AD_____

AWARD NUMBER: DAMD17-01-1-0421

TITLE: Prediction of Breast Cancer Risk by Aberrant Methylation in Mammary Duct Lavage

PRINCIPAL INVESTIGATOR: David Euhus, M.D.

CONTRACTING ORGANIZATION: The University of Texas
Southwestern Medical Center at Dallas
Dallas, Texas 75390-9105

REPORT DATE: July 2006

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
01-07-2006	Final	1 July 2001 – 30 Jun 2006

4. TITLE AND SUBTITLE

Prediction of Breast Cancer Risk by Aberrant Methylation in Mammary Duct Lavage

5a. CONTRACT NUMBER

5b. GRANT NUMBER
DAMD17-01-1-0421

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

David Euhus, M.D.

E-Mail: david.euhus@utsouthwestern.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

The University of Texas
Southwestern Medical Center at Dallas
Dallas, Texas 75390-9105

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Assessment of breast epithelial cells obtained by nipple duct lavage (NDL) may have value for breast cancer risk stratification. NDL was performed in 150 women: 67 with an incident breast cancer and 83 unaffected women. Promoter region methylation of Cyclin D2, APC, HIN1, RASSF1A, and RARβ2 was measured in NDL samples by quantitative methylation-specific real time PCR. Methylation of one or more of these genes was detected in 31/35 (88.6%)of primary tumors. NDL ipsilateral to these cancers retrieved atypical cells with methylation profiles that were similar to the cancer in only 9%. Unsupervised clustering revealed three distinct methylation clusters. 15.2% of ducts contralateral to a breast cancer or from unaffected high risk women belonged to one of these clusters as compared to only 7.8% of ducts from lower risk women (RR 1.95, P = 0.12). Identification of marked atypia OR TSG methylation provided the best discrimination between high and lower risk breasts (P < 0.01). Assessment of TSG methylation in benign breast cells may contribute to risk stratification.

15. SUBJECT TERMS
breast cancer, DNA methylation, Methylation Specific PCR, Nipple Duct Lavage, Risk assessment

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			USAMRMC
U	U	U	UU	46	**19b. TELEPHONE NUMBER** *(include area code)*

Table of Contents

Introduction

The Gail model[1] has previously been validated as a tool for breast cancer risk assessment. Cohort studies have consistently demonstrated that the model is well calibrated; that is, for a given population, the ratio of observed to expected breast cancers is near 1.0[2,3]. However, the model does not discriminate well between women who will develop breast cancer and women who will not. Biomarkers have been proposed as an approach for individualized risk stratification with the potential to improve on the discrimination of mathematical models. Characteristics of the ideal biomarker include biologic plausibility, differential expression in low- versus high-risk populations, presence in a reasonable proportion of the high-risk population, association with cancer in prospective studies, expression minimally influenced by normal physiologic processes, the ability to obtain the marker by minimally invasive techniques and an assessment method that provides reproducible results[4,5].

Studies in cell culture suggest that DNA methylation is a very early event in transformation, that precedes spontaneous immortalization[6]. Tumor suppressor gene (TSG) methylation is detectable in nearly all breast cancers but very rarely in histologically normal breast tissue. Its frequent occurrence in benign proliferative breast disease[7,8 ,9,10] in benign breast tissue adjacent to breast cancer[9,11] , and in LCIS[12] suggests that it is an early biomarker of carcinogenesis.

Nipple duct lavage (NDL) is a minimally invasive approach for obtaining breast epithelial cells. Cytological atypia identified in nipple aspirate fluid (NAF)[13] or in random periareolar FNA samples[14] is associated with increased breast cancer risk; cytological atypia diagnosed by NDL is currently being evaluated in a prospective multi-institutional clinical trial. Assessment of biomarkers, other than atypia, in benign breast epithelial cells may provide an individualized approach for breast cancer risk stratification. This study was designed to measure the prevalence of TSG methylation in benign breast epithelial cells obtained by NDL, and to determine whether TSG methylation occurs more frequently in NDL samples from high risk breasts as compared to lower risk breasts.

Study Subjects

Between 10/16/2001 and 6/21/2005 150 women were enrolled in the study. Nipple ducts were successfully cannulated in 149 women (99.3%). The nipple duct lavage procedure was performed in a total of 516 ducts from 291 breasts of 149 women. Samples adequate for cytological diagnosis were obtained for 134 of the 150 women (89.3%). The characteristics of the study sample are summarized in Table 1.

The intention had been to enroll 50 women with a 5 year Gail risk < 1.7%, 50 women with a 5-year Gail risk $\geq$ 1.7%, and 50 women with breast cancer. Accrual of women unaffected with breast cancer proved difficult and the final distribution of evaluable subjects was 44 with a 5-year Gail risk < 1.7%, 39 with a 5-year Gail risk $\geq$ 1.7%, and 67 with breast cancer (Figure 1). This is not felt to pose a serious limitation for completion of the specific aims as paired ipsilateral and contralateral lavages from breast cancer patients are particularly valuable for determining whether TSG methylation in duct lavage samples represents a field change associated with increased breast cancer risk.

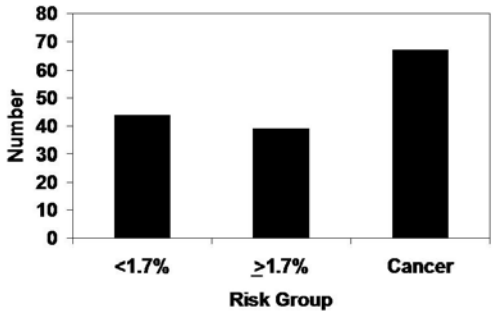


Figure 1: Distribution of subjects by risk categories

150 women were enrolled including 67 with breast cancer and 83 unaffected with breast cancer. Of the 67 patients with breast cancer, three had bilateral breast cancer. There were, potentially, 70 breasts ipsilateral to a breast cancer to lavage. There are only 65 lavages ipislateral to an intact primary cancer. Reasons include:

- Primary cancer excised prior to lavage – 4

- Unable to cannulate a duct – 1

There are 67 patients with breast cancer, but only 62 lavages contralateral to a breast cancer. Reasons include:

- Bilateral breast cancer – 3
- Unable to cannulate a duct – 2

Table 1: Characteristics of the Study Sample

Patients	150
Mean Age (Range)	48 (28-93)
Ethnicity (%)	
Caucasian	123 (82)
African American	20 (13)
Hispanic	5 (3)
Asian	2 (1)
Menopausal Status (%)	
Premenopausal	73 (49)
Perimenopausal	8 (5)
Postmenopausal	69 (46)
Oral Contraceptive Use (premenopausal)	18/73 (38)
Hormone Replacement (peri and postmenopausal)	25/77 (32)
Risk Groups	
Breast Cancer Patients	67 (45)
Breasts Ipsilateral to a Breast Cancer	[a]65
DCIS Only	6
Infiltrating Ductal Carcinoma	50
Infiltrating Lobular Carcinoma	7
Medullary Carcinoma	1
Metaplastic Carcinoma	1
Any Associated DCIS	53 (82)
Breasts Contralateral to a Breast Cancer	[b]62
Unaffected Risk Assessed Patients	83 (55)
History of ADH	4 (5)
BRCA Gene Mutation	3 (4)
5-Year Gail Risk	
0.01 – 0.85	26 (31)
0.86 – 1.69	18 (22)
1.70 – 2.54	23 (28)
>2.54	16 (19)

[a]3 bilateral cancers included; 4 excluded because cancer excised prior to enrollment; 1 excluded because unable to cannulate duct.

[b]3 bilateral cancer patients had no contralateral lavage; 2 excluded because unable to cannulate duct.

Eligibility Criteria

Patients with incident breast cancer, and unaffected women over the age of 18 presenting for breast cancer risk assessment were offered ductal lavage regardless of the calculated risk level. Exclusion criteria included: presence of an undefined palpable or mammographic breast lesion suspicious for malignancy, bilateral prophylactic mastectomy, any prior breasts irradiation, any systemic chemotherapy in the past, performance status that restricts normal activity for a significant portion of the day, current use of androgens, luteinizing-hormone-releasing-hormone (LHRH) analogs, prolactin inhibitors, antiandrogens, or glucocorticoids (women were eligible if these drugs were discontinued three months prior to lavage), ever use of tamoxifen, raloxifene, or other SERM therapy, or pregnancy or lactation within 6 months.

Menopausal Status

Premenopausal was defined as recent menstrual cycles of the usual frequency or, in the case of women who had undergone hysterectomy but not oophorectomy (N =1), age less than 40 and no estrogen withdrawal symptoms. Perimenopausal was defined as one or more menstrual cycles within the last 12 months occurring at an altered frequency as compared to the usual frequency for that patient and symptoms of estrogen withdrawal. Post menopausal was defined as no menstrual cycles in the prior 12 months or a history of bilateral oophorectomy.

Phase of the Menstrual Cycle

Sampling performed between day 1 and 14 (inclusive) of the menstrual cycle was classified as follicular phase sampling, while sampling between days 15 and 30 (inclusive) of the menstrual cycle was classified as luteal phase. If more than 30 days had elapsed since the last menstrual cycle, the phase was classified as unknown.

Hormonal Medication Usage

Hormonal medications were defined as systemic estrogen containing medications. Use within one month of the sampling was classified as “current use.” Vaginal estrogens were not included as hormone use. One premenopausal woman used Depo-Provera and this was not classified as OCP use.

The Duct Lavage Procedure

Local anesthetic cream (EMLA) was applied to the nipple which was then covered with an occlusive patch one to two hours prior to the procedure. At the start of the procedure, the patient performed a self-breast massage after which the nipple was dekeratinized with a mild abrasive gel (Nuprep, D. O. Weaver and Co., Aurora, CO). The operator then continued the breast massage in an effort to express nipple aspirate fluid (NAF). If no NAF was elicited manually a nipple aspirator (FirstCyte, Cytyc Health Corporation) was used. Fluid producing ducts were initially cannulated with a tapered dilator coated with 2% lidocaine jelly after which a ductal lavage microcatheter (FirstCyte Microcatheter, Cytyc Health Corporation) was inserted. Saline (10 ml) was infused into the duct in 0.5 ml increments and the effluent fluid aspirated. An attempt was made to lavage all fluid producing and at least one non-fluid producing duct from each breast. The location of each cannulated duct orifice was recorded on a circular grid with 45 cells so that the orifice of any duct yielding atypical cells could be identified in the future.

On the average, 1.4 dry ducts were lavaged per patient, and 2.1 NAF-producing ducts for a total of 3.5 ducts per patient.

Number and Risk Classification of Subjects, Breasts and Ducts Lavaged

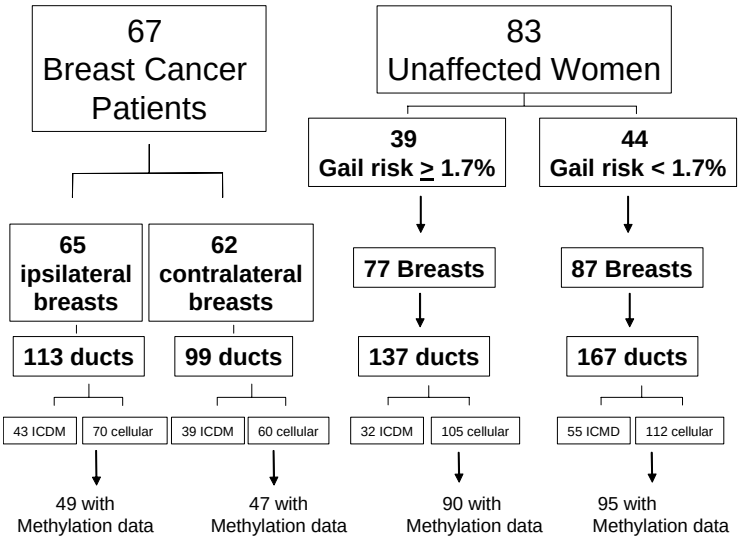


Figure 2: Distribution of study patients, breasts and ducts

Figure 2 shows the distribution of the patients, breasts and ducts that were lavaged. The lavage effluent was acellular for 33% of ducts limiting the samples available for methylation assays. Only samples with methylation results for ≥ 4 of the 5 markers were included in the final analysis.

The median age of the study population was 47 years. Figure 3 shows the distribution of the subjects by age.

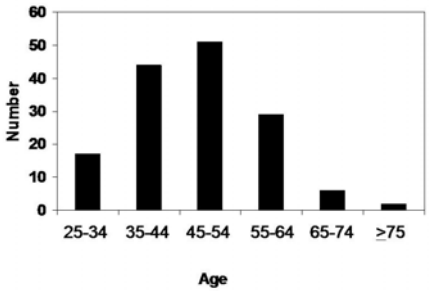


Figure 3: Age distribution of the study sample.

Sample Processing

Duct lavage samples were dispersed into 30 ml of CytoLyt solution as they were obtained. The samples were immediately split with half of the volume submitted for cytology and half submitted for the methylation assays. Samples for methylation were centrifuged at 2000 x G for 10 minutes and after the supernatant had been aspirated, the cell pellet was stored frozen at -80^{o}C until the time of DNA extraction.

DNA Extraction

To maximize the yield of amplifiable DNA in the paucicellular cases we compared five methods for DNA extraction in 50 – 50,000 cell samples of benign and malignant breast epithelium and lymphocytes[15]. Of the five methods evaluated, the QIAamp and Puregene kits showed the best linearity and highest rate of successful amplification. The Puregene method had the added benefit of lower cost. DNA was extracted from the duct lavage samples using the Puregene kit (Gentra, Cat # D-5500A). The final volume after DNA extraction was 20 μl.

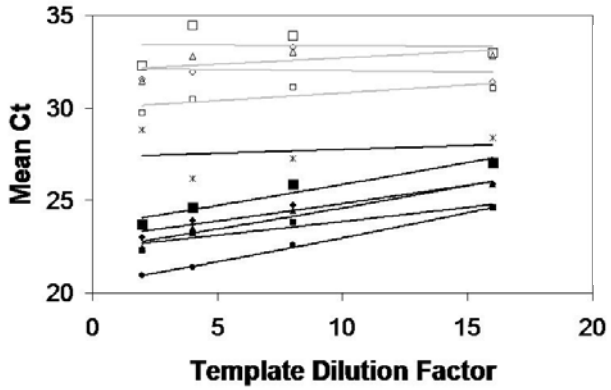


Figure 4: Mean GAPDH Ct values for 10 nipple duct lavage samples extracted using the Puregene method. Template was diluted from 1:2 to 1:16 for each extraction. Closed symbols and solid black lines are for six samples with <1000 cells; open symbols and grey lines are for four samples reported as acellular.

