translational informatics: enabling high-throughput research paradigms. *Physiol Genomics* 2009;39:131–40.
- [4] Sarkar IN. Biomedical informatics and translational medicine. *J Transl Med* 2010;8:22.
- [5] Nadkarni PM, Reeders ST, Zhou J. CECIL: a database for storing and retrieving clinical and molecular information on patients with Alport syndrome. In: *Proc annu symp comput appl med care*; 1993, p. 649–53.
- [6] Hu H, Brzeski H, Hutchins J, Ramaraj M, Qu L, Xiong R, et al. Biomedical informatics: development of a comprehensive data warehouse for clinical and genomic breast cancer research. *Pharmacogenomics* 2004;5:933–41.
- [7] Brammen D, Katzner C, Rohrig R, Weismuller K, Maier M, Hossain H, et al. An integrated data-warehouse-concept for clinical and biological information. *Stud Health Technol Inform* 2005;116:9–14.

- [8] Murphy SN, Mendis ME, Berkowitz DA, Kohane I, Chueh HC. Integration of clinical and genetic data in the i2b2 architecture. In: AMIA annu symp proc; 2006. p. 1040.
- [9] Hu H, Correll M, Osmond M, Gao J, Oleynikov A, Sheldon J, et al. A clinical data warehouse to support translational research. In: 15th annual international conference on Intelligent Systems for Molecular Biology (ISMB), Vienna, Austria; July 21–25, 2007.
- [10] Burgun A, Bodenreider O. Accessing and integrating data and knowledge for biomedical research. *Yearb Med Inform* 2008;91–101.
- [11] Rossille D, Burgun A, Pangault-Lorho C, Fest T. Integrating clinical, gene expression, protein expression and preanalytical data for in silico cancer research. *Stud Health Technol Inform* 2008;136:455–60.
- [12] Hu H. Data centralization. In: Hu H, Mural RJ, Liebman MN, editors. *Biomedical informatics in translational research*; 2008. p. 141–71 [chapter 8].
- [13] Lowe HJ, Ferris TA, Hernandez PM, Weber SC. STRIDE – an integrated standards-based translational research informatics platform. In: AMIA annu symp proc; 2009. p. 391–5.
- [14] Inmon WH, Kelley C. *Developing the data warehouse*. QED Publishing Group; 1993.
- [15] Kimball R, Ross M. *The data warehouse Toolkit*. 2nd ed. John Wiley and Sons; 2002.
- [16] Stead WW, Hammond WE, Straube MJ. A chartless record – is it adequate? *J Med Syst* 1983;7:103–9.
- [17] Niedner CD. The entity-attribute-value data model in radiology informatics. In: *Proceedings of the 10th conference on computer applications in radiology*, Anaheim, CA; 1990.
- [18] Nadkarni PM, Brandt C. Data extraction and ad hoc query of an entity-attribute-value database. *J Am Med Inform Assoc* 1998;5:511–27.
- [19] Codd F, Codd SB, Salley CT. *Providing OLAP (Online Analytical Processing) to user-analysts: an IT mandate*. San Jose: Codd & Date, Inc; 1993.
- [20] Rubin DL, Shah NH, Noy NF. Biomedical ontologies: a functional perspective. *Brief Bioinform* 2008;9:75–90.
- [21] Health Level 7. <<http://www.hl7.org/>>.
- [22] Lowe HJ, Barnett GO. Understanding and using the medical subject headings (MeSH) vocabulary to perform literature searches. *JAMA* 1994;271:1103–8.
- [23] SNOMED Clinical Terms® (SNOMED CT®). <[http://www.nlm.nih.gov/research/umls/Snomed/snomed\\_main.html](http://www.nlm.nih.gov/research/umls/Snomed/snomed_main.html)>.
- [24] DICOM – Digital Imaging Communications in Medicine. <<http://www.medical.nema.org/>>.
- [25] Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The gene ontology consortium. *Nat Genet* 2000;25:25–9.
- [26] Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, et al. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat Genet* 2001;29:365–71.
- [27] BioPAX: Biological Pathways Exchange. <<http://www.biopax.org/>>.
- [28] Sioutos N, de Coronado S, Haber MW, Hartel FW, Shaiu WL, Wright LW. NCI Thesaurus: a semantic model integrating cancer-related clinical and molecular information. *J Biomed Inform* 2007;40:30–43.
- [29] Rubin DL, Lewis SE, Mungall CJ, Misra S, Westerfield M, Ashburner M, et al. National Center for Biomedical Ontology: advancing biomedicine through structured organization of scientific knowledge. *OMICS* 2006;10:185–98.
- [30] Cote RG, Jones P, Apweiler R, Hermjakob H. The Ontology Lookup Service, a lightweight cross-platform tool for controlled vocabulary queries. *BMC Bioinform* 2006;7:97.
- [31] Lindberg DA, Humphreys BL, McCray AT. The unified medical language system. *Methods Inf Med* 1993;32:281–91.
- [32] Bard JB, Rhee SY. Ontologies in biology: design, applications and future challenges. *Nat Rev Genet* 2004;5:213–22.
- [33] Yu AC. Methods in biomedical ontology. *J Biomed Inform* 2006;39:252–66.
- [34] Cimino JJ, Zhu X. The practical impact of ontologies on biomedical informatics. *Yearb Med Inform* 2006;124–35.
- [35] Weedon-Fekjaer H, Lindqvist BH, Vatten LJ, Aalen OO, Tretli S. Breast cancer tumor growth estimated through mammography screening data. *Breast Cancer Res* 2008;10:R41.
- [36] Shoham Y. *Reasoning about change: time and causation from the standpoint of artificial intelligence*. Cambridge, MA, USA: MIT Press; 1988.
- [37] Shahar Y. A framework for knowledge-based temporal abstraction. *Art Intell* 1997;90:79–133.
- [38] Deshpande AM, Brandt C, Nadkarni PM. Temporal query of attribute-value patient data: utilizing the constraints of clinical studies. *Int J Med Inform* 2003;70:59–77.
- [39] Ceusters W, Steurs F, Zanstra P, Van Der Haring E, Rogers J. From a time standard for medical informatics to a controlled language for health. *Int J Med Inform* 1998;48:85–101.
- [40] Das AK, Shahar Y, Tu SW, Musen MA. A temporal-abstraction mediator for protocol-based decision-support systems. In: *Proc annu symp comput appl med care*; 1994. p. 320–4.
- [41] Shahar Y, Das AK, Tu SW, Kraemer FB, Musen MA. Knowledge-based temporal abstraction for diabetic monitoring. In: *Proc annu symp comput appl med care*; 1994. p. 697–701.
- [42] Nguyen JH, Shahar Y, Tu SW, Das AK, Musen MA. A temporal database mediator for protocol-based decision support. In: *Proc AMIA annu fall symp*; 1997. p. 298–302.
- [43] Post AR, Harrison Jr JH. PROTEMPA: a method for specifying and identifying temporal sequences in retrospective data for patient selection. *J Am Med Inform Assoc* 2007;14:674–83.
- [44] Post AR, Sovarel AN, Harrison JH, Jr. Abstraction-based temporal data retrieval for a Clinical Data Repository. In: AMIA annu symp proc; 2007. p. 603–7.
- [45] Chen J, Zhao P, Massaro D, Clerch LB, Almon RR, DuBois DC, et al. The PEPR GeneChip data warehouse, and implementation of a dynamic time series query tool (SGQT) with graphical interface. *Nucleic Acids Res* 2004;32:D578–81.
- [46] Yamamoto Y, Namikawa H, Inamura K. Development of a time-oriented data warehouse based on a medical information event model. *Igaku Butsuri* 2002;22:327–33.
- [47] Informatics for Integrating Biology and the Bedside (i2B2). <<https://www.i2b2.org/>>.
- [48] Mendis M, Wattanasin N, Kuttan R, Pan W, Phillips L, Hackett K, et al. Integration of hive and cell software in the i2b2 architecture. In: AMIA annu symp proc; 2007. p. 1048.
- [49] Mendis M, Phillips LC, Kuttan R, Pan W, Gainer V, Kohane I, et al. Integrating outside modules into the i2b2 architecture. In: AMIA annu symp proc; 2008. p. 1054.
- [50] Uzuner O, Luo Y, Szolovits P. Evaluating the state-of-the-art in automatic de-identification. *J Am Med Inform Assoc* 2007;14:550–63.
- [51] Barrett N, Weber-Jahnke JH. Applying natural language processing toolkits to electronic health records – an experience report. *Stud Health Technol Inform* 2009;143:441–6.
- [52] Deshmukh VG, Meystre SM, Mitchell JA. Evaluating the informatics for integrating biology and the bedside system for clinical research. *BMC Med Res Methodol* 2009;9:70.
- [53] Meystre SM, Deshmukh VG, Mitchell J. A clinical use case to evaluate the i2b2 Hive: predicting asthma exacerbations. In: AMIA annu symp proc; 2009. p. 442–6.
- [54] Murphy SN, Weber G, Mendis M, Gainer V, Chueh HC, Churchill S, et al. Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). *J Am Med Inform Assoc* 2010;17:124–30.
- [55] The stanford translational research integrated database environment (STRIDE). <<http://clinicalinformatics.stanford.edu/STRIDE/>>.
- [56] Hernandez P, Podchyska T, Weber S, Ferris T, Lowe H. Automated mapping of pharmacy orders from two electronic health record systems to RxNorm within the STRIDE clinical data warehouse. *AMIA Annu Symp Proc* 2009;2009:244–8.
- [57] Lowe HJ, Huang Y, Regula DP. Using a statistical natural language parser augmented with the UMLS specialist lexicon to assign SNOMED CT codes to anatomic sites and pathologic diagnoses in full text pathology reports. In: AMIA annu symp proc; 2009. p. 386–90.
- [58] caIntegrator. <<https://cabig.nci.nih.gov/tools/caIntegrator>>.
- [59] Repository of Molecular Brain Neoplasia Data (REMBRANDT). <<http://caIntegrator-info.nci.nih.gov/rembrandt/>>.
- [60] Cancer Genetic Markers of Susceptibility (CGEMS). <<http://cgems.cancer.gov/>>.
- [61] Clinical and Translational Science Award (CTSA, on the website of National Center for Research Resources). <[http://www.ncrr.nih.gov/clinical\\_research\\_resources/clinical\\_and\\_translational\\_science\\_awards](http://www.ncrr.nih.gov/clinical_research_resources/clinical_and_translational_science_awards)>.
- [62] The Cancer Biomedical Informatics Grid (caBIG). <<http://cabig.nci.nih.gov/>>.
- [63] von Eschenbach AC, Buetow KH. Cancer informatics vision: caBIGTM. *Cancer Inform* 2006;2:22–4.
- [64] The Cancer Biomedical Informatics Grid (caBIG): infrastructure and applications for a worldwide research community. *Stud Health Technol Inform* 2007;129:330–4.
- [65] Covitz PA, Hartel F, Schaefer C, De Coronado S, Fragoso G, Sahni H, et al. caCORE: a common infrastructure for cancer informatics. *Bioinformatics* 2003;19:2404–12.
- [66] Komatsoulis GA, Warzel DB, Hartel FW, Shanbhag K, Chilukuri R, Fragoso G, et al. caCORE version 3: implementation of a model driven, service-oriented architecture for semantic interoperability. *J Biomed Inform* 2008;41:106–23.
- [67] McCusker JP, Phillips JA, Gonzalez Beltran A, Finkelstein A, Krauthammer M. Semantic web data warehousing for caGrid. *BMC Bioinform* 2009;10(Suppl. 10):S2.
- [68] The Clinical Breast Care Project. <[www.cbcp.info/](http://www.cbcp.info/)>.
- [69] The Gynecological Disease Program. <<http://www.gyndisease.org/index.html>>.
- [70] Rector AL, Zanstra PE, Solomon WD, Rogers JE, Baud R, Ceusters W, et al. Reconciling users' needs and formal requirements: issues in developing a reusable ontology for medicine. *IEEE Trans Inf Technol Biomed* 1998;2:229–42.
- [71] Richesson RL, Krischer J. Data standards in clinical research: gaps, overlaps, challenges and future directions. *J Am Med Inform Assoc* 2007;14:687–96.
- [72] Mohanty SK, Mistry AT, Amin W, Parwani AV, Pople AK, Schmandt L, et al. The development and deployment of Common Data Elements for tissue banks for translational research in cancer – an emerging standard based approach for the Mesothelioma Virtual Tissue Bank. *BMC Cancer* 2008;8:91.
- [73] Min H, Manion FJ, Goralczyk E, Wong YN, Ross E, Beck JR. Integration of prostate cancer clinical data using an ontology. *J Biomed Inform* 2009;42:1035–45.
- [74] caBIG Vocabularies & Common Data Elements (VCDE) Workspace. <<https://cabig.nci.nih.gov/workspaces/VCDE/>>.
- [75] Brandt CA, Cohen DB, Shifman MA, Miller PL, Nadkarni PM, Frawley SJ. Approaches and informatics tools to assist in the integration of similar clinical research questionnaires. *Methods Inf Med* 2004;43:156–62.
- [76] Wolff AC, Hammond ME, Schwartz JN, Hagerty KL, Allred DC, Cote RJ, et al. American Society of Clinical Oncology/College of American Pathologists

- guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer. *Arch Pathol Lab Med* 2007;131:18–43.
- [77] Hammond ME, Hayes DF, Dowsett M, Allred DC, Hagerty KL, Badve S, et al. American Society of Clinical Oncology/College of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer. *Arch Pathol Lab Med* 2010;134:907–22.
- [78] Beaulah SA, Correll MA, Munro RE, Sheldon JG. Addressing informatics challenges in Translational Research with workflow technology. *Drug Discov Today* 2008;13:771–7.
- [79] Hu H, Kvecher L. Data tracking systems. In: Hu H, Mural RJ, Liebman MN, editors. *Biomedical informatics in translational research*; 2008. p. 111–36 [chapter 7].
- [80] Hu H, Zhang Y, Kvecher L, Sun W, Hooke J, Mural RJ, et al. Different characteristics of invasive breast cancers between Caucasian and African American women. In: *The 29th San Antonio breast cancer symposium*, San Antonio, TX; 14–17 December 2006.
- [81] Maskery SM, Hu H, Hooke J, Shriver CD, Liebman MN. A Bayesian derived network of breast pathology co-occurrence. *J Biomed Inform* 2008;41:242–50.
- [82] Bekhash A, Maskery SM, Kvecher L, Correll M, Zhang Y, Hooke J, et al. Clinical breast care project data warehouse as a research environment for breast cancer risk factor studies; submitted for publication.
- [83] Bekhash A, Maskery SM, Kvecher L, Hooke J, Liebman MN, Shriver CD, et al. A pilot study of known or controversial breast cancer risk factors using the Clinical Breast Care Project database as a research environment. In: *The 30th San Antonio breast cancer symposium*, San Antonio, TX; 13–16 December 2007.
- [84] Saini J, Kovatich A, Bekhash A, Hooke J, Mural RJ, Shriver CD, et al. Association of clinicopathologic characteristics with IHC-based breast cancer subtypes. *Cancer Res* 2009;69:635s.
- [85] Bekhash A, Saini J, Li X, Rapuri P, Hooke AJ, Mural RJ, et al. Ethnicity difference of benign breast diseases in breast cancer and non-cancer patients. *Cancer Res* 2010;288s–9s.
- [86] Saini J, Li X, Kvecher L, Larson C, Croft D, Yang YC, et al. Differential gene expression analysis among post-menopausal caucasian invasive breast cancer, benign and normal subjects. *Cancer Res* 2010;245s.
- [87] Li X, Rapuri P, Melley J, Brillhart G, Wu W, Kvecher L, et al. Comparative analysis of gene expression profiles in human breast cancer from microarray data using breast tissues and peripheral blood samples. In: *International conference on intelligent systems for molecular biology (ISMB)*, Boston, MA; 11–13 July 2010.
- [88] Edge SB, Byrd DR, Compton CC, Fritz AG, Greene FL, Trotti A. *AJCC cancer staging manual*. 7th ed. New York, NY: Springer-Verlag; 2010 [6th printing].
- [89] Lester SC, Bose S, Chen YY, Connolly JL, de Baca ME, Fitzgibbons PL, et al. Protocol for the examination of specimens from patients with invasive carcinoma of the breast. *Arch Pathol Lab Med* 2009;133:1515–38.
- [90] Lester SC, Bose S, Chen YY, Connolly JL, de Baca ME, Fitzgibbons PL, et al. Protocol for the examination of specimens from patients with ductal carcinoma in situ of the breast. *Arch Pathol Lab Med* 2009;133:15–25.
- [91] Printz C. New AJCC cancer staging manual reflects changes in cancer knowledge. *Cancer* 2010;116:2–3.
- [92] Andrews JE, Richesson RL, Krischer J. Variation of SNOMED CT coding of clinical research concepts among coding experts. *J Am Med Inform Assoc* 2007;14:497–506.
- [93] Childs LC, Enelow R, Simonsen L, Heintzelman NH, Kowalski KM, Taylor RJ. Description of a rule-based system for the i2b2 challenge in natural language processing for clinical data. *J Am Med Inform Assoc* 2009;16:571–5.
- [94] Murphy SN, Mendis M, Hackett K, Kuttan R, Pan W, Phillips LC, et al. Architecture of the open-source clinical research chart from Informatics for Integrating Biology and the Bedside. In: *AMIA annu symp proc*; 2007. p. 548–52.



## Gene expression alterations in the lymph node microenvironment in response to successful metastatic colonization

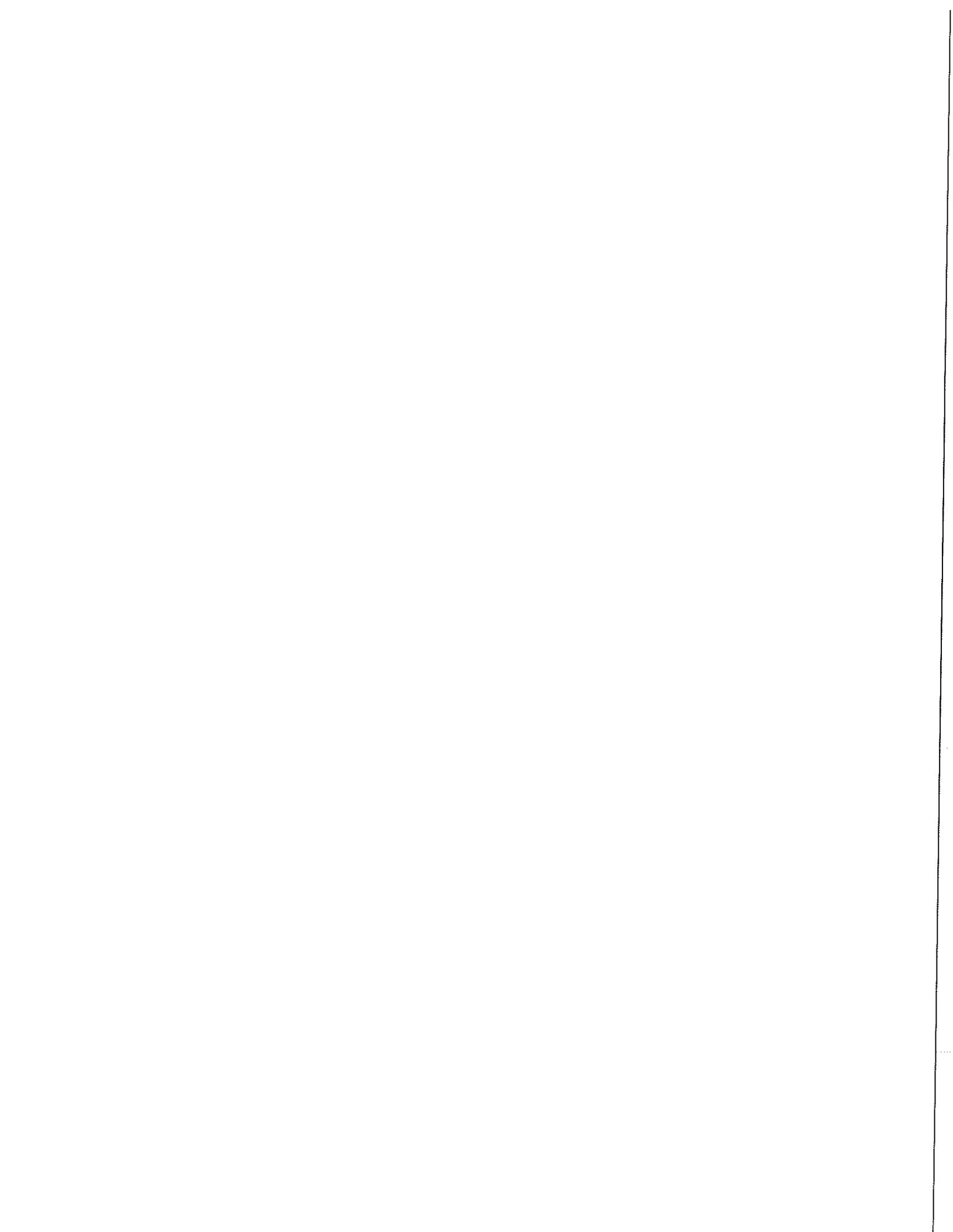
Rachel E. Ellsworth, Allyson L. Valente, Jennifer L. Kane, Darrell L. Ellsworth, Craig D. Shriver

**Background:** Breast stroma is known to play an active role in tumorigenesis, undergoing both phenotypic and molecular changes to facilitate and promote tumor development and growth. The metastatic microenvironment also plays a role in successful colonization; however, the genetic changes in these secondary microenvironments associated with metastasis are not well described.

**Methods:** Women with invasive breast cancer with at least one lymph node with macrometastases and one lymph node with no detectable metastases were identified from the Clinical Breast Care Project. Lymph node tissue was microdissected from both the positive and negative nodes and hybridized to U133A 2.0 gene expression arrays. Differential expression was detected using Partek<sup>®</sup> Genomics Suite<sup>™</sup> 6.5 using a cutoff of  $P < 0.001$ ,  $> 2$ -fold change.

**Results:** Nineteen genes were differentially expressed between negative lymph nodes and lymph node tissue microdissected from positive lymph nodes. Eleven genes, including EPCAM, KRT19 and MUC1 were expressed at significantly higher levels in lymph node tissue from metastatic lymph nodes while eight genes, such as CXCL2 and CXCL5, expressed at significantly higher levels in negative lymph nodes. Results have been validated in external sample sets for AZGP1, CLEC4M, CXCL2, EPCAM, MUC1, PIP and TFPI.

**Conclusions:** Lymph node tissue differs in gene expression between metastatic and non-metastatic lymph nodes. Genes expressed at higher levels in lymph nodes with macrometastases are involved in tumorigenesis, suggesting that like the breast stromal microenvironment, the metastatic microenvironment undergoes cross-talk with the tumor cells. In addition, a cluster of genes involved in immune function are expressed at lower levels in metastatic lymph nodes, suggesting that suppression of proper immune response may be required for successful metastatic colonization.



## Effect of obesity on gene expression in invasive breast tumors

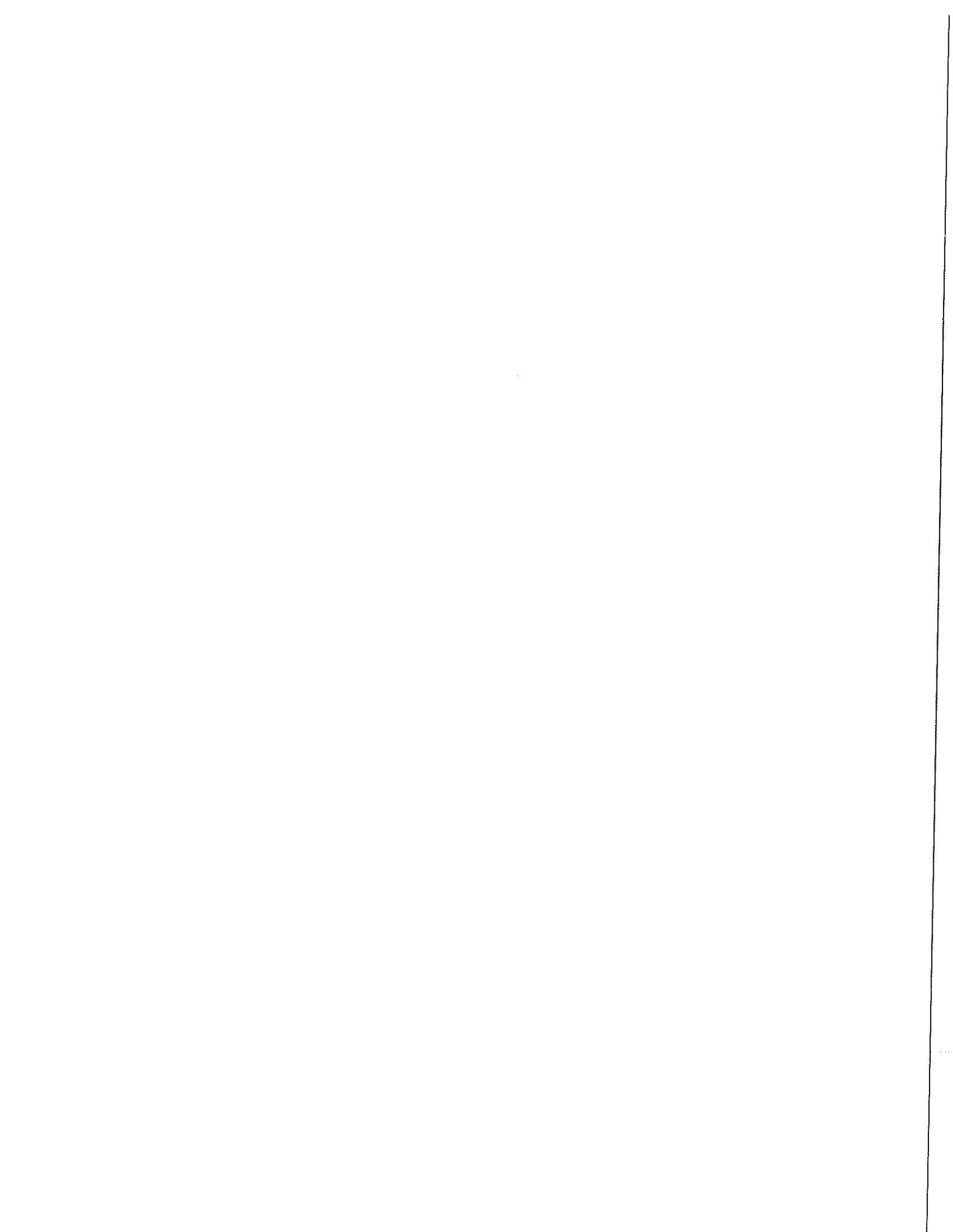
Rachel Ellsworth, Dan Croft, Darrell Ellsworth, Craig Shriver

**Background:** Obesity is a risk factor for breast cancer in postmenopausal women and weight gain after diagnosis is associated with decreased survival. Obesity is also associated with less favorable clinical characteristics such as greater tumor burden, higher grade and poor prognosis, regardless of menopausal status. How obesity affects breast cancer prognosis at the molecular level is, as yet, unknown.

**Methods:** The Clinical Breast Care Project database was queried to identify postmenopausal women with invasive breast cancer who were either a healthy weight (BMI = 18.5-24.9) or obese (BMI  $\geq$ 30) at diagnosis and had frozen tumor tissue available. Patients were matched by age, stage, ER and HER2 status and tumor grade. Gene expression data from U133 2.0 microarrays was generated using RNA from laser microdissected tumor sections. Data was analyzed using Partek Genomics Suite.

**Results:** PCA was able to cluster the specimens into two groups, either healthy or obese. Using a false-discover rate (FDR)  $<0.05$  with a 2.0-fold change, 119 probes from 106 were differentially expressed, with 35 expressed at significantly higher levels and 71 at significantly lower levels in tumors from obese women. Hierarchical clustering with these 119 probes was able to classify 100% of both the healthy-weight and obese samples.

**Conclusions:** Invasive breast tumors from obese women are genetically different from those of healthy-weight women. Genes differentially expressed include lower expression of several members of the histone cluster 1 family, APOD and IGF1, and higher expression of FABP7, members of the major histocompatibility complex and LTF in tumors from obese women. Altered expression of these genes may affect gene transcription, proliferation, cell survival, aggressive phenotypes, and immune response. Together, these underlying molecular differences may contribute to the less favorable prognosis in obese women with breast cancer.



## The effect of breastfeeding on molecular characteristics of invasive breast cancer

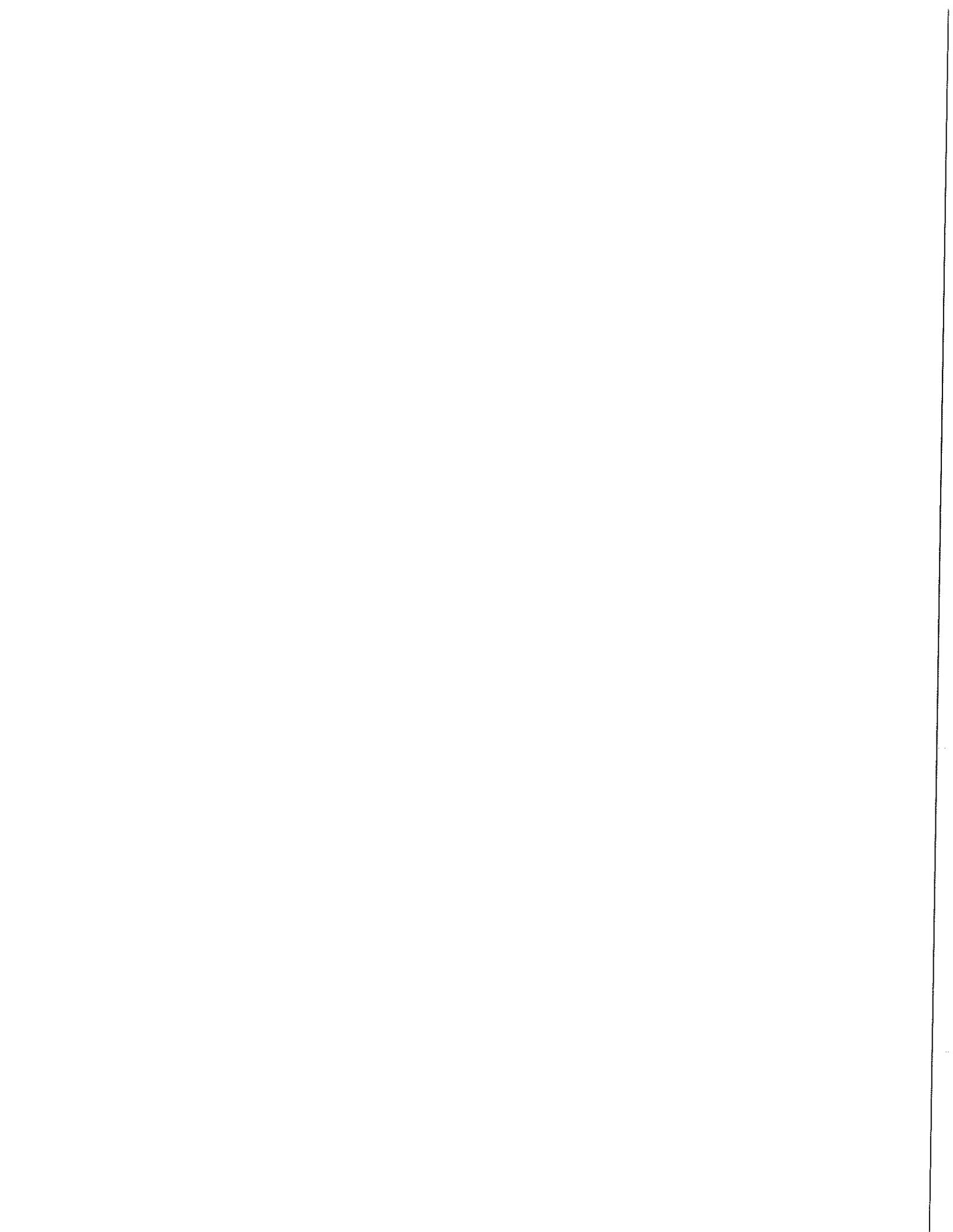
RE Ellsworth, AL Valente, JL Kane, CD Shriver

**INTRODUCTION:** Breastfeeding has been associated with an overall decreased risk of developing breast cancer and the protective effect of breastfeeding has been associated with decreased risk of triple negative and estrogen-responsive breast cancer. Although the mechanism by which breastfeeding provides a protective effect is not well-defined, physical changes in the mammary epithelium reflecting maximal differentiation, may be involved.

**METHODS:** The database of the Clinical Breast Care Project was queried to identify all patients born between 1946 and 1964 (“babyboomers”) with invasive breast cancer who had given birth to at least one living child. Clinicopathologic characteristics were compared between patients who breastfed for at least six cumulative months and those who never breastfed using chi-square and Fisher’s exact tests. RNA was isolated after laser-microdissection of 24 pairs of breastfeeding and non-breastfeeding tumors matched by stage, subtype and grade. Gene expression data was generated using HG U133A 2.0 microarrays and analyzed using the Partek Genomics Suite using a FDR  $P < 0.05$ , 2-fold expression difference to define significance.

**RESULTS:** Of the 1,008 parous women with invasive breast cancer in the database, 325 (33%) breastfed at least 6 months, 165 (16%) breastfed less than 6 months, and 518 (51%). Women who breastfed  $< 6$  months were not included in additional analyses. Age at menarche and first birth and number of children did not differ between groups; age at diagnosis was significantly ( $P < 0.0001$ ) lower (54.5 years) in women who breastfed compared to those who did not (61.0 years). No differences were detected between patients who breastfed and those who did not for ethnicity or for any tumor characteristics, including tumor stage, grade or size, hormone, HER2 or lymph node status, or subtype. At the molecular level, however, 21 probes representing 19 genes were differentially expressed between women who breastfed and those who did not. Eleven genes, including HAPLN1, had significantly higher and 8 genes, including LGALS7, had significantly lower expression in tumors from women who did not breastfeed.

**CONCLUSIONS:** Despite the matching of tumor samples, and lack of differences in pathological characteristics between women who breastfed  $\geq 6$  months and those who did not breastfeed, differentially expressed genes were identified between tumors the two groups. Many of the differentially expressed genes have been associated with prognosis, especially in ER negative breast tumors. These data suggest that breastfeeding does alter the underlying molecular characteristics of invasive breast tumors, and may reflect alterations in exposure to estrogen levels during breastfeeding.



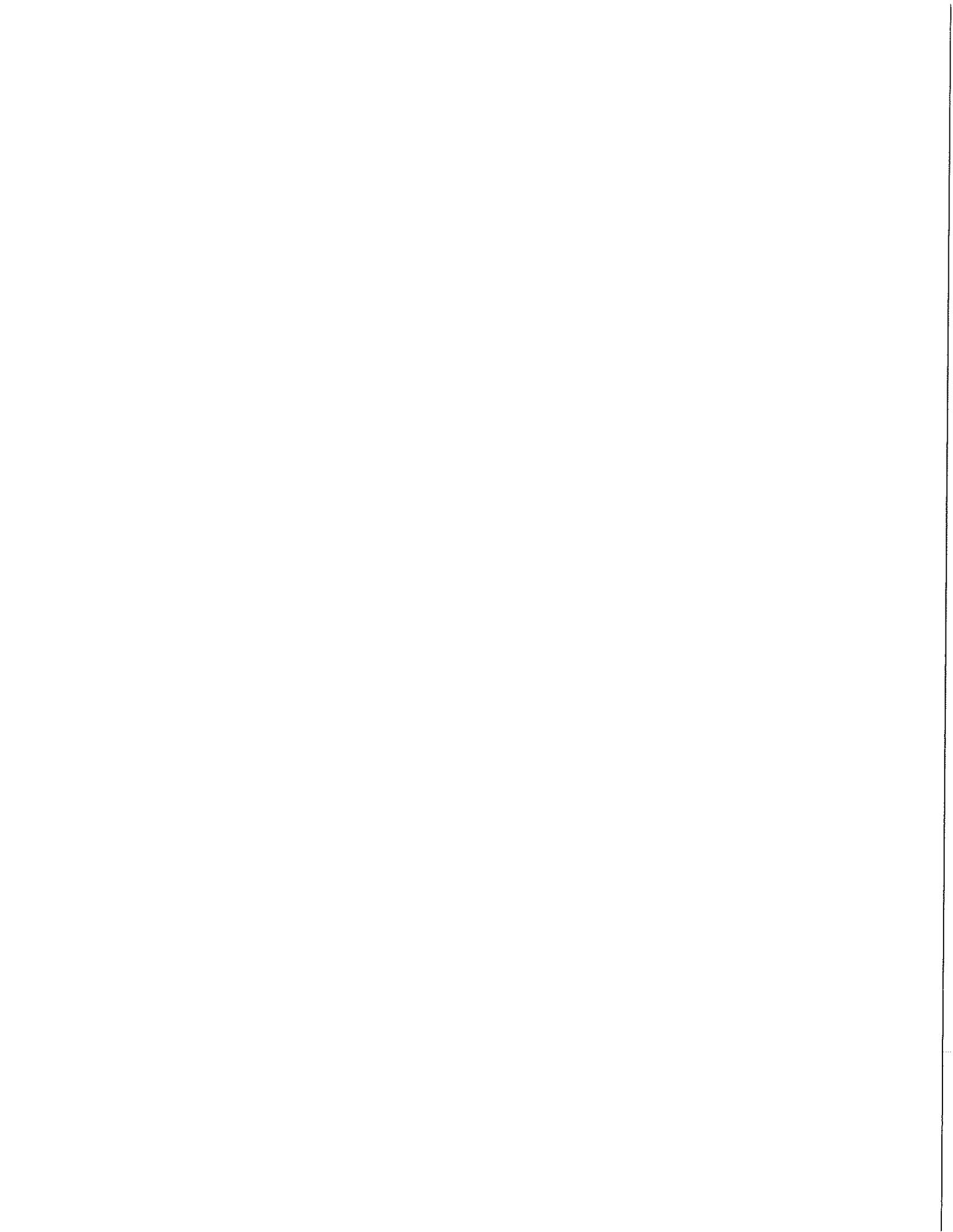
Molecular signature discriminating triple negative breast tumors between African American and Caucasian women  
Rachel Ellsworth, Lori Field, Craig Shriver

**Background:** Triple negative (TN) tumors are characterized by aggressive behavior and, the negative hormone receptor and HER2 status precludes the use of personalized therapies such as tamoxifen and trastuzumab. African American women have higher rates of both breast cancer mortality as well as TN breast tumors. What remains unknown is how much of these disparities can be attributed to molecular differences between populations.

**Methods:** The Clinical Breast Care Project database was queried to identify African American women with triple negative, high-grade tumors. Caucasian women with high-grade TN tumors were then matched by age at diagnosis, tumor stage, size and lymph node status. RNA was isolated after laser microdissection of the tumor and hybridized to HG U133A 2.0 microarrays. Data was analyzed using Partek Genomics Suite.

**Results:** Principal Component Analysis accurately clustered the triple negative specimens by ethnicity. Using a false-discovery rate (FDR) p-value  $<0.05$  with a minimum 2.0-fold change of expression, 132 probes from 117 genes were differentially expressed. Hierarchical clustering using a subset of these probes ( $P < 0.00001$ ) representing 23 known genes reliably partitioned all samples as African American or Caucasian. Genes differentially expressed include lower expression of AQP5, VEGF and RAD51C and higher expression of hemoglobin B, and complement factors CFD and C4A in tumors from African Americans.

**Discussion:** Despite matching of tumors by clinicopathological characteristics, molecular profiles of triple negative breast tumors differed between AAW and CW. These differences may serve as molecular targets for the development of novel prevention strategies and therapeutics to reduce the risk of developing and improve treatment of African American women with triple negative breast cancer.



## **Subtype-specific co-occurrence of atypical hyperplasia and in situ carcinoma with invasive breast cancers**

Albert J. Kovatich, Leonid Kvecher, Yaqin Chen, Anthony Bekhash, Jeffrey A. Hooke, Craig D. Shriver, Richard J. Mural, and Hai Hu

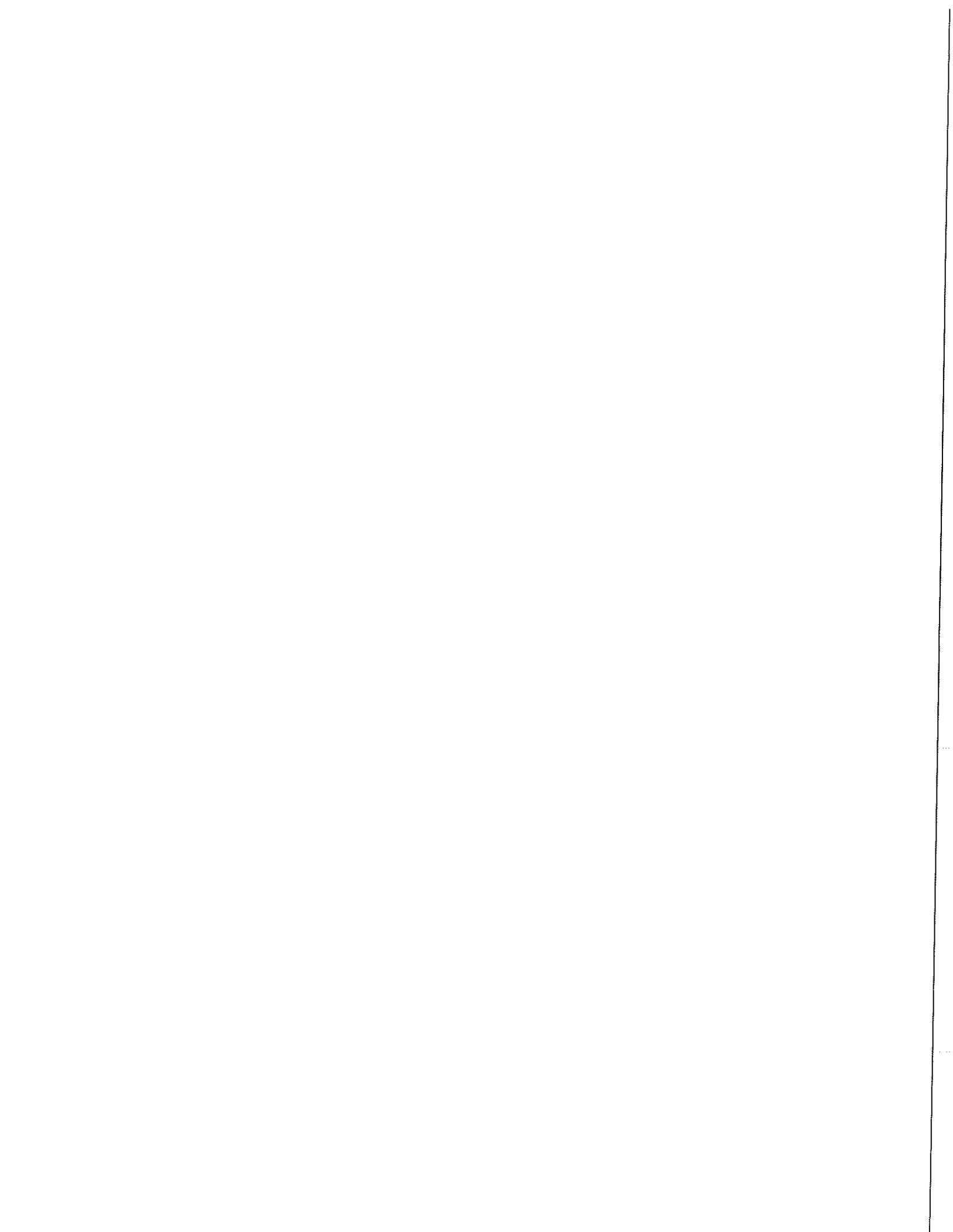
**Background:** Atypical ductal hyperplasia (ADH) and lobular carcinoma in situ (LCIS) are considered risk factors for the development of invasive breast cancer (IBC), while ductal carcinoma in situ (DCIS) is considered a precursor of IBC. Studying their co-occurrence patterns with IBC may provide insights into cancer initiation and development. IBC has been classified into multiple subtypes with distinct prognoses. A clinically practical IHC-based subtyping method has been developed based on the expression of ER, PR, and HER2, defining luminal, HER2+, and Triple Negative (TN) subtypes. Recently, the proliferation marker Ki-67 has been shown to refine the subtyping of luminal IBC into A (LA) and B (LB) subtypes. The Walter Reed Army Medical Center (WRAMC), through the Clinical Breast Care Project (CBCP), has enrolled over 500 IBC subjects. For these patients, all the pathology was reviewed by a single pathologist and the IHC analyses were performed by the same lab. These data provide an opportunity to study the relationship between different subtypes of IBC and the co-occurrence of ADH, LCIS, and DCIS.

**Methods:** Subjects were enrolled in the CBCP following IRB-approved protocols. IBC patients enrolled at WRAMC were selected and their clinical and pathology data were reviewed. A tumor is ER or PR positive if the corresponding nuclear staining is > 5%. The HER2 result is negative if the IHC=0 or 1, positive if IHC=3; for IHC=2, the FISH result determines the final HER2 status. Ki67 is positive if nuclear staining is  $\geq 15\%$ . For IBC subtypes, LA is ER+/HER2-/Ki67-; LB is either ER+/HER2-/Ki67+, or ER+/HER2+; HER2+ is ER-/PR-/HER2+; TN is ER-/PR-/HER2-. Statistical analysis was performed using SAS, and Chi-Square test was used for categorical data analysis supplemented by the Fisher's Exact test where appropriate. For age analysis, ANOVA was performed with Bonferroni adjustment for multi-pair t-test.

**Results:** A total of 459 IBC patients were identified, with the following subtype distribution: LA (41.6%), LB (27.7%), HER2+ (10.2%), and TN (20.5%). Many of the previously reported subtype-specific characteristics were confirmed. Age at diagnosis varied by subtype ( $p=0.0034$ ), with LA being the oldest (Mean $\pm$ SD=59.9 $\pm$ 12.5) and TN the youngest (54 $\pm$ 12.6,  $p=0.0048$ ). The ages for LB and HER2+ were 56.1 $\pm$ 14.3 and 55.5 $\pm$ 12.5 years, respectively. Occurrence of subtypes varied by ethnicity with African American patients having LA (18.13%), LB (30.7%), Her2+ (32.43%), and TN (42.17%) relative to Caucasian American patients ( $p=0.0008$ ). The grade, the AJCC stage and its components T and N were all significantly different among the subtypes, with  $p<0.0001$ ,  $p=0.0002$ ,  $p<0.0001$ , and  $p=0.0020$  respectively. The grades and stages were consistently lowest for LA, highest for HER2+ and TN, with LB being intermediate. We further found that the co-occurrence of ADH, DCIS, and LCIS with IBC were subtype-specific. A total of 78 subjects had ADH reported with LA (25.1%), LB (18.9%), HER2+ (0%), and TN

(6.4%) ( $p < 0.0001$ ); 311 subjects had DCIS distributed as LA (63.4%), LB (76.4%), HER2+ (80.9%), and TN (58.5%) ( $p = 0.0039$ ); 103 subjects had LCIS distributed as LA (36.7%), LB (19.7%), HER2+ (4.3%), and TN (6.4%) ( $p < 0.0001$ ).

**Discussion:** Inclusion of Ki67 in IHC-based IBC subtyping, we confirmed many subtype-specific clinic-pathological characteristics in the CBCP WRAMC population. Using breast prognostic markers, LB subtype was intermediate in grade and stage between the low (LA) and the high (HER2+/TN). We further report subtype-specific co-occurrence of ADH, DCIS, and LCIS. ADH co-occurred more often with LA and LB but rarely with HER2+ or TN. DCIS frequently co-occurred with all subtypes with HER2+ having the highest rate. LCIS co-occurred most often with LA, less frequently with LB, and rarely with HER2+ and TN. These findings may help unveil the development mechanisms of different subtypes of IBC.





Click here to print

[Go Back](#)

---

## Data warehouse for translational research

---

February 2012

by Amy Swinderman | [Email the author](#)

### EDIT CONNECT:

- [IDBS](#)
- [Windber Research Institute \(WRI\)](#)
- [Journal of Biomedical Informatics](#)
- [U.S. House of Representatives Defense Appropriations Subcommittee](#)
- [Clinical Breast Care Project](#)
- [Walter Reed Army Medical Center](#)

### SHARING OPTIONS:

Facebook | Twitter | Digg | LinkedIn | Delicious | Technorati

WINDBER, Pa.—A rare but successful translational research collaboration between data management provider IDBS and nonprofit research organization Windber Research Institute (WRI), detailed in a recent joint paper in the *Journal of Biomedical Informatics*, is expected to yield personalized cancer therapies and improve patient outcomes.

A multidisciplinary team of clinicians, scientists, software developers and biomedical informaticians tackled the daunting challenges of data gathering and analysis challenges, which are often said to place biomedical research years behind the results this growing field seeks.

The WRI is a nonprofit research organization founded more than a decade ago after U.S. Rep. John Murtha, D-Pa. and then-chairman of the U.S. House of Representatives Defense Appropriations Subcommittee, was approached by a concerned group of military spouses regarding the healthcare of women serving in the military—and in breast care, in particular. Murtha decided to support funding for this cause, and the Clinical Breast Care Project at Walter Reed Army Medical Center in Washington, D.C., was born. Using a multidisciplinary approach that brings together prevention, screening, diagnostics, treatment and continuing care, the project integrates advances in risk reduction, biomedical informatics, tissue banking and translational research.

The relationship between Windber and Walter Reed expanded rapidly, and the WRI is today considered to be one of the world's leading genetic research labs. Like many who work in the field of biomedical informatics, the WRI's researchers needed to integrate all of the data they acquired from multiple platforms and make them usable at the point of need—functions that are critical to the success of translational research. However, the organization encountered significant challenges in developing a comprehensive data warehouse system to meet the needs of translational research, including the importance of handling temporal information.

Now, in a study titled "DW4TR: A Data Warehouse for Translational Research" that was published in the December 2011 issue of *The Journal of Biomedical Informatics*, the groups describe the development of a novel system that meets those challenges from data model to interface.

Working with Walter Reed, the WRI and IDBS created a comprehensive data warehouse composed of a complex set of approximately 1,000 detailed breast cancer attributes, including risk factors, patient history, pathology and treatment. Called the Data Warehousing for Translational Research system, or DW4TR, the warehouse integrates multiple data points on cancer attributes such as risk factors, patient history and treatment, and makes this information available in a single source to give scientists the information they need for research into translational medicine or for clinical risk assessment.

With all these clinico-pathological data now available in a single source, scientists can query across the information to support multiple research goals, whether for translational medicine or clinical risk assessment, says Dr. Hai Hu, deputy chief scientific officer and senior director of biomedical informatics at the WRI.

"This is really a dream environment that many people would like to have," says Hu. "When you combine very different perspectives and different languages that have to be translated and achieve these two things, you have success. Such a collaborative environment is critical to the success of translational research, and to achieve that is not always easy."

The software development and implementation was done by InforSense, which was acquired by IDBS in July 2009. Designing and implementing the DW4TR posed a number of engineering challenges, but by borrowing many concepts from traditional business intelligence projects and applying those methods effectively to longitudinal clinical data, IDBS was able to combine data modeling and user interfaces, says Dr. Paul Denny-Gouldson, vice president of translational medicine at IDBS.

"This type of data management system enables you to get from bench to bedside, then back to the bench," he notes. "You may not have everyone in one single environment, but we're now able to link them together and provide an ecosystem of data capture and integration in one environment."

Now that the academic value of the DW4TR has been demonstrated, "the system has already enabled us to expand into other studies, and we are also seeking extramural funding," says Hu.

Headquartered in Guilford, U.K., IDBS' platform technologies are used by more than 200 pharmaceutical companies, major healthcare providers, global leaders in academic study and high-tech companies to increase efficiency, reduce costs and improve the productivity of industrial R&D and clinical research.

---

#### **IDBS, Royal Society of Chemistry automate online publishing of chemistry research**

GUILDFORD, U.K.—IDBS also recently announced the release of real-time publication capability for research chemists directly from its electronic notebook. According to IDBS, the approach enables researchers to publish their chemistry data to send it directly to the ChemSpider online compound database, providing "seamless access to the most comprehensive view of freely available chemical data."

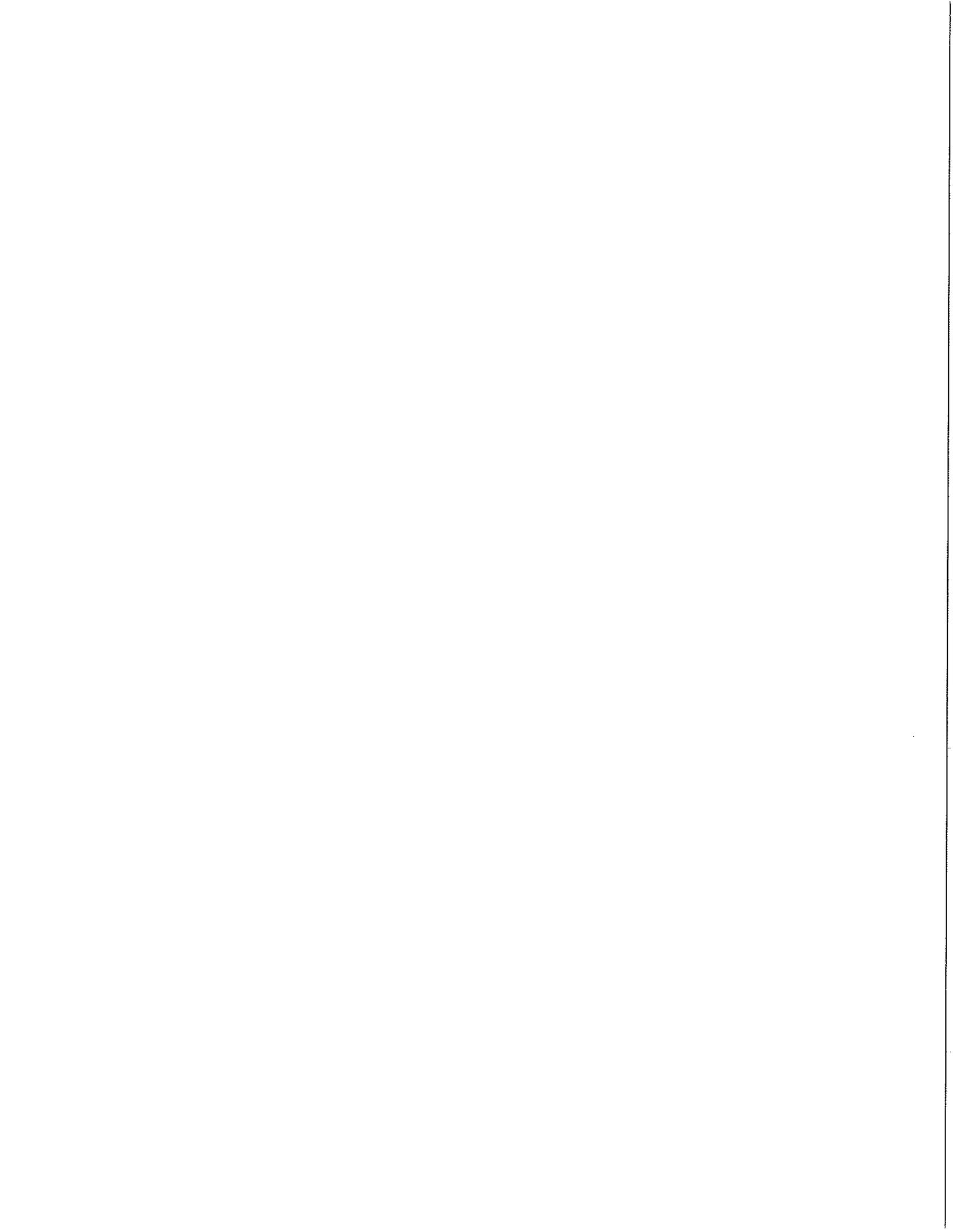
According to IDBS, researchers using E-WorkBook can now connect to ChemSpider to share and reuse their work, eliminating unnecessary repeated experiments "and providing valuable contributions to the world's knowledge of chemistry." The capability was co-developed by the Royal Society of Chemistry in partnership with the University of Cambridge, and is freely available to all users of E-WorkBook.

"Much of chemistry research is essentially lost, with many results simply not published and left languishing in forgotten lab notebooks," said Prof. Robert Glen, professor of molecular sciences informatics and director of the Unilever Centre for Molecular Sciences Informatics at University of Cambridge, in a press release. "Now, capturing novel molecules soon after synthesis on a searchable database like ChemSpider is an effortless process driven directly from the electronic laboratory notebook. This will encourage sharing of compounds, synthetic methods and all the associated data, and will significantly increase collaboration and cut research time. It's instant messaging for chemists."

The Royal Society of Chemistry is the United Kingdom's professional body for chemical scientists and an international learned society for the chemical sciences. With more than 48,000 members worldwide, the society is a major international publisher of chemical information.

Code: E021211

[Back](#)



## Should Immunohistochemical Markers be Performed on the Surgical Specimens Rather than the Core Needle Biopsy in Breast Cancer Patients?

**Authors:** Lauren T. Greer, M.D., Martin Rosman, M.D., Charles Mylander, Ph.D., Albert J. Kovatich, M.S., Jeffrey Hooke, M.D., Joan Woodward, M.D., J. Leigh Campbell, MS, PA, Janet Wareham, MS, PA, Wen Liang, D.O., Robert R. Buras, M.D., FACS, Craig D. Shriver, M.D., FACS, Lorraine Tafra, M.D., FACS

### **Objectives.**

Breast tumor heterogeneity, defined as distinct areas within single or multiple ipsilateral tumor(s) with either different histologic presentations on H&E or significant biomarker discrepancies (some leading to molecular subtype changes), may be present in up to 10% of breast cancer patients. Biomarkers from core needle biopsies (CNB) of the primary tumor may not be representative of the entire surgical specimen (SS) in heterogeneous tumors. Our aim is to investigate the level of concordance between the CNB and the surgical specimens (SS) in heterogeneous tumors.

### **Method**

A retrospective review of a prospective study was performed on women diagnosed with breast cancer from January 2009 to June 2011. Subjects treated with neoadjuvant therapy were excluded. Tumor type, grade, ER, PR, HER2, Ki67 expression by IHC were analyzed in the CNB and SS. Subjects with tumor heterogeneity were identified (at least 2 samples of IHC markers performed per subject). Contingency tables and agreement modelling were performed for heterogeneous subjects.

### **Results**

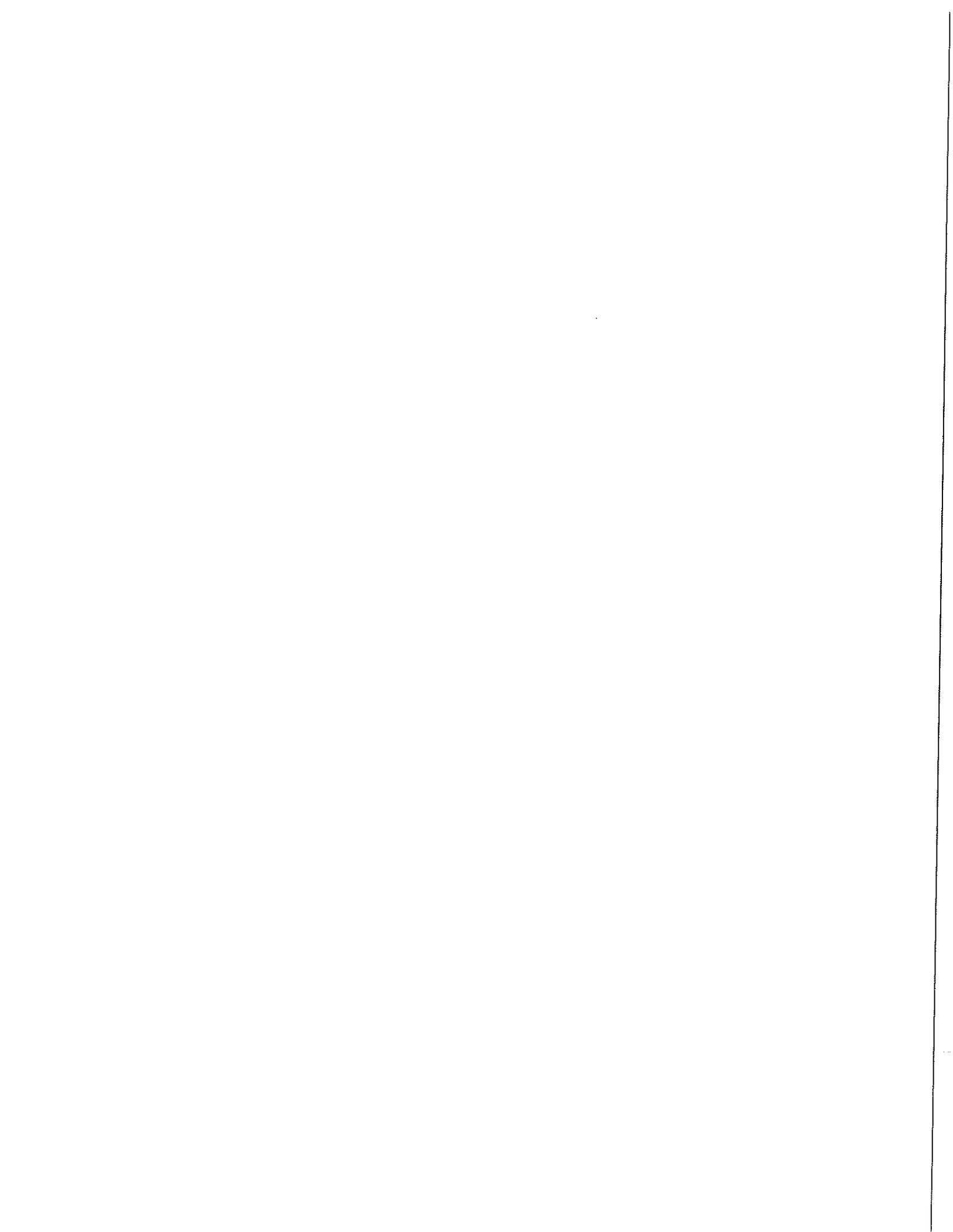
Tumor heterogeneity was identified in 12 (30 samples) out of 194 subjects. Four subjects had histologic differences on H&E and 8 had significant biomarker discrepancies allowing initial classification of heterogeneity. Tumor size ranged from 2mm to 7cm. One patient was noted to have 4 distinct molecular subtypes within the same tumor. There was substantial agreement between the CNB and SS for histologic grade; moderate agreement for PR%; fair agreement for tumor type and Ki67%; and slight agreement for HER2 (see Chart). A kappa value could not be generated for ER% since there were no ER negatives in the CNB. Four subjects had areas of their tumors that stained positive for HER2, which was not detected in their CNB. Three subjects had areas in their SS that were negative for ER and PR, but stained positively in their CNB.

Core Needle Biopsy (n)	Surgical Specimen (n)			Kappa ( $\kappa$ )	Landis-Koch Agreement Grade	
	Tumor Type	IDC	ILC			Mixed
	IDC	18	0	4	0.22	Fair
	ILC	0	3	0		
	Mixed	4	1	0		
Histologic Grade	1	2	3	0.61	Substantial	
1	2	0	0			
2	1	6	3			
3	0	2	15			
ER %	Negative (<1%)	Positive ( $\geq$ 1%)	N/A	N/A		
Negative (<1%)	0	0				
Positive ( $\geq$ 1%)	5	25				
PR %	Negative (<1%)	Positive ( $\geq$ 1%)	0.53	Moderate		
Negative (<1%)	4	0				
Positive ( $\geq$ 1%)	5	21				
HER2	Negative (0/1+/2+)	Positive (3+)	0.05	Slight		
Negative (0/1+/2+ FISH-)	21	5				
Positive (3+/2+ FISH+)	3	1				
Ki-67 %	Low: $\leq$ 14%	High: >14%	0.23	Fair		
Low: $\leq$ 14%	3	1				
High: >14%	7	11				

## Conclusions

Tumor heterogeneity in invasive breast cancer is difficult to identify on H&E alone in the SS and not routinely detected in CNB. The accuracy of biomarkers on the CNB of heterogeneous tumors is not always representative of the biology of the whole tumor(s). Since many heterogeneous tumors are detected by biomarkers alone in the SS, possibly reserving immunohistochemical staining only for the SS, would allow a more cost-effective and tailored treatment of breast cancer patients

"The views expressed in this abstract are those of the authors and do not reflect the official policy of the Department of the Army (DOA), Department of Defense (DOD), or US Government."



## Evaluation of BRCA1 mutations in patients with Triple Negative breast cancer

Seth Rummel, Erika Varner, Rachel Ellsworth, Craig Shriver

**Background:** Triple negative (TN) breast cancer (ER, PR and HER2 negative) is characterized by aggressive behavior and poor prognosis and until recently targeted treatments were not available. Ongoing trials of PARP inhibitors, which exploit defects in the DNA repair pathway caused by mutations in BRCA1, may offer the first tailored treatment for patients with TN disease, however, whether PARP inhibitors in both patients with and without BRCA1 mutations, remains to be determined. The prevalence and nature of BRCA1 mutations in a population of women with TN breast cancer with and without significant family history was thus evaluated.

**Methods:** The Clinical Breast Care Project database was queried to identify all female patients with TN tumors with and without significant family histories defined as being diagnosed  $\leq 35$  years, diagnosed  $\leq 50$  years with at least one primary family member with breast and/or ovarian cancer, or diagnosed  $< 50$  years with at least two secondary family members from the same parental branch from two generations diagnosed with breast and/or ovarian cancer. Genomic DNA was isolated from blood; 36 sets of PCR primers were designed amplify each exon including 100 bp of intronic sequence on either side of each exon. PCR products were sequenced on an ABI3730xl and sequence data was analyzed using Sequencher 4.10.1.

**Results:** Of the 162 women with TN tumors, genomic DNA was available for 154. Twenty-seven of these women had a significant family history of breast cancer, of which nine had clinically relevant mutations, including the C61G mutation in two Caucasian women, and the 943ins10 West African founder mutation in one African American woman. An additional two African American women with family histories harbored the 2090A/G and R841W variants of unknown clinical significance, while in the remaining 14 (52%) of women with significant family histories of breast cancer BRCA1 mutations were not detected. Of the 128 women without significant family history, five had clinically relevant mutations, including one Caucasian woman with the C61G mutation diagnosed at age 57, 15 had variants of unknown significance and 107 (84%) had no detectable mutations. Of the women without BRCA1 mutations, intrinsic subtype was determined for 37; 33 of which were basal-like.

**Conclusion:** The majority of TN tumors are not driven by germline BRCA1 mutations with only 9% and 11% of the overall TN population harboring clinically relevant mutations and variants of unknown significance, respectively. These results suggest that most TN tumors are sporadic in nature and may be driven by sporadic mutations or epigenetic changes in BRCA1, or disruptions of other genes in the DNA repair pathway. In addition, given the different genetic backgrounds of heritable compared to sporadic TN tumors not all TN tumors may respond in the same way to PARP inhibitors. Finally, the presence of BRCA1 mutations in TN patients without significant family histories and/or diagnosed at a later age and absence in patients with significant family histories suggests that current criteria for identifying patients likely to harbor BRCA1 mutations may need to be reevaluated in the triple negative patient populations.

**Impact of lifestyle factors on quality of life and prognosis in breast cancer survivors in the United States**

Rachel E. Ellsworth, Allyson L. Valente, Craig D. Shriver, Barry Bittman, Darrell L. Ellsworth

Rachel E. Ellsworth  
Clinical Breast Care Project  
Henry M. Jackson Foundation for the Advancement of Military Medicine  
620 Seventh Street  
Windber, PA 15963  
Phone: 814-361-6926  
FAX: 814-467-6334  
Email: r.ellsworth@wriwindber.org

Allyson L. Valente  
Clinical Breast Care Project  
Windber Research Institute  
620 Seventh Street  
Windber, PA 15963  
Phone: 814-361-6976  
FAX: 814-467-6334  
Email: a.valente@wriwindber.org

Craig D. Shriver  
Clinical Breast Care Project  
Walter Reed National Military Medical Center  
8901 Rockville Pike, Bldg 19, 3rd floor  
Bethesda, MD 20889  
Phone: 301-295-6333  
Fax: 301-295-9076  
Email: craig.shriver@med.navy.mil

Barry Bittman  
Mind-Body Wellness Center  
18201 Conneaut Lake Road  
Meadville, PA 16335  
Phone 814-333-5061  
FAX: 814-333-5067  
Email: bbittman@mmchs.org

Darrell L. Ellsworth  
Clinical Breast Care Project

Windber Research Institute  
620 Seventh Street  
Windber, PA 15963  
Phone: 814-361-6911  
FAX: 814-467-6334  
Email: [d.ellsworth@wriwindber.org](mailto:d.ellsworth@wriwindber.org)

## **Summary**

Advances in diagnostic screening and adjuvant therapy have dramatically increased the number of breast cancer survivors in the United States, who face numerous issues of physical and mental health, social support, quality of life, and economics. Women living with breast cancer are increasingly interested in lifestyle modification to decrease risk of recurrence and mortality while increasing physical and emotional well-being. Although a healthy diet, frequent physical activity, and stress reduction may be beneficial for decreasing breast cancer risk, studies examining the effects of lifestyle on clinical outcomes such as survival and prognosis have been inconclusive. With the number of breast cancer survivors predicted to increase to 3.4 million by 2015, it is important to develop more effective treatment paradigms that overcome barriers to behavioral modification to improve clinical outcomes and survivorship in breast cancer patients.

**Running Title**

Lifestyle modification in breast cancer survivors

**Keywords:** Breast cancer, lifestyle modification, survivors, diet, physical activity, stress management, quality of life

## **Introduction**

Breast cancer is the most frequently occurring cancer in women and is second only to lung cancer in lethality. Over the last fifty years, lifetime risk of breast cancer has increased such that today one in eight women are expected to develop breast cancer. Projections indicate that more than 230,000 new breast cancers will be diagnosed, and ~40,000 women will die from breast cancer, in the United States (US) in 2011 [1]. Mortality rates increased from 1970 to 1990, but in recent years, death rates have decreased by 28% [2]. This notable decline in mortality, likely attributable in large part to improvements in early detection and adjuvant therapy [3], has produced more than 2.5 million breast cancer survivors in the US today [201].

Breast cancer survivors face numerous issues that can degrade quality-of-life (QOL), including detrimental changes in social support, psychological health, personal finance, and physical activity [4]. Survivors may suffer from anxiety, depression, fear of recurrence, impaired body image, and increased risk of co-morbid conditions [5-7]. With the number of breast cancer survivors in the US predicted to increase to 3.4 million by 2015 [8], it is imperative to develop interventions that will improve QOL in breast cancer survivors.

Studies of behavioral modification in breast cancer survivors are providing new information about how lifestyle factors such as diet and related obesity, physical activity, and stress affect survivorship, as well as knowledge to develop new, effective intervention programs for survivors. In this article, we outline current scientific evidence demonstrating the effects of diet, physical activity, and stress management on QOL in breast cancer survivors, and discuss barriers to widespread implementation of lifestyle modification strategies.

## **The economic impact of breast cancer**

The economic impact on cancer survivors is often overlooked, but has great potential to negatively impact QOL. The National Cancer Institute (NCI) estimates that direct costs for care of women with breast cancer in the US was \$16.5 billion in 2010 [202]. Average lifetime costs for new personalized treatments such as trastuzumab (Herceptin<sup>®</sup>) for HER2 positive breast cancer are \$44,923 for early breast cancer patients and \$47,728 for women with metastatic disease [9]. Medical out-of-pocket costs can exceed \$1,000 per month during active treatment, and are ~\$500 per month at one year post-diagnosis, coupled with non-medical out-of-pocket costs typically ranging from \$200-\$500 per month [10]. For many patients living with breast cancer, the monetary burden associated with treatment and survival can accentuate stress, precipitate unhealthy behaviors, and impact QOL.

### **Lifestyle modification and human health**

Lifestyle choices and behavioral modifications can significantly impact overall health, as well as susceptibility to cancer, cardiovascular disease, and diabetes. While improvements in hygiene and sanitation have led to a decrease in infectious diseases, the transition from primarily plant-based diets to modern (Western) diets has been associated with an increased incidence of hypertension, heart disease, diabetes, and obesity [11]. Observations that women native to countries with low rates of breast cancer who migrate to the US and adopt a Western lifestyle develop a higher risk of breast cancer suggest that behavior and environmental factors also influence breast cancer etiology [12].

A variety of epidemiological study designs have been used to examine environmental factors and behaviors that contribute to risk of developing breast cancer or long-term prognosis (Table 1). Observational studies performed prospectively or retrospectively usually identify associations between behaviors and outcomes of interest [13]. Observational trials can assess the effectiveness of treatments in diverse patient populations, identify differences among therapeutic options, and examine long-term

effectiveness [14], but may be susceptible to bias and confounding, which limit their ability to measure causality or compare treatments. Conversely, randomized controlled trials (RCTs) are considered the gold standard for measuring the efficacy of interventions on particular outcomes, such as disease recurrence or prognosis. In an RCT, patients with similar characteristics are randomly assigned to a group receiving the intervention or to a control group who receive a placebo or standard treatment. At the conclusion of the trial, a significant difference between the intervention and control groups in the outcome of interest can be attributed to the intervention and not to other factors [15]. Much of the research presented in this paper was derived from observational studies and RCTs measuring the effects of diet, physical activity, and stress reduction on prognosis and QOL in breast cancer survivors.

## **Diet and breast cancer prognosis**

### ***Dietary fat, fiber, fruit, and vegetable intake***

The majority of dietary studies examining the impact of body weight on breast cancer survival observed that obesity at diagnosis is associated with poor prognosis [16]. Similarly, weight gain after diagnosis is common and is associated with mortality, disease recurrence, and development of co-morbid conditions including diabetes and cardiovascular disease [17]. In patients with breast cancer, each five-kilogram gain in weight has been associated with a 13% increase in risk of breast cancer mortality, a 12% increase in all-cause mortality, and a 19% increase in cardiovascular mortality [18]. Given the prediction that 41% of adults in the US will be obese by 2015 [19], improved understanding of the relationship between obesity and prognosis is crucial to developing effective strategies to improve survival and QOL.

The Nurses Health Study (NHS), one of the largest prospective cohort studies in the US, was established in 1976 to follow 121,700 female nurses over time through periodic health assessment

questionnaires. Contrary to earlier studies suggesting that a low-fat diet is associated with improved prognosis [20], food frequency data from 1,982 NHS participants with invasive breast cancer collected over 10 years showed no association between fat intake and mortality [21] (Table 2). A second NHS investigation of 2,619 women with dietary data collected before, and at least one year after, a diagnosis of invasive breast cancer compared women who consumed a “prudent” diet, high in fruits, vegetables, whole-grains, fiber, and protein to women following a “Western” diet, characterized by high amounts of refined grains, processed meat, high-fat dairy, and low amounts of protein and fiber. Although women consuming a Western diet before and after breast cancer diagnosis had a higher risk of all-cause mortality, diet did not impact breast cancer prognosis [22]. A third NHS study evaluating change in body mass index (BMI) in 5,204 women found that weight gain after breast cancer diagnosis was associated with less favorable prognosis, including higher rates of recurrence and mortality [23]. Therefore, weight gain, independent of individual dietary components, may be associated with poor prognosis in breast cancer survivors.

A second large, prospective cohort study, the Life After Cancer Epidemiology (LACE) Study, enrolled 2,321 women from a California-based health-maintenance organization (HMO) and the Utah cancer registry who were diagnosed with breast cancer between 1997 and 2000. Dietary intake of more than 100 foods, beverages, nutritional supplements, and medicinal herbs was assessed by food frequency questionnaires. Preliminary data suggest that dietary patterns of women with breast cancer were similar to those in women free from breast cancer in the general population [24]. Similar to the NHS study, a Western diet was associated with increased risk of all-cause mortality, but was not associated with breast cancer mortality [25].

In the Health, Eating, Activity, and Lifestyle (HEAL) Study, a significant number of African American and Hispanic participants were recruited from California, Washington, and New Mexico. In

this study, women gained an average of 1.7 kg of body weight within three years of breast cancer diagnosis. Overall, 68% of women gained weight and 74% increased their body fat [26]. Women (n=670) with early stage breast cancer, who consumed a better-quality diet (low in calories, added sugar, alcohol, and saturated fat) had a 60% lower risk of all-cause mortality and an 88% lower risk of breast cancer-related mortality [27]. A quality diet also was associated with increased mental and physical QOL, as well as decreased levels of circulating inflammatory markers [28,29].

In contrast to the observational studies described above, the Women's Intervention Nutrition Study (WINS) is a randomized trial measuring the effects of a low-fat diet on relapse-free survival. Women in the intervention group (n=975) undergoing standard treatment for stage I or stage II breast cancer followed a low-fat dietary intervention, while controls (n=1,462) received minimal nutrition counseling. The intervention group showed a significant decrease in fat intake and maintained lower fat consumption for more than five years, while patients in the control group showed no significant change in fat intake. After 60 months of follow-up, relapse-free survival was 24% higher in the intervention group [30]. Although this study focused solely on reducing dietary fat, participants in the intervention arm also increased fruit consumption and lost significantly more weight than controls.

In the Women's Healthy Eating and Living (WHEL) Study, women with stage I–IIIA breast cancer were randomized to an intervention arm (n=1,537) or control arm (n=1,551). The intervention group consumed a high-fruit and vegetable diet, with only 15%-20% of calories from fat, while controls were advised to follow the 5 A Day for Better Health Program diet recommended by the NCI. Nutritional intake for both groups was similar at baseline, but after four years, the intervention group increased vegetable consumption by 65%, fruit consumption by 25%, fiber by 30%, and decreased calories from fat by 13% compared to the control group. Despite improvements in nutrition, breast

cancer event-free survival and mortality did not differ significantly between the intervention and control groups [31].

These studies suggest that a low-fat, high-fiber diet may have only marginal effects on breast cancer mortality. Given the link between poor prognosis and obesity at diagnosis, as well as weight gain after diagnosis, positive effects on survival observed in the WINS may be attributable to incidental weight loss, rather than specific dietary components. Although diet may have no direct effect on breast cancer mortality, healthy eating has been shown to reduce the risk of developing other chronic diseases, such as cardiovascular disease and diabetes. Thus, the American Cancer Society (ACS) recommends a “heart healthy” diet similar to that recommended for cardiovascular health for breast cancer survivors.

### *Alcohol*

In a large meta-analysis that included 58,515 women with invasive breast cancer and 95,067 controls from 53 studies, alcohol consumption was associated with ~7% increased risk of developing breast cancer for every 10 g (unit or drink) of alcohol consumed on a daily basis [32]. Relationships between alcohol consumption and breast cancer survival, however, have not been well defined. One study of 1,268 women  $\leq 45$  years of age with invasive breast cancer assessed the effect of pre-diagnostic alcohol consumption on mortality and determined that women who consumed alcohol during the five-year period prior to diagnosis had a decreased risk of breast-cancer-related mortality compared to non-drinkers [33]. Similarly, the population-based Studies of Epidemiology and Risk Factors in Cancer Heredity (SEARCH) Study of 4,560 breast cancer patients detected a 2% reduction in mortality risk for every unit of alcohol consumed per week [34]. In contrast, the LACE Study detected a 1.3-fold increase in breast cancer recurrence and a 1.5-fold increase in mortality among overweight and obese women who had 3-4 alcoholic drinks per week [35]. Other studies have shown

no relationship between alcohol consumption and survival [36,37]. Given the protective effects of certain types of alcohol on cardiovascular disease, a personal decision on alcohol consumption should be balanced between breast cancer risk and cardiovascular health.

### *Nutraceuticals and functional foods*

Functional foods provide health benefits beyond nutrition. For example, yogurt is a functional food containing live microorganism cultures that do not add nutritive value, but inhibit pathogens and toxin-producing bacteria. In contrast, “nutraceuticals” represent hybrids between traditional nutrients and pharmaceutical compounds, usually available as pills or powders containing bioactive components that provide health benefits [38]. As 75%-87% of breast cancer survivors take dietary supplements to improve survival [39], it is important to understand how functional foods and nutraceuticals affect breast cancer prognosis.

### *Vitamins*

Nearly 50% of adults in the US use vitamin supplements, even though research has provided no convincing evidence of true beneficial effects [40]. At present, the US Preventive Services Task Force and the National Institutes of Health Office of Dietary Supplements and Office of Medical Applications do not recommend use of multivitamins for the prevention of chronic disease [41,42]. Furthermore, two large studies examining the impact of multivitamins on breast cancer risk, the Women’s Health Study (WHS) and Women’s Health Initiative (WHI), found that multivitamin use did not decrease risk of developing breast cancer or influence overall mortality [43,44].

Despite the negative results described above, other studies have observed positive associations between vitamin supplements and improved prognosis in breast cancer survivors. Vitamin A is known

to play a role in inflammation, cell growth, and tissue differentiation, and thus may influence recurrence and metastasis [45]. One study showed that premenopausal survivors diagnosed with node-positive breast cancer and low levels of retinol binding protein, the carrier for vitamin A in plasma, had early disease recurrence [46]. In a separate study of postmenopausal women, low vitamin A levels were associated with poor outcomes, such as distant metastasis and increased mortality [47].

Thiamine (B1), riboflavin (B2), niacin (B3), and folic acid (B9) belong to the water soluble B-complex group of vitamins. Certain B-complex vitamins promote cell growth and division, and thus have been the subject of research examining possible relationships between B vitamins and breast cancer. In a study of 516 postmenopausal women with breast cancer from Orange County, California, women with the highest folate intake had lower mortality risk. Nutrients derived from diet, rather than supplements, were believed to provide the survival advantage [48]. In the Long Island Breast Cancer Study Project (LIBCSP) (n=1,508 women), vitamins B1 and B3 were associated with a 46% and a 39% lower risk, respectively, of all-cause mortality. Vitamin B1 intake was associated with decreased breast cancer mortality [49].

Vitamin C (ascorbic acid) functions as an antioxidant and is important to healthy immune function. In the Orange County study highlighted above, vitamin C was significantly associated with decreased all-cause mortality, but not with breast cancer mortality [48]. In the Shanghai Breast Cancer Survivor Study (SBCSS) of 4,877 women with breast cancer, vitamin C intake was associated with a 44% decrease in all-cause mortality and a 38% decrease in breast cancer recurrence. Similar but less convincing results were observed for vitamin E [50]. An evaluation of the effects of vitamin E and its derivatives on breast cancer cell lines suggests that antioxidants such as vitamin E may prevent disease recurrence by inducing apoptosis in breast cancer cells [51].

Vitamin D is produced naturally by the body during exposure to sunlight, but may be consumed through fortified foods or vitamin supplements. In addition to bone health, vitamin D affects cell proliferation, cell differentiation, and apoptosis. Regional differences in sunlight exposure in the US have been suggested to influence breast cancer mortality — women with greater sunlight exposure had a lower risk of fatal breast cancer, and length of daylight at the time of diagnosis has been linked to breast cancer prognosis [52,53]. Although one group found significantly higher levels of circulating 25-hydroxyvitamin D in women with early stage disease compared to patients with locally advanced or metastatic breast cancer [54], other studies suggest that vitamin D consumption following treatment may have no survival advantage [55].

Nearly 75% of breast cancer patients report taking multivitamins to bolster their immune system and make treatment more effective. Although the results outlined above suggest that vitamin supplements may have beneficial effects on prognosis in breast cancer survivors, the link between vitamin supplements and breast cancer is complex and not well studied. Current recommendations from the ACS suggest that survivors should consider using multivitamins with 100% of the Daily Recommended Values to ensure proper nutrition.