Quantitative Methylation-specific Real Time PCR

Principles of the Methylation Assay

Expression of tumor suppressor genes is often regulated by covalently bonding methyl groups to cytosines in CpG islands found in the promoter regions of these genes. Unmethylated cytosines are converted to uracil when the DNA is treated with sodium bisulfite, but methylated cytosines are not. Methylation-specific PCR (MSP) assays use PCR primers specific for unconverted cytosines or converted cytosines to amplify promoter region DNA that was initially methylated or unmethylated. Qualitative MSP assays score a sample as positive or negative for methylation based on bands that can be visualized when the PCR products are resolved by electrophoresis on agarose gels. Real time quantitative MSP (qMSP) monitors the generation of PCR products from methyl-DNA-specific reactions and unmethylated DNA-specific reactions. The fraction of DNA that was initially methylated is calculated based on standard curves and results for internal standards. We used the multiplex qMSP method of Sukumar[16] to measure methylation for five genes in DNA from our duct lavage samples. Following sodium bisulfite treatment, selected regions of the DNA are amplified in a PCR that included primers for all five genes. PCR products from the initial multiplex reaction were then amplified again in uniplex reactions using primers specific for methylated and unmethylated DNA. Elaboration of PCR products is monitored real time as fluorochrome quenchers are released from gene-specific probes through the action of DNA polymerase.

Marker Selection

We have previously evaluated tumor suppressor gene methylation in benign and malignant breast cells obtained by fine needle aspiration biopsy[17]. We found that Cyclin D2 is methylated only in malignant cells, while genes like RASSF1A, APC, and RAR-β are also methylated in benign breast epithelium, but at a lower frequency. RAR-β is of interest, because it is methylated more frequently in benign breast cells from breast cancer patients (45%) than benign breast from women unaffected with breast cancer (9%) suggesting it may be a high risk field change acquired early in breast carcinogenesis. The frequency of APC and RASSF1A methylation in benign breast cells from unaffected women correlated with breast cancer risk calculated using the Gail model. The Gail risk factor driving this association was the number of prior breast biopsies, suggesting that methylation of these genes may be involved in preneoplasia. In the current study, we were interesting in determining whether evaluation of cancer-specific methylation markers (i.e. Cyclin D2) can improve the diagnostic accuracy of duct lavage for the detection of breast cancer, and whether evaluation of risk-related markers (i.e. APC, RASSF1A, or RAR-β) can be used to identify women at high risk for breast cancer. We added HIN-1 to our panel, because it is frequently methylated in benign breast epithelium. Its relation to breast cancer risk is currently unknown.

Each gene we selected has previously been shown to be regulated by promoter region methylation. Primers were specifically chosen to amplify a region of the promoter known to silence gene expression when methylated. Publications supporting our marker and primer selection include: Cyclin D2[18], APC (promoter A1)[19,20], HIN-1[21], RASSF1A[22], and RAR-β2[23].

Sodium Bisulfite Treatment

To estimate the quantity of amplifiable DNA in each sample, GAPDH was amplified from 1 μl of the DNA extraction and the PCR products were resolved by electrophoresis on an agarose gel. For samples producing strong bands relative to a 100ng/μl standard prepared from HCC1954 cells, 5 μl of DNA was sodium bisulfite treated; for samples producing bands similar to the standard, 10 μl was treated; and for samples producing bands that were weaker than the standard, the entire 20 μl of DNA was sodium bisulfite treated. Yeast tRNA was used as a carrier for all sodium bisulfite treatments so that product recovery would not pose a limitation for paucicellular samples.

Sodium bisulfite treatment was performed using the method of Clark[24]. Briefly, the DNA was denatured by adding NaOH to a final concentration of 0.3M in the presence of 1.25 μg of yeast tRNA and incubating for 15 minutes at 37°C. Hydroquinone, to a final concentration of 0.5mM, and sodium bisulphite, to a final concentration of 3.1M were added to the denatured DNA to a final volume of 600 μl. The sample was then gently mixed and incubated at 55°C for 16 hours. DNA was purified using a desalting column (Promega Magic DNA Clean-Up System), according to the manufacturer's instructions, and DNA was eluted in 50 μl of H_2O. Freshly prepared NaOH, to a final concentration of 0.3M, was added and the sample incubated at 37°C for 15 minutes. The solution was neutralized by addition of NH_4OAc, pH 7, to 3M and the DNA was ethanol precipitated in the presence of glycogen. The pellet was rinsed with 70% ethanol and then resuspended in 20 μl of DNA hydration solution (Gentra), aliquoted and stored at -80°C.

Multiplex PCR

Multiplex PCR was performed using the Qiagen multiplex PCR kit (Cat.No.206143). The reaction mixture was prepared by combining 2X Qiagen buffer 12.5μl, ultrapure water 9.8μl, first round primers (15μM, FW and REV together) 0.34μl x 5 (for 5 genes), and 1μl of DNA for a total volume is 25 μl.

We used an MJ Research PTC220 thermocycler with four independent heating blocks and the following PCR program: 95°C 15 min, and then 40 cycles of 94°C 30sec, 58°C 90sec, 72°C 90sec followed by, 72°C 10min with a final hold at 4°C.

First round multiplex primers were designed to bracket the region of interest external to the CpGs subject to methylation. The multiplex PCR primers we used are as follows:

Table 2: Primers for first round multiplex PCR		
Cyclin D2 F R1		tatttttttgtaaagatagttttgat
Cyclin D2 R R1		tacaactttctaaaaaataaccc
RASSF1A F R1		gttttatagtttttgtatttagg
RASSF1A R R1		aactcaataaactcaaactccc
RARβ2 F R1		gtaggagggtttatttttttgtt
RARβ2 R R1		aattacattttccaaacttactc
HIN1 F R1		gtttgttaagaggaagtttt
HIN1 R R1		ccgaaacatacaaaacaaaaccac
APC F R1		gggttagggttaggtaggttgtg
APC R R1		aactacaccaatacaaccacata

Uniplex PCR

The uniplex PCR reaction mixture was prepared as follows: pure water 16.49μl, primers (15μM FW and REV together) 0.83μl, Probe (15μM) 0.25μl, dNTP (1.25mM) 3.75μl, 10X buffer (HotstarTaq kit, Cat. No 203205) 2.5μl, Taq (5units/μl) 0.18μl and 1μl of the first round PCR products (diluted 1:100).

We used a Chromo4 real time PCR machine (MJ Research) running Opticon Monitor 3.00.367 on a dedicated HP workstation and the following program: 95°C for 10min followed by 40 cycles of 95°C 30sec, 60°C 45sec.

The second round uniplex PCR was nested within the region amplified by the first round primers. The uniplex primers were designed to bind specifically to methylated or unmethylated CpGs. The second round uniplex primers and probes were as follows.

Table 3: Second Round Uniplex PCR Primers and Probes		
Cyclin D2 R2-FM		tttgatttaaggatgcgttagagtacg
Cyclin D2 R2-RM		actttctccctaaaaaccgactacg
Cyclin D2 R2-FUM		ttaaggatgtgttagagtatgtg
Cyclin D2 R2-RUM		aaactttctccctaaaaaccaactacaat
RASSF1A R2-FM		gcgttgaagtcggggttc
RASSF1A R2-RM		cccgtacttcgctaactttaaacg
RASSF1A R2-FUM		ggtgttgaagttggggtttg
RASSF1A R2-RUM		cccatacttcactaactttaaac
RARβ2 R2-FM		agaacgcgagcgattcgagtag
RARβ2 R2-RM		tacaaaaaaccttccgaatacgtt
RARβ2 R2-FUM		ttgagaatgtgagtgatttgagtag
RARβ2 R2-RUM		ttacaaaaaaccttccaaatacattc
HIN1 R2-FM		tagggaaggggtacgggttt
HIN1 R2-RM		cgctcacgaccgtaccctaa
HIN1 R2-FUM		aagtttttgaggtttgggtaggga
HIN1 R2-RUM		accaacctcacccacactccta
APC R2-FM		tattgcggagtgcgggtc
APC R2-RM		tcgacgaactcccgacga
APC R2-FUM		gtgttttattgtggagtgtgggtt
APC R2-RUM		ccaatcaacaaactcccaacaa
Cyclin D2 UM probe		HEX-aatccaccaacacaatcaaccctaac-BHQ1
Cyclin D2 M Probe		6FAM-aatccgccaacacgatcgacccta-BHQ1
RASSF1A UM probe		HEX-ctaacaaacacaaaccaaacaaaacca-BHQ1
RASSF1A M Probe		6FAM-acaaacgcgaaccgaacgaaacca-BHQ1
RARβ2 UM probe		HEX-aaatcctaccccaacaatacccaaac-BHQ1
RARβ2 M probe		6FAM-atcctaccccgacgatacccaaac-BHQ1
HIN1 UM probe		HEX-caacttcctactacaaccaacaaacc-BHQ1
HIN1 M probe		6FAM-acttcctactacgaccgacgaacc-BHQ1
APC UM probe		6FAM-aacaccctaatccacatccaacaaat-BHQ1
APC M probe		6FAM-aacgccctaatccgcatccaacga-BHQ1

Real Time PCR Optimization

The real time PCR was initially optimized by monitoring melting curves generated using SYBR green. These curves were used to maximize the linearity and dynamic range of the quantitative assay and to exclude primer dimer formation.

Standard Curves and Internal Standards

A standard curve was run on every plate. The standards were prepared by mixing HCC1954 DNA (100% methylated for all markers) with normal lymphocyte DNA (0% methylated) 20:80. DNA from the cell mixture was extracted, sodium bisulfite treated, and then amplified in a multiplex PCR that included the five markers. This standard DNA solution was stored frozen at -80^{o}C. Standard curves were generated by diluting the stock DNA 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, and 10^{-7}. Standards were run in duplicate with the same primers and probes being used for the test samples on any given plate. The Ct threshold value established by the Opticon Monitor software was generally accepted, but was occasionally adjusted up or down to achieve a standard curve slope between -3.11 and -3.58, and an R^2 value $\geq$ 0.985.

Copy number standards were prepared for each gene (unmethylated and methylated) using lymphocytes for unmethylated DNA and the following cell lines for methylated DNA: HCC1954 (APC, HIN1 and RASSF1A), HCC1569 (Cyclin D2), and MCF7 (RAR-β). DNA was extracted, sodium bisulfite-treated, first round amplified by uniplex PCR and then amplified a

second time by nested uniplex PCR using the 10 primer pairs (methylated or unmethylated) corresponding to the five genes of interest. PCR products were subjected to gel electrophoresis and the DNA eluted from the resultant bands. DNA content was quantified using picogreen and the molecular weight of each PCR product was determined using the Biopolymer Calculator v4.1.1. The concentration of each standard was adjusted to 4 x 10^6 copies/μl in 1 mg/ml salmon sperm carrier DNA. These stock solutions were stored at -20°C and then diluted 100-fold to provide 40,000 copy standards (40K standards). Points along the standard curve were translated to copy number based on the Ct of the 40K standard run on the same plate.

Controls run on every plate included: a) serially diluted DNA for the standard curve, b) 40K standard DNA, c) positive control with known methylation fraction, d) a water blank that had been subjected to sodium bisulfite treatment and multiplex PCR, e) a no template (water blank) "product" from the multiplex PCR, f) a uniplex reaction mixture with no template, and g) a uniplex reaction mixture with a known DNA template, but no Taq.

Calculating Percent Methylation

The mean Ct value for duplicate test samples was converted to DNA copy number based on the 40K standard run on that plate. Percent methylation for a given sample was calculated as methylated copies/(methylated copies + unmethylated copies).

Quality Assurance Standards

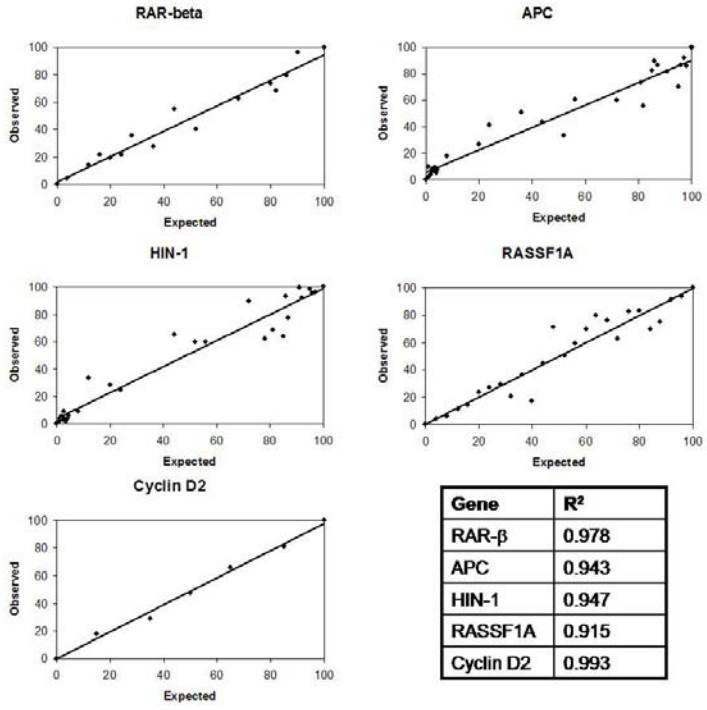
Assays failing to meet all of the quality assurance standards were repeated. The standards are as follows:

1. The slope of the standard curve must be within -3.11 and -3.58 without removing any points (this correlates with reaction efficiencies of 110% and 90% respectively).

2. R^2 for the standard curve must be $\geq$ 0.985.

3. The difference in duplicate Ct values must be less than 1.6.

4. The average Ct for the test sample must fall within the average Ct values of the standards.



Gene	R^2
RAR-β	0.978
APC	0.943
HIN-1	0.947
RASSF1A	0.915
Cyclin D2	0.993

Figure 5: Linearity of the Q-MSP assay.

Performance Characteristics of the Assay

The accuracy of the assay for measuring the fraction of DNA copies that are methylated in a given sample was initially assessed by mixing breast cancer cell lines that were 100% methylated for a given gene in various proportions with benign epithelial cells that were 0% methylated. Observed methylation fractions were plotted against expected (actual) methylation fractions.

The assay demonstrated excellent linearity across the dynamic range of 0 – 100% methylation (Figure 5). Additional experiments were performed to assess linearity for the low range (0 – 5%). The assay does not accurately distinguish between samples with, say, 1% methylation and samples with 3% methylation, but only 4% of samples with a known low level of DNA methylation (< 5%) returned a result of 0%.

Intra-assay and Inter-assay Reproducibility

Intra-assay reproducibility for methylation fraction was measured for samples with low methylation (<1%), moderate methylation (~10%) and high methylation (~80%).

Table 4: Intra-assay reproducibility. Coefficients of Variation for Methylation Fraction					
	APC	**Cyclin D2**	**HIN-1**	**RAR-β2**	**RASSF1A**
Low Methylation (< 1%)	0.436	0.423	0.148	0.284	0.298
Intermediate Methylation (~10%)	0.501	0.654	0.207	0.366	0.313
High Methylation (~80%)	0.119	0.305	0.003	0.244	0.143

Inter-assay reproducibility was measured by repeating the analysis for the same samples on five different days.

Table 5: Inter-assay reproducibility.					
	APC	**Cyclin D2**	**HIN-1**	**RAR-β2**	**RASSF1A**
Mean Methylation Fraction	0.113	0.112	0.432	0.070	0.027
Coefficient of Variation	0.450	0.245	0.159	0.297	0.555

In general, the reproducibility of the assay is lower than that commonly reported for assays like ELISA which usually show Coefficients of Variation $\leq$ 0.15. Nevertheless, for RASSF1A, which had the lowest inter-assay reproducibility, a methylation fraction of 0.06 would fall two standard deviations above a methylation fraction of 0.03 permitting reliable discrimination between these values. We could not identify any studies reporting intra- and inter-assay reproducibility for other quantitative methylation-specific PCR assays. Some investigators have reported coefficients of variation calculated from Ct values[25], but this is inappropriate as each Ct increment represents a log change in product quantity. A difference of only 2 Ct points will correspond to a >3-fold difference in product.

Sensitivity

The 40K methylated standards for each gene were mixed with unmethylated DNA in various proportions to achieve methylation fractions that ranged from 0.001 to 0.00001. Linearity was lower at these ultralow template concentrations, though it remained acceptable with R^2 values ranging from 0.780 – 0.995 (Figure 6). The sensitivity of the assay is measured at $\leq$ 1 methylated gene copy among 100,000 unmethylated copies.

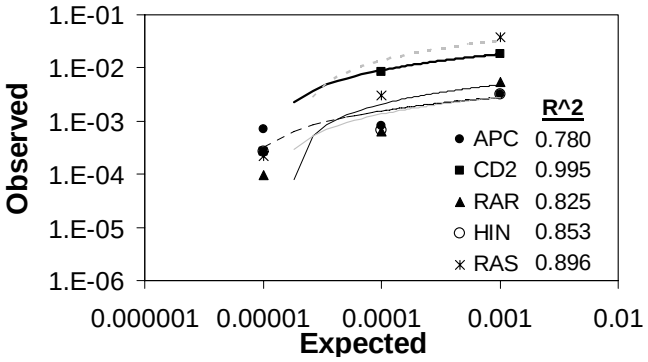


Figure 6: Linearity and sensitivity of the Q-MSP assay at ultra low methylation fractions.

Relationship between Qualitative and Quantitative Results

We have previously correlated methylation of RASSF1A or APC in random fine needle aspiration samples with breast cancer risk calculated using the Gail model[17]. This was done using a qualitative methylation-specific PCR. To understand the relationship between a qualitative positive (i.e. a band on a gel) and specific quantitative values 40 samples were evaluated using both assays. Every sample producing a methylated band on a gel was associated with a qMSP value >0. Seven of 22 samples that were negative by the qualitative assay showed methylation fractions that ranged from 0.0013 to 0.1865 (Figure 7). Results from this, and the linearity experiments, suggest that very low qMSP methylation fractions are not artifactual.

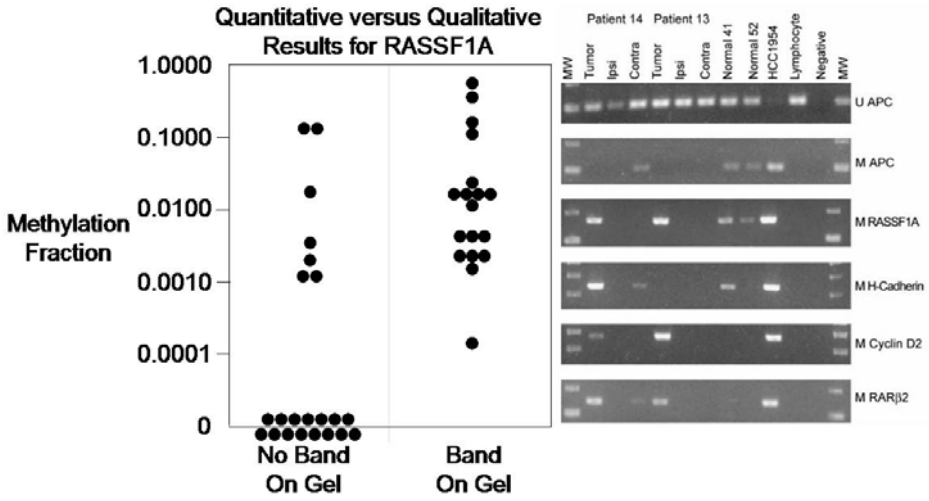


Figure 7: Comparison of results for qualitative MSP and quantitative RT-Q-MSP for RASSF1A.

Cytologic Evaluation

The lavage effluents from each duct were collected separately in 30 ml of CytoLyt solution (Cytyc Health Corporation). Cytology slides were prepared using the thin-prep method and stained using the Papanicolaou technique. All slides were evaluated by the same breast cytopathologist (Ashfaq). The epithelial cell yield for each sample was estimated as insufficient cellular material for diagnosis (ICMD), scant cellularity but sufficient for diagnosis (~10 cells), 11-99 cells, 100-999 cells, or ≥ 1000 cells. The cytopathologist classified each sample according to the most severe alterations identified: normal epithelium or apocrine metaplasia only, typical epithelial hyperplasia, mild atypia, or marked atypia. Cytological interpretation was performed according to the guidelines established by the Cytyc Health Corporation (http://www.ductallavage.com/professionals/cytologyTraining.cfm). Briefly, mild atypia was defined as clusters of crowded, overlapping cells with slight nuclear enlargement, mild anisonucleosis, permanent nucleoli, occasional myoepithelial cells and granular evenly distributed chromatin. Marked atypia was diagnosed when these same features were more pronounced and included marked anisonucleosis, significantly increased nuclear:cytoplasmic ratios and irregular, clumping chromatin.

Cytological preparations were also classified according to the Masood score[26]. Each of six cytological features is assigned a score of 1 – 4. These cytological features include cell arrangement, pleomorphism, number of myoepithelial cells, anisonucleosis, nucleoli, and chromatin clumping. Non-proliferative samples generally score in the 6–10 range, hyperplasia 11–14, and hyperplasia with atypia ≥15.

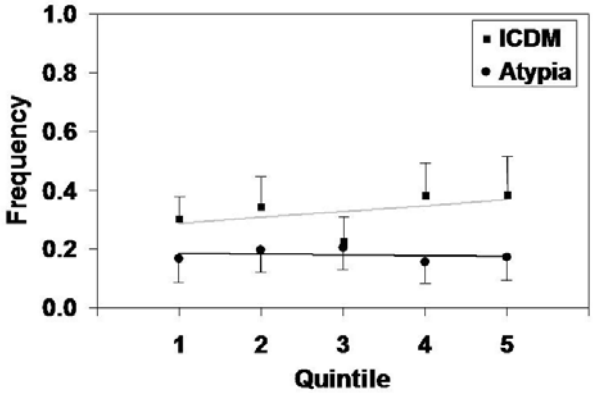


Figure 8: ICDM and atypia rates over time. ICDM is Insufficient Cellular Material for Diagnosis

Atypia Rates

The lavage returned insufficient cellular material for diagnosis (ICMD) in 168 of the 516 ducts (33%). NDL cytology was interpreted as atypical in 94 of the 516 ducts (18%). Mild atypia was diagnosed in 60 ducts (12%) and marked atypia in 34 (7%). ICDM and atypia rates in our series of women representing a wide range of breast cancer risk are nearly identical to those reported for a large series of high risk women[27]. Both ICMD and atypia rates were stable throughout the

study (Figure 8) suggesting that there were no significant changes in the technical performance of the lavage or in the cytological interpretation during the study period.

Table 6 summarizes the cytology results by duct and by breast.

Table 6: Frequency of Atypia by Sampling Group (%)

	Ducts				Breasts			
	ICMD	Mild Atypia	Marked Atypia	Any Atypia	ICMD	Mild Atypia	Marked Atypia	Any Atypia
Breast with Cancer	43/113 (38.1)	13/113 (11.5)	12/113 (10.6)	25/113 (22.1)	19/65 (29.2)	8/65 (12.3)	10/65 (15.4)	18/65 (27.7)
Contralateral to Cancer	38/99 (38.4)	7/99 (7.1)	9/99 (9.1)	16/99 (16.2)	17/62 (27.4)	7/62 (11.3)	8/62 (12.9)	15/62 (24.2)
Unaffected Right	45/156 (28.8)	18/156 (11.5)	7/156 (4.5)	25/156 (16.0)	15/83 (18.1)	14/83 (16.9)	5/83 (6.0)	19/83 (22.9)
Unaffected Left	42/148 (28.4)	22/148 (14.9)	6/148 (4.1)	28/148 (19.0)	18/81[a] (22.2)	15/81 (18.5)	4/81 (4.9)	19/81 (23.5)

ICMD is insufficient cellular material for diagnosis

[a]81 unaffected patients underwent bilateral lavage, but it was not possible to cannulate a duct in the left breast for two of the unaffected women.

Marked atypia was diagnosed in 10.6% of ducts from breasts with a known breast cancer, but only 4.3% of ducts from women unaffected with breast cancer (P = 0.03). Ducts from breasts contralateral to a breast cancer had an increased rate of marked atypia (9.1%) compared to ducts from women unaffected with breast cancer, but this difference was not statistically significant (P = 0.12). Mild atypia was diagnosed with approximately equal frequency in ducts from breasts with a known breast cancer, ducts contralateral to a breast cancer, and ducts from women without breast cancer. Marked atypia was diagnosed more frequently in breasts with a known cancer (15.4%) than breasts from women unaffected with breast cancer (5.5%, P = 0.02). The prevalence of marked atypia for breasts contralateral to a breast cancer (12.9%) was intermediate between that of breast affected with breast cancer and those from women without breast cancer.

The presence of a DCIS component did not predict lavage atypia as four of the 12 breast with breast cancer and NO DCIS component (33%) returned an atypical lavage as compared to 15 of 53 breasts with breast cancer that included a DCIS component (28%) (P = 0.98).

Ducts from patients with breast cancer were more likely to return a diagnosis of ICMD (38.2%) than ducts from women without breast cancer (28.6%, P = 0.03). ICMD rates ranged from 18.2% to 29.1% for breasts affected with breast cancer, breasts contralateral to a breast cancer, and breasts from women unaffected with breast cancer but none of these differences was statistically significant.

Nipple Aspirate Fluid (NAF) Production

NAF was expressible from 123 (82%) of the 150 patients. NAF was expressible from 49/65 (75.4%) of breasts with a known breast cancer, 48/62 (77.4%) of breasts contralateral to a breast cancer, and 116/164 (70.7%) of breasts from women unaffected by breast cancer (P = NS). The ICMD rate was 91/314 (29.0%) for NAF-producing ducts and 77/202 (38.1%) for ducts that did not produce NAF (P = 0.04). The ICMD rate was 42/213 (19.7%) for NAF-producing breasts and 30/78 (38.5%) for breasts that did not produce NAF (P = 0.002).

Factors Predicting Lavage Atypia

Published guidelines recommend that women with a 5 year Gail risk $\geq$ 1.7% consider having lavage performed for any fluid-yielding ducts[28,29], but it is not known whether increased

breast cancer risk or nipple fluid production predicts lavage atypia. After excluding 113 ducts from breasts with a known breast cancer, we compared atypia rates for fluid-producing and dry ducts. The cytological atypia rate was similar for the 240 NAF producing ducts (19%) and the 163 dry ducts (15%, P=0.36).

No significant differences were found for NAF-producing and dry ducts when atypia was categorized as mild (13% vs. 10%, P = 0.63) or marked (6% vs. 4%, P = 0.53). Among the 83 women unaffected with breast cancer, atypia was diagnosed in 15/44 (34%) with a 5-year Gail risk <1.7% and 10/39 (26%) with a 5-year Gail risk ≥ 1.7% (p=0.55). Among these women, the prevalence of mild atypia declined with advancing age while the prevalence of marked atypia increased (Figure 9). Overall, the prevalence of any atypia was inversely associated with increasing age, but this result was not statistically significant. Consistent with our earlier preliminary analysis[30] we conclude that neither NAF production nor 5 year Gail risk predicts lavage atypia.

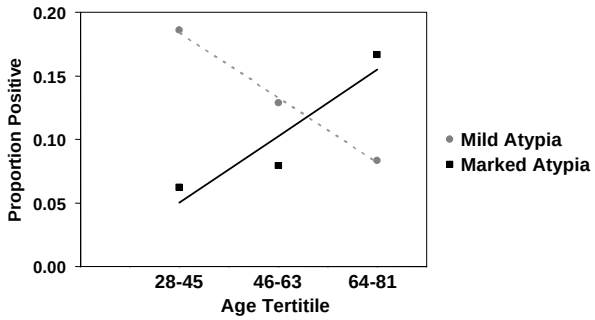


Figure 9: Atypia prevalence by age.