### ***Green tea***

Consumption of green tea has long been customary in Asia, but the purported health-promoting benefits of green tea have been recognized only recently in the US. Green tea contains a number of active ingredients including polyphenols, which may confer anti-mutagenic, anti-diabetic, anti-bacterial, anti-inflammatory, or hypo-cholesterolemic advantages [56]. In studies evaluating the effects of green tea consumption in 427 Japanese breast cancer survivors with stage I or stage II disease, recurrence was 16.7% in women who consumed  $\geq 5$  cups of green tea per day, compared to 24.3% in

those consuming <5 cups per day. This protective effect was not observed in patients with stage III breast cancer [57]. Likewise, in a second study of 1,160 Japanese women, consuming 3 or more cups of green tea each day was protective against recurrence in women with stage I disease, but showed no effect in more advanced cases [58]. A meta-analysis of these datasets showed that increased green tea consumption (more than 3 cups per day) was inversely associated with breast cancer recurrence [59]. Little information exists on the relationship between green tea consumption and survival in breast cancer patients; however, as no protective effects from green tea have been observed in advanced disease, the effects of green tea on overall breast cancer survival may be minimal.

### *Soy*

A variety of health benefits have been attributed to consumption of soy-based foods, primarily due to soybean isoflavones, such as genistein and diadzein. Major sources of isoflavones in the American diet are derived from soy-containing products such as oil, flour, soy-milk, and meat substitutes. The chemical structure of isoflavones may enable these compounds to interact with human estrogen receptors, causing effects in the body similar to those induced by estrogen. Experimental studies in cell lines and animal models suggest that genistein has the potential to stimulate cell proliferation, negatively influence prognosis, and impair the effectiveness of tamoxifen therapy in estrogen-receptor-positive (ER+) breast carcinomas [60,61].

Given the results of experimental studies outlined above, which suggest that soy constituents can be estrogenic and potentially risk enhancing, an important question remains: Do the estrogenic effects of soy influence prognosis and survival in humans? In population-based studies in the US, women from the LIBCSP diagnosed with a first primary invasive breast carcinoma (n=1,210) showed reduced hazard ratios for all-cause mortality in the highest quintile of dietary intake compared with the

lowest quintile for flavones, isoflavones, and anthocyanidins. A decrease in breast cancer mortality was detected only in postmenopausal patients [62]. Other studies of diverse populations also observed significant inverse associations between soy consumption and breast cancer recurrence in postmenopausal women [63,64].

The majority of studies examining the relationship between soy consumption and breast cancer prognosis have been performed in Asian populations where long-term soy intake is much higher than in Western countries. Available data suggest that dietary intake of soy-based foods may confer a protective advantage for postmenopausal women with breast cancer. Additional health benefits have been attributed to consumption of soy-based foods, such as lower cholesterol and improved bone health; however, intake of isoflavones at high concentrations typically found in dietary supplements may be detrimental. Until the issue is fully resolved, physicians should allow patients receiving hormonal therapy, or who have ER+ breast cancer, to consume moderate amounts of soy as part of a balanced diet, and may wish to recommend avoiding soy supplements.

As outlined above, most studies examining diet and survival in breast cancer patients were observational, population-based studies, rather than randomized trials. Although a prudent diet was not associated with breast cancer survival, a healthy diet has been shown to have beneficial effects on overall survival in conditions such as diabetes and heart disease, which are frequently seen in breast cancer patients. In addition, a sensible diet can result in significant weight loss, which has been associated with improved prognosis. Thus, patients should be encouraged to adopt a healthy diet that may include modest alcohol consumption, multivitamins, soy, and green tea.

### **Physical activity**

Physical activity is as important as diet for achieving optimal weight and maintaining a healthy lifestyle. In 2008, the US Department of Health and Human Services issued guidelines for physical activity, which included at least 150 minutes of moderate, or 75 minutes of intense, aerobic exercise per week, as well as strength-training at least twice a week to achieve significant health benefits [203]. Strenuous and recreational physical activities have been shown to confer numerous health benefits, including lower risk of developing breast cancer [65-67].

Treatment regimens consisting of surgery with or without chemotherapy and/or radiation have been associated with a number of significant side effects including fatigue, nausea, decreased cardiovascular health, and emotional distress. A number of randomized trials have determined that patients undergoing treatment who engaged in regular exercise had improved health and QOL compared to those who did not exercise [68-72]. Despite these positive findings, less than 50% of patients undergoing treatment, especially chemotherapy, engage in regular physical activity. As demonstrated in the HEAL study, physical activity decreased by approximately two hours per week after diagnosis compared to activity levels one year before diagnosis [73]. Similarly, Belgian women (n=267) with invasive breast cancer showed a significant decrease in physical activity during the first month post-diagnosis, and did not resume normal activity levels during the first year [74].

Risk of developing detrimental side effects, including cardiovascular disease, type 2 diabetes, fatigue, lymphedema, psychological distress, and poor QOL often persists in breast cancer survivors after treatment has been completed. Two small studies initially suggested that exercise may have beneficial effects in post-treatment survivors. In a cross-over study of 24 patients evaluating the effects of physical activity on psychological health, mild-to-moderate intensity exercise for 30 minutes, four days per week for 10 weeks, with or without positive reinforcement, was associated with lower depression and anxiety [75]. In a similar study, postmenopausal breast cancer survivors (n=53) who

performed moderate aerobic exercise three times per week for 15 weeks showed significant improvement in cardiopulmonary function and QOL [76].

Large cohort studies have clearly demonstrated the importance of physical activity in breast cancer patients (Table 3). Data from the NHS from 2,987 women with stage I-III breast cancer showed that any amount of exercise above 3-MET hours per week (the equivalent of walking at a moderate pace of 2-3 miles per hour for one hour) had a beneficial effect on survival. In patients with the highest activity, breast cancer recurrence and mortality were 26%-40% lower than in women with the lowest activity [77]. In the LACE trial, higher self-reported physical activity, including occupational, household, transportation-related, and recreational activity, was significantly associated with lower all-cause mortality in 1,970 breast cancer survivors [Sternfeld, 2009 1318 /id]. The Collaborative Women's Longevity Study (CWLS), which examined 4,482 women diagnosed with breast cancer between 1988 and 2001, suggested that women who engaged in >2.8 MET hours per week had a 35%-49% lower risk of breast cancer-related mortality compared to those who were physically inactive, even after adjusting for confounding factors such as disease stage, treatment modality, and age at diagnosis. Moderate activity appeared to provide the greatest benefit, with vigorous activity providing no additional survival advantage [79]. Likewise, women in the HEAL study who were physically active in the year immediately prior to diagnosis or two years post-diagnosis showed significantly lower breast cancer mortality than women who were inactive. Women who decreased their activity after diagnosis had a four-fold increase in mortality, while those who increased their activity had a 45% lower risk of death [80].

Most research to date strongly supports the link between physical activity and improved breast cancer prognosis and overall survival, but several issues, such as (1) the optimal form of physical activity, (2) the frequency and duration of activity, and (3) an individualized versus a standardized

physical activity regimen remain unresolved. Most studies indicate that moderate-intensity aerobic exercise provides the most benefit in terms of prognosis. The ACS currently recommends at least 30 minutes of physical activity five days a week [81]. In addition, a personalized activity plan may be more effective than a standardized regimen because an individual plan can be customized for different time periods from pre-diagnosis through cancer treatment based on individual needs and abilities. Perhaps the most important factor affecting breast cancer survivors is long-term compliance with any exercise plan. In order to derive maximum benefit, activity must become integrated into each patient's daily routine to ultimately be successful.

### ***Strength training and lymphedema***

Lymphedema results from the accumulation of excess fluid in the lymphatic system after breast cancer surgery and/or radiation therapy and may cause pain, swelling, numbness, stiffness, and fatigue in the affected arm [82]. Approximately 30% of breast cancer survivors suffer from lymphedema, which can adversely affect physical and psychosocial well being [83]. Although often incurable, treatments to lessen the physical effects of lymphedema such as manual drainage and limb compression have been developed.

Risk factors for lymphedema include the number of axillary lymph nodes surgically removed or damaged by radiation, obesity at diagnosis or weight gain after treatment, or activities that irritate or inflame the lymphatic system. To decrease risk of lymphedema, a number of breast cancer advocacy groups such as the ACS, Komen for the Cure, and the National Lymphedema Network have recommended limited use of the affected arm. Although these recommendations were designed as risk reduction strategies, limited use may make daily activities more challenging and actually increase risk for injury [84].

Two recent randomized clinical trials challenged the widespread recommendation that lymphedema patients should limit use of the affected/at-risk arm. The Physical Activity and Lymphedema (PAL) Trial randomized 141 breast cancer survivors with lymphedema and 154 survivors at-risk for lymphedema to either a non-exercising control group or to an intervention group who performed progressive weight training twice a week for one year. The intervention group showed greater improvements in self-reported symptoms and greater overall strength [85]. Importantly, progressive weight lifting did not exacerbate lymphedema. In at-risk patients who had five or more lymph nodes surgically removed, weight training had a noticeable effect on outcome — 22% of controls developed symptoms of lymphedema compared to only 7% of patients who participated in weight training [86]. These observations have revolutionized current concepts of lymphedema treatment and suggest that progressive strength training is in fact beneficial to patients.

### **Stress management**

Psychological distress associated with a diagnosis of breast cancer can evoke an array of emotional experiences, ranging from sadness and despair to anxiety and depression [87]. Although emotional symptoms may dissipate, an appreciable number of survivors continue to experience anxiety and depression for years after diagnosis [88]. Psychological stress may influence cancer risk and survival by altering endocrine responses, circadian rhythms, and immune functions (Figure 1).

Breast cancer survivors who have suffered stressful life events, such as divorce or death of a husband or child, had a significantly higher risk of relapse compared to survivors without such stressors [89]. Similarly, women with strong social ties to relatives, friends, and neighbors had significantly lower risk of breast cancer mortality compared to women lacking a social network [90]. In 2,835 women diagnosed with breast cancer, those who felt socially isolated (no close relationships to friends,

relatives, or adult children) before diagnosis had a 66% increased risk of all-cause mortality and a two-fold increase in breast cancer mortality [91]. Therefore, it is important for breast cancer patients to practice effective stress management techniques, develop a social network, and participate in group support sessions to improve prognosis and QOL.

### *Group support*

Breast cancer patients often experience social isolation during months of treatment and recovery, which precipitate feelings of loneliness, low self esteem, body image issues, and negatively impact their QOL [92]. Two types of group support have been evaluated for their ability to improve QOL and survival: (1) cognitive-behavioral therapy, which uses conversation and active participation to resolve dysfunctional emotions and behaviors in individual or group settings; and (2) psycho-educational therapy, which provides cancer education to patients, teaches coping mechanisms, and offers support and resources to improve psychosocial functioning [93]. Some studies suggest that these interventions have a positive effect on cancer patients for certain symptoms such as distress, depression, anxiety, fatigue, and QOL [94-97], while others conclude that psychosocial interventions are ineffective or limited in producing meaningful emotional benefit [98, {Fors, 2011 1415 /id}]. Limited evidence suggests that group support led by a trained psychologist focusing on strategies to reduce stress, improve mood, and maintain adherence to treatment can successfully lower risk of breast cancer recurrence and mortality [100].

Despite the inconclusive nature of current research, psychosocial interventions are rapidly becoming an important component of cancer care. For health care professionals to provide evidence-based guidance to their cancer patients, high quality well-controlled trials with pre-defined outcome measures are needed to determine the precise benefits of psychosocial interventions.

## *Yoga*

Yoga is a mind-body therapy that originated in India more than 5,000 years ago to reduce stress and promote overall physical and mental well-being. The practice of yoga relies on physical postures, focused breathing, and meditation to stretch muscles, control breathing, and minimize stress through visualization techniques and guided imagery. Yoga is believed to offer numerous psychological and physical health benefits to cancer patients [101], including disease prevention, improved cardiovascular health, weight loss, and stress reduction. Cancer survivors, in particular, tend to use complementary and alternative medicine (CAM) approaches [102] to alleviate common symptoms including fatigue, pain, and insomnia [103], which may persist long after treatment has been completed [104-106].

Initial scientific research has shown that yoga confers a variety of benefits to breast cancer patients. A 12-week yoga intervention significantly improved fatigue scores, as well as physical function, mood, and QOL in a small sample of breast cancer survivors [107]. Likewise, a small pilot study of menopausal women who completed a 10-week yoga program reported significant post-treatment improvement in sleep disturbance, sleep efficiency, and quality of sleep attained [108]. Yoga has been shown to improve psychosocial functioning in breast cancer survivors more than four years post-diagnosis [109], as well as those undergoing active clinical treatment [110].

Chronic pain and other sequelae such as paraesthesia (burning or tingling sensation of the skin), allodynia (pain caused by light touch), and lymphedema are significantly more frequent in breast cancer survivors compared to women in the general population [111-113]. Yoga has been shown to alleviate various forms of chronic pain experienced by many breast cancer survivors, including lower back pain [114], pain associated with fibromyalgia [115], and rheumatoid arthritis [116].

Chronic stress is another common ailment that can negatively impact QOL and survival in women with breast cancer. As an increasing number of women turn to CAM to improve their health after breast cancer diagnosis, it is critical to develop effective methods for managing stress. To date, group support has shown mixed overall results in patients suffering from anxiety and depression, but seems to improve survival in patients with the highest levels of anxiety and depression. Yoga has shown promising results in relieving psychological distress and decreasing fatigue, and may be effective in treating chronic pain associated with side effects of treatment. Most studies to date have been too small to adequately define treatment effects; thus, large-scale studies are urgently needed to quantify potential benefits of group support, yoga, and other measures of stress reduction in breast cancer patients.

### **Lifestyle interventions in patients with metastatic breast cancer**

The majority of intervention studies considering the effects of lifestyle factors on breast cancer prognosis often focus on QOL issues, and may suffer from design limitations that restrict their ability to adequately address survival or recurrence in diverse populations of breast cancer patients [117]. Epidemiologic studies often focus on women with early stage breast cancer and do not usually enroll large numbers of patients with late-stage, metastatic disease. Compared to women with localized disease who have a 98% five-year survival rate, women with stage IV breast cancer have five-year survival rates of only 23%. Patients living with metastatic disease have different health concerns than women free from metastasis. For example, women with stage IV breast cancer may need bone strengthening therapy, psychological counseling to deal with social isolation, nutritional counseling to help with appetite and weight loss, treatment for fatigue, and pain management. Intervention studies must be tailored to the specific needs of patients with metastatic disease, recognizing that diet, exercise,

and stress reduction techniques used in other breast cancer survivors may be less desirable or inappropriate.

Despite limited research on patients with metastatic disease, lifestyle interventions have shown benefit in women with stage IV breast cancer. Psychosocial interventions indicate that women randomized to weekly supportive-expressive group therapy demonstrate improved mood and perception of pain [118], as well as a significantly greater decline in mood disturbance and traumatic stress symptoms compared to controls [119]. Likewise, women randomized to receive cognitive therapy for depression had significantly lower levels of depressive symptoms, anxiety, fatigue, and insomnia following treatment compared to those assigned as wait-list controls [120]. Additional studies have shown that other lifestyle interventions, such as yoga-based palliative care, and integrative treatment with nutrition, fitness, and mind-spirit instruction [121], improve QOL and provide survival benefit.

## **Conclusions**

A multitude of studies investigating the impact of lifestyle modification on breast cancer survivors has produced highly variable and contradictory results. Large-scale dietary studies indicate that specific dietary components may not affect breast cancer prognosis, but achieving and/or maintaining a healthy body weight through diet provides health benefits to breast cancer survivors. Studies of physical activity in breast cancer patients have definitively linked physical activity to improved prognosis and survival, and suggest that women with, or at risk for, lymphedema should engage in progressive strength training to increase mobility and decrease risk of developing symptoms. Studies evaluating stress reduction and the effects of group therapy on QOL have been inconclusive, but many studies evaluated small groups of women and lacked adequate statistical power.

Nevertheless, stress reduction techniques may benefit women with the highest levels of anxiety and depression, and yoga in particular, appears to provide numerous health benefits that improve prognosis. These findings illustrate the need for continued, high-quality research into the effects of lifestyle interventions in breast cancer patients in order to provide breast cancer survivors with state-of-the-art knowledge and optimum methods for improving long-term prognosis and QOL.

### **Expert Commentary**

As the number of breast cancer survivors increases, interest in lifestyle intervention and CAM approaches continues to grow. Although the main focus for many breast cancer survivors is prevention of disease recurrence, survivors are at increased risk for developing chronic conditions such as osteoporosis, diabetes, and cardiovascular disease, as well as non-cancer-related mortality [122]. Behavioral modification may provide breast cancer patients with numerous health benefits, both during and after active treatment. In addition to potential health benefits, lifestyle modification may provide patients with feelings of control and self-determination because they become active participants in managing their own health.

Between September 2006 and February 2007, the President's Cancer Panel examined current evidence regarding the effects of lifestyle behaviors on cancer risk, as well as ways to reduce the national cancer burden by promoting healthy lifestyles [204]. The panel recommended that (1) health care providers coordinate and integrate education and prevention messages related to diet, nutrition, physical activity, and tobacco use with other chronic diseases to promote common risk reduction strategies, and (2) individuals assume personal responsibility for learning about cancer risk and make healthy lifestyle choices. Research has shown that breast cancer patients who receive healthy lifestyle recommendations, such as a recommendation to exercise, from their oncologist actually participate in

these behaviors significantly more than patients following standard treatment [123]; however, only about one-fourth of patients report actually receiving dietary or physical activity recommendations from their primary care physicians [124]. Health care providers must become more involved in recommending (and monitoring implementation of) healthy lifestyle behaviors for their cancer patients. Similarly, breast cancer survivors must be proactive in learning about options for healthier lifestyles and integrating appropriate behaviors into their daily routine.

Despite the large number of studies investigating the effects of lifestyle modification on breast cancer survivorship, few definitive conclusions have emerged. Many lifestyle behaviors may individually affect prognosis, but it is unlikely that a single comprehensive “lifestyle modification package” is appropriate for all survivors. Patients should consider general “common-sense” recommendations: (1) maintain a healthy weight by consuming a diet rich in plant-based foods and limited alcohol; (2) be physically active; (3) avoid smoking or second-hand smoke; and (4) minimize stress. Survivors should strive to incorporate as many healthy behaviors as possible, given their individual physical, emotional, and financial constraints.

### **Five-Year View**

Women with a history of breast cancer constitute the largest group of female cancer survivors, accounting for ~22% of all cancer survivors [205]. As the number of breast cancer survivors continues to grow, it is crucial to identify modifiable factors associated with breast cancer recurrence and survival and to fully understand optimum methods for risk factor reduction. This information will likely come from a variety of sources, but should include:

- Well-designed, high-quality randomized trials and observational studies with adequate statistical power — an example is the Pathways Study, a prospective study of breast cancer

survivorship, which is expecting to enroll 3,000 patients within two months of diagnosis to reduce the effect of survival-bias [125]. Recruiting women with late-stage disease, who frequently have different physical and emotional needs than early-stage patients, will provide much needed information for patients with metastasis.

- Comprehensive clinical programs with research components, such as the Clinical Breast Care Project, a breast cancer research program based at Walter Reed National Military Medical Center, which includes a state-of-the-art comprehensive breast center and research capabilities to examine environmental factors associated with prognosis [126].
- Survivors from diverse ethnic and socioeconomic backgrounds and culturally appropriate intervention programs [127], which would examine important ethnic differences in clinical care and cultural factors influencing behavior modification [128].

Specific issues of highest priority that need to be resolved over the next five years are:

- Who would derive the most benefit from, and thus should participate in, behavioral modification programs? Women of advanced age with significant co-morbidities or metastatic disease may not derive meaningful benefit from lifestyle changes, may be physically unable to participate in certain activity routines, and may be reluctant to adopt new dietary habits.
- Can an optimum combination of lifestyle choices can be determined on an individual patient basis that provides survival and QOL benefits that are superior to a standardized approach?
- What is the optimum way to overcome barriers to implementing and maintaining healthy lifestyle behaviors? Research has shown that (1) age [129] and ethnic [130] differences exist in the likelihood that survivors will adopt healthy lifestyle behaviors, and (2) patients may

not fully adhere over the longer-term to self-directed interventions and may miss potential health benefits [131]. Healthy lifestyle programs need to be developed for a diverse population of breast cancer survivors that have wider adaptability than those presently available.

- Evolution of primary care: Medical providers should not discount CAM approaches, given that a majority of patients believe that these alternative therapies improve their QOL. Physicians should be powerful catalysts for promoting behavior change in their patients, because they are optimally positioned to deliver guidance regarding health promotion.
- Do interactions between molecular factors and behaviors influence breast cancer recurrence and survival? Identifying genetic variation that may enhance or inhibit responses to lifestyle intervention would add an additional level of personalization to patient care.

Finally, integrating healthy behaviors into the daily lives of breast cancer patients faces the ongoing challenge of accessibility to health resources and cost. At present, the cost of treatment for metastatic breast cancer is \$128,556 (95% CI=\$118,409-\$137,644) per patient, and out-of-pocket expenses can reach \$1,500 per month [10]. Additional costs for incorporating healthy lifestyle components that use providers who charge for services [132], such as gym memberships, yoga classes, exercise physiologists, and dietician consultations may be prohibitive to many breast cancer survivors.

The Patient Protection and Affordable Care Act and the Health Care and Education Reconciliation Act were passed in March 2010 [206]. This long overdue overhaul of the American healthcare system is based on a “triple aim” philosophy (improved health in the general population, better patient experiences, and lower per-capita cost) that eventually may translate into improved care

for breast cancer patients. Under this new legislation, prevention will assume a greater role in health care [133,134], and evidence-based intervention programs may become covered services.

### **Key Issues**

- Due mainly to advances in early detection and treatment, there are an estimated 2.5 million breast cancer survivors in the US alone, and this number is projected to increase to 3.4 million within the next four years. Behavioral modification has great potential to improve prognosis and QOL in breast cancer survivors.
- Achieving or maintaining a healthy weight after diagnosis is associated with improved prognosis. A healthy diet may not affect breast cancer prognosis, but has been associated with increased overall survival.
- Limited research suggests that certain types of alcohol, vitamin supplements, soy, and green tea may have some beneficial properties for breast cancer survivors; however, large-scale studies are needed to provide conclusive results and specific recommendations for consumption.
- Moderate physical activity can ameliorate side effects during treatment and improve breast cancer prognosis and overall survival after treatment.
- Contrary to the long-held belief that patients affected with, or at-risk for, lymphedema should limit use of the affected/at-risk arm, progressive supervised strength training can decrease symptoms and improve outcomes.
- Stress reduction techniques such as group support and yoga have shown promising results in improving outcomes and QOL, but further research is essential for quantifying potential benefits in breast cancer patients.

- As more effective healthy lifestyle interventions evolve, important issues such as who will derive the most benefit, when lifestyle interventions should be initiated, how and where programs should be facilitated, and who will pay for healthy lifestyle programs must be resolved in order for lifestyle modification to have a meaningful impact on the lives of breast cancer survivors.

## References

- 1 American Cancer Society. *Cancer facts & figures 2011*. American Cancer Society, Atlanta, GA, USA (2011).
- 2 Jemal A, Ward E, Thun M. Declining death rates reflect progress against cancer. *Plos ONE* 5, e9584 (2010).
- 3 Berry DA, Cronin KA, Plevritis SK *et al*. Effect of screening and adjuvant therapy on mortality from breast cancer. *N. Engl. J. Med.* 353, 1784-1792 (2005).
- 4 Morgan MA. Cancer survivorship: history, quality-of-life issues, and the evolving multidisciplinary approach to implementation of cancer survivorship care plans. *Oncol. Nurs. Forum* 36, 429-436 (2009).
- 5 Stuart K, Brennan M, French J, Houssami N, Boyages J. Life after breast cancer. *Aust. Fam. Physician* 35, 219-224 (2006).
- 6 Rozenberg S, Antoine C, Carly B, Pastijn A, Liebens F. Improving quality of life after breast cancer: prevention of other diseases. *Menopause Int.* 13, 71-74 (2007).
- 7 Montazeri A. Health-related quality of life in breast cancer patients: a bibliographic review of the literature from 1974-2007. *J. Exp. Clin. Cancer Res.* 27, 32 (2008).
- 8 De Angelis R, Tavilla A, Verdecchia A *et al*. Breast cancer survivors in the United States: geographic variability and time trends, 2005-2015. *Cancer* 115, 1954-1966 (2009).
- 9 Garrison LP, Jr., Veenstra DL. The economic value of innovative treatments over the product life cycle: the case of targeted trastuzumab therapy for breast cancer. *Value Health* 12, 1118-1123 (2009).
- 10 Pisu M, Azuero A, McNees P, Burkhardt J, Benz R, Meneses K. The out of pocket cost of breast cancer survivors: a review. *J. Cancer Surviv.* 4, 202-209 (2010).

- 25 Kwan ML, Weltzien E, Kushi LH, Castillo A, Slattery ML, Caan BJ. Dietary patterns and breast cancer recurrence and survival among women with early-stage breast cancer. *J. Clin. Oncol.* 27, 919-926 (2009).
- 26 Irwin ML, McTiernan A, Baumgartner RN *et al.* Changes in body fat and weight after a breast cancer diagnosis: influence of demographic, prognostic, and lifestyle factors. *J. Clin. Oncol.* 23, 774-782 (2005).
- 27 George SM, Irwin ML, Smith AW *et al.* Postdiagnosis diet quality, the combination of diet quality and recreational physical activity, and prognosis after early-stage breast cancer. *Cancer Causes Control* 22, 589-598 (2011).
- 28 Wayne SJ, Baumgartner K, Baumgartner RN, Bernstein L, Bowen DJ, Ballard-Barbash R. Diet quality is directly associated with quality of life in breast cancer survivors. *Breast Cancer Res. Treat.* 96, 227-232 (2006).
- 29 George SM, Neuhaus ML, Mayne ST *et al.* Postdiagnosis diet quality is inversely related to a biomarker of inflammation among breast cancer survivors. *Cancer Epidemiol. Biomarkers Prev.* 19, 2220-2228 (2010).
- 30 Chlebowski RT, Blackburn GL, Thomson CA *et al.* Dietary fat reduction and breast cancer outcome: interim efficacy results from the Women's Intervention Nutrition Study. *J. Natl. Cancer Inst.* 98, 1767-1776 (2006).
- \*\* Observed significant improvement in prognosis of breast cancer patients on a low-fat diet.**
- 31 Pierce JP, Natarajan L, Caan BJ *et al.* Influence of a diet very high in vegetables, fruit, and fiber and low in fat on prognosis following treatment for breast cancer: the Women's Healthy Eating and Living (WHEL) randomized trial. *JAMA* 298, 289-298 (2007).
- \*\* Detected no association between a low-fat, high fruit and vegetable diet and outcomes in breast cancer patients.**
- 32 Hamajima N, Hirose K, Tajima K *et al.* Alcohol, tobacco and breast cancer — collaborative reanalysis of individual data from 53 epidemiological studies, including 58,515 women with breast cancer and 95,067 women without the disease. *Br. J. Cancer* 87, 1234-1245 (2002).
- 33 Reding KW, Daling JR, Doody DR, O'Brien CA, Porter PL, Malone KE. Effect of prediagnostic alcohol consumption on survival after breast cancer in young women. *Cancer Epidemiol. Biomarkers Prev.* 17, 1988-1996 (2008).
- 34 Barnett GC, Shah M, Redman K, Easton DF, Ponder BA, Pharoah PD. Risk factors for the incidence of breast cancer: do they affect survival from the disease? *J. Clin. Oncol.* 26, 3310-3316 (2008).

- 35 Kwan ML, Kushi LH, Weltzien E *et al.* Alcohol consumption and breast cancer recurrence and survival among women with early-stage breast cancer: the life after cancer epidemiology study. *J. Clin. Oncol.* 28, 4410-4416 (2010).
- 36 Franceschi S, Dal Maso L, Zucchetto A, Talamini R, Prospective Analysis of Case-control studies on Environmental factors and health (PACE) study group. Alcohol consumption and survival after breast cancer. *Cancer Epidemiol. Biomarkers Prev.* 18, 1011-1012 (2009).
- 37 Flatt SW, Thomson CA, Gold EB *et al.* Low to moderate alcohol intake is not associated with increased mortality after breast cancer. *Cancer Epidemiol. Biomarkers Prev.* 19, 681-688 (2010).
- 38 Espin JC, Garcia-Conesa MT, Tomas-Barberan FA. Nutraceuticals: facts and fiction. *Phytochemistry* 68, 2986-3008 (2007).
- 39 Velicer CM, Ulrich CM. Vitamin and mineral supplement use among US adults after cancer diagnosis. A systematic review. *J. Clin. Oncol.* 26, 665-673 (2008).
- 40 Nesheim MC. What is the research base for the use of dietary supplements? *Public Health Nutr.* 2, 35-38 (1999).
- 41 Morris CD, Carson S. Routine vitamin supplementation to prevent cardiovascular disease: a summary of the evidence for the U.S. Preventive Services Task Force. *Ann. Intern. Med.* 139, 56-70 (2003).
- 42 NIH State-of-the-Science Panel. National Institutes of Health State-of-the-Science Conference Statement: multivitamin/mineral supplements and chronic disease prevention. *Ann. Intern. Med.* 145, 364-371 (2006).
- 43 Ishitani K, Lin J, Manson JE, Buring JE, Zhang SM. A prospective study of multivitamin supplement use and risk of breast cancer. *Am. J. Epidemiol.* 167, 1197-1206 (2008).
- 44 Neuhauser ML, Wassertheil-Smoller S, Thomson C *et al.* Multivitamin use and risk of cancer and cardiovascular disease in the Women's Health Initiative cohorts. *Arch. Intern. Med.* 169, 294-304 (2009).
- 45 Sporn MB, Roberts AB, Goodman DS. *The Retinoids: Biology, Chemistry, and Medicine.* 2nd ed. Raven Press, New York, NY, USA (1994).
- 46 Mehta RR, Hart G, Beattie CW, Das Gupta TK. Significance of plasma retinol binding protein levels in recurrence of breast tumors in women. *Oncology* 44, 350-355 (1987).
- 47 Formelli F, Meneghini E, Cavadini E *et al.* Plasma retinol and prognosis of postmenopausal breast cancer patients. *Cancer Epidemiol. Biomarkers Prev.* 18, 42-48 (2009).

- 48 McEligot AJ, Largent J, Ziogas A, Peel D, Anton-Culver H. Dietary fat, fiber, vegetable, and micronutrients are associated with overall survival in postmenopausal women diagnosed with breast cancer. *Nutr. Cancer* 55, 132-140 (2006).
- 49 Xu X, Gammon MD, Wetmur JG *et al.* B-vitamin intake, one-carbon metabolism and survival among a population-based study of women with breast cancer. *Cancer Epidemiol. Biomarkers Prev.* 17, 2109-2116 (2008).
- 50 Nechuta S, Lu W, Chen Z *et al.* Vitamin supplement use during breast cancer treatment and survival: a prospective cohort study. *Cancer Epidemiol. Biomarkers Prev.* 20, 262-271 (2011).
- 51 Kline K, Yu W, Sanders BG. Vitamin E and breast cancer. *J. Nutr.* 134, 3458S-3462S (2004).
- 52 Garland FC, Garland CF, Gorham ED, Young JF. Geographic variation in breast cancer mortality in the United States: a hypothesis involving exposure to solar radiation. *Prev. Med.* 19, 614-622 (1990).
- 53 Robsahm TE, Tretli S, Dahlback A, Moan J. Vitamin D3 from sunlight may improve the prognosis of breast-, colon- and prostate cancer (Norway). *Cancer Causes Control* 15, 149-158 (2004).
- 54 Palmieri C, MacGregor T, Girgis S, Vigushin D. Serum 25-hydroxyvitamin D levels in early and advanced breast cancer. *J. Clin. Pathol.* 59, 1334-1336 (2006).
- 55 Jacobs ET, Thomson CA, Flatt SW *et al.* Vitamin D and breast cancer recurrence in the Women's Healthy Eating and Living (WHEL) Study. *Am. J. Clin. Nutr.* 93, 108-117 (2011).
- 56 Cabrera C, Artacho R, Giménez R. Beneficial effects of green tea — a review. *J. Am. Coll. Nutr.* 25, 79-99 (2006).
- 57 Nakachi K, Suemasu K, Suga K, Takeo T, Imai K, Higashi Y. Influence of drinking green tea on breast cancer malignancy among Japanese patients. *Jpn. J. Cancer Res.* 89, 254-261 (1998).
- 58 Inoue M, Tajima K, Mizutani M *et al.* Regular consumption of green tea and the risk of breast cancer recurrence: follow-up study from the Hospital-based Epidemiologic Research Program at Aichi Cancer Center (HERPACC), Japan. *Cancer Lett.* 167, 175-182 (2001).
- 59 Ogunleye AA, Xue F, Michels KB. Green tea consumption and breast cancer risk or recurrence: a meta-analysis. *Breast Cancer Res. Treat.* 119, 477-484 (2010).
- 60 Allred CD, Allred KF, Ju YH, Virant SM, Helferich WG. Soy diets containing varying amounts of genistein stimulate growth of estrogen-dependent (MCF-7) tumors in a dose-dependent manner. *Cancer Res.* 61, 5045-5050 (2001).
- 61 Ju YH, Doerge DR, Allred KF, Allred CD, Helferich WG. Dietary genistein negates the inhibitory effect of tamoxifen on growth of estrogen-dependent human breast cancer (MCF-7) cells implanted in athymic mice. *Cancer Res.* 62, 2474-2477 (2002).

- 62 Fink BN, Steck SE, Wolff MS *et al.* Dietary flavonoid intake and breast cancer survival among women on Long Island. *Cancer Epidemiol. Biomarkers Prev.* 16, 2285-2292 (2007).
- 63 Guha N, Kwan ML, Quesenberry CP, Weltzien EK, Castillo AL, Caan BJ. Soy isoflavones and risk of cancer recurrence in a cohort of breast cancer survivors: the Life After Cancer Epidemiology study. *Breast Cancer Res. Treat.* 118, 395-405 (2009).
- 64 Kang X, Zhang Q, Wang S, Huang X, Jin S. Effect of soy isoflavones on breast cancer recurrence and death for patients receiving adjuvant endocrine therapy. *CMAJ* 182, 1857-1862 (2010).
- 65 McTiernan A, Kooperberg C, White E *et al.* Recreational physical activity and the risk of breast cancer in postmenopausal women: the Women's Health Initiative Cohort Study. *JAMA* 290, 1331-1336 (2003).
- 66 Friedenreich CM, Cust AE. Physical activity and breast cancer risk: impact of timing, type and dose of activity and population subgroup effects. *Br. J. Sports Med.* 42, 636-647 (2008).
- 67 George SM, Irwin ML, Matthews CE *et al.* Beyond recreational physical activity: examining occupational and household activity, transportation activity, and sedentary behavior in relation to postmenopausal breast cancer risk. *Am. J. Public Health* 100, 2288-2295 (2010).
- 68 Mock V, Burke MB, Sheehan P *et al.* A nursing rehabilitation program for women with breast cancer receiving adjuvant chemotherapy. *Oncol. Nurs. Forum* 21, 899-907 (1994).
- 69 Dimeo FC, Stieglitz RD, Novelli-Fischer U, Fetscher S, Keul J. Effects of physical activity on the fatigue and psychologic status of cancer patients during chemotherapy. *Cancer* 85, 2273-2277 (1999).
- 70 Segal R, Evans W, Johnson D *et al.* Structured exercise improves physical functioning in women with stages I and II breast cancer: results of a randomized controlled trial. *J. Clin. Oncol.* 19, 657-665 (2001).
- 71 Campbell A, Mutrie N, White F, McGuire F, Kearney N. A pilot study of a supervised group exercise programme as a rehabilitation treatment for women with breast cancer receiving adjuvant treatment. *Eur. J. Oncol. Nurs.* 9, 56-63 (2005).
- 72 Hsieh CC, Sprod LK, Hydock DS, Carter SD, Hayward R, Schneider CM. Effects of a supervised exercise intervention on recovery from treatment regimens in breast cancer survivors. *Oncol. Nurs. Forum* 35, 909-915 (2008).
- 73 Irwin ML, Crumley D, McTiernan A *et al.* Physical activity levels before and after a diagnosis of breast carcinoma: the Health, Eating, Activity, and Lifestyle (HEAL) study. *Cancer* 97, 1746-1757 (2003).
- 74 Devoogdt N, Van Kampen M, Geraerts I *et al.* Physical activity levels after treatment for breast cancer: one-year follow-up. *Breast Cancer Res. Treat.* 123, 417-425 (2010).

- 75 Segar ML, Katch VL, Roth RS *et al.* The effect of aerobic exercise on self-esteem and depressive and anxiety symptoms among breast cancer survivors. *Oncol. Nurs. Forum* 25, 107-113 (1998).
- 76 Courneya KS, Mackey JR, Bell GJ, Jones LW, Field CJ, Fairey AS. Randomized controlled trial of exercise training in postmenopausal breast cancer survivors: cardiopulmonary and quality of life outcomes. *J. Clin. Oncol.* 21, 1660-1668 (2003).
- 77 Holmes MD, Chen WY, Feskanich D, Kroenke CH, Colditz GA. Physical activity and survival after breast cancer diagnosis. *JAMA* 293, 2479-2486 (2005).
- 78 Sternfeld B, Weltzien E, Quesenberry CP Jr *et al.* Physical activity and risk of recurrence and mortality in breast cancer survivors: findings from the LACE study. *Cancer Epidemiol. Biomarkers Prev.* 18, 87-95 (2009).
- 79 Holick CN, Newcomb PA, Trentham-Dietz A *et al.* Physical activity and survival after diagnosis of invasive breast cancer. *Cancer Epidemiol. Biomarkers Prev.* 17, 379-386 (2008).
- 80 Irwin ML, Smith AW, McTiernan A *et al.* Influence of pre- and post-diagnosis physical activity on mortality in breast cancer survivors: the Health, Eating, Activity, and Lifestyle study. *J. Clin. Oncol.* 26, 3958-3964 (2008).
- \* **Description of healthy benefits of exercise before and after breast cancer diagnosis.**
- 81 Byers T, Nestle M, McTiernan A *et al.* American Cancer Society guidelines on nutrition and physical activity for cancer prevention: Reducing the risk of cancer with healthy food choices and physical activity. *CA. Cancer J. Clin.* 52, 92-119 (2002).
- 82 Fu MR, Rosedale M. Breast cancer survivors' experiences of lymphedema-related symptoms. *J. Pain Symptom. Manage.* 38, 849-859 (2009).
- 83 Norman SA, Localio AR, Kallan MJ *et al.* Risk factors for lymphedema after breast cancer treatment. *Cancer Epidemiol. Biomarkers Prev.* 19, 2734-2746 (2010).
- 84 Schmitz KH. Balancing lymphedema risk: exercise versus deconditioning for breast cancer survivors. *Exerc. Sport Sci. Rev.* 38, 17-24 (2010).
- 85 Schmitz KH, Troxel AB, Chevillat A *et al.* Physical Activity and Lymphedema (the PAL trial): assessing the safety of progressive strength training in breast cancer survivors. *Contemp. Clin. Trials* 30, 233-245 (2009).
- \*\* **Sentinel study evaluating progressive, supervised weight training in women with lymphedema; led to paradigm shift in activity recommendations for lymphedema patients.**
- 86 Schmitz KH, Ahmed RL, Troxel AB *et al.* Weight lifting for women at risk for breast cancer-related lymphedema: a randomized trial. *JAMA* 304, 2699-2705 (2010).

**\*\* Sentinel study evaluating progressive, supervised weight training in women at risk for lymphedema; led to paradigm shift in treatment for patients who have had axillary lymph node removal, but have not developed lymphedema.**

87 Zabora J, BrintzenhofeSzoc K, Curbow B, Hooker C, Piantadosi S. The prevalence of psychological distress by cancer site. *Psychooncology* 10, 19-28 (2001).

88 Spiegel D. Psychosocial aspects of breast cancer treatment. *Semin. Oncol.* 24, S1-36-S1-47 (1997).

89 Ramirez AJ, Craig TK, Watson JP, Fentiman IS, North WR, Rubens RD. Stress and relapse of breast cancer. *BMJ* 298, 291-293 (1989).

90 Waxler-Morrison N, Hislop TG, Mears B, Kan L. Effects of social relationships on survival for women with breast cancer: a prospective study. *Soc. Sci. Med.* 33, 177-183 (1991).

91 Kroenke CH, Kubzansky LD, Schernhammer ES, Holmes MD, Kawachi I. Social networks, social support, and survival after breast cancer diagnosis. *J. Clin. Oncol.* 24, 1105-1111 (2006).

**\* Large study demonstrating detrimental effects of social isolation in breast cancer survivors.**

92 Rosedale M. Survivor loneliness of women following breast cancer. *Oncol. Nurs. Forum* 36, 175-183 (2009).

93 Jacobsen PB, Jim HS. Psychosocial interventions for anxiety and depression in adult cancer patients: achievements and challenges. *CA. Cancer J. Clin.* 58, 214-230 (2008).

94 Osborn RL, Démoncada AC, Feuerstein M. Psychosocial interventions for depression, anxiety, and quality of life in cancer survivors: meta-analyses. *Int. J. Psychiatry Med.* 36, 13-34 (2006).

95 Tatrow K, Montgomery GH. Cognitive behavioral therapy techniques for distress and pain in breast cancer patients: a meta-analysis. *J. Behav. Med.* 29, 17-27 (2006).

96 Zimmermann T, Heinrichs N, Baucom DH. "Does one size fit all?" moderators in psychosocial interventions for breast cancer patients: a meta-analysis. *Ann. Behav. Med.* 34, 225-239 (2007).

97 Duijts SF, Faber MM, Oldenburg HS, van BM, Aaronson NK. Effectiveness of behavioral techniques and physical exercise on psychosocial functioning and health-related quality of life in breast cancer patients and survivors — a meta-analysis. *Psychooncology* 20, 115-126 (2011).

98 Coyne JC, Stefanek M, Palmer SC. Psychotherapy and survival in cancer: the conflict between hope and evidence. *Psychol. Bull.* 133, 367-394 (2007).

99 Fors EA, Bertheussen GF, Thune I *et al.* Psychosocial interventions as part of breast cancer rehabilitation programs? Results from a systematic review. *Psychooncology* 20, 909-918 (2011).

- 100 Andersen BL, Yang HC, Farrar WB *et al.* Psychologic intervention improves survival for breast cancer patients: a randomized clinical trial. *Cancer* 113, 3450-3458 (2008).
- 101 Lin KY, Hu YT, Chang KJ, Lin HF, Tsauo JY. Effects of yoga on psychological health, quality of life, and physical health of patients with cancer: a meta-analysis. *Evid. Based Complement. Alternat. Med.* 2011, 659876 (2011).
- 102 Mao JJ, Palmer CS, Healy KE, Desai K, Amsterdam J. Complementary and alternative medicine use among cancer survivors: a population-based study. *J. Cancer Surviv.* 5, 8-17 (2011).
- 103 Lee ES, Lee MK, Kim SH *et al.* Health-related quality of life in survivors with breast cancer 1 year after diagnosis compared with the general population: a prospective cohort study. *Ann. Surg.* 253, 101-108 (2011).
- 104 Bower JE, Ganz PA, Desmond KA *et al.* Fatigue in long-term breast carcinoma survivors: a longitudinal investigation. *Cancer* 106, 751-758 (2006).
- 105 Minton O, Stone P. How common is fatigue in disease-free breast cancer survivors? A systematic review of the literature. *Breast Cancer Res. Treat.* 112, 5-13 (2008).
- 106 Otte JL, Carpenter JS, Russell KM, Bigatti S, Champion VL. Prevalence, severity, and correlates of sleep-wake disturbances in long-term breast cancer survivors. *J. Pain Symptom Manage.* 39, 535-547 (2010).
- 107 Bower JE, Garet D, Sternlieb B. Yoga for persistent fatigue in breast cancer survivors: results of a pilot study. *Evid. Based Complement. Alternat. Med.* 2011, 623168 (2011).
- 108 Booth-LaForce C, Thurston RC, Taylor MR. A pilot study of a Hatha yoga treatment for menopausal symptoms. *Maturitas* 57, 286-295 (2007).
- 109 Culos-Reed SN, Carlson LE, Daroux LM, Hatley-Aldous S. A pilot study of yoga for breast cancer survivors: physical and psychological benefits. *Psychooncology* 15, 891-897 (2006).
- 110 Vadiraja HS, Raghavendra RM, Nagarathna R *et al.* Effects of a yoga program on cortisol rhythm and mood states in early breast cancer patients undergoing adjuvant radiotherapy: a randomized controlled trial. *Integr. Cancer Ther.* 8, 37-46 (2009).
- 111 Gulluoglu BM, Cingi A, Cakir T, Gercek A, Barlas A, Eti Z. Factors related to post-treatment chronic pain in breast cancer survivors: the interference of pain with life functions. *Int. J. Fertil. Womens Med.* 51, 75-82 (2006).
- 112 Mao JJ, Armstrong K, Bowman MA, Xie SX, Kadakia R, Farrar JT. Symptom burden among cancer survivors: impact of age and comorbidity. *J. Am. Board Fam. Med.* 20, 434-443 (2007).
- 113 Peuckmann V, Ekholm O, Rasmussen NK *et al.* Chronic pain and other sequelae in long-term breast cancer survivors: nationwide survey in Denmark. *Eur. J. Pain* 13, 478-485 (2009).

- 114 Williams K, Abildso C, Steinberg L *et al.* Evaluation of the effectiveness and efficacy of Iyengar yoga therapy on chronic low back pain. *Spine (Phila Pa 1976)* 34, 2066-2076 (2009).
- 115 Carson JW, Carson KM, Porter LS, Keefe FJ, Shaw H, Miller JM. Yoga for women with metastatic breast cancer: results from a pilot study. *J. Pain Symptom Manage.* 33, 331-341 (2007).
- 116 Bosch PR, Traustadottir T, Howard P, Matt KS. Functional and physiological effects of yoga in women with rheumatoid arthritis: a pilot study. *Altern. Ther. Health Med.* 15, 24-31 (2009).
- 117 Kushi LH, Kwan ML, Lee MM, Ambrosone CB. Lifestyle factors and survival in women with breast cancer. *J. Nutr.* 137, 236S-242S (2007).
- 118 Goodwin PJ, Leszcz M, Ennis M *et al.* The effect of group psychosocial support on survival in metastatic breast cancer. *N. Engl. J. Med.* 345, 1719-1726 (2001).
- 119 Classen C, Butler LD, Koopman C *et al.* Supportive-expressive group therapy and distress in patients with metastatic breast cancer: a randomized clinical intervention trial. *Arch. Gen. Psychiatry* 58, 494-501 (2001).
- 120 Savard J, Simard S, Giguere I *et al.* Randomized clinical trial on cognitive therapy for depression in women with metastatic breast cancer: psychological and immunological effects. *Palliat. Support. Care* 4, 219-237 (2006).
- 121 Block KI, Gyllenhaal C, Tripathy D *et al.* Survival impact of integrative cancer care in advanced metastatic breast cancer. *Breast J.* 15, 357-366 (2009).
- \* **Describes beneficial effects of an integrative approach encompassing diet, exercise, and stress reduction in patients with advanced metastatic disease.**
- 122 Demark-Wahnefried W, Aziz NM, Rowland JH, Pinto BM. Riding the crest of the teachable moment: promoting long-term health after the diagnosis of cancer. *J. Clin. Oncol.* 23, 5814-5830 (2005).
- 123 Jones LW, Courneya KS, Fairey AS, Mackey JR. Effects of an oncologist's recommendation to exercise on self-reported exercise behavior in newly diagnosed breast cancer survivors: a single-blind, randomized controlled trial. *Ann. Behav. Med.* 28, 105-113 (2004).
- 124 Sabatino SA, Coates RJ, Uhler RJ, Pollack LA, Alley LG, Zauderer LJ. Provider counseling about health behaviors among cancer survivors in the United States. *J. Clin. Oncol.* 25, 2100-2106 (2007).
- 125 Kwan ML, Ambrosone CB, Lee MM *et al.* The Pathways Study: a prospective study of breast cancer survivorship within Kaiser Permanente Northern California. *Cancer Causes Control* 19, 1065-1076 (2008).

- 126 Ellsworth RE, Zhu K, Bronfman L, Gutchell V, Hooke JA, Shriver CD. The Clinical Breast Care Project: an important resource in investigating environmental and genetic contributions to breast cancer in African American women. *Cell Tissue Bank*. 9, 109-120 (2008).
- 127 Kelley MA. Culturally appropriate breast health educational intervention program for African-American women. *J. Natl. Black Nurses Assoc.* 15, 36-47 (2004).
- 128 Ashing-Giwa KT, Padilla G, Tejero J *et al.* Understanding the breast cancer experience of women: a qualitative study of African American, Asian American, Latina and Caucasian cancer survivors. *Psychooncology* 13, 408-428 (2004).
- 129 Patterson RE, Neuhouser ML, Hedderson MM, Schwartz SM, Standish LJ, Bowen DJ. Changes in diet, physical activity, and supplement use among adults diagnosed with cancer. *J. Am. Diet. Assoc.* 103, 323-328 (2003).
- 130 Demark-Wahnefried W, Peterson B, McBride C, Lipkus I, Clipp E. Current health behaviors and readiness to pursue life-style changes among men and women diagnosed with early stage prostate and breast carcinomas. *Cancer* 88, 674-684 (2000).
- 131 Pinto BM, Rabin C, Papandonatos GD, Frierson GM, Trunzo JJ, Marcus BH. Maintenance of effects of a home-based physical activity program among breast cancer survivors. *Support. Care Cancer* 16, 1279-1289 (2008).
- 132 Patterson RE, Neuhouser ML, Hedderson MM *et al.* Types of alternative medicine used by patients with breast, colon, or prostate cancer: predictors, motives, and costs. *J. Altern. Complement. Med.* 8, 477-485 (2002).
- 133 Schwartz K, Claxton G. The patient protection and affordable care act: how will it affect private health insurance for cancer patients? *Cancer J.* 16, 572-576 (2010).
- 134 Albright HW, Moreno M, Feeley TW *et al.* The implications of the 2010 Patient Protection and Affordable Care Act and the Health Care and Education Reconciliation Act on cancer care delivery. *Cancer* 117, 1564-1574 (2011).

### *Websites*

- 201 American Cancer Society. Learn About Cancer: Breast Cancer  
[www.cancer.org/cancer/breastcancer/detailedguide/breast-cancer-key-statistics](http://www.cancer.org/cancer/breastcancer/detailedguide/breast-cancer-key-statistics)
- 202 National Cancer Institute. Science Serving People: The Cost of Cancer  
[www.cancer.gov/aboutnci/servingpeople/cancer-statistics/costofcancer](http://www.cancer.gov/aboutnci/servingpeople/cancer-statistics/costofcancer)

- 203 US Department of Health and Human Services: Physical Activity Guidelines for Americans  
[www.health.gov/paguidelines/guidelines/default.aspx](http://www.health.gov/paguidelines/guidelines/default.aspx)
- 204 U.S. Department of Health and Human Services, National Institutes of Health, National Cancer Institute. President's Cancer Panel, 2006-2007 Annual Report. Promoting Healthy Lifestyles: Policy, Program, and Personal Recommendations for Reducing Cancer Risk  
<http://deainfo.nci.nih.gov/advisory/pcp/annualReports/pcp07rpt/pcp07rpt.pdf>
- 205 National Cancer Institute. Cancer Survivorship Research: Cancer Control and Population Sciences  
[www.cancercontrol.cancer.gov/ocs/](http://www.cancercontrol.cancer.gov/ocs/)
- 206 HealthCare.gov. Preventive Services Covered under the Affordable Care Act  
[www.healthcare.gov/law/about/provisions/services/lists.html](http://www.healthcare.gov/law/about/provisions/services/lists.html)

## **Acknowledgements**

This work was supported by the United States Department of Defense (Military Molecular Medicine Initiative MDA W81XWH-05-2-0075, protocol #01-20006) and was performed under the auspices of the Clinical Breast Care Project, a joint effort of many investigators and staff members whose contributions are gratefully acknowledged. The opinion and assertions contained herein are the private views of the authors and are not to be construed as official or as representing the views of the Department of the Army or the Department of Defense.

## **Figure Legend**

### **Figure 1. Mechanisms by which physiological stress may impact tumorigenesis.**

Stress resulting from a diagnosis of breast cancer may lead to secretion of hormones from the central nervous system, which alters the immune response and promotes angiogenesis, cellular proliferation, and invasion, thereby creating a microenvironment favorable to tumor growth and progression.

Reprinted by permission from Nature Publishing Group: *Nat. Rev. Cancer* 6, 240-248 (2006).

**Table 1. Study designs used to identify modifiable behaviors affecting incidence, prognosis, or QOL in breast cancer patients.**

Study type	Population(s)	Study design	Information provided
<i>Observational</i>			
Case report	Case(s) only	Clinical observation of patient(s) with outcome of interest	Hypothesis generating, no causality
Cross-sectional	Cases, controls	Outcome and behavior measured at same time point	Hypothesis generating, no causality
Case-control	Cases, controls	Past behaviors compared between cases and controls	Potential to establish causality
Cohort	Cases (with behavior), controls (without behavior)	Patients monitored over time to determine who developed outcome of interest	Can establish causality
<i>Interventional</i>			
Randomized Controlled Trial	Patients with similar characteristics, randomly assigned to intervention or non-intervention groups	Effect of intervention on outcome monitored over time	Provide evidence for causal effect

**Table 2. Large-scale studies evaluating the effects of diet on breast cancer outcomes.**

Study	Type	Participants	Dietary factor(s)	Results	Reference
NHS	Cohort (prospective)	1,982	Fat intake	No association	[21]
		2,619	Prudent vs Western diet	No association	[22]
LACE	Cohort (prospective)	2,321	100 foods, beverages, nutritional supplements, medicinal herbs	No association	[24]
		1,901	Prudent vs Western diet	No association	[25]
HEAL	Cohort (prospective)	670	Quality vs Poor diet	Reduced risk of breast cancer mortality with quality diet	[27]
		714	Quality vs Poor diet	Increased mental and physical functioning with quality diet	[28]
		746	Quality vs Poor diet	Lower CRP levels with quality diet	[29]
WINS	RCT	975 cases 1,462 controls	Low-fat diet vs minimal nutritional counseling	Cases had higher relapse-free survival	[30]
WHEL	RCT	1,537 cases 1,551 controls	High-fruit/vegetable vs 5 A Day diet	No association	[31]

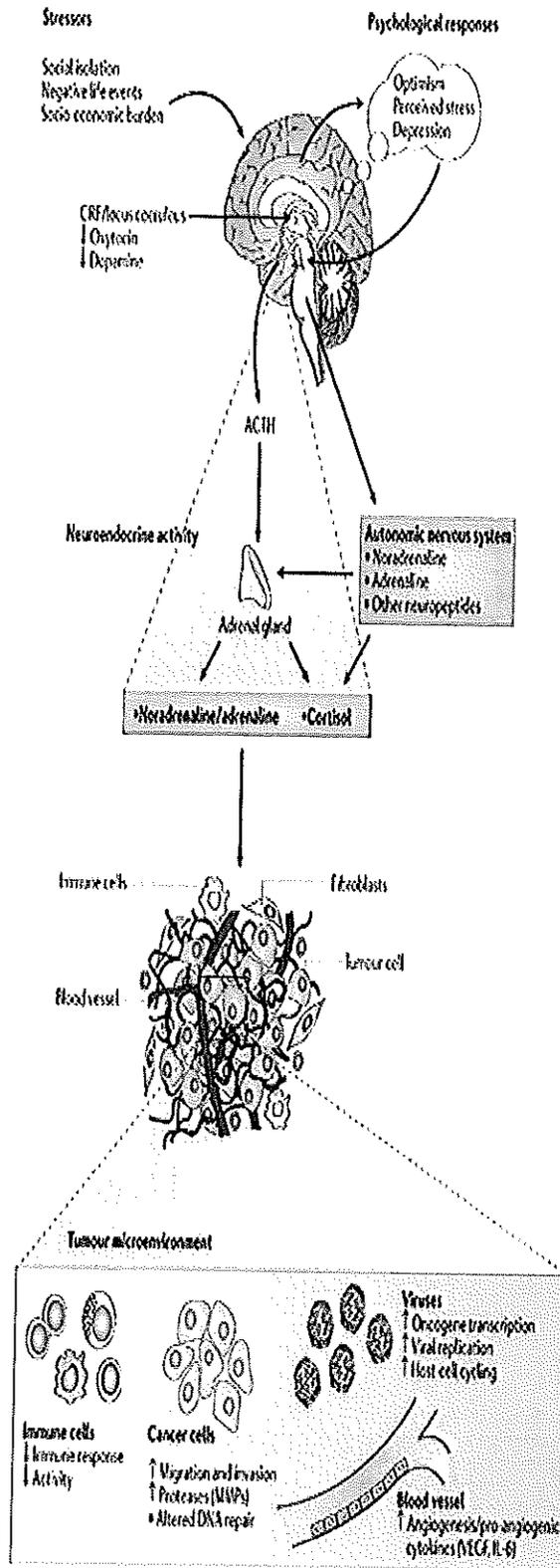
NHS: Nurses' Health Study; LACE: Life After Cancer Epidemiology; HEAL: Healthy Eating, Activity, and Lifestyle; WINS: Women's Intervention Nutrition Study; WHEL: Women's Healthy Eating and Living; RCT: randomized controlled trial; CRP: C-reactive protein.

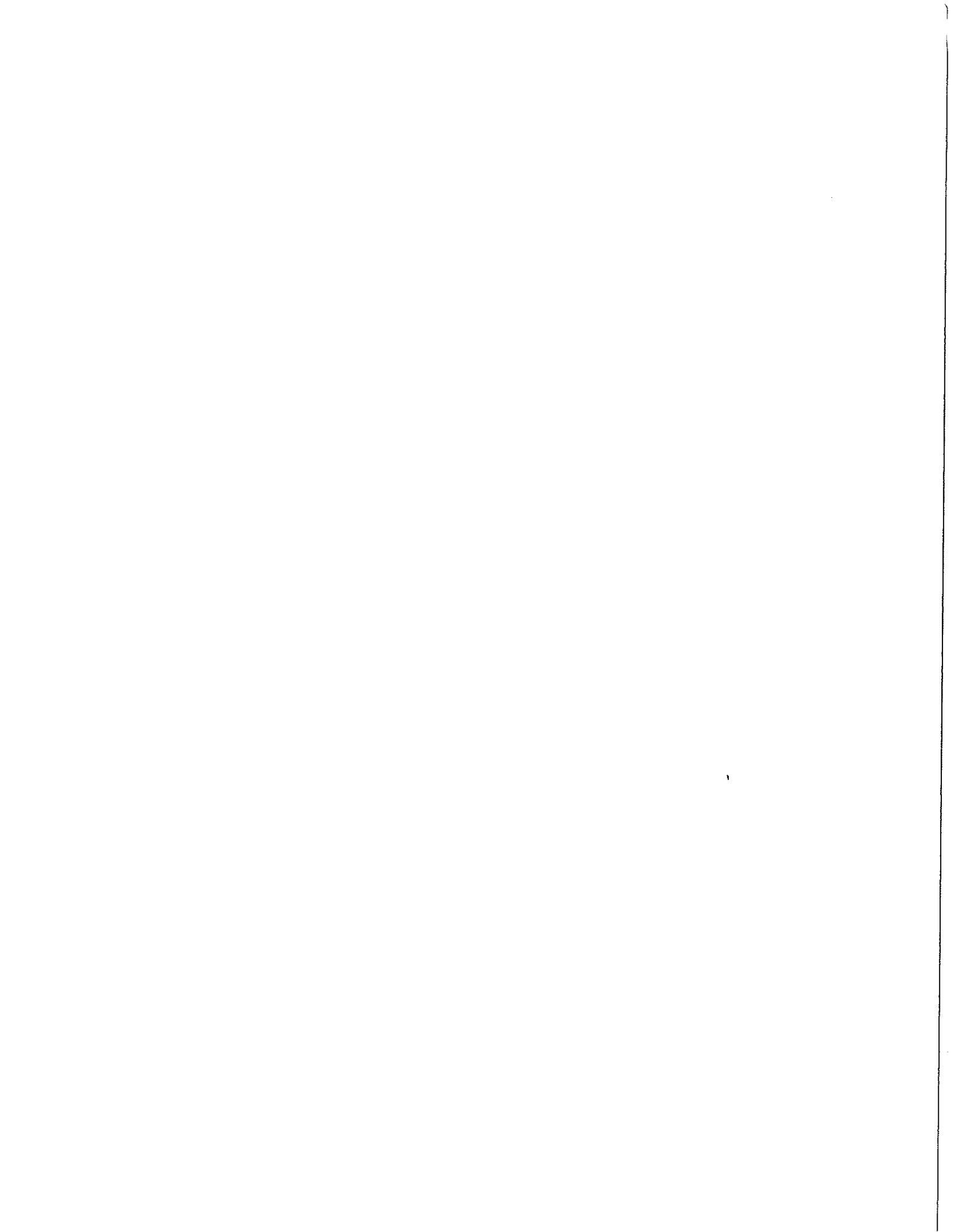
**Table 3. Large-scale studies evaluating the effects of physical activity on breast cancer outcomes.**

Study	Type	Participants	Physical Activity <sup>a</sup>	Results	Reference
NHS	Cohort (prospective)	2,987	MET hours per week	Optimal survival benefit in patients walking 3-5 hours/week at average pace	[77]
LACE	Cohort (prospective)	1,970	Work-, household-, recreation-, and transportation-related activity during six months prior to enrollment	No association with breast cancer recurrence/mortality; improvement in total mortality	[78]
CWLS	Cohort (prospective)	4,482	Post-diagnosis recreational physical activity	Women engaging in >2.8 MET hours/week had significantly lower risk of breast cancer mortality	[79]
HEAL	Cohort (prospective)	933	Physical activity in year before and two years after diagnosis	Increased physical activity after diagnosis associated with 45% lower risk of all-cause mortality; decreased physical activity after diagnosis associated with four-fold increased risk of all-cause mortality	[80]

NHS: Nurses' Health Study; LACE: Life After Cancer Epidemiology; CWLS: Collaborative Women's Longevity Study; HEAL: Healthy Eating, Activity, and Lifestyle; MET: metabolic equivalent task.

<sup>a</sup> All physical activity was self-reported.







## Development of a Bayesian Classifier for Breast Cancer Risk Stratification: A Feasibility Study

Alexander Stojadinovic, MD,<sup>a,b</sup> Christina Eberhardt,<sup>a</sup> Leonard Henry, MD,<sup>b,c</sup> John Eberhardt, BA,<sup>d</sup> Eric A. Elster, MD,<sup>c</sup> George E. Peoples, MD,<sup>b,e</sup> Aviram Nissan, MD,<sup>b,f</sup> and Craig D. Shriver, MD,<sup>a,b</sup>

<sup>a</sup>Department of Surgery, Division of Surgical Oncology, Walter Reed Army Medical Center, Washington, DC; <sup>b</sup>United States Military Cancer Institute, Clinical Trials Group, Washington, DC; <sup>c</sup>National Naval Medical Center, Bethesda, MD; <sup>d</sup>DecisionQ Corporation, Washington, DC; <sup>e</sup>Brooke Army Medical Center, Fort Sam Houston, TX; and <sup>f</sup>Hadassah Hebrew University Medical Center, Mount Scopus, Jerusalem, Israel

Correspondence: alexander.stojadinovic@amedd.army.mil

Published March 29, 2010

**Background:** Lifetime risk assessment tools are relatively limited in identifying breast cancer risk in younger women. The predictive value of mathematical models to estimate risk varies according to age, menopausal status, race/ethnicity, and family history. Current risk prediction models estimate population, *not* individual, levels of breast cancer risk; hence, individualized risk prediction models are needed to identify younger at-risk women who could benefit from timely risk reduction interventions. Clinical data collected as part of breast cancer screening studies may be modeled using Bayesian classification. **Purpose:** To train a proof-of-concept Bayesian classifier for breast cancer risk stratification. **Patients and Methods:** We trained a Bayesian belief network (BBN) model on cohort data (including risk factors, demographic, electrical impedance scanning (EIS), breast imaging, and biopsy data) from a prospective pilot screening trial in younger women ( $N = 591$ ). Receiver operating characteristic curve analysis and cross-validation of the model were used to derive preliminary guidance on the robustness of this approach and to gain insights into what a cross-validation exercise could provide in terms of risk stratification in a larger population. **Results:** Independent predictors of biopsy outcome in the BBN model included personal breast disease history, breast size, EIS (low vs high risk) and imaging results, and Gail cutoff (5-year risk:  $<1.66\%$  vs  $\geq 1.66\%$ ). Area under the receiver operating characteristic curve and positive predictive value for benign and malignant biopsy outcomes were 0.88 and 97% and 0.97 and 42%,

This clinical research effort was supported, in part, by the United States Military Cancer Institute and the Clinical Breast Care Project.

Our team comprises partly military service members and employees of the US Government. This work was prepared as part of our official duties. Title 17 USC 105 provides that "Copyright protection under this title is not available for any work of the United States Government." Title 17 USC 101 defines a US Government work as a work prepared by a military service member or employee of the US Government as part of that person's official duties.



respectively. Patient-specific probability of biopsy outcome given positive EIS result and Gail model 5-year risk  $\geq 1.66\%$  indicated that the combined effect of these predictors on likelihood that a biopsy would prove malignant exceeded the sum of the individual effects; breast cancer likelihood is as follows: 3% (EIS negative and Gail model 5-year risk  $< 1.66\%$ ) versus 9% (EIS positive and Gail model 5-year risk  $< 1.66\%$ ) versus 27% (EIS negative and Gail model 5-year risk  $\geq 1.66\%$ ) versus 45% (EIS positive and Gail model 5-year risk  $\geq 1.66\%$ ). **Conclusion:** Clinical data collected as part of breast cancer screening studies can be modeled using Bayesian classification. The BBN model may be predictive and may provide clinically useful incremental risk information for individualized breast cancer risk assessment in younger women.

---

Breast carcinoma is the most commonly diagnosed cancer and the second leading cause of cancer-related mortality among women in the United States.<sup>1</sup> In 2009, there were more than 192 000 estimated new cases of cancer of the breast and more than 40 000 disease-specific deaths.<sup>1</sup> Breast cancer-related mortality rates have steadily decreased over the past 2 decades, largely because of improved disease detection and therapy.<sup>2</sup>

As breast cancer in younger (age  $< 40$  years) women is infrequently diagnosed in the early stages by utilizing current screening guidelines, improved cancer screening and detection methods are important in current research, particularly in younger, at-risk women.<sup>3</sup> Breast cancer in younger women typically has unfavorable prognostic characteristics associated with increased disease-specific mortality; hence, early detection in younger women is imperative.<sup>4-6</sup> Younger women are not referred for periodic imaging unless they are identified as being "high risk."<sup>7</sup> "At risk" younger women with significant family history or genetic factors are encouraged to undergo frequent clinical and annual breast imaging surveillance and to consider chemoprevention.

While increased surveillance for at-risk women may be beneficial, the value of this approach is restricted by the rarity of breast cancer due to known genetic risk factors.<sup>8,9</sup> More than 90% of breast cancers are detected in women who are *not* identified as being high risk.<sup>3</sup> Furthermore, screening mammography is generally less accurate in younger women and those with increased breast tissue density commonly encountered in women younger than 40 years.<sup>10</sup> The reduced sensitivity of mammography for dense breasts impacts age groups in which a "life saved" often results in "higher" personal and societal costs in terms of altered life expectancy and personal productivity.<sup>11</sup>

Magnetic resonance imaging (MRI) is being used increasingly as a screening modality in high-risk women with a significant family history of breast cancer or those with *BRCA1* or *BRCA2* gene mutations, resulting in lifetime risk of cancer exceeding 20%.<sup>12,13</sup> Hence, breast MRI is currently applied to a relatively small proportion of all women. Because MRI is unaffected by breast tissue density, it is appealing to consider its use for screening young women in general; however, the high cost, requirement for intravenous contrast administration, and variable specificity limit its feasibility for widespread population-based screening.<sup>14,15</sup> Therefore, improved methods for risk prediction in younger women are needed to identify those at high risk for breast cancer.

Tamoxifen may be considered in both premenopausal and postmenopausal women, and raloxifene may be considered in postmenopausal women, with lobular carcinoma in situ or with a 5-year breast cancer risk estimate of 1.66% or higher (according to the Gail model



or the National Cancer Institute Breast Cancer Risk Assessment Tool), in order to reduce the risk of estrogen receptor-positive breast cancer.<sup>16</sup> In the National Surgical Adjuvant Breast and Bowel Project (NSABP) P-1 study tamoxifen (20 mg/day for 5 years) consistently reduced the incidence of breast cancer by 49% in at-risk women across all study age and risk groups (women age 35–59 with a  $\geq 1.66\%$  risk, those 60 years or older, or with those prior LCIS) thereby demonstrating the efficacy of chemoprevention for this disease.<sup>17</sup> The Multiple Outcomes of Raloxifene Evaluation (MORE), Continuing Outcomes Relevant to Evista (CORE), Raloxifene Use for the Heart (RUTH,) and NSABP Study of Tamoxifen and Raloxifene (STAR) trials demonstrated consistent significant reductions in estrogen receptor-positive breast cancer incidence in at-risk postmenopausal women.<sup>16</sup> Subsequent analyses of the NSABP P-1 study data suggested improved quality-adjusted survival and cost-effectiveness when tamoxifen was initiated as early as age 35 years in at-risk (Gail model 5-year risk  $\geq 1.66\%$ ) women.<sup>18,19</sup> Hence, identification of women who are at high risk and may benefit from chemoprevention is of particular importance.

Lifetime relative risk assessment tools (eg, Gail model) are available to identify women older than 35 years who are at risk for breast cancer. However, the predictive value of mathematical models to estimate breast cancer risk varies according to age, menopausal status, race/ethnicity, and family history of breast cancer. Instruments such as the Gail model are imperfect for identifying increased cancer risk in younger women.<sup>20</sup> Importantly, all current risk prediction models estimate population, *not* individual, levels of breast cancer risk. Currently, the only criterion generally used to identify high-risk young women who could benefit from chemoprevention is family/genetic history. The value of this risk estimation paradigm is limited by the rarity of breast cancer due to known gene mutations. Better individualized risk prediction models are needed to identify younger at-risk women who could benefit from risk reduction interventions and earlier chemoprevention.

Bayesian belief network (BBN) models have been used in research to better understand research data and biologic systems such as functional genomics. In recent years, these applications have been applied to better understand clinical problems, such as models developed to estimate breast cancer risk in mammographic microcalcifications and predict false-positive mammograms.<sup>21,22</sup> We believe that clinical data collected as part of breast cancer screening studies may be modeled using Bayesian classification. The objective of this study was to train a proof-of-concept machine-learned BBN model based on previously unpublished cohort data from this prospective pilot screening trial and to perform cross-validation for the purposes of evaluating the feasibility of using readily available data (including risk factors, demographic data, breast impedance, and breast imaging) and the Bayesian classification for breast cancer risk stratification (estimating biopsy outcome) in younger women.

## METHODS

### Patients

Between August 2002 and March 2005, a total of 591 female military healthcare beneficiaries were enrolled into this institution review board–approved, single-arm, prospective pilot screening trial. The clinical protocol was reviewed and approved by institutional review



boards of the Walter Reed Army Medical Center, Washington, DC, and the Keller Army Hospital, West Point, NY. Subjects who met the eligibility criteria were offered participation in this study. Subjects were recruited from the gynecology clinic or the Comprehensive Breast Center at Walter Reed Army Medical Center or the gynecology or family practice clinic at Keller Army Hospital. Study inclusion criteria consisted of younger women aged 18 to 49 years who provided informed consent prior to study enrollment and who were willing to be followed at the participating institution. Age was stratified for analysis as follows: younger than 30, 30 to 34, 35 to 39, and 40 to 49. Potential study subjects were excluded if they had breast surgery (including core biopsy) or were lactating within the preceding 3 months, had breast fine needle aspiration within the preceding 1 month, were pregnant, had electrically powered implanted devices (eg, pacemaker), or were undergoing chemotherapy or radiation treatment. Data collected for each study subject included age, race/ethnicity, clinical history (personal and family history of breast cancer, previous breast surgery or biopsy, and results of those interventions), hormonal information (age of menarche and first full-term pregnancy, menstrual status, date of last menstrual period, and exogenous hormone use), breast density and size (bra cup size), Gail model risk estimate, results of clinical breast examination (CBE), screening breast electrical impedance scanning (EIS), conventional imaging, and biopsy results. All study participants underwent EIS of the breast by using the T-Scan™ 2000ED (Mirabel Medical, Austin, TX) as previously described.<sup>23</sup>

### Statistical Methods

The BBN model was trained by using a priori variables to estimate the likely diagnostic outcome of breast biopsy. The BBN model was developed by using commercially available machine-learning algorithms (FasterAnalytics, DecisionQ, Washington, DC), which automatically learn network structures and joint probabilities from the prior probabilities in the data. BBN models are a type of directed acyclic graph, which means that they represent information in a hierarchical format. BBN models allow us to identify those variables that contain the most information and are thus most useful for estimating outcomes. The associations represented by BBN models are associations of conditional dependence, allowing us to estimate the posterior likelihood of a given outcome given prior observations.

In order to refine the model, a stepwise training process was used. Quantitative and qualitative assessments were used to optimize variable preparation and selection in order to produce the most robust and useful model. The objective was to produce the optimum biopsy outcome estimate through iterative quality assurance and reduction of confounding information. This process used to develop the model is summarized as follows: (1) preliminary modeling to identify appropriate machine-learning parameters, data quality issues, and confounding features and feature analogues that reduce model accuracy, (2) global modeling to set appropriate machine-learning parameters, remove identified analogs and confounders, and perform full "queue learning" to observe global data structure, (3) naive modeling of the outcome of interest to identify the relative contribution of covariates, and (4) focused modeling by using "queue learning" on subsets of variables identified in the prior steps to derive a more focused BBN model than that obtained in global modeling. By excluding marginal or noncontributory variables, the remaining ones are explored more exhaustively.



Cross-validation was performed on the final focused Bayesian classifier by using a train-and-test cross-validation methodology to produce classification accuracy estimates. Fivefold cross-validation was performed by randomizing the data set into 5 separate and unique train-and-test sets. Each set consists of a training set composed of 90% of patient records and a test set consisting of the remaining 10% of records. Once the model was constructed with a training set, the matching test set was entered into the model, generating a case-specific prediction for each record for independent variables of interest. A receiver operating characteristic curve was plotted for each test to calculate classification accuracy. The receiver operating characteristic curve was used to calculate area under the curve, a metric of overall model quality, and to calculate corresponding predictive values for biopsy outcome.

## RESULTS

The study population comprised an ethnically diverse group of younger women (41% non-Caucasian), healthcare beneficiaries in a free access system of military medical care. Of the 591 study participants, 67% were younger than 40 years (mean age:  $35 \pm 6.9$  years) and 90% were premenopausal (Table 1). Two percent of the study population was taking exogenous hormones at the time of study enrollment; however, there was no statistically significant association with disease ( $P = .95$ ). Fifty-five percent of participants had no family history of breast cancer, and family history was only marginally associated with biopsy outcome ( $P = .10$ ). The findings of CBE were statistically associated with both age ( $P = .01$ ) and disease ( $P \leq .001$ ); 31% of subjects had findings that were deemed not suspicious, whereas 4% of subjects had suspicious CBE findings. Five percent of study subjects had an estimated 5-year risk of breast cancer  $\geq 1.66\%$  according to the Gail model, and these findings were statistically associated with both disease and age of subject ( $P \leq .001$ ). Mammography was performed in 281 women and was found to be Breast Imaging Reporting and Data System (BIRADS) III or higher in 75 cases (27%); mammography was found to be statistically associated with both disease and age of subject ( $P \leq .001$ ). Breast ultrasound examination was performed in 258 women and was found to be BIRADS III or higher in 66 cases (26%); ultrasound was found to be statistically associated with disease ( $P \leq .001$ ) but not with age ( $P = .18$ ).

We also studied other well-known risk factors in our population and identified 3 risk factors that were not statistically associated with biopsy outcome: mean age at menarche ( $P = .12$ ), mean age at first pregnancy ( $P = .39$ ), and nulliparity ( $P = .93$ ). Finally, there was no statistically significant difference between the mean age of our population (35 years) and the mean age at time of cancer diagnosis (38 years,  $P = .35$ ) or diagnosis of premalignant histopathology (38 years,  $P = .56$ ). We tabulated data by age group and biopsy outcome as shown in Tables 1 and 2.

Of the 591 women enrolled in the study, 568 were found to be EIS negative (low risk) and 23 were found to be EIS positive (high risk). In the EIS-negative group, 95 underwent biopsy and 87 were benign on final histopathology. The remaining 8 were either premalignant ( $n = 4$ ) or malignant ( $n = 4$ ). In the EIS-positive group, 10 underwent biopsy; 5 were benign, whereas 5 were either premalignant ( $n = 3$ ) or malignant ( $n = 2$ ). Of 13 premalignant or malignant lesions, EIS identified 5 (38.5%). The negative predictive value (NPV) of the EIS-negative group was 92%, whereas the positive predictive value (PPV) of the EIS-positive group was 50%.



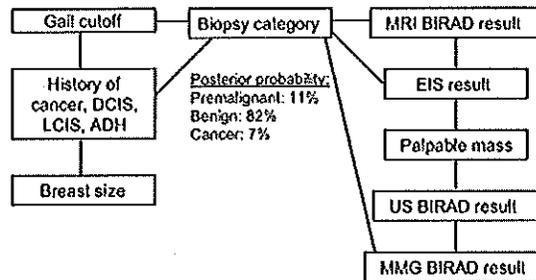
Table 1. Summary of study population characteristics by age\*

Characteristic	Age category, y				P
	<30	30-34	35-39	40-49	
Menopausal status					<.001
Premenopausal	124	114	122	171	
Postmenopausal	0	5	15	34	
Perimenopausal	0	0	1	4	
Not recorded	0	0	1	0	
Screening breast EIS result					.4515
Negative	120	117	132	199	
Positive	4	2	7	10	
Clinical breast examination result					.0077
No findings	78	88	77	141	
Not suspicious	45	28	52	57	
Suspicious	1	3	10	11	
Hormone replacement therapy					.0263
Current	0	2	2	11	
Past	0	1	0	3	
Never	124	116	137	195	
Mammogram results					<.001
BIRADS 0	1	0	3	10	
BIRADS I or II	6	16	61	109	
BIRADS III	1	6	8	21	
BIRADS IV	0	2	10	23	
BIRADS V	0	1	2	1	
No mammogram	116	94	55	45	
Breast biopsy category					.4076
Benign, no atypia	19	12	27	34	
Premalignant	1	0	2	4	
Infiltrating cancer or DCIS	0	1	2	3	
No biopsy (assumed benign)	104	106	108	168	
Family history category					.4080
One first degree	9	12	20	18	
One second degree	22	24	24	38	
One first and one or more second degree	9	9	12	11	
Two or more first degree	0	1	4	2	
Two or more second degree	7	12	10	21	
No significant family history	77	61	69	119	
Gail model 5-year risk category, %					<.001
<1.66	108	119	132	175	
≥1.66	0	0	3	27	

\*BIRADS indicates Breast Imaging Reporting and Data System; EIS, electrical impedance scanning; and ductal carcinoma in situ.

We trained a proof-of-concept BBN model on this pilot cohort data and performed analysis and cross-validation. The Bayesian Network shown in Figure 1 indicates that the six nearest independent associated features (direct relationship to breast biopsy diagnosis) used to estimate a breast biopsy diagnosis (Biopsy category) are screening breast EIS result, Gail model cutoff (5-year risk estimate <1.66% vs ≥1.66%), mammogram BIRAD result, MRI BIRAD result, breast size, and personal history of breast disease. This does not

mean, however, that “Any Palpable Mass” on CBE and ultrasound BIRAD results (indirect relationship to breast biopsy diagnosis—Fig 1) do not influence the estimate of likely biopsy diagnosis, but rather that they are conditionally independent of biopsy outcome, given knowledge of screening breast EIS and MMG BIRAD results.



**Figure 1.** Exemplar BBN model of breast biopsy outcome with expected frequency of histology in our study population. BBN indicates Bayesian belief network; LCIS, lobular carcinoma in situ; MRI, magnetic resonance imaging; BIRAD, Breast Imaging Reporting and Data System; EIS, electrical impedance scanning; US, ultrasound; MMG, mammography; DCIS, ductal carcinoma in situ; and ADH, atypical ductal hyperplasia.

The BBN feasibility model was validated using train-and-test cross-validation and produced strongly predictive areas under the curve (0.75–0.97) for differentiating malignancy and premalignant disease from benign findings (Table 3). Cross-validation also produces a 97% NPV and a 42% PPV for malignancy. It is important to note that with a relatively small set of outcomes, there is a high degree of variance in results between cross-validation exercises (Table 3). The BBN model is a recursive information structure, and the inclusion of conditional dependence between predictive variables guards against overinterpretation of data (overfitting). The model informs estimates not only through estimation of biopsy outcome but also through estimation of as-yet-unknown imaging results, wherein estimates of biopsy outcome can be derived from available clinical and imaging data, even if some imaging studies are unavailable at the time of biopsy outcome estimation.

To demonstrate the use of this type of model, we walk through an example case of how we can use available information to estimate clinically relevant outcomes using the network. Knowledge of breast size (bra cup) B (Fig 2: Evidence 1) results in slightly lower risk of cancerous biopsy result (–3.6%) for the subject compared with our reference population. When the additional knowledge of Gail model 5-year high-risk estimate (Fig 3: Evidence 2) is added to refine the posterior estimate of biopsy outcome given previously known breast size (bra cup) B, there is a 12% increased likelihood of cancerous biopsy and a 17% increase in the likelihood of premalignant histology, relative to our overall study cohort. Finally, adding knowledge of a positive (high risk) EIS screening result (Fig 4: Evidence 3) increases the posterior risk estimate of cancerous biopsy by 21% and the risk estimate of premalignant disease by 35%. Each posterior probability estimate is the result of adding evidence for each factor being used to make the estimate. As new evidence is “added” for a given factor, the existing evidence already input remains unchanged. However, other nodes



**Table 2. Summary of study population characteristics by biopsy category\***

Characteristic	Biopsy category				P
	Benign, no atypia	Infiltrating cancer or DCIS	Premalignant	No biopsy (benign)	
Mean age at menarche, y <sup>†</sup>	13	12	13	13	.9997
Mean age at first pregnancy, y <sup>†</sup>	25	22	24	24	.8622
% Nulliparous	79.1	100.0	66.7	75.8	.9299
Mean age at diagnosis, y <sup>†</sup>	36	38	38	35	.0959
Menopausal status					.0307
Premenopausal	86	6	6	433	
Postmenopausal	6	0	0	48	
Perimenopausal	0	0	1	4	
Screening breast EIS result					<.001
Negative	87	4	4	473	
Positive	5	2	3	13	
Clinical breast examination result					<.001
No findings	19	2	4	359	
Not suspicious	55	1	2	124	
Suspicious	18	3	1	3	
Hormone replacement therapy					.9460
Current	1	0	0	14	
Past	1	0	0	3	
Never	90	6	7	469	
Bra cup size					.0094
A	4	3	0	39	
B	25	1	3	123	
C	18	0	1	110	
D+	6	1	1	65	
Not recorded	39	1	2	149	
Mammogram results					<.001
BIRADS 0	3	0	3	8	
BIRADS I or II	27	0	2	163	
BIRADS III	6	0	0	30	
BIRADS IV	26	3	2	4	
BIRADS V	0	3	0	1	
No mammogram	30	0	0	280	
Family history category					.1035
One first degree	7	1	0	51	
One second degree	18	1	2	87	
One first degree and one or more second degree	0	1	0	6	
Two or more first degree	6	0	2	42	
Two or more second degree	55	2	3	266	
No significant family history	6	1	0	34	
Gail model 5-year risk category, %					<.001
<1.66	86	2	5	441	
≥1.66	4	3	1	22	
Not recorded	2	1	1	23	

\*BIRADS indicates Breast Imaging Reporting and Data System; EIS, electrical impedance scanning; and DCIS, ductal carcinoma in situ.

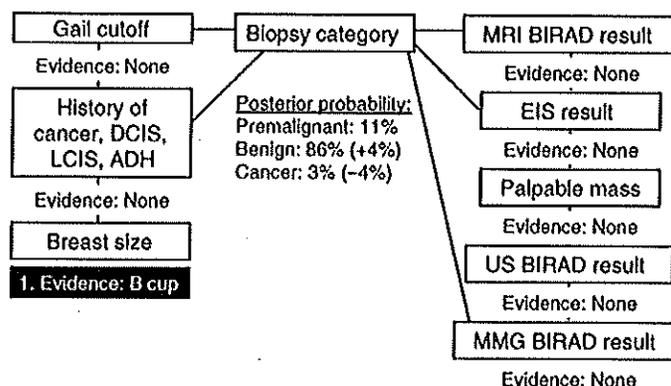
<sup>†</sup>P value of cancer and premalignant populations compared with benign population.



in the network for which no evidence is available have their posterior probabilities updated given other evidence that has been input, and these new posteriors in turn influence the predicted variable.

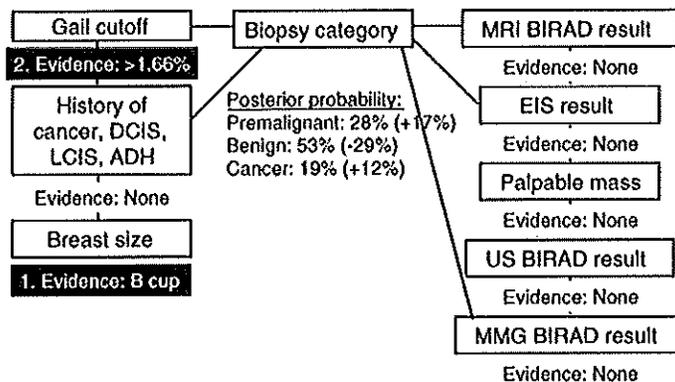
**Table 3.** Feasibility model cross-validation statistics

	Area under the curve			Predictive value, %		
	Benign	Cancer	Premalignant	Benign	Cancer	Premalignant
Internal	0.98	0.99	0.97	100.0	66.7	50.0
Exercise 1	0.94	0.98	0.58	100.0	33.3	0.0
Exercise 2	0.81	0.98	0.53	94.4	NA	0.0
Exercise 3	0.92	0.98	0.90	93.8	33.3	0.0
Exercise 4	0.89	0.95	0.95	100.0	50.0	16.7
Exercise 5	0.86	0.98	0.79	94.4	50.0	NA
<b>Mean</b>	<b>0.88</b>	<b>0.97</b>	<b>0.75</b>	<b>96.5</b>	<b>41.7</b>	<b>4.2</b>
95% CI low	0.82	0.96	0.51	92.6	26.4	0.0
95% CI high	0.95	0.99	0.98	100.0	57.0	17.4

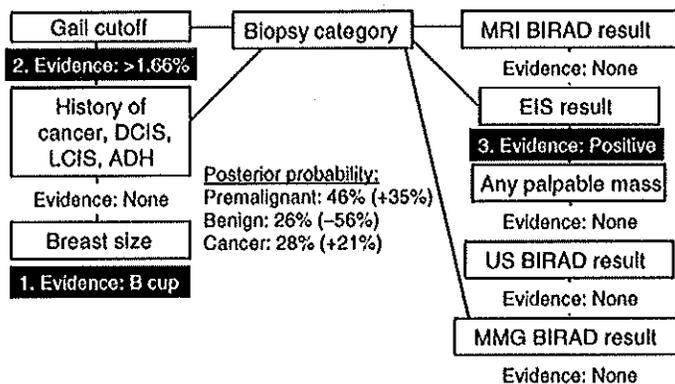


**Figure 2.** Probability of cancer diagnosis given only evidence of breast size (bra cup) B is 3%. LCIS indicates lobular carcinoma in situ; MRI, magnetic resonance imaging; BIRAD, Breast Imaging Reporting and Data System; EIS, electrical impedance scanning; US, ultrasound; MMG, mammography; DCIS, ductal carcinoma in situ; and ADH, atypical ductal hyperplasia.

As all features in the model are, at some level, conditionally dependent with biopsy outcome, those features available at the time of initial clinical visit (a priori knowledge) can be selected and applied to the model to estimate biopsy outcome. Subsets of features can also be used to generate an inference table (Table 4) that can be used to quickly estimate biopsy outcome for all known combinations of the identified features. The incremental values of both screening breast EIS and the Gail model data are shown in the inference table, Table 4. Under the most favorable circumstances (EIS negative and Gail model 5-year risk <1.66%) the risk of malignancy is 3%, and under the least favorable circumstances (screening EIS positive and Gail model 5-year risk  $\geq$ 1.66%), the risk of malignancy is 45%.



**Figure 3.** Probability of cancer diagnosis given only evidence of bra cup size B and Gail model 5-year risk  $\geq 1.66\%$  is 19%. LCIS indicates lobular carcinoma in situ; MRI, magnetic resonance imaging; BIRAD, Breast Imaging Reporting and Data System; EIS, electrical impedance scanning; US, ultrasound; MMG, mammography; DCIS, ductal carcinoma in situ; and ADH, atypical ductal hyperplasia.



**Figure 4.** Probability of cancer diagnosis given only evidence of bra cup size B, Gail model 5-year risk  $\geq 1.66\%$ , and EIS positive result is 28%. EIS indicates electrical impedance scanning; LCIS, lobular carcinoma in situ; MRI, magnetic resonance imaging; BIRAD, Breast Imaging Reporting and Data System; US, ultrasound; MMG, mammography; DCIS, ductal carcinoma in situ; and ADH, atypical ductal hyperplasia.

## DISCUSSION

For the majority of younger women, namely those considered to be average risk for developing breast cancer under the current risk assessment and screening model, the only generally available risk assessment modality is CBE, which is imperfect as a screening tool as it has an unacceptably low sensitivity and high false-positive rate compared with mammography.<sup>24</sup> CBE detects cancers only when they have advanced to the point of being



palpable. When cancers are clinically palpable, they have reached a more advanced stage of disease. Palpable breast cancers typically require more aggressive and costly treatments, with concomitant worse quality of life and oncological outcomes.<sup>25,26</sup>

**Table 4.** Probability of biopsy diagnosis given Gail model risk estimate and breast EIS result\*

Case frequency, %	Known evidence		Estimated outcome, %		
	EIS result	Gail cutoff <sup>a</sup>	Biopsy category		
			Benign, no atypia	Infiltrating cancer or DCIS	Premalignant
74	Negative	Negative	91	3	7
16	Positive	Negative	65	9	26
7	Negative	Positive	53	27	19
3	Positive	Positive	18	45	37

\*EIS indicates electrical impedance scanning; DCIS, ductal carcinoma in situ.

The early detection of breast cancer in younger women is very important, particularly because it demonstrates aggressive tumor biology with rapid tumor growth, demonstrates a relatively short preclinical disease phase, and has worse cancer-specific survival than in older women.<sup>6,27,28</sup> A risk stratification paradigm that improves upon the interpretation of existing clinical information should allow us to detect disease at an earlier stage of development. The Gail model offers an improvement in predicting risk, yet it is still an imperfect tool because it is designed using a primarily older, Caucasian population. Further, as more effective personalized detection, prevention, and treatment strategies become available for breast cancer in younger women, strategies and technologies that support truly personalized risk assessment and screening can favorably impact survival, especially if conducted at shorter intervals than in older women.<sup>28-33</sup>

Recognizing the need for individualized breast cancer risk assessment tools for younger women, we conducted a feasibility study to determine whether a machine-learned BBN model could be developed to support individualized breast cancer risk stratification. The model trained and cross-validated in this study was based on data from a prospective pilot screening trial in younger women ( $N = 591$ ) and produced receiver operating characteristic curves, when cross-validated, with areas under the curve of 0.88, 0.97 and 0.75 for benign, malignant, and premalignant findings, respectively. This proof-of-principle study shows that clinical data collected as part of routine, current breast cancer screening studies can be developed into enhanced screening tools with improved sensitivity and specificity by using machine-learned BBN models. These networks can use readily available information to estimate clinically relevant outcomes, providing clinically useful incremental risk information for individualized breast cancer risk assessment in younger women.

In our BBN model, the features showing direct conditional dependence with biopsy outcome include (1) personal history of breast disease, (2) breast size (bra cup), (3) EIS (low vs high risk), (4) breast imaging results, and (5) Gail model risk cutoff (5-year risk  $<1.66$  vs  $\geq 1.66$ ). Each of these variables is also significantly related to biopsy category when examined using the calculated  $\chi^2$ . CBE and breast ultrasound results were also determined to be statistically associated with biopsy outcome, and the Bayesian model



includes these features as well, but they are associated with biopsy outcome through EIS and mammography results. Other features that showed statistical significance, but dropped out of the training process in the model, include patient ethnicity, menopausal status, and prior breast biopsy. The machine-learning process is designed to produce a simpler (parsimonious) model whenever possible; thus, these additional factors were likely surpassed by more specific imaging and personal history risk factors.

Interestingly, a number of attributes were found to have statistically significant association with patient age, including CBE findings, mammography BIRADS category, nulliparity, and Gail model 5-year risk score. Conversely, certain factors found to be significantly associated with biopsy outcome, using both bivariate statistical tests and the machine-learning process, were not associated with patient age: breast (bra cup) size, EIS screening examination, or MRI. There was no statistically significant difference in mean age at diagnosis of premalignant or malignant disease compared with the mean age of our study population, and when biopsy results were examined by age category, they still did not demonstrate any statistically significant associations. Finally, features considered as well-established breast cancer risk factors in the general population were not found to have statistically significant associations with biopsy outcome in our younger population, including family history of breast cancer, age at menarche, nulliparity, and age at first pregnancy. While these are considered as common risk factors for developing breast cancer, data suggest that these variables are not effective to determine individual risk of cancerous or premalignant lesions in our study population. Although we must be cautious not to overinterpret these findings, they do raise important questions about the appropriate risk measures in a younger population of ethnically diverse women when well-established risk factors have statistically significant association with subjects' age while our disease outcome appears to be age independent.

Having trained, encoded, and validated the machine-learned BBN model, we can estimate the likely biopsy outcome given readily available clinical and imaging data. However, the BBN model not only allows the posterior estimation of the likely biopsy outcome but also identifies a hierarchy of conditional dependence that allows us to identify which pieces of information are most useful in calculating our estimate. This hierarchy also defines how independent variables influencing biopsy outcome also influence one another, providing a better understanding of how the estimate is derived and providing an opportunity to estimate missing parameters by using those currently available for any given patient. It is notable that the combined effect of these independent predictors on the likelihood of disease is greater than the sum of the individual effects. By way of example, mammography finding of BIRAD IV increases the likelihood of a malignant biopsy result in our study population by 5%, whereas a Gail model 5-year risk score of greater than 1.66% increases the likelihood of malignancy by 26%, yet together these findings increase the likelihood of disease by 42%—greater than the sum of their individual effects.

Importantly, the current proof-of-principle model estimates probability of having a breast abnormality biopsied and having it show the underlying malignancy. The ultimate clinical utility of such a model with appropriate sample size, population disease incidence, and follow-up will be based on robust predictive value of the model for developing breast cancer. The most clinically relevant model will be based on easily obtainable nonimaging parameters to identify at-risk women, who could benefit from breast imaging-based screening and risk reduction interventions.



## CONCLUSION

A need exists for a breast cancer risk estimation paradigm that can be used along with relevant demographic, clinical, and other readily obtainable patient-specific data in younger women in order to provide an individualized cancer risk assessment, direct screening efforts that can lead to prophylaxis, and detect breast cancer at an early stage. The computational complexity of designing such risk stratification algorithms for the average-risk woman necessitates a large, multidimensional cohort and requires a selection and encoding methodology that is both robust and transparent so as to sustain clinical scrutiny and improve clinical practice. For our study, we integrated multidimensional clinical, imaging, and pathological data from a prospective cohort to test the feasibility of model development. Using this cohort, we trained a BBN model by using machine-learning algorithms and developed a risk classification model for the average-risk younger woman, with promising cross-validation results. Our proof-of-concept study shows that this type of model could be used to perform individualized screening on a regular basis, using available clinical data at low cost, and that it warrants further assessment and independent testing.

## Acknowledgment

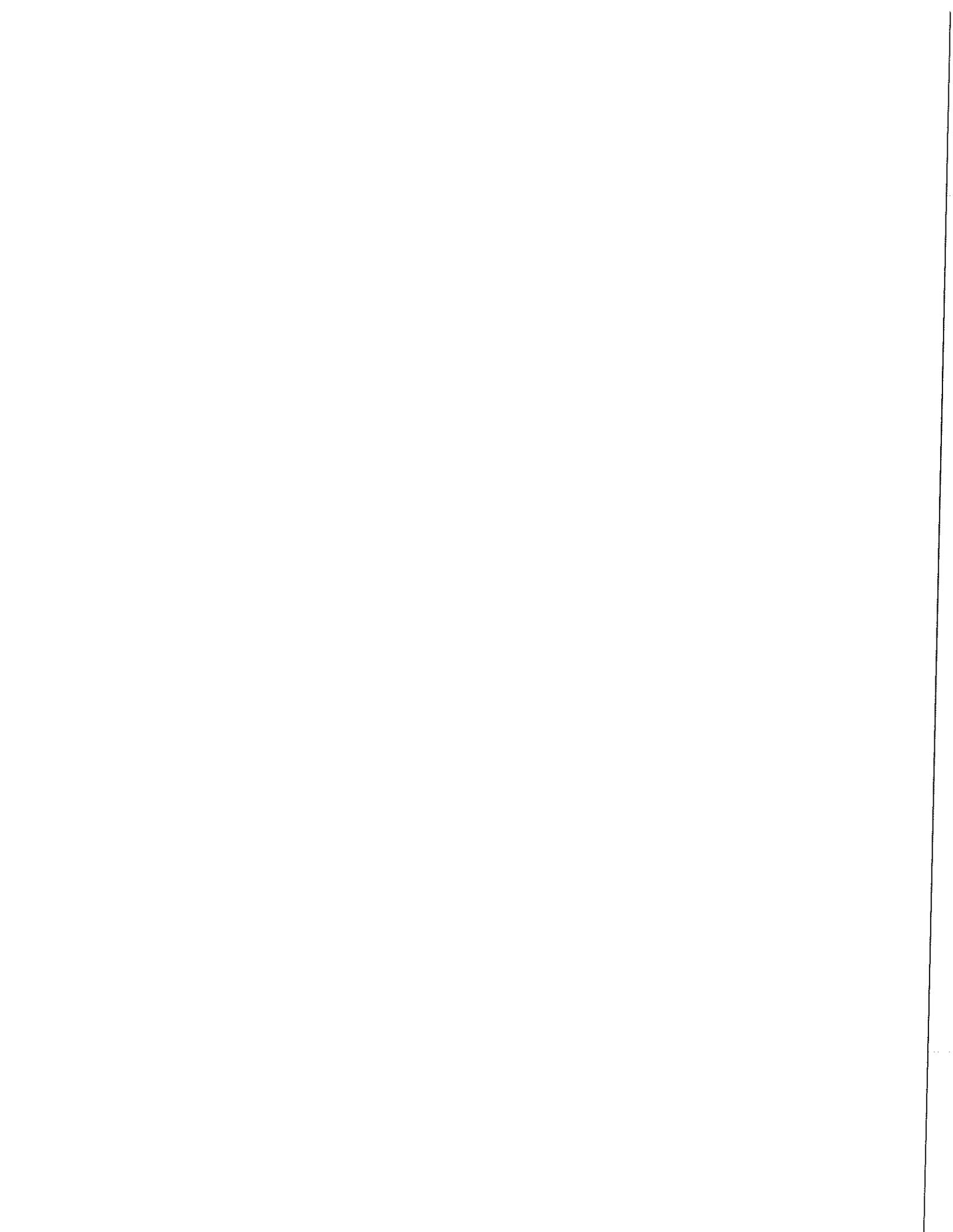
The authors thank Tiffany Felix for her invaluable assistance supported in part by the Henry M. Jackson Foundation for the Advancement of Military Medicine and the members and staff of the United States Military Cancer Institute and the Clinical Breast Care Project for their consistent support of this collaborative research effort.

## REFERENCES

1. Jemal A, Siegel R, Ward E, Hao Y, Xu J, Thun MJ. Cancer statistics 2009. *CA Cancer J Clin.* 2009;59(4):225-49.
2. American Cancer Society. *Cancer Facts and Figures 2008*. Atlanta, GA: American Cancer Society; 2009.
3. Joy JE, Penhoet EE, Petitti DB, eds. *Saving Women's Lives: Strategies for Improving Breast Cancer Detection and Diagnosis*. Washington, DC: National Academies Press; 2004.
4. Winchester DP, Osteen RT, Menck HR. The National Cancer Data Base report on breast carcinoma characteristics and outcome in relation to age. *Cancer.* 1996;78:1838-43.
5. Chung M, Chang HR, Bland KI, Wanebo HJ. Younger women with breast carcinoma have a poorer prognosis than older women. *Cancer.* 1996;77:97-103.
6. Xiong Q, Valero V, Kau V, et al. Female patients with breast carcinoma age 30 years and younger have a poor prognosis: the M.D. Anderson Cancer Center experience. *Cancer.* 2001;92(10):2523-8.
7. Smith RA, Saslow D, Sawyer KA, et al. American Cancer Society guidelines for breast cancer screening: update 2003. *CA Cancer J Clin.* 2003;53(3):141-69.
8. Kollias J, Sibbering DM, Blamey RW, et al. Screening women aged less than 50 years with a family history of breast cancer. *Eur J Cancer.* 1998;34:878-83.
9. Tilanus-Lindhorst MM, Bartels CC, Obedijn AL, Oudkerk M. Earlier detection of breast cancer by surveillance of women at familial risk. *Eur J Cancer.* 2000;36:514-9.
10. Carney PA, Miglioretti DL, Yankaskas BC, et al. Individual and combined effects of age, breast density, and hormone replacement therapy use on the accuracy of screening mammography. *Ann Intern Med.* 2003;138(3):168-75.



11. Kroenke CH, Rosner B, Chen WY, Kawachi I, Colditz GA, Holmes MD. Functional impact of breast cancer by age at diagnosis. *J Clin Oncol.* 2004;22(10):1849-56.
12. Kriege M, Brekelmans CTM, Boetes C, et al. Efficacy of MRI and mammography for breast-cancer screening in women with a familial or genetic predisposition. *N Engl J Med.* 2004;351:427-37.
13. Warner E, Plewes DB, Hill KA, et al. Surveillance of BRCA1 and BRCA2 mutation carriers with magnetic resonance imaging, ultrasound, mammography, and clinical breast examination. *JAMA.* 2004;292:1317-25.
14. Morris EA, Schwartz LH, Dershaw DD, van Zee KJ, Abramson AF, Liberman L. MR imaging of the breast in patients with occult primary breast carcinoma. *Radiology.* 1997;205:437-40.
15. Orel SG, Schnall MD, LiVolsi VA, Troupin RH. Suspicious breast lesions: MR imaging with radiologic-pathologic correlation. *Radiology.* 1994;190:485-93.
16. Visvanathan K, Chlebowski RT, Hurley P, et al. American Society Of Clinical Oncology clinical practice guideline update on the use of pharmacologic interventions including tamoxifen, raloxifene, and aromatase inhibition for breast cancer risk reduction. *J Clin Oncol.* 2009;27(19):3235-58.
17. Fisher B, Costantino JP, Wickerham DL, et al. Tamoxifen for prevention of breast cancer: report of the National Surgical Adjuvant Breast and Bowel Project P-1 Study. *J Natl Cancer Inst.* 1998;90(18):1371-88.
18. Grann VR, Sundararajan V, Jacobson JS, et al. Decision analysis of tamoxifen for the prevention of invasive breast cancer. *Cancer J Sci Am.* 2000;6:169-78.
19. Hershman D, Sundararajan V, Jacobson JS, Heitjan DF, Neugut AI, Grann VR. Outcomes of tamoxifen chemoprevention for breast cancer in very high-risk women: a cost-effectiveness analysis. *J Clin Oncol.* 2001;20:9-16.
20. MacKarem G, Roche CA, Hughes KS. The effectiveness of the Gail model in estimating risk for development of breast cancer in women under 40 years of age. *Breast J.* 2001;7(1):34-9.
21. Burnside ES, Rubin DL, Fine JP, Shachter RD, Sisney GA, Leung WK. Bayesian network to predict breast cancer risk of mammographic microcalcifications and reduce number of benign biopsy results: initial experience. *Radiology.* 2006;240(3):666-73.
22. Christiansen CL, Wang F, Barton MB, et al. Predicting the cumulative risk of false-positive mammograms. *J Natl Cancer Inst.* 2000;92(20):1657-66.
23. Stojadinovic A, Nissan A, Gallimidi Z, et al. Electrical impedance scanning for the early detection of breast cancer in young women: preliminary results of a multicenter prospective trial. *J Clin Oncol.* 2005;23(12):2703-15.
24. Kopans DB. *Breast Imaging.* Philadelphia, PA: Lippincott-Raven; 1998.
25. Barth RJ Jr, Gibson GR, Carney PA, Mott LA, Becher RD, Poplack SP. Detection of breast cancer on screening mammography allows patients to be treated with less-toxic therapy. *AJR Am J Roentgenol.* 2005;184:324-9.
26. McPherson CP, Swenson KK, Jolitz G, Murray CL. Survival of women ages 40-49 years with breast carcinoma according to method of detection. *Cancer.* 1997;79:1923-32.
27. Gnerlich JL, Deshpande AD, Jeffe DB, Sweet A, White N, Margenthaler JA. Elevated breast cancer mortality in women younger than age 40 years compared with older women is attributed to poorer survival in early stage disease. *J Am Coll Surg.* 2009;208:341-7.
28. Yildirim E, Dalgıç T, Berberoğlu U. Prognostic significance of young age in breast cancer. *J Surg Oncol.* 2000;74:267-72.
29. Feig SA. Mammographic screening of women aged 40-49 years: benefit, risk, and cost considerations. *Cancer.* 1995;76:2097-106.
30. Moskowitz M. Breast cancer: age-specific growth rates and screening strategies. *Radiology.* 1986;161:37-41.
31. Pelikan S, Moskowitz M. Effects of lead time, length bias, and false-negative assurance on screening for breast cancer. *Cancer.* 1993;71:1998-2005.
32. Tabár L, Chen HH, Fagerberg G, Duffy SW, Smith TC. Recent results from the Swedish two-county trial: the effects of age, histologic type, and mode of detection on the efficacy of breast cancer screening. *J Natl Cancer Inst Monogr.* 1997;22:43-7.
33. Tabár L, Fagerberg G, Duffy SW, Day NE, Gad A, Grönroft O. Update of the Swedish two-county program of mammographic screening for breast cancer. *Radiol Clin North Am.* 1992;30:187-210.



# Comprehensive molecular portraits of human breast tumours

The Cancer Genome Atlas Network\*

We analysed primary breast cancers by genomic DNA copy number arrays, DNA methylation, exome sequencing, messenger RNA arrays, microRNA sequencing and reverse-phase protein arrays. Our ability to integrate information across platforms provided key insights into previously defined gene expression subtypes and demonstrated the existence of four main breast cancer classes when combining data from five platforms, each of which shows significant molecular heterogeneity. Somatic mutations in only three genes (*TP53*, *PIK3CA* and *GATA3*) occurred at >10% incidence across all breast cancers; however, there were numerous subtype-associated and novel gene mutations including the enrichment of specific mutations in *GATA3*, *PIK3CA* and *MAP3K1* with the luminal A subtype. We identified two novel protein-expression-defined subgroups, possibly produced by stromal/microenvironmental elements, and integrated analyses identified specific signalling pathways dominant in each molecular subtype including a HER2/phosphorylated HER2/EGFR/phosphorylated EGFR signature within the HER2-enriched expression subtype. Comparison of basal-like breast tumours with high-grade serous ovarian tumours showed many molecular commonalities, indicating a related aetiology and similar therapeutic opportunities. The biological finding of the four main breast cancer subtypes caused by different subsets of genetic and epigenetic abnormalities raises the hypothesis that much of the clinically observable plasticity and heterogeneity occurs within, and not across, these major biological subtypes of breast cancer.

Breast cancer is one of the most common cancers with greater than 1,300,000 cases and 450,000 deaths each year worldwide. Clinically, this heterogeneous disease is categorized into three basic therapeutic groups. The oestrogen receptor (ER) positive group is the most numerous and diverse, with several genomic tests to assist in predicting outcomes for ER<sup>+</sup> patients receiving endocrine therapy<sup>1,2</sup>. The *HER2* (also called *ERBB2*) amplified group<sup>3</sup> is a great clinical success because of effective therapeutic targeting of *HER2*, which has led to intense efforts to characterize other DNA copy number aberrations<sup>4,5</sup>. Triple-negative breast cancers (TNBCs, lacking expression of ER, progesterone receptor (PR) and *HER2*), also known as basal-like breast cancers<sup>6</sup>, are a group with only chemotherapy options, and have an increased incidence in patients with germline *BRCA1* mutations<sup>7,8</sup> or of African ancestry<sup>9</sup>.

Most molecular studies of breast cancer have focused on just one or two high information content platforms, most frequently mRNA expression profiling or DNA copy number analysis, and more recently massively parallel sequencing<sup>10–12</sup>. Supervised clustering of mRNA expression data has reproducibly established that breast cancers encompass several distinct disease entities, often referred to as the intrinsic subtypes of breast cancer<sup>13,14</sup>. The recent development of additional high information content assays focused on abnormalities in DNA methylation, microRNA (miRNA) expression and protein expression, provide further opportunities to characterize more completely the molecular architecture of breast cancer. In this study, a diverse set of breast tumours were assayed using six different technology platforms. Individual platform and integrated pathway analyses identified many subtype-specific mutations and copy number changes that identify therapeutically tractable genomic aberrations and other events driving tumour biology.

## Samples and clinical data

Tumour and germline DNA samples were obtained from 825 patients. Different subsets of patients were assayed on each platform:

466 tumours from 463 patients had data available on five platforms including Agilent mRNA expression microarrays ( $n = 547$ ), Illumina Infinium DNA methylation chips ( $n = 802$ ), Affymetrix 6.0 single nucleotide polymorphism (SNP) arrays ( $n = 773$ ), miRNA sequencing ( $n = 697$ ), and whole-exome sequencing ( $n = 507$ ); in addition, 348 of the 466 samples also had reverse-phase protein array (RPPA) data ( $n = 403$ ). Owing to the short median overall follow up (17 months) and the small number of overall survival events (93 out of 818), survival analyses will be presented in a later publication. Demographic and clinical characteristics are presented in Supplementary Table 1.

## Significantly mutated genes in breast cancer

Overall, 510 tumours from 507 patients were subjected to whole-exome sequencing, identifying 30,626 somatic mutations comprised of 28,319 point mutations, 4 dinucleotide mutations, and 2,302 insertions/deletions (indels) (ranging from 1 to 53 nucleotides). The point mutations included 6,486 silent, 19,045 missense, 1,437 nonsense, 26 read-through, 506 splice-site mutations, and 819 mutations in RNA genes. Comparison to COSMIC and OMIM databases identified 619 mutations across 177 previously reported cancer genes. Of 19,045 missense mutations, 9,484 were predicted to have a high probability of being deleterious by Condel<sup>15</sup>. The MuSiC package<sup>16</sup>, which determines the significance of the observed mutation rate of each gene based on the background mutation rate, identified 35 significantly mutated genes (excluding LOC or Ensembl gene IDs) by at least two tests (convolution and likelihood ratio tests) with false discovery rate (FDR) <5% (Supplementary Table 2).

In addition to identifying nearly all genes previously implicated in breast cancer (*PIK3CA*, *PTEN*, *AKT1*, *TP53*, *GATA3*, *CDH1*, *RBI*, *MLL3*, *MAP3K1* and *CDKN1B*), a number of novel significantly mutated genes were identified including *TBX3*, *RUNX1*, *CBBF*, *AFF2*, *PIK3R1*, *PTPN22*, *PTPRD*, *NFI*, *SE3B1* and *CCND3*. *TBX3*, which is mutated in ulnar-mammary syndrome and involved in mammary gland development<sup>17</sup>, harboured 13 mutations (8 frame-shift indels,

\*A list of participants and their affiliations appears at the end of the paper.

1 in-frame deletion, 1 nonsense, and 3 missense), suggesting a loss of function. Additionally, 2 mutations were found in *TBX4* and 1 mutation in *TBX5*, which are genes involved in Holt–Oram syndrome<sup>18</sup>. Two other transcription factors, *CTCF* and *FOXA1*, were at or near significance harbouring 13 and 8 mutations, respectively. *RUNX1* and *CBFβ*, both rearranged in acute myeloid leukaemia and interfering with haematopoietic differentiation, harboured 19 and 9 mutations, respectively. *PIK3RI* contained 14 mutations, most of which clustered in the PIK3CA interaction domain similar to previously identified mutations in glioma<sup>19</sup> and endometrial cancer<sup>20</sup>. We also observed a statistically significant exclusion pattern among *PIK3RI*, *PIK3CA*, *PTEN* and *AKT1* mutations ( $P = 0.025$ ). Mutation of splicing factor *SF3B1*, previously described in myelodysplastic syndromes<sup>21</sup> and chronic lymphocytic leukaemia<sup>22</sup>, was significant with 15 non-silent mutations, of which 4 were a recurrent K700E substitution. Two protein tyrosine phosphatases (*PTPN22* and *PTPRD*) were also significantly mutated; frequent deletion/mutation of *PTPRD* is observed in lung adenocarcinoma<sup>23</sup>.

Mutations and mRNA-expression subtype associations

We analysed the somatic mutation spectrum within the context of the four mRNA-expression subtypes, excluding the normal-like group owing to small numbers ( $n = 8$ ) (Fig. 1). Several significantly mutated genes showed mRNA-subtype-specific (Supplementary Figs 1–3) and clinical-subtype-specific patterns of mutation (Supplementary Table 2). Significantly mutated genes were considerably more diverse and recurrent within luminal A and luminal B tumours than within basal-like and HER2-enriched (HER2E) subtypes; however, the overall mutation rate was lowest in luminal A subtype and highest in the basal-like and HER2E subtypes. The luminal A subtype harboured the most significantly mutated genes, with the most frequent being *PIK3CA* (45%), followed by *MAP3K1*, *GATA3*, *TP53*, *CDH1* and *MAP2K4*. Twelve per cent of luminal A tumours contained likely inactivating mutations in *MAP3K1* and *MAP2K4*, which represent two contiguous

steps in the p38–JNK1 stress kinase pathway<sup>24</sup>. Luminal B cancers exhibited a diversity of significantly mutated genes, with *TP53* and *PIK3CA* (29% each) being the most frequent. The luminal tumour subtypes markedly contrasted with basal-like cancers where *TP53* mutations occurred in 80% of cases and the majority of the luminal significantly mutated gene repertoire, except *PIK3CA* (9%), were absent or near absent. The HER2E subtype, which has frequent *HER2* amplification (80%), had a hybrid pattern with a high frequency of *TP53* (72%) and *PIK3CA* (39%) mutations and a much lower frequency of other significantly mutated genes including *PIK3RI* (4%).

Intrinsic mRNA subtypes differed not only by mutation frequencies but also by mutation type. Most notably, *TP53* mutations in basal-like tumours were mostly nonsense and frame shift, whereas missense mutations predominated in luminal A and B tumours (Supplementary Fig. 1). Fifty-eight somatic *GATA3* mutations, some of which were previously described<sup>25</sup>, were detected including a hotspot 2-base-pair deletion within intron 4 only in the luminal A subtype (13 out of 13 mutants) (Supplementary Fig. 2). In contrast, 7 out of 9 frame-shift mutations in exon 5 (DNA binding domain) occurred in luminal B cancers. *PIK3CA* mutation frequency and spectrum also varied by mRNA subtype (Supplementary Fig. 3); the recurrent *PIK3CA* E545K mutation was present almost exclusively within luminal A (25 out of 27) tumours. *CDH1* mutations were common (30 out of 36) within the lobular histological subtype and corresponded with lower *CDH1* mRNA (Supplementary Fig. 4) and protein expression. Finally, we identified 4 out of 8 somatic variants in *HER2* within lobular cancers, three of which were within the tyrosine kinase domain.

We performed analyses on a selected set of genes<sup>26</sup> using the normal tissue DNA data and detected a number of germline predisposing variants. These analyses identified 47 out of 507 patients with deleterious germline variants, representing nine different genes (*ATM*, *BRCA1*, *BRCA2*, *BRIP1*, *CHEK2*, *NBN*, *PTEN*, *RAD51C* and *TP53*; Supplementary Table 3), supporting the hypothesis that ~10% of sporadic breast cancers may have a strong germline contribution.

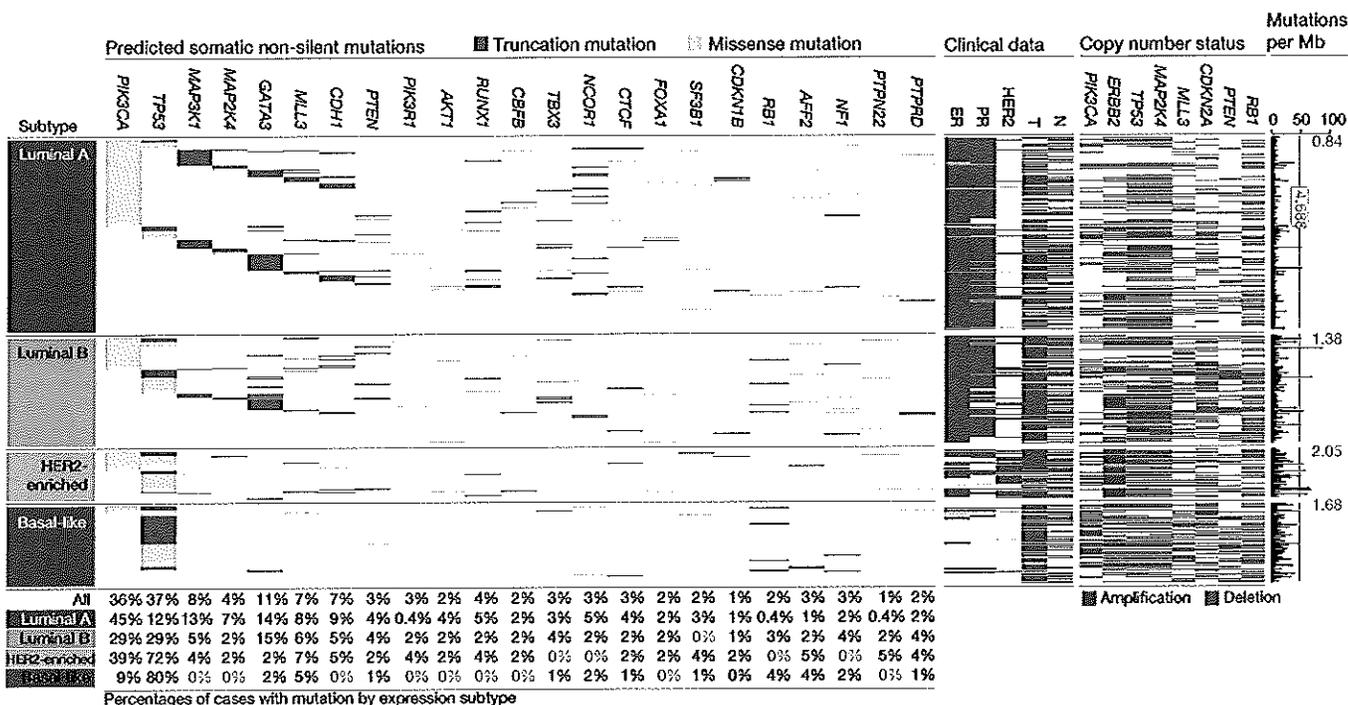


Figure 1 | Significantly mutated genes and correlations with genomic and clinical features. Tumour samples are grouped by mRNA subtype: luminal A ( $n = 225$ ), luminal B ( $n = 126$ ), HER2E ( $n = 57$ ) and basal-like ( $n = 93$ ). The left panel shows non-silent somatic mutation patterns and frequencies for significantly mutated genes. The middle panel shows clinical features: dark grey, positive or T2–4; white, negative or T1; light grey, N/A or equivocal.

N, node status; T, tumour size. The right panel shows significantly mutated genes with frequent copy number amplifications (red) or deletions (blue). The far-right panel shows non-silent mutation patterns and frequencies for significantly mutated genes. The average mutation rate for each expression subtype is indicated. Hypermutated: mutation rates > 3 s.d. above the mean (>4.688, indicated by grey line).

These data confirmed the association between the presence of germline *BRCA1* mutations and basal-like breast cancers<sup>7,8</sup>.

### Gene expression analyses (mRNA and miRNA)

Several approaches were used to look for structure in the mRNA expression data. We performed an unsupervised hierarchical clustering analysis of 525 tumours and 22 tumour-adjacent normal tissues using the top 3,662 variably expressed genes (Supplementary Fig. 5); SigClust analysis identified 12 classes (5 classes with >9 samples per class). We performed a semi-supervised hierarchical cluster analysis using a previously published 'intrinsic gene list'<sup>14</sup>, which identified 13 classes (9 classes with >9 samples per class) (Supplementary Fig. 6). We also classified each sample using the 50-gene PAM50 model<sup>14</sup> (Supplementary Fig. 5). High concordance was observed between all three analyses; therefore, we used the PAM50-defined subtype predictor as a common classification metric. There were only eight normal-like and eight claudin-low tumours<sup>27</sup>, thus we did not perform focussed analyses on these two subtypes.

MicroRNA expression levels were assayed via Illumina sequencing, using 1,222 miRBase<sup>28</sup> v16 mature and star strands as the reference database of miRNA transcripts/genes. Seven subtypes were identified by consensus non-negative matrix factorization (NMF) clustering using an abundance matrix containing the 25% most variable miRNAs (306 transcripts/genes or MIMATs (miRNA IDs)). These subtypes correlated with mRNA subtypes, ER, PR and HER2 clinical status (Supplementary Fig. 7). Of note, miRNA groups 4 and 5 showed high overlap with the basal-like mRNA subtype and contained many *TP53* mutations. The remaining miRNA groups (1–3, 6 and 7) were composed of a mixture of luminal A, luminal B and HER2E with little correlation with the PAM50 defined subtypes. With the exception of *TP53*—which showed a strong positive correlation—and *PIK3CA* and *GATA3*—which showed negative associations with groups 4 and 5, respectively—there was little correlation with mutation status and miRNA subtype.

### DNA methylation

Illumina Infinium DNA methylation arrays were used to assay 802 breast tumours. Data from HumanMethylation27 (HM27) and HumanMethylation450 (HM450) arrays were combined and filtered to yield a common set of 574 probes used in an unsupervised clustering analysis, which identified five distinct DNA methylation groups (Supplementary Fig. 8). Group 3 showed a hypermethylated phenotype and was significantly enriched for luminal B mRNA subtype and under-represented for *PIK3CA*, *MAP3K1* and *MAP2K4* mutations. Group 5 showed the lowest levels of DNA methylation, overlapped with the basal-like mRNA subtype, and showed a high frequency of *TP53* mutations. HER2-positive (HER2<sup>+</sup>) clinical status, or the HER2E mRNA subtype, had only a modest association with the methylation subtypes.

A supervised analysis of the DNA methylation and mRNA expression data was performed to compare DNA methylation group 3 ( $N = 49$ ) versus all tumours in groups 1, 2 and 4 (excluding group 5, which consisted predominantly of basal-like tumours). This analysis identified 4,283 genes differentially methylated (3,735 higher in group 3 tumours) and 1,899 genes differentially expressed (1,232 downregulated); 490 genes were both methylated and showed lower expression in group 3 tumours (Supplementary Table 4). A DAVID (database for annotation, visualization and integrated discovery) functional annotation analysis identified 'extracellular region part' and 'Wnt signalling pathway' to be associated with this 490-gene set; the group 3 hypermethylated samples showed fewer *PIK3CA* and *MAP3K1* mutations, and lower expression of Wnt-pathway genes.

### DNA copy number

A total of 773 breast tumours were assayed using Affymetrix 6.0 SNP arrays. Segmentation analysis and GISTIC were used to

identify focal amplifications/deletions and arm-level gains and losses (Supplementary Table 5). These analyses confirmed all previously reported copy number variations and highlighted a number of significantly mutated genes including focal amplification of regions containing *PIK3CA*, *EGFR*, *FOXA1* and *HER2*, as well as focal deletions of regions containing *MLL3*, *PTEN*, *RBI* and *MAP2K4* (Supplementary Fig. 9); in all cases, multiple genes were included within each altered region. Importantly, many of these copy number changes correlated with mRNA subtype including characteristic loss of 5q and gain of 10p in basal-like cancers<sup>5,29</sup> and gain of 1q and/or 16q loss in luminal tumours<sup>4</sup>. NMF clustering of GISTIC segments identified five copy number clusters/groups that correlated with mRNA subtypes, ER, PR and HER2 clinical status, and *TP53* mutation status (Supplementary Fig. 10). In addition, this aCGH subtype classification was highly correlated with the aCGH subtypes recently defined by ref. 30 (Supplementary Fig. 11).

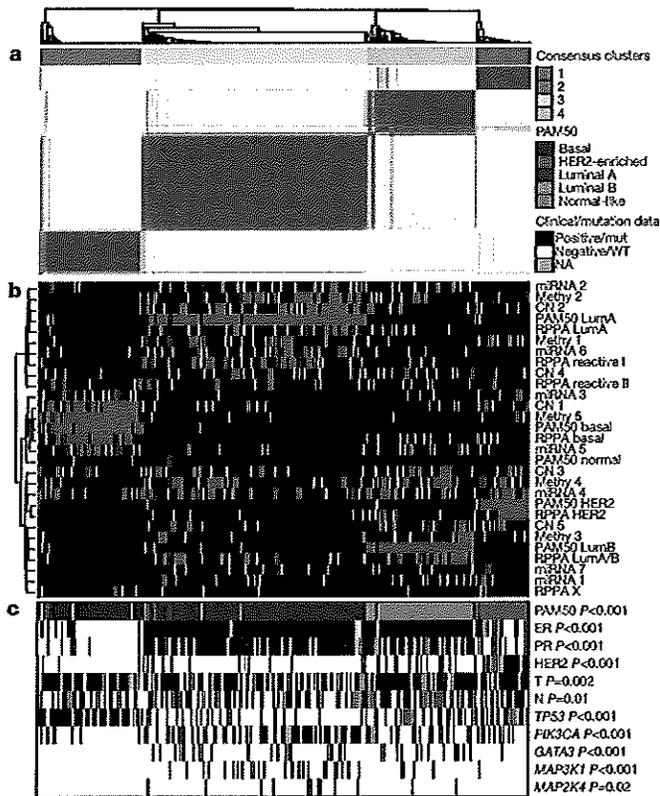
### Reverse phase protein arrays

Quantified expression of 171 cancer-related proteins and phospho-proteins by RPPA was performed on 403 breast tumours<sup>31</sup>. Unsupervised hierarchical clustering analyses identified seven subtypes; one class contained too few cases for further analysis (Supplementary Fig. 12). These protein subtypes were highly concordant with the mRNA subtypes, particularly with basal-like and HER2E mRNA subtypes. Closer examination of the HER2-containing RPPA-defined subgroup showed coordinated overexpression of HER2 and EGFR with a strong concordance with phosphorylated HER2 (pY1248) and EGFR (pY992), probably from heterodimerization and cross-phosphorylation. Although there is a potential for modest cross reactivity of antibodies against these related total and phospho-proteins, the concordance of phosphorylation of HER2 and EGFR was confirmed using multiple independent antibodies.

In RPPA-defined luminal tumours, there was high protein expression of ER, PR, AR, BCL2, *GATA3* and INPP4B, defining mostly luminal A cancers and a second more heterogeneous protein subgroup composed of both luminal A and luminal B cancers. Two potentially novel protein-defined subgroups were identified: reactive I consisted primarily of a subset of luminal A tumours, whereas reactive II consisted of a mixture of mRNA subtypes. These groups are termed 'reactive' because many of the characteristic proteins are probably produced by the microenvironment and/or cancer-activated fibroblasts including fibronectin, caveolin 1 and collagen VI. These two RPPA groups did not have a marked difference in the percentage tumour cell content when compared to each other, or the other protein subtypes, as assessed by SNP array analysis or pathological examination. In addition, supervised analyses of reactive I versus II groups using miRNA expression, DNA methylation, mutation, or DNA copy number data identified no significant differences between these groups, whereas similar supervised analyses using protein and mRNA expression identified many differences.

### Multiplatform subtype discovery

To reveal higher-order structure in breast tumours based on multiple data types, significant clusters/subtypes from each of five platforms were analysed using a multiplatform data matrix subjected to unsupervised consensus clustering (Fig. 2). This 'cluster of clusters' (C-of-C) approach illustrated that basal-like cancers had the most distinct multiplatform signature as all the different platforms for the basal-like groups clustered together. To a great extent, the four major C-of-C subdivisions correlated well with the previously published mRNA subtypes (driven, in part, by the fact that the four intrinsic subtypes were one of the inputs). Therefore, we also performed C-of-C analysis with no mRNA data present (Supplementary Fig. 13) or with the 12 unsupervised mRNA subtypes (Supplementary Fig. 14), and in each case 4–6 groups were identified. Recent work identified ten copy-number-based subgroups in a 997 breast cancer set<sup>30</sup>. We evaluated



**Figure 2** | Coordinated analysis of breast cancer subtypes defined from five different genomic/proteomic platforms. **a**, Consensus clustering analysis of the subtypes identifies four major groups (samples,  $n = 348$ ). The blue and white heat map displays sample consensus. **b**, Heat-map display of the subtypes defined independently by miRNAs, DNA methylation, copy number (CN), PAM50 mRNA expression, and RPPA expression. The red bar indicates membership of a cluster type. **c**, Associations with molecular and clinical features.  $P$  values were calculated using a chi-squared test.

this classification in a C-of-C analysis instead of our five-class copy number subtypes, with either the PAM50 (Supplementary Fig. 15) or 12 unsupervised mRNA subtypes (Supplementary Fig. 16); each of these C-of-C classifications was highly correlated with PAM50 mRNA subtypes and with the other C-of-C analyses (Fig. 2). The transcriptional profiling and RPPA platforms demonstrated a high correlation with the consensus structure, indicating that the information content from copy number aberrations, miRNAs and methylation is captured at the level of gene expression and protein function.

### Luminal/ER<sup>+</sup> summary analysis

Luminal/ER<sup>+</sup> breast cancers are the most heterogeneous in terms of gene expression (Supplementary Fig. 5), mutation spectrum (Fig. 1), copy number changes (Supplementary Fig. 9) and patient outcomes<sup>1,14</sup>. One of the most dominant features is high mRNA and protein expression of the luminal expression signature (Supplementary Fig. 5), which contains *ESR1*, *GATA3*, *FOXA1*, *XBPI* and *MYB*; the luminal/ER<sup>+</sup> cluster also contained the largest number of significantly mutated genes. Most notably, *GATA3* and *FOXA1* were mutated in a mutually exclusive fashion, whereas *ESR1* and *XBPI* were typically highly expressed but infrequently mutated. Mutations in *RUNX1* and its dimerization partner *CBFβ* may also have a role in aberrant ER signalling in luminal tumours, as *RUNX1* functions as an ER DNA tethering factor<sup>32</sup>. PARADIGM<sup>33</sup> analysis comparing luminal versus basal-like cancers further emphasized the presence of a hyperactivated FOXA1-ER complex as a critical network hub differentiating these two tumour subtypes (Supplementary Fig. 17).

A confirmatory finding here was the high mutation frequency of *PIK3CA* in luminal/ER<sup>+</sup> breast cancers<sup>34,35</sup>. Through multiple

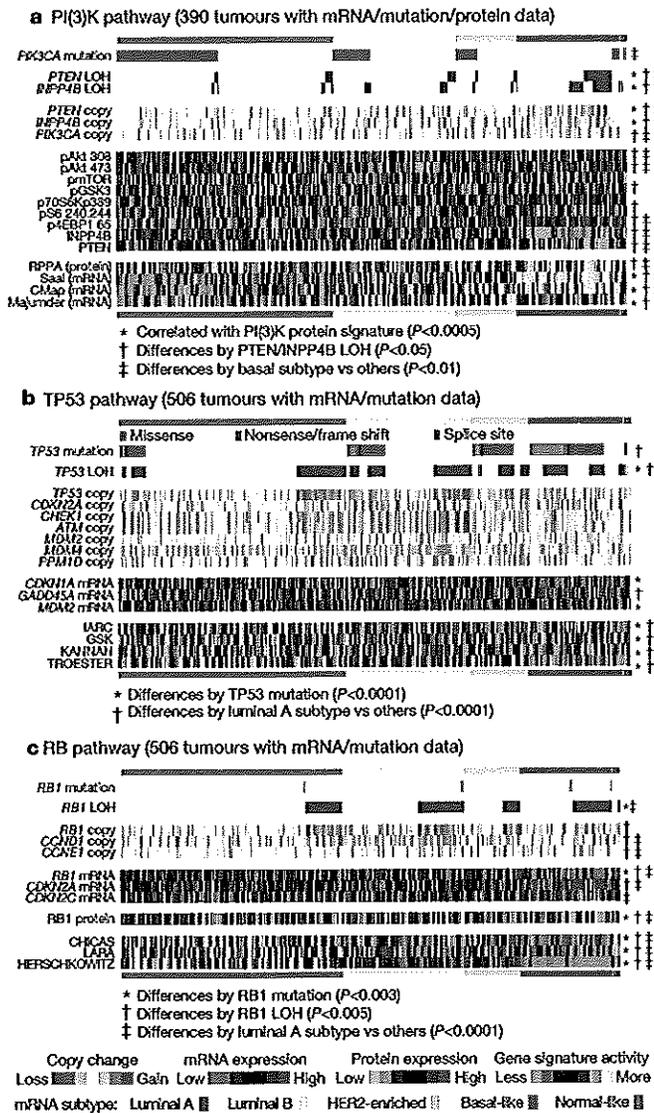
technology platforms, we examined possible relationships between *PIK3CA* mutation, *PTEN* loss, *INPP4B* loss and multiple gene and protein expression signatures of pathway activity. RPPA data demonstrated that pAKT, pS6 and p4EBP1, typical markers of phosphatidylinositol-3-OH kinase (PI(3)K) pathway activation, were not elevated in *PIK3CA*-mutated luminal A cancers; instead, they were highly expressed in basal-like and HER2E mRNA subtypes (the latter having frequent *PIK3CA* mutations) and correlated strongly with *INPP4B* and *PTEN* loss, and to a degree with *PIK3CA* amplification. Similarly, protein<sup>36</sup> and three mRNA signatures<sup>37-39</sup> of PI(3)K pathway activation were enriched in basal-like over luminal A cancers (Fig. 3a). This apparent disconnect between the presence of *PIK3CA* mutations and biomarkers of pathway activation has been previously noted<sup>36</sup>.

Another striking luminal/ER<sup>+</sup> subtype finding was the frequent mutation of *MAP3K1* and *MAP2K4*, which represent two contiguous steps within the p38-JNK1 pathway<sup>24,40</sup>. These mutations are predicted to be inactivating, with *MAP2K4* also a target of focal DNA loss in luminal tumours (Supplementary Fig. 9). To explore the possible interplay between *PIK3CA*, *MAP3K* and *MAP2K4* signalling, MEMo analysis<sup>41</sup> was performed to identify mutually exclusive alterations targeting frequently altered genes likely to belong to the same pathway (Fig. 4). Across all breast cancers, MEMo identified a set of modules that highlight the differential activation events within the receptor tyrosine kinase (RTK)-PI(3)K pathway (Fig. 4a); mutations of *PIK3CA* were very common in luminal/ER<sup>+</sup> cancers whereas *PTEN* loss was more common in basal-like tumours. Almost all *MAP3K1* and *MAP2K4* mutations were in luminal tumours, yet *MAP3K1* and *MAP2K4* appeared almost mutually exclusive relative to one another.

The TP53 pathway was differentially inactivated in luminal/ER<sup>+</sup> breast cancers, with a low *TP53* mutation frequency in luminal A (12%) and a higher frequency in luminal B (29%) cancers (Fig. 1). In addition to *TP53* itself, a number of other pathway-inactivating events occurred including *ATM* loss and *MDM2* amplification (Figs 3b and 4b), both of which occurred more frequently within luminal B cancers. Gene expression analysis demonstrated that individual markers of functional TP53 (*GADD45A* and *CDKN1A*), and TP53 activity<sup>42,43</sup> signatures, were highest in luminal A cancers (Fig. 3b). These data indicate that the TP53 pathway remains largely intact in luminal A cancers but is often inactivated in the more aggressive luminal B cancers<sup>44</sup>. Other PARADIGM-based pathway differences driving luminal B versus luminal A included hyperactivation of transcriptional activity associated with *MYC* and *FOXM1* proliferation.

The critical retinoblastoma/RB1 pathway also showed mRNA-subtype-specific alterations (Fig. 3c). RB1 itself, by mRNA and protein expression, was detectable in most luminal cancers, with highest levels within luminal A. A common oncogenic event was cyclin D1 amplification and high expression, which preferentially occurred within luminal tumours, and more specifically within luminal B. In contrast, the presumed tumour suppressor *CDKN2C* (also called *p18*) was at its lowest levels in luminal A cancers, consistent with observations in mouse models<sup>45</sup>. Finally, RB1 activity signatures were also high in luminal cancers<sup>46-48</sup>. Luminal A tumours, which have the best prognosis, are the most likely to retain activity of the major tumour suppressors RB1 and TP53.

These genomic characterizations also provided clues for druggable targets. We compiled a drug target table in which we defined a target as a gene/protein for which there is an approved or investigational drug in human clinical trials targeting the molecule or canonical pathway (Supplementary Table 6). In luminal/ER<sup>+</sup> cancers, the high frequency of *PIK3CA* mutations suggests that inhibitors of this activated kinase or its signalling pathway may be beneficial. Other potential significantly mutated gene drug candidates include AKT1 inhibitors (11 out of 12 *AKT1* variants were luminal) and PARP inhibitors for *BRCA1/BRCA2* mutations. Although still unapproved as biomarkers, many potential copy-number-based drug targets



**Figure 3 | Integrated analysis of the PI(3)K, TP53 and RB1 pathways.** Breast cancer subtypes differ by genetic and genomic targeting events, with corresponding effects on pathway activity. a–c, For PI(3)K (a), TP53 (b) and RB1 (c) pathways, key genes were selected using prior biological knowledge. Multiple mRNA expression signatures for a given pathway were defined (details in Supplementary Methods); PI(3)K: Saal, PTEN loss in human breast tumours; CMap, PI(3)K/mTOR inhibitor treatment *in vitro*; Majumder, Akt overexpression in mouse model; TP53: IARC, expert-curated p53 targets; GSK, TP53 mutant versus wild-type cell lines; KANNAN, TP53 overexpression *in vitro*; TROESTER, TP53 knockdown *in vitro*; RB: CHICAS, RB1 mouse knockdown versus wild type; LARA, RB1 knockdown *in vitro*; HERSCHKOWITZ, RB1 loss of heterozygosity (LOH) in human breast tumours) and applied to the gene expression data, in order to score each tumour for relative signature activity (yellow, more active). The PI(3)K panel includes a protein-based (RPPA) proteomic signature. Tumours were ordered first by mRNA subtype, although specific ordering differs between the panels.  $P$  values were calculated by a Pearson's correlation or a Chi-squared test.

were identified including amplifications of fibroblast growth factor receptors (FGFRs) and *IGF1R*, as well as cyclin D1, *CDK4* and *CDK6*. A summary of the general findings in luminal tumours and the other subtypes is presented in Table 1.

### HER2-based classifications and summary analysis

DNA amplification of *HER2* was readily evident in this study (Supplementary Fig. 9) together with overexpression of multiple

HER2-amplicon-associated genes that in part define the HER2E mRNA subtype (Supplementary Fig. 5). However, not all clinically HER2<sup>+</sup> tumours are of the HER2E mRNA subtype, and not all tumours in the HER2E mRNA subtype are clinically HER2<sup>+</sup>. Integrated analysis of the RPPA and mRNA data clearly identified a HER2<sup>+</sup> group (Supplementary Fig. 12). When the HER2<sup>+</sup> protein and HER2E mRNA subtypes overlapped, a strong signal of EGFR, pEGFR, HER2 and pHER2 was observed. However, only ~50% of clinically HER2<sup>+</sup> tumours fall into this HER2E-mRNA-subtype/HER2-protein group, the rest of the clinically HER2<sup>+</sup> tumours were observed predominantly in the luminal mRNA subtypes.

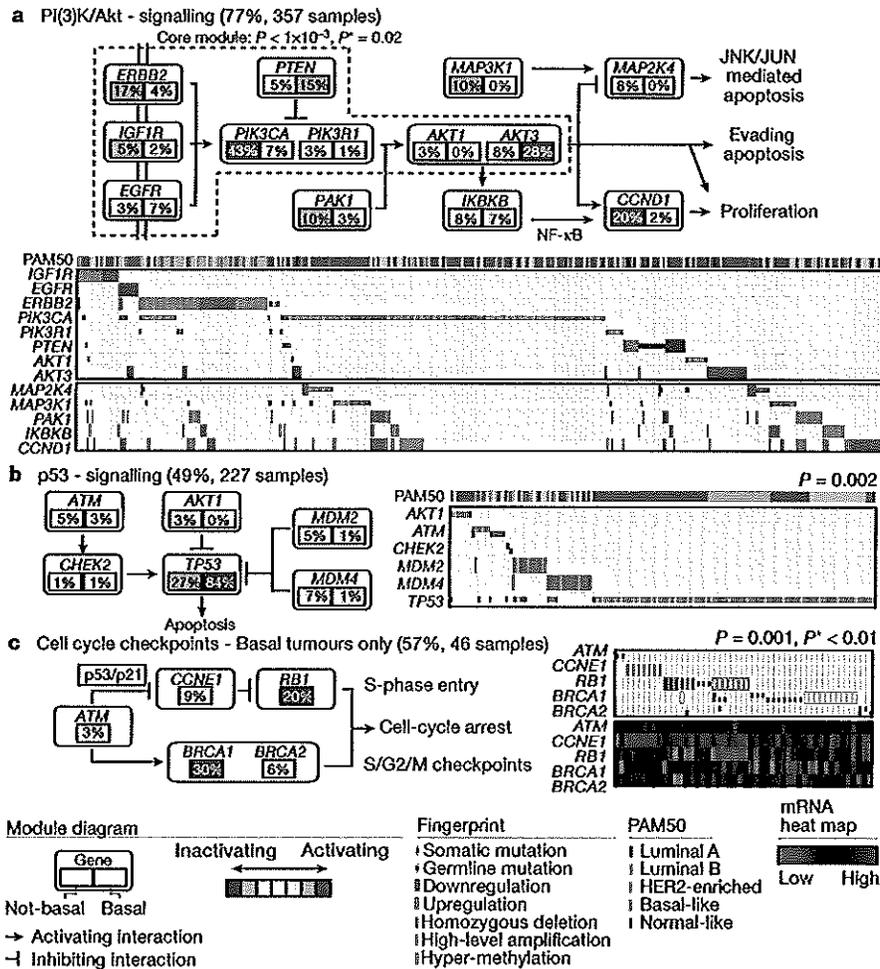
These data indicate that there exist at least two types of clinically defined HER2<sup>+</sup> tumours. To identify differences between these groups, a supervised gene expression analysis comparing 36 HER2E-mRNA-subtype/HER2<sup>+</sup> versus 31 luminal-mRNA-subtype/HER2<sup>+</sup> tumours was performed and identified 302 differentially expressed genes ( $q$ -value = 0%) (Supplementary Fig. 18 and Supplementary Table 7). These genes largely track with ER status but also indicated that HER2E-mRNA-subtype/HER2<sup>+</sup> tumours showed significantly higher expression of a number of RTKs including *FGFR4*, *EGFR*, *HER2* itself, as well as genes within the HER2 amplicon (including *GRB7*). Conversely, the luminal-mRNA-subtype/HER2<sup>+</sup> tumours showed higher expression of the luminal cluster of genes including *GATA3*, *BCL2* and *ESR1*. Further support for two types of clinically defined HER2<sup>+</sup> disease was evident in the somatic mutation data supervised by either mRNA subtype or ER status; TP53 mutations were significantly enriched in HER2E or ER-negative tumours whereas *GATA3* mutations were only observed in luminal subtypes or ER<sup>+</sup> tumours.

Analysis of the RPPA data according to mRNA subtype identified 36 differentially expressed proteins ( $q$ -value < 5%) (Supplementary Fig. 18G and Supplementary Table 8). The EGFR/pEGFR/HER2/pHER2 signal was again observed and present within the HER2E-mRNA-subtype/HER2<sup>+</sup> tumours, as was high pSRC and pS6; conversely, many protein markers of luminal cancers again distinguished the luminal-mRNA-subtype/HER2<sup>+</sup> tumours. Given the importance of clinical HER2 status, a more focused analysis was performed based on the RPPA-defined protein expression of HER2 (Supplementary Fig. 19)—the results strongly recapitulated findings from the RPPA and mRNA subtypes including a high correlation between HER2 clinical status, HER2 protein by RPPA, pHER2, EGFR and pEGFR. These multiple signatures, namely HER2E mRNA subtype, HER2 amplicon genes by mRNA expression, and RPPA EGFR/pEGFR/HER2/pHER2 signature, ultimately identify at least two groups/subtypes within clinically HER2<sup>+</sup> tumours (Table 1). These signatures represent breast cancer biomarker(s) that could potentially predict response to anti-HER2 targeted therapies.

Many therapeutic advances have been made for clinically HER2<sup>+</sup> disease. This study has identified additional somatic mutations that represent potential therapeutic targets within this group, including a high frequency of *PIK3CA* mutations (39%), a lower frequency of *PTEN* and *PIK3R1* mutations (Supplementary Table 6), and genomic losses of *PTEN* and *INPP4B*. Other possible druggable mutations included variants within HER family members including two somatic mutations in *HER2*, two within *EGFR*, and five within *HER3*. Pertuzumab, in combination with trastuzumab, targets the HER2–HER3 heterodimer<sup>49</sup>; however, these data suggest that targeting EGFR with HER2 could also be beneficial. Finally, the HER2E mRNA subtype typically showed high aneuploidy, the highest somatic mutation rate (Table 1), and DNA amplification of other potential therapeutic targets including FGFRs, *EGFR*, *CDK4* and cyclin D1.

### Basal-like summary analysis

The basal-like subtype was discovered more than a decade ago by first-generation cDNA microarrays<sup>13</sup>. These tumours are often referred to as triple-negative breast cancers (TNBCs) because most basal-like tumours are typically negative for ER, PR and HER2.



**Figure 4 | Mutual exclusivity modules in cancer (MEMo) analysis.** Mutual exclusivity modules are represented by their gene components and connected to reflect their activity in distinct pathways. For each gene, the frequency of alteration in basal-like (right box) and non-basal (left box) is reported. Next to each module is a fingerprint indicating what specific alteration is observed for each gene (row) in each sample (column). a, MEMo identified several overlapping modules that recapitulate the RTK–PI(3)K and p38–JNK1

signalling pathways and whose core was the top-scoring module. b, MEMo identified alterations to TP53 signalling as occurring within a statistically significant mutually exclusive trend. c, A basal-like only MEMo analysis identified one module that included ATM mutations, defects at BRCA1 and BRCA2, and deregulation of the RB1 pathway. A gene expression heat map is below the fingerprint to show expression levels.

However, ~75% of TNBCs are basal-like with the other 25% comprised of all other mRNA subtypes<sup>6</sup>. In this data set, there was a high degree of overlap between these two distinctions with 76 TNBCs, 81 basal-like, and 65 that were both TNBCs and basal-like. Given the known heterogeneity of TNBCs, and that the basal-like subtype proved to be distinct on every platform, we chose to use the basal-like distinction for comparative analyses.

Basal-like tumours showed a high frequency of TP53 mutations (80%)<sup>9</sup>, which when combined with inferred TP53 pathway activity suggests that loss of TP53 function occurs within most, if not all, basal-like cancers (Fig. 3b). In addition to loss of TP53, a MEMo analysis reconfirmed that loss of RB1 and BRCA1 are basal-like features (Fig. 4c)<sup>17,50</sup>. PIK3CA was the next most commonly mutated gene (~9%); however, inferred PI(3)K pathway activity, whether from gene<sup>37–39</sup>, protein<sup>36</sup>, or high PI(3)K/AKT pathway activities, was highest in basal-like cancers (Fig. 3a). Alternative means of activating the PI(3)K pathway in basal-like cancers probably includes loss of PTEN and INPP4B and/or amplification of PIK3CA. A recent paper<sup>12</sup> performed exome sequencing of 102 TNBCs. Five of the top six most frequent TNBC mutations in ref. 12 were also observed at a similar frequency in our TNBC subset (Myo3A not present here); of those five, three passed our test as a significantly mutated gene in TNBCs (Supplementary Table 2).

Expression features of basal-like tumours include a characteristic signature containing keratins 5, 6 and 17 and high expression of genes associated with cell proliferation (Supplementary Fig. 5). A PARADIGM<sup>33</sup> analysis of basal-like versus luminal tumours emphasized the importance of hyperactivated FOXM1 as a transcriptional driver of this enhanced proliferation signature (Supplementary Fig. 17). PARADIGM also identified hyperactivated MYC and HIF1- $\alpha$ /ARNT network hubs as key regulatory features of basal-like cancers. Even though chromosome 8q24 is amplified across all subtypes (Supplementary Fig. 9), high MYC activation seems to be a basal-like characteristic<sup>51</sup>.

Given the striking contrasts between basal-like and luminal/HER2E subtypes, we performed a MEMo analysis on basal-like tumours alone. The top-scoring module included ATM mutations, BRCA1 and BRCA2 inactivation, RB1 loss and cyclin E1 amplification (Fig. 4c). Notably, these same modules were identified previously for serous ovarian cancers<sup>41</sup>. Furthermore, the basal-like (and TNBC) mutation spectrum was reminiscent of the spectrum seen in serous ovarian cancers<sup>52</sup> with only one gene (that is, TP53) at >10% mutation frequency. To explore possible similarities between serous ovarian and the breast basal-like cancers, we performed a number of analyses comparing ovarian versus breast luminal, ovarian versus breast basal-like, and breast basal-like versus breast luminal cancers

**Table 1 | Highlights of genomic, clinical and proteomic features of subtypes**

Subtype	Luminal A	Luminal B	Basal-like	HER2E
ER <sup>+</sup> /HER2 <sup>-</sup> (%)	87	82	10	20
HER2 <sup>+</sup> (%)	7	15	2	68
TNBCs (%)	2	1	80	9
TP53 pathway	TP53 mut (12%); gain of <i>MDM2</i> (14%)	TP53 mut (32%); gain of <i>MDM2</i> (31%)	TP53 mut (84%); gain of <i>MDM2</i> (14%)	TP53 mut (75%); gain of <i>MDM2</i> (30%)
PIK3CA/PTEN pathway	PIK3CA mut (49%); <i>PTEN</i> mut/loss (13%); <i>INPP4B</i> loss (9%)	PIK3CA mut (32%); <i>PTEN</i> mut/loss (24%); <i>INPP4B</i> loss (16%)	PIK3CA mut (7%); <i>PTEN</i> mut/loss (35%); <i>INPP4B</i> loss (30%)	PIK3CA mut (42%); <i>PTEN</i> mut/loss (19%); <i>INPP4B</i> loss (30%)
RB1 pathway	Cyclin D1 amp (29%); <i>CDK4</i> gain (14%); low expression of <i>CDKN2C</i> ; high expression of <i>RB1</i>	Cyclin D1 amp (58%); <i>CDK4</i> gain (25%)	<i>RB1</i> mut/loss (20%); cyclin E1 amp (9%); high expression of <i>CDKN2A</i> ; low expression of <i>RB1</i>	Cyclin D1 amp (38%); <i>CDK4</i> gain (24%)
mRNA expression	High ER cluster; low proliferation	Lower ER cluster; high proliferation	Basal signature; high proliferation	HER2 amplicon signature; high proliferation
Copy number	Most diploid; many with quiet genomes; 1q, 8q, 8p11 gain; 8p, 16q loss; 11q13.3 amp (24%)	Most aneuploid; many with focal amp; 1q, 8q, 8p11 gain; 8p, 16q loss; 11q13.3 amp (51%); 8p11.23 amp (28%)	Most aneuploid; high genomic instability; 1q, 10p gain; 8p, 5q loss; <i>MYC</i> focal gain (40%)	Most aneuploid; high genomic instability; 1q, 8q gain; 8p loss; 17q12 focal <i>ERBB2</i> amp (71%)
DNA mutations	PIK3CA (49%); TP53 (12%); GATA3 (14%); MAP3K1 (14%)	TP53 (32%); PIK3CA (32%); MAP3K1 (5%)	TP53 (84%); PIK3CA (7%)	TP53 (75%); PIK3CA (42%); PIK3R1 (8%)
DNA methylation	—	Hypermethylated phenotype for subset	Hypomethylated	—
Protein expression	High oestrogen signalling; high MYB; RPPA reactive subtypes	Less oestrogen signalling; high FOXM1 and MYC; RPPA reactive subtypes	High expression of DNA repair proteins, <i>PTEN</i> and <i>INPP4B</i> loss signature (pAKT)	High protein and phospho-protein expression of EGFR and HER2

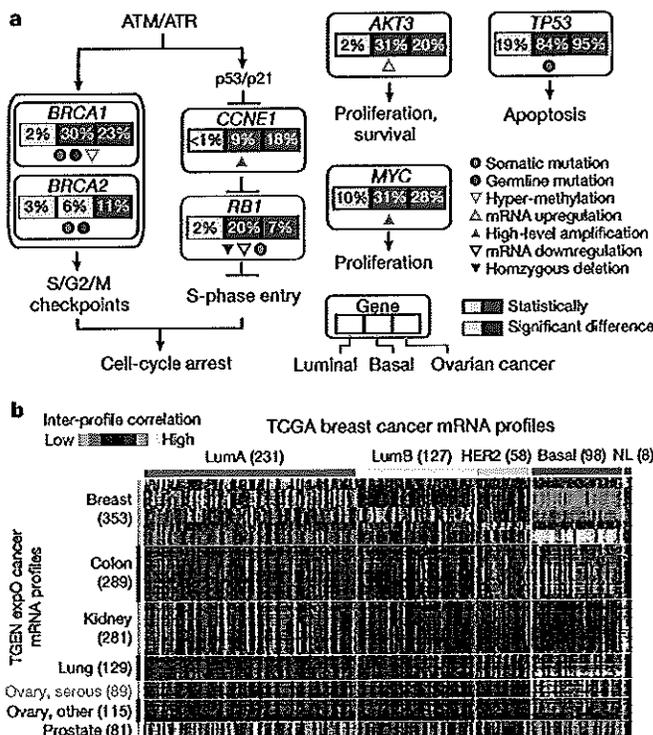
Percentages are based on 466 tumour overlap list. Amp, amplification; mut, mutation.

(Fig. 5). Comparing copy number landscapes, we observed several common features between ovarian and basal-like tumours including widespread genomic instability and common gains of 1q, 3q, 8q and 12p, and loss of 4q, 5q and 8p (Supplementary Fig. 20A). Using a more global copy number comparison, we examined the overall fraction of the genome altered and the overall copy number correlation of ovarian cancers versus each breast cancer mRNA subtype

(Supplementary Fig. 20A, B); in both cases, basal-like tumours were the most similar to the serous ovarian carcinomas.

We systematically looked for other common features between serous ovarian and basal-like tumours when each was compared to luminal. We identified: (1) *BRCA1* inactivation; (2) *RB1* loss and cyclin E1 amplification; (3) high expression of *AKT3*; (4) *MYC* amplification and high expression; and (5) a high frequency of *TP53* mutations (Fig. 5a). An additional supervised analysis of a large, external multitumour type transcriptomic data set (Gene Expression Omnibus accession GSE2109) was performed where each TCGA (The Cancer Genome Atlas) breast tumour expression profile was compared via a correlation analysis to that of each tumour in the multitumour set. Basal-like breast cancers clearly showed high mRNA expression correlations with serous ovarian cancers, as well as with lung squamous carcinomas (Fig. 5b). A PARADIGM analysis that calculates whether a gene or pathway feature is both differentially activated in basal-like versus luminal cancers and has higher overall activity across the TCGA ovarian samples was performed; this identified comparably high pathway activity of the HIF1- $\alpha$ /ARNT, *MYC* and *FOXM1* regulatory hubs in both ovarian and basal-like cancers (Supplementary Fig. 20C). The common findings of *TP53*, *RB1* and *BRCA1* loss, with *MYC* amplification, strongly suggest that these are shared driving events for basal-like and serous ovarian carcinogenesis. This suggests that common therapeutic approaches should be considered, which is supported by the activity of platinum analogues and taxanes in breast basal-like and serous ovarian cancers.

Given that most basal-like cancers are TNBCs, finding new drug targets for this group is critical. Unfortunately, the somatic mutation repertoire for basal-like breast cancers has not provided a common target aside from *BRCA1* and *BRCA2*. Here we note that ~20% of basal-like tumours had a germline ( $n = 12$ ) and/or somatic ( $n = 8$ ) *BRCA1* or *BRCA2* variant, which suggests that one in five basal-like patients might benefit from PARP inhibitors and/or platinum compounds<sup>53,54</sup>. The copy number landscape of basal-like cancers showed multiple amplifications and deletions, some of which may provide therapeutic targets (Supplementary Table 6). Potential targets include losses of *PTEN* and *INPP4B*, both of which have been shown to sensitize cell lines to PI(3)K pathway inhibitors<sup>55,56</sup>. Interestingly, many of the components of the PI(3)K and RAS-RAF-MEK pathway were amplified (but not typically mutated) in basal-like cancers including *PIK3CA* (49%), *KRAS* (32%), *BRAF* (30%) and *EGFR* (23%). Other RTKs that are plausible drug targets and amplified in



**Figure 5 | Comparison of breast and serous ovarian carcinomas.** a, Significantly enriched genomic alterations identified by comparing basal-like or serous ovarian tumours to luminal cancers. b, Inter-sample correlations (yellow, positive) between gene transcription profiles of breast tumours (columns; TCGA data, arranged by subtype) and profiles of cancers from various tissues of origin (rows; external 'TGEN expO' data set, GSE2109) including ovarian cancers.

some basal-like cancers include *FGFR1*, *FGFR2*, *IGF1R*, *KIT*, *MET* and *PDGFRA*. Finally, the PARADIGM identification of high HIF1- $\alpha$ /ARNT pathway activity suggests that these malignancies might be susceptible to angiogenesis inhibitors and/or bioreductive drugs that become activated under hypoxic conditions.

### Concluding remarks

The integrated molecular analyses of breast carcinomas that we report here significantly extends our knowledge base to produce a comprehensive catalogue of likely genomic drivers of the most common breast cancer subtypes (Table 1). Our novel observation that diverse genetic and epigenetic alterations converge phenotypically into four main breast cancer classes is not only consistent with convergent evolution of gene circuits, as seen across multiple organisms, but also with models of breast cancer clonal expansion and *in vivo* cell selection proposed to explain the phenotypic heterogeneity observed within defined breast cancer subtypes.

### METHODS SUMMARY

Specimens were obtained from patients with appropriate consent from institutional review boards. Using a co-isolation protocol, DNA and RNA were purified. In total, 800 patients were assayed on at least one platform. Different numbers of patients were used for each platform using the largest number of patients available at the time of data freeze; 466 samples (463 patients) were in common across 5 out of 6 platforms (excluding RPPA) and 348 patients were in common on 6 out of 6 platforms. Technology platforms used include: (1) gene expression DNA microarrays<sup>35</sup>; (2) DNA methylation arrays; (3) miRNA sequencing; (4) Affymetrix SNP arrays; (5) exome sequencing; and (6) reverse phase protein arrays. Each platform, except for the exome sequencing, was used in a *de novo* subtype discovery analysis (Supplementary Methods) and then included in a single analysis to define an overall subtype architecture. Additional integrated across-platform computational analyses were performed including PARADIGM<sup>33</sup> and MEMo<sup>41</sup>.

Received 22 March; accepted 11 July 2012.

Published online 23 September 2012.

- Paik, S. *et al.* A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N. Engl. J. Med.* **351**, 2817–2826 (2004).
- van 't Veer, L. J. *et al.* Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415**, 530–536 (2002).
- Slamon, D. J. *et al.* Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science* **235**, 177–182 (1987).
- Chin, K. *et al.* Genomic and transcriptional aberrations linked to breast cancer pathophysiology. *Cancer Cell* **10**, 529–541 (2006).
- Bergamaschi, A. *et al.* Distinct patterns of DNA copy number alteration are associated with different clinicopathological features and gene-expression subtypes of breast cancer. *Genes Chromosomes Cancer* **45**, 1033–1040 (2006).
- Perou, C. M. Molecular stratification of triple-negative breast cancers. *Oncologist* **16** (suppl. 1), 61–70 (2011).
- Sortie, T. *et al.* Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc. Natl Acad. Sci. USA* **100**, 8418–8423 (2003).
- Foulkes, W. D. *et al.* Germline BRCA1 mutations and a basal epithelial phenotype in breast cancer. *J. Natl Cancer Inst.* **95**, 1482–1485 (2003).
- Carey, L. A. *et al.* Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. *J. Am. Med. Assoc.* **295**, 2492–2502 (2006).
- Ding, L. *et al.* Genome remodelling in a basal-like breast cancer metastasis and xenograft. *Nature* **464**, 999–1005 (2010).
- Shah, S. P. *et al.* Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature* **461**, 809–813 (2009).
- Shah, S. P. *et al.* The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature* **486**, 395–399 (2012).
- Perou, C. M. *et al.* Molecular portraits of human breast tumours. *Nature* **406**, 747–752 (2000).
- Parker, J. S. *et al.* Supervised risk predictor of breast cancer based on intrinsic subtypes. *J. Clin. Oncol.* **27**, 1160–1167 (2009).
- González-Pérez, A. & Lopez-Bigas, N. Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score. *Condel. Am. J. Hum. Genet.* **88**, 440–449 (2011).
- Dees, N. D. *et al.* MuSiC: Identifying mutational significance in cancer genomes. *Genome Res.* **22**, 1589–1598 (2012).
- Bamshad, M. *et al.* Mutations in human TBX3 alter limb, apocrine and genital development in ulnar-mammary syndrome. *Nature Genet.* **16**, 311–315 (1997).
- Li, Q. Y. *et al.* Holt-Oram syndrome is caused by mutations in TBX5, a member of the Brachyury (T) gene family. *Nature Genet.* **15**, 21–29 (1997).
- The Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **455**, 1061–1068 (2008).
- Cheung, L. W. *et al.* High frequency of PIK3R1 and PIK3R2 mutations in endometrial cancer elucidates a novel mechanism for regulation of PTEN protein stability. *Cancer Discov.* **1**, 170–185 (2011).
- Malcovati, L. *et al.* Clinical significance of SF3B1 mutations in myelodysplastic syndromes and myelodysplastic/myeloproliferative neoplasms. *Blood* **118**, 6239–6246 (2011).
- Wang, L. *et al.* SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. *N. Engl. J. Med.* **365**, 2497–2506 (2011).
- Ding, L. *et al.* Somatic mutations affect key pathways in lung adenocarcinoma. *Nature* **455**, 1069–1075 (2008).
- Johnson, G. L. & Lapadat, R. Mitogen-activated protein kinase pathways mediated by ERK, JNK, and p38 protein kinases. *Science* **298**, 1911–1912 (2002).
- Usary, J. *et al.* Mutation of GATA3 in human breast tumors. *Oncogene* **23**, 7669–7678 (2004).
- Walsh, T. *et al.* Detection of inherited mutations for breast and ovarian cancer using genomic capture and massively parallel sequencing. *Proc. Natl Acad. Sci. USA* **107**, 12629–12633 (2010).
- Prat, A. *et al.* Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. *Breast Cancer Res.* **12**, R68 (2010).
- Kozomara, A. & Griffiths-Jones, S. miRBase: Integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res.* **39**, D152–D157 (2011).
- Weigman, V. J. *et al.* Basal-like breast cancer DNA copy number losses identify genes involved in genomic instability, response to therapy, and patient survival. *Breast Cancer Res. Treat.* **133**, 865–880 (2011).
- Curtis, C. *et al.* The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* **486**, 346–352 (2012).
- Hennessy, B. T. *et al.* A technical assessment of the utility of reverse phase protein arrays for the study of the functional proteome in non-microdissected human breast cancers. *Clin. Proteomics* **6**, 129–151 (2010).
- Daub, H. *et al.* Kinase-selective enrichment enables quantitative phosphoproteomics of the kinome across the cell cycle. *Mol. Cell* **31**, 438–448 (2008).
- Vaske, C. J. *et al.* Inference of patient-specific pathway activities from multi-dimensional cancer genomics data using PARADIGM. *Bioinformatics* **26**, 237–245 (2010).
- Campbell, I. G. *et al.* Mutation of the PIK3CA gene in ovarian and breast cancer. *Cancer Res.* **64**, 7678–7681 (2004).
- Bachman, K. E. *et al.* The PIK3CA gene is mutated with high frequency in human breast cancers. *Cancer Biol. Ther.* **3**, 772–775 (2004).
- Stemke-Hale, K. *et al.* An integrative genomic and proteomic analysis of PIK3CA, PTEN, and AKT mutations in breast cancer. *Cancer Res.* **68**, 6084–6091 (2008).
- Creighton, C. J. *et al.* Proteomic and transcriptomic profiling reveals a link between the PI3K pathway and lower estrogen-receptor (ER) levels and activity in ER<sup>+</sup> breast cancer. *Breast Cancer Res.* **12**, R40 (2010).
- Majumder, P. K. *et al.* mTOR inhibition reverses Akt-dependent prostate intraepithelial neoplasia through regulation of apoptotic and HIF-1-dependent pathways. *Nature Med.* **10**, 594–601 (2004).
- Saal, L. H. *et al.* Recurrent gross mutations of the PTEN tumor suppressor gene in breast cancers with deficient DSB repair. *Nature Genet.* **40**, 102–107 (2008).
- Wagner, E. F. & Nebreda, A. R. Signal integration by JNK and p38 MAPK pathways in cancer development. *Nature Rev. Cancer* **9**, 537–549 (2009).
- Ciriello, G., Cerami, E., Sander, C. & Schultz, N. Mutual exclusivity analysis identifies oncogenic network modules. *Genome Res.* **22**, 398–406 (2012).
- Kannan, K. *et al.* DNA microarrays identification of primary and secondary target genes regulated by p53. *Oncogene* **20**, 2225–2234 (2001).
- Troester, M. A. *et al.* Gene expression patterns associated with p53 status in breast cancer. *BMC Cancer* **6**, 276 (2006).
- Deisenroth, C., Thorne, A. R., Enomoto, T., Perou, C. M. & Zhang, Y. Mitochondrial Hep27 is a c-Myc target gene that inhibits Mdm2 and stabilizes p53. *Mol. Cell Biol.* **30**, 3981–3993 (2010).
- Pei, X. H. *et al.* CDK inhibitor p18<sup>INK4c</sup> is a downstream target of GATA3 and restrains mammary luminal progenitor cell proliferation and tumorigenesis. *Cancer Cell* **15**, 389–401 (2009).
- Chicas, A. *et al.* Dissecting the unique role of the retinoblastoma tumor suppressor during cellular senescence. *Cancer Cell* **17**, 376–387 (2010).
- Herschkovitz, J. I., He, X., Fan, C. & Perou, C. M. The functional loss of the retinoblastoma tumour suppressor is a common event in basal-like and luminal B breast carcinomas. *Breast Cancer Res.* **10**, R75 (2008).
- Lara, M. F. *et al.* Gene profiling approaches help to define the specific functions of retinoblastoma family in epidermis. *Mol. Carcinog.* **47**, 209–221 (2008).
- Baselga, J. *et al.* Pertuzumab plus trastuzumab plus docetaxel for metastatic breast cancer. *N. Engl. J. Med.* **366**, 109–119 (2012).
- Jiang, Z. *et al.* Rb deletion in mouse mammary progenitors induces luminal-B or basal-like/EMT tumor subtypes depending on p53 status. *J. Clin. Invest.* **120**, 3296–3309 (2010).
- Chandriani, S. *et al.* A core MYC gene expression signature is prominent in basal-like breast cancer but only partially overlaps the core serum response. *PLoS ONE* **4**, e6693 (2009).
- The Cancer Genome Atlas Research Network. Integrated genomic analyses of ovarian carcinoma. *Nature* **474**, 609–615 (2011).
- Audeh, M. W. *et al.* Oral poly(ADP-ribose) polymerase inhibitor olaparib in patients with BRCA1 or BRCA2 mutations and recurrent ovarian cancer: a proof-of-concept trial. *Lancet* **376**, 245–251 (2010).
- Fong, P. C. *et al.* Inhibition of poly(ADP-ribose) polymerase in tumors from BRCA mutation carriers. *N. Engl. J. Med.* **361**, 123–134 (2009).