Methylation Markers

Methylation data is available for 320 (62.0%) of the 516 ducts that were lavaged. Reasons for excluded methylation data include:

No epithelial cells in the sample: 168 (32.6%)
Sample mislabeled or lost: 16 (3.1%)
DNA did not amplify after multiple attempts: 12 (2.3%)

Five markers were evaluated: Cyclin-D2, APC, HIN-1, RAR-β, and RASSF1A

For the purposes of data analysis and summarization, we only included ducts where methylation results could be obtained for four or five markers. The rationale being, that if the Q-MSP assay was not successful for two or more markers, then the results for the other markers were questionable. The following analysis includes methylation data for 281 ducts from 185 breasts of 118 subjects.

Duct Lavage for Breast Cancer Detection

Cytological assessment of NDL samples has a very low sensitivity for the detection of breast cancer[31]. Application of biomarker assays, such as tumor suppressor gene methylation, may improve the sensitivity of the test for early detection of breast cancer. Q-MSP data was available for 35 tumor tissue FNAs from patients participating in the duct lavage study. Figure 10 shows the methylation pattern for the cancers and for duct lavage samples obtained from the breast ipsilateral to these cancers. Each patient had results for at least one ipsilateral duct, 10 patients had results for two ducts and two patients had results for three ducts. The data is sorted by the extent of methylation of the cancers (from most methylated to least methylated).

Methylation was detectable in 31 (89%) of 35 tumor samples and 30 (64%) of 47 ducts ipsilateral to these breast cancers. Duct lavage convincingly retrieved cancer cells in only three of these 35 cases (9%). The lavage samples from patients 4 and 19 in the figure contained markedly atypical cells with a methylation profile that was very similar to that of the cancer. The cells for patient 1 were only mildly atypical, but the methylation pattern in the ipsilateral lavage sample was similar to that of the tumor FNA. These data confirm that the Q-MSP assay is reliably detecting tumor suppressor gene methylation and the selected marker panel is relevant to breast cancer. Based on these methylation signatures, we conclude that duct lavage in breasts ipsilateral to a breast cancer rarely retrieves cancer cells.

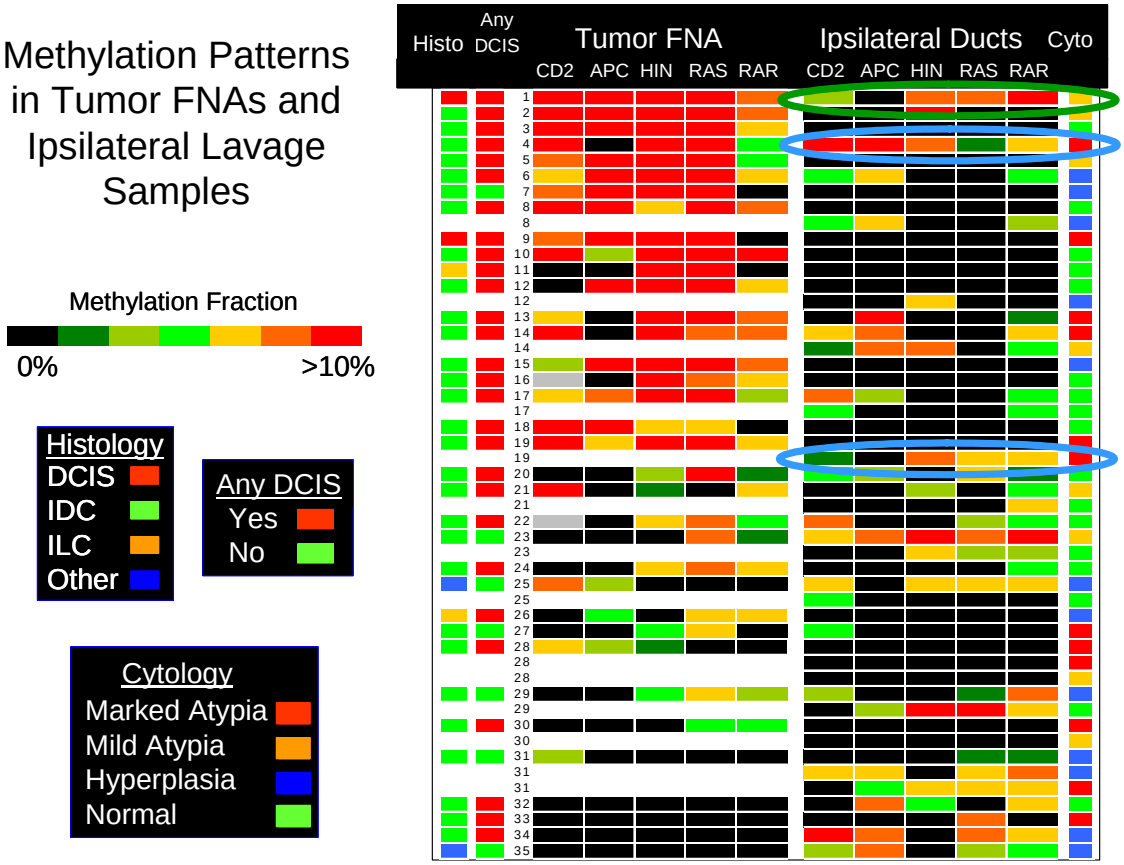


Figure 10: Methylation profiles of tumor cells obtained by FNA as compared to ducts ipsilateral to the same cancer.

Methylation and Duct Cytology

Masood has previously described a cytological scoring system in which each of six cytological features is assigned a score of 1 – 4[26]. These cytological features include cell arrangement, pleomorphism, number of myoepithelial cells, anisonucleosis, nucleoli, and chromatin clumping. Non-proliferative samples generally score in the 6–10 range, hyperplasia 11–14, and hyperplasia with atypia 15–18.

Table 7: Masood Score: Prevalence of methylation positivity by gene and cytology (Threshold = > 95th percentile for ducts with Masood Score ≤ 10)

	CD2	APC	HIN1	RASSF1A	RAR-β2
Threshold for Positive	>2.0%	>5.0%	>10%	>2.2%	>1.9%
≤10	0.051	0.051	0.053	0.052	0.051
11 - 14	0.085	0.113	0.015	0.127	0.043
≥ 15	0.106	0.106	0.130	0.128	0.087
P-Values for Distribution	NS	≤0.20	≤0.05	≤0.10	NS

There is a trend for increasing methylation prevalence with increasing cytological alterations for each of genes. The distribution is significant for HIN1 and approaches significance for RASSF1A.

A composite methylation score was calculated for each duct by averaging the methylation fractions for each gene. The threshold for classifying a sample as methylation positive was set to the 95th percentile for ducts with a Masood score $\leq$ 10. It is apparent from Figure 11 that most ducts with atypical cytology were negative for methylation, but Masood scores >15 were associated with methylation positivity more frequently than lower Masood scores.

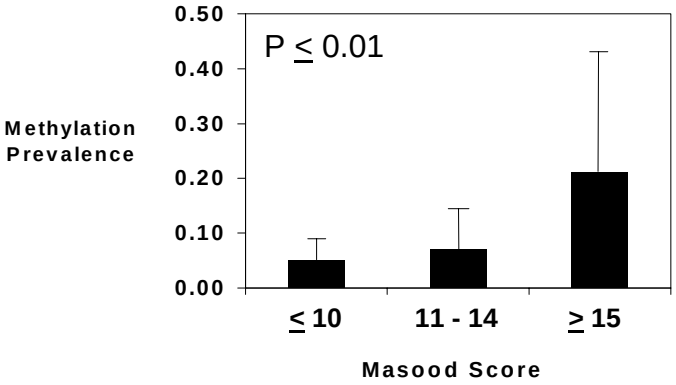


Figure 11: Ducts classified as atypical according to the Masood score were more likely to show methylation than ducts with more normal cytology.

	RAR	CD2	APC	HIN
RAS	0.41 <0.0001	0.39 <0.0001	0.37 <0.0001	0.44 <0.0001
RAR		0.26 <0.0001	0.24 <0.0001	0.41 <0.0001
CD2			0.46 <0.0001	0.27 <0.0001
APC				0.19 0.002

Table 8: Correlation between genes assessed in the same duct

Correlations Between Genes

Spearman correlation coefficients were calculated for each gene combination to determine whether methylation of one gene or another occurred independently of methylation of the other genes within a given duct. Table 8 shows that, in general, if methylation for one gene is detected in a duct, other genes are likely to be methylated as well. Although the correlation is significant for all gene combinations, the methylation status of RASSF1A most strongly predicted the methylation status of all other genes.

Correlation Between Ducts and Breasts – The Methylator Phenotype

If some patients are more prone to TSG methylation than others (i.e. a methylator phenotype), we would expect that the methylation status of one duct would predict the methylation status of other ducts in the same breast or in the same patient. Methylation scores were calculated for each duct (or breast) by averaging the methylation fractions for each gene. Only breasts with results for two or more ducts were included in this analysis, and breasts with a known cancer were excluded. The methylation status of individual ducts correlated poorly with the methylation status of other ducts from the same patient (this analysis included ducts from either breast). Correlation coefficients for Spearman pair-wise comparisons ranged from 0.104 – 0.351 for various duct combinations with P-values ranging from 0.057 – 0.586. There was better correlation when the analysis was restricted to ducts from the same breast, with correlation coefficients ranging from 0.282 – 0.744 and P-values ranging from 0.001 – 0.289 for the various duct combinations. Nevertheless, the methylation status of one breast (expressed as an average methylation score for all of the ducts from that breast) was highly correlated with the methylation status of the other breast (correlation coefficient = 0.646, P < 0.0001). This analysis suggests that TSG methylation within a given breast is a fairly generalized phenomenon likely to affect (or not affect) multiple ducts in that breast and that the global methylation status of one breast is highly correlated with that of the other breast. These observations support the existence of a methylator phenotype.

Factors Predicting Methylation

It has previously been shown that benign breast disease is associated with NAF production[32]. In our study, the prevalence of methylation positivity was similar for NAF-producing and dry ducts. Setting the threshold for classifying a duct as positive for methylation to the 95th percentile for NAF (-) ducts (5-marker average = 0.20), 5% of NAF (-) ducts were positive for methylation as compared to 1% of NAF (+) ducts (P = 0.19). Clearly, methylation is not related to NAF production.

Setting the threshold for classifying a breast as methylation positive to the 95th percentile for lower risk breasts (0.05), 5% of breasts from women <44.3 years of age (the median age for the sample) were methylation positive as compared to 11% of breasts from women ≥44.3 (P = 0.35).

Methylation of HIN-1 and RASSF1A correlated with cell yield, but the composite methylation score (mean for all markers) did not (Table 9).

Table 9: Methylation Prevalence by Gene and Cell Yield (ducts from unaffected breasts only)

	Cyclin D2	**APC**	**HIN-1**	**RASSF1A**	**RAR-β2**	**Duct Avg**
*Threshold	0.032	0.050	0.029	0.015	0.019	0.062
10 - 99	3/64 (0.047)	3/64 (0.047)	3/56 (0.054)	3/62 (0.048)	3/63 (0.048)	3/64 (0.047)
100 - 999	5/109 (0.046)	8/109 (0.073)	9/105 (0.086)	10/109 (0.092)	5/107 (0.047)	7/107 (0.065)
≥1000	5/55 (0.091)	6/55 (0.109)	13/55 (0.236)	13/54 (0.241)	4/55 (0.074)	5/55 (0.091)
P-Value	NS	NS	≤0.025	≤0.01	NS	NS

*95th percentile for the 10 – 99 cell yields. Only 11 duct had a yield ~10 cells.

Methylation as a Marker of Breast Cancer Risk

Definition of Risk Classifications

The absolute risk values calculated by the Gail model are strongly affected by age and race. 1.7% is simply the 5-year Gail risk for an average 60 year old woman; and, because breast cancer incidence is lower in African-Americans than Caucasian women, many African-Americans with the same risk factors as a "high risk" Caucasian woman will not reach the 1.7% threshold. To define a truly high risk group, we calculated a Gail Risk Index by dividing the absolute 5-year Gail risk by the general population risk for age- and race-matched women. 5-year general population risk values were obtained from SEER data[33].

Breasts were classified by risk level as follows:

- **High Risk**
  - Contralateral to a breast cancer
  - 5-year Gail risk ≥ twice age- and race-matched general population risk

- **Lower Risk**
  - 5-year Gail risk < twice age- and race-matched general population risk

- **Ipsilateral to a breast cancer**

Unsupervised Clustering

It is apparent from the unsupervised clustering plot that most of the samples did not show methylation for any of the five genes. APC and Cyclin D2 formed one gene cluster, and HIN-1

and RASSF1A a second cluster that was weakly associated with RAR-β. Three distinct methylation clusters were identified. Samples from high risk breasts, defined as those contralateral to a breast cancer or those from unaffected high risk women, were over represented in each of the methylation clusters. 15.2% of high risk samples belonged to a methylation cluster as compared to only 7.8% of lower risk samples (RR 1.95, P = 0.12, Figure 12).