55. Fedele, C. G. *et al.* Inositol polyphosphate 4-phosphatase II regulates PI3K/Akt signaling and is lost in human basal-like breast cancers. *Proc. Natl Acad. Sci. USA* **107**, 22231–22236 (2010).
56. Gewinner, C. *et al.* Evidence that inositol polyphosphate 4-phosphatase type II is a tumor suppressor that inhibits PI3K signaling. *Cancer Cell* **16**, 115–125 (2009).

Supplementary Information is available in the online version of the paper.

**Acknowledgements** We thank M. Sheth and S. Lucas for administrative coordination of TCGA activities, and C. Gunter for critical reading of the manuscript. This work was supported by the following grants from the USA National Institutes of Health: U24CA143883, U24CA143858, U24CA143840, U24CA143799, U24CA143835, U24CA143845, U24CA143882, U24CA143867, U24CA143866, U24CA143848, U24CA144025, U54HG003079, P50CA116201 and P50CA58223. Additional support was provided by the Susan G. Komen for the Cure, the US Department of Defense through the Henry M. Jackson Foundation for the Advancement of Military Medicine, and the Breast Cancer Research Foundation. The views expressed in this paper are those of the authors and do not reflect the official policy of the Department of Defense, or US Government.

**Author Contributions** TCGA research network contributed collectively to this study. Biospecimens were provided by tissue source sites and processed by a Biospecimens Core Resource. Data generation and analyses were performed by genome sequencing centres, cancer genome characterization centres, and genome data analysis centres. RPPA analysis was performed at the MD Anderson Cancer Center in association with the genome data analysis centre. All data were released through the Data Coordinating Center. Project activities were coordinated by NCI and NHGRI project teams. We also acknowledge the following TCGA investigators of the Breast Analysis Working Group who contributed substantially to the analysis and writing of this manuscript: Project leaders, C.M.P., M.J.E.; manuscript coordinator, C.M.P., K.A.H.; data coordinator, K.A.H.; analysis coordinator, C.M.P., K.A.H.; DNA sequence analysis, D.C.K., L.D.; mRNA microarray analysis, K.A.H., C.F.; miRNA sequence analysis, A.G.R., A.C.; DNA methylation analysis, S. Malik, S. Mahurkar, P.W.L.; copy number analysis, A.D.C., M.M.; protein analysis, W.L., R.G.W.Y., G.B.M.; pathway/integrated analysis, C.J.C., C.Y., J.M.S., C.C.B., G.C., C.S., S.R., I.S.; bioinformatics core resource, T.L., J.B., J.M.G.; pathology and clinical expertise, T.A.K., H.H., R.J.M., J.N.J., T.S., F.W.

**Author Information** All of the primary sequence files are deposited in CGHub (<https://cghub.ucsc.edu/>); all other data including mutation annotation file are deposited at the Data Coordinating Center (<http://cancergenome.nih.gov/>). Sample lists, data matrices and supporting data can be found at [http://tcga-data.nci.nih.gov/docs/publications/brca\\_2012/](http://tcga-data.nci.nih.gov/docs/publications/brca_2012/). The data can be explored via the ICB Regulome Explorer (<http://explorer.canceregulome.org/>) and the cBio Cancer Genomics Portal (<http://cbioportal.org/>). Data descriptions can be found at <https://wiki.nci.nih.gov/display/TCGA/TCGA+Data+Primer+and+in+Supplementary+Methods>. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). This paper is distributed under the terms of the Creative Commons Attribution-Non-Commercial-Share Alike licence, and the online version of this paper is freely available to all readers. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.M.P. ([cprou@med.unc.edu](mailto:cprou@med.unc.edu)).

## The Cancer Genome Atlas Network

**Genome sequencing centres:** Washington University in St Louis Daniel C. Koboldt<sup>1</sup>, Robert S. Fulton<sup>1</sup>, Michael D. McLellan<sup>1</sup>, Heather Schmidt<sup>1</sup>, Joelle Kalicki-Weizer<sup>1</sup>, Joshua F. McMichael<sup>1</sup>, Lucinda L. Fulton<sup>1</sup>, David J. Dooling<sup>1</sup>, Li Ding<sup>1,2</sup>, Elaine R. Mardis<sup>1,2,3</sup>, Richard K. Wilson<sup>1,2,3</sup>

**Genome characterization centres:** BC Cancer Agency Adrian Ally<sup>4</sup>, Miruna Balasundaram<sup>4</sup>, Yaron S. N. Butterfield<sup>4</sup>, Rebecca Carlsen<sup>4</sup>, Candace Carter<sup>4</sup>, Andy Chu<sup>4</sup>, Eric Chuah<sup>4</sup>, Hye-Jung E. Chun<sup>4</sup>, Robin J. N. Coope<sup>4</sup>, Noreen Dhalla<sup>4</sup>, Ranabir Guin<sup>4</sup>, Carrie Hirst<sup>4</sup>, Martin Hirst<sup>4</sup>, Robert A. Holt<sup>4</sup>, Darlene Lee<sup>4</sup>, Haiyan L. Li<sup>4</sup>, Michael Mayo<sup>4</sup>, Richard A. Moore<sup>4</sup>, Andrew J. Mungall<sup>4</sup>, Erin Pleasance<sup>4</sup>, A. Gordon Robertson<sup>4</sup>, Jacqueline E. Schein<sup>4</sup>, Arash Shafiei<sup>4</sup>, Payal Sipahimalani<sup>4</sup>, Jared R. Slobodan<sup>4</sup>, Dominik Stoll<sup>4</sup>, Angela Tam<sup>4</sup>, Nina Thiessen<sup>4</sup>, Richard J. Varhol<sup>4</sup>, Natasja Wye<sup>4</sup>, Thomas Zeng<sup>4</sup>, Yongjun Zhao<sup>4</sup>, Inanc Birol<sup>4</sup>, Steven J. M. Jones<sup>4</sup>, Marco A. Marra<sup>4</sup>; Broad Institute Andrew D. Cherniack<sup>5</sup>, Gordon Saksena<sup>5</sup>, Robert C. Onofrio<sup>5</sup>, Nam H. Pho<sup>5</sup>, Scott L. Carter<sup>5</sup>, Steven E. Schumacher<sup>5,6</sup>, Barbara Tabak<sup>5,6</sup>, Bryan Hernandez<sup>5</sup>, Jeff Gentry<sup>5</sup>, Huy Nguyen<sup>5</sup>, Andrew Crenshaw<sup>5</sup>, Kristin Ardlie<sup>5</sup>, Rameen Beroukhi<sup>5,7,8</sup>, Wendy Winckler<sup>5</sup>, Gad Getz<sup>5</sup>, Stacey B. Gabriel<sup>5</sup>, Matthew Meyerson<sup>5,9,10</sup>, Brigham & Women's Hospital & Harvard Medical School Lynda Chin<sup>9,11</sup>, Peter J. Park<sup>12</sup>, Raju Kuchertapati<sup>13</sup>, University of North Carolina, Chapel Hill Katherine A. Hoadley<sup>14,15</sup>, J. Todd Auman<sup>14,17</sup>, Cheng Fan<sup>15</sup>, Yidi J. Turman<sup>15</sup>, Yan Shi<sup>15</sup>, Ling Li<sup>15</sup>, Michael D. Topal<sup>15,18</sup>, Xiaping He<sup>14,15</sup>, Hann-Hsiang Chao<sup>14,15</sup>, Aleix Prat<sup>14,15</sup>, Grace O. Silva<sup>14,15</sup>, Michael D. Iglesia<sup>14,15</sup>, Wei Zhao<sup>14,15</sup>, Jerry Usary<sup>15</sup>, Jonathan S. Berg<sup>14,15</sup>, Michael Adams<sup>14</sup>, Jessica Brooker<sup>18</sup>, Junyuan Wu<sup>15</sup>, Anisha Gulabani<sup>15</sup>, Tom Bodenheimer<sup>15</sup>, Alan P. Hoyle<sup>15</sup>, Janae V. Simons<sup>15</sup>, Matthew G. Soloway<sup>15</sup>, Lisle E. Mose<sup>15</sup>, Stuart R. Jefferys<sup>15</sup>, Saianand Balu<sup>15</sup>, Joel S. Parker<sup>15</sup>, D. Neil Hayes<sup>15,19</sup>, Charles M. Perou<sup>14,15,18</sup>, University of Southern California/Johns Hopkins Simeen Malik<sup>20</sup>, Swapna Mahurkar<sup>20</sup>, Hui Shen<sup>20</sup>, Daniel J. Weisenberger<sup>20</sup>, Timothy Triche Jr<sup>20</sup>, Phillip H. Lai<sup>20</sup>, Moiz S. Bootwalla<sup>20</sup>, Dennis T. Maglinte<sup>20</sup>, Benjamin P. Berman<sup>20</sup>, David J. Van Den Berg<sup>20</sup>, Stephen B. Bayliss<sup>21</sup>, Peter W. Laird<sup>20</sup>

**Genome data analysis:** Baylor College of Medicine Chad J. Creighton<sup>22,23</sup>, Lawrence A. Donehower<sup>22,23,24,25</sup>, Broad Institute Gad Getz<sup>26</sup>, Michael Noble<sup>26</sup>, Doug Voegele<sup>26</sup>, Gordon Saksena<sup>26</sup>, Nils Gehlenborg<sup>12,26</sup>, Daniel DiCara<sup>26</sup>, Junhua Zhang<sup>27</sup>, Hailie Zhang<sup>26</sup>, Chang-Jun Wu<sup>28</sup>, Spring Yingshun Liu<sup>26</sup>, Michael S. Lawrence<sup>26</sup>, Lihua Zou<sup>26</sup>, Andrey Sivachenko<sup>26</sup>, Pei Lin<sup>26</sup>, Petar Stojanov<sup>26</sup>, Rui Jing<sup>26</sup>, Juok Cho<sup>26</sup>, Raktim Sinha<sup>26</sup>, Richard W. Park<sup>26</sup>, Marc-Danle Nazaire<sup>26</sup>, Jim Robinson<sup>26</sup>, Helga Thorvaldsdottir<sup>26</sup>, Jill Mesirov<sup>26</sup>, Peter J. Park<sup>12,29,30</sup>, Lynda Chin<sup>31,32</sup>, Institute for Systems Biology Sheila Reynolds<sup>31</sup>, Richard B. Kreisberg<sup>31</sup>, Brady Bernard<sup>31</sup>, Ryan Bressler<sup>31</sup>, Timo Erkkila<sup>32</sup>, Jake Lin<sup>31</sup>, Vesteinn Thorsson<sup>31</sup>, Wei Zhang<sup>33</sup>, Ilya Shmulevich<sup>31</sup>, Memorial Sloan-Kettering Cancer Center Giovanni Ciriello<sup>34</sup>, Nils Weinhold<sup>34</sup>, Nikolaus Schultz<sup>34</sup>, Jianjiong Gao<sup>34</sup>, Ethan Cerami<sup>34</sup>, Benjamin Gross<sup>34</sup>, Anders Jacobsen<sup>34</sup>, Rileen Sinha<sup>34</sup>, B. Arman Aksoy<sup>34</sup>, Yevgeniy Antipin<sup>34</sup>, Boris Reva<sup>34</sup>, Ronglai Shen<sup>35</sup>, Barry S. Taylor<sup>34</sup>, Marc Ladanyi<sup>36</sup>, Chris Sander<sup>34</sup>; Oregon Health & Science University Pavana Anur<sup>37</sup>, Paul T. Spellman<sup>37</sup>; The University of Texas MD Anderson Cancer Center Ying Lu<sup>38,39</sup>, Wenbin Liu<sup>40</sup>, Roel R. G. Verhaak<sup>40</sup>, Gordon B. Mills<sup>38,39</sup>, Rehan Akbani<sup>40</sup>, Nianxiang Zhang<sup>40</sup>, Bradley M. Broom<sup>40</sup>, Tod D. Casasent<sup>40</sup>, Chris Wakefield<sup>40</sup>, Anna K. Unruh<sup>40</sup>, Keith Baggerly<sup>40</sup>, Kevin Coombes<sup>40</sup>, John N. Weinstein<sup>40</sup>; University of California, Santa Cruz/Buck Institute David Haussler<sup>41,42</sup>, Christopher C. Benz<sup>43</sup>, Joshua M. Stuart<sup>41</sup>, Stephen C. Benz<sup>41</sup>, Jingchun Zhu<sup>41</sup>, Christopher C. Szeto<sup>41</sup>, Gary K. Scott<sup>43</sup>, Christina Yau<sup>43</sup>, Evan O. Paull<sup>41</sup>, Daniel Carlini<sup>41</sup>, Christopher Wong<sup>41</sup>, Artem Sokolov<sup>41</sup>, Janita Thusbarg<sup>43</sup>, Sean Mooney<sup>43</sup>, Sam Ng<sup>41</sup>, Theodora C. Goldstein<sup>41</sup>, Kyle Ellrott<sup>41</sup>, Mia Griffoard<sup>41</sup>, Christopher Wilks<sup>41</sup>, Singer Ma<sup>41</sup>, Brian Craft<sup>41</sup>; NCI Chunhua Yan<sup>44</sup>, Ying Hu<sup>44</sup>, Daoud Meerzaman<sup>44</sup>

**Biospecimen core resource:** Nationwide Children's Hospital Biospecimen Core Resource Julie M. Gastier-Foster<sup>45,46,47</sup>, Jay Bowen<sup>47</sup>, Nilsa C. Ramirez<sup>45,47</sup>, Aaron D. Black<sup>47</sup>, Robert E. Pyatt<sup>45,47</sup>, Peter White<sup>46,47</sup>, Erik J. Zmuda<sup>47</sup>, Jessica Frick<sup>47</sup>, Tara M. Lichtenberg<sup>47</sup>, Robin Brookers<sup>47</sup>, Myra M. George<sup>47</sup>, Mark A. Gerken<sup>47</sup>, Hollie A. Harper<sup>47</sup>, Kristen M. Leraas<sup>47</sup>, Lisa J. Wise<sup>47</sup>, Teresa R. Tabler<sup>47</sup>, Cynthia McAllister<sup>47</sup>, Thomas Barr<sup>47</sup>, Melissa Hart-Kochari<sup>47</sup>

**Tissue source sites:** ABS-IUPUI Katie Tarvin<sup>48</sup>, Charles Saller<sup>49</sup>, George Sandusky<sup>50</sup>, Colleen Mitchell<sup>50</sup>, Christiana Mary V. Iacocca<sup>51</sup>, Jennifer Brown<sup>51</sup>, Brenda Rabeno<sup>51</sup>, Christine Czerniewski<sup>51</sup>, Nicholas Petrelli<sup>51</sup>, Celine Oleg Dolzhansky<sup>52</sup>, Mikhail Abramov<sup>53</sup>, Olga Voronina<sup>54</sup>, Olga Potapova<sup>54</sup>; Duke University Medical Center Jeffrey R. Marks<sup>55</sup>; The Greater Poland Cancer Centre Wiktorija M. Suchorska<sup>56</sup>, Dawid Murawa<sup>56</sup>, Witold Kycler<sup>56</sup>, Matthew Ibbes<sup>56</sup>, Konstanty Korski<sup>56</sup>, Arkadiusz Spychala<sup>56</sup>, Pawel Murawa<sup>56</sup>, Jacek J. Brzeziński<sup>56</sup>, Hanna Perz<sup>56</sup>, Radoslaw Lazniak<sup>56</sup>, Marek Teresiak<sup>56</sup>, Honorata Tatka<sup>56</sup>, Ewa Leporowska<sup>56</sup>, Marta Bogusz-Czerniewicz<sup>56,57</sup>, Julian Malicki<sup>56,57</sup>, Andrzej Mackiewicz<sup>56,57</sup>, Maciej Wiznerowicz<sup>56,57</sup>, ILSBio Xuan Van Le<sup>58</sup>, Bernard Kohl<sup>58</sup>, Nguyen Viet Tien<sup>59</sup>, Richard Thorp<sup>60</sup>, Nguyen Van Bang<sup>61</sup>, Howard Sussman<sup>62</sup>, Bui Duc Phu<sup>61</sup>, Richard Hajek<sup>63</sup>, Nguyen Phi Hung<sup>64</sup>, Tran Viet The Phuong<sup>65</sup>, Huynh Quyet Thang<sup>66</sup>, Khurram Zaki Khan<sup>60</sup>; International Genomics Consortium Robert Penny<sup>67</sup>, David Mallery<sup>67</sup>, Erin Curb<sup>68</sup>, Candace Shelton<sup>67</sup>, Peggy Yena<sup>67</sup>; Mayo Clinic James N. Ingle<sup>69</sup>, Fergus J. Couch<sup>68</sup>, Wilma L. Lingle<sup>68</sup>, MSKCC Tari A. King<sup>69</sup>; MD Anderson Cancer Center Ana Maria Gonzalez-Angulo<sup>38,70</sup>, Gordon B. Mills<sup>70</sup>, Mary D. Dyer<sup>70</sup>, Shuying Liu<sup>70</sup>, Xiaolong Meng<sup>70</sup>, Modesto Patangan<sup>70</sup>; University of California San Francisco Frederic Waldman<sup>71,72</sup>, Hubert Stoppier<sup>73</sup>; University of North Carolina W. Kimryn Rathmell<sup>15</sup>, Leigh Thorne<sup>15,74</sup>, Mei Huang<sup>15,74</sup>, Lori Boice<sup>15,74</sup>, Ashley Hill<sup>15</sup>; Roswell Park Cancer Institute Carl Morrison<sup>75</sup>, Carmelo Gaudioso<sup>75</sup>, Wiam Bshara<sup>75</sup>; University of Miami Kelly Daily<sup>76</sup>, Sophie C. Egea<sup>76</sup>, Mark D. Pegram<sup>76</sup>, Carmen Gomez-Fernandez<sup>76</sup>; University of Pittsburgh Rajiv Dhir<sup>77</sup>, Rohit Bhargava<sup>78</sup>, Adam Brufsky<sup>78</sup>, Walter Reed National Military Medical Center Craig D. Shriver<sup>79</sup>, Jeffrey A. Hooke<sup>79</sup>, Jamie Leigh Campbell<sup>79</sup>, Richard J. Mural<sup>80</sup>, Hai Hu<sup>80</sup>, Stella Somari<sup>80</sup>, Caroline Larson<sup>80</sup>, Brenda Deyarmin<sup>80</sup>, Leonid Kvecher<sup>80</sup>, Albert J. Kovatich<sup>81</sup>

**Disease working group:** Matthew J. Ellis<sup>3,82,83</sup>, Tari A. King<sup>69</sup>, Hai Hu<sup>80</sup>, Fergus J. Couch<sup>68</sup>, Richard J. Mural<sup>80</sup>, Thomas Stricker<sup>84</sup>, Kevin White<sup>84</sup>, Olufunmilayo Olopade<sup>85</sup>, James N. Ingle<sup>68</sup>, Chunqing Luo<sup>80</sup>, Yaqin Chen<sup>80</sup>, Jeffrey R. Marks<sup>55</sup>, Frederic Waldman<sup>71,72</sup>, Maciej Wiznerowicz<sup>56,57</sup>, Ron Bose<sup>3,82,83</sup>, Li-Wei Chang<sup>86</sup>, Andrew H. Beck<sup>10</sup>, Ana Maria Gonzalez-Angulo<sup>38,70</sup>

**Data coordination centre:** Todd Pihl<sup>87</sup>, Mark Jensen<sup>87</sup>, Robert Sfeir<sup>87</sup>, Ari Kahn<sup>87</sup>, Anna Chu<sup>87</sup>, Prachi Kothiyal<sup>87</sup>, Zhining Wang<sup>87</sup>, Eric Snyder<sup>87</sup>, Joan Pontius<sup>87</sup>, Brenda Ayala<sup>87</sup>, Mark Backus<sup>87</sup>, Jessica Walton<sup>87</sup>, Julien Babou<sup>87</sup>, Dominique Berton<sup>87</sup>, Matthew Nicholls<sup>87</sup>, Deepak Srinivasan<sup>87</sup>, Rohini Ramon<sup>87</sup>, Stanley Girshik<sup>87</sup>, Peter Kigonya<sup>87</sup>, Shelley Alonso<sup>87</sup>, Rashmi Sanbhadri<sup>87</sup>, Sean Barletta<sup>87</sup>, David Pot<sup>87</sup>

**Project team:** National Cancer Institute Margi Sheth<sup>88</sup>, John A. Demchok<sup>88</sup>, Kenna R. Mills Shaw<sup>88</sup>, Liming Yang<sup>88</sup>, Greg Eley<sup>89</sup>, Martin L. Ferguson<sup>90</sup>, Roy W. Tamuzzer<sup>88</sup>, Jiashan Zhang<sup>88</sup>, Laura A. L. Dillon<sup>88</sup>, Kenneth Buetow<sup>44</sup>, Peter Fielding<sup>88</sup>, National Human Genome Research Institute Bradley A. Ozenberger<sup>91</sup>, Mark S. Guyer<sup>91</sup>, Heidi J. Sofia<sup>91</sup>, Jacqueline D. Palchik<sup>91</sup>

<sup>1</sup>The Genome Institute, Washington University, St Louis, Missouri 63108, USA.

<sup>2</sup>Department of Genetics, Washington University, St Louis, Missouri 63110, USA. <sup>3</sup>Siteman Cancer Center, Washington University, St Louis, Missouri 63110, USA. <sup>4</sup>Canada's Michael Smith Genome Sciences Centre, BC Cancer Agency, Vancouver, British Columbia V5Z, Canada. <sup>5</sup>The Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA.

<sup>6</sup>Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA. <sup>7</sup>Department of Medicine, Harvard Medical School, Boston, Massachusetts 02215, USA. <sup>8</sup>Departments of Cancer Biology and Medical Oncology, and the Center for

Cancer Genome Discovery, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA. <sup>9</sup>Department of Medical Oncology and the Center for Cancer Genome Discovery, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA. <sup>10</sup>Department of Pathology, Harvard Medical School, Boston, Massachusetts 02115, USA. <sup>11</sup>Belfer Institute for Applied Cancer Science, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA. <sup>12</sup>The Center for Biomedical Informatics, Harvard Medical School, Boston, Massachusetts 02115, USA. <sup>13</sup>Department of Genetics, Harvard Medical School and Division of Genetics, Brigham and Women's Hospital, Boston, Massachusetts 02115, USA. <sup>14</sup>Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA. <sup>15</sup>Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA. <sup>16</sup>Eshelman School of Pharmacy, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA. <sup>17</sup>Institute for Pharmacogenetics and Individualized Therapy, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA. <sup>18</sup>Department of Pathology and Laboratory Medicine, University of North Carolina at Chapel Hill, Chapel Hill, Chapel Hill, North Carolina 27599, USA. <sup>19</sup>Department of Internal Medicine, Division of Medical Oncology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA. <sup>20</sup>USC Epigenome Center, University of Southern California, Los Angeles, California 90033, USA. <sup>21</sup>Cancer Biology Division, The Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins University, Baltimore, Maryland 21231, USA. <sup>22</sup>Dan L. Duncan Cancer Center, Baylor College of Medicine, Houston, Texas 77030, USA. <sup>23</sup>Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas 77030, USA. <sup>24</sup>Department of Molecular and Cellular Biology, Baylor College of Medicine, Houston, Texas 77030, USA. <sup>25</sup>Department of Molecular Virology and Microbiology, Baylor College of Medicine, Houston, Texas 77030, USA. <sup>26</sup>The Eli and Edythe L. Broad Institute of Massachusetts Institute of Technology and Harvard University, Cambridge, Massachusetts 02142, USA. <sup>27</sup>Institute for Applied Cancer Science, Department of Genomic Medicine, University of Texas MD Anderson Cancer Center, Houston, Texas 77054, USA. <sup>28</sup>Department of Genomic Medicine, University of Texas MD Anderson Cancer Center, Houston, Texas 77054, USA. <sup>29</sup>Division of Genetics, Brigham and Women's Hospital, Boston, Massachusetts 02115, USA. <sup>30</sup>Informatics Program, Children's Hospital, Boston, Massachusetts 02115, USA. <sup>31</sup>Institute for Systems Biology, Seattle, Washington 98109, USA. <sup>32</sup>Tampere University of Technology, Tampere, Finland. <sup>33</sup>Cancer Genomics Core Laboratory, MD Anderson Cancer Center, Houston, Texas 77030, USA. <sup>34</sup>Computational Biology Center, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA. <sup>35</sup>Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA. <sup>36</sup>Human Oncology and Pathogenesis Program, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA. <sup>37</sup>Oregon Health and Science University, 3181 Southwest Sam Jackson Park Road, Portland, Oregon 97239, USA. <sup>38</sup>Department of Systems Biology, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030, USA. <sup>39</sup>Kleberg Center for Molecular Markers, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030, USA. <sup>40</sup>Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030, USA. <sup>41</sup>Department of Biomolecular Engineering and Center for Biomolecular Science and Engineering, University of California Santa Cruz, Santa Cruz, California 95064, USA. <sup>42</sup>Howard Hughes Medical Institute, University of California Santa Cruz, Santa Cruz, California 95064, USA. <sup>43</sup>Buck Institute for Research on Aging, Novato, California 94945, USA. <sup>44</sup>Center for Bioinformatics and Information Technology, National Cancer Institute, Rockville, Maryland 20852, USA. <sup>45</sup>The Ohio State University College of Medicine, Department of Pathology, Columbus, Ohio 43205, USA. <sup>46</sup>The Ohio State University College of Medicine, Department of Pediatrics, Columbus, Ohio 43205, USA. <sup>47</sup>The Research Institute at Nationwide Children's Hospital, Columbus, Ohio 43205, USA. <sup>48</sup>ABS Inc. Indianapolis, Indiana 46204, USA. <sup>49</sup>ABS Inc. Wilmington, Delaware 19801, USA. <sup>50</sup>Indiana University School of Medicine, Indianapolis, Indiana 46202, USA. <sup>51</sup>Helen F. Graham Cancer Center, Christiana Care, Newark, Delaware 19713, USA. <sup>52</sup>Moscow City Clinical Oncology Dispensary 1 and the Central IHC Laboratory of the Moscow Health Department, Moscow 105005, Russia. <sup>53</sup>Russian Cancer Research Center, Moscow 115478, Russia. <sup>54</sup>Cureline, Inc., South San Francisco, California 94080, USA. <sup>55</sup>Department of Surgery, Duke University Medical Center, Durham, North Carolina 27710, USA. <sup>56</sup>The Greater Poland Cancer Centre, Poznań 61-866, Poland. <sup>57</sup>Poznan University of Medical Sciences, Poznań 61-701, Poland. <sup>58</sup>LSBio, LLC, Chestertown, Maryland 21620, USA. <sup>59</sup>Ministry of Health, Hanoi, Vietnam. <sup>60</sup>LSBio LLC, Karachi, Pakistan. <sup>61</sup>Hue Central Hospital, Hue City, Vietnam. <sup>62</sup>Stanford University Medical Center, Stanford, California 94305, USA. <sup>63</sup>Center for Minority Health Research, University of Texas, MD Anderson Cancer Center, Houston, Texas 07703, USA. <sup>64</sup>National Cancer Institute, Hanoi, Vietnam. <sup>65</sup>Ho Chi Minh City Cancer Center, Vietnam. <sup>66</sup>Can Tho Cancer Center, Can Tho, Vietnam. <sup>67</sup>International Genomics Consortium, Phoenix, Arizona 85004, USA. <sup>68</sup>Mayo Clinic, Rochester, Minnesota 55905, USA. <sup>69</sup>Department of Surgery, Breast Service, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA. <sup>70</sup>Department of Breast Medical Oncology, The University of Texas, MD Anderson Cancer Center, Houston, Texas 77030, USA. <sup>71</sup>University of California at San Francisco; San Francisco, California 94143, USA. <sup>72</sup>Cancer Diagnostics; Nichols Institute, Quest Diagnostics; San Juan Capistrano, California 92675, USA. <sup>73</sup>Helen Diller Family Comprehensive Cancer Center, University of California San Francisco, San Francisco, California 94115, USA. <sup>74</sup>UNC Tissue Procurement Facility, Department of Pathology, UNC Lineberger Cancer Center, Chapel Hill, North Carolina 27599, USA. <sup>75</sup>Department of Pathology, Roswell Park Cancer Institute, Buffalo, New York 14263, USA. <sup>76</sup>Department of Pathology, University of Miami Miller School of Medicine, Sylvester Comprehensive Cancer Center, Miami, Florida 33136, USA. <sup>77</sup>University of Pittsburgh, Pittsburgh, Pennsylvania 15213, USA. <sup>78</sup>Magee-Womens Hospital of University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania 15213, USA. <sup>79</sup>Walter Reed National Military Medical Center, Bethesda, Maryland 20899-5600, USA. <sup>80</sup>Windber Research Institute, Windber, Pennsylvania 15963, USA. <sup>81</sup>MDR Global, LLC, Windber, Pennsylvania 15963, USA. <sup>82</sup>Breast Cancer Program, Washington University, St Louis, Missouri 63110, USA. <sup>83</sup>Department of Internal Medicine, Division of Oncology, Washington University, St Louis, Missouri 63110, USA. <sup>84</sup>Institute for Genomics and Systems Biology, University of Chicago, Chicago, Illinois 60637, USA. <sup>85</sup>Center for Clinical Cancer Genetics, The University of Chicago, Chicago, Illinois 60637, USA. <sup>86</sup>Department of Pathology and Immunology, Washington University School of Medicine, St Louis, Missouri 63110, USA. <sup>87</sup>SRA International, 4300 Fair Lakes Court, Fairfax, Virginia 22033, USA. <sup>88</sup>The Cancer Genome Atlas Program Office, Center for Cancer Genomics, National Cancer Institute, Bethesda, Maryland 20852, USA. <sup>89</sup>TCGA Consultant, Scimientis, LLC, Statham, Georgia 30666, USA. <sup>90</sup>MLF Consulting, Arlington, Massachusetts 02474, USA. <sup>91</sup>National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892, USA.

## Identification of differentially expressed genes in breast tumors from African American compared to Caucasian women

RE Ellsworth, LA Field, B Love, B Deyarmin, JA Hooke, CD Shriver

**Background:** Breast tumors from African American women (AAW) have less favorable pathological characteristics and higher mortality rates than Caucasian women. While socioeconomic status may influence prognosis, biological factors are also likely to contribute to tumor behavior.

**Methods:** Patients treated within an equal-access military medical center were matched by age, grade and ER status. Tumors were subjected to laser microdissection and RNA hybridized to HG U133A 2.0 microarrays. Data was analyzed using Partek Genomics Suite using a cutoff of  $P < 0.001$ , 1.5-fold change. Microarray results were validated by qRT-PCR.

**Results:** Twenty-six pairs of matched tumor samples were identified. Clinico-pathological factors did not differ significantly for age at diagnosis, tumor size or stage, lymph node, hormone receptor, or HER2 status. Mortality was similar between the two groups. One putative and nineteen known genes were differentially expressed between populations with five genes expressed at higher and 15 genes at lower levels in tumors from AAW; hierarchical clustering classified 23/26 AAW and 26/26 CW correctly. Differential expression was validated ( $P < 0.001$ ) for PSPHL (higher in AAW) and LTF and PSD3 (lower in AAW).

**Conclusions:** Despite matching of tumors by pathological characteristics and provision of medical care in an equal-access health care system, a molecular signature differentiating AAW from CW tumors was detected, supporting a model in which aggressive tumor phenotypes are driven by molecular differences within the tumors. These differentially expressed genes are involved in immune response, cell cycle and metastasis and thus may contribute to the poor outcome in AAW.



## Molecular characterization of breast tumor-associated adipose

Lori Field, Brenda Deyarmin, Ryan van Laar, Jeff Hooke, Craig Shriver, Rachel Ellsworth

**Background:** Having long been thought to function only as an inert energy storage depot, the role of adipose tissue in tumorigenesis has been largely ignored; however, **adipose is an active**

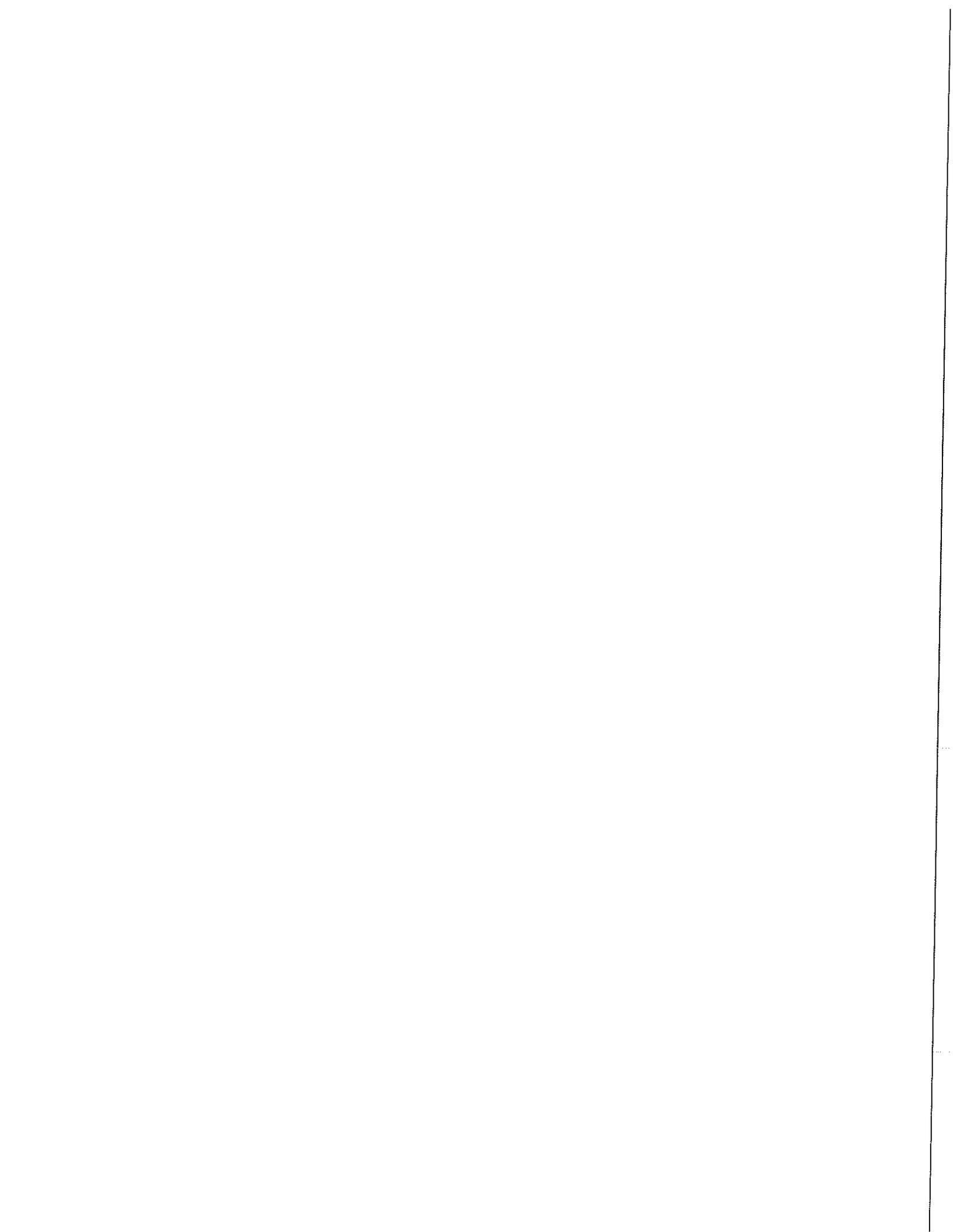
**endocrine organ that can directly influence tumor growth** Improved understanding of the role of adipose in tumorigenesis is crucial given the association between obesity and breast cancer risk in post-menopausal women, increasing rates of obesity and use of autologous fat transfer in breast reconstruction.

**Methods:** Adipose, adjacent to and distant from (>3 cm from the closest tumor margin) invasive breast tumors, was laser microdissected from 20 post-menopausal women, and from 20 post-menopausal women with non-malignant breast disease. Gene expression data were generated using U133A 2.0 microarrays. Data were analyzed to identify significant patterns of differential expression between adipose classes, at the individual gene and molecular pathway level. Gene expression differences were validated using qRT-PCR in an additional set of 30 specimens.

**Results:** SPP1, RRM2, MMP9 and PLA2G7 were expressed at >3-fold ( $P < 0.01$ ) higher levels in adjacent adipose compared to distant adipose from the same breast. A number of immune response genes including MARCO, FABP7, ELF5, MYBPC1, MMP7, CLDN8, HLA-DQB1 and HLA-DQA1 were differentially expressed in distant adipose compared to adipose from non-malignant breasts. The most significant gene expression differences were detected between tumor-adjacent and non-malignant adipose, with >3-fold higher expression of EGFL6 and ITGB2 and >3-fold lower levels of PIP, which are involved in growth, proliferation, and cellular adhesion in adjacent compared to non-malignant adipose. Pathway analysis revealed that immune response differs between non-malignant, distant and tumor adjacent adipose; with an

enhanced B- and T-cell response detected in adjacent compared to distant or non-malignant adipose. Inflammatory response as well as DNA transcription and replication pathways were differentially expressed in distant compared to non-malignant adipose.

**Conclusions:** Gene expression levels differ in breast adipose, depending on presence of and proximity to tumor cells. Adipose adjacent to the tumor demonstrated the largest immune response. Although an enhanced immune response may reflect a reaction to surgical insult from the original biopsy, genes involved in cellular proliferation, degradation of the extracellular matrix and angiogenesis differentially expressed in adjacent compared to distant or non-malignant adipose, suggest that tumor-adjacent adipose may be contributing to the growth and invasion of the primary tumor. These data thus suggest that adipose is not an inert component of the breast microenvironment but plays an active role in tumorigenesis.



**Laura M. Voegtly**, PhD, Windber Research Institute; **J. Leigh Campbell**, MS, Henry M. Jackson Foundation for the Advancement of Military Medicine; **Craig D. Shriver**, MD, Walter Reed National Military Medical Center; **Rachel E. Ellsworth**, PhD, Henry M. Jackson Foundation for the Advancement of Military Medicine

Corresponding author:

Rachel E. Ellsworth

Clinical Breast Care Project

620 Seventh Street

Windber, PA 15963

[r.ellsworth@wriwindber.org](mailto:r.ellsworth@wriwindber.org)

Phone: (814) 361-6926

FAX: (814) 46706334

Title: Molecular alterations associated with early and late breast cancer mortality

Shortened title: AI in lethal breast tumors

Disclosures: The authors have no conflicts to declare. This research was supported by a grant from the United States Department of Defense (Military Molecular Medicine Initiative MDA W81XWH-05-2-0075, Protocol 01-20006).

## **Synopsis**

Analysis of allelic imbalance (AI) in tumors from patients who died of disease are genetically distinct from those from patients with >5 year overall survival. In addition, different frequencies of AI at chromosomes 13q14 and 16q22-q24 in patients who died of disease within five years of diagnosis suggests that patients with short- and long-term mortality have distinct genetic diseases.

## **Abstract**

**Background:** Breast cancer is a heterogeneous disease and patients with similar pathologies and treatments may have different clinical outcomes. Identification of molecular alterations associated with disease outcome may improve risk assessment and treatments for aggressive breast cancer.

**Methods:** Allelic imbalance (AI) data was generated for 122 invasive breast tumors with known clinical outcome. Levels and patterns of AI were compared between patients who died of disease (DOD) and those with long-term ( $\geq 5$  years) survival (LTS) using Student t-test and chi-square analysis with a significance value of  $P < 0.05$ .

**Results:** Levels of AI were significantly higher in tumors from the 31 DOD patients (28.6%) compared to the 91 LTS patients (20.1%). AI was found significantly more frequently in DOD compared to LTS tumors at chromosomes 7q31, 8p22-p21, 13q14, 17p13.3, 17p13.1, 18q21 and 22q12.3. When patients who died within 5 years of diagnosis and those who died  $> 5$  years were evaluated separately, the frequency of AI at chromosome 13q14 was significantly higher in long-term DOD patients than short-term DOD or LTS patients and AI at chromosome 16q22-q24 was significantly lower in short-term compared to long-term DOD or LTS patients.

**Conclusion:** Tumors from DOD compared to LTS patients are marked by increased genomic instability and AI at chromosomes 7q31, 8p22-p21, 17p13.3, 17p13.1, 18q21 and 22q12.3. In addition, AI at chromosome 16q22-q24 is associated with early mortality and at chromosome 13q14 with late mortality, suggesting that patients with short- and long-term mortality may have distinct genetic diseases.

## Introduction

Although mortality rates from breast cancer have decreased since 1990, it is estimated that 39,520 women will die from breast cancer in 2011 (1). Pathological characteristics such as architecture, differentiation, size and presence of local or distant metastasis, estrogen receptor (ER), progesterone receptor (PR) and HER2 expression are all considered when predicting prognosis and determining the most effective treatment options. At the molecular level, breast tumors have additional levels of variability, with at least five breast subtypes based on patterns of differential gene expression (2;3). This pathological and molecular heterogeneity influences prognosis and treatment: in a study of 2,929 breast cancer patients, despite use of identical treatment modalities for patients with similar pathological characteristics, clinical outcomes were highly variable (4).

This complexity suggests that the current standard of care approach to medicine, which is based on averaging responses across large cohorts, is insufficient to manage breast cancer. In contrast, personalized medicine provides care and treatment based on the individual characteristics of the tumor. For example, molecular profiling has improved pathological classification of tumors; molecular tests such as MammaPrint® and *Oncotype DX*<sup>TM</sup> can provide estimates of the odds of recurrence and be used to discriminate those patients requiring aggressive treatment from those who would not benefit from cytotoxic regimens (5-7). These RNA-based signatures are, however, not 100% accurate and can be affected by confounding factors such as RNA integrity and contamination of the tumor signature by stroma (8).

Because DNA is more stable than RNA and genomic alterations are not transiently affected by physiological conditions, identification of copy number alterations in primary breast tumors may provide a more stable assessment of the underlying genetic composition of tumors with different phenotypes and behaviors. To identify chromosomal alterations associated with poor prognosis, we subjected primary breast tumors from patients who died of disease (DOD) or survived >5 years post-diagnosis to allelic imbalance (AI) analysis. Data from DOD patients was further stratified by whether patients had short- (DOD within five years of diagnosis) or long-term (DOD > five years post-diagnosis) mortality to determine whether short- and long-term mortality has different molecular characteristics.

## **Materials and Methods**

Tissue and blood samples from CBCP patients (n=149) were collected with approval from the Walter Reed Army Medical Center Human Use Committee and Institutional Review Board. All subjects enrolled in the CBCP voluntarily agreed to participate and gave written informed consent. Clinical information was collected for all CBCP samples using questionnaires designed by and administered under the auspices of the CBCP.

Paraffin-embedded primary breast tumors from 149 female patients were obtained from the Clinical Breast Care Project (CBCP) Pathology Laboratory at Walter Reed Army Medical Center (WRAMC). To ensure consistency, diagnosis of all tumor samples were made by one pathologist from hemotoxylin and eosin (H&E) stained slides; staging

was performed using guidelines defined by the *AJCC Cancer Staging Manual* seventh edition (9) and grade was assigned using the Bloom-Richardson system of classification (10;11). Estrogen receptor (ER), progesterone receptor (PR) and HER2 status were determined by immunohistochemical (IHC) analysis (MDR Global, Windber, PA); cases with HER2 scores = 2+ were then evaluated by fluorescence in situ hybridization using the PathVysion® HER-2 DNA Probe kit (Abbott Laboratories, Abbott Park, IL).

DNA was obtained from homogeneous populations of primary breast tumor cells following laser-assisted microdissection on an *ASLMD* laser microdissection system (Leica Microsystems, Wetzlar, Germany) (12). All microdissected sections were examined by the CBCP pathologist, who identified and marked regions of tumor before dissection. The integrity of multiple serial sections was established by pathological verification of the first and last sections stained with H&E. Referent DNA for the CBCP samples was obtained from blood clots using Clotspin and Puregene DNA purification kits (Qiagen, Valencia, CA).

Microsatellite markers were amplified as previously described (13). Genotypes were determined using Genetic Profiler version 2.0 software and AI detected using the formula  $(T1/T2)/(N1/N2)$  where T1 and N1 represent the peak heights of the less intense alleles and T2 and N2 represent the peak heights of the more intense alleles of the tumor and referent samples, respectively (14). Each chromosomal region was represented by two polymorphic markers, and AI was defined according to the following criteria: (1) when at least one marker for a given region showed an allelic ratio  $\leq 0.35$ , the region was considered to show AI, (2) when neither marker had an allelic ratio  $\leq 0.35$  and at least one

marker was informative, the region was considered normal, and (3) when both markers were homozygous, the region was considered uninformative.

Comparison of the clinicopathological factors and levels and patterns of AI clinical status were performed using Student's t-test and Fisher's exact tests. Comparison of survival curves was determined using a log-rank (Mantel-Cox) test (GraphPad Prism™ v. 5.04). A significance value of  $P < 0.05$  was used for all analyses.

## **Results**

### *Clinicopathological characteristics*

Of the 149 patients whose tumor DNA was subjected to AI analysis, 13 were lost to follow-up before five years and 14 died of other causes (DOC); these patients were not included in further analyses. Thirty-one patients were DOD; seven died more than five years post-diagnosis (average 82 months, range 62-90 months) while 24 died within five years of diagnosis (average 26 months, range 1-50 months). The average survival time in the 91 patients who were alive >5 years (LTS) after diagnosis was 78 months (range 60-130 months). When comparing clinicopathological factors between DOD and LTS patients, ethnicity and histological type did not differ significantly, however, tumors in DOD patients were diagnosed at a younger age (average 53.3 years in DOD and 59.2 years in LTS patients) and were significantly larger, of higher stage and grade, and more likely to have positive lymph node status and to be of the triple negative subtype (Table 1).

### *AI in DOD and LTS tumors*

Overall levels of AI were significantly higher in DOD (29%) compared to LTS (20%) tumors. When stratified by chromosomal region, AI was found significantly more frequently in DOD compared to LTS tumors at chromosomes 7q31, 8p22-p21, 13q14, 17p13.3, 17p13.1, 18q21 and 22q12.3 and more significantly in LTS compared to DOD tumors at chromosome 16q22-q24 (Table 2).

Survival curves were generated for each chromosome region. Survival analysis through 80 months supported a significant ( $P<0.05$ ) association between AI at regions 7q31, 8p22-p21, 17p13.3, 17p13.1, 18q21 and 22q12.3 and decreased survival (Figure 1).

#### *AI in short-term DOD compared to long-term DOD*

Although the sample size of patients who died of disease more than five years after diagnosis was small ( $n=7$ ), levels and patterns of AI were compared between short- and long-term DOD patients. Levels of AI did not differ significantly between short-term (27%) and long-term DOD (34%). The frequency of AI at chromosome 13q14 was significantly higher in long-term DOD (71%) compared to either short-term DOD (26%) or LTS (15%) patients. The effect of AI at chromosome 13q14 on long-term survival is seen in Figure 2 where the survival curves begin to differ significantly only after 60 months. In contrast, AI at chromosome 16q22-q24 was significantly higher in both in long-term DOD (43%) and LTS (36%) compared to short-term DOD (10%) patients. When survival analysis was performed using only patients who died within five years of diagnosis, patients with AI at chromosome 16q22-q24 had significantly improved survival compared to those with retention of chromosome 16q22-q24 (Figure 2).

## Discussion

Although breast cancer treatment regimens are based largely on tumor characteristics such as tumor size and grade, ER, PR and HER2 status, traditional pathological evaluation is imperfect in predicting outcome (4). Molecular signatures may be more accurate prognostic tools than histological evaluation; identification of intrinsic subtypes based on gene expression data separates breast tumors into different groups that have different biological characteristics, behaviors, rates of recurrence and sites of metastasis. Breast tumors can be classified by their intrinsic subtypes based on different patterns of gene expression (3), and these subtypes have been associated with differences in relapse-free and overall survival with basal-like and HER2 positive subtypes having the shortest survival times (15). Not only do intrinsic subtypes have different prognoses, but recent studies have shown that each subtype has preferential sites of metastasis: bone was the predominant site of relapse in luminal and HER2 positive tumors but was infrequent in basal-like tumors. In contrast, basal-like tumors had frequent relapse in brain, lung and distant lymph nodes (16;17).

While these data highlight the differences in behavior between different molecularly-defined groups, it is important to determine whether tumors with aggressive behavior share common genetic alterations, regardless of phenotype. Within our dataset, a set of seven chromosomal regions were altered at significantly higher frequencies in tumors from patients who died of disease, despite these tumors representing a range of pathological and molecular subtypes. Chromosomal alterations in a number of these

regions have been previously associated with decreased survival in breast cancer patients, including loss of heterozygosity of chromosome 7q31, 8p22-p21, and 17p13.3, and copy number alterations of 8p, 13q and 18q (18-21). In addition to these chromosomal changes, altered expression of genes from these regions including the MET proto-oncogene on chromosome 7q31, mitochondrial tumor suppressor 1 (MTUS1) gene on chromosome 8p22, tumor protein 53 (TP53) on 17p13.1 and B-cell CLL/lymphoma 2 (BCL2) on chromosome 18q21 have been associated with survival differences (22-25).

Two LTS patients were diagnosed with stage IV breast cancer: patient 1, diagnosed with metastasis to the liver, and patient 2, diagnosed with bone metastasis; have survived 85 months and 86 months, respectively. Although patient 1 has levels of AI similar to those in the DOD patients (29%), levels of AI were low (9%) in patient 2. Importantly, neither patient had detectable AI at 7q31, 8p22-p21, 13q14, 17p13.3, 17p13.1, 18q21 or 22q12.3. In contrast to these late-stage patients with LTS, nine patients were diagnosed with early-stage (I or II) disease but died of disease. Within these patients, the average level of AI was 37.6% (range 9-67%). All but one of these early-stage DOD patients had AI at at least one of the seven loci associated with DOD.

The genetic differences between patients who were DOD within five years diagnosis and those that died > five years may provide a model for early and late breast cancer mortality. The two groups of patients had similar overall levels and patterns of AI except at chromosomes 13q14 and 16q22-q24. Chromosomal alterations at 16q22-q24 have been associated with favorable tumor characteristics such as low-grade, higher age at diagnosis, positive ER status and the luminal A subtype, which has the most favorable

prognosis (26-28). Loss of chromosome 16q22 and decreased expression of genes from chromosome 16q22 and 16q24 have also been associated with improved survival (29;30). These data suggest metastasis promoting genes exist within these regions; loss or AI at these regions such as seen at significantly higher levels in LTS (36%) and DOD patients who died >5 years diagnosis (43%), would therefore be protective, while in the short-term DOD (10%) patients, retention of this region may promote early DOD. In contrast, a high frequency of AI at chromosome 13q14 was detected in 71% of late-term DOD patients but only 26% of short-term DOD and 15% of LTS patients. Dysfunction of the retinoblastoma (RB1) gene, which maps to chromosome 13q14, has been associated with increased risk for cancer-specific death, in patients who have survived at least five years from diagnosis (31). Thus, AI at chromosome 16q22-q24 may protect against early lethality while AI at chromosome 13q14 drives long-term DOD.

In conclusion, tumors from patients who died from breast cancer differ at the chromosomal level from patients with good prognosis. The higher frequency of AI in DOD tumors suggests that increased genomic instability is associated with poor prognosis. In addition, AI at chromosomes 7q31, 8p22-p21, 17p13.3, 17p13.1, 18q21 and 22q12.3 may contribute to overall less favorable prognosis, while retention of chromosome 16q22-q24 is associated with early mortality and at chromosome 13q14 with late mortality. These data improve our understanding of the molecular underpinnings of breast cancer mortality and suggest patients with short- and long-term mortality may have distinct genetic diseases.

## **Acknowledgments**

We thank Marilyn Means for her assistance in collecting follow-up data for this study.

This research was supported by a grant from the United States Department of Defense

(Military Molecular Medicine Initiative MDA W81XWH-05-2-0075, Protocol 01-

20006). The opinion and assertions contained herein are the private views of the authors

and are not to be construed as official or as representing the views of the Department of

the Army or the Department of Defense.

## References

- (1) American Cancer Society. Breast cancer facts and figures 2011-2012. 2011. Atlanta, GA, American Cancer Society.
- (2) Sorlie T, Perou CM, Tibshirania R, Aas T, Geisler S, Johnsen H et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci USA* 2001; 98(19):10869-10874.
- (3) Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA et al. Molecular portraits of human breast tumours. *Nature* 2000; 406(6797):747-752.
- (4) Gort M, Broekhuis M, Otter R, Klazinga NS. Improvement of best practice in early breast cancer: actionable surgeon and hospital factors. *Breast Cancer Res Treat* 2007; 102(2):219-226.
- (5) van't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002; 415(6871):530-536.
- (6) Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med* 2004; 351(27):2817-2826.
- (7) Paik S, Tang G, Shak S, Kim C, Baker J, Kim W et al. Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. *J Clin Oncol* 2006; 24(23):3276-3274.
- (8) Eden P, Ritz C, Rose C, Ferno M, Peterson C. "Good Old" clinical markers have similar power in breast cancer prognosis as microarray gene expression profilers. *Eur J Cancer* 2004; 40(12):1837-1841.
- (9) American Joint Committee on Cancer. *AJCC Cancer Staging Manual*. 7 ed. New York: Springer; 2010.
- (10) Bloom HJ, Richardson WW. Histological grading and prognosis in breast cancer. *Br J Cancer* 1957; 11:359-377.
- (11) Elston CW, Ellis IO. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow up. *Histopathology* 1991; 19:403-410.
- (12) Ellsworth DL, Shriver CD, Ellsworth RE, Deyarmin B, Somiari RI. Laser capture microdissection of paraffin-embedded tissues. *Biotechniques* 2003; 34(1):42-46.
- (13) Ellsworth RE, Ellsworth DL, Lubert SM, Hooke J, Somiari RI, Shriver CD. High-throughput loss of heterozygosity mapping in 26 commonly deleted regions in breast cancer. *Cancer Epidemiol Biomarkers Prev* 2003; 12(9):915-919.

- (14) Medintz IL, Lee C-CR, Wong WW, Pirkola K, Sidransky D, Mathies RA. Loss of heterozygosity assay for molecular detection of cancer using energy-transfer primers and capillary array electrophoresis. *Genome Res* 2000; 10:1211-1218.
- (15) Sorlie T, Tibshirania R, Parker J, Hastie T, Marron JS, Nobel A et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci USA* 2003; 100(14):8418-8423.
- (16) Kennecke H, Yerushalmi R, Woods R, Cheang MC, Voduc D, Speers CH et al. Metastatic behavior of breast cancer subtypes. *J Clin Oncol* 2010; 28(20):3271-3277.
- (17) Smid M, Wang Y, Zhang Y, Sieuwerts AM, Yu J, Klijn JG et al. Subtypes of breast cancer show preferential site of relapse. *Cancer Res* 2008; 68(9):3108-3114.
- (18) Dellas A, Torhorst J, Schultheiss E, Mihatsch MJ, Moch H. DNA sequence losses on chromosomes 11p and 18q are associated with clinical outcome in lymph node-negative ductal breast cancer. *Clin Cancer Res* 2002; 8(5):1210-1216.
- (19) Bieche I, Champeme MH, Matifas F, Hacene K, Callahan R, Lidereau R. Loss of heterozygosity on chromosome 7q and aggressive primary breast cancer. *Lancet* 1992; 339(8786):139-143.
- (20) Emi M, Utada Y, Yoshimoto M, Sato T, Minobe K, Matsumoto S et al. Correlation of allelic loss with poor postoperative survival in breast cancer. *Breast Cancer* 1999; 6(4):351-356.
- (21) Rennstam K, Ahlstedt-Soini M, Baldetorp B, Bendahl PO, Borg A, Karhu R et al. Patterns of chromosomal imbalances defines subgroups of breast cancer with distinct clinical features and prognosis. A study of 305 tumors by comparative genomic hybridization. *Cancer Res* 2003; 63(24):8861-8868.
- (22) Garcia S, Dales JP, Charafe-Jauffret E, Carpentier-Meunier S, Andrac-Meyer L, Jacquemier J et al. Overexpression of c-Met and of the transducers PI3K, FAK and JAK in breast carcinomas correlates with shorter survival and neoangiogenesis. *Int J Oncol* 2007; 31(1):49-58.
- (23) Chang J, Clark GM, Allred DC, Mohsin S, Chamness G, Elledge RM. Survival of patients with metastatic breast carcinoma: importance of prognostic markers of the primary tumor. *Cancer* 2003; 97(3):545-553.
- (24) Rodrigues-Ferreira S, Di Tommaso A, Dimitrov A, Cazaubon S, Gruel N, Colasson H et al. 8p22 MTUS1 gene product ATIP3 is a novel anti-mitotic protein underexpressed in invasive breast carcinoma of poor prognosis. *Plos ONE* 2009; 4(10):e7239.

- (25) Dawson SJ, Makretsov N, Blows FM, Driver KE, Provenzano E, Le Quesne J et al. BCL2 in breast cancer: a favourable prognostic marker across molecular subtypes and independent of adjuvant therapy received. *Br J Cancer* 2010; 103(5):668-675.
- (26) Downing TE, Oktay MH, Fazzari MJ, Montagna C. Prognostic and predictive value of 16p12.1 and 16q22.1 copy number changes in human breast cancer. *Cancer Genet Cytogenet* 2010; 198(1):52-61.
- (27) Ellsworth RE, Hooke JA, Love B, Kane JL, Patney HL, Ellsworth DL et al. Correlation of levels and patterns of genomic instability with histological grading of invasive breast tumors. *Breast Cancer Res Treat* 2008; 107(2):259-265.
- (28) Natarajan R, Lambros MB, Geyer FC, Marchio C, Tan DS, Vatcheva R et al. Loss of 16q in high grade breast cancer is associated with estrogen receptor status: Evidence for progression in tumors with a luminal phenotype? *Genes Chromosomes Cancer* 2009; 48(4):351-365.
- (29) Nordgard SH, Johansen FE, Alnaes GI, Bucher E, Syvanen AC, Naume B et al. Genome-wide analysis identifies 16q deletion associated with survival, molecular subtypes, mRNA expression, and germline haplotypes in breast cancer patients. *Genes Chromosomes Cancer* 2008; 47(8):680-696.
- (30) Thomassen M, Tan Q, Kruse TA. Gene expression meta-analysis identifies chromosomal regions and candidate genes involved in breast cancer metastasis. *Breast Cancer Res Treat* 2009; 113(2):239-249.
- (31) Chano T, Ikebuchi K, Tomita Y, Jin Y, Inaji H, Ishitobi M et al. RB1CC1 together with RB1 and p53 predicts long-term survival in Japanese breast cancer patients. *Plos ONE* 2010; 5(12):e15737.

References Annals uses the American Medical Association Manual of Style, 10th Edition (New York: Oxford University Press, 2007) References must be cited in consecutive numerical order at first mention in the text and arranged numerically, not alphabetically, on pages preceded by the head "References." Reference in the text should be identified by numbers in brackets (e.g. [1]) In each reference, list all authors' names when there are seven or fewer; if there are more than seven, list the first three authors followed by et al.

The author is responsible for the accuracy of the references. References must be printed double-spaced. Material cited in the reference list "in press" must have been accepted for publication, not merely submitted for review. Ibid references are not permitted.

Unpublished data and personal communications should not be included in the reference list; rather, this information may be included in the text, with pertinent identification (A. Author, unpublished data) or (B. Author, personal communication). The unpublished data of others and personal communications can be used only when written authorization from the data owner or communicator is submitted with the original manuscript.

### **Sample References for Style Reference**

#### ***Journal***

1. Badgwell B, Cormier JN, Xing Y, et. al. Attempted Salvage Resection for Recurrent Gastric or Gastroesophageal Cancer. *Ann Surg Oncol* 2009; 16:42-50.

Journal article with DOI reference

2. Xing M: BRAF Mutation in Papillary Thyroid Microcarcinoma: The Promise of Better Risk Management. *Ann Surg Oncol* DOI: 10.1245/s10434-008-0298-z [Online January 22, 2009].

***Book***

3. Bland KI, Sarr MG, Büchler, MW, et. al. *General Surgery Principles and International Practice*, Second Edition. New York: Springer-Verlag, 2009.

Book Chapter

4. Jones SM, Roh MS. Results of Surgical Resection for Hepatocellular Carcinoma.. In: Talamonti MS, Pappas SG, eds. *Liver-Directed Therapy for Primary and Metastatic Liver Tumors*. Massachusetts:: Kluwer Academic Publishers, 2001:59-76.

Web-Based Resource

5. Society of Surgical Oncology (2007). Position Statement on Prophylactic Mastectomy. Available: <http://www.surgonc.org/default.aspx?id=47> [accessed January 28, 2009]

Authors should save each figure as a single image file in either uncompressed TIFF (Tag Image File Format) or EPS (Encapsulated PostScript) format. The JPEG format is acceptable if the image is saved at the highest quality (without or with lossless compression). Images created in slide presentation programs, such as Microsoft PowerPoint, are not recommended. Charts created with Microsoft Excel are not

acceptable in any circumstances. If you are unable to submit your manuscript via Manuscript Central, please contact the Editorial Office .

## Figure legends

Figure 1. Survival curves of the six chromosomal regions associated with significant differences in survival between patients with and without AI. Red circles = tumors demonstrating AI, blue squares = tumors without detectable AI.

Figure 2. Survival curves for patients with short-term and long-term survival. For chromosome 13q14, survival curves between patients with and without AI only being to separate after 60 months, suggesting that alteration of 13q14 is associated with long-term mortality. For chromosome 16q22-q24, only the patients with short-term mortality were included in survival analysis; here, those patients with AI at chromosome 16q22-q24 (red circles) have significantly better survival than those with retention of chromosome 16q22-q24 (blue squares).

Table 1. Clinical and pathological features of 122 invasive breast tumor specimens. P-values listed as NS (not significant) >0.05. P-value (LTS vs. DOD)

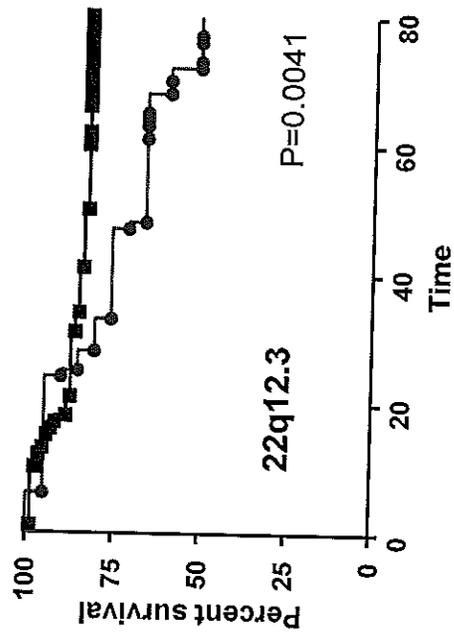
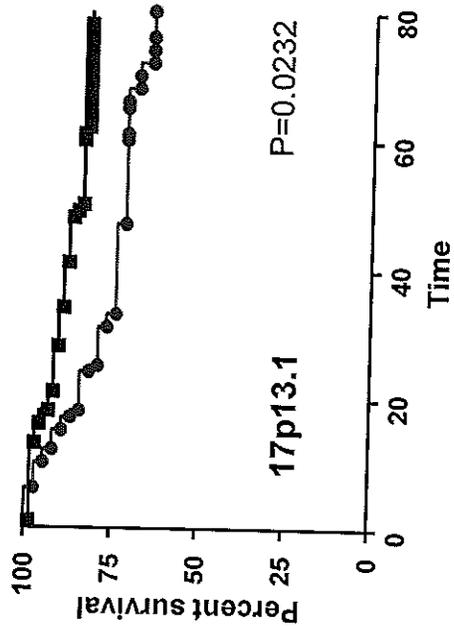
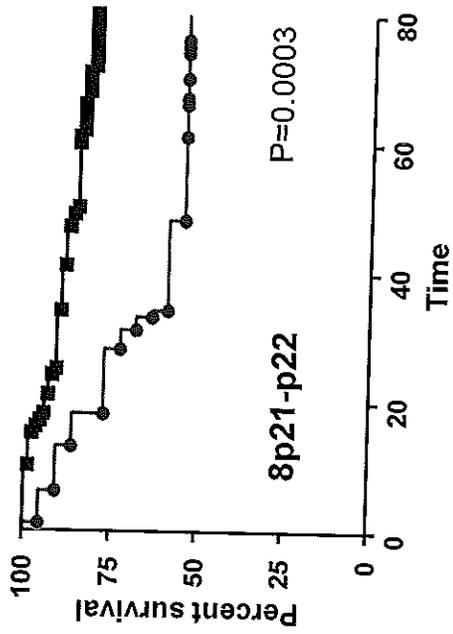
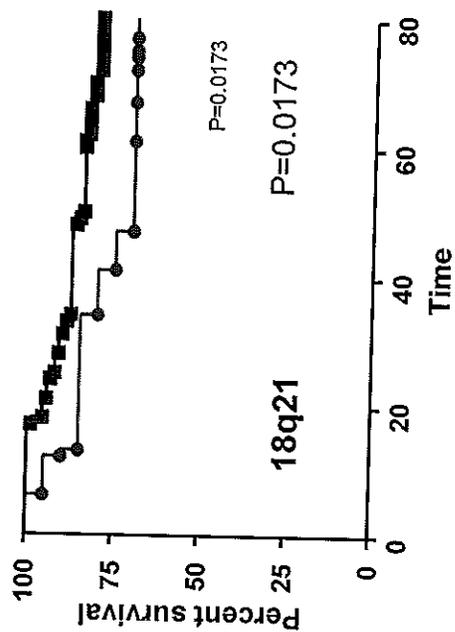
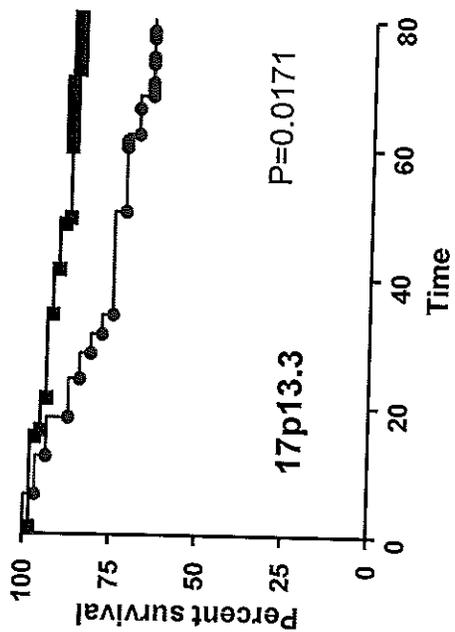
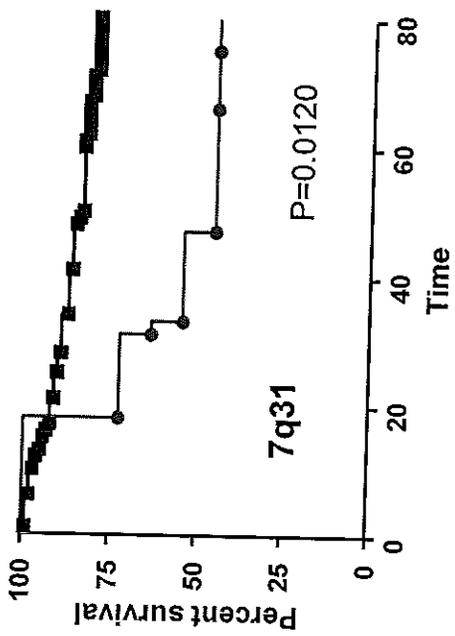
	LTS (n=91)	DOD (n=31)	P-value (LTS vs. DOD)
<b>Age at diagnosis</b>			<b>P&lt;0.005</b>
	<40 years	13%	
	40-49 years	32%	
	≥50 years	55%	
<b>Ethnicity</b>			NS
	White	68%	
	African American	32%	
	Asian	0%	
	Hispanic	0%	
<b>Histology</b>			NS
	IDCA	90%	
	ILCA	6%	
	Mixed	4%	
	Other <sup>a</sup>	0%	
<b>Tumor size</b>			<b>P&lt;0.05</b>
	T1	38%	
	T2	52%	
	T3	10%	
<b>TNM Stage</b>			<b>P&lt;0.001</b>
	I	10%	
	II	19%	
	III	23%	
	IV	48%	
<b>Tumor Grade</b>			<b>P&lt;0.005</b>
	Low	6%	
	Moderate	29%	
	High	65%	
<b>Lymph Node Status</b>			<b>P&lt;0.001</b>

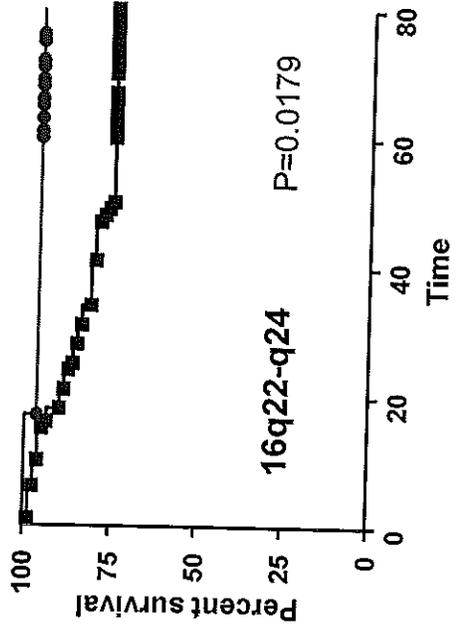
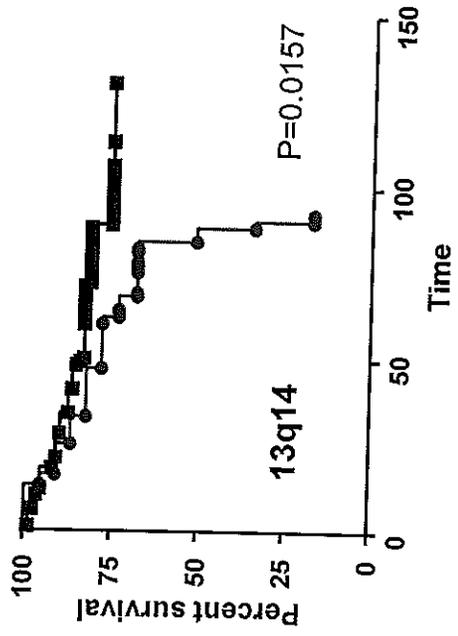
<i>Subtype<sup>b</sup></i>			
Negative	64%	13%	<b>P&lt;0.05</b>
Positive	36%	87%	
Luminal A	68%	42%	
Luminal B	9%	6%	
HER2 enriched	7%	10%	
Triple negative	16%	42%	

<sup>a</sup>Other included histological types including tubular, medullary, apocrine, and mucinous carcinomas  
<sup>b</sup>IHC markers were used as surrogates to define subtype as follows: luminal A = ER and or PR+/HER2-; luminal B = ER and or PR+/HER2+; HER2-enriched = ER and PR-/HER2+ and triple negative = ER, PR and HER2-

Table 2. Frequency of AI in DOD and LTS patients at 26 chromosomal regions. NS = not significant. Note that AI at chromosome 16q22-q24 is significantly higher in the LTS compared to DOD patients.

Chromosomal Region	DOD (n=31)	LTS (n=91)	P DOD v LTS
1p36.1-p36.2	0.29	0.21	NS
2q21.3-23.3	0.14	0.20	NS
3p14.1	0.31	0.21	NS
5q21.1-q21.3	0.14	0.21	NS
6q15	0.28	0.19	NS
6q22.1-q23.1	0.34	0.21	NS
6q25.2-q27	0.31	0.24	NS
7q31.1-q31.31	0.25	0.06	0.021
8p22-p21.3	0.41	0.12	0.001
8q24	0.23	0.19	NS
9p21	0.28	0.12	NS
10q23.31-q23.33	0.29	0.19	NS
11p15	0.23	0.24	NS
11q13.1	0.20	0.19	NS
11q23	0.29	0.23	NS
13q12.3	0.28	0.27	NS
13q14.2-q14.3	0.38	0.15	0.009
14q32.11-q31	0.28	0.18	NS
16q11.2-q22.1	0.20	0.30	NS
16q22.3-q24.3	0.14	0.36	0.033
17p13.3	0.55	0.28	0.019
17p13.1	0.54	0.28	0.015
17q12-q21	0.22	0.19	NS
18q21.1-q21.3	0.36	0.13	0.009
22q12.3	0.39	0.14	0.005
22q13.1	0.33	0.28	NS





## Development of DNA aptamers that inhibit invasion, migration, and growth of breast cancer cells.

Rebecca Kopco<sup>1</sup>, Joji Iida<sup>1</sup>, John Lehman<sup>1</sup>, Stella Somiari<sup>1</sup>, Richard Somiari<sup>2</sup>, Richard Mural<sup>1</sup> and Craig Shriver<sup>3</sup>

1: Department of Cell Biology, Windber Research Institute, Windber, PA 15963

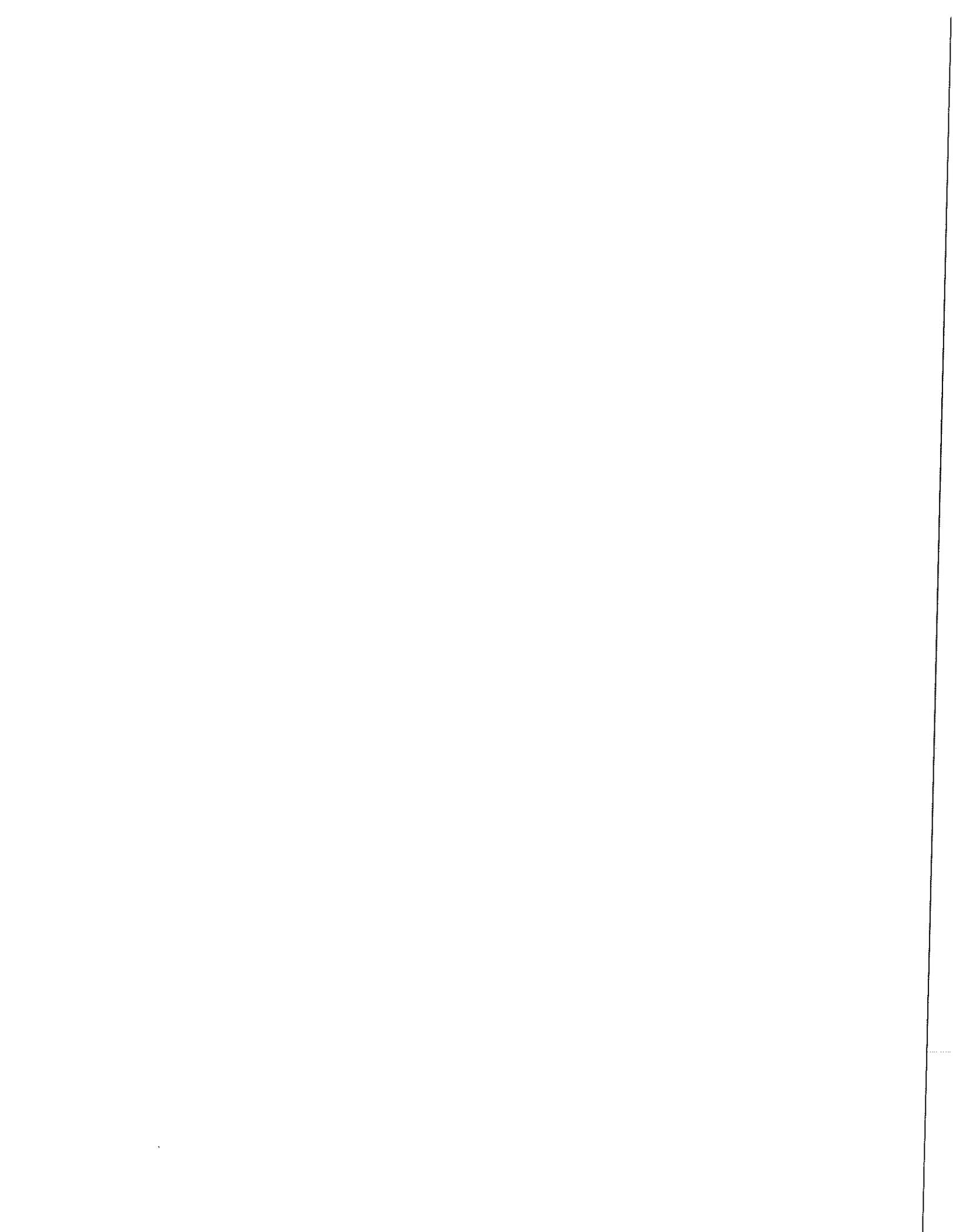
2: ITSI-Biosciences, Johnstown, PA 15901

3: Department of Surgery, Walter-Reed Army Medical Center, Washington, DC 20902

Triple negative (TN) breast cancers are defined by a lack of expression of estrogen, progesterone, and her-2/neu receptors. It is widely recognized that TN breast cancers have a poorer prognosis than other subtypes of breast cancer. Given the lack of effective targeted therapies for TN breast cancer patients, understanding of the mechanisms of growth and invasion of these tumors will provide insight into developing novel approaches to lower the mortality from TN breast cancer.

We utilized three TN breast cancer cell lines (HCC38, HCC1806, HCC1937) as model systems to characterize mechanisms of growth, migration, and invasion. FACS analyses demonstrated that these TN cell lines showed a phenotype of CD44<sup>+</sup>CD24<sup>-low</sup> epithelial specific antigen (ESA)<sup>+</sup>. An inhibitory antibody against exon v10 of CD44 significantly inhibited three-dimensional (3D) growth of tumor cells. Importantly, this antibody also inhibited alpha2beta1/alpha3beta1 integrins-mediated migration and invasion into matrigel. These results suggest that CD44 harboring exon 10 facilitates progression and metastasis of TN breast cancer cells in vivo by enhancing growth, migration, and invasion. We also demonstrated that this exon interacts with a cell surface tyrosine kinase, EphA2. Thus, we hypothesize that CD44 would form a molecular complex with EphA2 through exon v10 on TN breast cancer cells that plays a key role in facilitating invasion and metastasis.

In order to develop novel reagents to inhibit TN breast cancer invasion and metastasis, we have utilized SELEX (systematic evolution of ligands by exponential enrichment) to isolate DNA aptamers that specifically recognize exon v10 of CD44. DNA aptamers are functional molecules with the appropriate sequence and structure to form a complex with a target molecule. Typically, dissociation constants for these aptamer-target complexes are in the high pico-molar to low nano-molar range, which is comparable to antibodies. Given the ease of chemical synthesis and modifications of oligonucleotides, DNA aptamers that recognize cancer cells will be easier to develop and have a great impact for the next generation of strategies for cancer detection, diagnosis, and therapy. We describe the development of "function-blocking" DNA aptamers specific for exon v10 of CD44 that inhibits TN breast cancer invasion and metastasis.





# The Journal

Vol. 24

No. 16

www.bethesda.med.navy.mil

April 19, 2012

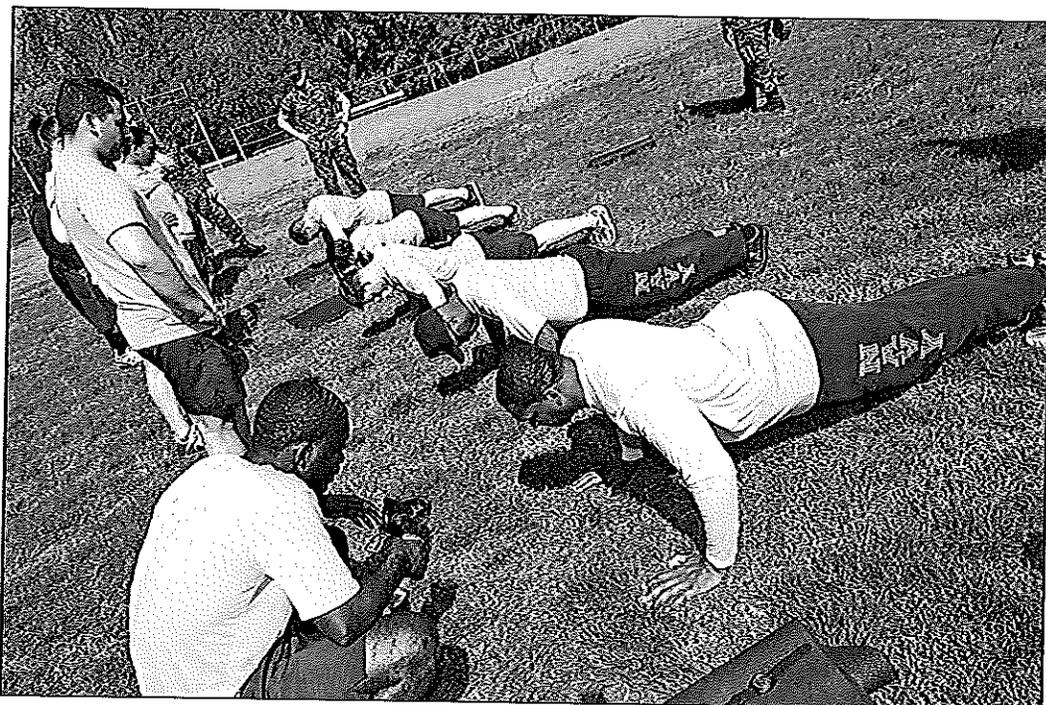
## NSAB Conducts First Independent PRT Cycle

By Mass Communication Specialist Seaman Dion Dawson  
NSAB Public Affairs staff writer

With physical readiness tests (PRT) season here once again, Naval Support Activity Bethesda (NSAB) is using the current PRT cycle as an opportunity to stand its first cycle independent of the Walter Reed National Military Medical Center (WRNMMC).

"NSAB independently conducting its own physical readiness testing represents an important milestone in further establishing the installation as a member of Naval Installations Command," said Capt. Frederick Kass, commanding officer of NSAB.

All personnel assigned to NSAB will now fall under the new program with the exception of those currently enrolled in the Fitness Enhancement Program (FEP); they will still fall under the WRNMMC. At the completion of this cycle, the Command Fitness Leader (CFL)



Naval Support Activity Bethesda command fitness leaders watch closely as Sailors complete the push-up portion of the Navy's semi-annual physical readiness test.

Photo by Mass Communication Specialist Seaman Dion Dawson

See PRT page 7

## Shriver Leads U.S. Military Cancer Institute to Excellence

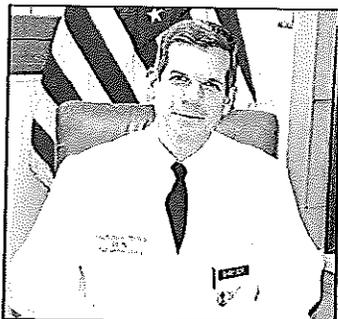


Photo by Sarah Fortney

Col. Craig Shriver was recently named interim director of the U.S. Military Cancer Institute. He is also the director and principal investigator of the Clinical Breast Care Project and Chief of General Surgery at Walter Reed Bethesda.

By Sarah Fortney  
Journal staff writer

Col. (Dr.) Craig Shriver knew he wanted to make a difference in breast cancer research, long before it hit home.

Named interim director of the Cancer Center at Walter Reed Bethesda in February 2010, Shriver graduated cum laude from Albright College, in Reading, Pa., with a bachelor's degree in biochemistry in 1980. He then earned his medical degree in 1984 from Temple School of Medicine, in Philadelphia.

"I knew I wanted to make an impact as a cancer surgeon. There was a lot we didn't know about cancer in general," he said. Considering breast cancer is, and has been, the most common cancer amongst women, "I felt early in my career

that's an area where I could make a difference," Shriver said.

While attending medical school on an Army scholarship, he found his passion for serving in the military and completed his general surgery residency at the former Walter Reed Army Medical Center (WRAMC) in 1989. Shriver was later selected for an advanced fellowship in surgical oncology at the Memorial Sloan-Kettering Cancer Center in New York. During his career, he has published more than 200 academic papers including those on lymph node biopsy in breast cancer. His publications are included in the *New England Journal of Medicine*.

In 1993, after completing his fellowship, he returned to WRAMC as a general surgeon, and later became chief of surgical oncology two years later.

After his first of two deployments

to Afghanistan, in 2007, he returned home to learn his wife had been diagnosed with breast cancer.

"It's different being on this side, versus that side," he said, pointing across his desk where a patient might sit, then to his lab coat.

The advancements in research aided his wife's successful recovery, he said. "She was the beneficiary of everything we've learned in the last 20 years."

He recalled how far breast cancer treatment has advanced.

"When I first started as a young surgeon, a woman with breast cancer would come in with a large tumor and would need a mastectomy, and all of her lymph nodes removed. We would treat every patient with breast cancer the same

See CANCER page 7

# Preparing to Weather the Storm

By Joe Macri  
NSAB Public Affairs Officer

Every season has its share of potentially destructive weather events and summer is no exception. Naval District Washington installations must stay aware of these threats and ensure the region is properly prepared for natural disasters.

The Navy prepares for the threat of tropical cyclones with a yearly exercise called HURREX/Citadel Gale. The purpose of this annual exercise is to prepare commands to respond to weather threats to U.S. coastal regions, and to maintain the ability to deploy forces even under the most adverse weather conditions. HURREX 12 is currently underway and runs until April 27.

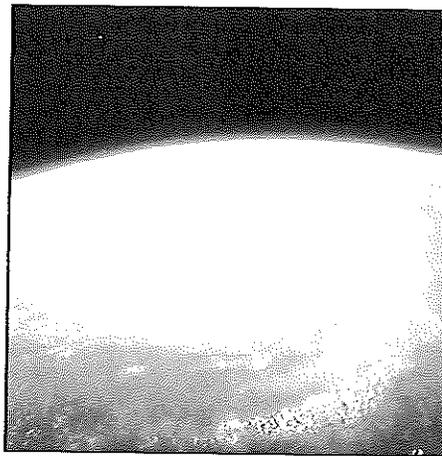
Naval Support Activity Bethesda (NSAB) will participate in HURREX 12 through several "behind the scenes" events designed to test the installation's ability to prepare for destructive weather.

According to NSAB Director of Emergency Management Ron Kunz, many people in the region often focus on adverse winter weather but forget to prepare for hurricane season which generally runs from June 1 to Nov. 30.

"People seem to remember events like 'Snowmageddon', but because we feel like we're inland here, we don't tend to get too worried about hurricanes, but even inland they can generate very strong wind and rain," said Kunz.

An event scheduled to test the installation's response to destructive weather, is the evacuation exercise of Tranquility Hall, which houses wounded warriors.

One of the key criteria's for success during the evacuation is the 100 percent accountability of all personnel in Building 62, because the command considers the safety of its service members and their families the highest possible priority, especially during violent weather.



Accountability is emphasized before, during and after destructive weather.

"We will be coordinating with local fire and rescue to test our Mutual Aid Response agreements and our ability to immediately respond to a required evacuation in difficult conditions," said Kunz.

The evacuation drill will take place on Tuesday. Affected residents of Building 62 will receive fliers to help them prepare for the event.

"The intent of this event is to balance our ability to properly provide for the safety and security of our wounded warriors during an incident, while also trying to make the exercise as minimally disruptive as possible," said NSAB Executive Director Bill Meekins.

For more information on preparing for destructive weather, please visit the Naval Safety Center website at: <http://www.public.navy.mil/navsafecen/Documents/media/safetips/fm/hurricane.doc>

For more information on Navy Family Emergency Preparedness, please visit: [http://www.usnic.navy.mil/CNIC\\_HQ\\_Site/WhatWeDo/FamilyLine/FamilyPreparedness/index.htm](http://www.usnic.navy.mil/CNIC_HQ_Site/WhatWeDo/FamilyLine/FamilyPreparedness/index.htm)

opment of a culture of fitness within the Navy has become an ever-increasing focus to the Navy. Physical readiness as a whole has taken on a bigger spotlight when it comes to retention tools such as Perform to Serve, a member's evaluation and a part of shaping that culture. Understand, as the CFL, I am overall responsible for the management of the program, but it is a team effort to complete the administration and coordination of the events," he said. The same requirements I had to meet, had to be met by the ACFLs as well."

The ACFLs assist in the coordination of PRTs, tracking required documents, and leading coordinated physical training sessions.

"Being an ACFL is a rewarding position to have," said Master-At-Arms 3rd Class Ian Hamilton with NSAB security forces. "I am able to stay in shape and help my command at the same time. The same hard work and effort I apply for my own PRT score, I use to help my shipmates. If there is anything I can do to make sure they are successful, I do it."

With the cycle currently underway until May 18, Sailors are encouraged to contact Gamba at (301) 295-2558 to schedule their PRT.

## CANCER

Continued from 1

and the cultural awareness," Shriver said. People were ashamed of having breast cancer due to lack of awareness and screening programs that are available today 25 years later.

"Women know they have to take care of themselves, and screening programs are available," he continued. "In addition to that, we've gone from an era when everybody's treated the same, to individualized treatment."

It's now common knowledge that every woman's breast cancer is unique, he added.

"We've gone from one-size-fits-all, to very focused treatment," he said. "We can be much more targeted. We can identify exactly what type of tumor it is, save the breast, just take out the tumor and be very targeted with the lymph nodes, assess those, and then do a genetic analysis of the tumor itself, then give the treatment based on that person's tumor, not somebody else's tumor."

Shriver continued, it's peace of mind knowing these available resources will continue to advance for his teenage daughter and son's generation. Pointing to a picture of his children, he said they'll have the opportunity to undergo testing now to learn whether they might carry the gene for breast cancer later in life.

To continue enhancing breast cancer care, as well as oncology, at Walter Reed Bethesda, the USMCI will be incorporated, and dissolved, into the medical center's Comprehensive Cancer Center over the next six months, he said. Comprised of military and civilian oncologists who develop cancer research, education and care for the military, the USMCI will enhance the services already provided at the cancer center. This includes the Prostate Project, GYN oncology project, the Breast Project, Medical Oncology, Radiation Oncology, Pediatric Oncology, and the Biobanking Initiatives under Pathology.

"It's an organization where many elements are coming together with one goal of [providing] the best cancer care in the Department of Defense, and in the nation, for our beneficiaries," Shriver added. "We seek to continue to improve that."

He noted that the cancer center's core elements - now a culmination of best practices from WRAMC and the former National Naval Medical Center - have been in existence for more than a decade, and is the military's only cancer center. Funded directly by the Department of Defense, the center has applied for National Cancer Institute (NCI) designation, an achievement only 40 civilian centers in the nation have earned, he added.

Kay Kelley, research protocol coordinator for the center, echoed Shriver's sentiments, stating, "The cancer center has so much to offer."

A former Navy nurse, Kelley said the center strives to offer convenience, allowing patients to receive all of their care under one roof, all the while with their family alongside.

The patient can stay in one spot, instead of traveling to different locations, she said, adding that patients receiving care in the center are making a contribution to the knowledge of cancer treatment. Kay explained each patient completes a questionnaire, focusing on topics, such as medical and family history, which is kept in a database and used as part of research.

Kay continued, she's learned a lot about breast cancer research, having worked at the center for six years, and has enjoyed working under Col. Shriver's leadership.

Shriver said he couldn't do his job without the support from his staff, or from leadership throughout the command and region.

"There's nothing better than leading an organization where everybody wants you to succeed. The wind is at our back," he said. "It's really exciting."

## PRT

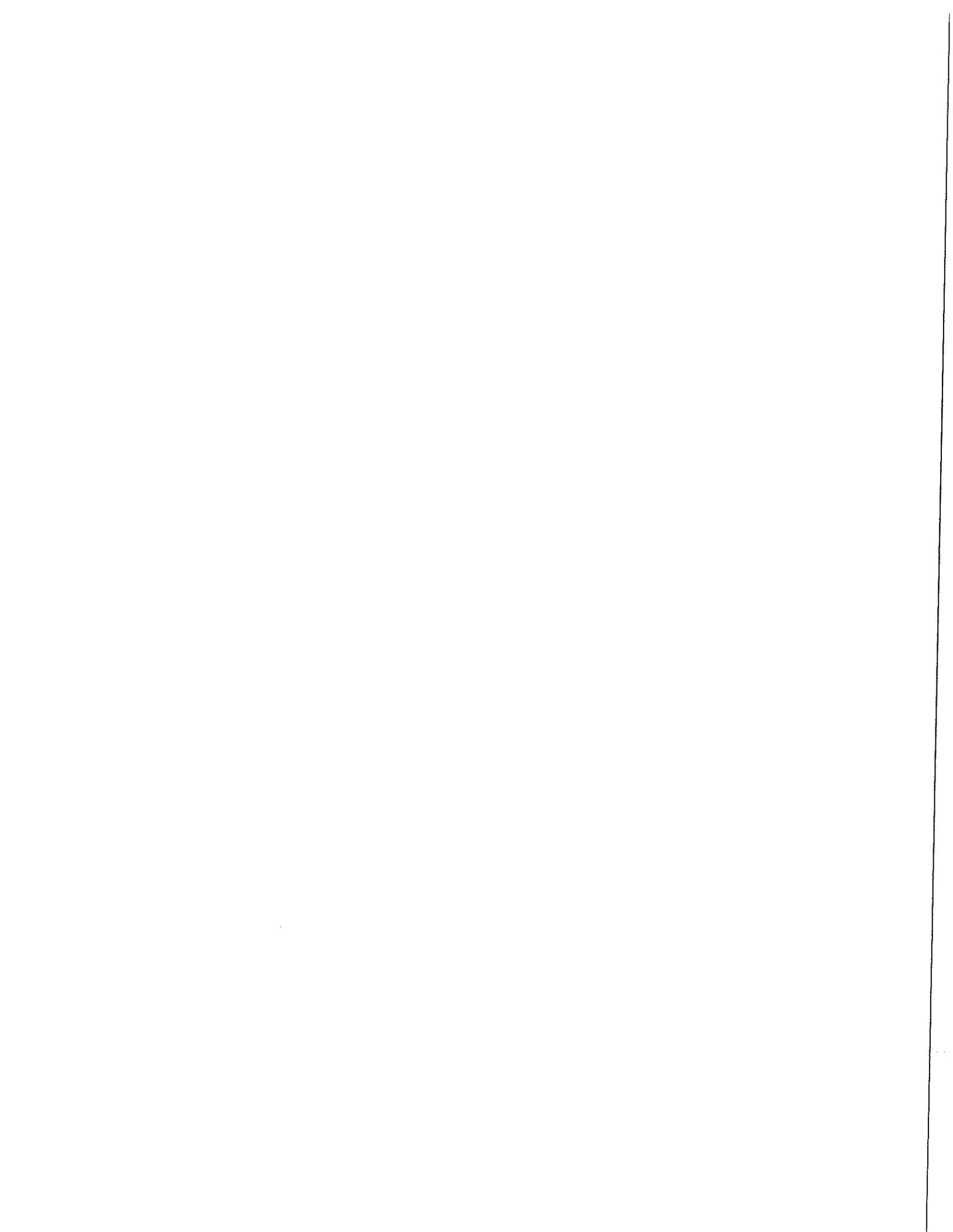
Continued from 1

and Assistant Command Fitness Leaders (ACFLs) will work together to stand up NSAB's own FEP.

"This is an amazing opportunity to show pride in our command," said Master-At-Arms 1st Class Michael Gamba, NSAB CFL. "Our first time conducting a PRT shows the continued growth as a command. Besides the fact it's the first one, the program itself is critical to both the Navy and NSAB as a way to assess the physical readiness of Sailors. At the completion of this cycle, myself and the ACFLs will work together to provide the Commanding Officer an assessment of his Sailors."

To become a CFL, Gamba had to attend a Command Fitness Leader Course. The pre-requisites for the class were an excellent or above on the last PRT, recommended by the command in the form of a Letter of Designation and a non-smoke status.

"It gives me a good feeling overall," Gamba said about his selection as the CFL. "The devel-



## **Should Immunohistochemical (IHC) Markers be Performed on Axillary Lymph Node Metastases in View of the Lack of Concordance Between The Primary Tumor and Axillary Lymph Node Metastases?**

**Authors:** Greer, Rosman, Mylander, Wareham, Campbell, Hooke, Kovatich, Shriver, Tafra

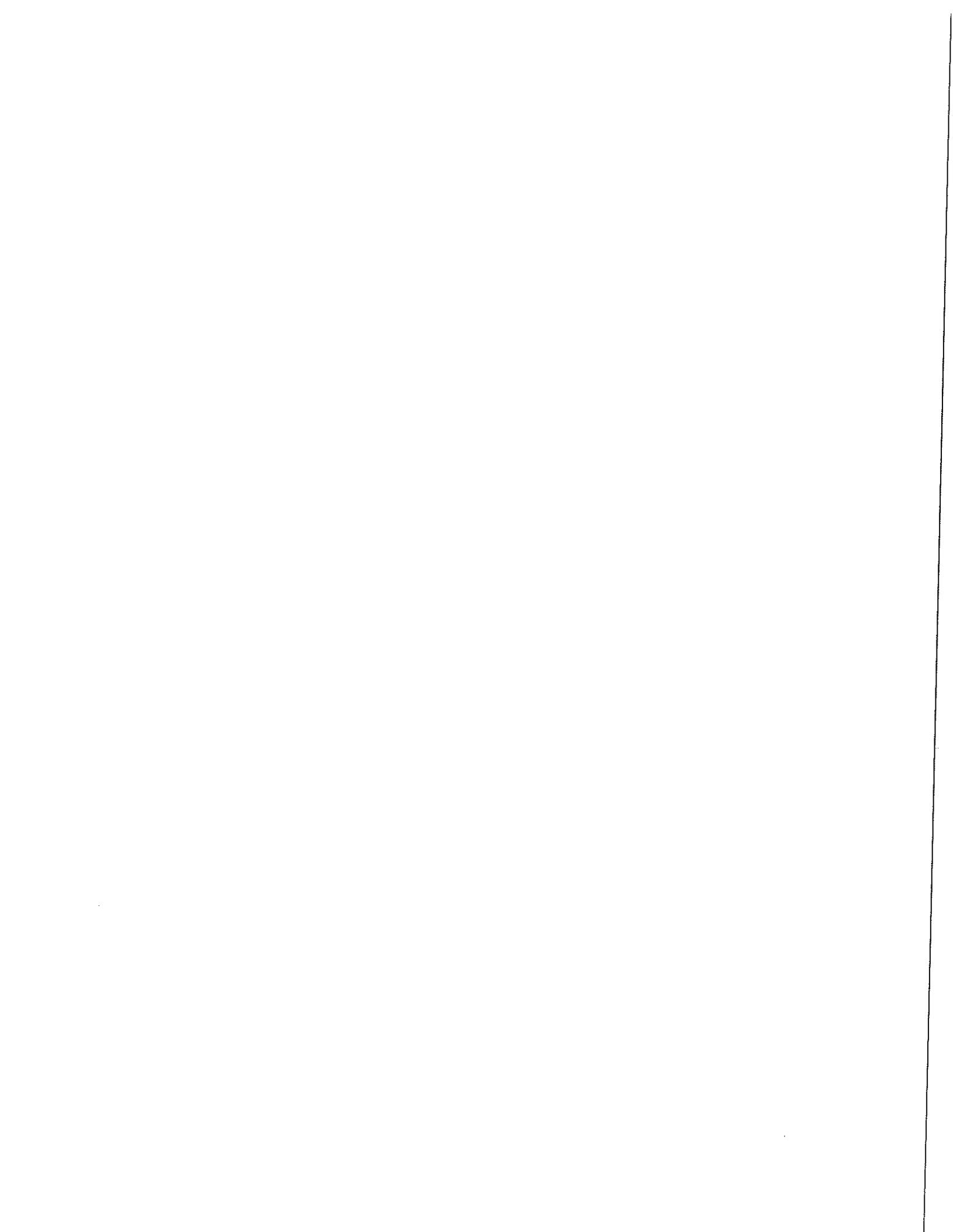
**Introduction:** Immunohistochemical (IHC) stains are performed on breast cancer primary tumors (PT) to tailor systemic therapy and they are not routinely repeated on axillary lymph node (ALN) metastases. Our aim was to examine differences in receptor expression between the PT and metastatic ALNs.

**Methods:** A prospective study was performed on 165 women (205 samples) with invasive breast cancer diagnosed from January 2009 to July 2011. Patients with ALN macro metastases were identified. Hormone receptor markers (ER, PR, HER2, and Ki67) by IHC were analyzed on the core needle biopsy and ALN metastases. Contingency tables and agreement modeling were performed.

**Results:** Thirty women with ALN macro metastases were identified. There was almost perfect concordance between PT and ALN for PR ( $\kappa=0.92$ , 97%) and HER2 ( $\kappa=0.87$ , 97%) staining; and moderate concordance for ER ( $\kappa=0.47$ , 83%) and Ki67 ( $\kappa=0.50$ , 70%). All biomarkers, especially ER and Ki67, had large quantitative decreases in the ALN compared to the PT. Five cases (17%) changed from ER positive in the PT to ER negative in the ALN. One case changed from HER2 positive in the PT to HER2 negative in the ALN.

**Conclusions:** Up to 20% of patients had significant biomarker changes between the PT and ALN metastases, which may alter the effectiveness of systemic therapy in these patients. There may be an added benefit to performing IHC stains on ALN metastases to increase therapeutic sensitivity and decrease resistance to therapy, but further validation in clinical trials is needed.

"The views expressed in this abstract are those of the authors and do not reflect the official policy of the Department of the Army (DOA), Department of Defense (DOD), or US Government."



## Effect of ASCO/CAP guidelines for determining ER status on molecular subtype

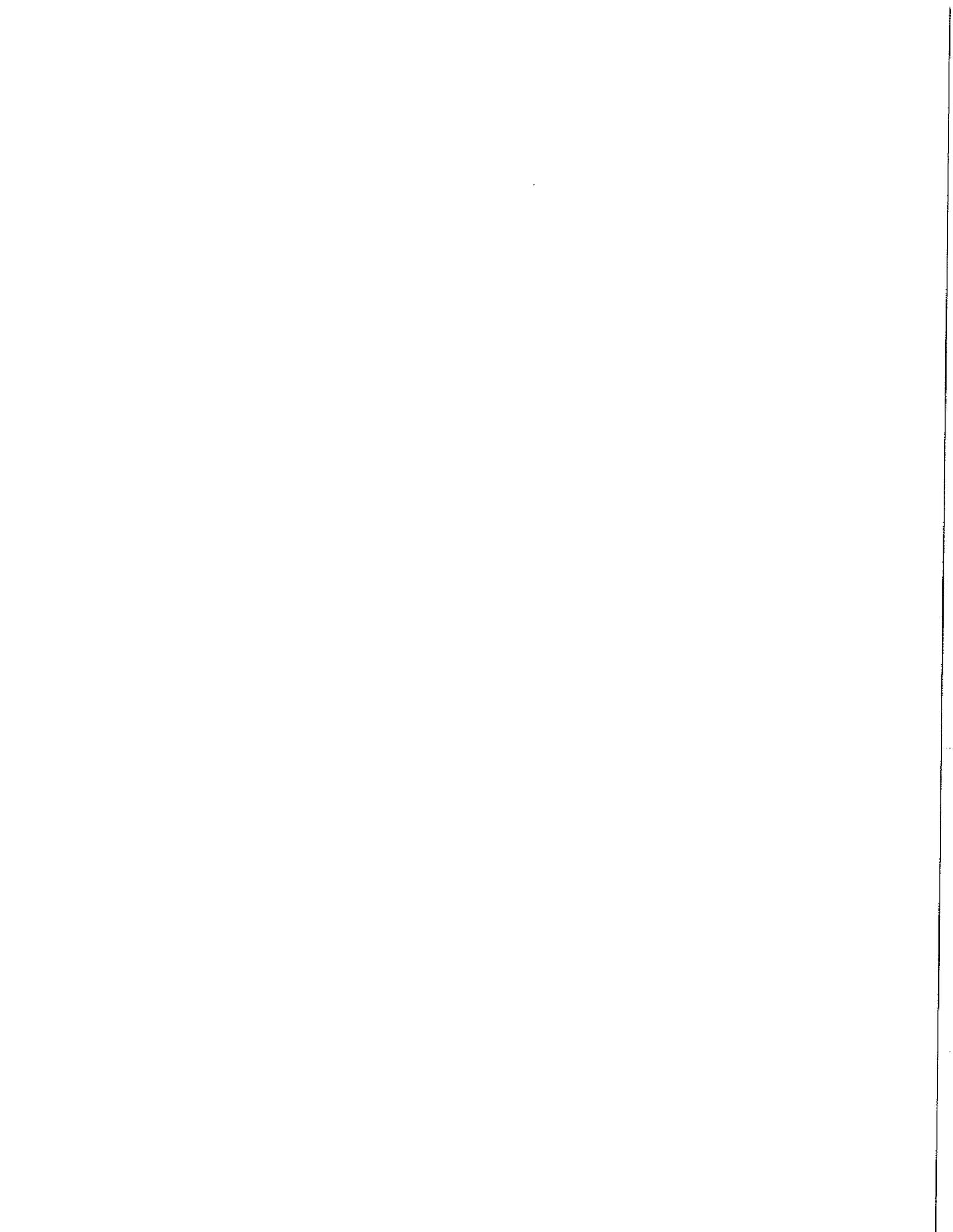
Rachel E. Ellsworth, Ryan van Laar, Brenda Deyarmin, Craig D. Shriver

**Background:** Determination of ER, PR and HER2 status are standard for predicting prognosis and determining treatment options for patients with breast cancer. In 2010, the American Society of Clinical Oncology (ASCO) and College of the American Pathologists (CAP) issued guidelines for scoring ER and PR status: the cutoff for defining ER/PR positive tumors was defined as  $\geq 1\%$  positively staining cells and endocrine therapy should be considered for these patients. Here, we determined how this cutoff affects molecular subtype.

**Methods:** Gene expression data was generated for 258 primary breast tumors from women enrolled in the Clinical Breast Care Project and intrinsic subtype determined using the nearest centroid method (ChipDx LLC). Surrogate IHC-based subtypes were defined as luminal A = ER and or PR+/HER2-; luminal B = ER and or PR+/HER2+; HER2-enriched = ER and PR-/HER2+ and triple negative = ER, PR and HER2-. IHC-defined subtypes using the new ASCO/CAP definitions of positive ER status were compared to molecular subtypes.

**Results:** The percent of positive staining cells for ER was available for 91% of tumors. Of the tumors with 0% ER staining, 69/70 (98.5%) were either basal-like or HER2-enriched. Of the 10 tumors with ER staining of 1-5%, all were ER negative at the molecular level: 7 basal-like and three HER2-enriched. Within the group of 13 tumors with ER staining of 6-50%, those tumors characterized as basal-like had ER values 7-20%, HER2 enriched tumors had ER values 15-46% and luminal A tumors had ER values  $>20\%$ .

**Conclusions:** Under the new ASCO/CAP guidelines, the 23 tumors assayed here with ER staining between 1-50% would be classified as ER positive, yet 83% are basal-like or HER2-enriched at the molecular level. While it is possible that patients with a small minority of ER positive cells would derive benefit from endocrine treatment, classification of these tumors as ER positive by IHC may preclude the use of PARP inhibitors in those patients with molecularly-defined basal-like breast tumors. Thus, as ER status is a critical element in the choice of treatments for patients with breast cancer, it is imperative that the most effective method for classifying tumors be developed.

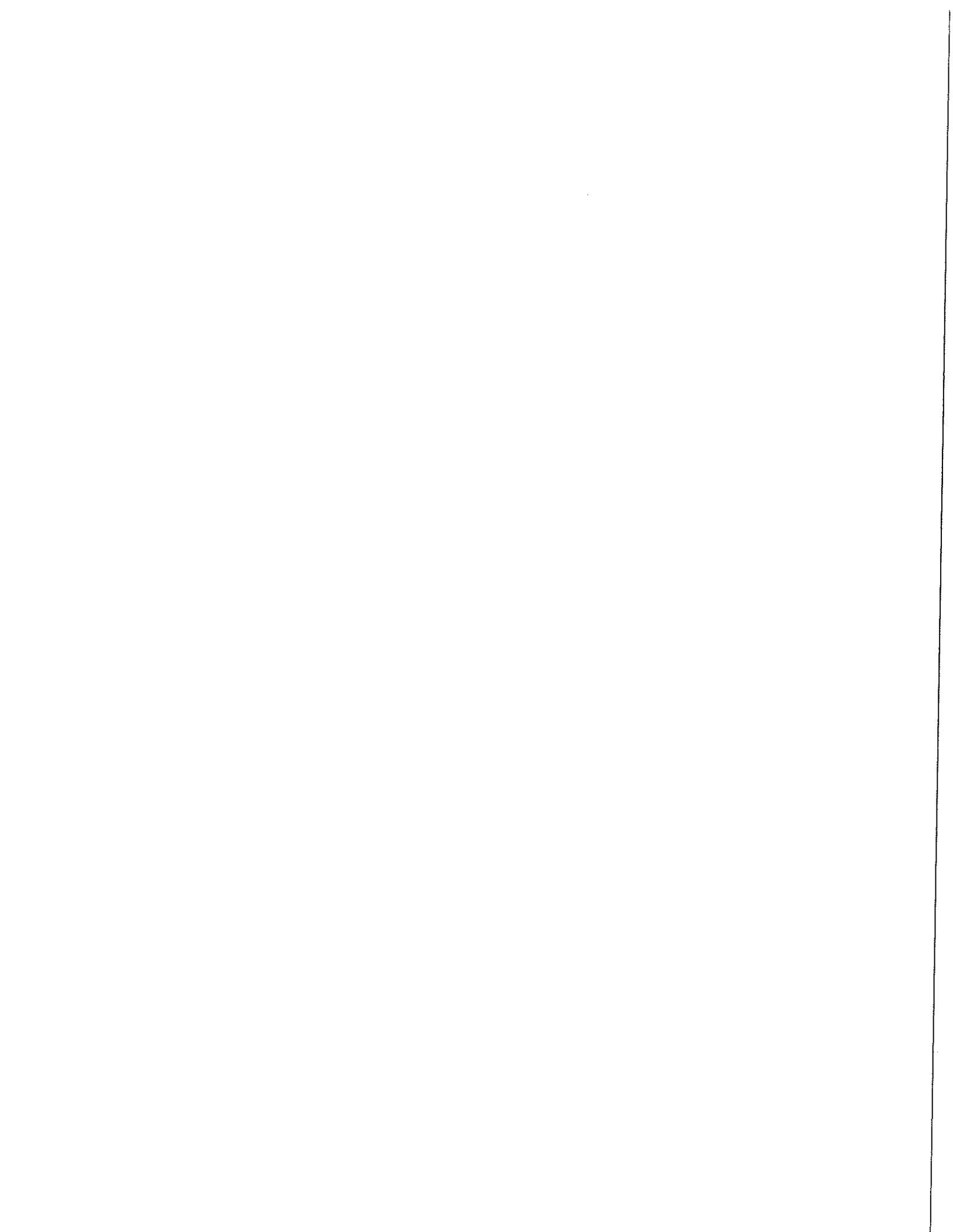


## **Challenges in data warehousing for translational research**

For presentation at the 2012 IDBS Translational Medicine

Hai Hu, Windber Research Institute

In the last few years I have led a project team composed of clinicians, lab scientists, and biomedical informaticians from the Windber Research Institute and its clinical and research partners, in collaboration with IT developers from InforSense/IDBS, to develop a Data Warehouse for Translational Research, which lead to the commercialization of the product ClinicalSense. The system is based on its unique ontology including a patient-centric clinical data model, a specimen-centric molecular data model, a temporal data model, and an image data model. The system is also characterized by two intuitive interfaces targeting non-informatics specialists, for aggregated biomedical information browsing and individual subject information viewing. Currently, translational research is in rapid development in multi-facets. Take The Cancer Genome Atlas project as an example; more than 20 cancer types have been identified for study. For each cancer types, multiple platforms of data have been generated, including DNA sequencing, DNA copy number, DNA methylation, gene expression, RNA-sequencing, etc. Analysis of each platform of the data, and integrative analysis of these data, imposes challenges not only in data integration but also application integration. In this presentation, I will use concrete examples to lay out example challenges we are facing in data warehousing for translational research, and the solutions we have developed or conceived.



DRAFT NEWS RELEASE

In addition to finding leads on diagnostic tools and therapies, the study might reveal proteins and molecular pathways that have gone astray and led to the cancer in the first place.

A recent, unrelated study reported in the news from the journal Nature re-grouped breast cancers into 10 sub-groups based on the cancer's genes and which genes were turned on or off in the cancerous cells. But genes are like a raw movie script -- how the movie turns out depends on many details beyond the words in the script. This new study will look beyond genes to see how cancer cells translated their scripts into live action.

The Army is a good place to find the understudied women who are more likely to get these breast cancers. Blacks make up about 20 percent of the Army, and more than 91 percent of the Army's soldiers are under 40.

The PNNL research effort will draw on the unique instruments and expertise developed at PNNL and EMSL in support of DOE-funded research in biofuels and bioremediation, which are also applicable to biological questions related to human health.

###

Interdisciplinary teams at Pacific Northwest National Laboratory address many of America's most pressing issues in energy, the environment and national security through advances in basic and applied science. PNNL employs 4,700 staff, has an annual budget of nearly \$1.1 billion, and has been managed for the U.S. Department of Energy by Ohio-based Battelle since the laboratory's inception in 1965. For more, visit the PNNL's News Center, or follow PNNL on Facebook, LinkedIn and Twitter.

EMSL, the Environmental Molecular Sciences Laboratory, is a national scientific user facility sponsored by the Department of Energy's Office of Science. Located at Pacific Northwest National Laboratory in Richland, Wash., EMSL offers an open, collaborative environment for scientific discovery to researchers around the world. Its integrated computational and experimental resources enable researchers to realize important scientific insights and create new technologies. Follow EMSL on Facebook, LinkedIn and Twitter.

Notes (ignore everything below this): <http://www.pnl.gov/news/release.asp?id=TK#>

<http://www.slideshare.net/pastinson/us-military-active-duty-demographic-profile-presentation>

Comment [t1]: If anyone knows the percentage of black women under 40 in the army or military, that would be great. But I'm not just multiplying 20 percent by 91 percent because there's no guarantee that black women are represented equally in both populations.

DRAFT NEWS RELEASE

Release date: Month TK

Contact: Mary Beckman, mary.beckman@pnnl.gov, (509) 375-3688

Collaborative study looks for clues on hard-to-treat breast cancer  
*PNNL, Walter Reed & Windber to scour triple negative breast cancer for leads on diagnostics and drugs*

RICHLAND, Wash. -- Some types of breast cancer can be successfully treated with specific drugs such as tamoxifen, but treatment for a type of breast cancer more common in young and black women is still limited to radiation and general chemotherapy. Called triple negative breast cancer, this type of cancer is the focus of a 20-month, \$8.6-million research project that aims to find new diagnostic tools and options for drugs.

The project takes advantage of one of the most comprehensive collections of breast cancer clinical samples in the U.S. -- the Clinical Breast Care Project located at the Walter Reed National Military Medical Center in Bethesda, Md., and the Windber Research Institute in Winder, Pa.

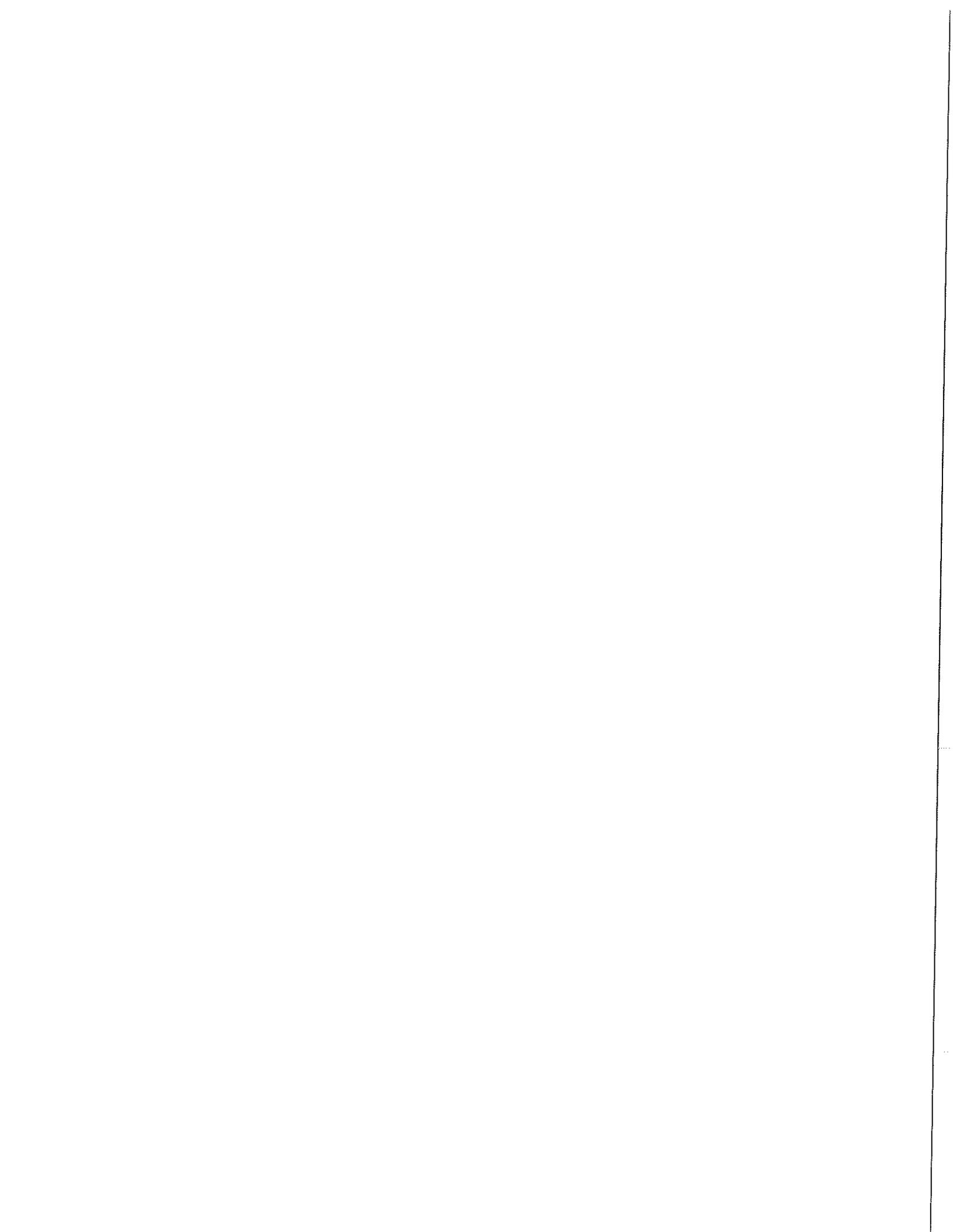
Researchers will explore these samples using advanced proteomics technology at the Department of Energy's Pacific Northwest National Laboratory and EMSL, DOE's Environmental Molecular Sciences Laboratory in Richland, Wash. Led by PNNL proteomics researcher Richard Smith, the study is funded by the Department of Defense.

"Triple negative cancers are more likely to hit young women and African American women. That's a health disparity issue. We need a better understanding of this disease," said team member Karin Rodland, a cancer biologist at PNNL. "And what's been holding that up has been getting enough samples to thoroughly examine how triple negative cancers operate."

Because the Army has such a large population of women that receive health care for years, as well as a higher percentage of blacks than the U.S. population, the Walter Reed-Windber breast cancer repository will provide many high quality samples with well-documented health histories.

One of the first things doctors check when a woman is diagnosed with breast cancer is whether her cancer will grow in response to any or all of three hormones: one that stimulates cell growth and two sex hormones, estrogen or progesterone -- cancers that do can be treated with particular drugs. But many other breast cancers don't respond. Called triple negative breast cancers, these types represent a wide variety of cancers and are typically more aggressive and harder to treat.

The research team will profile the complement of proteins -- known as the proteome -- that the breast cancer tissues produce, looking for proteins that triple negative cancers share. The shared proteins could suggest new options for drug therapies. In addition, comparing how aggressive the cancers are to the complement of proteins the cancers make or other metabolic products could lead to new diagnostic tools.



## Does Breast Tumor Heterogeneity Necessitate Further Immunohistochemical Staining on Surgical Specimens?

**Authors:** Lauren T. Greer, M.D., Martin Rosman, M.D., Charles Mylander, Ph.D., Albert J. Kovatich, M.S., Jeffrey Hooke, M.D., Wen Liang, D.O., Robert R. Buras, M.D., FACS, Craig D. Shriver, M.D., FACS, Lorraine Tafra, M.D., FACS

**Introduction:** Prognostic and predictive tumor markers (ER, PR, HER2, Ki-67) in breast cancer are most commonly performed on core needle biopsies (CNB) of the primary tumor. Because treatment recommendations are influenced by these markers, it is imperative to verify strong concordance between tumor markers derived from IHC testing on CNB specimens and the corresponding surgical specimens (SS).

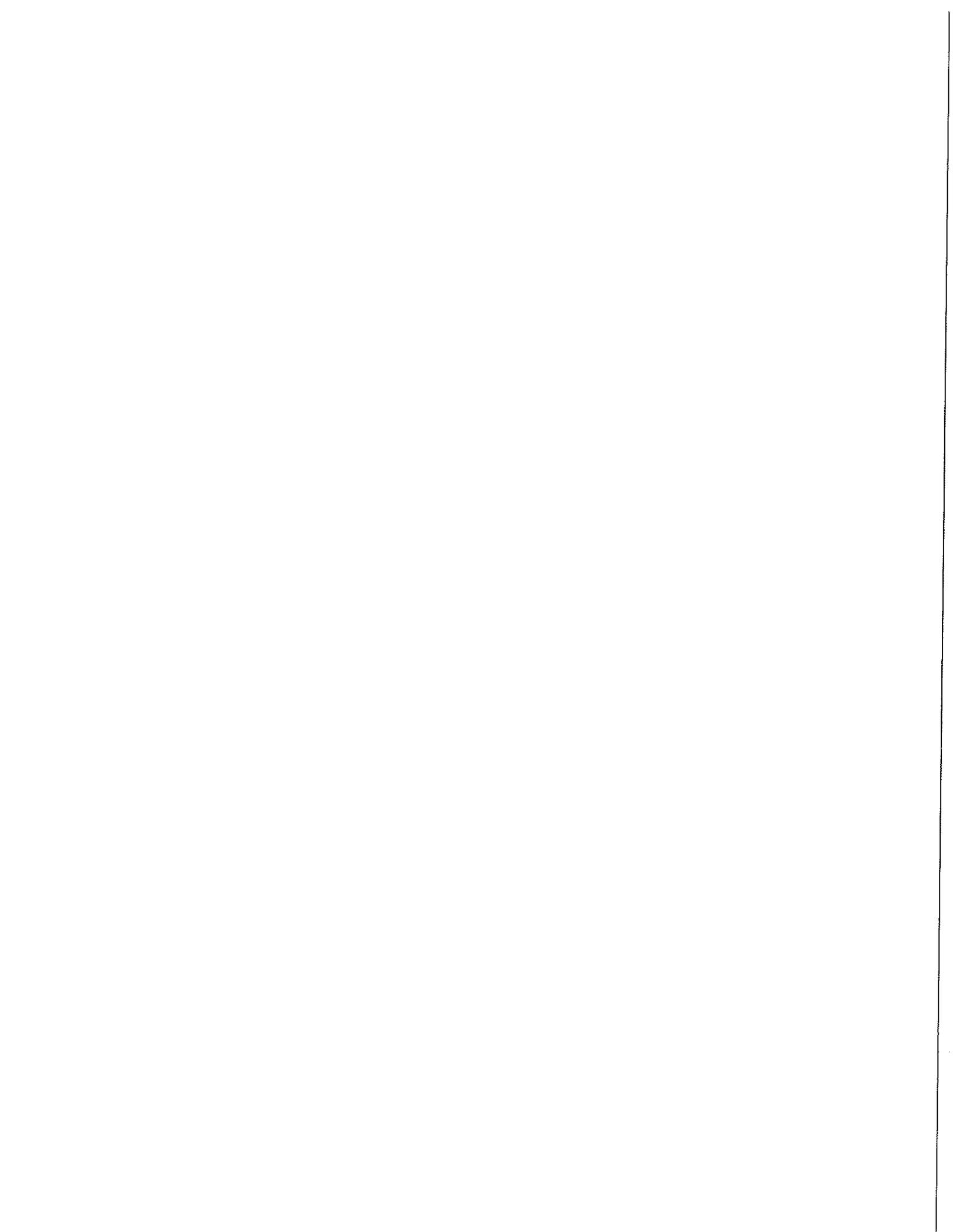
**Methods:** A prospective study was performed on 132 women (151 samples) with breast cancer diagnosed from January 2009 to March 2011. Tumor type, grade, ER, PR, HER2, Ki67 expression by IHC were retrospectively analyzed in the CNB and SS. Contingency tables and agreement modeling were performed to generate Kappa values. ER% and PR% were calculated based on positivity being  $\geq 1\%$ , HER2 was considered positive if CNB showed 2+(FISH positive) or 3+ but on SS was based on 3+ only (no FISH performed on SS).

**Results:** There was substantial agreement between the CNB and SS for tumor type, ER%, PR%, HER2; moderate agreement for grade; and fair agreement for Ki67% (see Chart). The CNB contained higher percentages of ER, PR and Ki67 than the SS. Distinct molecular subtypes (tumor heterogeneity) by H&E and IHC were noted in the SS of 14 subjects (11%). These patients had multiple IHC testing performed on different areas of tumor. In 3 of these subjects (21%), HER2 was found to be positive in an area of the SS but was negative on the CNB. No cases were seen where ER or PR were negative on CNB but became positive in the SS. However, 2 subjects in the heterogeneous group had distinctly different regions in their tumor where one was ER positive and another was ER negative.

**Conclusion:** The heterogeneous distribution of antigens in breast cancer tumors raises concern that the CNB may not adequately represent the true biologic profile in all patients. There is strong concordance for tumor type, ER, and PR between CNB and SS (although a decline was noted from CNB to SS); however, HER2 activity does not appear to be adequately detected on CNB in patients with heterogeneous tumors. This data highlights that the IHC testing on the CNB alone may not be adequate to tailor targeted therapy in all patients.

Biomarker	# Agreements	Kappa ( $\kappa$ )	Landis-Koch Agreement Grade	95% CI
Tumor Type	124/151 (82%)	0.66	Substantial	(0.55, 0.77)
Histologic Grade	84/133 (63%)	0.44	Moderate	(0.31, 0.56)
ER%	142/151 (94%)	0.73	Substantial	(0.56, 0.89)
PR%	134/151 (89%)	0.69	Substantial	(0.55, 0.82)
HER2	128/134 (96%)	0.68	Substantial	(0.43, 0.92)
Ki67 %	52/119 (44%)	0.21	Fair	(0.10, 0.31)

"The views expressed in this abstract are those of the authors and do not reflect the official policy of the Department of the Army (DOA), Department of Defense (DOD), or US Government."



## Evaluation of BRCA1 mutations in patients with family history of breast cancer

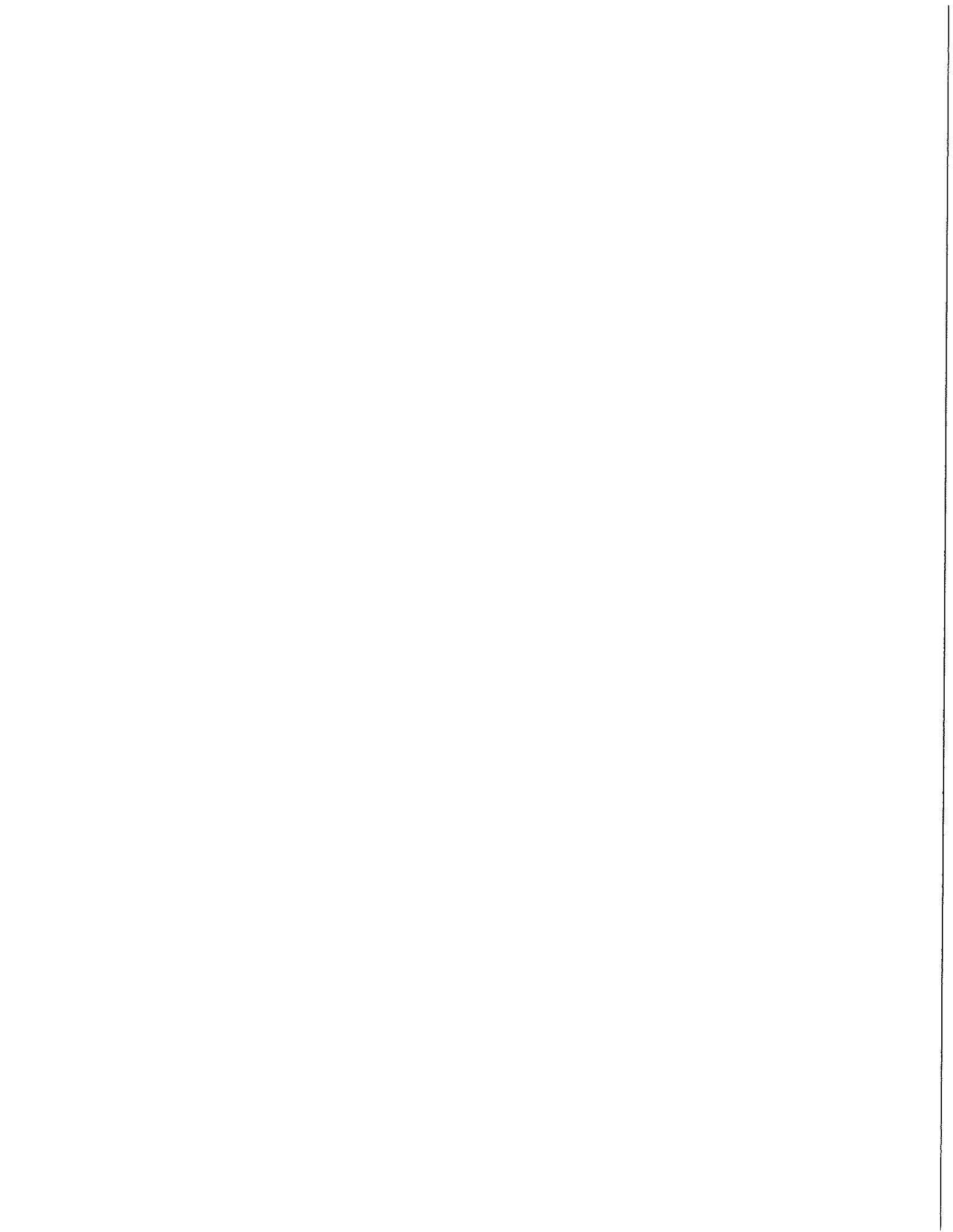
Seth Rummel, Craig Shriver, Rachel Ellsworth

**Background:** Identification of BRCA1 in 1994 has allowed patients at high-risk of developing breast cancer to be screened for causative mutations. Current criteria used to identify patients for genetic testing include family history of early-onset breast cancer with or without a family history of ovarian and/or male breast cancer, or a personal history of breast cancer at an early age. Here, we determined the prevalence of BRCA1 mutations in a cohort of female invasive breast cancer patients with a significant family history.

**Methods:** The Clinical Breast Care Project database was queried to identify all female patients with significant family histories defined as being diagnosed  $\leq 35$  years, diagnosed  $\leq 50$  years with at least one primary family member with breast and/or ovarian cancer, or diagnosed  $< 50$  years with at least two secondary family members from the same parental branch from two generations diagnosed with breast and/or ovarian cancer. Genomic DNA was isolated from blood and exonic regions of the BRCA1 genes were amplified and sequenced on an ABI3730xl. Sequence data was analyzed using Sequencher 4.10.1.

**Results:** Of the 115 women with significant family histories, three had known deleterious BRCA2 mutations and one had a mutation p53 with a family history of Li-Fraumeni syndrome; these patients were excluded from further analyses. The most common (49%) group of women with significantly family histories are those diagnosed between 35-50 years with a single primary family member with either pre- or post-menopausal breast or ovarian cancer. Luminal A was the most common tumor subtype (52%), followed by triple negative (32%), luminal B (12%) and HER2-enriched (4%). Of the eleven causative mutations detected, all were in patients with triple negative tumors: three diagnosed  $\leq 35$  years, six diagnosed 36-50 years with at least one primary relative with breast cancer and two diagnosed 36-50 with at least two secondary family members with breast cancer.

**Conclusion:** Presence of causative BRCA1 mutations did not segregate with early age of onset or number of relatives with breast cancer. Rather, only patients with triple negative breast disease harbored BRCA1 mutations. These results have important genetic counseling and risk reduction implications as the association between BRCA1 mutations and triple negative breast cancer suggest tumor subtype may be added to the current criteria used to identify patients who should be offered testing for BRCA1 mutations.



## CDH1 mutations in patients with lobular carcinoma of the breast

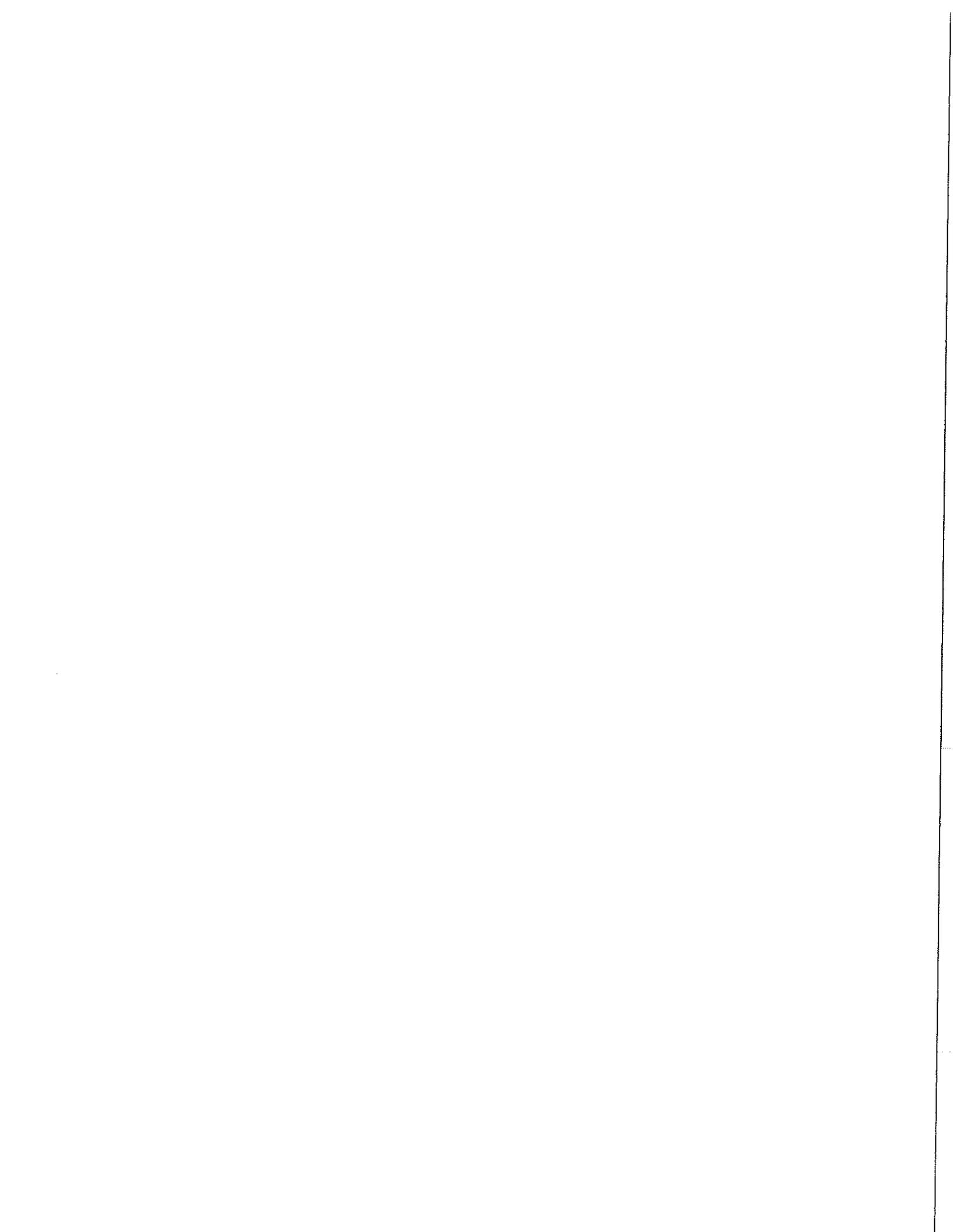
Allyson L. Valente, Seth Rummel, Craig D. Shriver, Rachel E. Ellsworth

**Background:** The E-cadherin (CDH1) gene is involved in cell adhesion and maintenance of tissue architecture, and loss of CDH1 has been detected in a number of cancers characterized by loose cell-cell adhesion such as hereditary diffuse gastric, endometrial, ovarian and invasive lobular carcinoma (ILCA) of the breast. In addition, germline mutations in CDH1 have been associated with increased risk of developing ILCA but not other breast tumor types. To improve understanding of the role of CDH1 mutations in breast cancer, mutation status was determined in a breast cancer cohort of patients with ILCA with and without family history of other tumor types.

**Methods:** The Clinical Breast Care Project database was queried to identify all patients with ILCA with or without a family history of gastric/stomach, endometrial or ovarian cancer. Genomic DNA was isolated from peripheral blood samples. DNA variants were detected for each exon of CDH1 using high-resolution melting technology and the underlying change identified by direct sequencing. Variant A617T was assayed using TaqMan SNP assay rs33935154.

**Results:** Of the 72 patients with ILCA, five had a family history of stomach cancer. Two private mutations were detected: P825Q in a patient with a family history of stomach cancer, and G879S in a patient with no family history of cancer. In addition, the A592T variant, which has been reported as non-pathogenic, was detected in a Hispanic woman while the causative mutations 1137G-A, identified in a Caucasian woman, and A617T identified in two African American women were detected. The A617T mutation was then evaluated in an additional 166 African American patients: 12 patients carried the mutation, nine of which had IDCA, two with microinvasive disease and one with mixed ductal and lobular features.

**Conclusions:** The frequency of CDH1 mutations in this patient population was low (<10%), with only two mutations with known clinical consequences identified. In addition, the detection of the A617T mutation previously associated with endometrial cancer in African American women, in patients with predominantly ductal carcinomas questions whether this mutation is associated with increased risk of ILCA. The low frequency of germline mutations in CDH1 suggests that inherited mutations are not the predominant mechanism but rather alternate methods, such as epigenetic modification or miRNA silencing, may be driving the suppression of E-cadherin expression and the development of ILCA.



**Brenda Deyarmin**, Windber Research Institute; **Jennifer L. Kane**, Windber Research Institute; **Allyson L. Valente**, Windber Research Institute; **Ryan van Laar**, PhD, ChipDx LLC; **Christopher Gallagher**, MD, Walter Reed National Military Medical Center; **Craig D. Shriver**, MD, Walter Reed National Military Medical Center; **Rachel E. Ellsworth**, PhD, Henry M. Jackson Foundation for the Advancement of Military Medicine

Corresponding author:

Rachel E. Ellsworth

Clinical Breast Care Project

620 Seventh Street

Windber, PA 15963

[r.ellsworth@wriwindber.org](mailto:r.ellsworth@wriwindber.org)

Phone: (814) 361-6926

FAX: (814) 46706334

**Title:** Effect of ASCO/CAP guidelines for determining ER status on molecular subtype

**Shortened title:** ASCO guidelines and molecular subtype

**Disclosures:** The authors have no conflicts to declare. This research was supported by a grant from the United States Department of Defense (Military Molecular Medicine Initiative MDA W81XWH-05-2-0075, Protocol 01-20006).

**Synopsis**

Low staining (ER positive cells <10%) are either basal-like or HER2-enriched, both subtypes considered to be ER negative. The ASCO/CAP guidelines that use a threshold of >1% positively staining cells to define ER positive status may not accurately reflect the underlying molecular behavior of the tumor and a value of <10% may be more accurate.

## **ABSTRACT**

**Background:** Determination of ER, PR and HER2 status are standard for predicting prognosis and determining treatment options for patients with breast cancer. In 2010, the American Society of Clinical Oncology (ASCO) and College of the American Pathologists (CAP) issued guidelines that tumors with  $\geq 1\%$  positively staining cells should be considered ER positive. Here, we determined how this cutoff relates to molecular subtype.

**Methods:** Gene expression data were generated for 299 primary breast tumors and intrinsic subtype determined (ChipDx LLC). IHC-defined surrogate subtypes using the new ASCO/CAP guidelines were compared to molecular subtypes.

**Results:** Tumors with low ER staining (1-10%) were more clinicopathologically similar to ER negative (ER staining  $< 1\%$ ) than ER positive (ER staining  $> 10\%$ ) tumors. At the molecular level, 88% of low ER staining tumors were classified as either basal-like or HER2-enriched. Only those tumors expressing 10% ER positive cells demonstrated the ER positive luminal A subtype.

**Conclusions:** Under the new ASCO/CAP guidelines, tumors with ER staining 1-10% would be classified as ER positive, yet most are basal-like or HER2-enriched and have similar pathological features, including mortality rates, with ER negative tumors.

Although it is possible that these patients may derive benefit from endocrine treatment, classification of these tumors as ER positive would preclude the use of treatments such as PARP inhibitors in patients with molecularly-defined basal-like breast tumors. As ER status is a critical element in the choice of treatments for patients with breast cancer, it is imperative that the most effective method for classifying tumors be developed.

## INTRODUCTION

Hormone receptor status is one of the most important clinical characteristics in predicting outcome and guiding treatment options for patients with invasive breast cancer. The estrogen receptor (ER) plays a critical role in tumorigenesis as binding of estrogen to ER leads to altered expression of genes involved in cell growth, differentiation, apoptosis and angiogenesis [1]. Hormone receptor status can be used as a prognostic factor with positive receptor status inversely associated with cancer-related mortality, and patients with positive ER and progesterone receptor (PR) status having 1.2-2.1-fold lower risk than patients with ER+/PR- or ER-/PR+ tumors and 2.1-2.6-fold lower risks than patients with or ER-/PR- tumors [2, 3]. The anti-estrogen drug tamoxifen, which blocks the action of estrogen by binding to the ER, represents one of the first successful targeted treatments for cancer. Meta-analysis of 55 clinical trials demonstrated that tamoxifen use in patients with ER positive tumors can decrease recurrence and improve mortality, however, patients with “ER poor” tumors (zero to low expression of ER) did not derive benefit from short- or long-term tamoxifen use [4]. These data underscore the importance of correctly determining ER status.

A number of methods have been developed to determine hormone receptor status. Immunohistochemical (IHC) analysis is currently the most widely used technology for evaluating hormone receptor status. IHC has advantages over older techniques such as the ligand-binding assay, as it is less labor intensive and expensive, does not require the use of radioactivity, preserves morphology, minimizes the effect of stromal contamination, and can be used to assess small tumors and retrospectively collected formalin-fixed, paraffin-embedded specimens [5]. Despite its widespread use, however,

IHC analysis is not quantitative and can be influenced by a number of pre-analytical and analytical variables, as well as subjectivity in interpretation and the use of different scoring systems [6, 7].

Determination of what percentage of positively staining cells should be used to classify tumors as ER positive or ER negative also remains problematic: cutoff values have ranged from 1% to 20% stained cells defining positive receptor status. In 2005, the St. Gallen guidelines suggested three categories for scoring ER status: endocrine responsive, with strong expression of ER, endocrine response uncertain, with low expression of ER, and endocrine non-responsive, with no expression of ER. In developing this three-tiered classification, an exact cutoff to differentiate between strong and low ER expression was not provided although the authors suggest that tumors with 1-10% positive cells are “usually considered” as having low ER expression [8]. In an effort to improve the accuracy of hormone receptor testing by IHC, the American Society of Clinical Oncology (ASCO) and College of American Pathologists (CAP) recommended that a cutoff of 1% positive cells be used to define ER positive status [9]. In this study, we determined the intrinsic subtype of tumors with low ER expression (1-10%) to determine whether using a cutoff of 1% to define positive ER status accurately reflects the molecular biology of the tumor.

## **MATERIALS AND METHODS**

### *Human tissue samples*

Enrollment of patients into the Clinical Breast Care Project (CBCP) began in 2001. For inclusion in the CBCP, all patients must have met the following criteria: 1)

adult over the age of 18 years, 2) mentally competent and willing to provide informed consent, and 3) presenting to the breast centers with evidence of possible breast disease, for routine screening mammograms or elective reductive mammoplasty. Tissue and blood samples were collected with approval from the Walter Reed Army Medical Center Human Use Committee and Institutional Review Board. All subjects enrolled in the CBCP voluntarily agreed to participate and gave written informed consent. Clinical information was obtained for all CBCP samples using questionnaires designed by and administered under the auspices of the CBCP.

Tissue was collected from patients undergoing surgical procedures, including lumpectomy or mastectomy. Within 5-15 minutes of surgical removal, breast tissue was taken on crushed, wet ice to the pathology laboratory where a licensed pathologist or pathologists' assistant performed routine pathology analyses. Diagnosis of every specimen was performed by a breast pathologist from hematoxylin and eosin (H&E) stained slides; staging was performed using guidelines defined by the *AJCC Cancer Staging Manual* seventh edition [10] and grade assigned using the Nottingham Histologic Score [11, 12]. ER and PR status were determined by IHC analysis at a clinical laboratory (MDR Global, Windber, PA) and the percent stained cells were recorded. For HER2 status, IHC analysis was performed in the same clinical laboratory as ER and PR status (MDR Global, Windber, PA); cases with HER2 scores = 2+ were then evaluated by fluorescence *in situ* hybridization using the PathVysion® HER-2 DNA Probe kit (Abbott Laboratories, Abbott Park, IL).

*RNA Isolation, Amplification, aRNA Labeling and Hybridization*

For each tumor with frozen tissue available, H&E stained slides were examined by a breast pathologist and tumor areas marked for laser microdissection. Two to twelve serial sections (8  $\mu\text{m}$  thick) were cut, mounted on glass PEN foil slides (Leica Microsystems, Wetzlar, Germany), stained using the LCM staining kit (Applied Biosystems, Foster City, CA) and laser microdissected on an *ASLMD* laser microdissection system (Leica Microsystems, Wetzlar, Germany). Slide preparation, staining and cutting were performed within 15 minutes to preserve RNA integrity.

RNA was isolated from laser microdissected tumor cells using the RNAqueous-Micro kit (Applied Biosystems, Foster City, CA) and treated with DNase I to remove any contaminating genomic DNA. RNA integrity was assessed using the 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA), converted to biotin-labeled aRNA using two rounds of amplification with the MessageAmpII aRNA Amplification kit (Applied Biosystems, Foster City, CA), and the concentration and quality of the samples was measured with the NanoDrop 1000 (NanoDrop Products, Wilmington, DE) and 2100 Bioanalyzer. Hybridization, washing, staining and scanning were performed using HG U133A 2.0 arrays (Affymetrix, Santa Clara, CA) according to manufacturer's protocols.

### *Molecular Subtyping*

GeneChip CEL files were processed and scaled using the MAS5 algorithm and analyzed using BreastPRST<sup>TM</sup> (Signal Genetics, New York, NY) which uses separate genomic signatures to assess of recurrence risk (prognosis), molecular grade and subtype [13]. Samples were classified as one of five subtypes using a two-step process; first a two-class, 100-gene support vector machine (SVM) algorithm was applied to stratify all

patients as normal-like or 'other'. This classifier was developed by using recursive feature elimination on a randomly partitioned training series, derived from a previously published study [14]. A set of 100 genes were selected that were able to classify normal-like versus other breast tumors with 100% and 95% accuracy in cross validation and independent validation series, respectively. Next those samples classified as 'other' were further analyzed using a four-class, 180-gene nearest centroid (mean) algorithm. This method was developed by cross validation of all non-normal-like tumors from the same previously published study to determine the gene set that resulted in the highest agreement with the existing molecular subtype labels of each sample. Genes differentially expressed between subtypes at  $P \leq 1 \times 10^{-5}$  were identified as the optimal selection threshold for classification purposes.

Mean cross validation results for the 180-gene nearest centroid algorithm were 90% sensitivity, 97% specificity, 89% positive predictive value and 97% negative predictive value, across the four classes. The algorithm was then applied to a number of independent published datasets for which predicted molecular subtype information was available. Comparison of BreastPRS molecular subtype assignments with those generated from other methods, resulted in high concordance (>90%) in all cases.

## **RESULTS**

### *Sample Characteristics*

In the ten year period between 2001 and 2011, 1,459 women with invasive breast cancer enrolled in the CBCP. Of these, 221 were classified as ER positive or negative

but a percent of stained cells was not provided. Within the 1,238 tumors with staining information, 19% would be defined as ER negative (ER <1%) using the ASCO/CAP guidelines. Within the ER positive group, 5% had low (1-10%) levels of ER staining (Figure 1).

Clinical characteristics for the four groups are shown in Table 1. Clinical characteristics differed significantly between low staining and ER negative tumors only for tumor grade. Further evaluation of tumor grade revealed that 6/8 (75%) of the low-staining well-differentiated tumors had 10% positively staining cells. When comparing low ER tumors to tumors with >10% ER staining, low-staining tumors had significantly less favorable characteristics for age at diagnosis, ethnicity, tumor stage and grade and HER2 status.

### *Molecular Subtypes*

Microarray data was generated from 70 tumors with 0% ER staining. Of these, 99% were of either the HER2-enriched (16%) or basal-like (83%) subtypes, both of which are considered ER negative. Tissue was available for 26 low-staining tumors (Table 2). Under the ASCO/CAP guidelines, all low-staining tumors would be considered ER positive, however, 62% were classified as basal-like and 27% as HER2-enriched, both are considered ER negative subtypes. Of note, the three luminal A tumors had an ER staining value of 10%, while all of the basal-like and all but one of the HER2-enriched tumors had ER staining <10%.

### *Treatment and Clinical Outcome*

A chart review revealed that 47% patients with low ER staining tumors were treated with tamoxifen, anastrozole, letrozole, or exemestane. Patient outcomes did not differ significantly between patients who were treated compared to those not treated with endocrine therapy.

Differences in mortality rates were statistically significant between ER positive (1.3%) and low-staining (7.5%) tumors ( $P < 0.001$ ), but not between the low-staining and ER negative (8.6%) tumors. None of the patients ( $n=10$ ) with 10% staining have died of disease and one has since been diagnosed with distant metastasis. In contrast, 9% of low-staining patients with ER  $< 10\%$  ( $n=44$ ) died of disease and two patients have progressed.

## **DISCUSSION**

The ASCO/CAP guidelines were issued to standardize determination of IHC-determined hormone receptor status, thus improving the prognostic and predictive ability of these markers. The threshold of  $\geq 1\%$  positive staining cells to define positive ER status was selected after a number of studies reported that patients with ER staining as low as 1% could achieve meaningful clinical responses when treated with endocrine therapies [9]. Although these ASCO/CAP recommendations maximize the number of women who are eligible for hormone therapy, our data suggests that the threshold of  $\geq 1\%$  stained cells defining ER positive status does not accurately reflect the underlying molecular behavior of the tumor. Every low-staining tumor subjected to molecular subtyping in our study with ER staining  $< 10\%$  was either basal-like (70%) or HER2-enriched (30%); of those tumors with 10% staining, 75% were luminal A. We therefore suggest that  $< 10\%$  ER staining is a more accurate threshold for defining ER status.

Intrinsic subtypes can provide prognostic information beyond that of hormone receptor status. The ER positive luminal subtypes can be divided into the highly ER positive luminal A group which also has low expression of HER2 and proliferation genes, and the low ER positive luminal B group with high expression of proliferation genes and variable expression of HER2. The ER negative subtypes include basal-like and HER2-enriched tumors with the HER2-enriched group having high expression of HER2 and low expression of ER and luminal genes and the basal-like subtype having low expression of HER2, ER and luminal genes [15, 16]. Basal-like and HER2-enriched subtypes tend to relapse early, during the first five years, and have less favorable prognosis than the luminal subtypes [16, 17]. Each subtype also has different preferential sites of metastasis: luminal tumors metastasize most frequently to bone while HER2-enriched and basal-like metastasize to brain, bone and lung [17]. As seen in Table 2, all tumors with ER expression <10% had either the basal-like or HER2-enriched subtype, yet under the ASCO/CAP guidelines, these tumors would now be considered ER positive/luminal. This discrepancy could have significant impact on patient prognosis and management, especially in the basal-like patients, by upgrading these low staining basal-like tumors to the most favorable luminal A subtype.

Tumor classification can also impact treatment. Patients with positive ER status can be treated with hormonal therapies such as tamoxifen or aromatase inhibitors that target the ER pathway. The ASCO/CAP Expert Panel evaluated data from numerous studies that, when evaluating the efficacy of endocrine therapy in patients whose tumors were evaluated by IHC analysis, included patients with low ER staining tumors in their ER positive tumor groups. The number of patients with low ER staining in these studies,

was, however, small. A recent study evaluated outcome and response to hormone treatment in a larger number of patients with low ER staining, including 897, 241 and 119 tumors with 0%, 1-5% and 6-10% ER staining, respectively. Similar to data from our study, clinicopathological characteristics did not differ significantly between groups, except that patients with 0% staining had a higher frequency of high-grade tumors. Recurrence-free survival and overall survival did not differ between the three groups. Within the subgroup of patients who received hormonal therapy, overall survival did not differ between the three patient groups and use of endocrine treatments had a marginal impact in the 6-10% group compared to the 0% and 1-5% groups [18].

Special consideration must be given to basal-like tumors which represent >60% of the low staining tumors in our series. Historically, patients with basal-like tumors, which have little to no expression of ER and HER2, have not been eligible for targeted treatments such as hormonal agents or trastuzumab, leaving chemotherapy as the only option. More recently, new treatment approaches have been developed that exploit the underlying genomic instability characterizing some basal-like tumors, especially those harboring BRCA1 mutations. In tumor cells with DNA repair deficiencies, treatment with poly(ADP-ribose) polymerase (PARP) inhibitors causes additional DNA repair damage, leading to synthetic lethality of the tumor cells [19]. Clinical trials are underway to determine the efficacy of PARP inhibitors in patients with triple negative and/or basal-like breast tumors with and without BRCA1 or BRCA2 mutations [20-23]. As new agents are developed that provide tailored treatment options to patients with basal-like breast tumors, it is critical to correctly classify patients so that they receive the optimum treatment options.

In conclusion, although a number of molecular assays, such as the PAM50 Breast Cancer Intrinsic Subtype Classifier (ARUP Laboratories Salt Lake City, UT) are now commercially available, these tests are not used routinely in clinical practice [24] . Because IHC analysis remains the standard prognostic approach, it is critical to determine the correct cutoff value for determining ER status. Here, all tumors with <10% ER staining were either basal-like or HER2-enriched; moreover, these low staining ER tumors were clinically and pathologically more similar to ER negative than ER positive tumors. These data suggest that ASCO/CAP use of  $\leq 1\%$  staining cells to define positive ER status may not accurately reflect the underlying molecular biology of these tumors, thereby precluding patients from receiving appropriately tailored treatment options.

## **ACKNOWLEDGEMENTS**

This research was supported by a grant from the United States Department of Defense (Military Molecular Medicine Initiative MDA W81XWH-05-2-0075, Protocol 01-20006). The opinion and assertions contained herein are the private views of the authors and are not to be construed as official or as representing the views of the Department of the Army or the Department of Defense.

## **FIGURE LEGENDS**

Figure 1. H&E images of low ER staining tumors at 20X magnification. Figure 1a is from patient 262 who was diagnosed with a stage IIA, HER2 positive, poorly-differentiated invasive ductal carcinoma with 1-5% ER positive staining cells. Figure 1b is from patient 107 who was diagnosed with stage I, HER2 positive, moderately-differentiated invasive ductal carcinoma with 6-10% positive staining cells.

## REFERENCES

- (1) Pietras RJ, Marquez-Garban DC. Membrane-associated estrogen receptor signaling pathways in human cancers. *Clin Cancer Res* 2007;13:4672-4676.
- (2) Aaltomaa S, Lipponen P, Eskelinen M et al. Hormone receptors as prognostic factors in female breast cancer. *Ann Med* 1991;23(6):643-648.
- (3) Dunnwald LK, Rossing MA, Li CI. Hormone receptor status, tumor characteristics, and prognosis: a prospective cohort of breast cancer patients. *Breast Cancer Res* 2007;9:R6.
- (4) Early Breast Cancer Trialists' Collaborative Group (EBCTCG). Tamoxifen for early breast cancer: an overview of the randomised trials. *Lancet* 1998;351(9114):1451-1467.
- (5) Regan MM, Viale G, Mastropasqua MG et al. Re-evaluating adjuvant breast cancer trials: assessing hormone receptor status by immunohistochemical versus extraction assays. *J Natl Cancer Inst* 2006;98(21):1571-1581.
- (6) Osborne CK. Steroid hormone receptors in breast cancer management. *Breast Cancer Res Treat* 1998;51:227-238.
- (7) Engel KB, Moore HM. Effects of preanalytical variables on the detection of proteins by immunohistochemistry in formalin-fixed, paraffin-embedded tissue. *Arch Pathol Lab Med* 2011;135:537-543.
- (8) Goldhirsch A, Glick JH, Gelber RD et al. Meeting highlights: international expert consensus on the primary therapy of early breast cancer 2005. *Ann Oncol* 2005;16(10):1569-1583.
- (9) Hammond MEH, Hayes DF, Dowsett M et al. American Society of Clinical Oncology/College of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer. *J Clin Oncol* 2010;28(16):2784-2795.
- (10) American Joint Committee on Cancer. *AJCC Cancer Staging Manual*. 7 ed. New York: Springer; 2010.
- (11) Bloom HJ, Richardson WW. Histological grading and prognosis in breast cancer. *Br J Cancer* 1957;11:359-377.
- (12) Elston CW, Ellis IO. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow up. *Histopathology* 1991;19:403-410.

- (13) van Laar R. Design and multiseres validation of a web-based gene expression assay for predicting breast cancer recurrence and patient survival. *J Mol Diagn* 2011;13(3):297-304.
- (14) Sabatier R, Finetti P, Cervera N et al. A gene expression signature identifies two prognostic subgroups of basal breast cancer. *Breast Cancer Res Treat* 2011;126(2):407-420.
- (15) Perou CM, Sorlie T, Eisen MB et al. Molecular portraits of human breast tumours. *Nature* 2000;406(6797):747-752.
- (16) Sorlie T, Perou CM, Tibshirania R et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci USA* 2001;98(19):10869-10874.
- (17) Kennecke H, Yerushalmi R, Woods R et al. Metastatic behavior of breast cancer subtypes. *J Clin Oncol* 2010;28(20):3271-3277.
- (18) Raghav KPS, Hernandez-Aya LF, Lei X et al. Impact of low estrogen/progesterone receptor expression on survival outcomes in breast cancers previously classified as triple negative breast cancers. *Cancer* 2011;epub ahead of print.
- (19) Ashworth A. A synthetic lethal therapeutic approach: poly(ADP) ribose polymerase inhibitors for the treatment of cancers deficient in DNA double-strand break repair. *J Clin Oncol* 2008;26:3785-3790.
- (20) Fong PC, Boss DS, Yap TA et al. Inhibition of poly(ADP-ribose) polymerase in tumors from BRCA mutation carriers. *N Engl J Med* 2009;361(2):123-134.
- (21) O'Shaughnessy J, Osborne C, Pippen JE et al. Iniparib plus chemotherapy in metastatic triple-negative breast cancer. *N Engl J Med* 2011;364(3):205-214.
- (22) Tutt A, Robson M, Garber JE et al. Oral poly(ADP-ribose) polymerase inhibitor olaparib in patients with BRCA1 or BRCA2 mutations and advanced breast cancer: a proof-of-concept trial. *Lancet* 2010;376(9737):235-244.
- (23) Gelmon KA, Tischkowitz M, Mackay H et al. Olaparib in patients with recurrent high-grade serous or poorly differentiated ovarian carcinoma or triple-negative breast cancer: a phase 2, multicentre, open-label, non-randomised study. *Lancet Oncol* 2011;12(9):852-861.
- (24) Paik S. Is gene array testing to be considered routine now? *Breast* 2011;20(Suppl 3):S87-S91.
- (25) Carey LA, Perou CM, Livasy CA, et al. Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. *JAMA* 2006;295:2492-502.

Table 1. Clinical and pathological features of 1,238 invasive breast tumor specimens. Negative = 0% staining cells, Low-staining = 1-10%, Positive = >10%. P-values listed as NS (not significant) >0.05.

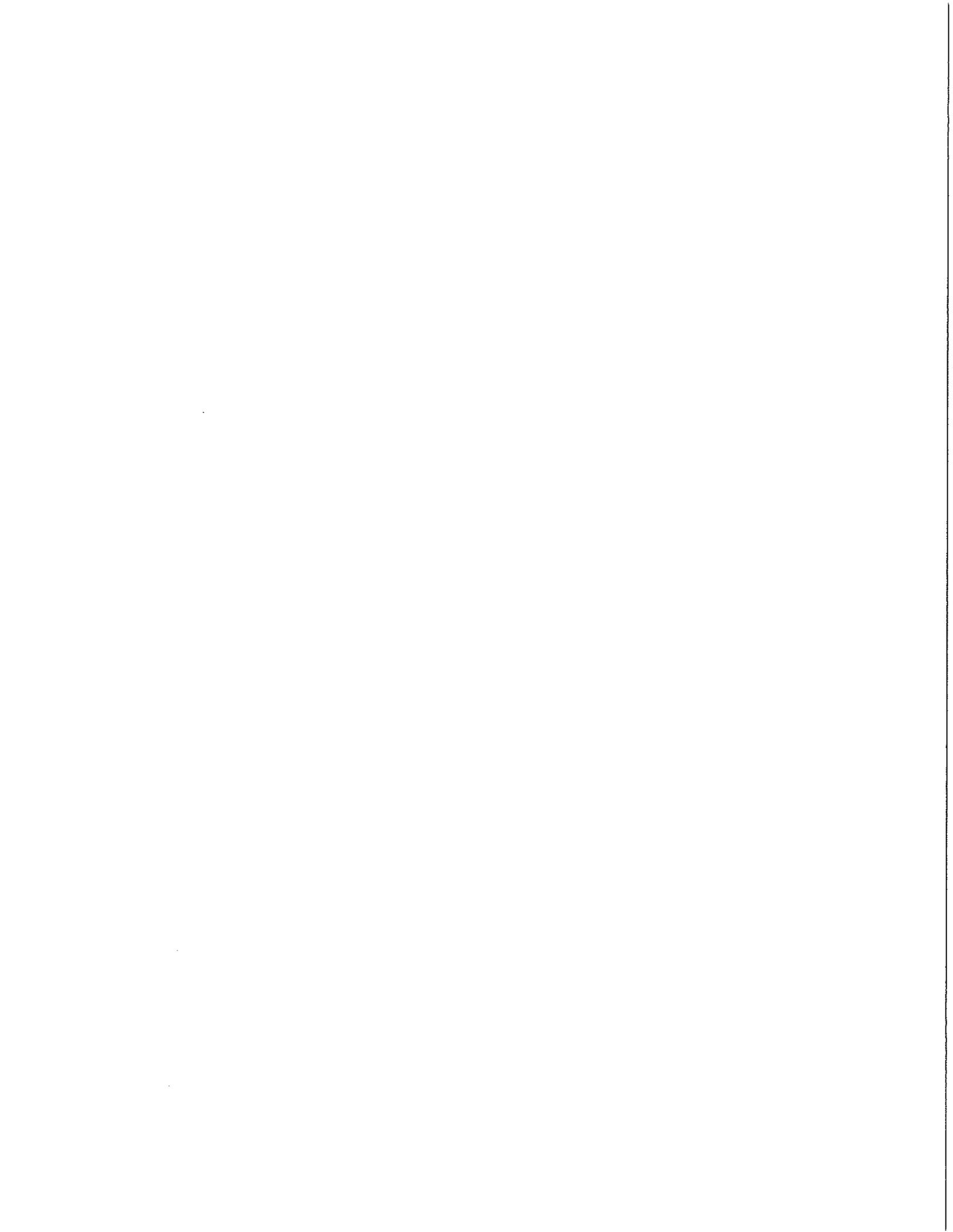
	Negative (n=233)	Low- staining (n=54)	Positive (n=951)	<i>P</i> -value Negative vs. Low- staining	<i>P</i> -value Positive vs. Low- staining
<b><i>Age at diagnosis</i></b>				NS	0.004
<40 years	9%	14%	4%		
40-49 years	30%	24%	20%		
≥50 years	61%	62%	76%		
<b><i>Ethnicity</i></b>				NS	0.002
White	71%	61%	83%		
African American	25%	29%	13%		
Asian	1%	4%	2%		
Hispanic	1%	6%	2%		
Other	2%	0%	0%		
<b><i>Tumor size</i></b>				NS	NS
T1	49%	60%	71%		
T2	41%	30%	24%		
T3	10%	10%	5%		
<b><i>TNM Stage</i></b>				NS	0.001
I	37%	42%	58%		
II	41%	38%	30%		
III	18%	9%	10%		
IV	4%	11%	2%		
<b><i>Tumor Grade</i></b>				>0.001	>0.001
Low	3%	16%	36%		
Moderate	14%	24%	45%		
High	83%	60%	19%		
<b><i>Lymph Node Status</i></b>				NS	NS
Negative	61%	65%	71%		
Positive	39%	35%	29%		
<b><i>HER2 Status</i></b>				NS	0.003
Negative	71%	76%	88%		
Positive	29%	24%	12%		

Table 2. Pathological characteristics of low ER expressing tumors with microarray data available. IHC surrogate subtypes were defined as luminal A = ER and or PR+/HER2-; luminal B = ER and or/PR+/HER2+; HER2-enriched = ER and PR-/HER2+ and triple negative = ER, PR and HER2- [25].