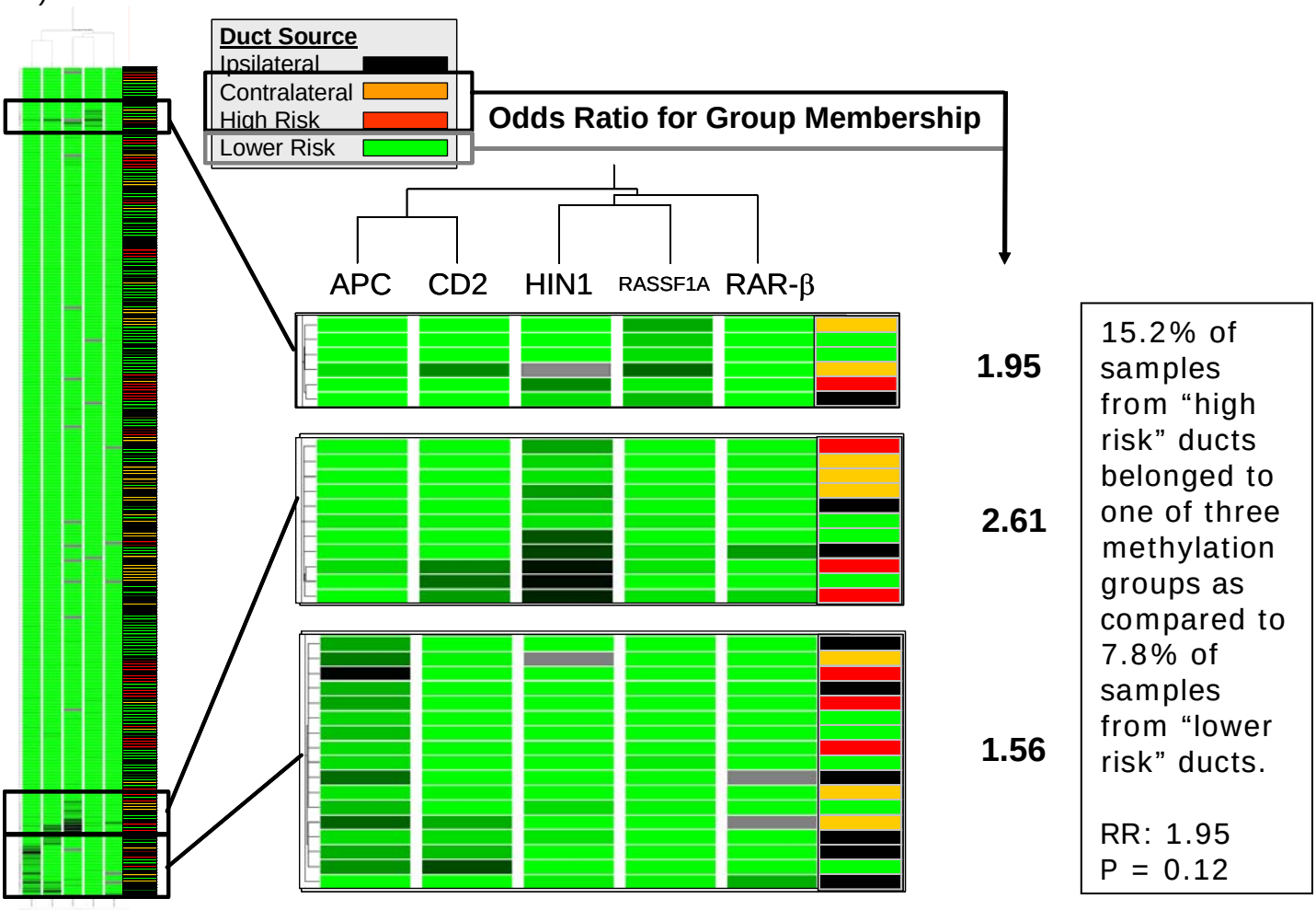


Figure 12: Unsupervised cluster analysis for all ducts and genes.

Thresholds for Calculating Methylation Prevalence

In order to calculate methylation prevalence from quantitative methylation data, a threshold must be established for scoring a particular result as positive or negative. Figure 13 shows methylation prevalence by sample source and gene at various thresholds. Values represent the mean methylation fraction from all ducts contributing to a given breast. Scoring a sample as positive if the methylation value exceeds the sensitivity of the assay (> 0.00001) results in a methylation prevalence of about

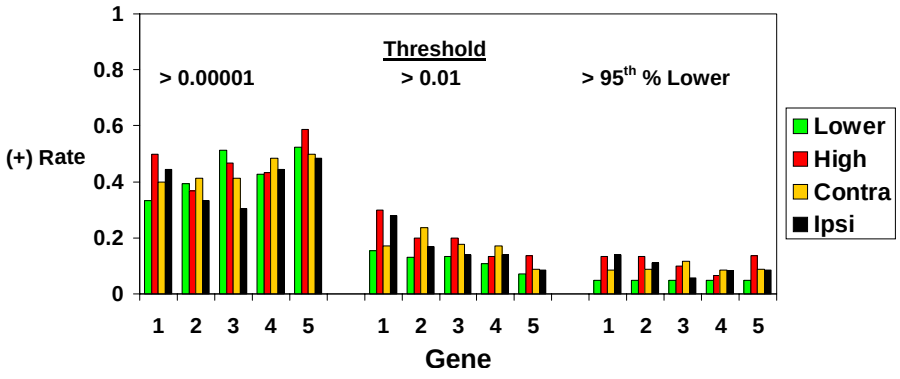


Figure 13: Methylation prevalence for each gene at selected threshold levels by risk level of the breast providing the sample. Genes: 1 – APC, 2 – HIN1, 3 – RASSF1A, 4 – Cyclin D2, 5 – RARβ2.

40% for all five genes, with little discrimination between high and lower risk breasts. If the threshold for classifying a result as positive is arbitrarily set to 0.01 (1%) methylation prevalence ranges from 10 – 20%, depending on the gene, and is generally greater for samples from high risk breasts than lower risk breasts. Similarly, setting the threshold for classifying a result as positive to the 95th percentile of the lower risk samples produces a methylation prevalence of about 10% for the high risk samples providing some discrimination between the high risk and lower risk breasts.

Methylation of Multiple Genes

About 40% of breasts showed methylation of at least one gene at the >0.01 threshold, irrespective of risk level. High risk breasts showed methylation of more genes than lower risk breasts, and breasts from unaffected high risk women were very similar to breasts ipsilateral or contralateral to a breast cancer in this regard. This suggests that TSG methylation is an early, risk-associated, event in breast carcinogenesis that does not progress until the time of overt malignant transformation. Though 12% of high risk breasts showed methylation of 3 or more genes as compared to only 4% of lower risk breasts, this difference did not reach statistical significance (P = 0.09, Figure 14).

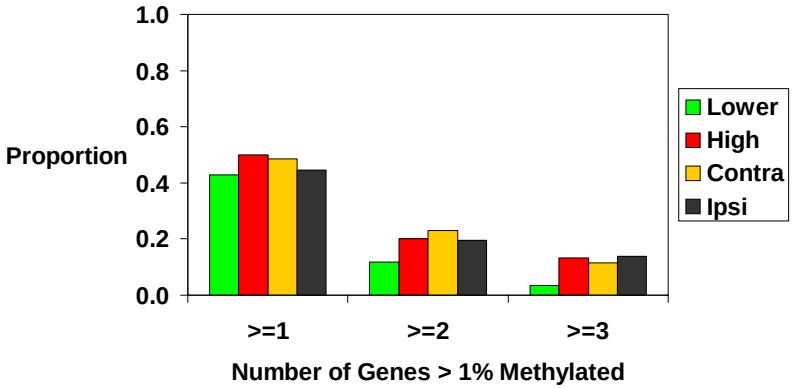


Figure 14: Methylation of multiple genes by sample source.

Methylation Prevalence by Gene

Quantitative methylation fractions were dichotomized into positive and negative using the 95th percentile value for the lower risk breasts as the threshold for classifying a sample as positive. Table 10 shows the proportion of samples that were positive for each gene by the risk classification of the breast providing the sample. In cases where more than one duct provided methylation results, the values were averaged for that breast. A composite methylation score (combined) was calculated as the average for all ducts and markers for a given breast.

Table 10: Prevalence of Methylation Positivity by Breast Class and Gene (Threshold for positive = 95th percentile for lower risk breasts)

	CD2	APC	HIN1	RAS	RAR	Combined
Threshold for Positive	**>6.8%**	**>7.3%**	**>5.9%**	**>3.7%**	**>1.2%**	**>5.0%**
Unaffected Lower Risk	0.048	0.048	0.048	0.049	0.048	0.048
Unaffected High Risk	0.067	0.133	0.133	0.100	0.138	0.133
Contralateral to Cancer	0.086	0.086	0.088	0.118	0.088	0.114
Ipsilateral to Cancer	0.083	0.139	0.111	0.056	0.086	0.194
P-Value (for distribution)	**NS**	**NS**	**NS**	**NS**	**NS**	**<0.1**

In general, methylation prevalence is lowest for the lower risk samples and greater for all other samples but the distribution was not statistically significant for any gene. The composite methylation score provided the best discrimination between risk categories. This distribution approached statistical significance ($P \leq 0.1$). It is notable that, based on the composite score,

breasts from unaffected high risk women had a similar methylation rate as breasts contralateral to a breast cancer.

Methylation rates were compared for high risk versus lower risk breasts (Figure 15). The high risk category includes breasts contralateral to a breast cancer and breasts from unaffected high risk women. The threshold for classifying a result as positive was set to the 95th percentile for the lower risk breasts. Differences were not statistically significant. (P = 0.14 for the combined , ALL, comparison).

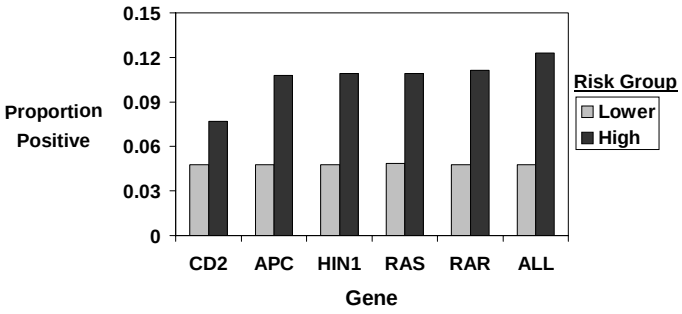


Figure 15: Methylation rates for lower risk as compared to high risk breasts. High risk includes breasts contralateral to a breast cancer and breasts from unaffected high risk women.

Prevalence of Methylation and Atypia by Risk Level

A composite methylation score was calculated for each breast based on the average methylation fraction for all genes and all ducts contributing to that breast. The threshold for classifying a breast as positive for methylation was set to the 95th percentile for the lower risk breasts. Figure 16 shows the prevalence of methylation or marked atypia by the classification of the breast providing the sample(s).

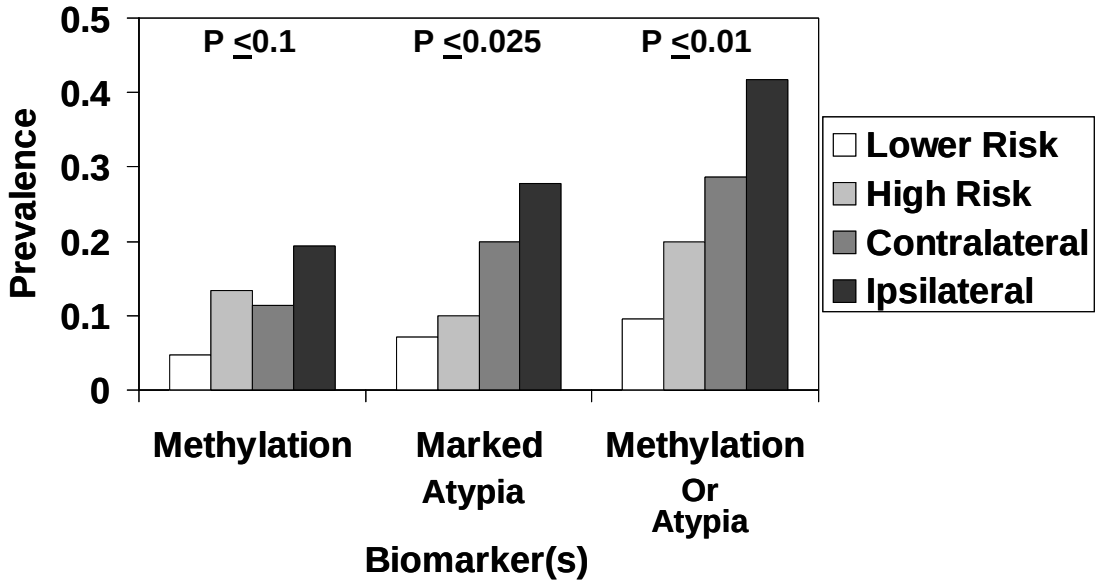


Figure 16: Biomarker prevalence by classification of the breast providing the sample.

Breasts from unaffected high risk women are affected by TSG methylation at the same rate as breasts contralateral to a breast cancer. Methylation occurs more frequently in breasts ipsilateral to a breast cancer, but the difference between ipsilateral and contralateral breasts is accounted for by the retrieval of cancer cells in 9% of lavages ipsilateral to a breast cancer. This distribution suggests that TSG methylation is an early, risk-associated change but the differences did not reach statistical significance.

Marked atypia, defined as a Masood score > 15, occurs at about the same rate in breasts from unaffected high risk and lower risk women, but occurs more frequently in breasts contralateral to a breast cancer. Comparison of the distributions for methylation positivity and marked atypia suggests that the cytological changes recognized as atypia occur subsequent to, or independent of the TSG methylation events.

Most atypical samples are negative for TSG methylation (Figure 11), suggesting that these biomarkers are largely independent. Classifying samples as methylation positive OR atypical, provides the best biomarker discrimination between samples from unaffected lower risk breasts, unaffected high risk breasts, breast contralateral to a breast cancer, and breasts ipsilateral to a breast cancer.

Our primary hypothesis was that measures of TSG methylation in duct lavage samples would discriminate high risk from lower risk breasts. The high risk category includes breasts contralateral to a breast cancer and breasts from women with a 5-year Gail risk that is ≥ twice age- and race-matched general population risk. Though there is a suggestion that methylation or the combination of methylation or marked atypia discriminates high risk from lower risk breasts, these differences were not statistically significant (Figure 17).

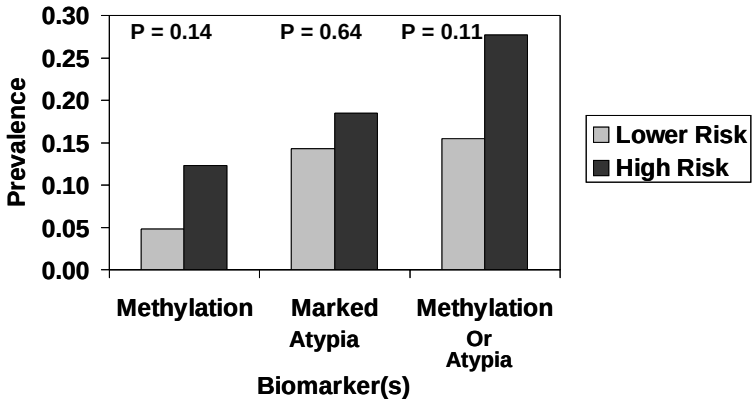


Figure 17: Biomarker prevalence by risk level of the breast providing the sample.