Sample	ER Group <sup>a</sup>	PR status <sup>b</sup>	HER2 Status	IHC subtype	Molecular subtype
22	1-5%	+	-	Luminal A	Basal-like
23	10%	+	-	Luminal A	Luminal A
43	5%	-	-	Luminal A	Basal-like
61	1%	+	-	Luminal A	Basal-like
98	1-5%	+	+	Luminal B	HER2-enriched
100	7%	-	-	Luminal A	Basal-like
102	10%	+	-	Luminal A	Luminal A
107	8%	-	+	Luminal B	HER2-enriched
120	1-5%	+	-	Luminal A	Basal-like
121	7%	+	-	Luminal A	Basal-like
157	10%	-	+	Luminal B	HER2-enriched
159	10%	-	-	Luminal A	Luminal A
171	1-5%	+	-	Luminal A	Basal-like
174	1-5%	+	-	Luminal A	Basal-like
176	3%	-	-	Luminal A	Basal-like
189	1-5%	-	-	Luminal A	Basal-like
257	1-5%	-	+	Luminal B	HER2-enriched
261	2%	-	-	Luminal A	Basal-like
262	1-5%	+	+	Luminal B	HER2-enriched
291	1-5%	-	-	Luminal A	Basal-like
292	4%	-	+	Luminal B	HER2-enriched
298	1-5%	+	-	Luminal A	Basal-like
306	2%	-	+	Luminal B	HER2-enriched
307	6%	-	-	Luminal A	Basal-like
308	8%	-	-	Luminal A	Basal-like
309	1%	-	-	Luminal A	Basal-like

<sup>a</sup>Actual percent staining is reported for those cases where the absolute value was available.

<sup>b</sup>PR values reported here are classified according to the ASCO/CAP guidelines where tumors with PR staining >1% are PR positive.



Genomic (in)stability of the breast tumor microenvironment

Seth Rummel<sup>1</sup>, Allyson L. Valente<sup>1</sup>, Jennifer L. Kane<sup>1</sup>, Craig D. Shriver<sup>2</sup>, Rachel E. Ellsworth<sup>3</sup>

From the Clinical Breast Care Project at the <sup>1</sup>Windber Research Institute, 620 Seventh Street, Windber, PA 15963 (SR: [s.rummel@wriwindber.org](mailto:s.rummel@wriwindber.org); ALV: [a.valente@wriwindber.org](mailto:a.valente@wriwindber.org); JLK: [j.kane@wriwindber.org](mailto:j.kane@wriwindber.org); DLE: [d.ellsworth@wriwindber.org](mailto:d.ellsworth@wriwindber.org)); <sup>2</sup>Walter Reed Army Medical Center, 6900 Georgia Avenue NW, Washington, DC 20307 (CDS: [craig.shriver@us.army.mil](mailto:craig.shriver@us.army.mil)); <sup>3</sup>Henry M. Jackson Foundation for the Advancement of Military Medicine, 620 Seventh Street, Windber, PA 15963 (REE: [r.ellsworth@wriwindber.org](mailto:r.ellsworth@wriwindber.org))

Corresponding Author:

Rachel E. Ellsworth

Clinical Breast Care Project

Henry M. Jackson Foundation for the Advancement of Military Medicine

620 Seventh Street, Windber, PA 15963

Phone: (814) 361-6926

FAX: (814) 467-6334

Email: [r.ellsworth@wriwindber.org](mailto:r.ellsworth@wriwindber.org)

## **Abstract**

**Introduction:** The breast tumor microenvironment plays an active role in tumorigenesis. Molecular alterations, including epigenetic modifications to DNA, and changes in RNA and protein expression have been identified in tumor-associated stroma; however, there is considerable debate as to whether the stroma is characterized by genomic instability or whether detection of chromosomal alterations in the breast stroma is a reflection of technological artifact rather than the true genomic content of the tumor microenvironment.

**Methods:** Surgically-removed breast stroma specimens from 112 women undergoing reductive mammoplasty (n=7), prophylactic mastectomy (N=6) or mastectomy for a diagnosis of breast disease (n=99) were frozen in optimal cutting temperature medium. Allelic imbalance (AI) analysis was performed in 484 stromal specimens from 98 women using a panel of 52 microsatellite markers; SNP data was generated from a subset of 86 stromal specimens using 250K SNP arrays (Affymetrix). Copy number alterations were identified using Partek Genomics Suite.

**Results:** AI was not detected in 92% (444/484) of stroma specimens. When compared to previously generated AI data from 77 formalin-fixed, paraffin-embedded stroma specimens, 32 (42%) of which harbored at least one detectable AI event, the frequency of AI in the FFPE specimens (4.62%) was significantly higher ( $P < 0.0001$ ) than that found in frozen specimens (0.45%). Of the stroma specimens assayed using SNP arrays 95% (82/86) had no detectable alterations and the 11 copy number changes were small and not shared between specimens.

**Conclusions:** The data presented here support a model in which the tumor microenvironment is genetically stable. The direct comparison of copy number alterations between FFPE and frozen research-grade specimens using identical methodologies suggests that past reports of significant AI in breast stroma, both adjacent to and distant from the tumor, reflects artifact in the archival specimens caused by formalin-fixation, paraffin-embedding and tissue storage.

## INTRODUCTION

Research in the past decade has increased our understanding of how the breast microenvironment influences tumor development. Stroma, comprised of fibroblasts, endothelial, smooth muscle, inflammatory and nerve cells, adipocytes, and macromolecules of the extracellular matrix (ECM), serves a supportive role to epithelial cells [1]. A number of physiological changes have been detected in stroma surrounding epithelial neoplasms including altered migratory potential and changes in expression of growth factors and cellular polarity, degradation of the ECM and increased angiogenesis [1-5].

One of the hallmarks of breast carcinogenesis is global genomic instability, which may include copy number gains or losses of whole or parts of chromosomes. Over 10 years ago, chromosomal alterations were detected as loss of heterozygosity (LOH) or allelic imbalance (AI) events in morphologically normal terminal ductal lobular units (TDLUs) adjacent to breast carcinomas [8]. A number of additional studies, the majority of which were performed using microsatellites markers and formalin-fixed paraffin-embedded (FFPE) archival specimens, have also detected genomic instability in tumor-associated stroma, furthering the notion that the tumor microenvironment is a dynamic entity that contributes to tumorigenesis through morphological and molecular alterations [9-15]. In contrast, recent studies using DNA from fresh frozen tissues that failed to detect chromosomal changes in stromal cell, leading some investigators to speculate that the LOH/AI studies performed on DNA from FFPE specimens do not accurately reflect the chromosomal content of cells of the tumor associated stroma but rather are the result

of technical artifacts resulting from the use of microsatellite assays and fixation methods [6-12].

Determination of whether the breast microenvironment is genetically stable is not merely an academic pursuit; genomic instability in tumor-associated stromal cells may impact clinical care for patients with breast cancer. For example, stroma adjacent to breast carcinomas from patients treated with chemotherapy demonstrated increased and persistent LOH post-treatment, especially in those patients who were non-responsive to treatment, suggesting that patients with stable stroma may be more likely to benefit from chemotherapy than those with unstable stroma [13]. Likewise, stromal instability may have important ramifications for defining surgical margins: the presence of LOH on chromosome 3p in normal TDLUs adjacent to early-stage breast carcinomas was associated with a 3.9-5.2 fold increased risk of recurrence [14].

To determine whether breast tumor stroma is genetically stable, high-quality breast specimens from mastectomy and reductive mammoplasty (RM) were frozen in optimal cutting temperature (OCT) compound. AI events were detected using the same 52-marker microsatellite panel used in earlier studies that detected AI in FFPE stroma adjacent to and distant from tumors. A subset of high-quality specimens was also subjected to copy number analysis using single nucleotide polymorphism (SNP) array data. These analyses demonstrated that AI levels were significantly higher in FFPE compared to frozen specimens and that stroma from both diseased and non-diseased breasts is genetically stable.

## **MATERIALS AND METHODS**

### *Human tissue samples*

Enrollment of patients into the Clinical Breast Care Project (CBCP) began in 2001. For inclusion in the CBCP, all patients must have met the following criteria: 1) adult over the age of 18 years, 2) mentally competent and willing to provide informed consent, and 3) presenting to the breast centers with evidence of possible breast disease, for routine screening mammograms or elective reductive mammoplasty. Tissue and blood samples were collected with approval from the Walter Reed Army Medical Center Human Use Committee and Institutional Review Board. All subjects enrolled in the CBCP voluntarily agreed to participate and gave written informed consent. Clinical information was obtained for all CBCP samples using questionnaires designed by and administered under the auspices of the CBCP.

Stroma was collected from patients undergoing surgical procedures, including RM, prophylactic mastectomy (PM) for BRCA1 [GenBank:NM\_007294] or BRCA2 [GenBank:NM\_000059] mutations and single or double mastectomy after a breast disease diagnosis. Within 5-15 minutes of surgical removal, breast tissue was taken on crushed, wet ice to the pathology laboratory where a licensed pathologist or pathologists' assistant performed routine pathology analyses (gross characterization, margin status assessment, and other indicated purposes). Excess stromal tissues were frozen in OCT (Sakura Finetek, Torrance, CA) on dry ice. Once preserved, frozen tissue samples were stored in liquid nitrogen freezers. All sections in this study were examined by a dedicated pathologist, who diagnosed all regions of disease within the specimen. The integrity of multiple serial sections was established by pathological verification of the first and last sections stained with hematoxylin and eosin (H&E) (Figure 1).

### *DNA isolation*

Excess OCT was trimmed from each specimen before cutting 10-20 12  $\mu$ m sections on the CM3050 S Cryostat (Leica Microsystems, Wetzlar, Germany). The 200-400 mg of tissue was homogenized using the PRO250 laboratory homogenizer (PRO Scientific Inc., Oxford, CT) for 30-60 seconds and processed using the Gentra Puregene Tissue kit (Qiagen, Valencia, CA) according to manufacturer's protocol. Genomic DNA, used as the referent, was isolated from blood clots using Clotspin and Puregene DNA purification kits (Qiagen, Valencia, CA) according to manufacturer's protocols.

### *AI analysis*

Microsatellite markers were amplified as previously described [15]. Genotypes were determined using Genetic Profiler version 2.0 software and AI detected using the formula  $(T1/T2)/(N1/N2)$  where T1 and N1 represent the peak heights of the less intense alleles and T2 and N2 represent the peak heights of the more intense alleles of the tumor and referent samples, respectively [16]. Each chromosomal region was represented by two polymorphic markers, and AI was defined according to the following criteria: (1) when at least one marker for a given region showed an allelic ratio  $\leq 0.35$ , the region was considered to show AI, (2) when neither marker had an allelic ratio  $\leq 0.35$  and at least one marker was informative, the region was considered normal, and (3) when both markers were homozygous, the region was considered uninformative. Comparisons of AI frequencies between frozen specimens and previously reported data from FFPE specimens [17] were calculated using Student's t-test with  $P < 0.05$  defining significance.

### *Copy number analysis*

To generate SNP data, 250 ng of stroma or genomic DNA was hybridized to GeneChip Human Mapping 250K Sty arrays (Affymetrix, Santa Clara, CA) following manufacturer's protocols. SNP data was imported into Partek® Genomics Suite™ 6.5 (Partek, Inc, St Louis, MO) as CEL files using default Partek parameters. Copy number was calculated as a paired analysis, using the genomic DNA from each patient as the referent sample. Deletions and amplifications were detected using a Hidden Markov Model using a minimum of five consecutive altered markers.

## RESULTS

### *Sample characteristics*

Samples were collected from patients undergoing RM (n=7), PM after identification of a deleterious BRCA1 or BRCA2 mutation (n=6), mastectomy (n=55), or double mastectomy (n=44) for a diagnosis of invasive or *in situ* breast disease; of the patients with invasive breast cancer, BRCA1 mutations were identified in six patients, BRCA2 in four patients and a seventh patient had a TP53 [GenBank:NM\_000546] mutation resulting in Li-Fraumeni Syndrome. For patients undergoing mastectomies, stroma samples from the Upper Outer (UOQ), Upper Inner (UIQ), Lower Outer (LOQ) and Lower Inner (LIQ) quadrants, as well as a piece from the central area of the breast were collected whenever possible. Single stromal samples were collected from RM. In total, 523 breast stroma specimens were subjected to AI (n=484) and/or SNP copy number analysis (n=86).

Patients undergoing RM and PM were significantly younger ( $P < 0.0001$ ) than those with breast disease; the average ages were 37.8 years (range 25-48 years), 36.2

years (range 23-42 years) and 53.51 (range 25-82 years) for RM, PM, and mastectomies, respectively (Table 1). Ethnicity did not differ significantly between groups, with the majority being self-described as white. Within the RM group, two women had fibroadenoma, one had fibrocystic changes, and four had no abnormalities detected; none of these women have returned to the breast clinic for subsequent breast disease. Disease progression and breast cancer mortality did not differ significantly between women with PM and those undergoing mastectomy.

#### *Allelic imbalance in research-grade breast stroma specimens*

In 484 stromal specimens, 53 AI events were detected; these alterations were detected in 40/484 (8%) breast specimens and the majority (92%) of specimens harbored no detectable AI events. The average AI frequency per specimen was <1% and the range was 0-25% (median = 0%). In the five quadrant specimens with DCIS, a single AI event was detected in two specimens with an average AI frequency of 1.6%.

Levels of AI were not significantly different in the breast stroma from patients who died of disease (0.6%), those with recurrence (0.2%) or those who are disease-free (0.5%). Breast stroma from patients with BRCA1 and BRCA2 mutations (0.3%) did not have higher levels of AI than those without known mutations (0.5%). Presence of AI did not differ significantly between quadrants (LIQ = 5%, LOQ = 7%, UIQ = 13%, UOQ = 9%).

The stromal specimen with the highest level of AI (25%) was from the UOQ of patient 1 who was diagnosed at age 47 with stage II, poorly differentiated basal-like breast tumor. The other three quadrants did not have detectable AI. The tumor for this

patient was located in the mid-outer region. Tumor was not detected for this patient in either the first or last section; and the section was comprised primarily of adipose cells. After two years of follow-up, this patient was disease-free.

Frequencies of AI from these research-grade frozen stromal specimens were compared to data previously generated in our laboratory from 77 FFPE archival breast quadrants [18]. Thirty-two (42%) archival quadrants had at least one detectable AI event and the frequency of AI in the FFPE specimens (4.62%) was significantly higher ( $P < 0.0001$ ) than that found in frozen specimens (0.45%) (Figure 2).

#### *Copy number alterations in research grade breast stroma*

SNP array data was generated from 27 women undergoing mastectomy ( $n=18$ ), PM ( $n=2$ ) or RM ( $n=7$ ) (Table 2). Of the 86 breast specimens assayed, no copy number alterations were detected in 82/86 (95%). The 11 copy number changes detected in the four stroma specimens from three patients are listed in Table 3. Amplification of chromosome 10q22.3 was detected in the LIQ from patient 1 and the UIQ from patient 2.

SNP data was available from the tumor component of patient 1. In contrast to the stromal specimens which had little to no copy number alterations, the tumor was characterized by numerous amplification and deletion events detected throughout the genome (Figure 3).

## **DISCUSSION**

One of the earliest studies that reported chromosomal changes in the breast tumor microenvironment assessed LOH/AI in DNA isolated from tumor epithelial cells and

adjacent normal terminal ductal-lobular units (TDLUs) from FFPE specimens. LOH/AI was detected in TDLUs adjacent to the tumor but not in distant TDLUs [19]. These results led to the idea that molecular data may be more accurate than surgical or pathological data in defining clean margins. Additional data from breast stroma close to and distant from the tumor have led to clinically important conclusions such as the precession of molecular changes in stroma compared to epithelial cells, possibly facilitating tumor invasion [18, 20], identification of an array of hotspots in the stromal cells that influence variability among patients in tumor biology, response to treatment, and outcome [21, 22], and that BRCA1/2 mutations may lead to genomic instability in the stroma, promoting neoplastic transformation [23]. Together, these data suggest that the tumor microenvironment is genetically unstable and this instability contributes to risk and prognosis of breast pathogenesis.

Technological improvements in identifying copy number changes and access to non-FFPE tissue specimens has generated new data that questions the conclusions, and indeed, the validity of the data generated from AI/LOH approaches. Gene expression differences were detected in purified stromal cell types (e.g., epithelial, myoepithelial and endothelial cells, leukocytes, myofibroblasts and fibroblasts) from normal and diseased breasts; copy number alterations, while detected in the tumor epithelial cells, were not detected in stromal cells [6]. In a follow-up study, stromal cells were found to be hypomethylated compared to their normal counterparts [9]. In the absence of chromosomal alterations, epigenetic, rather than chromosomal, changes in stromal cells may alter the tumor microenvironment and contribute to breast tumorigenesis [10, 11]. Finally, copy number analysis performed on DNA from microdissected epithelial and

cancer-associated fibroblasts (CAFs) failed to detect any chromosomal alterations in the CAFs. Additionally, microsatellite analysis from a region of chromosome 11 considered a hotspot for LOH in breast stroma from earlier studies was also performed using these CAF DNA samples and LOH was not detected [12]. The inability of these studies to detect copy number changes in breast stromal cells has led to the suggestion that the LOH/AI events detected in stroma from FFPE specimens represent technical artifact.

To determine whether AI events detected in archival stroma specimens using polymerase chain reaction (PCR)-based microsatellite analysis represents technical artifact, a direct comparison of technology and/or tissue types must be performed. In our study, we compared AI detected in research-grade stroma specimens compared to that detected in FFPE stroma specimens; both data sets were generated with the same set of 52 microsatellite markers using the same amplification conditions, with the same criteria to define AI [15, 18]. Of note, the use of highly multiplexed PCR, failure to optimize PCR conditions when amplifying DNA of low quantity and quality from FFPE samples and lack of replicate PCR reactions have been proposed as contributions to the erroneous detection of chromosomal alterations in breast stroma [11]. All 52 microsatellite markers used were optimized to robustly amplify DNA from FFPE specimens and multiplexed PCR was not performed. In addition, all AI events detected were confirmed using a second microsatellite marker or by performing a second amplification on a separate DNA aliquot from the same specimen [8]. Despite these precautions, levels of AI were >10-fold higher in stroma from FFPE specimens compared to those from frozen stroma specimens. Given that the 52 microsatellite markers represented small and defined regions of the genome, the use of global SNP arrays provided an enhanced coverage as

well as an agnostic approach to identify chromosomal alterations in breast stroma.

Although DNA from FFPE stroma specimens was not assayed on SNP arrays, the paucity of copy number changes in DNA from frozen stroma specimens suggest that the tumor microenvironment is genetically stable.

Given the complexity of breast stroma, it is possible that chromosomal alterations present in certain stromal cell types are masked by the genomic content of other components. Elevated frequencies of AI/LOH in fibroblasts laser microdissected from patients with BRCA1/2 mutations (59.7%) and sporadic tumors (36.7%) suggest that the fibroblast component may contribute to the instability detected in whole stroma [23], yet assessment of fibroblasts isolated from frozen breast stroma did not support these findings [12]. In addition, evaluation of the stromal cell types individually failed to detect copy number alterations [6]. Paradoxically, while evaluation of DNA from isolated cell types should provide the most accurate estimation of genomic alterations, the majority of copy number changes were detected in whole stroma FFPE specimens.

## CONCLUSIONS

The data presented here support a model in which the tumor microenvironment is genomically stable. The direct comparison of copy number alterations between FFPE and frozen research-grade specimens using identical methodologies suggests that past reports of significant AI/LOH in breast stroma, both adjacent to and distant from the tumor, reflects artifact in the archival specimens caused by formalin-fixation, paraffin-embedding and tissue storage. Importantly, these data suggest that the tumor microenvironment from patients with aggressive tumors or inherited susceptibility to

breast cancer does not harbor significant instability. Finally, this stability suggests that therapeutics targeting the tumor microenvironment may be effective tools in the treatment of breast cancer.

## **LIST OF ABBREVIATIONS**

AI = allelic imbalance

CAF = cancer-associated fibroblasts

CBCP = Clinical Breast Care Project

CGH = comparative genomic hybridization

ECM = extracellular matrix

FFPE = formalin-fixed, paraffin-embedded

H&E = hematoxylin and eosin

LIQ = lower inner quadrant

LOH = loss of heterozygosity

LOQ = lower outer quadrant

OCT = optimal cutting temperature

PCR = polymerase chain reaction

PM = prophylactic mastectomy

RM = reductive mammoplasty

SNP = single nucleotide polymorphism

TDLU = terminal ductal lobular unit

UIQ = upper inner quadrant

UOQ = upper outer quadrant

## **COMPETING INTERESTS**

The author(s) declare that they have no competing interests.

## **AUTHORS' CONTRIBUTIONS**

JLK performed specimen identification and processing and reviewed the manuscript, SR and ALV generated the AI and SNP data, CDS oversaw the sample collection and REE conceived of the SNP study design, performed the statistical analysis and drafted the manuscript.

## **ACKNOWLEDGEMENTS AND FUNDING.**

This research was supported by a grant from the United States Department of Defense (Military Molecular Medicine Initiative MDA W81XWH-05-2-0075, Protocol 01-20006). The opinion and assertions contained herein are the private views of the authors and are not to be construed as official or as representing the views of the Department of the Army or the Department of Defense.

## REFERENCES

1. Shekhar MP, Pauley R, Heppner G: Host microenvironment in breast cancer development: extracellular matrix-stromal cell contribution to neoplastic phenotype of epithelial cells in the breast. *Breast Cancer Res* 2003, 5:130-135.
2. Chiquet-Ehrissmann R, Mackie EJ, Pearson CA, Sakakura T: Tenascin: an extracellular matrix protein involved in tissue interactions during fetal development and oncogenesis. *Cell* 1986, 47:131-139.
3. Singer C, Rasmussen A, Smith HS, Lippman ME, Lynch HT, Cullen KJ: Malignant breast epithelium selects for insulin-like growth factor II expression in breast stroma: evidence for paracrine function. *Cancer Res* 1995, 55:2448-2454.
4. Yee D, Rosen N, Favoni RE, Cullen KJ: The insulin-like growth factors and their receptors, and their binding proteins in human breast cancer. *Cancer Treat Res* 1991, 53:93-106.
5. Boudreau N, Myers C: Breast cancer-induced angiogenesis: multiple mechanisms and the role of the microenvironment. *Breast Cancer Res* 2003, 5:140-146.
6. Allinen M, Beroukhi R, Cai L, Brennan C, Lahti-Domenici J, Huang H, Porter D, Hu M, Chin L, Richardson A, Schnitt S, Sellers WR, Polyak K: Molecular characterization of the tumor microenvironment in breast cancer. *Cancer Cell* 2004, 6:17-32.
7. Campbell I, Qiu W, Haviv I: Genetic changes in tumor microenvironments. *J Pathol* 2011, 223:450-458.
8. Holliday C, Rummel S, Hooke JA, Shriver CD, Ellsworth DL, Ellsworth RE: Genomic instability in the breast microenvironment? A critical evaluation of the evidence. *Expert Rev Mol Diagn* 2009, 9:667-678.
9. Hu M, Yao J, Cai L, Bachman KE, van den Brule F, Velculescu V, Polyak K: Distinct epigenetic changes in the stromal cells of breast cancers. *Nature Genet* 2005, 37:899-905.
10. Hu M, Polyak K: Molecular characterisation of the tumour microenvironment in breast cancer. *Eur J Cancer* 2008, 44:2760-2765.
11. Polyak K, Haviv I, Campbell IG: Co-evolution of tumor cells and their microenvironment. *Trends Genet* 2009, 25:30-38.

12. Qiu W, Hu M, Sridhar A, Opeskin K, Fox S, Shipitsin M, Trivett M, Thompson ER, Ramakrishna M, Gorringer KL, Polyak K, Haviv I, Campbell IG: No evidence of clonal somatic genetic alterations in cancer-associated fibroblasts from human breast and ovarian carcinomas. *Nat Genet* 2008, 40:650-655.
13. Oudin C, Bonnetain F, Boidot R, Vegran F, Soubeyrand M-S, Arnould L, Riedinger J-M, Lizard-Nacol S: Patterns of loss of heterozygosity in breast carcinoma during neoadjuvant chemotherapy. *Int J Oncol* 2007, 30:1145-1151.
14. Li Z, Moore DH, Meng ZH, Ljung B-M, Gray JW, Dairkee SH: Increased risk of local recurrence is associated with allelic loss in normal lobules of breast cancer patients. *Cancer Res* 2002, 62:1000-1003.
15. Ellsworth RE, Ellsworth DL, Lubert SM, Hooke J, Somiari RI, Shriver CD: High-throughput loss of heterozygosity mapping in 26 commonly deleted regions in breast cancer. *Cancer Epidemiol Biomarkers Prev* 2003, 12:915-919.
16. Medintz IL, Lee C-CR, Wong WW, Pirkola K, Sidransky D, Mathies RA: Loss of heterozygosity assay for molecular detection of cancer using energy-transfer primers and capillary array electrophoresis. *Genome Res* 2000, 10:1211-1218.
17. Ellsworth DL, Ellsworth RE, Liebman MN, Hooke JA, Shriver CD: Genomic instability in histologically normal breast tissues: implications for carcinogenesis. *Lancet Oncol* 2004, 5:753-758.
18. Ellsworth DL, Ellsworth RE, Love B, Deyarmin B, Lubert SM, Mittal V, Hooke JA, Shriver CD: Outer breast quadrants demonstrate increased levels of genomic instability. *Annals Surg Oncol* 2004, 11:861-868.
19. Deng G, Lu Y, Zlotnikov G, Thor AD, Smith HS: Loss of heterozygosity in normal tissue adjacent to breast carcinomas. *Science* 1996, 274:2057-2059.
20. Moinfar F, Man YG, Arnold L, Bratthauer GL, Ratschchek M, Tavassoli FA: Concurrent and independent genetic alterations in stromal and epithelial cells of mammary carcinoma: implications for tumorigenesis. *Cancer Res* 2000, 60:2562-2566.
21. Fukino K, Shen L, Matsumoto S, Morrison CD, Mutter GL, Eng C: Combined total genome loss of heterozygosity scan of breast cancer stroma and epithelium reveals multiplicity of stromal targets. *Cancer Res* 2004, 64:7231-7236.
22. Fukino K, Shen L, Patocs A, Mutter GL, Eng C: Genomic instability within tumor stroma and clinicopathological characteristics of sporadic primary invasive breast carcinoma. *J A M A* 2007, 297:2103-2111.

23. Weber F, Shen L, Fukino K, Patocs A, Mutter GL, Caldes T, Eng C: Total-genome analysis of BRCA1/2-related invasive carcinomas of the breast identifies tumor stroma as potential landscaper for neoplastic initiation. *Am J Hum Genet* 2006, 78:961-972.

## FIGURE LEGENDS

Figures 1a and 1b. H&E images of representative breast stroma sections from patient 19, quadrant UOQ. Figure 1a is an H&E image from the original section; figure 1b represents the final cut, taken after the intervening 240  $\mu\text{m}$  were used for DNA isolation.

Figure 2. Graph of the number of AI events in FFPE (n=77) compared to frozen (n=484) breast specimens. Blue and pink bars represent number of AI events in frozen and FFPE specimens, respectively. The number of specimens with no AI detected was significantly ( $P<0.0001$ ) higher in frozen (91%) compared to FFPE (58%) stromal specimens.

Figure 3. Chromosomal map of copy number alterations detected in tumor and stromal components from patient 1. Amplifications are depicted in red, deletions in blue. Teal and orange arrows mark copy number alterations detected in patient 1 UOQ and LIQ, respectively.

Table 1. Patient characteristics from 112 women with frozen breast stroma.

	Invasive/in situ diseases		Prophylactic (n=6)	Reduction (n=7)
	M (n=55)	DM (n=44)		
Age Diagnosis				
<40 years	0.12	0.16	0.83	0.71
40-49 years	0.24	0.32	0.17	0.29
≥50 years	0.64	0.52	0.00	0.00
Ethnicity				
Caucasian	0.60	0.79	0.67	0.50
African American	0.29	0.16	0.17	0.33
Hispanic	0.06	0.00	0.00	0.00
Asian	0.05	0.03	0.00	0.17
Other	0.00	0.02	0.16	0.00
Disease Status				
Disease-free	0.93	0.84	1.00	NA
Progression	0.02	0.09	0.00	NA
Died of disease	0.05	0.07	0.00	NA

Table 2. Characteristics of all stroma specimens assayed using SNP arrays. When the patient had a double mastectomy, LT = left breast, RT = right breast.

Patient #	Coded Patient Diagnosis	Procedure Type	Quadrant	Tumor Location	Pathologies Present in Specimen
1	IDCA	Mastectomy		Mid-outer	
			LIQ		Fibrocystic changes, Stromal fibrosis, Cysts, Apocrine metaplasia, microcalcifications, moderate intraductal hyperplasia, Columnar cell change
			LOQ		Fibrocystic changes, Stromal fibrosis, Cysts, Columnar cell change
			UIQ		Fibrocystic changes, Stromal fibrosis, Cysts, Microcalcifications, Columnar cell change
			UOQ		No abnormalities
2	Mucinous	Mastectomy		LIQ	
			LIQ		Post-surgical changes
			LOQ		Stromal fibrosis, Microcalcifications
			UIQ		No abnormalities
3	IDCA	Double mastectomy		UOQ LT	
			LIQ LT		Fibrocystic changes, Stromal fibrosis, Cysts, Microcalcifications, Moderate Intraductal hyperplasia
			LOQ LT		focal DCIS (probably not present in OCT block), microcalcifications, Stromal fibrosis
			UOQ LT		Fibrocystic changes, Stromal fibrosis, Cysts, Sclerosing adenosis, Microcalcifications
			Central LT		Microcalcifications, Stromal fibrosis
4	IDCA	Double mastectomy		Mid Inner RT	
			LIQ RT		No abnormalities
			LOQ RT		No abnormalities

5	Prophylactic mastectomy	UIQ RT	No abnormalities
		LIQ LT	No abnormalities
		UIQ LT	No abnormalities
		UOQLT	Microcalcifications
		LIQ RT	No abnormalities
		LOQ RT	No abnormalities
	Double mastectomy	UOQ RT	No abnormalities
		LIQ RT, All	Fibrocystic changes, Stromal fibrosis, Cysts, Columnar cell change, Sclerosing adenosis
6	IDCA	Quadrants LT	Fibrocystic changes, Stromal fibrosis, Cyst
		LIQ RT	Fibrocystic changes, Stromal fibrosis, Cyst
		LOQ RT	Fibrocystic changes, Stromal fibrosis, Cysts, Fibroadenomatoid nodule
		UIQ RT	Fibrocystic changes, Stromal fibrosis, Cysts, Sclerosing adenosis
		UOQ RT	Fibrocystic changes, Stromal fibrosis, Cysts
		LIQ LT	Fibrocystic changes, Stromal fibrosis, Cysts, Sclerosing adenosis, florid intraductal hyperplasia, Columnar cell change with atypia
	Prophylactic mastectomy	UIQ LT	No abnormalities
7	IDCA	UOQ LT	No abnormalities
		UOQ	Fibrocystic changes, Stromal fibrosis, Cysts
		Central	Fibrocystic changes, Stromal fibrosis, Cysts, Sclerosing adenosis
	Mastectomy	Multiple areas	No abnormalities
8	IDCA	LOQ	Stromal fibrosis
		UOQ	Stromal fibrosis

	In Situ	Mastectomy	Multiple areas	
9	In Situ	Mastectomy	Multiple areas	Fibrocystic changes, Stromal fibrosis, Cysts, Columnar cell change Fibrocystic changes, Stromal fibrosis, Cysts, Columnar cell change, Mild intraductal hyperplasia Fibrocystic changes, Stromal fibrosis, Cysts, Columnar cell change
10	In Situ	Mastectomy	Multiple areas	No abnormalities Fibrocystic changes, Stromal fibrosis, Cysts, Apocrine metaplasia, Columnar cell change No abnormalities
11	IDCA	Mastectomy	LOQ	Fibrocystic changes, Stromal fibrosis, Cysts, Apocrine metaplasia Fibrocystic changes, Stromal fibrosis, Cysts DCIS (focal), Fibrocystic changes, Stromal fibrosis, Cysts, Apocrine metaplasia, Columnar cell hyperplasia with atypia
12	IDCA	Mastectomy	Central	No abnormalities No abnormalities No abnormalities
13	ILCA	Double Mastectomy	Upper Mid RT, Upper Outer LT	Fibrocystic changes, Stromal fibrosis, Cysts, Apocrine metaplasia, Columnar cell change, Columnar cell hyperplasia Stromal fibrosis, Columnar cell change, Columnar cell hyperplasia
			LIQ RT	
			LOQ RT	

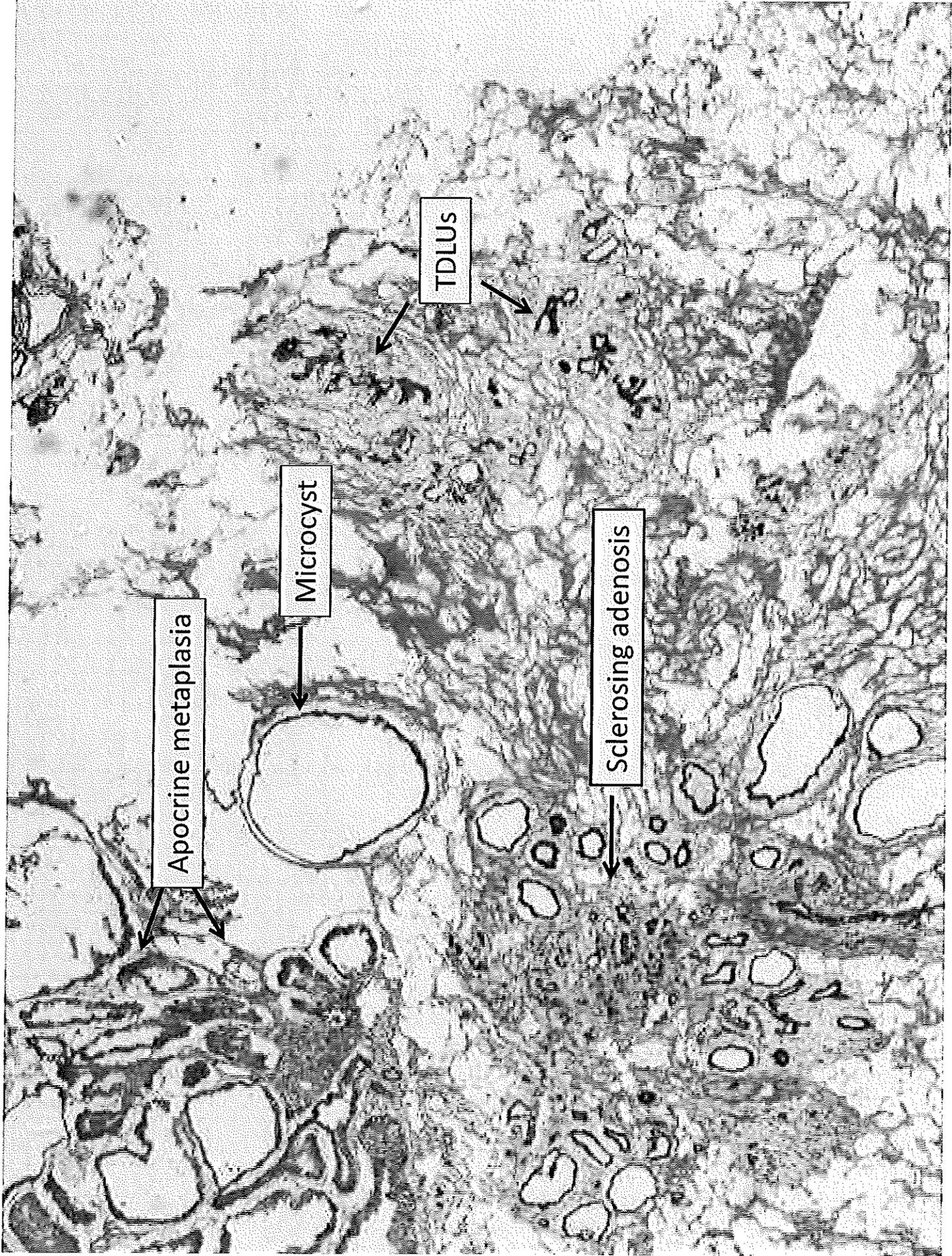
		UIQ RT	Stromal fibrosis
		UOQ RT	Fibrocystic changes, Stromal fibrosis, Cysts, Columnar cell hyperplasia, moderate Intraductal hyperplasia,
		LIQ LT	Fibrocystic changes, Stromal fibrosis, Cysts
		LOQ LT	Fibrocystic changes, Stromal fibrosis, Cysts, Columnar cell hyperplasia, moderate Intraductal hyperplasia,
		UIQ LT	Fibrocystic changes, Stromal fibrosis, Cysts
14	IDCA	LOQ LT	
	Double mastectomy		
		LIQ RT	Sclerosing adenosis, Fibroadenoma, Intraductal hyperplasia moderate
		LOQ RT	Stromal fibrosis
		UIQ RT	Stromal fibrosis
		UOQ RT	No abnormalities
		LOQ LT	Fibrocystic changes, Stromal fibrosis, Cysts, Columnar cell hyperplasia with atypia
		UIQ LT	Microcalcifications, Fibrocystic changes, Stromal fibrosis, Cysts, Sclerosing adenosis, Columnar cell hyperplasia with atypia, Columnar cell change with atypia
		UOQ LT	Fibrocystic changes, Stromal fibrosis, Cysts, Columnar cell change
15	IDCA	Central	
	Mastectomy		
		LIQ	No abnormalities
		LOQ	No abnormalities
		UIQ	No abnormalities
		Central	No abnormalities
16	IDCA	LOQ LT	
	Double mastectomy		
		LIQ LT	No abnormalities
		LOQ LT	No abnormalities

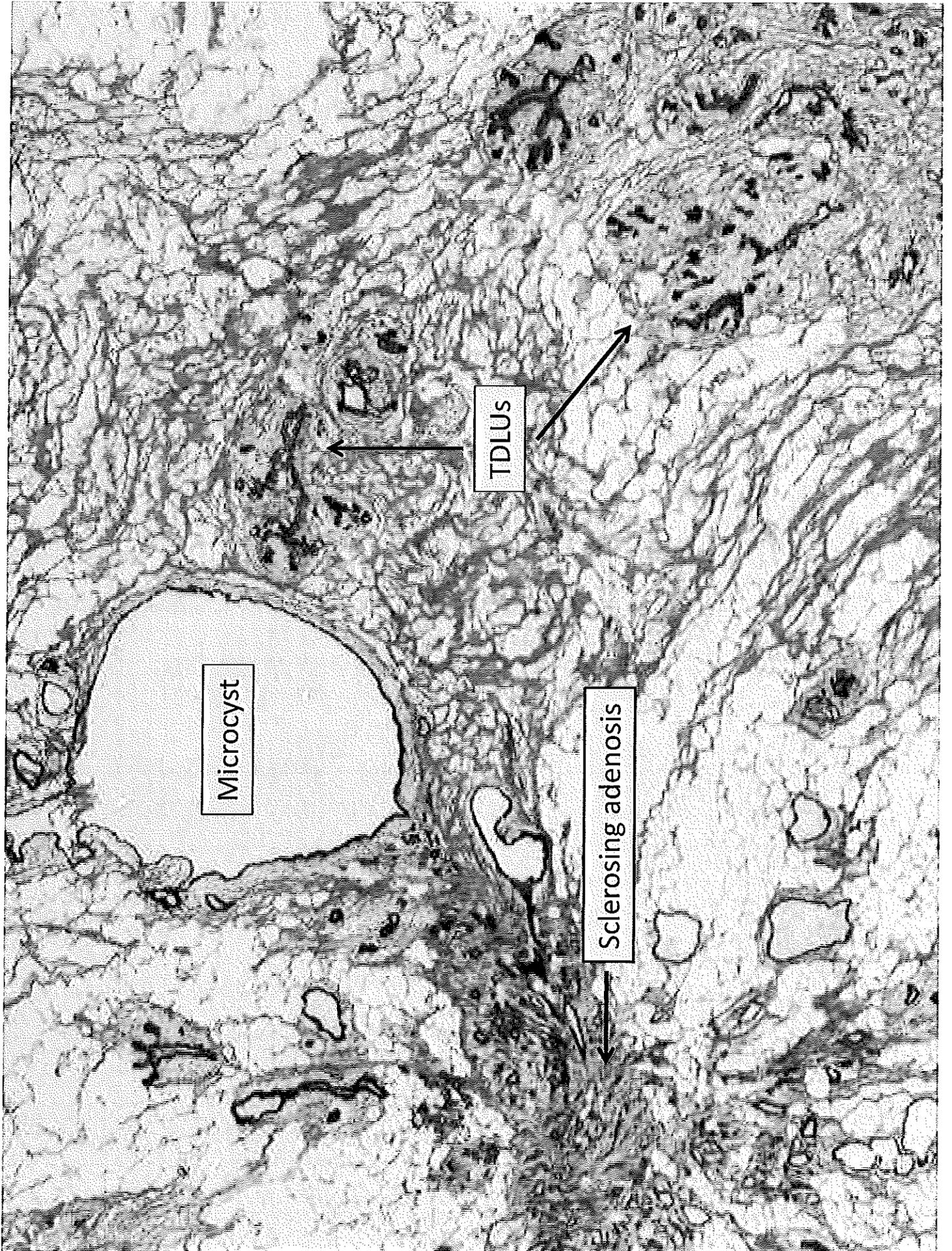
17	ILCA	Mastectomy	UIQ LT UOQ LT	UOQ	No abnormalities No abnormalities
18	ILCA	Mastectomy	LIQ LOQ UIQ UOQ	UOQ	No abnormalities No abnormalities No abnormalities No abnormalities
19	IDCA	Mastectomy	LIQ LOQ UIQ UOQ	Central	No abnormalities No abnormalities Fibrocystic changes, Stromal fibrosis, Cysts, Microcalcifications, Columnar cell change No abnormalities
20	IDCA	Mastectomy	UOQ Central	LOQ	Microcalcifications, Sclerosing adenosis, Columnar cell change, Radial scar Fibroadenoma, Multiple peripheral papillomas, Sclerosing adenosis, Radial scar
21	Disease-free	Mammoplasty	LIQ		No abnormalities
22	Disease-free	Mammoplasty	LOQ		Columnar cell changes with atypia,
23	Disease-free	Mammoplasty	UIQ		Microcalcifications, Sclerosing adenosis
24	Disease-free Fibrocystic changes	Mammoplasty	UOQ		Apocrine metaplasia No abnormalities
25	Fibroadenoma	Mammoplasty			No abnormalities
26		Mammoplasty			No abnormalities

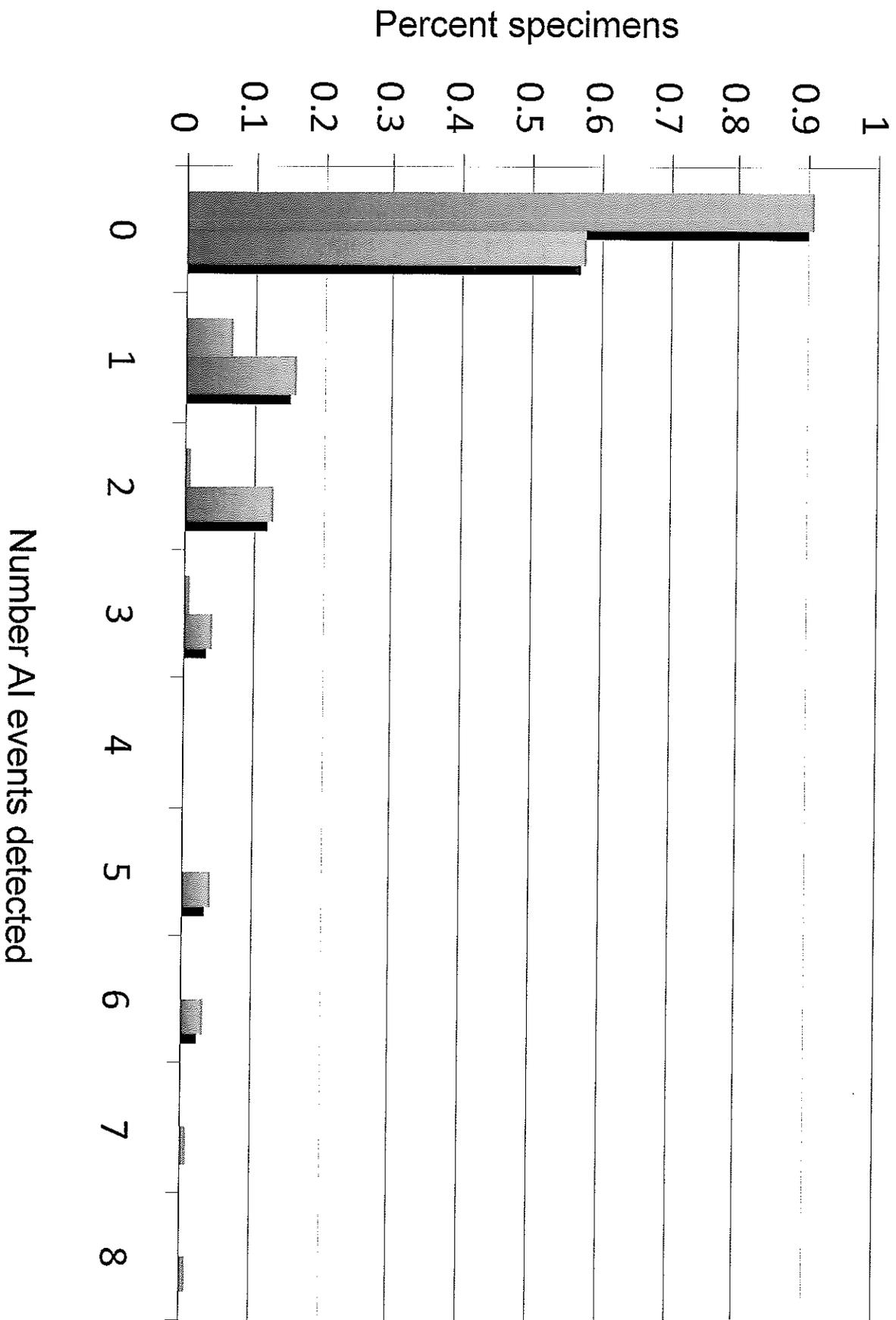
27	Fibroadenoma	Mammoplasty	No abnormalities
----	--------------	-------------	------------------

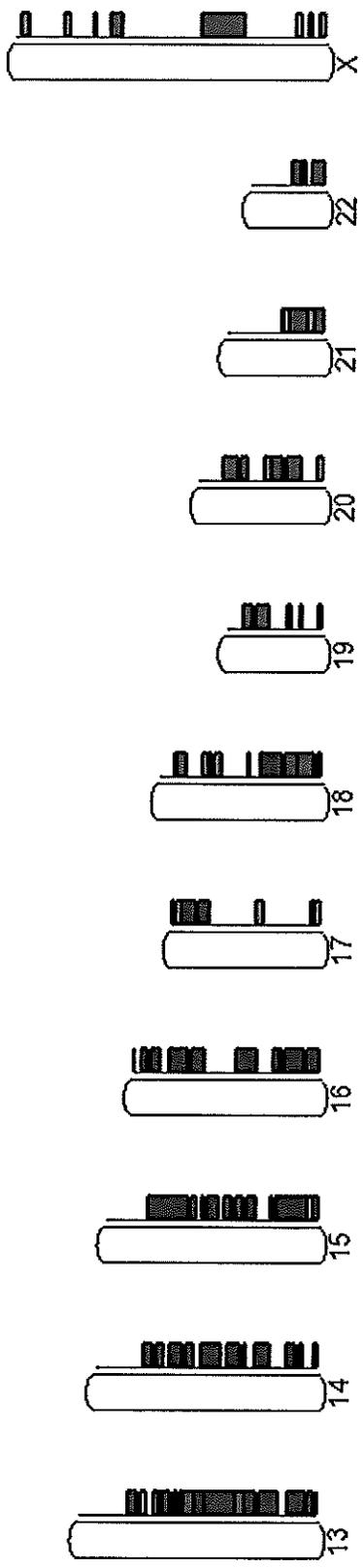
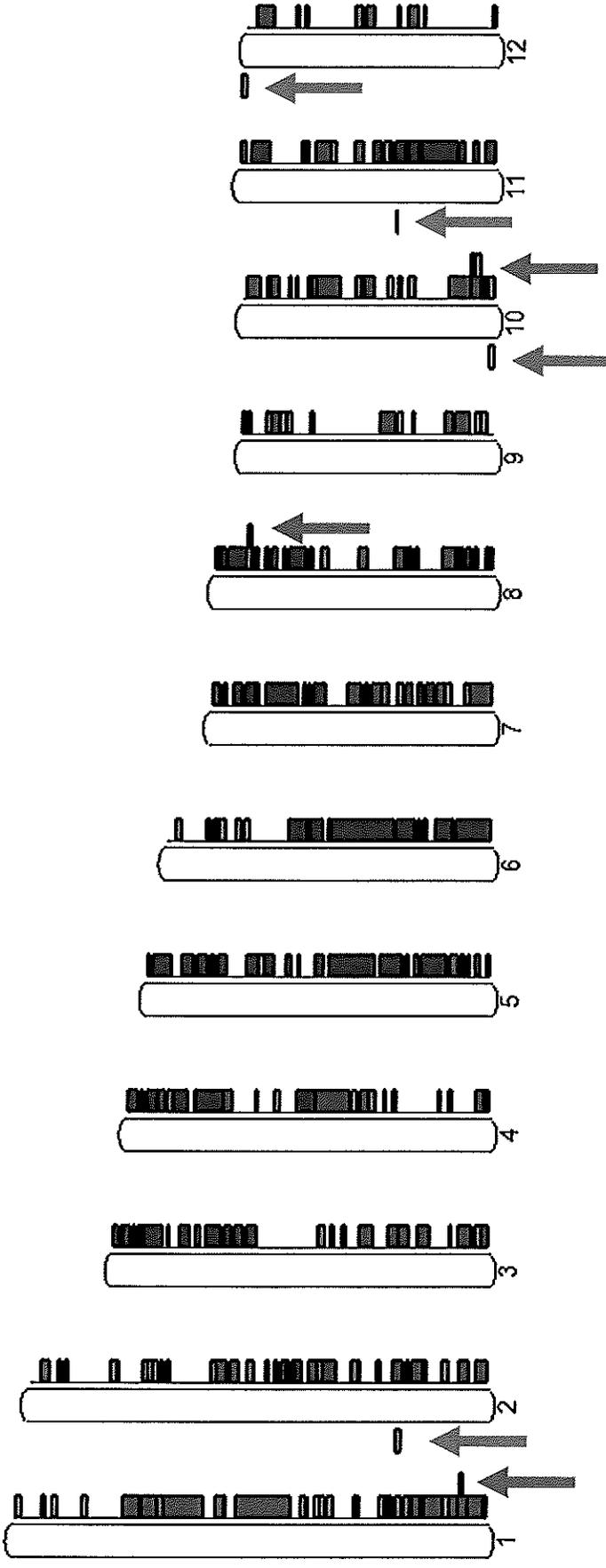
Table 3. Chromosomal regions with altered copy number in breast stroma specimens. One chromosomal region, 10q22.3, was amplified in two different stromal specimens. The shared 284 kb region harbors the genes retinoic acid induced 17 (ZMIZ1), peptidylprolyl isomerase F precursor (PPIF) and Homo sapiens zinc finger, CCHC domain containing 24 (ZCCHC24).

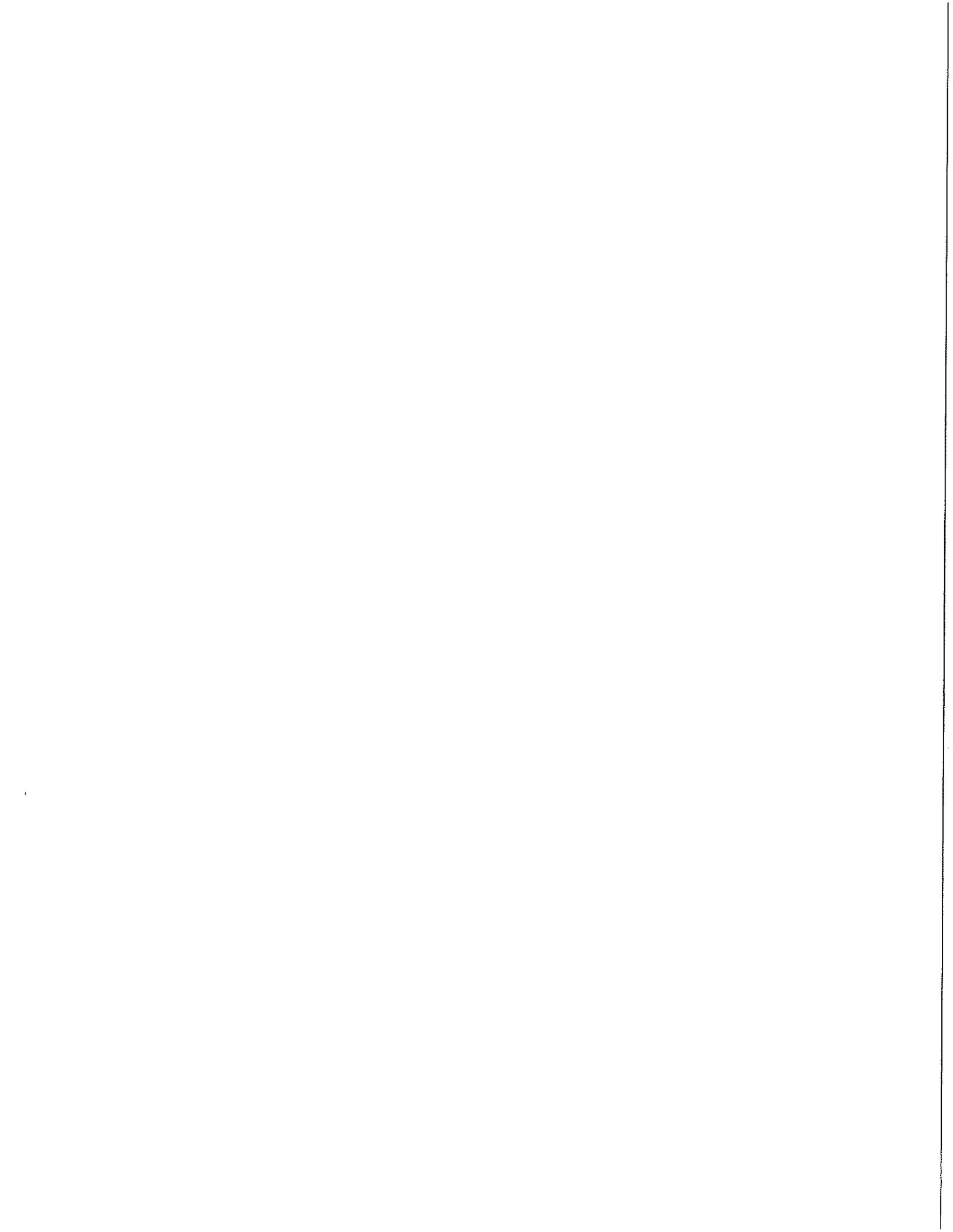
Specimen	Copy Number	Location	Size
1 UOQ	2.8	1q42.2-1q43	2.3 Mb
1 UOQ	2.6	8p22-8p21.3	2.6 Mb
1 UOQ	5.5	10q26.11-10q26.13	5.1 Mb
1LIQ	2.6	1q32.1	2.6 Mb
1LIQ	2.7	9q33.3-9q34.11	2.7 Mb
1LIQ	3.1	10q22.3	0.7 Mb
1LIQ	2.6	11p15.5-11p15.4	3.0 Mb
2 UIQ	4.0	10q22.3	0.3 Mb
2 UIQ	2.8	11p11.2	0.9 Mb
2 UIQ	3.1	14q24.1	0.7 Mb
23	2.7	8q24.3	5.9 Mb











Role for wnt/beta-catenin pathways for facilitating triple-negative breast cancer migration and growth.

Joji Iida

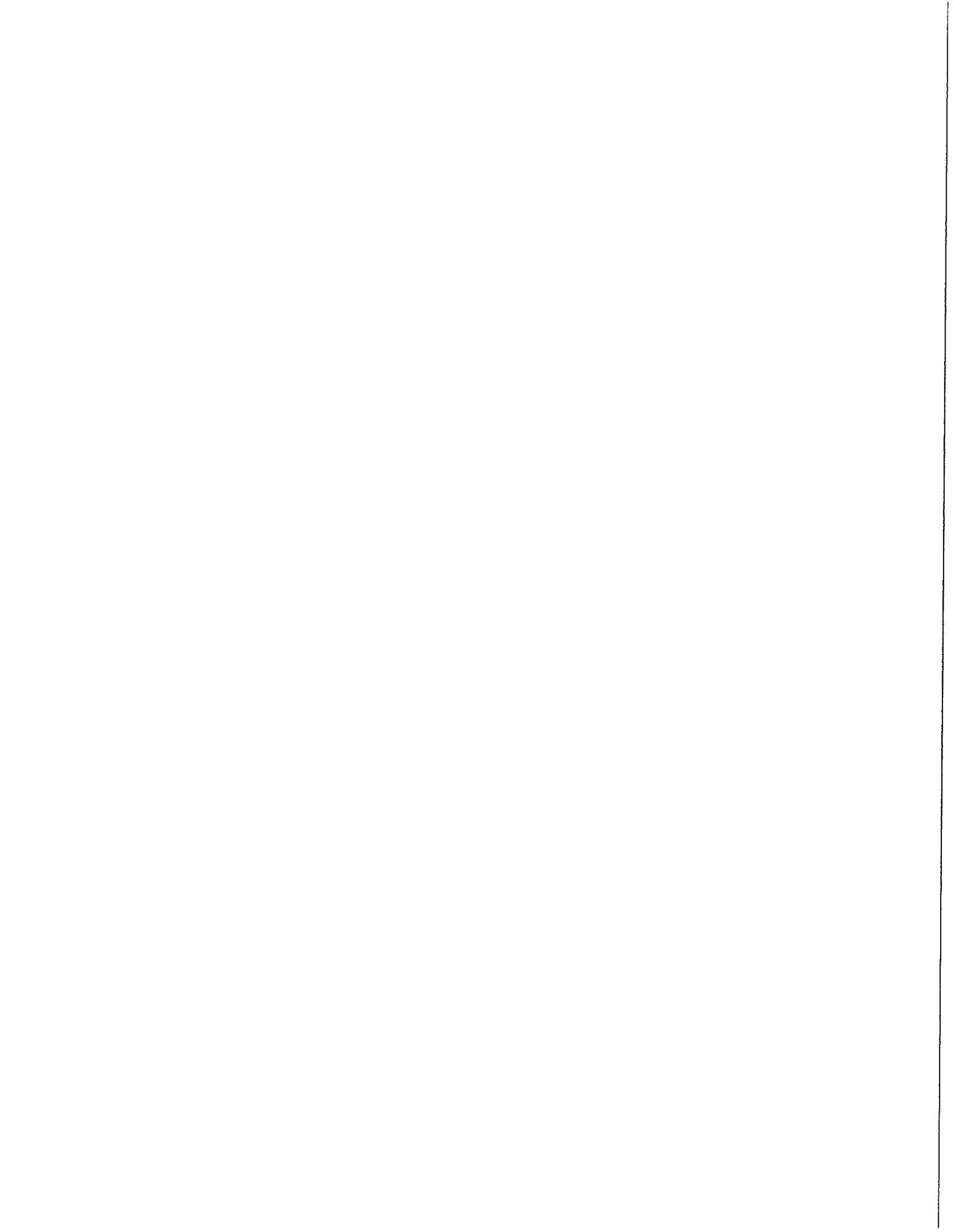
Director of Cancer Biology Program, Windber Research Institute

There is substantial evidence indicating that wnt signaling pathway is activated in various cancer cell types including breast cancer. Previous studies reported that FH535, a small molecule inhibitor of wnt signaling pathway, decreased growth of cancer cells but not normal fibroblast cells, suggesting this pathway plays a role in tumor progression and metastasis. In this study, we tested FH535 as a potential inhibitor for malignant phenotypes of breast cancer cells including migration, invasion, and growth. FH535 significantly inhibited growth, migration, and invasion of triple negative (TN) breast cancer cells (MDA-MB231 and HCC38) *in vitro*. Under our experimental conditions, FH535 did not affect expression of mRNAs of MMP-2 or MMP-9, however, significantly inhibited secretion of MMP-2 and MMP-9 from cells. Importantly, amounts of cell-associated MMP-2 and MMP-9 were not affected by FH535, implicating a role for canonical wnt-signaling pathways in regulating intracellular trafficking and/or secretion of MMP-2 and MMP-9. Finally, we demonstrated that FH535 was a potent growth inhibitor for triple-negative phenotype of breast cancer cells (HCC38 and MDA-MB-231) but not other (MCF-7, T47D or SK-Br3) when cultured in three dimensional (3D) type I collagen gel. We demonstrated that treatment of MDA-MB-231 cells with FH535 markedly inhibited the expression of NEDD9 but not in HCC38, implicating cell type specific migratory mechanisms in TN breast cancer cells. We also demonstrate that CD44 is associated with NEDD9 by co-immunoprecipitation assays, Given the information of NEDD9 and CD44 in cancer progression and metastasis by enhancing migration and invasion, our results suggest that targeting canonical wnt-signaling pathway would be beneficial strategies for treatment of TN breast cancer phenotype and possibly other cancers in which this signaling pathway is activated.

Discuss breast cancer progression from the view point of biological functions of CD44.

Discuss wnt/beta-catenin pathways in regulating tumor progression and metastasis.

The views expressed in this article are those of the author and do not reflect the official policy of the Department of Defense, or U.S. Government.



## Demographic differences in African American compared to Caucasian women with luminal A breast cancer

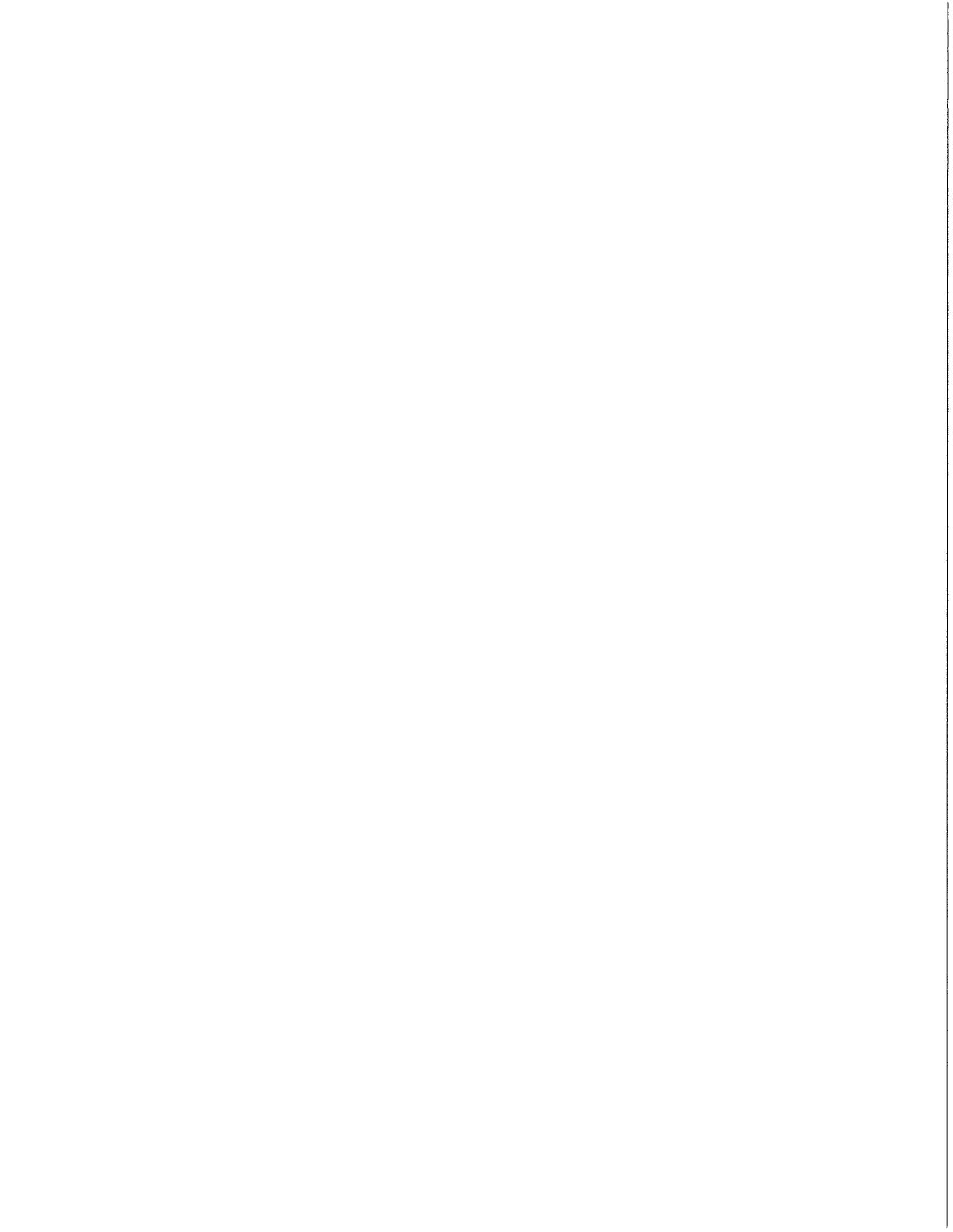
Rachel E. Ellsworth, Craig D. Shriver

**Background:** In a recent study, the Carolina Breast Cancer Study found that mortality rates were significantly higher in African American women (AAW) compared to Caucasian women (CW) with luminal A breast cancer. This disparity may be attributable to social and/or molecular differences. Here, we evaluated demographic data from AAW and CW with luminal A breast cancer to determine whether any factors differ between populations.

**Methods:** The database of the Clinical Breast Care Project was queried to identify all AAW (n=115) and CW patients (n=596) with ER+/HER2- breast cancer. Demographic data recorded at the time of diagnosis were compared between populations using chi-square or Student's t-test. A cutoff of  $P < 0.05$  was used to define significance.

**Results:** Age at diagnosis or menarche, parity, breastfeeding, education levels, exercise, fat intake, smoking and presence of heart disease and diabetes did not differ significantly between populations. Significantly more AAW began using oral contraceptives and were parous before age 18 (28% and 16%) compared to CW (7% and 4%). AAW were more likely to be obese (51% compared to 29%) and to perform at least monthly breast self-exams (97% compared to 80%) and to have hypertension (50% compared to 40%). AAW consumed significantly lower levels of caffeine (439mg/day compared to 983 mg/day) and alcohol (43% compared to 21% non-drinkers) than CW and were less likely to be married (52% compared to 74%). Pathologically, tumors from AAW and CW did not differ for stage or lymph node status, however, luminal A tumors from AAW were significantly more likely to be poorly-differentiated (25% compared to 14%) and were larger (33% compared to 21%) than those from CW.

**Discussion:** Although luminal A tumors are considered the most favorable breast tumor subtype, ER+/HER2- tumors from AAW were larger and of higher grade than those from CW. It is not clear how differences in caffeine and alcohol consumption may impact outcome in women with luminal A breast tumors, however, use of oral contraceptives during adolescence, pregnancy at an early age (<20 years) and increased inflammation in the breasts of obese women have been associated with higher grade and poorer prognosis. In addition, women who are married have a more favorable prognosis than those who are single, divorced or widowed and patients with hypertension has been associated with increased breast cancer mortality. Together, these demographic factors may contribute to the outcome disparity detected in AAW compared to CW with luminal A breast cancer.



# FH535 Inhibited Migration and Growth of Breast Cancer Cells

<http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0044418>

**Joji Iida<sup>1\*</sup>, Jesse Dorchak<sup>1</sup>, John R. Lehman<sup>1</sup>, Rebecca Clancy<sup>1</sup>, Chunqing Luo<sup>2</sup>,  
Yaqin Chen<sup>2</sup>, Stella Somiari<sup>3</sup>, Rachel E. Ellsworth<sup>4</sup>, Hai Hu<sup>2</sup>, Richard J. Mural<sup>5</sup>, Craig  
D. Shriver<sup>6</sup>**

**1** Department of Cell Biology, Windber Research Institute, Windber, Pennsylvania, United States of America, **2** Department of Biomedical Informatics, Windber Research Institute, Windber, Pennsylvania, United States of America, **3** Department of Tissue Bank, Windber Research Institute, Windber, Pennsylvania, United States of America, **4** Department of Genetics, Windber Research Institute, Windber, Pennsylvania, United States of America, **5** Windber Research Institute, Windber, Pennsylvania, United States of America, **6** Department of Surgery, Walter-Reed National Military Medical Center, Bethesda, Maryland, United States of America

## Abstract

---

There is substantial evidence indicating that the WNT signaling pathway is activated in various cancer cell types including breast cancer. Previous studies reported that FH535, a small molecule inhibitor of the WNT signaling pathway, decreased growth of cancer cells but not normal fibroblasts, suggesting this pathway plays a role in tumor progression and metastasis. In this study, we tested FH535 as a potential inhibitor for malignant phenotypes of breast cancer cells including migration, invasion, and growth. FH535 significantly inhibited growth, migration, and invasion of triple negative (TN) breast cancer cell lines (MDA-MB231 and HCC38) *in vitro*. We demonstrate that FH535 was a potent growth inhibitor for TN breast cancer cell lines (HCC38 and MDA-MB-231) but not for other, non-TN breast cancer cell lines (MCF-7, T47D or SK-Br3) when cultured in three dimensional (3D) type I collagen gels. Western blotting analyses suggest that treatment of MDA-MB-231 cells with FH535 markedly inhibited the expression of NEDD9 but not activations of FAK, Src, or downstream targets such as p38 and Erk1/2. We demonstrated that NEDD9 was specifically associated with CSPG4 but not with  $\beta$ 1 integrin or CD44 in MDA-MB-231 cells. Analyses of gene expression profiles in breast cancer tissues

suggest that CSPG4 expression is higher in Basal-type breast cancers, many of which are TN, than any other subtypes. These results suggest not only a mechanism for migration and invasion involving the canonical WNT-signaling pathways but also novel strategies for treating patients who develop TN breast cancer.

**Citation:** Iida J, Dorchak J, Lehman JR, Clancy R, Luo C, et al. (2012) FH535 Inhibited Migration and Growth of Breast Cancer Cells. PLoS ONE 7(9): e44418. doi:10.1371/journal.pone.0044418

**Editor:** Masaru Katoh, National Cancer Center, Japan

**Received:** September 22, 2011; **Accepted:** August 6, 2012; **Published:** September 11, 2012

**Copyright:** © 2012 Iida et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This research was supported by a grant from the United States Department of Defense (Military Molecular Medicine Initiative MDA W81XWHH-05-2-0075, Protocol 01-20006). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

\* E-mail: [g.iida@wriwindber.org](mailto:g.iida@wriwindber.org)

## INTRODUCTION

---

Triple negative (TN) breast cancers are defined by lack of expression of estrogen, progesterone, and HER-2 (ERBB2) receptors. It is widely recognized that TN breast cancers have a poorer prognosis than any other subtypes of breast cancer. Given the lack of effective targeted therapies for TN breast cancer patients, a better understanding of the mechanisms of growth and invasion of these tumors and their interaction with the tissue microenvironment will provide valuable insight into developing novel approaches to lower the mortality associated with TN breast cancer.

There is increasing evidence demonstrating both  $\beta$ -catenin-dependent (canonical) and independent (non-canonical) WNT-signaling pathways play a key role in regulating pathological processes by facilitating tumor growth, migration, and invasion. For example, Matsuda et al. reported that inhibition of the WNT-signaling pathway suppressed tumor growth in xenograft experimental systems in MDA-MB-231 cells [1]. Genetic and histological studies demonstrated that changes in the WNT-signaling pathway genes are common in metastatic breast cancer cells and enriched in epithelial-mesenchymal transition and stem cell phenotypes [2], [3], [4]. Importantly, recent studies analyzing gene expression profiles suggest that specific classes of TN breast cancer cells express genes regulating tumor migration, invasion, and differentiation, including TGF- $\beta$  signaling pathways, extracellular matrix (ECM) reorganization, and the WNT-signaling pathway [5]. Consistent with this notion, immunohistochemical studies demonstrated that the canonical WNT-signaling pathway is activated in TN breast cancer cells compared to cells other of other molecular subtypes [6], [7]. Thus, it is important for generating therapeutic strategies by evaluating how signaling pathways regulate tumor migration and invasion in TN breast cancer cells.

Recent studies have demonstrated that FH535 is a synthetic inhibitor of the canonical WNT-signaling pathway and inhibits growth of colon, lung, and hepatocellular carcinoma line but not normal fibroblasts [8], suggesting that the targeting of the canonical WNT-signaling pathway by small molecules could be a promising therapeutic approach for cancer cells in which this pathway is activated. Indeed, Vaid et al. reported that FH535 significantly inhibited malignant melanoma growth and migration [9]. In this study, we hypothesize that the canonical WNT-signaling enhances tumor progression and metastasis in TN breast cancer cells. In order to test this hypothesis, we utilized FH535 as a tool to evaluate its ability to inhibit growth, migration, and invasion of TN breast cancer cell lines (MDA-MB-231 and HCC38). Here, we show results suggesting that targeting the canonical WNT-signaling pathway by compounds like FH535 may be a promising therapeutic approach for TN breast cancers.

## MATERIALS AND METHODS

---

### Cell Lines

HCC38, T47D, MDA-MB-231, MCF-7, SK-Br3 cells were purchased from ATCC (Manassas, VA, USA) and maintained in RPMI1640 containing 10% FBS.

## **Antibodies and Chemicals**

MOPC-21 (IgG<sub>1</sub>), UPC-10 (IgG2a), and MOPC141 (IgG2b) were purchased from Sigma-Aldrich (St. Louis, MO, USA). Rat tail type I collagen, anti-CSPG4 (clone 9.2.27), anti-β1(clone P4C10), anti-α2 (clone P1E6), anti-α3 (clone P1B5) integrin, anti-α3 (clone P1D5), anti-α6 (clone NK1-GoH3), anti-αvβ3 (clone LM609) anti-actin, anti-Ki67, anti-MMP-2, anti-MMP9, and anti-NEDD9 antibodies were purchased from Millipore (Billerica, MA, USA). Anti-Axin, anti-phospho <sup>397</sup>Tyr-FAK, anti-FAK, anti-phospho<sup>416</sup>Tyr-Src, anti-Src, anti-p38MAPK, anti-phospho-p38 MAPK, anti-Erk1/2, anti-phospho-Erk1/2, and anti-CD44 (clone 156-3C11) antibodies were purchased from Cell Signaling Technology (Beverly, MA, USA). FITC-conjugated rabbit anti-mouse IgG and FITC-conjugated rabbit anti-rat IgG, HRP-conjugated rabbit anti-mouse IgG, and HRP-conjugated rat anti-rabbit IgG were purchased from Jackson ImmunoResearch (West Grove, PA, USA). Antigen-retrieval solution was purchased from Biogenex (Fremont, CA, USA). Matrigel and rat tail type I collagen were purchased from BD Bioscience (Franklin Lakes, NJ, USA). FH535 was purchased from Calbiochem, and dissolved in DMSO at a concentration of 10 mM as a stock solution and used by diluting in DMSO to give a final concentration of 0.1% (v/v). Thus, control groups in this study contained DMSO at this concentration. For inhibiting adhesion, migration, and invasion of tumor cells by anti-integrin antibodies, we utilized each antibody at a final of concentration of 1 µg/ml. Other chemicals were purchased from Sigma-Aldrich unless otherwise mentioned.

## **FACS Analysis**

Expression of integrin subunits was evaluated by FACS analysis according to the methods previously described [10]. Briefly, cells were harvested with PBS containing 5 mM EDTA and immediately neutralized in FACS buffer (RPMI1640 containing 1% BSA and 0.025% NaN<sub>3</sub>). Tumor cells (10<sup>5</sup> cells) were incubated with 1 µg/ml of the primary antibodies by shaking for 1 hour at 4°C. After extensive washing with FACS buffer, cells were incubated with FITC-conjugated secondary antibody (1:1000 dilutions) by shaking for 1 hour at 4°C. Cells were then washed with FACS buffer 5 to 6 times and

fixed in PBS containing 1% paraformaldehyde. Expression of integrin subunit was measured on a FACS Calibur (Becton-Dickinson, NJ USA) by collecting 10,000 events and analyzed by CellQuest.

### **Adhesion Assays**

Adhesion of tumor cells to type I collagen was evaluated as described previously with minor modifications [10]. 96-well plates were coated with type I collagen at a concentration of 5 µg/ml and blocked with PBS containing 1 mg/ml BSA. Cells were harvested with PBS containing 5 mM EDTA, washed, resuspended in RPMI1640 containing 1 mg/ml BSA (adhesion buffer) at a concentration of 10<sup>5</sup> cell/ml. Cells (100 µl/well) were incubated on plates coated with type I collagen as described above for 1 hour in the presence or absence of FH535. Plates were washed with adhesion buffer 4–5 times and remaining cells were fixed followed by staining with crystal violet. After extensive washing to remove excess dye, cells were lysed with 100 µl of PBS containing 2% SDS and then the absorbance was measured at 550 nm. Anti-β1 integrin antibody (P4C10), anti-α2 integrin antibody (P1E6), anti-α3 integrin antibody (P1B5) antibodies were used to ensure that tumor adhesion was mediated by α2β1 and α3β1 integrins. MOPC-21 antibody was used as an isotype-matched negative control antibody. Experiments were performed in quadruplicates and repeated three times. The representative data were shown as mean ± S.D. of absorbance values.

### **Migration Assays**

Tumor cell migration was characterized using the Transwell systems (Costar, 8.0 µm pore size) as described previously [11]. Briefly, cells were harvested with PBS containing 5 mM EDTA, washed, and resuspended in serum-free RPMI1640 media at a concentration of 5×10<sup>5</sup> cells/ml. An aliquot of 100 µl of the cell suspension was added into the upper chamber of Transwell system). Type I collagen was dissolved in serum-free RPMI1640 media at a concentration of 3 µg/ml and added into the lower chamber of the system. Antibodies (1 µg/ml) or FH535 (0.01 to 1 µM) were added in both upper and lower chambers. Migration assays were performed at 37°C for 4–5 hours. The inserts were removed and then fixed in 10% formalin in PBS for five minutes and stained with 0.5% crystal violet for five minutes. The inserts were washed and the upper surface of the membranes was wiped with a cotton swab to remove non-migratory cells. Migrated cells were counted in three randomly selected fields.

Each experiment was performed in triplicate and repeated three times. The results presented are shown as mean cell numbers  $\pm$  S.D. per mm<sup>2</sup>.

### **Invasion Assays**

Invasion of tumor cells into Matrigel was evaluated as described previously [10]. Tumor cells were harvested and washed as described above, and resuspended in RPMI1640 containing 2% FBS at a concentration of 10<sup>6</sup>/ml. Transwells (Costar, cat# 3422, 8  $\mu$ m pore-size) were coated with 100  $\mu$ l of Matrigel at a concentration of 5 mg/ml overnight and washed with RPMI1640 containing 2% FBS to remove unpolymerized proteins. Type I collagen (3  $\mu$ g/ml) were added to the lower chambers of the Transwell as an adhesive protein. Cells (100  $\mu$ l/well) were incubated on Matrigel overnight at 37°C in the presence or absence of FH535 in both upper and lower chambers of the Transwell. Anti- $\beta$ 1 integrin antibody (P4C10), anti- $\alpha$ 2 integrin antibody (P1E6), anti- $\alpha$ 3 integrin antibody (P1B5) antibodies were used as a control. MOPC-21 antibody was used as an isotype-matched negative control antibody. The inserts were removed and fixed, stained, as described above. Migrated cells underneath of the membranes were counted at three randomly selected fields. Each experiment was performed in triplicate and repeated three times. The represented results were shown as mean cell numbers  $\pm$  S.D. per mm<sup>2</sup>.

### **Western Blotting**

Cells were cultured in 6 well plates (7 $\times$ 10<sup>5</sup> cells/well/1 ml) coated with 3  $\mu$ g/ml of type I collagen for 4 hours in the presence or absence of FH535 (1  $\mu$ M). Plates were washed with PBS and lysed in SDS-sample buffer. After fragmenting DNA by sonication, proteins were separated on SDS-PAGE and transferred onto Immobilon-P membrane as we described previously [11]. The membranes were blotted with primary antibodies followed by HRP-conjugated secondary antibodies. The proteins were detected with enhanced chemiluminescence (ECL) reactions. For serial immunodetection by different antibodies, membranes were stripped and ensured that there was no carry over signals from the last detection by ECL.

### **Co-immunoprecipitation Assays**

Co-immunoprecipitation assays were performed as described previously [12]. Briefly, cells were cultured in dishes (10 cm diameter) overnight in RPMI1640 containing 10% FBS. Cells were briefly washed with PBS one time and then directly lysed in 1 ml of 100 mM Tris-HCl (pH7.4) containing 1% Brij35, 0.14 M NaCl, 1 mM CaCl<sub>2</sub>, 1 mM MgCl<sub>2</sub>, 1 mM MnCl<sub>2</sub>, and a protease inhibitor cocktail. Cell lysates were centrifuged and the supernatants were precleared with mouse IgG-protein G-agarose for 4 hours by shaking at 4°C. The cleared lysates were immunoprecipitated with control IgG, anti-CD44, anti-β1, or anti-CSPG4 antibody-conjugated protein G-agarose for 4 hours by shaking at 4°C. The beads were extensively washed with Tris-HCl (pH7.4) containing 1% Brij35, 0.14 M NaCl, 1 mM CaCl<sub>2</sub>, 1 mM MgCl<sub>2</sub>, 1 mM MnCl<sub>2</sub>, and a protease inhibitor cocktail to remove any proteins that nonspecifically bound on the beads. The bound proteins were released by boiling at 95°C in SDS-sample buffer under reducing conditions and separated on SDS-PAGE followed by western blotting analysis as described above.

## **Collagen Gel Culture**

Cell growth was studied in three-dimensional collagen gels as described previously [13]. Briefly, purified type I collagen gels were prepared by adding 10X RPMI, FBS (final concentration 10%) and neutralized by adding 1 N NaOH to give 2.5 mg/ml of type I collagen. Immediately, cells were added at a concentration of  $5 \times 10^5$  cells/ml and the collagen/cell mixture was added to wells, and warmed at 37°C for completing gelling for 4–5 hours. When gels had formed, culture media (RPMI1640 containing 10% FBS) was added and cultured for eight days by changing media every three days. FH535 was added in both gel and medium throughout the experiments at final concentrations as described in the text.

## **Morphological Studies**

After eight days of incubation, cell/collagen matrix was fixed by immersion in PBS-10% formalin for 4–5 hours and embedded in paraffin. The paraffin section was serially cut at 5 μm and mounted on glass slides. Every seven sections, cell/collagen matrices were stained with hematoxylin and eosin to determine the presence of cells in the serial sections. The sections were deparaffinized and treated with antigen-retrieval solution according to the manufacturer's instructions. Sections were stained with

various antibodies and visualized by DAB staining with counter staining by DAPI to localize cells. The results are shown by the (mean +/- standard deviation) of % of positive cells (DAB-positive) out of the total cells (DAPI-positive) from three independent experiments.

### **Gene expression Analysis**

The log<sub>2</sub> transformed gene expression data of 514 breast cancers were downloaded from The Cancer Genome Atlas (TCGA) project data portal (<https://tcga-data.nci.nih.gov/>). PAM50 classification results of all samples were obtained from the TCGA breast cancer AWG group [14], distributed as follows: Luminal A (LumA) = 231, Luminal B (LumB) = 128, Her2-like = 58, and Basal = 97. All statistical analysis was performed using SAS and/or R software.

### **Statistical Analysis**

Statistical significance was calculated by Student's two-tailed t-test (paired). *P* values less than 0.05 were considered as statistically significant. For analyzing gene expression profiles, analysis of variance (ANOVA) was performed.

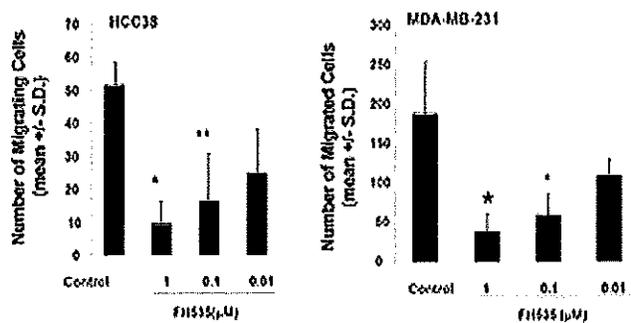
## **RESULTS**

---

### **FH535 Inhibited Triple-negative (TN) Breast Cancer cells, MDA-MB-231 and HCC38, Migration Toward Type I Collagen**

Migration of breast cancer cells into surrounding connective tissue architecture is important for establishing metastasis. Type I collagen is a major component of connective tissues and migration toward this protein is implicated as a key step for invading into tissues. Previous studies demonstrated that tumor migration to type I collagen is mediated by either of  $\alpha 2\beta 1$  or  $\alpha 3\beta 1$ -integrin or both [15], [16]. Indeed, inhibitory antibodies against  $\alpha 2$ ,  $\alpha 3$ , or  $\beta 1$  integrin subunits significantly inhibited migration toward type I collagen using MDA-MB-231 and HCC38 cells (not shown). Under these experimental conditions, we tested FH535 for its ability to regulate the migration of HCC38 and MDA-MB-231 cells to type I collagen. Our results demonstrated that FH535 inhibited migration in a concentration dependent manner and statistically significant inhibition was observed even at a

concentration of 0.1  $\mu\text{M}$  in both cell lines (**Figure 1**), consistent with the previous studies using human malignant melanoma cells [9]. Previous studies demonstrated that FH535 is a potent inhibitor for the canonical WNT-signaling pathway without affecting the amount of  $\beta$ -catenin [8]. When MDA-MB-231 cells were treated with FH535 at a concentration of 1  $\mu\text{M}$ , the amount of  $\beta$ -catenin was not affected, nor was axin (**Figure 2**) consistent with previous studies [8]. The same treatment, however, reduced the expression of  $\beta$ -catenin while increasing the amount of axin in HCC38 cells (**Figure 2**). Given the key role of axin in regulating degradation of  $\beta$ -catenin [17], these results imply that FH535 may inhibit the canonical WNT-signaling pathway through the stabilization of axin, which leads to a degradation of  $\beta$ -catenin. Thus, regardless of the significant inhibition of migration in the presence of FH535 in both cell lines, these results suggest that FH535 may affect migratory abilities of these cell lines through different mechanisms.



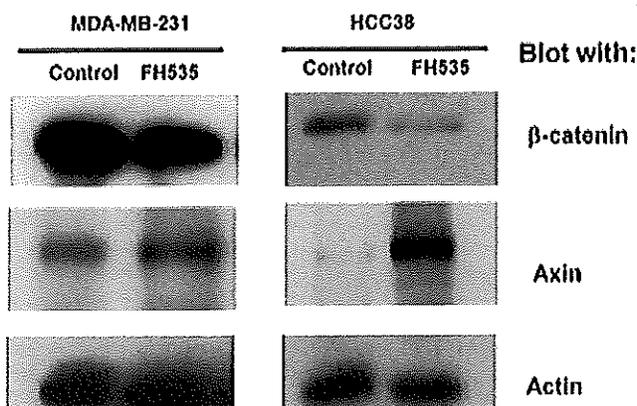
**Figure 1.** FH535 inhibited migration of MDA-MB231 and HCC38 cells to type I collagen.

Cells were harvested, washed, and resuspended in RPMI-serum free media at a concentration of  $5 \times 10^5$  cells/ml.

Type I collagen was used as a chemoattractant at a concentration of 3 g/ml. FH535 (0.01–1  $\mu\text{M}$ ) were added in both cell suspension and type I collagen solution and incubated for 4 hours at 37°C. Migrated cells were manually counted and expressed as mean  $\pm$  S.D. Experiments were repeated three times. \*  $p < 0.001$ , \*\*  $p < 0.05$  (by Student's two-tailed paired  $t$ -test).

doi:10.1371/journal.pone.0044418.g001

**Download:** PowerPoint slide | larger image (56KB PNG) | original image (129KB TIFF)



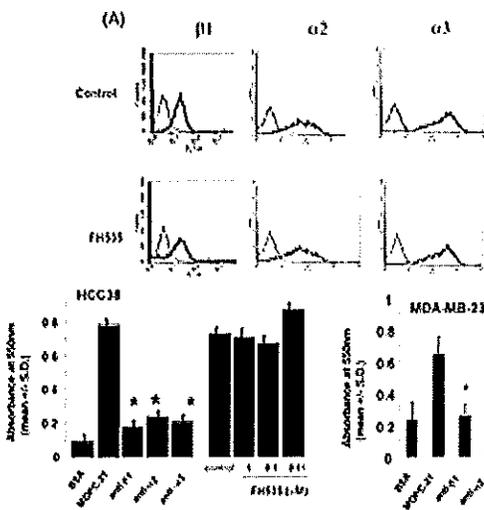
**Figure 2. Effect of FH535 on the expression of  $\beta$ -catenin and axin in MDA-MB-231 and HCC38 cells.**

Cells were cultured on type I collagen for 4 hours in the presence or absence of FH535 (1  $\mu$ M) at 37°C. Cells were directly lysed in SDS-sample buffer. Proteins were separated on SDS-PAGE followed by transferred onto Immobilon-P membranes. Membranes were blotted with anti-  $\beta$ -catenin or anti-axin antibody. The membranes were then striped and blotted with anti-actin antibody as a loading control.

doi:10.1371/journal.pone.0044418.g002

**Download:** [PowerPoint slide](#) | [larger image \(352KB PNG\)](#) | [original image \(327KB TIFF\)](#)

We, then, asked whether FH535 inhibited expression of integrin subunits that promote cell adhesion to type I collagen. Treatment of MDA-MB-231 cells with FH535 at a concentration of 1  $\mu$ M did not affect expression of  $\alpha$ 2,  $\alpha$ 3, or  $\beta$ 1 integrin subunit evidenced by FACS analysis (**Figure 3A**). The same results were obtained when HCC38 cells were treated with FH535 (**not shown**). Consistent with these results, adhesion of MDA-MB-231 or HCC38 cells to type I collagen was not inhibited in the presence of FH535 at a concentration of 1  $\mu$ M (**Figure 3B**). These results demonstrate that FH535 inhibited cell migration without affecting adhesive abilities of cells to type I collagen, suggesting that signaling pathways important for promoting migration would be attenuated in the presence of FH535.



**Figure 3. FH535 did not inhibit adhesion of MDA-MB-231 and HCC38 cells on type I collagen.**

(A) MDA-MB-231 cells were cultured in the presence or absence of FH535 (1  $\mu$ M) for 4 hours at 37°C. Cells were harvested with PBS containing 5 mM EDTA, washed with RPMI1640 containing 0.025%  $\text{NaN}_3$  and 1% BSA. Cells were incubated with the indicated antibodies (1  $\mu$ g/ml) for 1 hour, washed, and then incubated with secondary

antibodies (1:1000 dilutions) for 1 hour. Expression of integrins was analyzed on FACScalibur with CellQuest by collecting 10,000 events. (B) Plates were coated with type I collagen at a concentration of 5  $\mu$ g/ml and blocked with BSA. Cells were

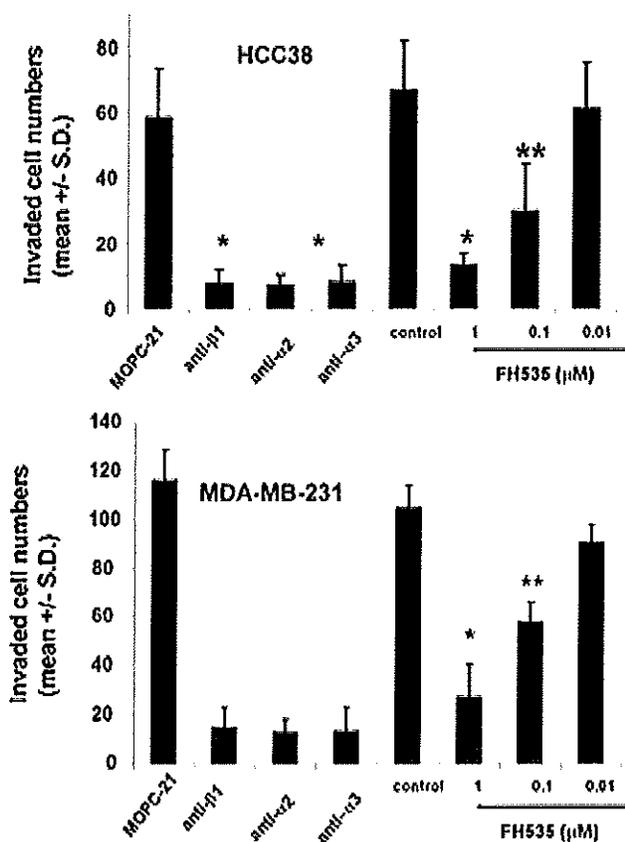
resuspended in RPMI1640 containing 1 mg/ml BSA and incubated on the plates for 1 hour at 37°C in the presence or absence of FH535 (0.01–1  $\mu$ M). Antibodies (1  $\mu$ g/ml) were incubated with cells. After extensive washing, adherent cells were fixed and stained with crystal violet. Cell adhesion was expressed as Absorbance at 550 nm (mean  $\pm$  S.D). \*  $p < 0.001$  (by Student's two-tailed paired  $t$ -test).

doi:10.1371/journal.pone.0044418.g003

**Download:** PowerPoint slide | larger image (277KB PNG) | original image (474KB TIFF)

## FH535 Inhibited Invasion of MDA-MB-231 and HCC38 Cells

In order to establish metastasis, tumor cells must transverse basement membrane to reach connective tissues. Invasion of tumor cells through matrigel has been used as a model system to evaluate migratory abilities of tumor cells through the basement membrane [18]. Invasion of HCC38 and MDA-MB-231 cells through Matrigel was significantly inhibited by anti-  $\alpha$ 2,  $\alpha$ 3 and -  $\beta$ 1 integrin antibodies, supporting that  $\alpha$ 2 $\beta$ 1 and  $\alpha$ 3 $\beta$ 1 integrins play a key role in promoting tumor invasion into matrigel. These results are consistent with the fact that type IV collagen is one of the major components that specifically binds to  $\alpha$ 2 $\beta$ 1 and  $\alpha$ 3 $\beta$ 1 integrins expressed on tumor cell surfaces [8]. Importantly, FH535 inhibited invasion of both MDA-MB-231 and HCC38 cells into matrigel in a concentration-dependent manner (Figure 4). As observed in the migration assays as described



above, statistically significant inhibition of invasion was achieved even at a concentration of 0.1  $\mu$ M (Figure 4).

**Figure 4. FH535 inhibited invasion of MDA-MB-231 and HCC38 cells into matrigel.**

Filters were coated with Matrigel. FH535 (0.01–1  $\mu$ M) were added in both cell suspension and type I collagen solution and incubated overnight at 37°C. Antibodies (1  $\mu$ g/ml) were

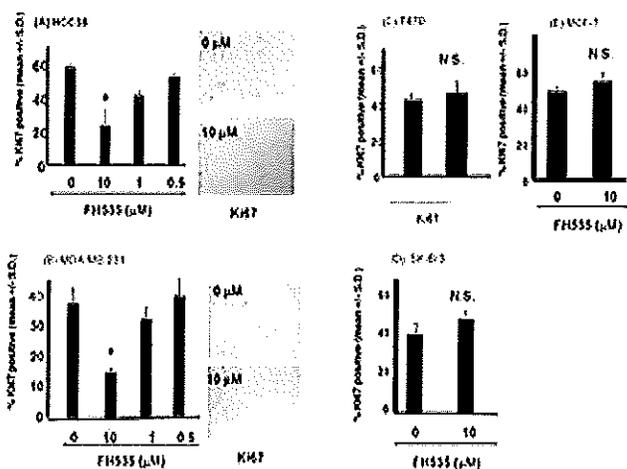
incubated with cells. Migrated cells were manually counted and expressed as mean  $\pm$  S.D. Experiments were repeated three times. \*  $p < 0.001$ , \*\*  $p < 0.05$  (by Student's two-tailed paired  $t$ -test).

doi:10.1371/journal.pone.0044418.g004

Download: PowerPoint slide | larger image (71KB PNG) | original image (154KB TIFF)

## FH535 Regulated Growth of HCC38 in 3D type I Collagen Culture

Although *in vivo* tissues are highly complex architecture consisting of ECM proteins, stromal fibroblasts, and soluble growth factors, recent studies suggest that the ECM could approximate tissues and provide a model for growth of various tumor cell including breast cancer cells [19], [20], [21]. In order to test if the canonical WNT-signaling pathway is involved in growth of breast cancer cells, various breast cancer cell lines (MDA-MB-231, HCC38, SkBr3, MCF-7, and T47D) were cultured in three dimensional (3D) type I collagen matrices as described previously [22], [23]. When HCC38 cells were cultured for eight days in the presence of FH535 at a concentration of 10  $\mu$ M, cell proliferation was significantly inhibited compared to control cells (**Figure 5A**). Similarly, growth of the other TN breast cancer cells, MDA-MB-231, was also significantly inhibited in the presence of FH535 (**Figure 5B**). Interestingly, proliferation of breast cancer cells with other from other subtypes, T47D (ER<sup>+</sup>, PR<sup>+</sup>, Her2<sup>-</sup>) (**Figure 5C**), SK-Br3 (ER<sup>-</sup>, PR<sup>-</sup>, Her2<sup>+</sup>) (**Figure 5D**), or MCF-7 (ER<sup>+</sup>, PR<sup>+</sup>, Her2<sup>-</sup>) (**Figure 5E**) was not inhibited by FH535 under our experimental conditions (**Figure 5B**). These results suggest that FH535 selectively inhibited growth of TN breast cancer cells under conditions of an artificial three dimensional collagen matrix and thus this experimental system may provide a useful model for evaluating molecules that affect the growth of TN breast cancer.



**Figure 5. Proliferation of tumor cells in three-dimensional type I collagen gel in the presence of FH535.**

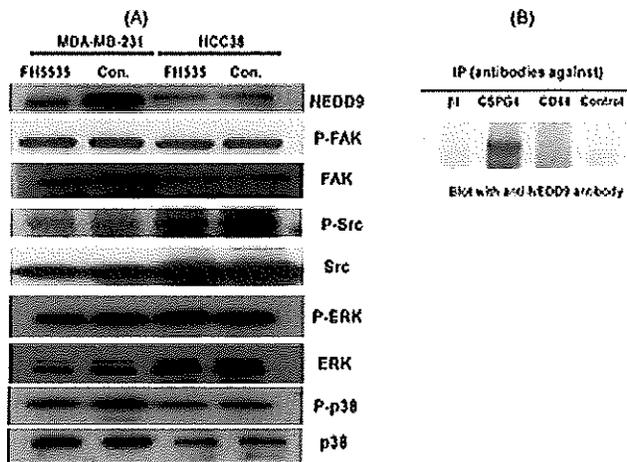
Cells (HCC38, MDA-MB-231, T47D, Sk-Br3, MCF-7) were cultured in type I collagen gel as described in material and methods. Cell/gel matrices were fixed and embedded in paraffin. The paraffin section was serially cut at 5  $\mu$ m and mounted on slide glasses. Tissue sections were stained with anti-Ki67 antibody or control IgG followed by HRP-conjugated secondary antibody as visualizing by DAB staining with counter staining by DAPI to localize cells. The results were demonstrated by the (mean  $\pm$  standard deviation) of % of positive cells for DAB from three independent experiments. Insets showed the typical staining pattern of anti-Ki67 antibody in HCC38 and MDA-MB-231 cells in the presence or absence of FH535 at a concentration of 10  $\mu$ M. Magnification x200. \*  $p < 0.001$  (by Student's two-tailed paired  $t$ -test). n.s: not significance (by Student's two-tailed paired  $t$ -test).

doi:10.1371/journal.pone.0044418.g005

**Download:** [PowerPoint slide](#) | [larger image \(722KB PNG\)](#) | [original image \(900KB TIFF\)](#)

## Expression of NEDD9 was Inhibited by FH535 in MDA-MB-231 Cells

In order to characterize mechanisms of tumor migration, invasion, and growth which were inhibited by FH535, we evaluated expression of various proteins important for promoting these processes by western blotting analysis. Among proteins tested, our results suggest that expression of NEDD9 was markedly inhibited in the presence of FH535 in MDA-MB-231 cells (**Figure 6A**). Previous studies suggest that activated FAK and Src phosphorylate NEDD9 [24]. However, FH did not affect expression or activation of these tyrosine kinases (**Figure 6A**). Similarly, expressions of Pyk2 or p130Cas were not inhibited in the presence of FH535 (**not shown**). Importantly, FH535 inhibited neither phosphorylation nor protein expression of ERK1/2 or p38MAPK (**Figure 6A**). These results suggest that the canonical WNT-signaling pathway would regulate specific gene expressions in MDA-MB-231 cells without affecting integrin-mediated cell adhesion followed by the activation of FAK and Src. Despite of the significant inhibition of migration and growth of HCC38 by FH535, we did not observe inhibition of NEDD9 in these cells (**Figure 6A**). Given the fact that NEDD9 promotes tumor cell migration and invasion [24], these results suggest that NEDD9 plays a key role in facilitating MDA-MB-231 cell migration, invasion, and growth.



**Figure 6. FH535 inhibited expression of NEDD9 in MDA-MB-231 cells.**

(A) MDA-MB-231 or HCC38 cells were cultured in the presence or absence of FH535 (1  $\mu$ M) overnight at 37°C. Cells were directly lysed in SDS sample buffer and separated on SDS-PAGE followed by transferred on to membranes. Membranes were blotted with anti-NEDD9,

anti-FAK, anti-phospho FAK, anti-phospho Src, anti-Src, anti-phospho Erk1/2, anti-Erk1/2, anti-phospho p38, and anti-p38 antibodies. (B) MD-MB-231 cells were cultured as described above. Cells were directly lysed in 100 mM Tris-HCl (pH7.4) containing 1% Brij35, 0.14 M NaCl, 1 mM CaCl<sub>2</sub>, 1 mM MgCl<sub>2</sub>, 1 mM MnCl<sub>2</sub>, and protease inhibitor cocktails. Cell lysates were precleared and immunoprecipitated with control, anti- $\beta$ 1, anti-CD44, or anti-CSPG antibodies for 4 hours. Bound proteins were released, separated, and transferred onto membranes. Membranes were blotted with anti-NEDD9 antibody.

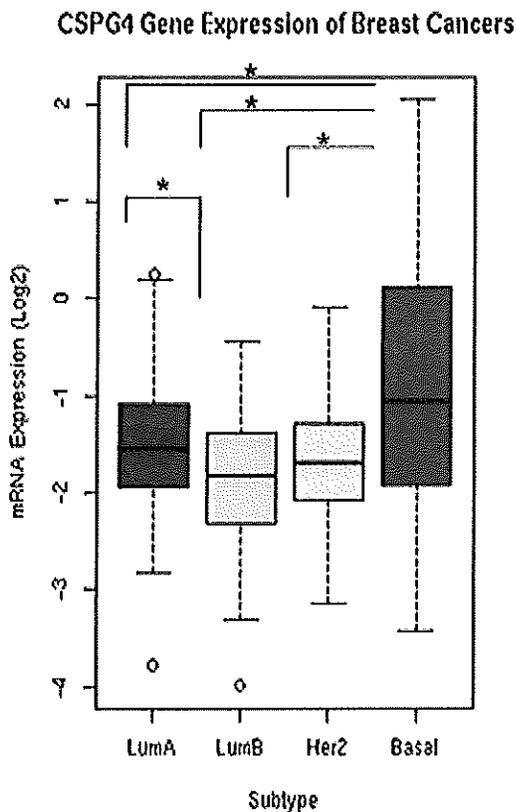
doi:10.1371/journal.pone.0044418.g006

**Download:** [PowerPoint slide](#) | [larger image \(907KB PNG\)](#) | [original image \(777KB TIFF\)](#)

It has been reported that NEDD9 associates with integrin-mediated signaling pathways by scaffolding various signaling molecules important for promoting migration and growth [25]. In order to test whether NEDD9 associates with cell adhesion receptors in MDA-MB-231, we performed co-immunoprecipitation assays as described in materials and methods. Cell lysates were precleared and precipitated with control antibody, anti-CD44, anti- $\beta$ 1 integrin, or anti-CSPG4 antibodies. The precipitates were separated on SDS-PAGE and blotted with anti-NEDD9 antibody. NEDD9 was specifically detected in anti-CSPG4 precipitate but not in precipitates of the other cell adhesion receptors (**Figure 6B**). Thus, our results suggest a novel model in which NEDD9 and CSPG4 form a molecular complex in MDA-MB-231 cells that stimulates formation of a signaling complex that contains various kinases such as FAK and Src, which play a key role in promoting migration, invasion, and proliferation in a type I collagen matrix.

## Gene Expression of CSPG4 in Human Breast Cancer Tissues

One-way between-subjects analysis of variance (ANOVA) was performed on the CSPG4 gene expression, as a function of breast cancer subtypes for basal, Her2-like, Luminal A, and Luminal B subtypes as determined by PAM50 [14]. The assumption of normality of error was met for each subtype. The assumption of homogeneity of variance was not met with Brown-Forsythe-test  $F(3, 510) = 37.63$ , ( $p < .0001$ ); however, using the Welch test of means to adjust for heterogeneity of variance did not change the final results. Therefore, the results of ANOVA are shown (**Figure 7**). We observed that there was a significant difference in the expression of CSPG4 among the subtypes,  $F(3, 510) = 24.62$ , ( $p < .0001$ ). Post hoc pairwise comparisons were performed using the Bonferroni adjustment. Tumors from patients with the basal subtype had significantly higher CSPG4 expression than the patients with other subtypes, with all the  $p$  values  $< .0001$ . In addition, Luminal A tumors also had a significantly higher CSPG4 expression than Luminal B tumors ( $p = 0.0006$ ).



**Figure 7. Gene expression profile of CSPG4 in breast cancer tissues.**

Gene expression data were retrieved from The Cancer

Genome Atlas (TCGA) project data portal (<https://tcga-data.nci.nih.gov/>). \*  $p < 0.001$  (by ANOVA).

doi:10.1371/journal.pone.0044418.g007

**Download:** [PowerPoint slide](#) | [larger image](#) (244KB PNG) | [original image](#) (455KB TIFF)

## DISCUSSION [Top](#)

It is widely recognized that triple negative (TN) breast cancers have a poorer prognosis than any other subtypes of breast cancer. In order to develop strategies and therapeutic reagents for this phenotype of breast cancers, we tested FH535 a compound known to inhibit invasion, migration, and

growth in vitro in other types of cancer. In this study, we utilized MDA-MB-231 and HCC38 cells to evaluate the canonical WNT-signaling pathway in triple negative (TN) breast cancers since this pathway is activated evaluated by LRP6 expression [26], suggesting that the WNT-signaling pathway is activated in these cell lines. In this study, we demonstrated that FH535 inhibited migration, invasion, and growth of breast cancer cell lines. In a three-dimensional collagen matrix, which is considered as a model system for evaluating cancer growth in vivo, FH535 selectively inhibited TN breast cancer cell lines (i.e. HCC38 and MDA-MB-231) but not breast cancer cell lines from other subtypes (i.e. SK-Br3, T47D, and MCF-7). These results suggest that the canonical WNT-signaling pathway could be a target for patients diagnosed with TN breast cancer.

Previous studies suggest that FH535 inhibits the transcription of TCF4, which is a critical protein regulating gene expression in colon cancer cells [8]. We did not observe similar inhibition by quantitative RT-PCR analysis in HCC38 cells (not shown). Instead, our results suggest that FH535 treatment reduced the protein levels of total  $\beta$ -catenin, which was different from what was found in previous studies [8]. As one mechanism for reducing  $\beta$ -catenin protein, we demonstrated that FH535 treatment increased the levels of axin. Axin has been characterized as a tumor-suppressor gene and plays a key role in inhibiting the canonical WNT-pathway by forming molecular complexes with other proteins such as GSK3 and APC [17]. In contrast, FH535 did not affect expression of  $\beta$ -catenin or axin in MDA-MB-231 cells, consistent with the previous studies [8]. Thus, it is possible that FH535 would exert its inhibitory activity for migration, invasion, and growth through distinct mechanisms in a cell-type dependent manner.

Enhanced migration and invasion through connective tissue is one of the key steps for establishing metastasis. Since type I collagen is a major component of connective tissues, we tested FH535 as a potential inhibitors in migration and growth in a type I collagen matrix. In addition to the inhibitory activity of FH535 on cell migration, our results demonstrated that FH535 significantly inhibited invasion into basement membrane (matrigel) in both MDA-MB-231 and HCC38 cells, consistent with the previous studies with human melanoma [9]. Our results suggest that the canonical WNT-signaling pathway would play a key role in tumor dissemination from the primary tumor by enhancing migratory and invasive abilities of tumor cells *in vivo*. Invasion of breast cancer cells into matrigel is mediated by

complex molecular mechanisms involving matrix degradation and/or cell shape change and includes activation of small GTPases and cytoskeletal reorganization. For example, it has been reported that collagenases such as MMP-1, MMP-13, and MMP-14 play a key role in promoting invasion and metastasis possibly by facilitating pericellular degradation of matrix proteins [27], [28], [29]. Recently, Poincloux et al, [30] proposed a unique model for invasion of MDA-MB-231 cells into matrigel, in which traction forces for cell movement in the matrix are generated by actomyosin-mediated formation of molecular complexes of  $\beta$ 1integrin and f-actin on the cell surface. In this regard, our current research focus is to characterize the mechanisms of the canonical WNT-pathway signaling that induce expression of matrix degrading enzymes and/or organizes cytoskeletal architectures in HCC38 and MDA-MB-231 cells toward a better understanding of cancer cell migration and invasion.

Recent studies suggest that three-dimensional (3D) cell culture models emulates a more physiologically relevant microenvironment [20]. For example, down-regulation of  $\beta$ 1integrin and growth factor signaling pathways in 3D culture systems resulted in reversion of the malignant phenotypes which were demonstrated as growth arrest and reformation of tissue polarity [31]. Keely and colleagues reported that a tumor microenvironment consisting of a type I collagen matrix regulates breast cancer proliferation and differentiation [23], [32], [33]. In this study, we demonstrated that FH535 significantly inhibited growth of TN breast cancer cell lines (i.e. HCC38 and MDA-MB-231) but not cell lines from other subtypes, suggesting a selectivity of growth inhibition of FH535 in breast cancer cells. Interestingly, FH535 did not induce apoptosis in MDA-MB-231 nor HCC38 cells evaluated staining with Apoptag (**not shown**). The mechanisms of tumor growth in a 3D type I collagen matrix is not clear at present, however, the requirement of signaling pathways to promote growth is likely dependent on cell types. For example T47D cells grow in the matrix through the activation of GTPase (i.e. Rho) and Rho kinase (ROCK) [23]. In human breast cancer tissues, immunohistochemical studies suggest that the canonical WNT signaling pathways are activated in TN compared to other subtypes [6], [34]. Given the fact that FH535 is an inhibitor for the canonical WNT signaling pathway [8], [9], [35], our results suggest a key role of this pathway in facilitating progression and metastasis of TN breast cancers.

NEDD9 has been evaluated in tumor progression in various tumor types including breast, lung, and colon cancers [36], [37], [38]. The fact that the NEDD9-null genetic background significantly limits mammary tumor initiation in mice strongly supports the notion of a key role of NEDD9 in promoting breast cancer progression and metastasis [39]. We demonstrated that FH535 inhibited the expression of NEDD9 by western blotting analysis in MDA-MB-231 cells, consistent with the recent studies identifying NEDD9 as a downstream target of the canonical WNT-signaling pathways [38]. Importantly, activation or expression of tyrosine kinases demonstrated to phosphorylate NEDD9 such as FAK and src were not affected by the treatment with FH535. These results suggest that the expression of NEDD9 would be responsible for promoting migration, invasion, and growth of MDA-MB-231 cells. Recent studies suggest that changes in the expression of NEDD9 were a potent prometastatic stimulus in various tumor cells. Thus, our results support the notion that NEDD9 promotes migration and invasion of TN breast cancer cells and would be a promising target for therapy of TN breast cancer [37].

Previous studies demonstrated that integrin and proteoglycan expressed in tumor cells promotes migration and invasion [40], [41]. In this study, we demonstrated that NEDD9 was specifically precipitated with CSPG4 but not with CD44 or  $\beta$ 1 integrin, suggesting that NEDD9 would form a molecular complex with CSPG4 in MDA-MB-231 cells. CSPG4 was first identified in human melanoma as a melanoma specific antigen and has been evaluated as a target for melanoma therapy [42], [43], [44]. We previously reported that CSPG4 plays a key role in promoting spreading, migration, invasion, as well as growth of malignant human melanoma cells [10], [12], [45], [46], [47], [48], [49]. CSPG4 is also expressed in various cancer cells including breast cancer cells and anti-CSPG4 antibody has been demonstrated to inhibit breast cancer growth and metastasis [50], [51], suggesting a key role in promoting breast cancer progression and metastasis. Thus, it is possible that CSPG4 expressed on breast cancer cells would transduce signals by forming a complex with NEDD9 for promoting migration, invasion, and growth in surrounding tissues. We are currently identifying key domains of NEDD9 that binds to CSPG4 to further understand the mechanisms of cancer progression and metastasis. Recent studies suggest that CSPG4 plays a key role in maintaining stem cell phenotypes in various tumor types such as melanoma, glioblastoma, and breast cancer cells [44], [52], [53]. The

characterization of CSPG4 interaction with NEDD9 may provide insights for understanding mechanisms of growth and survival of these tumor stem cells.

Using the TCGA breast cancer gene expression data and the PAM50-based subtypes [14], we confirmed that the basal subtype breast cancers express significantly higher levels of CSPG4 transcripts compared to tumors of other subtypes as described previously [51]. In addition, there is an apparent higher variance of the CSPG4 expression level in the basal subtype compared to the levels in other subtypes, which may be useful in the study of the heterogeneity of the basal subtype and/or the mechanism for the regulation of CSPG4 expression. Furthermore, Luminal A tumors express significantly higher CSPG4 compared to the Luminal B tumors. It will be interesting to find out whether there is a correlation between CSPG4 expression level and the outcome, which can be examined later when sufficient outcome data are available from the TCGA breast cancer project. Although the mechanism of expression of CSPG4 in basal and Luminal A tumors was not fully characterized, these results suggest that CSPG4 may serve as a target not only for basal but also Luminal A breast tumors.

In this study, we demonstrated a pivotal role for the canonical WNT-signaling pathway in enhancing tumor migration, invasion, and growth of TN breast cancer cells, MDA-MB-231 and HCC38. We identified NEDD9 as a potential downstream target of the canonical WNT-signaling pathway and showed that it form a complex with CSPG4 in MDA-MB-231 cells. We are currently identifying target molecules of FH535 in HCC38 by gene expression analyses. Recent studies suggest that the non-canonical WNT-signaling pathway plays a key role in promoting brain metastasis in breast cancer [54]. Thus, more comprehensive characterization of both the canonical and non-canonical WNT-signaling pathways by genomic and biological studies of breast cancer cell lines are clearly required for generating better treatments for patients who diagnosed with TN breast cancer.

## ACKNOWLEDGMENTS

---

The authors thank Dr. Patricia Keely (University of Wisconsin, Madison) for guidance with the three dimensional collagen gel assays. We thank to Brenda Deyarmin and Jennifer Kane for advising histology and immunohistochemistry procedures. We thank to Allyson Valente for running qRT-PCR reactions. The opinion and assertions contained herein are the private views of the authors and are

not to be construed as official or as representing the views of the Department of the Army or the Department of Defense.

## AUTHOR CONTRIBUTIONS

---

Conceived and designed the experiments: JI. Performed the experiments: JRL RC JD. Analyzed the data: JI RE SS CL YC HH. Wrote the paper: JI. Critical discussion and revising the draft: RJM CDS.

## REFERENCES

---

1. Matsuda Y, Schlange T, Oakeley EJ, Boulay A, Hynes NE (2009) WNT signaling enhances breast cancer cell motility and blockade of the WNT pathway by sFRP1 suppresses MDA-MB-231 xenograft growth. *Breast Cancer Res* 11: R32. [FIND THIS ARTICLE ONLINE](#)
2. Hayes MJ, Thomas D, Emmons A, Giordano TJ, Kleer CG (2008) Genetic changes of Wnt pathway genes are common events in metaplastic carcinomas of the breast. *Clin Cancer Res* 14: 4038–4044. [FIND THIS ARTICLE ONLINE](#)
3. Hennessy BT, Gonzalez-Angulo AM, Stemke-Hale K, Gilcrease MZ, Krishnamurthy S, et al. (2009) Characterization of a naturally occurring breast cancer subset enriched in epithelial-to-mesenchymal transition and stem cell characteristics. *Cancer Res* 69: 4116–4124. [FIND THIS ARTICLE ONLINE](#)
4. Lacroix-Triki M, Geyer FC, Lambros MB, Savage K, Ellis IO, et al. (2010) beta-catenin/Wnt signalling pathway in fibromatosis, metaplastic carcinomas and phyllodes tumours of the breast. *Mod Pathol* 23: 1438–1448. [FIND THIS ARTICLE ONLINE](#)
5. Lehmann BD, Bauer JA, Chen X, Sanders ME, Chakravarthy AB, et al. (2011) Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *J Clin Invest* 121: 2750–2767. [FIND THIS ARTICLE ONLINE](#)
6. Geyer FC, Lacroix-Triki M, Savage K, Arnedos M, Lambros MB, et al. (2011) beta-Catenin pathway activation in breast cancer is associated with triple-negative phenotype but not with CTNNB1 mutation. *Mod Pathol* 24: 209–231. [FIND THIS ARTICLE ONLINE](#)

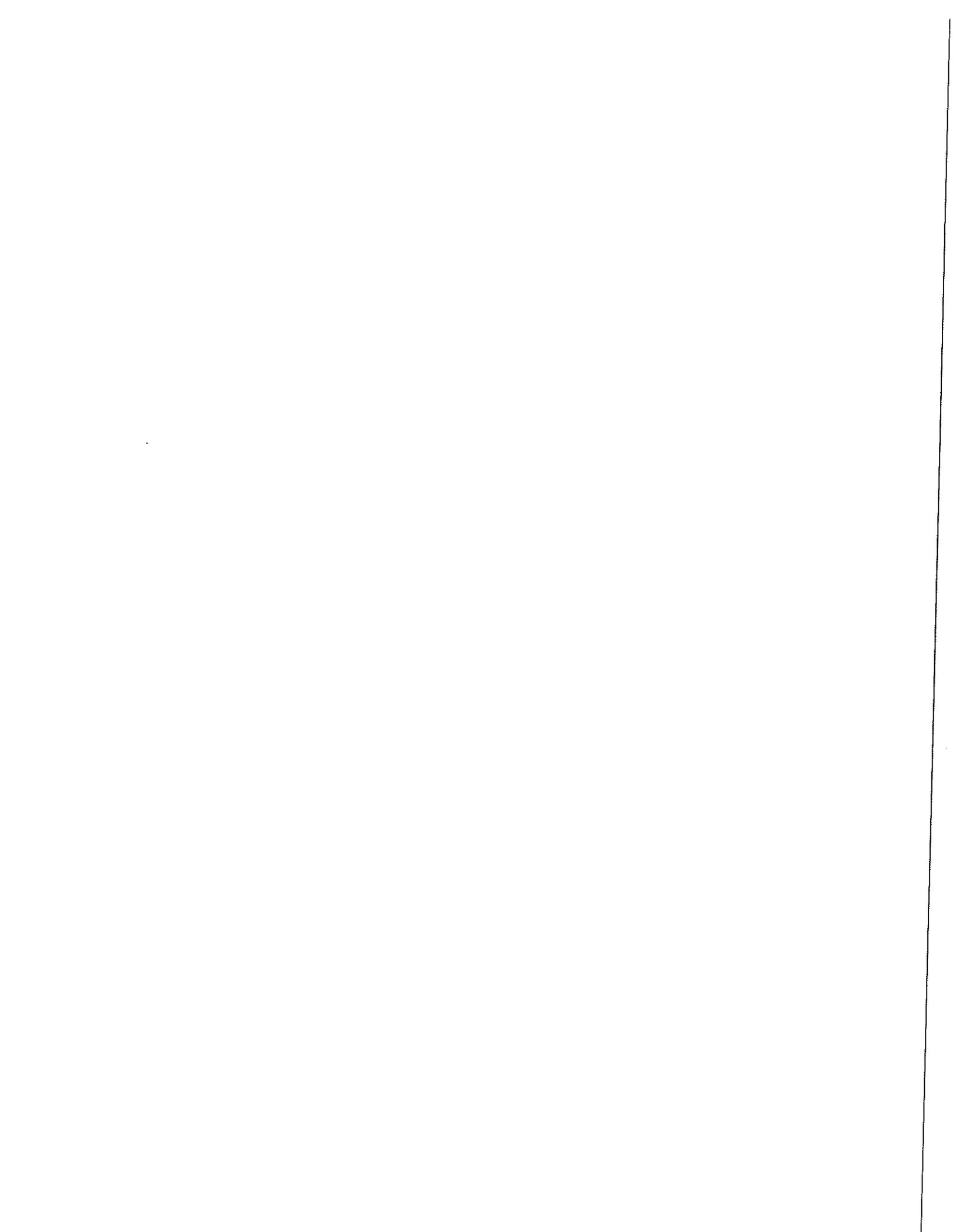
7. Khramtsov AI, Khramtsova GF, Tretiakova M, Huo D, Olopade OI, et al. Wnt/beta-catenin pathway activation is enriched in basal-like breast cancers and predicts poor outcome. *Am J Pathol* 176: 2911–2920. [\*\*FIND THIS ARTICLE ONLINE\*\*](#)
8. Handeli S, Simon JA (2008) A small-molecule inhibitor of Tcf/beta-catenin signaling down-regulates PPARgamma and PPARdelta activities. *Mol Cancer Ther* 7: 521–529. [\*\*FIND THIS ARTICLE ONLINE\*\*](#)
9. Vaid M, Prasad R, Sun Q, Katiyar SK (2011) Silymarin Targets beta-Catenin Signaling in Blocking Migration/Invasion of Human Melanoma Cells. *PLoS One* 6: e23000. [\*\*FIND THIS ARTICLE ONLINE\*\*](#)
10. Iida J, Pei D, Kang T, Simpson MA, Herlyn M, et al. (2001) Melanoma chondroitin sulfate proteoglycan regulates matrix metalloproteinase-dependent human melanoma invasion into type I collagen. *J Biol Chem* 276: 18786–18794. [\*\*FIND THIS ARTICLE ONLINE\*\*](#)
11. Iida J, Wilhelmson KL, Price MA, Wilson CM, Pei D, et al. (2004) Membrane type-1 matrix metalloproteinase promotes human melanoma invasion and growth. *J Invest Dermatol* 122: 167–176. [\*\*FIND THIS ARTICLE ONLINE\*\*](#)
12. Eisenmann KM, McCarthy JB, Simpson MA, Keely PJ, Guan JL, et al. (1999) Melanoma chondroitin sulphate proteoglycan regulates cell spreading through Cdc42, Ack-1 and p130cas. *Nat Cell Biol* 1: 507–513. [\*\*FIND THIS ARTICLE ONLINE\*\*](#)
13. Keely PJ, Fong AM, Zutter MM, Santoro SA (1995) Alteration of collagen-dependent adhesion, motility, and morphogenesis by the expression of antisense alpha 2 integrin mRNA in mammary cells. *J Cell Sci* 108 (Pt 2): 595–607. [\*\*FIND THIS ARTICLE ONLINE\*\*](#)
14. Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, et al. (2009) Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol* 27: 1160–1167. [\*\*FIND THIS ARTICLE ONLINE\*\*](#)
15. Carter WG, Wayner EA, Bouchard TS, Kaur P (1990) The role of integrins alpha 2 beta 1 and alpha 3 beta 1 in cell-cell and cell-substrate adhesion of human epidermal cells. *J Cell Biol* 110: 1387–1404. [\*\*FIND THIS ARTICLE ONLINE\*\*](#)

16. Languino LR, Gehlsen KR, Wayner E, Carter WG, Engvall E, et al. (1989) Endothelial cells use alpha 2 beta 1 integrin as a laminin receptor. *J Cell Biol* 109: 2455–2462. [\*\*FIND THIS ARTICLE ONLINE\*\*](#)
17. MacDonald BT, Tamai K, He X (2009) Wnt/beta-catenin signaling: components, mechanisms, and diseases. *Dev Cell* 17: 9–26. [\*\*FIND THIS ARTICLE ONLINE\*\*](#)
18. Knutson JR, Iida J, Fields GB, McCarthy JB (1996) CD44/chondroitin sulfate proteoglycan and alpha 2 beta 1 integrin mediate human melanoma cell migration on type IV collagen and invasion of basement membranes. *Mol Biol Cell* 7: 383–396. [\*\*FIND THIS ARTICLE ONLINE\*\*](#)
19. Bissell MJ (1999) Tumor plasticity allows vasculogenic mimicry, a novel form of angiogenic switch. A rose by any other name? *Am J Pathol* 155: 675–679. [\*\*FIND THIS ARTICLE ONLINE\*\*](#)
20. Bissell MJ, Labarge MA (2005) Context, tissue plasticity, and cancer: are tumor stem cells also regulated by the microenvironment? *Cancer Cell* 7: 17–23. [\*\*FIND THIS ARTICLE ONLINE\*\*](#)
21. Bissell MJ, Radisky D (2001) Putting tumours in context. *Nat Rev Cancer* 1: 46–54. [\*\*FIND THIS ARTICLE ONLINE\*\*](#)
22. Provenzano PP, Inman DR, Eliceiri KW, Knittel JG, Yan L, et al. (2008) Collagen density promotes mammary tumor initiation and progression. *BMC Med* 6: 11. [\*\*FIND THIS ARTICLE ONLINE\*\*](#)
23. Wozniak MA, Desai R, Solski PA, Der CJ, Keely PJ (2003) ROCK-generated contractility regulates breast epithelial cell differentiation in response to the physical properties of a three-dimensional collagen matrix. *J Cell Biol* 163: 583–595. [\*\*FIND THIS ARTICLE ONLINE\*\*](#)
24. Singh M, Cowell L, Seo S, O'Neill G, Golemis E (2007) Molecular basis for HEF1/NEDD9/Cas-L action as a multifunctional co-ordinator of invasion, apoptosis and cell cycle. *Cell Biochem Biophys* 48: 54–72. [\*\*FIND THIS ARTICLE ONLINE\*\*](#)
25. Cabodi S, del Pilar Camacho-Leal M, Di Stefano P, Defilippi P (2010) Integrin signalling adaptors: not only figurants in the cancer story. *Nat Rev Cancer* 10: 858–870. [\*\*FIND THIS ARTICLE ONLINE\*\*](#)
26. Liu CC, Prior J, Piwnicka-Worms D, Bu G (2010) LRP6 overexpression defines a class of breast cancer subtype and is a target for therapy. *Proc Natl Acad Sci U S A* 107: 5136–5141. [\*\*FIND THIS ARTICLE ONLINE\*\*](#)

27. Bravo-Cordero JJ, Marrero-Diaz R, Megias D, Genis L, Garcia-Grande A, et al. (2007) MT1-MMP proinvasive activity is regulated by a novel Rab8-dependent exocytic pathway. *EMBO J* 26: 1499–1510. **FIND THIS ARTICLE ONLINE**
28. Eck SM, Hoopes PJ, Petrella BL, Coon CI, Brinckerhoff CE (2009) Matrix metalloproteinase-1 promotes breast cancer angiogenesis and osteolysis in a novel in vivo model. *Breast Cancer Res Treat* 116: 79–90. **FIND THIS ARTICLE ONLINE**
29. Paquette B, Therriault H, Desmarais G, Wagner R, Royer R, et al. (2011) Radiation-enhancement of MDA-MB-231 breast cancer cell invasion prevented by a cyclooxygenase-2 inhibitor. *Br J Cancer* 105: 534–541. **FIND THIS ARTICLE ONLINE**
30. Poincloux R, Collin O, Lizarraga F, Romao M, Debray M, et al. (2011) Contractility of the cell rear drives invasion of breast tumor cells in 3D Matrigel. *Proc Natl Acad Sci U S A* 108: 1943–1948. **FIND THIS ARTICLE ONLINE**
31. Park CC, Zhang H, Pallavicini M, Gray JW, Baehner F, et al. (2006) Beta1 integrin inhibitory antibody induces apoptosis of breast cancer cells, inhibits growth, and distinguishes malignant from normal phenotype in three dimensional cultures and in vivo. *Cancer Res* 66: 1526–1535. **FIND THIS ARTICLE ONLINE**
32. Kenny PA, Lee GY, Bissell MJ (2007) Targeting the tumor microenvironment. *Front Biosci* 12: 3468–3474. **FIND THIS ARTICLE ONLINE**
33. Provenzano PP, Eliceiri KW, Keely PJ (2009) Shining new light on 3D cell motility and the metastatic process. *Trends Cell Biol* 19: 638–648. **FIND THIS ARTICLE ONLINE**
34. King TD, Suto MJ, Li Y (2011) The Wnt/beta-catenin signaling pathway: A potential therapeutic target in the treatment of triple negative breast cancer. *J Cell Biochem*.
35. Lim SK, Orhant-Prioux M, Toy W, Tan KY, Lim YP (2011) Tyrosine phosphorylation of transcriptional coactivator WW-domain binding protein 2 regulates estrogen receptor {alpha} function in breast cancer via the Wnt pathway. *Faseb J* 25: 3004–3018. **FIND THIS ARTICLE ONLINE**
36. Chang JX, Gao F, Zhao GQ, Zhang GJ (2012) Expression and clinical significance of NEDD9 in lung tissues. *Med Oncol*.

37. Kong C, Wang C, Wang L, Ma M, Niu C, et al. (2011) NEDD9 is a positive regulator of epithelial-mesenchymal transition and promotes invasion in aggressive breast cancer. PLoS One 6: e22666. [\*\*FIND THIS ARTICLE ONLINE\*\*](#)
38. Li Y, Bavarva JH, Wang Z, Guo J, Qian C, et al. (2011) HEF1, a novel target of Wnt signaling, promotes colonic cell migration and cancer progression. Oncogene 30: 2633–2643. [\*\*FIND THIS ARTICLE ONLINE\*\*](#)
39. Izumchenko E, Singh MK, Plotnikova OV, Tikhmyanova N, Little JL, et al. (2009) NEDD9 promotes oncogenic signaling in mammary tumor development. Cancer Res 69: 7198–7206. [\*\*FIND THIS ARTICLE ONLINE\*\*](#)
40. Iida J, Meijne AM, Knutson JR, Furcht LT, McCarthy JB (1996) Cell surface chondroitin sulfate proteoglycans in tumor cell adhesion, motility and invasion. Semin Cancer Biol 7: 155–162. [\*\*FIND THIS ARTICLE ONLINE\*\*](#)
41. Iozzo RV, Sanderson RD (2011) Proteoglycans in cancer biology, tumour microenvironment and angiogenesis. J Cell Mol Med 15: 1013–1031. [\*\*FIND THIS ARTICLE ONLINE\*\*](#)
42. Campoli M, Ferrone S, Wang X (2010) Functional and clinical relevance of chondroitin sulfate proteoglycan 4. Adv Cancer Res 109: 73–121. [\*\*FIND THIS ARTICLE ONLINE\*\*](#)
43. Price MA, Colvin Wanshura LE, Yang J, Carlson J, Xiang B, et al. (2011) CSPG4, a potential therapeutic target, facilitates malignant progression of melanoma. Pigment Cell Melanoma Res 24: 1148–1157. [\*\*FIND THIS ARTICLE ONLINE\*\*](#)
44. Wang X, Wang Y, Yu L, Sakakura K, Visus C, et al. (2010) CSPG4 in cancer: multiple roles. Curr Mol Med 10: 419–429. [\*\*FIND THIS ARTICLE ONLINE\*\*](#)
45. Iida J, Meijne AM, Oegema TR, Yednock TA, Kovach NL, et al. (1998) A role of chondroitin sulfate glycosaminoglycan binding site in alpha4beta1 integrin-mediated melanoma cell adhesion. J Biol Chem 273: 5955–5962. [\*\*FIND THIS ARTICLE ONLINE\*\*](#)
46. Iida J, Meijne AM, Spiro RC, Roos E, Furcht LT, et al. (1995) Spreading and focal contact formation of human melanoma cells in response to the stimulation of both melanoma-associated proteoglycan (NG2) and alpha 4 beta 1 integrin. Cancer Res 55: 2177–2185. [\*\*FIND THIS ARTICLE ONLINE\*\*](#)

47. Iida J, Wilhelmson KL, Ng J, Lee P, Morrison C, et al. (2007) Cell surface chondroitin sulfate glycosaminoglycan in melanoma: role in the activation of pro-MMP-2 (pro-gelatinase A). *Biochem J* 403: 553–563. **[FIND THIS ARTICLE ONLINE](#)**
48. Yang J, Price MA, Li GY, Bar-Eli M, Salgia R, et al. (2009) Melanoma proteoglycan modifies gene expression to stimulate tumor cell motility, growth, and epithelial-to-mesenchymal transition. *Cancer Res* 69: 7538–7547. **[FIND THIS ARTICLE ONLINE](#)**
49. Yang J, Price MA, Neudauer CL, Wilson C, Ferrone S, et al. (2004) Melanoma chondroitin sulfate proteoglycan enhances FAK and ERK activation by distinct mechanisms. *J Cell Biol* 165: 881–891. **[FIND THIS ARTICLE ONLINE](#)**
50. Cooney CA, Jousheghany F, Yao-Borengasser A, Phanavanh B, Gomes T, et al. (2011) Chondroitin sulfates play a major role in breast cancer metastasis: a role for CSPG4 and CHST11 gene expression in forming surface P-selectin ligands in aggressive breast cancer cells. *Breast Cancer Res* 13: R58. **[FIND THIS ARTICLE ONLINE](#)**
51. Wang X, Osada T, Wang Y, Yu L, Sakakura K, et al. (2010) CSPG4 protein as a new target for the antibody-based immunotherapy of triple-negative breast cancer. *J Natl Cancer Inst* 102: 1496–1512. **[FIND THIS ARTICLE ONLINE](#)**
52. Legg J, Jensen UB, Broad S, Leigh I, Watt FM (2003) Role of melanoma chondroitin sulphate proteoglycan in patterning stem cells in human interfollicular epidermis. *Development* 130: 6049–6063. **[FIND THIS ARTICLE ONLINE](#)**
53. Svendsen A, Verhoeff JJ, Immervoll H, Brogger JC, Kmiecik J, et al. (2011) Expression of the progenitor marker NG2/CSPG4 predicts poor survival and resistance to ionising radiation in glioblastoma. *Acta Neuropathol* 122: 495–510. **[FIND THIS ARTICLE ONLINE](#)**
54. Klemm F, Bleckmann A, Siam L, Chuang HN, Rietkotter E, et al. (2011) beta-catenin-independent WNT signaling in basal-like breast cancer and brain metastasis. *Carcinogenesis* 32: 434–442. **[FIND THIS ARTICLE ONLINE](#)**



# A biomedical informatics infrastructure to support translational research

Hai Hu, PhD

Senior Director of Biomedical Informatics  
Deputy Chief Scientific Officer  
Windber Research Institute

NCICBIT Speaker Series • September 5, 2012

# Windber Research Institute

---

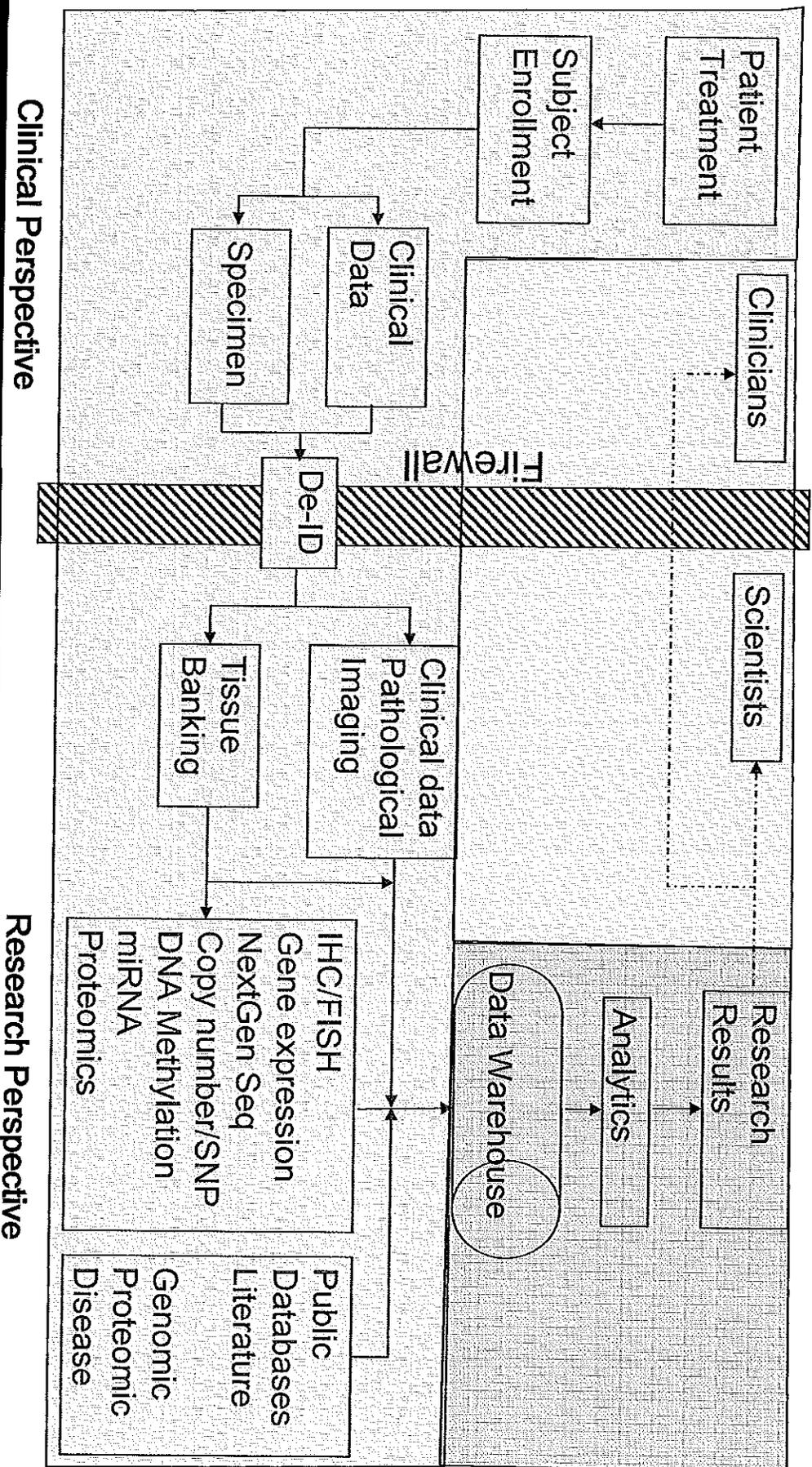
Founded: 2001

Staff: ~ 50

## Main Programs:

- Breast cancer—The Clinical Breast Care Project (CBCP)
  - Clinical Care
  - Risk Reduction
  - Tissue Banking
  - Focused Research
  - Biomedical Informatics
- Gynecological cancer—The Gynecological Disease Program (GDP)
- Cardiovascular diseases
- Health promotion and disease prevention

# Biomedical Informatics for Translational Research



Adapted from Hu H. et al. 2008

# Biomedical Informatics

---

- Data collection and generation
  - Laboratory Information Management System (LIMS)
  - Quality Assurance (QA) programs
- Data centralization
  - Data Warehouse for Translational Research (DW4TR)
  - Data models
  - Interfaces
- Applications
  - Data analysis and mining using the DW4TR
  - A Komen Promise Grant
  - The Cancer Genome Atlas Project—Breast Cancer

# Clinical LIMS for CBCP—CLWS

Designed by WRI

Implemented by Cimarron

Subject enrollment

Questionnaire printing / barcode linking

Questionnaire completion

Double data entry

Specimen processing / barcode linking

Specimen transfer

Specimen receiving / tissue banking

Laser capture microdissection

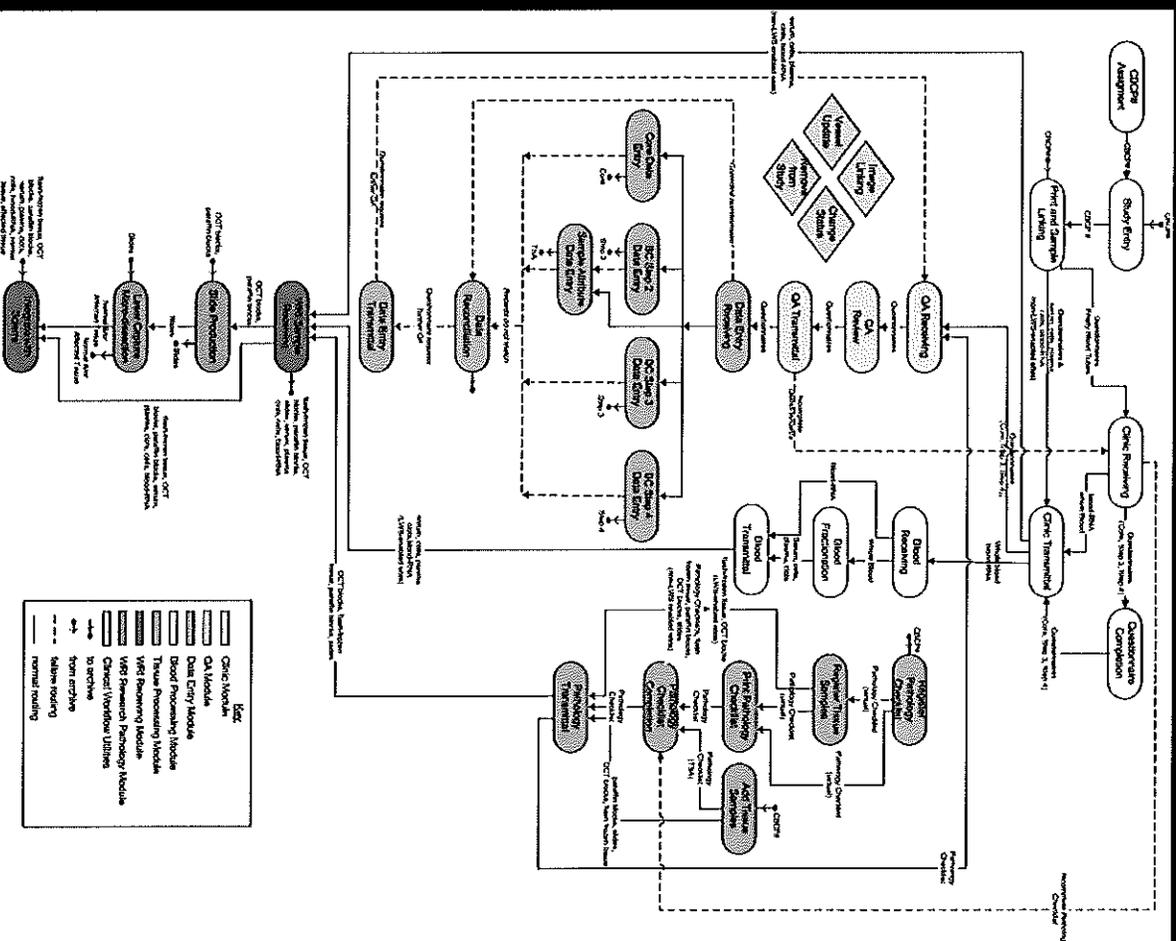
Specimen info integration with lab modules

Reporting utilities

Problem management

Error correction

Deployed in November, 2004



# Questionnaire Data Entry

Scierra

File Edit View Applications Tools Window Help

Start Work: <can>



## Workflow Navigation

- QA
- QA Receiving
- QA Review
- QA Transmittal
- Data Entry
- Data Entry Receiving
- Core Data Entry
- BC Step 2 Data Entry
- Sample Attribute Data Entry
- BC Step 3 Data Entry
- BC Step 4 Data Entry
- Data Reconciliation
- Data Entry Transmittal
- Blood Processing
- Blood Receiving
- Blood Fractionation
- Blood Transmittal
- Tissue Processing
- Register Pathology Checklis

App Status Login successful for user hhu.

## QMS: Core Data Entry

Barcode	Filter	Selected Items	Selected	Completion Date	Status	Work Item Queue Time
00002DZN		100000348		Sep 30, 2005	None	6d 04:30
00002DZT		100000348		Dec 5, 2005	None	6d 04:30
00002DZV		100000351		Dec 13, 2005	None	6d 04:30
00002DZS		100000347		Nov 7, 2005	None	6d 04:30

Protocol: Core 4/14/2003  
 Questions 1 - 170 | Questions 171 - 207: BC | Questions 208 - 225: HR | Questions 226 - 245: Geo

1. CBCP# 100000351

2. Date

3. Date of Birth

4. Age

5. Sex

6. Ethnic Background

6.1. Other Ethnic Background

7. Marital Status

8. Education

9. Occupation

10. Size of Family

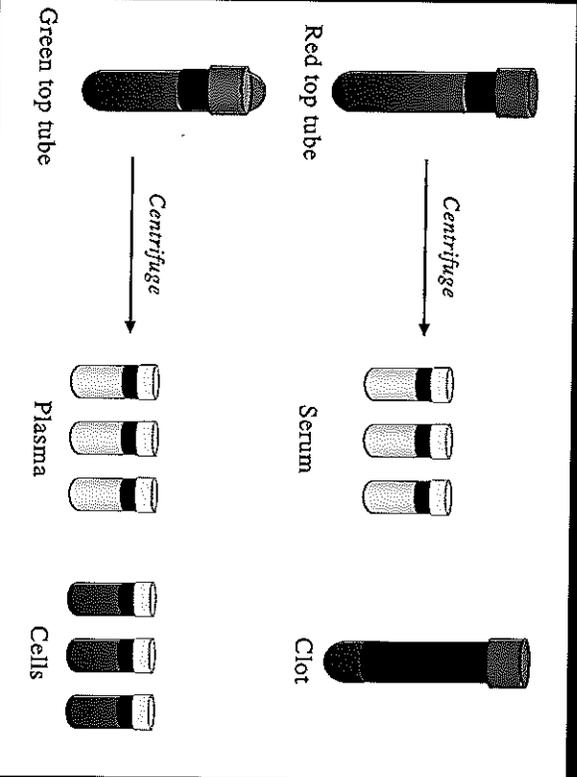
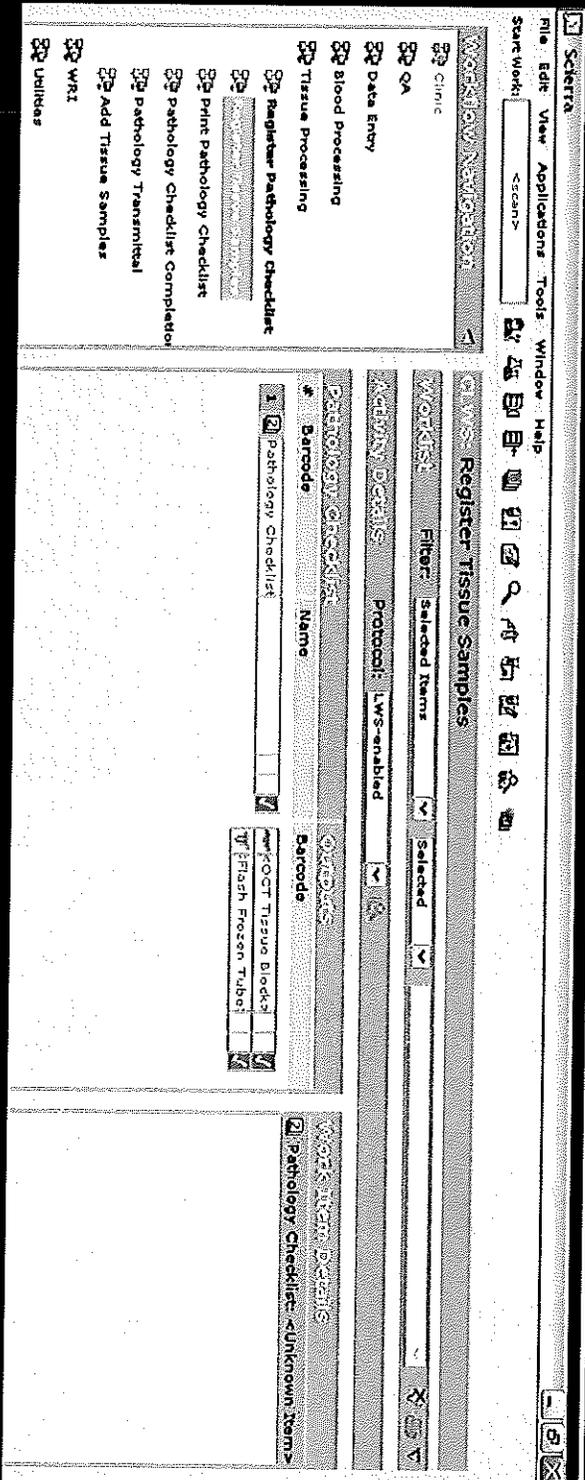
10.1. Number of brothers

Barcode: <00002DZV>

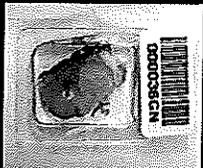
Work Item Details  
 Core Questionnaire: 00002DZV  
 Route To:

- Finish
- Clear
- Close
- Help
- Save

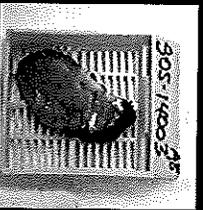
# Biospecimen Processing



Tissue



OCT



FFPE

Courtesy of Dr. Stella Somlari

# Tissue Banking

Start  Edit  View  Applications  Tools  Window  Help  
 Start Work:

**Location Contents**  
 Path:  
 IS Freezer 3(1-03)  
   06  
     IS Freezer 3, Rack 06(1-0306)  
       12  
         IS030612(1-030612)

Find Location:   
 Barcode:   
 Location Name:   
 Object Barcode:   
 Find contained location of type:

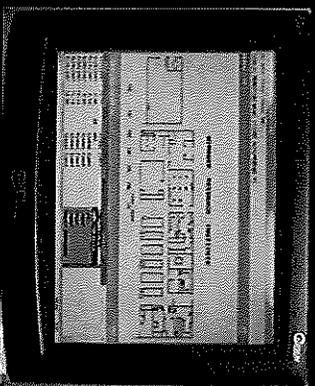
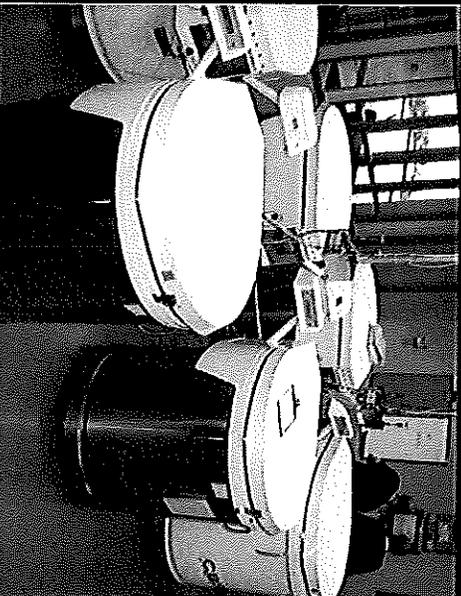
Name: IS030612  
 Barcode: IS030612  
 Type: Shelf  
 Comments:

Name	Barcode	Type	Position
0000280M	0000280M	Plasma Tube	5I
0000271Y	0000271Y	Plasma Tube	7H
000027C6	000027C6	Plasma Tube	8G
0000270K	0000270K	Plasma Tube	1E

Worklists:     
 Integrates...

Assign:   
 Remove

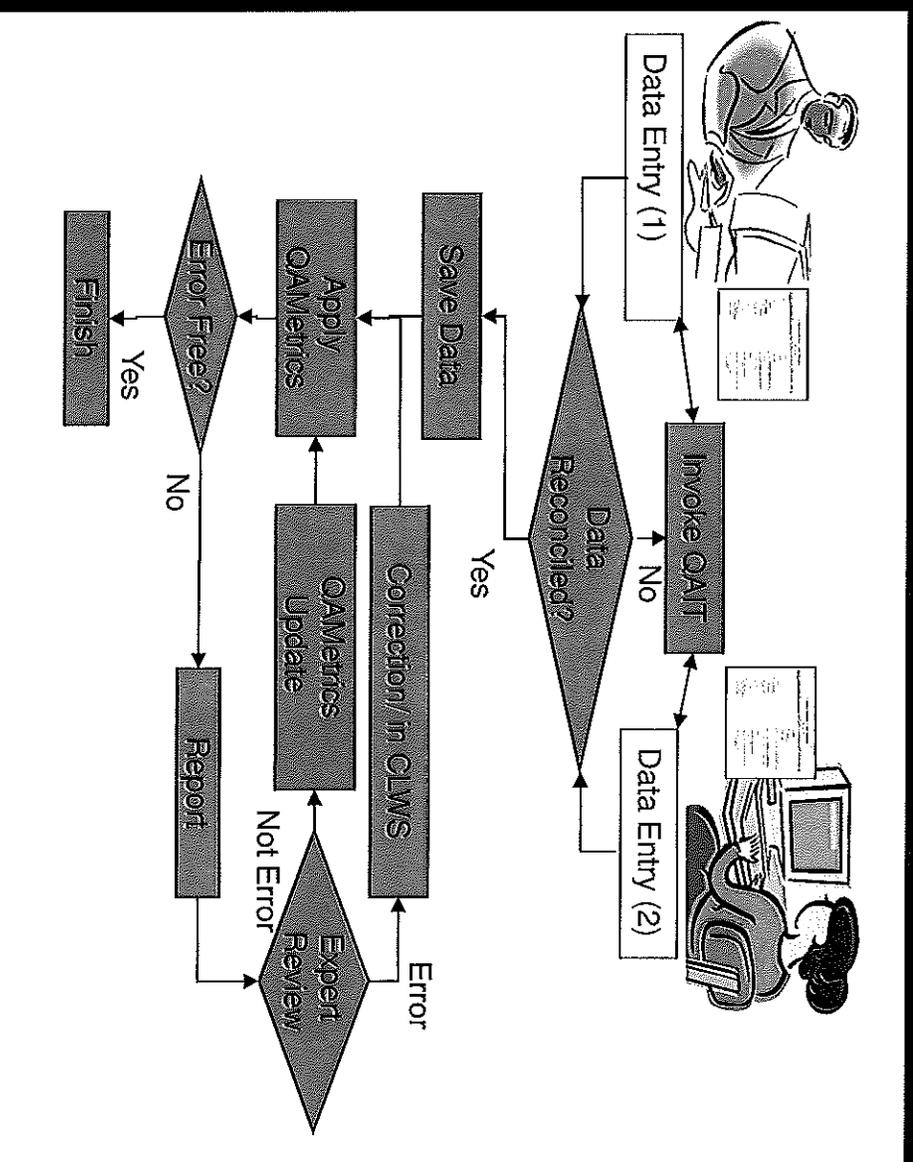
Order Layer Positions by:  Row  Column



Courtesy of Dr. Stella Somiari

# Clinical Data QA

SOPs  
Clinical QA review  
Double-data entry  
QAMetrics  
Editing utility for CLWS  
QA Issue Tracking System



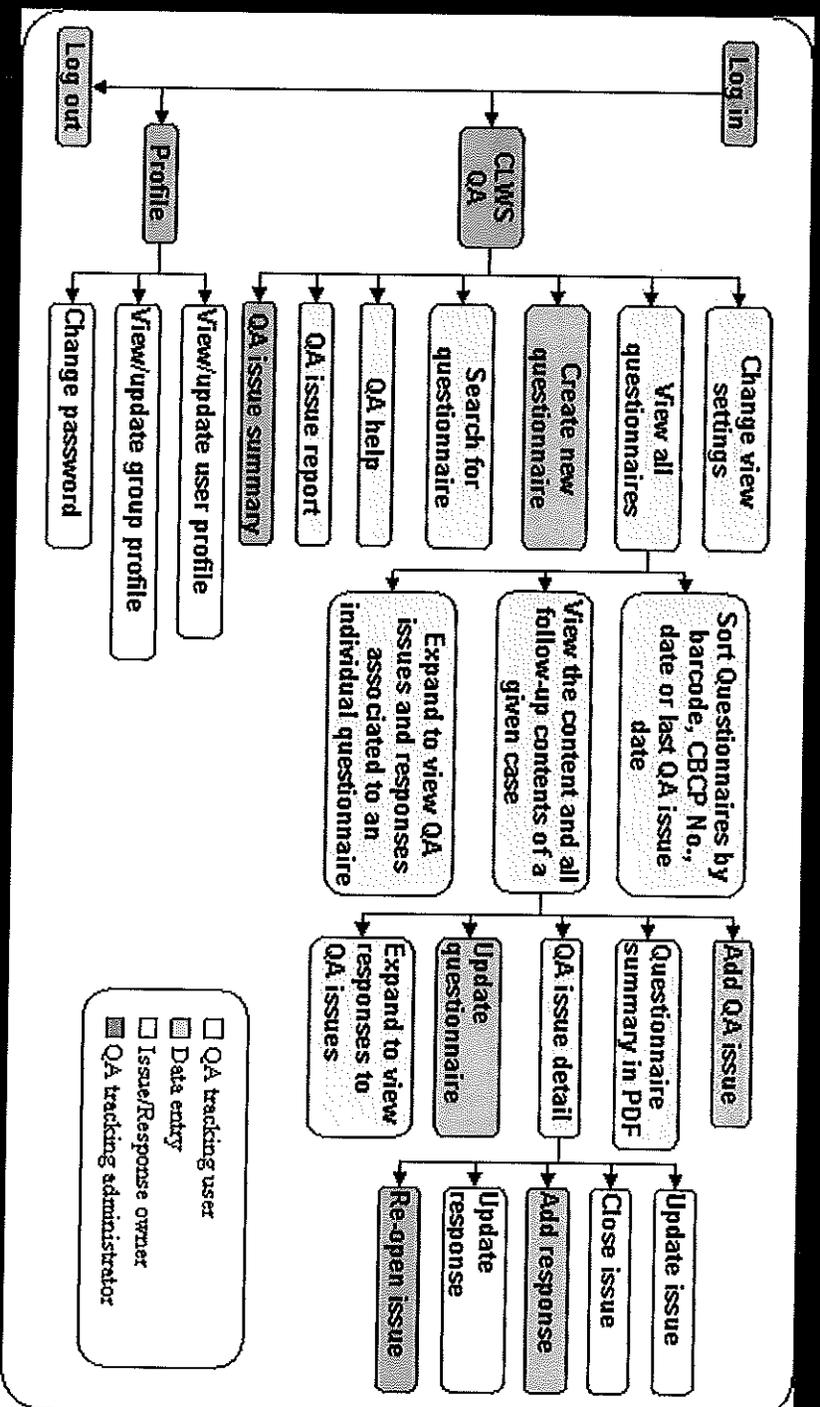
# QA Issue Tracking System (QAITS)

To facilitate QA communications

- Web-based
- Backend MySQL
- Front end IIS/PHP/JavaScript

Main functions:

- QA issue general view
- QA issue summary and report
- User management

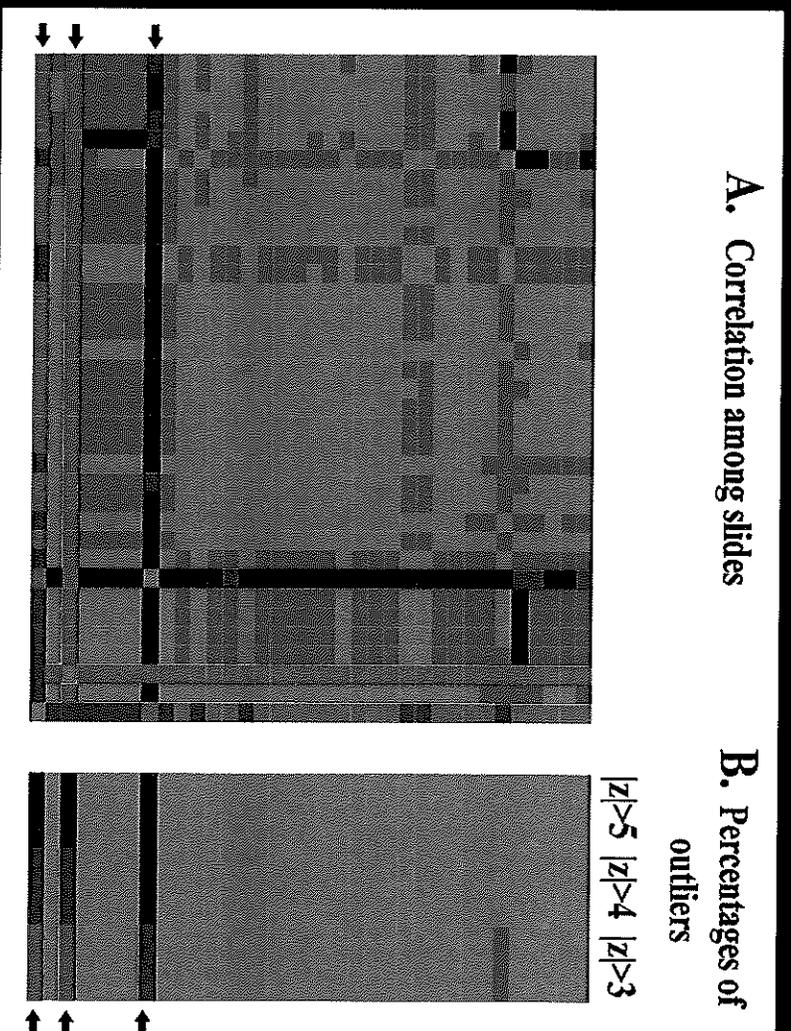


QA tracking user  
 Data entry  
 Issue/Response owner  
 QA tracking administrator

# Microarray Data QA

---

1. In-chip QA, following manufacturer's manual and published methods
2. Cross-chip QA, using Pearson Correlation and z-score based outlier%



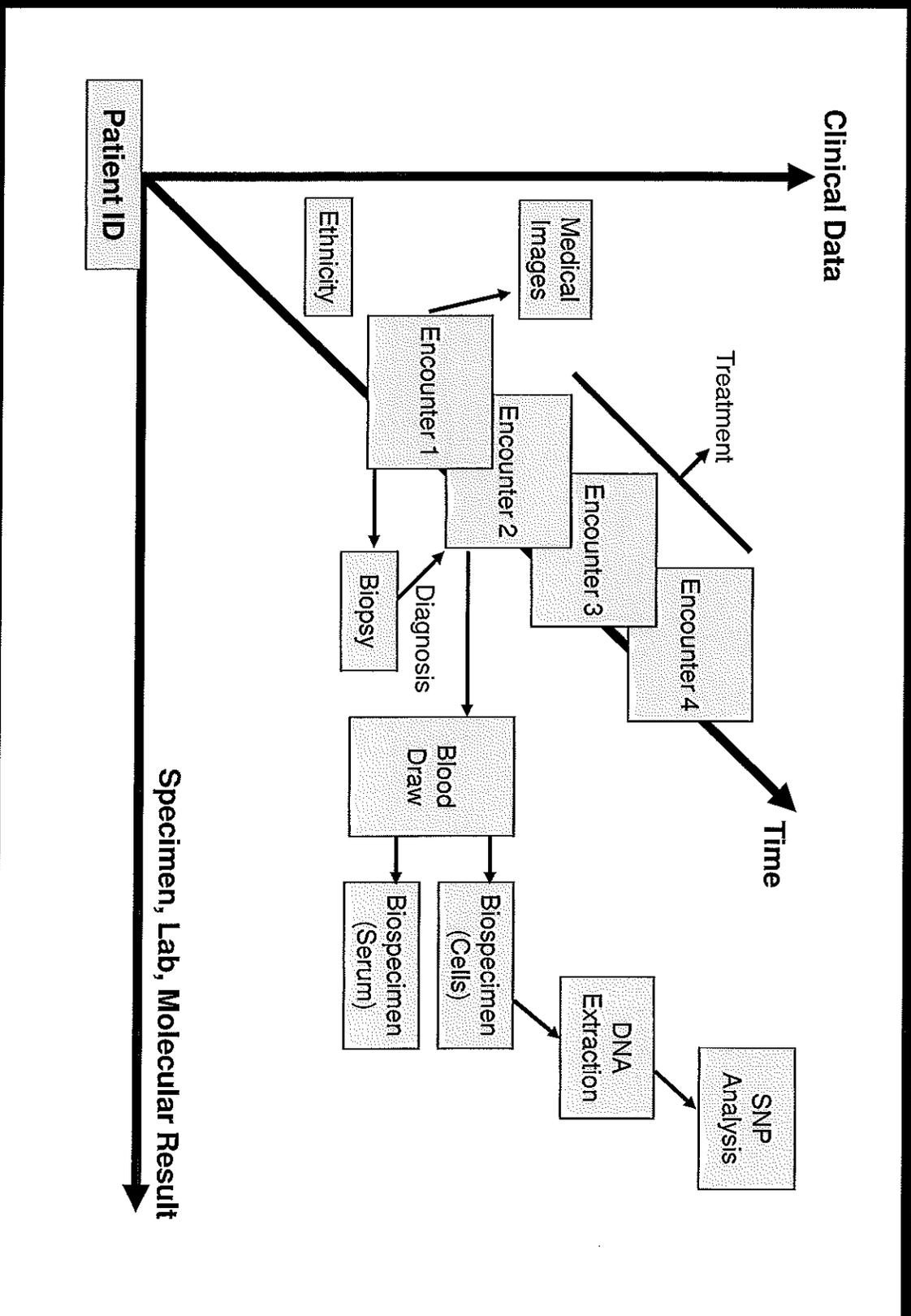
Yang S, Guo X, and Hu H. Proteomics, Genomics, and Bioinformatics 2009  
Yang S, ....., and Hu H. Cancer Informatics 2006

# Biomedical Informatics

---

- Data collection and generation
  - Laboratory Information Management System (LIMS)
  - Quality Assurance (QA) programs
- Data centralization
  - Data Warehouse for Translational Research (DW4TR)
  - Data models
  - Interfaces
- Applications
  - Data analysis and mining using the DW4TR
  - A Komen Promise Grant
  - The Cancer Genome Atlas Project—Breast Cancer

# DW4TR—Data Relationships



# Classifications of Breast Cancer

## MeSH

### 3. Diseases (C)

- Skin and Connective Tissue Diseases [C17]
- Skin Diseases [C17.800]
  - Breast Diseases [C17.800.090]
  - Breast Neoplasms [C17.800.090.500]

### 3. Diseases (C)

- Neoplasms [C04]
- Neoplasms by Site [C04.588]
- Breast Neoplasms [C04.588.180]

## SNOMED CT Concept (SNOMED RT+CTV3)

- Clinical finding (finding)
- Finding by site (finding)
  - Finding of body region (finding)
    - Finding of trunk structure (finding)
      - Finding of region of thorax (finding)
        - Breast finding (finding)
          - Disorder of breast (disorder)
            - Neoplasm of breast (disorder)

# Patient-Centric Clinical Data Model

---

## Modules

- Attributes
- Sub-modules
- Disease independent/specific

## References

- Medical textbooks
- MD Clinical experiences
- MESH, SNOMED

- ☞ Medical History
- ☞ Physical Examination
- ☞ Diagnostics
  - ☞ Lab Tests
  - ☞ Biomarker Studies
  - ☞ Ancillary studies
  - ☞ ER
  - ☞ HER2
  - ☞ Ki67
  - ☞ p53
  - ☞ PR
- ☞ Microarray Studies
- ☞ Pathological Diagnosis
- ☞ Prognosis
- ☞ Surgical Procedures
- ☞ Tissue Sample
- ☞ Therapeutics
- ☞ Surveys and Consents

# Specimen-Centric Molecular Data Model

Sample		Microarray	
CBCP#	cannot enter duplicate samples in genomics	User Name	Person doing the Lab experiment
Date	Date tissue was obtained	Project Name	Study Name
Sample type	DNA/RNA/Protein	Investigator	Study Investigator
Volume	volume of sample used	Date	Date of Microarray exp readout
Conc. of DNA	DNA conc. In the sample volume	Batch/Lot Number	Microarray chip Number
Sample Barcode	CLWS Sample Barcode	Vendor of Chip	affymetrix
DNA Barcode	Freezer Works Barcode	Tissue	tissue used for study
Derivative	Tissue -> Sectioning/ Shaving/Flash Frozen: Blood -> Pax/Red/Green	Call Rate	Depends upon tissue type
Experiment		3/5 ratio	House keeping genes
Experimental Design	.pdf file location	Pass/Fail	Signal Detected - Y/N
RIN	PCR -number	MDR	?Multi Dimensional Ranger
DNA Isolation Protocol	Text	MCR	? multivariate curve resolution
260/280 ratio	DNA conc	Confidence Factor	
		Image Quality	burns in image or any other mishap
		Image of the chip	.jpg file
		Report File	.rpt location of report file
		Data Analysis	
		Software Used	GeneChip
		Sample File	Sample Annotation file Location - .xls
		.CEL or .GPR	raw Cell Intensity data file
		.DAT	Image of Scanned probe array file
		.CHP	Analysis/results file
		XML	
		.JPG	Image file Location
		*MR.CHP	
		.rpt	Report file ~ .xml?