The prevalence of methylation positivity was 12.3% for high risk breasts as compared to 4.8% for lower risk breasts (P = 0.14). The prevalence of marked atypia was 18.5% for high risk breasts as compared to 14.3% for lower risk breasts (P = 0.64). Only 17% of the breasts with marked atypia were scored as methylation positive, suggesting that, for the most part, methylation and marked atypia are not occurring in the same breasts. The prevalence of methylation OR marked atypia was 27.7% for the high risk breasts as compared to 15.5% for the lower risk breasts (P = 0.11).

Reproducibility of Atypia and Methylation in Repeat Samples

Reproducibility of Lavage Cytology

Twenty-four patients whose initial lavage was classified as mildly or markedly atypical underwent a repeat lavage 2.3 to 16.6 months (median 8.8 months) later. At the time of the repeat lavage, every effort was made to re-lavage the same ducts that had been classified as atypical as well as any other ducts that could be cannulated. A total of 88 ducts from 48 breasts were included in the repeat lavage. Atypia was scored as reproducible for a duct if both the initial and the repeat lavage were classified as atypical for that duct. Atypia was scored as reproducible for a breast if any duct from that breast returned atypical cells on the repeat lavage even if it was not the same duct that had initially been interpreted as atypical. Table 11 summarizes the reproducibility of mild and marked atypia for ducts, breasts and patients.

Table 11: Reproducibility of Lavage Atypia by Patient, Breast and Duct

Initial Lavage	Mild Atypia					Marked Atypia					Any Atypia
Repeat Lavage	Marked	Mild	Normal or EH	ICMD	Any Atypia	Marked	Mild	Normal or EH	ICMD	Any Atypia	Any Atypia
By Patient	3/15 (20.0)	4/15 (26.7)	7/15 (46.7)	0/15 (0)	7/15 (46.7)	2/9 (22.2)	3/9 (33.3)	3/9 (33.3)	1/9 (11.1)	5/9 (55.5)	13/24 (54.2)
By Breast	4/25 (16.0)	5/25 (20.0)	12/25 (48.0)	4/25 (16.0)	9/25 (36.0)	1/9 (11.1)	2/9 (22.2)	5/9 (55.6)	1/9 (11.1)	3/9 (33.3)	12/34 (35.3)
By Duct	3/32 (9.4)	4/32 (12.5)	15/32 (46.9)	10/32 (31.3)	7/32 (21.9)	0/12 (0)	2/12 (16.7)	9/12 (75.0)	1/12 (8.3)	2/12 (16.7)	9/44 (20.5)

Only 20.5% of ducts initially classified as atypical were classified as atypical on repeat lavage. Atypia was reproduced for 35.3% of the breasts initially classified as atypical, and 54.2% of the patients who had at least one atypical duct on the initial lavage. The ICMD rate on repeat lavage was 14.7% for breasts initially classified as atypical as compared to 22.1% of all initial lavages in breasts unaffected with breast cancer (P = 0.45). Marked atypia was no more reproducible than mild atypia. More than half of the ducts initially classified as atypical returned cells that were classified as normal or epithelial hyperplasia only on repeat lavage.

Another way to evaluate reproducibility is to calculate correlation coefficients for paired data. Table 12 shows the correlation coefficients for the initial and repeat lavages, calculated using the method most appropriate for the continuous or categorical variables.

Table 12: Reproducibility of Cytology and Methylation on Repeat Sampling. Data by Breast

	Coefficient	**P-Value**
Methylation Value – Spearman Correlation		
Cyclin D2	0.56	0.001
APC	0.49	0.006
HIN-1	0.34	0.06
RASSF1A	0.16	0.38
RAR-β2	0.12	0.53
Subjective Cytology Classification – Simple kappa		
	0.084 (95% CI: -0.06 – 0.23)	
Masood Score – Pearson Correlation		
	0.18	0.22

The cytological classification, whether subjective or according to the Masood score, was poorly reproducible, while measures of Cyclin D2 and APC methylation were moderately consistent between samplings.

Summary Results by Contract Tasks

Task #1: *To assemble a panel of genes which are frequently aberrantly methylated in breast cancers and apply it to the study of risk assessment.*

Methylation of Cyclin D2, APC, HIN1, RASSF1A or RAR-β1 is detected in 89% of breast cancers (Figure 10). This gene panel is judged to be highly relevant to breast cancer.

Task #2: *To determine whether the methylation profile of breast epithelial cells obtained by nipple duct lavage correlates with the degree of breast cancer risk as estimated by computerized modeling and cytological abnormalities as determined by the Masood score.*

TSG Methylation occurs at a similar frequency in breasts of unaffected high risk women and breasts contralateral to a breast cancer (about 12%), but at a lower frequency (5%) in breasts from unaffected lower risk women. This difference was not statistically significant (Figure 17).

Lavage samples exhibiting marked cytological atypia, as classified by the Masood score, are more likely to be positive for TSG methylation than samples with more normal cytology, but most atypical samples are negative for methylation (Figure 11).

The distribution of TSG methylation or cytological atypia according to the risk level of the breast providing the NDL sample suggests that these markers have some value for risk stratification (Figure 16).

Task #3: *To determine and compare the methylation profile of NDL cells obtained from women at different degrees of breast cancer development.*

High risk breasts were more frequently methylated for each of the five genes than lower risk breasts (Table 10, Figure 15). The methylation prevalence ratio for high risk versus lower risk breasts was 2.3 for APC, HIN1, RASSF1A, and RAR-β2, but only 1.6 for Cyclin D2. These prevalence ratios were not statistically significant. Methylation of one gene is frequently associated with methylation of other genes (Table 8). We were not able to identify a specific combination of genes that was more associated with high risk samples than other combinations.

Key Research Accomplishments

- A quantitative methylation-specific real time PCR was optimized and validated for Cyclin D2, APC, HIN1, RASSF1A, and RAR-β2.
- Neither NAF production nor a 5-year Gail risk >1.7% predicts lavage atypia.
- Marked cytological atypia is identified more frequently in breasts from cancer patients than those from unaffected women.
- Marked cytological atypia tracks well with the risk level of the breast providing the NDL sample.
- Assessment of TSG methylation patterns does not improve the diagnostic accuracy of NDL for the detection of breast cancer.
- The methylation status of one duct predicts the methylation status of other ducts in the same breast.
- TSG methylation is identified more frequently in high risk breasts than lower risk breasts, but this difference was not statistically significant with our sample size.
- The combination of TSG methylation or marked atypia provides better risk stratification than either marker alone.
- Measures of TSG methylation are more reproducible over time than cytological assessments.

Reportable Outcomes

National Presentations

Euhus DM, Ashfaq R, Milchgrub S, Naftalis E, Leitch AM, Virmani A. Comparison of Nipple Duct Lavage & Random Fine Needle Aspiration Biopsy for the Detection of Atypical Breast Epithelium. Soc Surg Oncol, Los Angeles, CA, March 2003.

Cler LR, Ashfaq R, Naftalis E, Leitch AM, Hoover S, Euhus DM. Cytological atypia diagnosed by nipple duct lavage: reproducibility and MRI findings. Association of Academic Surgery, Sacramento, CA, November 2003.

Johnson A, Ashfaq R, Cler L, Naftalis E, Leitch AM, Hoover S, Euhus DM. Cytological Atypia Diagnosed by Nipple Duct Lavage: Reproducibility and MRI Findings. Soc Surg Oncology, New York, NY, March 2004.

Johnson A, Ashfaq R, Naftalis E, Leitch AM, Hoover S, Euhus DM. Patient and Duct Selection for Nipple Duct Lavage: Challenging the Current Paradigm. American Society of Breast Surgeons, Las Vegas, NV, April 2004.

Euhus D, Johnson Maddux A, Cler L, Naftalis E, Leitch A, Ashfaq R. Patient And Duct Selection For Nipple Duct Lavage. 4th International Santa Barbara Symposium: The Intraductal Approach to Breast Cancer, Santa Barbara, CA, 2005.

Bushnaq Z, Ashfaq R, Leitch M, Euhus D. Patient Variables Predicting Atypical Cytology by Nipple Duct Lavage. American Society of Breast Diseases, Las Vegas, NV, 2006.

Euhus DM, Ashdaq R, Bu D, Leitch AM, Lewis CM. Assessment of tumor suppressor gene methylation for breast cancer risk screening. American Society of Clinical Oncology (Oral Presentation),Atlanta, GA, 2006.

Published Manuscripts

Johnson Maddux A, Ashfaq R, Naftalis E, Leitch A, Hoover S, Euhus D. Patient and duct selection for nipple duct lavage. Am J Surg 2004;188:390-394.

Johnson-Maddux A, Ashfaq R, Cler L, Naftalis E, Leitch A, Hoover S, Euhus D. Reproducibility of cytologic atypia in repeat nipple duct lavage. Cancer 2005;103:1129-1136.

Cler L, Bu D, Lewis C, Euhus, D. A Comparison of Five Methods for Extracting DNA from Paucicellular Clinical Samples. Molecular and Cellular Probes, 2006;20:191-196.

Euhus D, Ashfaq R, Lewis C, Bu D. Assessment of Tumor Suppressor Gene Methylation in Benign Breast Epithelial Cells Obtained by Nipple Duct Lavage. In preparation.

Conclusions

The breast is composed of 12 – 15 independent ductal systems making duct lavage unsuitable for the detection of focal changes like carcinoma. TSG methylation appears to be a field change that occurs early in breast carcinogenesis and that does not progress until the time of overt malignant transformation. Because TSG methylation is identified in a reasonable proportion of at risk breasts, is detected more frequently in high risk as compared to lower risk breasts, is reasonably reproducible, and can be serially measured in samples obtained by minimally invasive techniques, it may be a suitable marker for risk stratification and may have value as a surrogate endpoint biomarker in phase II prevention trials. Because duct lavage is expensive, time consuming, uncomfortable, and associated with a high insufficient sample rate, it may not be the best approach for obtaining breast epithelial cells for biomarker assays.

References

1. Gail MH, Brinton LA, Byar DP, et al. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *J Natl Cancer Inst.* 1989;81:1879-1886.
2. Costantino JP, Gail MH, Pee D, et al. Validation studies for models projecting the risk of invasive and total breast cancer incidence. *J Natl Cancer Inst.* 1999;91:1541-1548.
3. Rockhill B, Spiegelman D, Byrne C, Hunter DJ, Colditz GA. Validation of the Gail et al. model of breast cancer risk predicition and implications for chemoprevention. *J Natl Cancer Inst.* 2001;93:358-366.
4. Kelloff GJ, Boone CW, Crowell JA, et al. Risk biomarkers and current strategies for cancer chemoprevention. *Journal of Cellular Biochemistry.* 1996;25:1–14.
5. Boone CW, Bacus JW, Bacus JV, Steele VE, Kelloff GJ. Properties of intraepithelial neoplasia relevant to cancer chemoprevention and to the development of surrogate end points for clinical trials. *Proceedings of the Society for Experimental Biology and Medicine.* 1997;216:151–165.

6. Tlsty TD, Crawford YG, Holst CR, et al. Genetic and epigenetic changes in mammary epithelial cells may mimic early events in carcinogenesis. *Journal of Mammary Gland Biology & Neoplasia.* 2004;9(3):263-274.
7. Parrella P, Poeta ML, Gallo AP, et al. Nonrandom distribution of aberrant promoter methylation of cancer-related genes in sporadic breast tumors. *Clinical Cancer Research.* 2004;10(16):5349-5354.
8. Lehmann U, Langer F, Feist H, Glockner S, Hasemeier B, Kreipe H. Quantitative assessment of promoter hypermethylation during breast cancer development. *American Journal of Pathology.* 2002;160(2):605-612.
9. Umbricht CB, Evron E, Gabrielson E, Ferguson A, Marks J, Sukumar S. Hypermethylation of 14-3-3 sigma (stratifin) is an early event in breast cancer. *Oncogene.* 2001;20:3348–3353.
10. Jeronimo C, Costa I, Martins MC, et al. Detection of gene promoter hypermethylation in fine needle washings from breast lesions. *Clinical Cancer Research.* 2003;9(9):3413-3417.
11. Li Z, Meng ZH, Chandrasekaran R, et al. Biallelic inactivation of the thyroid hormone receptor beta1 gene in early stage breast cancer. *Cancer Res.* 2002;62:1939–1943.
12. Fackler MJ, McVeigh M, Evron E, et al. DNA methylation of RASSF1A, HIN-1, RAR-beta, Cyclin D2 and Twist in in situ and invasive lobular breast carcinoma. *International Journal of Cancer.* 2003;107(6):970-975.
13. Wrensch MR, Petrakis NL, Miike R, et al. Breast cancer risk in women with abnormal cytology in nipple aspirates of breast fluid. *J Natl Cancer Inst.* 2001;93:1791-1798.
14. Fabian CJ, Kimler BF, Zalles CM, al e. Short-term breast cancer prediction by random periareolar fine-needle aspiration cytology and the Gail risk model. *J Natl Cancer Inst.* 2000;92:1217 –1227.
15. Lewis C, Cler L, Bu D, et al. Promoter Hypermethylation in Benign Breast Epithelium in Relation to Predicted Breast Cancer Risk. *Clin Cancer Res.* 2005;11(1):166-172.
16. Fackler MJ, McVeigh M, Mehrotra J, et al. Quantitative Multiplex Methylation-Specific PCR Assay for the Detection of Promoter Hypermethylation in Multiple Genes in Breast Cancer. *Cancer Res.* 2004;64:4442-4452.
17. Lewis CM, Cler LR, Bu D-w, et al. Promoter Hypermethylation in Benign Breast Epithelium in Relation to Predicted Breast Cancer Risk. *Clin Cancer Res.* 2005;11:166-172.
18. Evron E, Umbricht CB, Korz D, et al. Loss of Cyclin D2 Expression in the Majority of Breast Cancers Is Associated with Promoter Hypermethylation. *Cancer Res.* 2001;61:2782-2787.
19. Zochbauer-Muller S, Fong KM, Virmani AK, Geradts J, Gazdar AF, Minna JD. Aberrant promoter methylation of multiple genes in non-small cell lung cancers. *Cancer Research.* 2001;61:249-255.
20. Virmani AK, Rathi A, Sathyanarayana UG, et al. Aberrant methylation of the adenomatous polyposis coli (APC) gene promoter 1A in breast and lung carcinomas. *Clinical Cancer Research.* 2001;7:1998-2004.
21. Krop IE, Sgroi D, Porter DA, et al. HIN-1, a putative cytokine highly expressed in normal but not cancerous mammary epithelial cells. *Proc Natl Acad Sci USA.* 2001;98(17):9796-9801.
22. Burbee DG, Forgacs E, Zochbauer-Muller S, et al. Epigenetic inactivation of RASSF1A in lung and breast cancers and malignant phenotype suppression. *Journal of the National Cancer Institute.* 2001;93:691-699.
23. Widschwendter M, Berger J, Hermann M, et al. Methylation and silencing of the retinoic acid receptor-beta2 gene in breast cancer. *Journal of the National Cancer Institute.* 2001;92:826-832.
24. Clark SJ, Harrison J, Paul CL, Frommer M. High sensitivity mapping of methylated cytosines. *Nucleic Acids Res.* 1994;22:2990-2997.
25. Bais AJ, Gardner AE, McKenzie OL, Callen DF, Sutherland GR, Kremmidiotis G. Aberrant CBFA2T3B gene promoter methylation in breast tumors. *Mol Cancer.* 2004;3:22-38.

26. Masood S, Frykberg ER, McLellan GL, Scalapino MC, Mitchum DG, Bullard JB. Prospective evaluation of radiologically directed fine-needle aspiration biopsy of nonpalpable breast lesions. *Cancer.* 1990;66:1480–1487.
27. Dooley WC, Ljung B, Veronesi U, et al. Ductal lavage for the detection of cellular atypia in women at high risk for breast cancer. *J Natl Cancer Inst.* 2001;93:1624-1632.
28. Morrow M, Vogel V, Ljung BM, O'Shaughnessy JA. Evaluation and management of the woman with an abnormal ductal lavage. *J Am Coll Surg.* 2002;194:648-656.
29. O'Shaughnessy JA, Ljung BM, Dooley WC, et al. Ductal lavage and the clinical management of women at high risk for breast carcinoma. *Cancer.* 2002;94(2):292-298.
30. Maddux AJ, Ashfaq R, Naftalis E, Leitch AM, Hoover S, Euhus DM. Patient and Duct Selection for Nipple Duct Lavage. *Am J Surg.* 2004;188(4):390-394.
31. Khan SA, Wiley EL, Rodriguez N, et al. Ductal Lavage Findings in Women With Known Breast Cancer Undergoing Mastectomy. *J Natl Cancer Inst.* 2004;96:1510-1517.
32. Petrakis NL, Ernster VL, Sacks ST, et al. Epidemiology of breast fluid secretion...cancer risk factors and cerumen type. *J Natl Cancer Inst.* 1981;67:277-284.
33. SEER Cancer Statistics Review, 1975-2003. *National Cancer Institute*. Available from: http://seer.cancer.gov/csr/1975_2003/,.

Excerpta Medica

The American Journal of Surgery

The American Journal of Surgery 188 (2004) 390–394

Scientific paper

Patient and duct selection for nipple duct lavage

April Johnson Maddux, M.D.[a], Raheela Ashfaq, M.D.[b], Elizabeth Naftalis, M.D., F.A.C.S.[a], Ann Marilyn Leitch, M.D., F.A.C.S.[a], Susan Hoover, M.D., F.A.C.S.[a] David Euhus, M.D., F.A.C.S.[a,*]

[a]*Department of Surgery, University of Texas Southwestern Medical Center at Dallas, 5323 Harry Hines Blvd., Dallas, TX 75390-9155, USA*
[b]*Department of Pathology, University of Texas Southwestern Medical Center at Dallas, Dallas, TX, USA*

Manuscript received May 19, 2004; revised manuscript June 6, 2004

Presented at the Fifth Annual Meeting of the American Society of Breast Surgeons, March 31–April 4, 2004, Las Vegas, Nevada

Abstract

Background: Nipple ductal lavage (NDL) is a new minimally invasive procedure with the potential to help identify women who could benefit from breast cancer risk intervention. NDL is currently encouraged for women with fluid-producing ducts and a 5-year Gail risk ≥1.7%. The purpose of this study was to evaluate the atypia rate by NDL in fluid-producing ducts compared with non–fluid-producing ducts and the atypia rate in high-risk verses low-risk patients to determine if current recommendations are supported.
Methods: Fifty-nine women were studied with NDL. The 226 ducts lavaged included all fluid-producing ducts (n = 136) and any dry ducts we could cannulate (n = 90). Breast cancer risk was calculated using mathematic models.
Results: There were 26 (44%) women with a 5-year Gail risk ≥1.7% and 33 (56%) with a 5-year Gail risk <1.7%. Cytologic atypia was diagnosed in 20 of 59 (34%) of patients. The atypia rate was similar for women with a 5-year Gail risk ≥1.7% (9 of 26 or 35%) compared with lower-risk women (11 of 33 or 33%, $P = 1.0$) and for fluid-producing ducts (26 of 136 or 19%) compared with dry ducts (14 of 90 or 15%, $P = 0.61$). No significant differences were found when the atypia was categorized as mild versus marked. Of note, the insufficient sample rate was higher for dry ducts (33%) compared with fluid-producing duct (22%, $P = 0.07$).
Conclusions: If NDL results are found to correlate with breast cancer incidence, it will be important to apply the test in a way that maximizes sensitivity for the detection of atypia in a screened population. We were unable to identify patient or duct characteristics that predict NDL atypia rates.

Keywords: Atypia; Cancer risk; Nipple duct lavage

The National Surgical Adjuvant Breast and Bowel Project (NSABP)-sponsored Breast Cancer Prevention Trial (BCPT-P1) reported that 5 years of tamoxifen therapy decreased the incidence of breast cancer by nearly 50% in women at increased risk for the disease [1]. For the purposes of this trial, increased risk was defined as a ≥1.7% probability of developing breast cancer during 5 years as calculated by the Gail model [2]. The ≥1.7% five-year Gail risk has been accepted by the Food and Drug Administration as providing a reasonable margin of benefit for tamoxifen therapy when balanced against the risks of thromboembolic complications and endometrial cancer. However, most women with a calculated 5-year breast cancer risk ≥1.7% refuse tamoxifen therapy. There is evidence, however, that women at higher risk levels, who are likely to enjoy greater benefits from tamoxifen, are more likely to accept this intervention. In this regard, women with previous breast biopsy specimens showing atypical ductal hyperplasia have a 3- to 5-fold greater risk of breast cancer than women with breast biopsy specimens showing nonproliferative fibrocystic changes only [3]. In addition, the NSABP BCPT-P1 trial recorded an 86% decrease in breast cancer incidence for these women. Based on these data, it is reasonable to direct chemoprevention resources toward women with atypical hyperplasia.

Previously, atypical ductal hyperplasia was only diagnosed incidentally when a palpable or mammographic abnormality was assessed by surgical biopsy. There has been

* Corresponding author. Tel.: +1-214-648-6467; fax: +1-214-648-7965
E-mail address: david.euhus@utsouthwestern.edu

0002-9610/04/$ – see front matter
doi:10.1016/j.amjsurg.2004.06.021

an interest, however, in developing more widely applicable, less-invasive approaches for identifying women with atypical breast epithelium. Wrensch et al [4,5] followed-up 2300 women for 12.7 years and reported that cytologic atypia in nipple aspirate fluid was associated with a relative risk for breast cancer of 4.9 and that the combination of cellular atypia with a family history of breast cancer was associated with a relative risk of 18. In the most recently reported follow-up, however, atypical cells in nipple aspirate fluid were associated with a relative risk for breast cancer of only 2.8 [6]. Fabian et al [7] found that high-risk women with atypical cells diagnosed by random fine-needle aspiration biopsy of the breast were 5 times more like to develop breast cancer than women without atypical cells.

Nipple duct lavage (NDL) has been proposed as a minimally invasive technique for obtaining breast epithelium for cytologic assessment. In a multicenter trial, Dooley et al [8] compared NDL with nipple duct aspiration (NDA) alone in 507 women and found that NDL was associated with a much lower insufficient sample rate than NDA (29% vs. 73%) because they retrieved an average of 13,500 epithelial cells compared with only 120 cells by NDA . Of note, the atypia rate for the increased risk women evaluated in this trial was 23% by NDL and 9% by NDA. It is currently unclear, however, whether atypia diagnosed by NDL confers the same risk of breast cancer as atypical hyperplasia diagnosed by surgical breast biopsy. Nevertheless, the test has been promoted as a method for breast cancer risk stratification and is currently recommended for women with a 5-year Gail risk ≥1.7% who would consider tamoxifen if they were found to have atypical cells. Because it is impractical to lavage each of the 6 to 16 duct orifices in each breast, it has been suggested that only fluid-producing ducts be lavaged. This study was designed to determine whether fluid-producing ducts are more likely to return cytologically atypical cells than dry ducts and to determine whether women with a 5-year Gail risk ≥1.7% are more likely to have atypia than lower-risk women.

Methods

Eligibility criteria

The Institutional Review Board at the University of Texas Southwestern Medical Center at Dallas approved this study, and written informed consent as well as Health Insurance Portability and Accountability Act authorizations were documented for each patient. Patients were recruited from the Mary L. Brown Breast Cancer Genetics and Risk Assessment Clinic in the University of Texas Southwestern Center for Breast Care. Comprehensive risk factor information was collected for each patient, and breast cancer risk calculated using the models of Gail, Claus, Bodian and BRCAPRO using software we developed (Breast C.A.R.E.). Certain components of this software are generally available in the CancerGene package we distribute [9]. All patients >18 years presenting for comprehensive breast cancer risk assessment were offered NDL on this protocol regardless of calculated risk level. Exclusion criteria included previous invasive breast cancer of any type; ductal carcinoma in situ or previous lobular carcinoma in situ treated by mastectomy; presence of a histologically undefined palpable or mammographic breast lesion suspicious for malignancy; bilateral prophylactic mastectomy; participation in a cancer prevention study (NSABP Protocol P-1 subjects who received placebo were eligible); any previous breast irradiation; any systemic chemotherapy in the past; performance status that restricted normal activity for a significant portion of each day; current use of androgens, luteinizing hormone–releasing hormone analogs, prolactin inhibitors, antiandrogens, or steroids (women who discontinue these drugs at least 3 months before duct lavage were eligible); any use of tamoxifen, raloxifene, or other selective estrogen-receptor modulator therapy; current use of coumadin; pregnant or lactating (within 6 month); presence of saline or silicone breast implants; or active bleeding disorder.

NDL procedure

EMLA cream (AstraZeneca, London, United Kingdom) was applied to the nipples and sealed with an occlusive plastic patch 2 hours before the procedure. The nipple area was dekeratinized by scrubbing with a mild abrasive gel. Breast massage was performed by the patient initially and then by the operator in an attempt to elicit nipple duct discharge. A nipple duct aspirator was used to encourage discharge; however, it was found that manual expression of fluid was generally more successful. An attempt was made to cannulate all fluid-producing ducts using a tapered dilator coated with 2% lidocaine jelly. If this was successful, a duct lavage catheter (Cytyk Health Corp., Boxborough, Massachusetts) was inserted, and a total of 10 mL physiologically buffered saline instilled and aspirated in 0.5-mL increments. When all fluid-producing ducts had been lavaged, an attempt was made to cannulate and lavage at least 1 non–fluid-producing duct in each breast.

Cytologic assessment

Lavage effluents were collected separately for each duct in 30 mL CytoLyt solution (Cytyk Health). Cytology slides were prepared using the thin-prep method, stained using the Papanicolaou technique, and then evaluated by a breast cytologist (R.A.). Cellularity was estimated for each sample as no cells or 1 to 10, 11 to 99, 100 to 999, or ≥1000 cells. A score (1 to 4) was assigned for each of 10 cytologic features: cellular arrangement, cell pleomorphism, myoepithelial cells, anisonucleosis, nucleoli, chromatin clumping, nuclear diameter, mitoses, nuclear molding, and cellular polarity. A composite score was calculated as the sum of the component scores. The cytologist was also asked to subjectively classify the sample according to all cell patterns

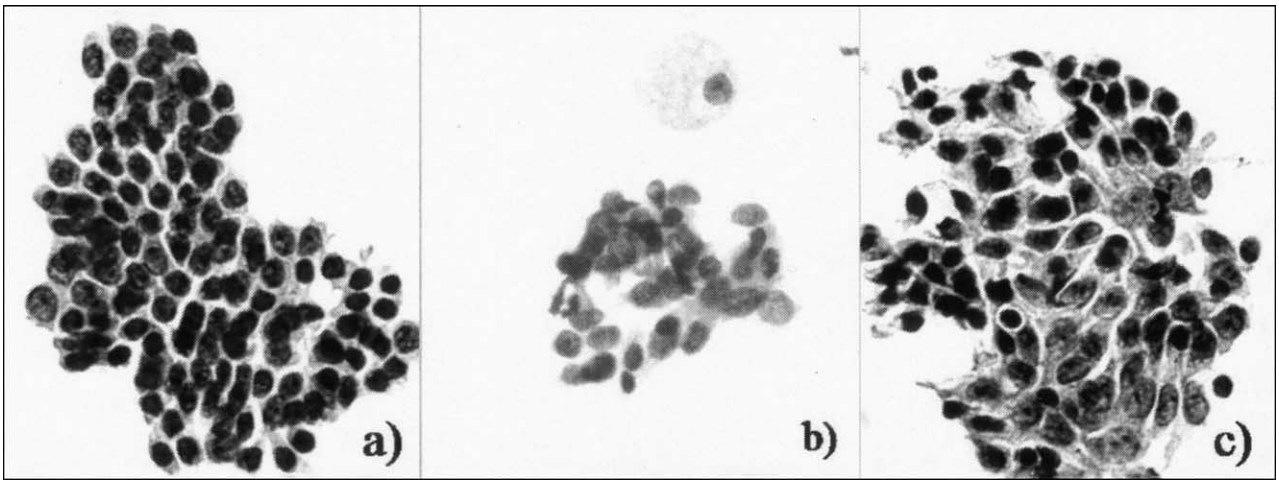


Fig. 1. Lavage cytology showing (**a**) normal cells, (**b**) mild atypia, and (**c**) marked atypia.

observed: normal, apocrine metaplasia, typical hyperplasia, mild atypia, marked atypia, or suspicious for cancer. Examples of normal cells, cells with mild atypia, and cells with marked atypia are shown in Fig. 1.