# Temporal Data Model

## Three data types:

Static – not related to a time point

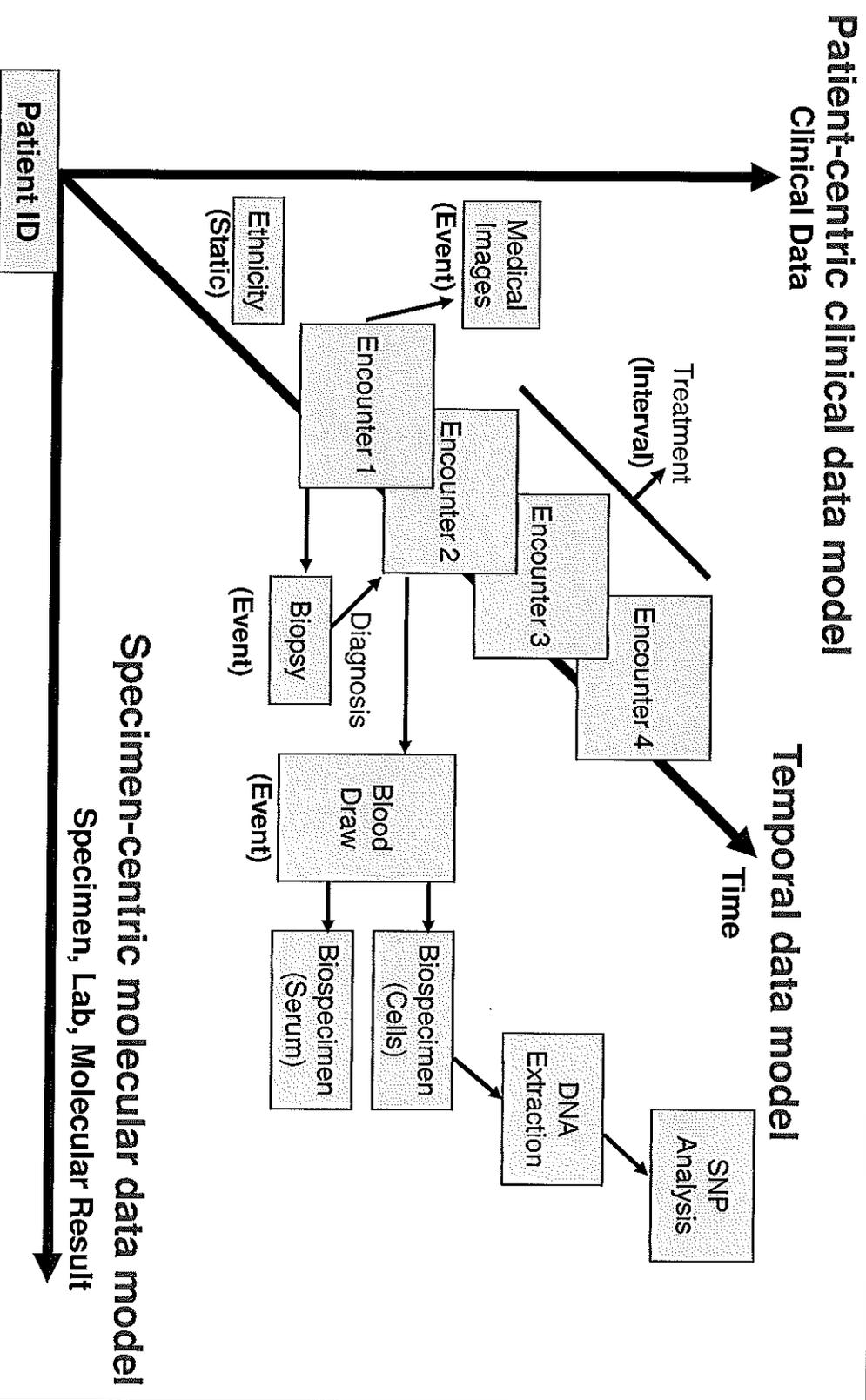
Event – with a specific time point

Interval – with a starting and ending time point

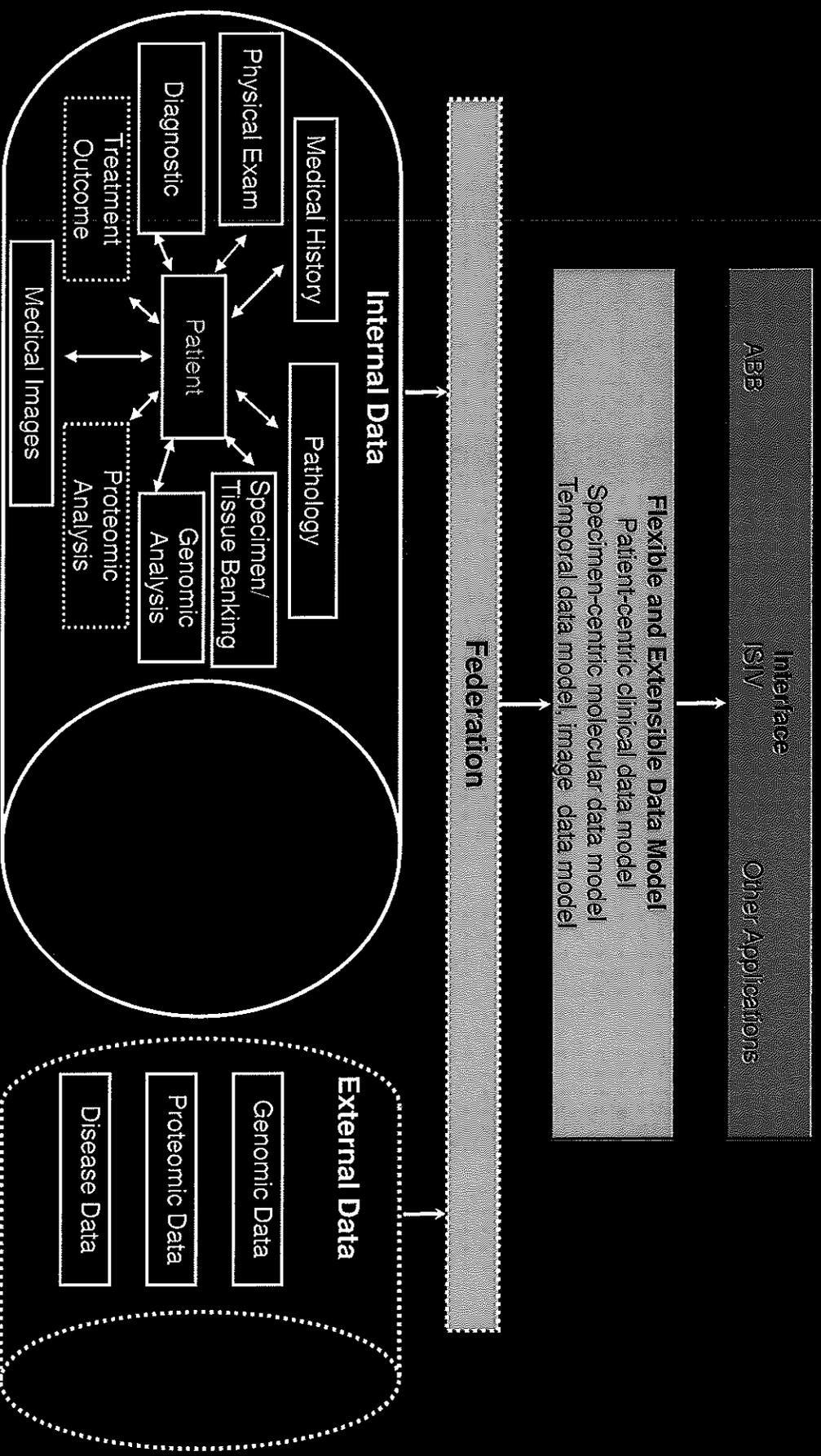
## Illustrative data model examples:

ID	Attribute	Facet	Data Type	Units	Multiple Values	From Date	To Date	Temporal	Parent ID
23	Ethnic Group		VARCHAR2		Yes			Static	
1541	Surgical Proc		COMPLEX		No	mm/dd/yyyy		Event	
1542	Surgical Proc	Name	VARCHAR2		No	mm/dd/yyyy		Event	1541
1543	Surgical Proc	Code	VARCHAR2		No	mm/dd/yyyy		Event	1541
1544	Surgical Proc	Laterality	VARCHAR2		No	mm/dd/yyyy		Event	1541
103	Alcohol Use		COMPLEX		No	mm/dd/yyyy	mm/dd/yyyy	Interval	
104	Alcohol Use	Name	VARCHAR2		No	mm/dd/yyyy	mm/dd/yyyy	Interval	103
105	Alcohol Use	Frequency	NUMBER	/day	No	mm/dd/yyyy	mm/dd/yyyy	Interval	103
106	Alcohol Use	Amount	NUMBER	drinks	No	mm/dd/yyyy	mm/dd/yyyy	Interval	103

# DW4TR—Data Models



# Data Warehouse for Translational Research (DW4TR)



# DW4TR Interfaces

---

## Users

- General users: clinicians and lab scientists
- Data analysis specialists: statisticians
- Intermediate users: bioinformaticians

## Data types

- Clinicopathological
- Molecular—specialized software
- Imaging—specialized software

## Analytical needs

- Aggregated information for overall studies
- Individual subject's information

The whole DW4TR was a joint development with InforSense (now part of IDBS)

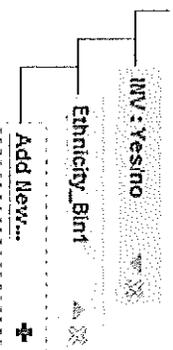
# Aggregated Biomedical-Information Browser (ABB)

Home | Explore Patient Data | Edit Timelines | Cohort View | Current Study: vxru\_cohort\_51



⌘ Add and edit rows

The row hierarchy determines which characteristics is shown next when you expand a row. Click on Attribute names and then right-click to save the selection as a new Cohort, or send to an IntraSense service.



All items		5664	100	52.3
INV : Yes/no		2523	41.46	49.76
NO				
Ethnicity_Bin1	INV : Yes/no = NO			
Caucasian American		1554	61.59	51.51
African American		516	20.45	46.62
Other		213	8.44	44.89
YES		1614	26.52	58.36
Ethnicity_Bin1	INV : Yes/no = YES			
Caucasian American		1207	74.78	59.2
African American		235	14.56	54.07
Other		75	4.65	57.96

To get started, click the Edit Rows or Edit Columns buttons on the toolbar above.

# Aggregated Biomedical-Information Browser (ABB)

☒ Patient Count

Pathologic category

Choose an attribute

Plain Text

Define Time Period

All Items	Patient Count		Pathologic category					
	ER: Result	PR: Result	NEGATIVE	POSITIVE	NOT PERFORMED	Value not set	Value not set	Value not set
ER: Result	5507	1616	131	1701	12	1	355	1743
PR: Result	381	345	1	7	3	n/a	34	n/a
NEGATIVE	362	329	1	5	3	n/a	32	n/a

HER2 by IHC : Result

ER: Result	PR: Result	NEGATIVE	POSITIVE	NOT PERFORMED	Value not set
ER: Result = NEGATIVE	ER: Result = NEGATIVE	176	171	1	3
PR: Result = NEGATIVE	PR: Result = NEGATIVE	118	115	n/a	n/a
NEGATIVE	NEGATIVE	71	48	n/a	2
POSITIVE	POSITIVE	15	12	n/a	1
NOT PERFORMED	NOT PERFORMED	35	31	1	2
Value not set	Value not set	2	1	n/a	1
Value not set	Value not set	1	1	n/a	n/a
Value not set	Value not set	1392	1220	13	14
Value not set	Value not set	5	3	n/a	n/a
Value not set	Value not set	3758	74	118	1681

ER: Result = NEGATIVE,  
PR: Result = NEGATIVE

ER: Result	PR: Result	NEGATIVE	POSITIVE	NOT PERFORMED	Value not set
ER: Result = NEGATIVE	ER: Result = NEGATIVE	176	171	1	3
PR: Result = NEGATIVE	PR: Result = NEGATIVE	118	115	n/a	n/a
NEGATIVE	NEGATIVE	71	48	n/a	2
POSITIVE	POSITIVE	15	12	n/a	1
NOT PERFORMED	NOT PERFORMED	35	31	1	2
Value not set	Value not set	2	1	n/a	1
Value not set	Value not set	1	1	n/a	n/a
Value not set	Value not set	1392	1220	13	14
Value not set	Value not set	5	3	n/a	n/a
Value not set	Value not set	3758	74	118	1681

# Individual Subject Information Viewer (ISIV)

Events More...

Gall Model: Lifetime patient average risk More: More...

Occupation: Occupation  Gall Model: 5-year risk ratio More...

Body Mass Index (BMI)  Microarray Experiment: Sample ID More...

Blood Sample: Preservation Type  Mammogram: Result More...

Hormonal treatment for bcr  Gall Model: Lifetime patient average risk (to age 90) More...

Deodorant or antiperspirant usage More...

Blood Sample: Preservation Type More...

Family Cancer Diagnosis: Menopausal status at diagnosis More...

Family Cancer Diagnosis: Cancer Type  Gall Model: 5-year patient risk More...

Medical Disorders: Other Medical Disorder  Medical Disorders: History of More...

Gall Model: 5-year risk ratio  Family Cancer Diagnosis: Car More...

Blood Sample: Preservation T More...

Events

Mammogram: Result More...

Total SLN: # Neg More...

Blood Sample: Location in the freezer More...

Tissue Sample: Location in the freezer More...

Tissue Sample: Pathologic Category More...

Invasive Breast Carcinoma: Final margins More...

Menopausal Status More...

Routinely wear an under-w More...

Breast Cancer Treatment: Surgery More...

Breast Cancer recur same breast More...

Medical Disorders: History of Other More...

Breast Cancer recur same breast More...

Medical Disorders: History of Other More...

Breast Cancer recur same breast More...

Events

Blood Sample: Location in the freezer More...

Invasive Breast Carcinoma: Multiple ipsilateral More...

ER: Result More...

Biopsies on right breast: Total number More...

Tissue Sample: Diagnosis Codes More...

Tissue Sample: Type More...

# Biomedical Informatics

---

- Data collection and generation
  - Laboratory Information Management System (LIMS)
  - Quality Assurance (QA) programs
- Data centralization
  - Data Warehouse for Translational Research (DW4TR)
  - Data models
  - Interfaces
- Applications
  - Data analysis and mining using the DW4TR
  - A Komen Promise Grant
  - The Cancer Genome Atlas Project—Breast Cancer

# Use of the DW4TR for Research

Fibroadenomatoid changes are more prevalent in middle-aged women and have a positive association with invasive breast cancer.

Chen Y, Bekhash A, Kovatich AJ, Hooke JA, Kvecher L, Mitchell EP, Rui H, Mural RJ, Shriver CD, and Hu H: *The 35<sup>th</sup> San Antonio Breast Cancer Symposium*. December 4-8, 2012. San Antonio, TX. (MS in preparation)

**Subtype-specific co-occurrence of atypical hyperplasia and in situ carcinoma with invasive breast cancers.**

Kovatich AJ, Kvecher L, Chen Y, Bekhash A, Hooke JA, Shriver CD, Mural RJ, and Hu H: *The 34<sup>th</sup> San Antonio Breast Cancer Symposium*. December 6-10, 2011. San Antonio, TX.

**Ethnicity Difference of Benign Breast Diseases in Breast Cancer and Non-Cancer Patients.**

Bekhash A, Saini J, Li X, Rapuri P, Hooke JA, Kovatich AJ, Mural RJ, Shriver CD, and Hu H: *The 33<sup>rd</sup> San Antonio Breast Cancer Symposium*. December 8-12, 2010. San Antonio, TX. (MS in preparation)

# Use of the DW4TR for Research

**Fibroadenomatoid changes are more prevalent in middle-aged women and have a positive association with invasive breast cancer.**

Chen Y, Bekhash A, Kovatich AJ, Hooke JA, Kvecher L, Mitchell EP, Rui H, Mural RJ, Shriver CD, and Hu H: *The 35<sup>th</sup> San Antonio Breast Cancer Symposium*. December 4-8, 2012. San Antonio, TX. (MS in preparation)

**Subtype-specific co-occurrence of atypical hyperplasia and in situ carcinoma with invasive breast cancers.**

Kovatich AJ, Kvecher L, Chen Y, Bekhash A, Hooke JA, Shriver CD, Mural RJ, and Hu H: *The 34<sup>th</sup> San Antonio Breast Cancer Symposium*. December 6-10, 2011. San Antonio, TX.

**Ethnicity Difference of Benign Breast Diseases in Breast Cancer and Non-Cancer Patients.**

Bekhash A, Saini J, Li X, Rapuri P, Hooke JA, Kovatich AJ, Mural RJ, Shriver CD, and Hu H: *The 33<sup>rd</sup> San Antonio Breast Cancer Symposium*. December 8-12, 2010. San Antonio, TX. (MS in preparation)

# Use of the DW4TR for Research

---

**Fibroadenomatoid changes are more prevalent in middle-aged women and have a positive association with invasive breast cancer.**

Chen Y, Bekhash A, Kovatich AJ, Hooke JA, Kvecher L, Mitchell EP, Rui H, Mural RJ, Shriver CD, and Hu H: *The 35<sup>th</sup> San Antonio Breast Cancer Symposium*. December 4-8, 2012. San Antonio, TX. (MS in preparation)

**Subtype-specific co-occurrence of atypical hyperplasia and in situ carcinoma with invasive breast cancers.**

Kovatich AJ, Kvecher L, Chen Y, Bekhash A, Hooke JA, Shriver CD, Mural RJ, and Hu H: *The 34<sup>th</sup> San Antonio Breast Cancer Symposium*. December 6-10, 2011. San Antonio, TX.

**Ethnicity Difference of Benign Breast Diseases in Breast Cancer and Non-Cancer Patients.**

Bekhash A, Saini J, Li X, Rapuri P, Hooke JA, Kovatich AJ, Mural RJ, Shriver CD, and Hu H: *The 33<sup>rd</sup> San Antonio Breast Cancer Symposium*. December 8-12, 2010. San Antonio, TX. (MS in preparation)

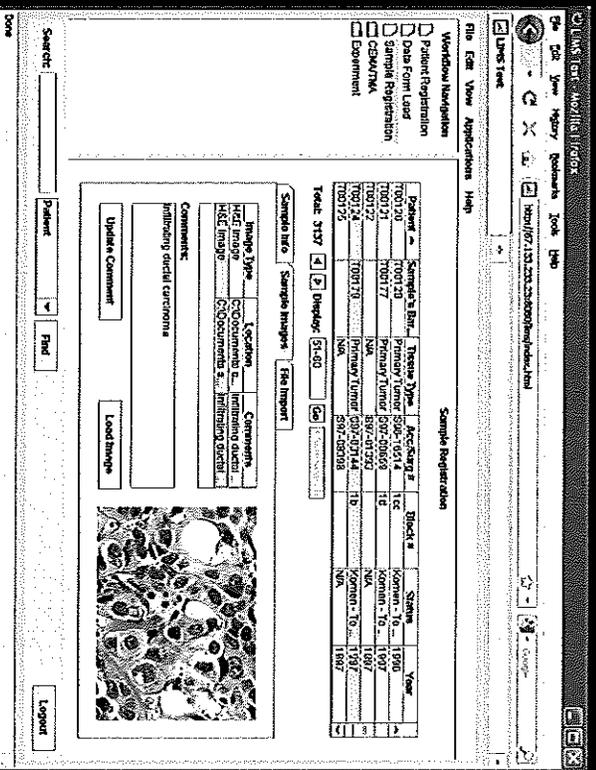
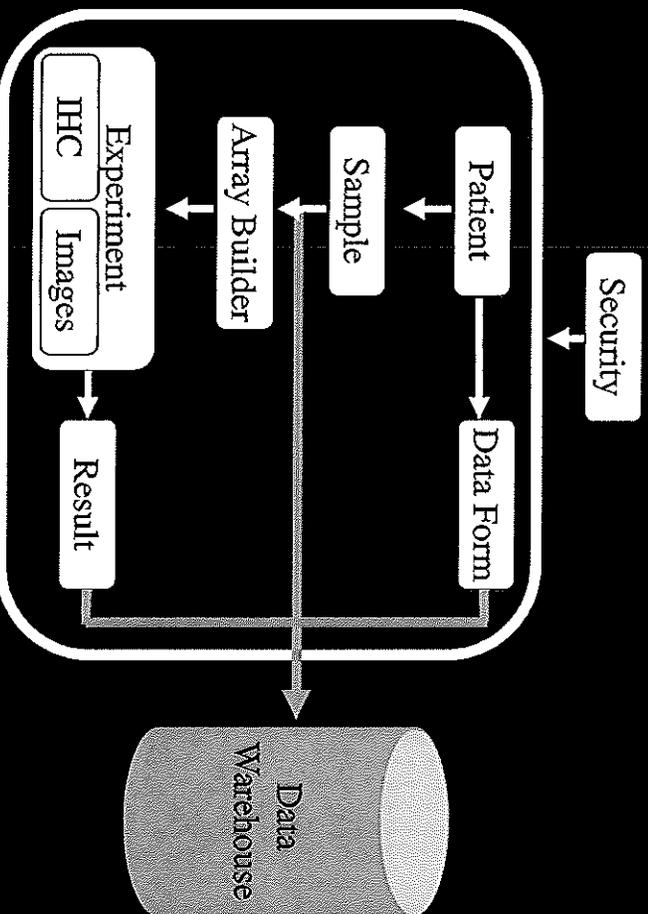
# A Komen Promise Grant

Consortium: 1) Thomas Jefferson University, 2) Windber Research Institute, 3) WRNIMC, 4) MDR Global, 5) DecisionQ.

Two aims:

1. Mapping 250 therapy-relevant proteins levels across 5,000 breast cancer tumors
2. Phase II trial of a PARP inhibitor on triple negative breast cancer

WRI BMX Responsibilities: 1) Infrastructure support, 2) Data QA and analysis



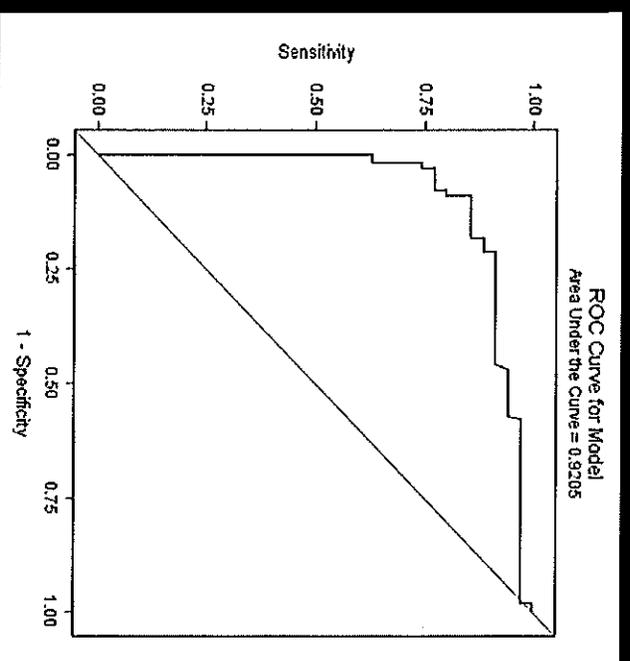
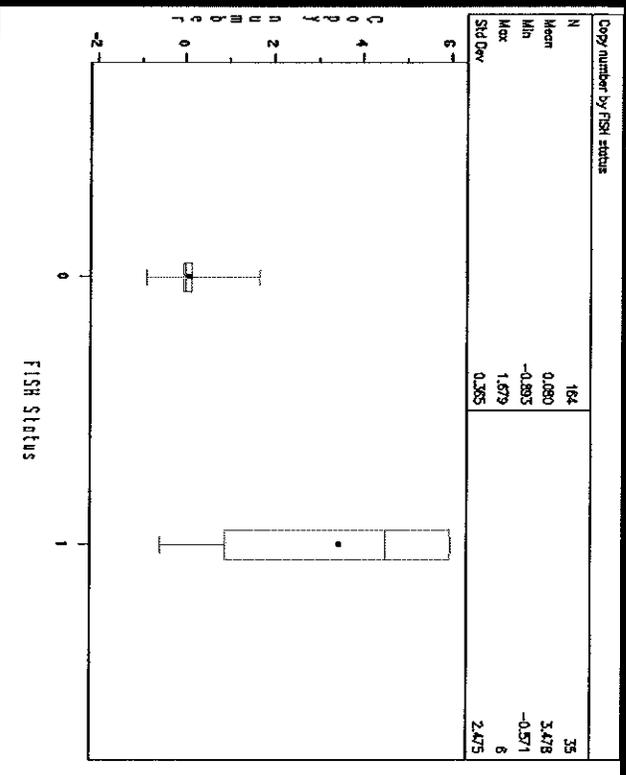
# The Cancer Genome Atlas—Breast Cancer

Supplying breast cancer tumors

- 106 cases
- Highest passing rate (89.6%)

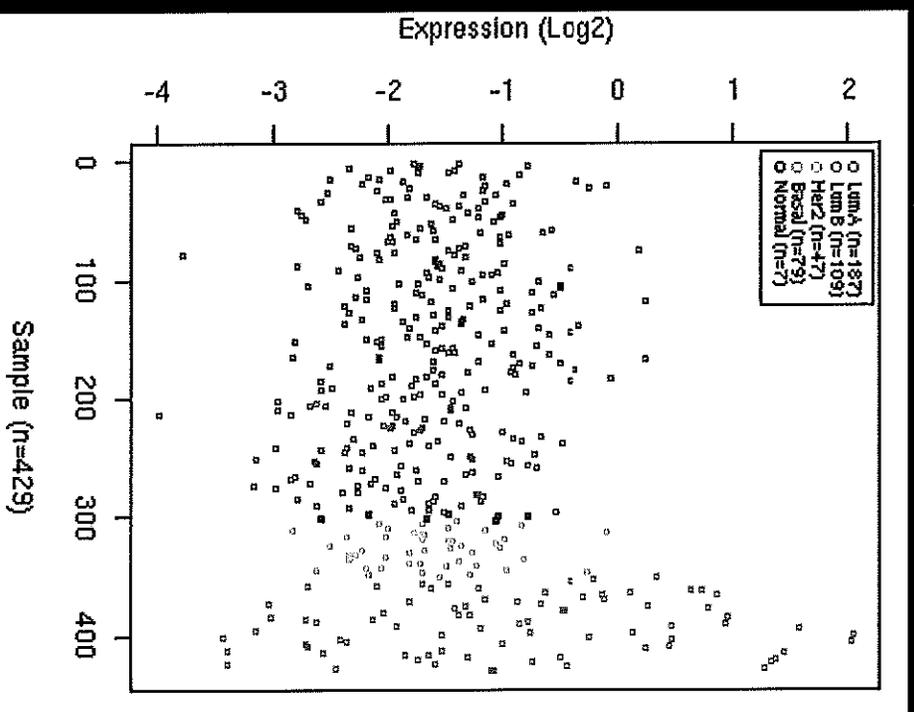
Clinical data QA

- Assessed critical data fields, developed methods to improve data quality
- Authored Supplemental Methods Section II



# Use of the TCGA—BC Data

CSPG4, known association with metastasis  
(Collaboration with Dr. Joji Iida of Cell Biology)



- Subtype analysis  
Basal higher expression,  $p=0.0014$
- Basal subtype
- Differential gene expression analysis:  
MIA, SOX10 etc.
- New lab cell line experiments

- TNM quartile analysis:  
Only N significant with  $p=0.04$

CSPG4 Quartile	1	2	3	4
LN+ rate	45.80%	54.20%	32.00%	16.70%

New questions: 1) Is CSPG4 positively associated with metastasis? 2) If true, are these metastases not via the lymphatic system?

Iida J et al. PLOS ONE 2012 (in press)

Luo C et al. SABCS 2012 (accepted); Dorchak J et al. SABCS 2012 (accepted)

# Summary

---

We have developed a biomedical informatics infrastructure composed of

- A data tracking system (LIMS)
- A set of QA programs
- A Data Warehouse for Translational Research (DW4TR)

To support translational research programs of

- the Clinical Breast Care Project (CBCP)
- the Gynecological Disease Program (GDP)

Which enabled data analysis and data mining of the clinicopathologic and molecular data.

Our infrastructure development and research capabilities also enabled us to play important roles in new translational research projects.

# Acknowledgments

---

## WRI

Leonid Kvecher  
Anthony Bekash  
Vincent Wu  
Hua Wang  
Chungqing Luo  
Yaqin Chen  
Joni Kohr  
Apryl Shaffer

Richard J. Mural  
Stella Somiari  
Caroline Larson  
Joji (George) Iida

**Previous members**  
Susan Maskery  
Henry Brzeski  
Michael N. Liebman

**Clinical and research collaborators**  
Craig D. Shriver, Jeffrey A. Hooke  
And the rest of the CBCP team

Larry G. Maxwell, Kathleen Darcy & GDP team  
Hallgeir Rui, Thomas Jefferson University  
Albert J. Kovatich, MDR Global

**InforSense/IDBS**  
Mick Correll, Jonathan Sheldon & the DW team  
Robin Munro and others

**Cimarron Software**  
Rob Sargent, Debi Koltenuk, and others

**Funding**  
The CBCP and the GDP were funded by the US Department of Defense through Henry Jackson Foundation for the Advancement of Military Medicine. The views expressed in this presentation are those of the authors and do not reflect the official policy of the Department of Defense, or U.S. Government.  
Susan Komen for the Cure

**All human subjects enrolled in the studies**

---

*Thank you all!*

[h.hu@wriwindber.org](mailto:h.hu@wriwindber.org)

## **FH535 inhibited migration and growth of breast cancer cells**

**Jesse Dorchak<sup>1</sup>, Joji Iida<sup>1</sup>, Rebecca Clancy<sup>1</sup>, Chunqing Luo<sup>1</sup>, Yaqin Chen<sup>1</sup>, Hai Hu<sup>1</sup>, Richard J. Mural<sup>1</sup>, and Craig D. Shriver<sup>2</sup>**

**1** Windber Research Institute, Windber, PA 15963, **2** Department of Surgery, Walter Reed National Military Medical Center, Bethesda, MD 20889

**Background:** Given the lack of effective targeted therapies for triple-negative (TN) breast cancer patients, a better understanding of the mechanisms of growth and invasion of these tumors will provide valuable insight into developing novel approaches to lower the mortality associated with TN breast cancer. There is substantial evidence indicating that the canonical Wnt signaling pathway is activated in TN breast cancer tissues. Previous studies report that FH535, a small molecule inhibitor of the Wnt signaling pathway, decreases growth of cancer cells but not normal fibroblast cells, suggesting this pathway plays a role in tumor progression and metastasis. In this study, we test FH535 as a potential inhibitor of migration, invasion, and growth against malignant phenotypes of breast cancer.

**Methods and Results:** We show that FH535 significantly inhibits growth, migration, and invasion of TN breast cancer cells (MDA-MB-231 and HCC38) using established protocols, implicating the Wnt signaling pathway plays a role in tumor progression and metastasis. FH535 is a potent growth inhibitor for TN breast cancer cells (MDA-MB-231 and HCC38), but not in other breast cancer cells (MCF-7, T-47D, or SK-BR-3) when cultured in three dimensional type I collagen gel. Additionally, western blotting analysis shows that treatment of MDA-MB-231 cells with FH535 markedly inhibits the expression of NEDD9, but not FAK, Src, or p130Cas, suggesting specificity of the Wnt signaling pathway in regulating gene expressions. Co-immunoprecipitation studies suggest that NEDD9 is specifically immunoprecipitated with CSPG4, but not with  $\beta$ 1 integrin or CD44, implying CSPG4 and NEDD9 form a molecular complex for facilitating tumor migration, invasion, and growth. Statistical analysis (ANOVA) of The Cancer Genome Atlas (TCGA) project gene expression data (<https://tcga-data.nci.nih.gov/>) suggests that tumors from patients with basal-type breast cancer had significantly higher CSPG4 expression than tumors from patients with other subtypes, with all the  $p$  values  $<0.0001$ . Moreover, Luminal A tumors also had a significantly higher CSPG4 expression than Luminal B tumors ( $p=0.0006$ ).

**Discussion:** In this study, we demonstrate the pivotal role of the canonical Wnt signaling pathway in enhancing tumor migration, invasion, and growth of TN breast cancer cells MDA-MB-231 and HCC38. We identified NEDD9 as a potential downstream target of the Wnt signaling pathway and found an association of NEDD9 with CSPG4 in MDA-MB-231 cells. Our results suggest a novel mechanism of migration, invasion, and growth of tumor cells involving a molecular complex between CSPG4 and NEDD9, which could be drug targets in TN and possibly Luminal A breast cancer patients. More comprehensive characterization of canonical Wnt signaling pathway by genomic and biological studies of breast cancer cells are clearly required for generating better treatments. The views expressed in this article are those of the author and do not reflect the official policy of the Department of Defense, or U.S. Government.

**Role for wnt signaling pathways in promoting migration and growth of triple negative breast cancer cells.**

**Joji Iida<sup>1a</sup>, Stella Somiari<sup>1</sup>, John Lehman<sup>1</sup>, Rebecca Clancy<sup>1</sup>, Gerianne Aberts<sup>1</sup>, Amy O'donnell<sup>1</sup>, Wannian Yang<sup>2</sup>, Richard J. Mural<sup>1</sup>, and Craig D. Shriver<sup>2</sup>**

- 1. Windber Research Institute, Windber, PA 15963**
- 2. Center for Research, Geisinger Clinic, Danville, PA 17822**
- 3. Department of Surgery, Walter-Reed Army Medical Center, Washington, DC20902**

- Running title: Wnt signaling in migration and growth of breast cancer cells**
- Keywords: wnt, integrin, MMP, metastasis, breast**
- This research was supported by a grant from the United States Department of Defense (Military Molecular Medicine Initiative MDA W81XWHH-05-2-0075, Protocol 01-20006)**
- Corresponding Author: Joji Iida, Windber Research Institute, 620 7<sup>th</sup> Street, Windber, PA15963**

**Tel: 814-361-6985, Fax: 814-467-6334, e-mail: [g.iida@wriwinber.org](mailto:g.iida@wriwinber.org)**

- There is no conflict of interests.**
- Word count = 4979. 5 figures and 1 table.**

## Abstract

Enhanced migration and growth are the hallmark of malignant tumor phenotypes, thus regulating these processes would be a promising approach to inhibiting tumor invasion and metastasis. The wnt signaling pathway is important in normal development and in tumor progression. Previous studies reported that FH535 and XAV939, which are synthetic inhibitors for wnt signaling pathway, inhibited cancer cell but not normal fibroblast cell growth, implicating their efficacy for inhibiting malignant tumor cell phenotypes. The purpose of this study is to elucidate wnt signaling pathway in promoting tumor migration and growth in breast cancer cells. We demonstrated that both FH535 and XAV939 significantly inhibited  $\alpha2\beta1/\alpha3\beta1$  integrin-mediated cell migration of HCC38 (triple negative breast cancer cells). Further studies showed that treatment of cells with FH535 and XAV939 inhibited tyrosine phosphorylation of paxillin at Tyr<sup>118</sup>, but not of focal adhesion kinase (FAK) at Tyr<sup>397</sup>, suggesting that wnt signaling pathway regulates integrin-mediated tyrosine kinase pathways. Interestingly, FH535 but not XAV939 significantly inhibited integrin-independent three dimensional (3D) tumor growth of HCC38 cells. 3D growth was also inhibited in the presence of GM6001, which is a selective inhibitor for matrix metalloproteinases (MMPs). Reverse transcriptase (RT)-PCR analysis showed that MMP-2 and MMP-9 mRNAs are downstream target of FH535 treatment. These results demonstrated the key role for wnt signaling pathway in regulating multiple malignant phenotypes including migration, growth, and gelatinase productions, thus suggesting that targeting the wnt signaling pathway may be a promising approach for inhibiting invasion, growth, and metastasis of triple-negative breast cancer cells.

## **Introduction**

Breast cancer is one of the leading causes of death in women today in Western countries. Although the mortality rate of this cancer has been decreasing as a result of the development of early detection methods such as mammography as well as adjuvant therapy with tamoxifen and chemo-therapeutic drugs, metastatic breast cancer continues to be the most fatal condition for women. Triple negative (TN) breast cancers are defined by a lack of expression of estrogen, progesterone, and her-2/neu receptors. It is widely recognized that TN breast cancers have a poorer prognosis than any other subtypes of breast cancer. Given the lack of effective targeted therapies for TN breast cancer patients, a better understanding of the mechanisms of growth and invasion will provide valuable insight into developing novel approaches to lower the mortality associated with TN breast cancer.

Integrins are a family of cell adhesion receptors expressed on the cell surface that mediate cellular interaction with extracellular matrix (ECM) components (1). There is substantial evidence suggesting a pivotal role for integrin-mediated cell adhesion and subsequent signaling pathways including tyrosine kinases for facilitating tumor migration, invasion and growth in the tissue architectures (2). Thus, further characterizing mechanisms of integrin-mediated cell adhesion and signaling pathways would provide insight for not only understanding the processes of tumor progression but also generating therapeutic strategies for TN cancer patients.

The wnt signaling pathway is important in normal development and in tumor progression (3). This signaling pathway is regulated by Wnt ligands, the adenomatous polyposis coli (APC)-axin complex and  $\beta$ -catenin among others (4). While  $\beta$ -catenin is constitutively phosphorylated and degraded in the

absence of activation of wnt signaling, stabilization of  $\beta$ -catenin by stimulating the wnt signaling pathway leads to its accumulation and subsequent localization to the nucleus, where it forms complexes with transcription factors of the Tcf/Lef families (5). Although recent studies demonstrated that activation of wnt/ $\beta$ -catenin pathway was preferentially found in TN phenotype and is associated with poor clinical outcome (6), role for wnt/ $\beta$ -catenin pathway in regulating migration and growth of TN breast cancer cells are not clear present.

Recently, selective inhibitors for wnt signaling pathway were developed and characterized (7, 8). FH535 is a synthetic inhibitor for TCF/ $\beta$ -catenin pathway probably through its inhibition of the complex formation containing  $\beta$ -catenin and Tcf/Lef in the nucleus, which leads to decreased gene transcription activity for genes such as Tcf-4 (7). The other wnt signaling pathway inhibitor is XAV939, which selectively inhibits  $\beta$ -catenin-mediated transcription by stimulating  $\beta$ -catenin degradation through the stabilization of axin protein in cytosol (8). Importantly, these two synthetic inhibitors were demonstrated to inhibit growth of colon cancer cells (7, 8), suggesting that the wnt-signaling pathway would play a key role for facilitating tumor growth and metastasis.

In this study, we evaluated the wnt signaling pathway in promoting migration and growth of TN breast cancer cells. By utilizing FH535 and XAV939, we demonstrated that wnt signaling pathway is important for facilitating  $\alpha 2\beta 1/\alpha 3\beta 1$  integrins-mediated TN breast cancer migration. Our results showed that FH535 but not XAV939 inhibited tumor growth and production of gelatinases (i.e. MMP-2 and MMP-9), thus suggesting that specific molecules sensitive to FH535 play a key role in promoting these processes. These results further suggest that the wnt-signaling pathway play a role in promoting TN breast cancer

progression and therefore, this pathway may be a promising therapeutic target for this phenotype of breast cancer as well as other breast cancer phenotypes.

## **Materials and Methods**

**Cell lines:** HCC38, HCC1937, HCC1806, MCF-7, MDA-MB-231, HCC 1419, HCC2218, SK-Br3, and HT1080 were purchased from ATCC (Manassas, VA, USA) and maintained in RPMI1640 containing 10% FBS.

**Antibodies and chemicals:** Poly-HEMA, FH535, XAV939, and MOPC-21 (IgG<sub>1</sub>) were purchased from Sigma-Aldrich (St. Louis, MO, USA). Anti- $\beta$ 1 (clone P4C10 IgG<sub>1</sub>), anti- $\alpha$ 2 (clone P1E6 IgG<sub>1</sub>), anti- $\alpha$ 3 (clone P1B5 IgG<sub>1</sub>) integrin antibodies were purchased from Millipore (Billerica, MA, USA). Anti-phosphopaxillin (Tyr<sup>118</sup>), anti-paxillin, anti-phosphotyrosine, and anti-actin antibodies were purchased from Cell Signaling Technology (Beverly, MA, USA). HRP-conjugated secondary antibodies were purchased from Jackson ImmunoResearch Laboratories (West Grove, PA, USA). GM-6001 and WST-1 were purchased from Calbiochem (Gibbstown, NJ, USA). Type I collagen solution (PureCol) was purchased from Inamed Biomaterial (Fremont, CA, USA).

**Migration assays:** Tumor cell migration was characterized using Transwell systems (Costar, 8.0 $\mu$ m pore size) as described previously (9). Briefly, cells were harvested with 5 mM EDTA, washed, and resuspended in serum free RPMI1640 media at a concentration of  $5 \times 10^5$  cells/ml. An aliquot of 100  $\mu$ l of the cell suspension was added into the upper chamber of Transwell system. Type I collagen was dissolved in serum Free RPMI1640 media at a concentration of 3  $\mu$ g/ml and added into the lower chamber of the system. Antibodies (5 $\mu$ g/ml), FH535 (0.01 to 1  $\mu$ M), or XAV939 (0.001 to 0.1  $\mu$ M) was added in both upper and lower chambers. Migration assays were performed at 37°C for 4-5 hours. The inserts were removed and then fixed in 10% formalin in PBS for 5 min and stained with 0.5% crystal violet for 5 min. The inserts were washed and the upper surface of the membranes was wiped with a cotton swab to remove

non-migratory cells. Migrated cells were counted at randomly selected three fields. Each experiment was performed by triplicate. The data was expressed as mean cell numbers +/- S.D. per mm<sup>2</sup>.

Growth Assays: 96-well plates were coated with 100 µg/ml poly-HEMA for inhibiting attachment of cells as described previously (9, 10). Cells were harvested with PBS containing 5 mM EDTA and washed three times with RPMI1640-10%FBS. Cells were then resuspended in the same media at a concentration of 10<sup>5</sup> cells/ml. An aliquot of 100 µl was added into each well and then mixed with 100 µl of RPMI-10%FBS containing FH535, XAV939, or GM6001 at concentrations described in the text. The plates were incubated for overnight and then cell growth was measured using colorimetric assay system with WST-1. Each experiment was performed quadruplets and the data were shown as mean +/- S.D. of Absorbance at 450 nm.

Western blotting: Cells were cultured in 6 well plates (7 x 10<sup>5</sup> cells/well/1ml) coated with 3 µg/ml of type I collagen for 4 hours and then lysed in SDS-sample buffer. FH535 and XAV939 were used as inhibitors at concentrations of 1 µM and 0.1 µM, respectively. After fragmenting DNA by a sonicator, proteins were separated on SDS-PAGE and transferred onto Immobilon-P membrane as we described previously (9). The membranes were blotted with primary antibodies followed by HRP-conjugated secondary antibodies. The proteins were detected with enhanced chemiluminescence (ECL) reactions.

Gelatin zymography: The visualizations of MMP-2 and MMP-9 in the serum free conditioned media of HCC38 cells were performed using gelatin zymography as we previously described (9). Briefly, cells (7 x 10<sup>5</sup> cells per well) were cultured in the growth media in 6 well plates overnight. Cells were washed with warmed serum free RPMI1640 five times and then incubated in the presence or absence of FH535 (20

$\mu\text{M}$ ) or XAV939 (2  $\mu\text{M}$ ) in serum free RPMI 1640 overnight. Conditioned media (10  $\mu\text{l}$ ) were mixed with SDS sample buffer (6% SDS, 0.16 M Tris, 30% glycerol, and 0.2% bromophenol blue) and then separated on SDS-PAGE containing 1 mg/ml gelatin. The gels were washed two times with 2.5% Triton-X100 and then incubated in buffer containing 20 mM Tris-HCl (pH 7.5), 1.25% Triton-X100, 5 mM  $\text{CaCl}_2$ , and 1 mM  $\text{ZnCl}_2$  overnight at 37°C. The gels were stained with Coomassie blue and destained. As a standard, conditioned media prepared from ConA-stimulated HT1080 cells was used to localize pro- (p), intermediate- (i), and active- (a) forms of MMP-2 as well as pro-MMP-9 (9).

Cells were briefly washed with PBS and total RNA was extracted as described below.

Luminex assays: The amount of MMP-1, MMP-2, and MMP-9 secreted in culture media was measured using Fluorokine MAP, Human MMP MultiAnalyte Profiling Base Kit and Bead Sets (R&D Systems; Minneapolis, MN) in 96-well formats according to the manufacturer's protocol. Briefly, serum free conditioned media prepared as described in the previous section were diluted 5-fold. 50  $\mu\text{l}$  of microparticle bead mixtures for MMP-1, MMP-2, and MMP-9 were plated into the 96-well micro plates. The plates were then incubated for 2 hours at room temperature. The plates were washed three times and then 50  $\mu\text{l}$  of Biotin Antibody Cocktail was added to each well followed by incubation at room temperature for 1 hour. Then, 50  $\mu\text{l}$  of diluted Streptavidin-phycoerythrin conjugate (PE) was added to each well and incubated for additional 30 minutes. After extensive washing, fluorescence intensity of MMP-1, MMP-2, and MMP-9 was detected with the Luminex 100 system. Experiments were repeated three times and data were shown as pg/ml by calculating based on the standard curves of fluorescence intensity of MMP-1, MMP-2, and MMP-9.

RT-PCR analysis: Total RNA was prepared from cells treated with FH535 or XAV939 as described in the section of gelatin/zymography using RNA extraction kit (Qiagen, Maryland, USA). The same amount of RNA (1µg) was transcribed into DNA using random hexamer primers according to the manufactures• protocols (Invitrogen, Carlsbad, CA, USA)). The sequences of the primers were as follows.

MMP-2: forward 5•-CAA TAC CTG AAC ACC TT-3• and reverse 5•- CTG TAT GTG ATC TGG TT-5•

MMP-9: forward 5•- AGA TTC CAA ACC TTT GAG-3• and reverse 5•- GGC CTT GGA AGA TGA

ATG-3•. TCF-4: forward 5•- TTC AAA GAC GAC GGC GAA CAG-3• and reverse 5•- TTG CTG

TAC GTG ATA AGA GGC G-3•. GAPDH: forward 5•- TGA TGA CAT CAA GAA GGT GGT GAA G-

3• and 5•- TCC TTG GAG GCC ATG TGG GCC A-3•.PCR reactions were performed by heating at 94°C

(2min) followed by 30 cycles of 94°C (30 sec), 55°C (30 sec), and 72°C (1min) followed by a final

extension at 72°C 7min). The PCR products were separated on 2% agarose gel and visualized using ethidium bromide.

Statistical analysis: Statistical significance was calculated by Student•s two-tailed t-test. *p* values less than 0.05 are considered as statistically significant.

## **Results:**

### **FH535 significantly inhibited $\alpha2\beta1/\alpha3\beta1$ integrins-mediated HCC38 triple negative (TN) breast cancer cells**

Enhanced migration has been recognized as one of the hallmarks of malignant tumor cell phenotypes. Migration of breast cancer cells into surrounding connective tissue architecture is the important process for establishing metastasis. Type I collagen is a major component of connective tissues and migration toward this protein is implicated to be a key step for invading into tissues. Previous studies demonstrated that tumor migration to type I collagen is mediated by both or either of  $\alpha2\beta1$  and  $\alpha3\beta1$  integrin (11, 12). Since HCC38 cells express both  $\alpha2$  and  $\alpha3$  integrin subunits evaluated by FACS analysis (not shown), we tested inhibitory antibodies against these subunits for their abilities to inhibit HCC38 cell migration to type I collagen. Both anti- $\alpha2$  and  $\alpha3$  integrin subunits antibodies as well as anti- $\beta1$  integrin subunit antibody almost completely inhibited tumor migration (**Figure 1A**), while isotype matched control antibody (MOPC-21) did not affect migration of tumor cells (**Figure 1A**). These results indicate that  $\alpha2\beta1$  and  $\alpha3\beta1$  integrins mediated HCC38 migration to type I collagen.

Under these experimental conditions, we tested FH535, which is a specific inhibitor for wnt signaling pathways, for its ability to regulate HCC38 migration to type I collagen. In addition to the previous studies demonstrating that FH535 selectively inhibits growth of colon and lung cancer cells, our results indicated that this compound was a significant inhibitor for tumor cell migration even at 0.1  $\mu\text{M}$  (**Figure 1B**). We also tested XAV939, which is an inhibitor for wnt signaling pathway by inhibiting tankyrases (8), for its ability to inhibit cell migration. XAV939 significantly inhibited HCC38 cell migration to type I collagen in a concentration-dependent manner (**Figure 1B**). Under these experimental conditions, treatment of

cells with FH535 or XAV939 did not cause cell death as evaluated by trypan blue exclusion assays (not shown). These results suggest that wnt signaling pathways regulate  $\alpha2\beta1/\alpha3\beta1$  integrins-mediated tumor migration.

Previous studies demonstrated that tyrosine kinase pathways stimulated through ligation of integrins to their specific ligands plays a key role in facilitating cell migration (13, 14). In order to test whether wnt signaling pathways were involved in tyrosine kinase pathways in HCC38 cells, cells were treated with FH535 (1  $\mu\text{M}$ ) or XAV939 (0.1  $\mu\text{M}$ ) on type I collagen-coated substrate for 4 hours. Total cell lysates were separated on SDS-PAGE and transferred on Immobilon-P membranes. Blotting with anti-phosphotyrosine (PY) antibody suggest that phosphorylation of at least two proteins (see arrows in the Figure 2A) was inhibited in the presence of FH535 or XAV939 compared to control cells (Figure 2A). Actin was used as a loading control (Figure 2A). One of the prominent proteins migrated at 68 kDa (Figure 2A). Previous studies suggest that tyrosine phosphorylation of Tyr<sup>118</sup> of paxillin is implicated to play a role in tumor migration (15). Blotting with anti-phosphopaxillin (Tyr<sup>118</sup>) antibody suggests that the phosphorylation was inhibited by treatment with both FH535 and XAV939 (Figure 2B). However, the amounts of paxillin were not affected by these treatments (Figure 2B). These results suggest that FH535 and XAV939 inhibited  $\alpha2\beta1/\alpha3\beta1$ -integrin mediated migration to type I collagen through regulating tyrosine kinase pathways involving phosphorylation of Tyr<sup>118</sup> of paxillin. These synthetic compounds did not inhibit cell adhesion to type I collagen (not shown), suggesting that initial attachment of cells mediated by  $\alpha2\beta1/\alpha3\beta1$ -integrin would not be targets of FH535 and XAV939. Indeed, treatment of cells with either FH535 or XAV939 did not alter phosphorylation of Tyr<sup>397</sup> of focal adhesion kinase (FAK) under these experimental conditions (Figure 2C). We are currently evaluating tyrosine kinases that phosphorylate paxillin and characterizing mechanisms of regulation of their kinases activities by wnt signaling.

#### FH535 inhibited integrin-independent growth of HCC38 cells

FH535 and XAV939 have been demonstrated to inhibit growth of various tumor cells including colon, lung, and hepatocellular carcinomas (7, 8). We next tested the ability of these synthetic compounds for inhibiting HCC38 cells to grow in three dimensional (3D) culture as described in Materials and Methods, since increased growth is associated with tumor progression and invasion (16, 17). While FH535 partially but significantly inhibited tumor growth, XAV939 failed to do so (**Figure 3A**), implicating overlapping but distinct roles of wnt signaling pathways for regulating migration and growth. XAV939 did not inhibit tumor growth even at a concentration of 10  $\mu$ M (not shown). FH535 inhibited growth of HCC38 in a concentration-dependent manner and significant inhibition was observed at 20 and 2  $\mu$ M under our experimental conditions (**Figure 3B**), consistent with the previous studies (7). Furthermore, our data showed that a synthetic MMP inhibitor, GM6001, also partially but significantly inhibited tumor growth (**Figure 3B**). The combination of treatment of HCC38 cells with FH535 did not enhance the growth inhibitory activity by GM6001 (not shown), suggesting catalytic activity of MMPs would be involved in 3D tumor growth which is mediated by wnt signaling pathways.

#### FH535 inhibited MMP-2 and MMP-9 production from HCC38 cells.

We, then, hypothesized that the wnt signaling pathway facilitated tumor growth by regulating expression of MMPs. There is substantial evidence demonstrating MMP-1, MMP-2, and MMP-9 plays a key role in various tumor invasion and metastasis such as breast cancer (18). We first evaluated production of these MMPs in serum-free conditioned media by Luminex as described in Materials and Methods. Although TNBC cells (HCC1937, HCC1806, HCC38, and MDA-MB-231) produced at least one MMP in the conditioned media, non-TNBC cells (SK-Br3, HCC1419, and HCC2218) did not produce MMP-1, MMP-

2, or MMP-9 under our experimental conditions (**Table 1**). In order to ask the question whether wnt signaling pathway was involved in MMP-2 and MMP-9 production from HCC38 cells, cells were cultured in the presence of FH535, XAV939 in serum-free media overnight and the conditioned media were analyzed for MMP-2 and MMP-9 productions by both gelatin zymography (**Figure 4A**) and a multiple assay system (Luminex, xMAP technology) (**Figures 4B, 4C, and 4D**). When cells were treated with FH535, the production of both MMP-2 and MMP-9 was almost completely suppressed evaluated on gelatin-zymography, while XAV939 did not affect these gelatinase productions (**Figure 4A**). These results were further confirmed by the Luminex assay systems. FH535 treatment significantly inhibited the production of MMP-2 (**Figure 4B**) and MMP-9 (**Figure 4C**). MMP-1 production was not observed in control cells or inhibitor treated cells (**Figure 4D**). Importantly, XAV939 did not inhibit productions of MMP-2 and MMP-9 from HCC38 cells (**Figures 4A, 4B, 4C**), indicating specific signaling pathways sensitive to treatment with FH535 would play a key role in the production of these gelatinases.

We then tested whether MMP-2 and MMP-9 were downstream targets of FH535 treatment by reverse transcriptase (RT)-PCR analysis. When cells were treated with FH535 at a concentration of 10 $\mu$ M, the expression of mRNA for MMP-2 and MMP-9 were greatly inhibited (**Figure 5**), suggesting that FH535 regulates the production of MMP-2 and MMP-9 by inhibiting transcription of these genes. It is, however, the inhibition of MMP-2mRNA by FH535 was partial (**Figure 5**), while the MMP-2 protein production was almost completely inhibited (**Figure 4B**). These results may implicate additional functions of FH535 for regulating intracellular trafficking of MMP-2 in breast cancer cells. Previous studies showed that treatment of colon cancer cells (HCT116) with FH535 cells inhibited the transcription of TCF-4 (7), and this notion was also partially supported in HCC38 cells by demonstrating a decreased expression of TCF-4 in cells treated with FH535 (**Figure 5**).

## Discussion

There is increasing evidence demonstrating that wnt signaling pathways play key roles not only in the physiological but also in the disease processes (19). In the normal breast, expressions of wnt-signal components are tightly regulated, dictating its role in balancing stem cell-renewal, maintenance, and differentiation during embryonic and postnatal development (19). Thus, it is not surprising that dysregulation of the expressions and biological activities of wnt- signaling components would be one of the key steps in the progression of breast cancer. Recent studies suggest the aberrant activation of wnt/b-catenin pathway in triple-negative (TN) breast cancer phenotype (6). In this study, we demonstrated that wnt signaling is a key pathway for facilitating both integrin-mediated migration and -independent growth and gelatinase productions. Our results not only suggest the importance of wnt-signaling in breast cancer progression but also provide additional insights of targeting this signaling pathway for the therapy of breast cancer including TN phenotype.

In this study, we characterized mechanisms of migration and growth of HCC38 cells as a model system of TN phenotype breast cancer cells. FACS analyses suggest that all TN cell lines used in this study (HCC38, HCC1806, HCC1937, MDA-MB-231) were CD44<sup>+</sup>CD24<sup>-low</sup>ESA<sup>+</sup> phenotype (Iida et al., manuscript in preparation), which is originally identified as a phenotype of breast cancer initiating cells (20, 21). Using these systems, we demonstrated that both FH535 and XAV939 significantly inhibited  $\alpha2\beta1/\alpha3\beta1$ -integrin mediated tumor migration, suggesting that these would regulate overlapping pathway(s) that promote  $\alpha2\beta1/\alpha3\beta1$  integrin-mediated signaling pathways important for cell migration. Indeed, our results demonstrated that both FH535 and XAV939 inhibited phosphorylation of Tyr<sup>118</sup> of paxillin without affecting phosphorylation of Tyr<sup>397</sup> of FAK. These results suggest that wnt pathway would regulate downstream signaling of integrin ligation for facilitating tumor migration. Previous studies suggest that

phosphorylations of Tyr<sup>38</sup> and Tyr<sup>118</sup> of paxillin occur upon cell adhesion to extracellular matrix (ECM) (22), and FAK (23), Src kinase families (24), and c-Abl (25) mediate these phosphorylations. We are further characterizing mechanisms of wnt signaling-mediated tumor migration by identifying tyrosine kinases(s) regulated by FH535 and XAV939 in TN breast cancer cells.

The processes of three dimensional (3D) tumor growth have been characterized as anchorage-independent growth and one of the malignant phenotypes of cancer cells (16, 17). Handell et al. reported that FH535 inhibited tumor growth but not normal fibroblasts in a conventional two dimensional culture systems (7). Consistent with this notion, our results also support that FH535 inhibited growth of TN breast cancer cells in 3D conditions, suggesting that the inhibition of wnt signaling pathway would be a beneficial approach for breast cancer therapy. Although previous studies demonstrated that XAV939 inhibited tumor growth (8), we could not observe growth inhibitory activity under our experimental conditions. Although the reason of this discrepancy is not clear at present, it is possible that XAV939 may affect tumor growth in a cell type-dependent manner. We demonstrated that a synthetic inhibitor, GM6001, significantly inhibited tumor growth under our experimental conditions. Although the details of the mechanisms involving the catalytic activities of MMPs in facilitating tumor growth is unclear at present, it is possible that growth factors which could be activated by MMPs may play a role in the growth systems. Previous studies suggest that MMPs can directly regulate growth factor activities. For example, MMP-2, MMP-9, and MMP-14 have been demonstrated to proteolytically cleave pro-form of TGF- $\beta$  to convert to its active form (26, 27). MMP-9 and MMP-11 could release IGF proteins from their growth-inhibitory binding proteins (IGFBPs), which in turn stimulates tumor growth (28, 29). Moreover, MMPs could regulate tumor cell proliferation by altering pericellular matrix assembly which releases ECM-bound growth factors stimulates as well as integrin-mediated growth (30-32). Thus, it is of importance to identify growth

factors that promote 3D tumor growth for further understanding the mechanism of tumor growth at the microenvironments.

Production and activation of MMP-2 and MMP-9 have been characterized as key processes in both physiological and pathological processes (33). In breast cancer progression, there is substantial evidence demonstrating MMP-1, MMP-2 and MMP-9 promotes tumor invasion and metastasis (34-37). It is of interest that all TN cell lines used in this study produced MMP-1, MMP-2, or MMP-9, while the non-TN cell lines failed to produce these matrix degrading enzymes. The productions of MMP-1, MMP-2, and MMP-9 from either TN or non-TN breast cancer cells was not altered with soluble growth factors including TGF- $\beta$ , EGF or IGF-1 (not shown), which have been implicated to play a role in promoting tumor invasion (38-40). In TN breast cancer cells, our results using FH535 provide an insight for controlling the production of MMP-2 and MMP-9. On the other hand, XAV939 did not affect the production of these gelatinases, implicating specific pathways sensitive to FH535 but not XAV939 would play a key role in the processes of the productions. Recent studies showed that the inhibition of  $\beta$ -catenin by specific siRNA inhibited MMP-2 and MMP-9 productions in colon cancer cells, inflamed tissue, and mesenchymal stem cells (41-43), however, the role for wnt signaling pathway in producing these gelatinase from breast cancer cells is not clear. In this regard, our results provide evidence of a pivotal role of TCF/ $\beta$ -catenin pathway in the production of MMP-2 and MMP-9 from TN breast cancer cells. Previous studies suggest that FH535 antagonizes both peroxisome-proliferator activator receptor (PPAR)  $\gamma$  and  $\delta$ , while the structurally related compound, GW9662, antagonizes only PPAR $\gamma$  (7). We observed that GW9662 failed to inhibit productions of MMP-2 and MMP-9 from TN breast cancer cells evaluated by gelatin/zymography and Luminex studies (Iida et al., manuscript in preparation), implicating PPAR $\delta$  may play a key role in the productions of these gelatinases from TN breast cancer cells. Consistent with this

notion, previous studies implicate role for PPARs in productions of MMP-2 and MMP-9 (44, 45). In this regard, we are currently characterizing mechanisms of MMP-2 and MMP-9 production from TN breast cancer cells involving biological functions of PPAR $\delta$  in the process.

Taken together, these results suggest that inhibition of wnt signaling with FH535 inhibited migration, growth and gelatinase productions. Given the fact that the enhanced motility, growth, and invasion were key steps for establishing metastasis, our results not only support the notion of importance of wnt signaling pathways in regulating these processes but also provide a novel therapeutic strategies for patients diagnosed as a TN phenotype (46).

**Acknowledgment:**

The authors thank to Dr. Rachel Ellsworth for critical reading of this manuscript. The opinion and assertions contained herein are the private views of the authors and are not to be construed as official or as representing the views of the Department of the Army or the Department of Defense.

**Grant Support:**

This research was supported by a grant from the United States Department of Defense (Military Molecular Medicine Initiative MDA W81XWHH-05-2-0075, Protocol 01-20006)

**Figure legends:**

Figure 1: Both FH535 and XAV939 inhibited migration of HCC38 cells to type I collagen.

(A) Cells were harvested, washed, and resuspended in RPMI-serum free media at a concentration of  $5 \times 10^5$  cells/ml. Type I collagen was used as a chemoattractant at a concentration of 3  $\mu\text{g/ml}$ . Antibodies (5 $\mu\text{g/ml}$ ) were added in both cell suspension and type I collagen solution and incubated for 4 hours at 37°C. Migrated cells were manually counted and expressed as mean  $\pm$  S.D. Experiments were repeated three times. Statistical significance was calculated by Student's two-tailed *t*-test (paired).

(B) Cells were prepared as described above. FH535 (0.01-1 $\mu\text{M}$ ) and XAV939 (0.001 to 0.1  $\mu\text{M}$ ) were added in both cell suspension and type I collagen solution and incubated for 4 hours at 37°C. Migrated cells were manually counted and expressed as mean  $\pm$  S.D. Experiments were repeated three times. Statistical significance was calculated by Student's two-tailed *t*-test (paired).

Figure 2: Effect of FH535 and XAV939 on the tyrosine phosphorylation pathways in HCC38 cells on type I collagen.

(A) Cells were harvested and cultured on substrate coated with type I collagen for 4 hours in the presence or absence of FH535 (1  $\mu\text{M}$ ) or XAV939 (0.1  $\mu\text{M}$ ). Cells were lysed and the same amount of proteins was separated on SDS-PAGE followed by transferred onto Immobilon-P membranes. Membranes were blotted with anti-phosphotyrosine antibody (PY) followed by anti-actin (actin) antibody. Molecular weights (kDa) were shown in the left column.

(B) The same cell lysates were separated and transferred as described above. Membranes were blotted with anti-phosphopaxillin (Tyr<sup>118</sup>) antibody, anti-paxillin antibody followed by anti-actin antibody.

(C) The same cell lysates were separated and transferred as described above. Membranes were blotted with anti-phosphoFAK (Tyr<sup>397</sup>) antibody, anti-actin antibody followed by anti-actin antibody.

Figure 3: Effect of FH535 and XAV939 on the growth of HCC38 cells.

(A) Cells were harvested and resuspended in RPMI1640-10%FBS at a concentration of  $10^5$  cells/ml. An aliquot (100  $\mu$ l) of cell suspension was incubated with FH535 (20  $\mu$ M) or XAV939 (2  $\mu$ M) for 24 hours. Cell growth was measured as an Absorbance at 450 nm of WST-1 as described in manufacture's instruction and expressed as mean +/- S.D. Experiments were repeated three times. Statistical significance was calculated by Student's two-tailed *t*-test (paired).

(B) Cells were prepared as described above and incubated in the presence of FH535 or GM6001 for 24 hours. Cell growth was measured and showed as described above. Experiments were repeated three times. Statistical significance was calculated by Student's two-tailed *t*-test (paired).

Figure 4: FH535 but not XAV939 inhibited production of MMP-2 and MMP-9 from HCC38 cells.

(A) Cells ( $7 \times 10^5$  cells/well) were cultured overnight in RPMI-1640-10%FBS, washed, and replaced with serum-free RPMI1640 media in the presence or absence of inhibitors overnight. The conditioned media were analyzed on gelatin/zymography. The concentrations of inhibitors were FH535 at 20  $\mu$ M and XAV939 at 2  $\mu$ M. The localization of pro-MMP-9, pro (p), intermediate (i), and active (a) forms of MMP-2 were shown in the left column using serum free conditioned media prepared from ConA-stimulated HT1080 cells.

The same serum free conditioned media were analyzed for the production of MMP-2 (B), MMP-9 (C), and MMP-1 (D) by Luminex. Statistical significance was calculated by Student's two-tailed *t*-test (paired).

Figure 5: Both MMP-2 and MMP-9 were downstream targets of FH535.

Cells ( $7 \times 10^5$  cells/well) were cultured overnight in RPMI-1640-10%FBS, washed, and replaced with serum-free RPMI1640 media in the presence or absence of inhibitors overnight as described in the figure legend of Figure 4. Total RNA was extracted and the same amount of RNA was transcribed to DNA using random hexamer. The DNA was used as a template for PCR analysis for MMP-2, MMP-9, and Tcf-4. GAPDH was used as a control.

Table 1: Secretion of MMP-1, MMP-2, and MMP-9 from various breast cancer cell lines

<sup>a</sup> Cell lines	<sup>b</sup> pg/ml (Mean +/- S.D.)		
	MMP-1	MMP-2	MMP-9
SK-Br3	< 10.7	< 95.3	< 60.4
HCC1419	< 10.7	< 95.3	< 60.4
HCC2218	< 10.7	< 95.3	< 60.4
HCC1937	< 10.7	< 95.3	<u>310 +/- 5</u>
HCC1806	<u>10360 +/- 443</u>	< 95.3	< 60.4
HCC38	< 10.7	<u>600 +/- 28</u>	<u>833 +/- 71</u>
MDA-MB-231	<u>2937 +/- 22</u>	< 95.3	< 60.4

<sup>a</sup>Serum free conditioned media was prepared from each of confluent cell culture.

<sup>b</sup>Production of MMPs were measured by Luminex as described in the Materials and Methods.

## References

1. Hynes RO. The extracellular matrix: not just pretty fibrils. *Science (New York, NY)* 2009;326(5957):1216-9.
2. Cox D, Brennan M, Moran N. Integrins as therapeutic targets: lessons and opportunities. *Nature reviews*;9(10):804-20.
3. Wend P, Holland JD, Ziebold U, Birchmeier W. Wnt signaling in stem and cancer stem cells. *Seminars in cell & developmental biology*;21(8):855-63.
4. Segditsas S, Tomlinson I. Colorectal cancer and genetic alterations in the Wnt pathway. *Oncogene* 2006;25(57):7531-7.
5. Behrens J. Control of beta-catenin signaling in tumor development. *Annals of the New York Academy of Sciences* 2000;910:21-33; discussion -5.
6. Geyer FC, Lacroix-Triki M, Savage K, *et al.* beta-Catenin pathway activation in breast cancer is associated with triple-negative phenotype but not with CTNNB1 mutation. *Mod Pathol*.
7. Handeli S, Simon JA. A small-molecule inhibitor of Tcf/beta-catenin signaling down-regulates PPARgamma and PPARdelta activities. *Molecular cancer therapeutics* 2008;7(3):521-9.
8. Huang SM, Mishina YM, Liu S, *et al.* Tankyrase inhibition stabilizes axin and antagonizes Wnt signalling. *Nature* 2009;461(7264):614-20.
9. Iida J, Wilhelmson KL, Price MA, *et al.* Membrane type-1 matrix metalloproteinase promotes human melanoma invasion and growth. *The Journal of investigative dermatology* 2004;122(1):167-76.
10. Drury LJ, Wendt MK, Dwinell MB. CXCL12 chemokine expression and secretion regulates colorectal carcinoma cell anoikis through bim-mediated intrinsic apoptosis. *PloS one*;5(9):e12895.

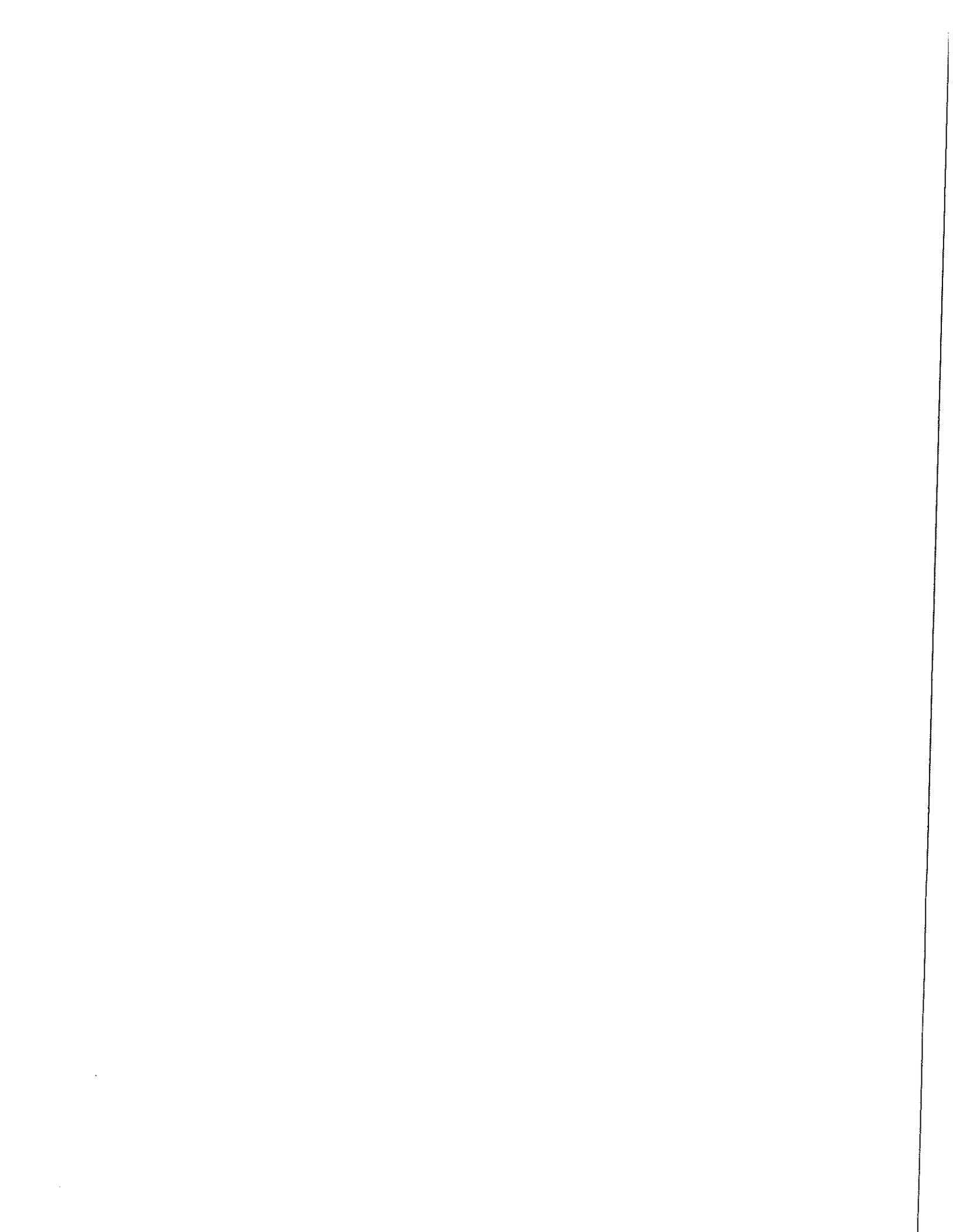
11. Carter WG, Wayner EA, Bouchard TS, Kaur P. The role of integrins alpha 2 beta 1 and alpha 3 beta 1 in cell-cell and cell-substrate adhesion of human epidermal cells. *The Journal of cell biology* 1990;110(4):1387-404.
12. Languino LR, Gehlsen KR, Wayner E, Carter WG, Engvall E, Ruoslahti E. Endothelial cells use alpha 2 beta 1 integrin as a laminin receptor. *The Journal of cell biology* 1989;109(5):2455-62.
13. Tavora B, Batista S, Reynolds LE, *et al.* Endothelial FAK is required for tumour angiogenesis. *EMBO molecular medicine*;2(12):516-28.
14. Zhao X, Guan JL. Focal adhesion kinase and its signaling pathways in cell migration and angiogenesis. *Advanced drug delivery reviews*.
15. Azuma K, Tanaka M, Uekita T, *et al.* Tyrosine phosphorylation of paxillin affects the metastatic potential of human osteosarcoma. *Oncogene* 2005;24(30):4754-64.
16. Cifone MA. In vitro growth characteristics associated with benign and metastatic variants of tumor cells. *Cancer metastasis reviews* 1982;1(4):335-47.
17. Freedman VH, Shin SI. Cellular tumorigenicity in nude mice: correlation with cell growth in semi-solid medium. *Cell* 1974;3(4):355-9.
18. Radisky ES, Radisky DC. Matrix metalloproteinase-induced epithelial-mesenchymal transition in breast cancer. *Journal of mammary gland biology and neoplasia*;15(2):201-12.
19. Micalizzi DS, Farabaugh SM, Ford HL. Epithelial-mesenchymal transition in cancer: parallels between normal development and tumor progression. *Journal of mammary gland biology and neoplasia*;15(2):117-34.
20. Harrison H, Farnie G, Howell SJ, *et al.* Regulation of breast cancer stem cell activity by signaling through the Notch4 receptor. *Cancer research*;70(2):709-18.

21. Mine T, Matsueda S, Li Y, *et al.* Breast cancer cells expressing stem cell markers CD44+ CD24 are eliminated by Numb-1 peptide-activated T cells. *Cancer Immunol Immunother* 2009;58(8):1185-94.
22. Burridge K, Turner CE, Romer LH. Tyrosine phosphorylation of paxillin and pp125FAK accompanies cell adhesion to extracellular matrix: a role in cytoskeletal assembly. *The Journal of cell biology* 1992;119(4):893-903.
23. Schaller MD, Parsons JT. pp125FAK-dependent tyrosine phosphorylation of paxillin creates a high-affinity binding site for Crk. *Molecular and cellular biology* 1995;15(5):2635-45.
24. Klinghoffer RA, Sachsenmaier C, Cooper JA, Soriano P. Src family kinases are required for integrin but not PDGFR signal transduction. *The EMBO journal* 1999;18(9):2459-71.
25. Lewis JM, Schwartz MA. Integrins regulate the association and phosphorylation of paxillin by c-Abl. *The Journal of biological chemistry* 1998;273(23):14225-30.
26. Mu D, Cambier S, Fjellbirkeland L, *et al.* The integrin alpha(v)beta8 mediates epithelial homeostasis through MT1-MMP-dependent activation of TGF-beta1. *The Journal of cell biology* 2002;157(3):493-507.
27. Yu Q, Stamenkovic I. Cell surface-localized matrix metalloproteinase-9 proteolytically activates TGF-beta and promotes tumor invasion and angiogenesis. *Genes & development* 2000;14(2):163-76.
28. Manes S, Llorente M, Lacalle RA, *et al.* The matrix metalloproteinase-9 regulates the insulin-like growth factor-triggered autocrine response in DU-145 carcinoma cells. *The Journal of biological chemistry* 1999;274(11):6935-45.
29. Manes S, Mira E, Barbacid MM, *et al.* Identification of insulin-like growth factor-binding protein-1 as a potential physiological substrate for human stromelysin-3. *The Journal of biological chemistry* 1997;272(41):25706-12.

30. Agrez M, Chen A, Cone RI, Pytela R, Sheppard D. The alpha v beta 6 integrin promotes proliferation of colon carcinoma cells through a unique region of the beta 6 cytoplasmic domain. *The Journal of cell biology* 1994;127(2):547-56.
31. Dallas SL, Rosser JL, Mundy GR, Bonewald LF. Proteolysis of latent transforming growth factor-beta (TGF-beta)-binding protein-1 by osteoclasts. A cellular mechanism for release of TGF-beta from bone matrix. *The Journal of biological chemistry* 2002;277(24):21352-60.
32. Tatti O, Vehvilainen P, Lehti K, Keski-Oja J. MT1-MMP releases latent TGF-beta1 from endothelial cell extracellular matrix via proteolytic processing of LTBP-1. *Experimental cell research* 2008;314(13):2501-14.
33. Kessenbrock K, Plaks V, Werb Z. Matrix metalloproteinases: regulators of the tumor microenvironment. *Cell*;141(1):52-67.
34. Bohn OL, Nasir I, Brufsky A, *et al.* Biomarker profile in breast carcinomas presenting with bone metastasis. *International journal of clinical and experimental pathology* 2009;3(2):139-46.
35. Del Casar JM, Gonzalez-Reyes S, Gonzalez LO, *et al.* Expression of metalloproteases and their inhibitors in different histological types of breast cancer. *Journal of cancer research and clinical oncology*;136(6):811-9.
36. Garcia MF, Gonzalez-Reyes S, Gonzalez LO, *et al.* Comparative study of the expression of metalloproteases and their inhibitors in different localizations within primary tumours and in metastatic lymph nodes of breast cancer. *International journal of experimental pathology*;91(4):324-34.
37. Gonzalez LO, Gonzalez-Reyes S, Junquera S, *et al.* Expression of metalloproteases and their inhibitors by tumor and stromal cells in ductal carcinoma in situ of the breast and their relationship with microinvasive events. *Journal of cancer research and clinical oncology*;136(9):1313-21.

38. Perks CM, Holly JM. IGF binding proteins (IGFBPs) and regulation of breast cancer biology. *Journal of mammary gland biology and neoplasia* 2008;13(4):455-69.
39. Rosfjord EC, Dickson RB. Growth factors, apoptosis, and survival of mammary epithelial cells. *Journal of mammary gland biology and neoplasia* 1999;4(2):229-37.
40. Roussidis AE, Theocharis AD, Tzanakakis GN, Karamanos NK. The importance of c-Kit and PDGF receptors as potential targets for molecular therapy in breast cancer. *Current medicinal chemistry* 2007;14(7):735-43.
41. Karow M, Popp T, Egea V, Ries C, Jochum M, Neth P. Wnt signalling in mouse mesenchymal stem cells: impact on proliferation, invasion and MMP expression. *Journal of cellular and molecular medicine* 2009;13(8B):2506-20.
42. Kim TH, Kim SH, Seo JY, *et al.* Blockade of the Wnt/beta-catenin pathway attenuates bleomycin-induced pulmonary fibrosis. *The Tohoku journal of experimental medicine*;223(1):45-54.
43. Wang XH, Sun X, Meng XW, *et al.* beta-catenin siRNA regulation of apoptosis- and angiogenesis-related gene expression in hepatocellular carcinoma cells: potential uses for gene therapy. *Oncology reports*;24(4):1093-9.
44. Park BC, Thapa D, Lee JS, Park SY, Kim JA. Troglitazone inhibits vascular endothelial growth factor-induced angiogenic signaling via suppression of reactive oxygen species production and extracellular signal-regulated kinase phosphorylation in endothelial cells. *Journal of pharmacological sciences* 2009;111(1):1-12.
45. Yiqin Y, Meilin X, Jie X, Keping Z. Aspirin inhibits MMP-2 and MMP-9 expression and activity through PPARalpha/gamma and TIMP-1-mediated mechanisms in cultured mouse celiac macrophages. *Inflammation* 2009;32(4):233-41.

46. Liu C, He X. Destruction of a destructor: a new avenue for cancer therapeutics targeting the Wnt pathway. *Journal of molecular cell biology*;2(2):70-3.



## **Prestigious Journal to Publish Collaborative Breast Cancer Research**

A landmark paper on breast cancer from National Cancer Institute/The Cancer Genome Atlas project, which includes scientists from the Clinical Breast Care Project (CBCP), was recently published in the highly regarded journal Nature.

The paper, "Comprehensive Molecular Portraits of Human Breast Tumors," is the culmination of three years of work of a consortium of more than 25 institutions including the Windber Research Institute and Walter Reed National Military Medical Center in Bethesda, Md. The consortium collected, sequenced and analyzed data from approximately 400 breast cancer cases.

"It's been a very exciting collaboration for us," said Richard Mural, Ph.D., chief scientific officer with the Windber Research Institute in Windber, Pa. "We've felt very honored to be a part of this collaboration."

Under the leadership of Army Colonel Craig Shriver, M.D., the CBCP's goal is to decrease morbidity and mortality of breast cancer among American women. The multidisciplinary project is unique in incorporating advances in risk reduction, biomedical informatics, tissue banking and translational research.

Because of Shriver's vision and foresight, Mural said, the team was able to produce the quality of samples that have now become the standard. As chief scientific officer, Mural oversees the laboratory work of 35 researchers on the team. His experience working on human genomic projects at corporations such as Celera Genomics made him ideal for leading the effort, which began when he joined the Windber team in 2005.

"I think the paper will be a watershed in breast cancer research," Mural said. Publication in the prestigious journal will lead to citations, media attention and other breast cancer work. He believes this publication will set off a "large collection of these kinds of papers in the future."

The Windber Institute, recognized as one of the world's leading genetic research laboratories, is responsible for maintaining an approximately 45,000-sample tissue repository, performing analysis and providing the bioinformatic structure for all of the clinical and laboratory data of 5,200 volunteer patients.

The CBCP team, which has been working together since 2000, has participated in the National Institutes of Health's Cancer Genome Atlas, a project that charts the genomic changes involved in more than 20 types of cancer.