Results

Patients

NDL was performed for 59 asymptomatic women whose risk of breast cancer had been defined using mathematic models. Most of the patients were white (95%), and most were premenopausal (56%). The median age was 43 years. Twenty-six (44%) had a 5-year Gail risk ≥1.7%, and 33 (56%) had a 5-year Gail risk <1.7% (Table 1). A total of 226 ducts were lavaged. Of these, 136 were fluid producing, and 90 were dry. The insufficient sample rate was 30 of 136 (22.1%) for the fluid-producing ducts and 30 of 90 (33.3%) for the dry ducts ($P = 0.07$).

Table 1
Patient demographics

No. of patients	59
No. age in years (%)	
21–30	4 (7)
31–40	19 (32)
41–50	25 (42)
51–60	8 (14)
61–70	3 (5)
No. race (%)	
White	56 (95)
Hispanic	2 (3)
Asian	1 (2)
African American	0
No. menopausal status (%)	
Premenopausal	35 (59)
Perimenopausal	6 (10)
Postmenopausal	18 (31)
No. 5-year Gail risk	
≥1.7%	26 (44)
<1.7%	33 (56)

Atypia rate by Gail risk calculation

Overall, atypia was diagnosed in 20 of 59 (34%), mild atypia in 13 of 59 (22%), and marked atypia in 7 of 59 (12%) patients. The atypia rate was similar for women with a 5-year Gail risk ≥1.7% compared with women having a 5-year Gail risk <1.7% (9 of 26 or 35% vs. 11 of 33 or 33%, respectively, $P = 1.0$). Marked atypia was more common in women with a 5-year Gail risk ≥1.7% (4 of 26, 15%) than in women with a 5-year Gail risk <1.7% (3 of 33 9%), but this result was not statistically significantly ($P = 0.73$, Table 2).

Atypia rate by fluid-producing status of ducts

Overall, atypia was diagnosed in 40 of 226 (18%) ducts, mild atypia in 28 of 226 (12%), and marked atypia in 12 of 226 (5%). The atypia rate was similar for fluid-producing (26 of 136 or 19%) and dry ducts (14 of 90 or 15%, $P = 0.61$) with no differences noted for mild or marked atypia (Table 3).

Atypia rate by fluid-producing status and calculated Gail risk

It is currently recommended that only patients with a 5-year Gail risk ≥1.7% and fluid-producing ducts undergo

Table 2
Atypia rate by calculated Gail risk

Gail risk	All atypia	Mild atypia	Marked atypia
No. all patients (%)	20/59 (34)	13/59 (22)	7/59 (12)
No. 5-year Gail risk (%)			
≥1.7%	9/26 (35)*	5/26 (19)†	4/26 (15)‡
<1.7%	11/33 (33)*	8/33 (24)†	3/33 (9)‡

* $P = 1.00$; †$P = 0.89$; ‡$P = 0.73$.

Table 3
Atypia rate by fluid-producing status of ducts

Duct status	Any atypia	Mild atypia	Marked atypia
No. all ducts (%)	40/226 (18)	28/226 (12)	12/226 (5)
No. fluid producing (%)	26/136 (19)*	18/136 (13)†	8/136 (6)‡
No. dry ducts (%)	14/90 (15)*	10/90 (11)†	4/90 (4)‡

* $P = 0.61$; †$P = 0.80$; ‡$P = 0.88$.

ductal lavage for additional risk stratification. We calculated atypia rates for patients with a 5-year Gail risk ≥1.7% and <1.7% considering results only for fluid-producing or only for non–fluid-producing ducts. The atypia rate for patients with a 5-year Gail risk ≥1.7% considering only fluid-producing ducts was 7 of 22 (32%). The atypia rate for patients with a 5-year Gail risk <1.7% considering only non–fluid-producing ducts was 8 of 27 (30%, $P = 1.00$). Of note, although atypia rates for dry and fluid-producing ducts among patients with a 5-year Gail risk <1.7% were similar (25% vs. 30%, $P = 0.93$), the atypia rate was higher for fluid-producing than dry ducts for patients with a 5-year Gail risk ≥1.7% (32% vs. 11%, Fig. 2), but this result did not reach statistical significance ($P = 0.09$).

Comments

NDL is currently proposed as a minimally invasive approach for identifying atypical breast epithelial cells for the purpose of individualized breast cancer risk stratification. If NDL results are shown to correlate with breast cancer incidence, it will be important to apply the test in a way that maximizes its sensitivity for detection of atypical cells in the screened population. Current recommendations, however, limit the test to women with fluid-producing ducts and a 5-year Gail risk ≥1.7%. Both of these criteria are reasonably challenged based on previously published studies.

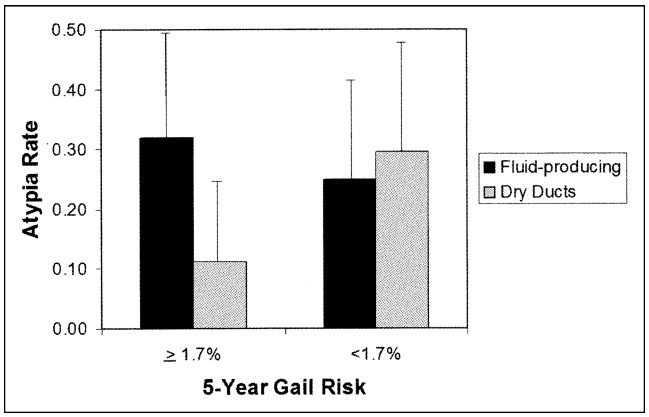


Fig. 2. Atypia rates by calculated Gail risk and fluid-producing status of the ducts. Atypia rates were similar for all categories except for fluid-producing versus dry ducts among women with a calculated 5-year Gail risk ≥1.7% (32% vs. 11%, $P = 0.09$).

First, data from the Nurses Health Study demonstrated that during a 5-year period, 753 of 54,844 women with a 5-year Gail risk <1.7% developed breast cancer compared with 601 of 27,225 women with a 5-year Gail risk ≥1.7% [10]. That is, most of the breast cancers observed in this cohort (55%) occurred in women who would not have been considered eligible for ductal lavage. Second, the etiology of expressible nipple duct discharge is multifactorial (e.g., duct ectasia, apocrine metaplasia, papilloma), and the presence of expressible nipple duct fluid does not reliably distinguish patients with benign breast disease from control patients [11]. Based on these observations, it is reasonable to ask whether the atypia rate, as measured by NDL, is higher for patients with 5-year Gail risks ≥1.7% than for lower risk women or for fluid-producing ducts compared with dry ducts. Our data suggested that these criteria do not identify women that are more likely to have atypical cells diagnosed by NDL.

The most obvious limitation of this study was the small sample size and the possibility of a type II error. It should be noted, however, that with respect to atypia rates in fluid-producing versus dry ducts, with α set at 0.05, our study had a power of 0.979 to recognize a 20% difference (30% vs. 10%). With respect to atypia rates in women with a 5-year Gail risk ≥1.7% versus <1.7%, our power was only 0.486 to recognize a similar difference. Even if a larger study were to demonstrate a statistically significant difference between atypia rates in high- and low-risk women, it is unlikely that the difference would be clinically significant in the context of population screening.

The difference in atypia rates between fluid-producing and dry ducts among the increased risk women (Fig. 2) is intriguing although not statistically significant. It is possible that there are important biologic differences between the epithelial cells retrieved from women at different risk levels that cannot be recognized under the microscope. If this is the case, limiting NDL to women determined to be at increased risk based on epidemiologic models would be reasonable. This can only be known, however, as the results of several on going studies evaluating biomarker expression in lavage cells become available.

NDL is an intriguing technology for sampling breast epithelial cells from selected nipple ducts. It remains to be determined, however, how best to select the duct(s) to lavage, how best to select the patients to lavage, and, most important, whether atypia diagnosed by NDL predicts an increased risk for breast cancer. Clearly, additional study is required.

References

[1] Fisher B, Constantino 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: 1371–88.

[2] Gail MH, Brinton LA, Byar DP, et al. Projecting individualized

probabilities of developing breast cancer for white females who are being examined annually. J Natl Cancer Inst 1989;81:1879–86.
[3] Dupont WD, Parl FF, Hartman WH, et al. Breast Cancer risk associated with proliferative breast disease and atypical hyperplasia. Cancer 1993;71:1258–65.
[4] Wrensch MR, Petrakis NL, King EB, et al. Breast cancer incidence in women with abnormal cytology in nipple aspirates of breast fluid. Am J Epidemiol 1992;135:130–41.
[5] Wrensch MR, Petrakis NL, King EB, et al. Breast cancer risk associated with abnormal cytology in nipple aspirates of breast fluid and prior history of breast biopsy. Am J Epidemiol 1993;137:829–33.
[6] Wrensch MR, Petrakis NL, Miike R, et al. Breast cancer risk in women with abnormal cytology in nipple aspirates of breast fluid. J Natl Cancer Inst 2001;93:1791–8.
[7] Fabian CJ, Kimler BF, Zalles CM, et al. Short-term breast cancer prediction by random periareolar fine-needle aspiration cytology and the Gail risk model. J Natl Cancer Inst 2000;92:1217–27.
[8] Dooley WC, Ljung B, Veronesi U, et al. Ductal lavage for the detection of cellular atypia in women at high risk for breast cancer. J Natl Cancer Inst 2001;93:1624–32.
[9] Euhus DM. Understanding mathematical models for breast cancer risk assessment and counseling. Breast J 2001;7:224–32.
[10] Rockhill B, Spiegelman D, Byrne C, et al. Validation of the Gail et al. model of breast cancer risk prediction and implications for chemoprevention. J Natl Cancer Inst 2001;93:358–66.
[11] Wynder EL, Lahti H, Laakso K, et al. Nipple aspirates of breast fluid and the epidemiology of breast diseases. Cancer 1985;56:1473–8.

Reproducibility of Cytologic Atypia in Repeat Nipple Duct Lavage

April Johnson-Maddux, M.D.[1]
Raheela Ashfaq, M.D.[2]
Leslie Cler, B.S.[1]
Elizabeth Naftalis, M.D.[1]
Ann Marilyn Leitch, M.D.[1]
Susan Hoover, M.D.[1]
David M. Euhus, M.D.[1]

[1] Department of Surgery, University of Texas Southwestern Medical Center, Dallas, Texas.

[2] Department of Pathology, University of Texas Southwestern Medical Center, Dallas, Texas.

Supported by a grant from the Department of Defense Breast Cancer Research Program (DAMD17-01-1-0421).

Address for reprints: David M. Euhus, M.D., Division of Surgical Oncology, E6.222, 5323 Harry Hines Boulevard, Dallas, TX 75390-9155; Fax: (214) 648-7965; E-mail: david.euhus@utsouthwestern.edu

Received August 4, 2004; revision received October 19, 2004; accepted October 28, 2004.

BACKGROUND. It is believed that atypical cells identified by nipple duct lavage (NDL) indicate an increased risk for breast carcinoma similar to atypical ductal hyperplasia diagnosed by tissue biopsy, but many basic performance characteristics of NDL currently are undefined.

METHODS. NDL was performed in 108 patients unselected for breast carcinoma risk and then was repeated after 2–14 months (median, 8 months) if the initial lavage was classified as atypical. Breast magnetic resonance images (MRIs) were obtained from a subset of patients who had atypical lavage results.

RESULTS. Marked atypia was diagnosed in 22% of 36 breasts with an incident carcinoma compared with 7% of 172 unaffected breasts ($P = 0.01$). After excluding breasts with an incident carcinoma, there were 32 patients (30%) with either mild or marked atypia. The lavage was repeated in 23 of these women, and the second lavage was classified as atypical in 48%. Neither marked atypia on the initial lavage nor a 5-year Gail risk $\geq$ 1.7% predicted atypia on repeat lavage, but there was a trend for improved reproducibility when the atypia initially was diagnosed in a fluid-producing duct. MRIs were abnormal in 13% of 24 breasts with an atypical lavage, and ductal carcinoma in situ was diagnosed subsequently in 1 breast.

CONCLUSIONS. Atypia frequently is diagnosed by NDL, but the reproducibility of repeat lavage is low. Lavage atypia may be physiologic or artifactual rather than pathologic in many instances. Marked atypia occasionally may represent mammographically occult ductal carcinoma in situ. ***Cancer* 2005;103:1129–36.**

KEYWORDS: breast neoplasms, precancerous conditions, nipples, epithelial cells.

Most breast carcinomas arise from the epithelial cells lining the ductal system, and atypical ductal epithelium is a marker of increased risk for the development of breast carcinoma. Wresch et al. followed 2300 women over 12.7 years and found that cytologic atypia in nipple aspirate fluid (NAF) was associated with a relative risk for breast carcinoma of 4.9.[1] An updated analysis of those data adjusted the relative risk down to 2.8.[2] Fabian et al. reported that women with a personal history of invasive or in situ breast carcinoma or with a 10-year Gail breast carcinoma risk $>$ 4% who were found to have atypical cells on random fine-needle aspiration breast biopsy (FNAB) were 5 times more likely to develop breast carcinoma than women with a 10-year Gail risk $<$ 4% and no atypia.[3] Dupont et al. found that the risk of breast carcinoma was increased 4.3-fold in women who were diagnosed with atypical hyperplasia by surgical biopsy.[4] Nipple duct lavage (NDL) has been proposed as a minimally invasive technique for obtaining breast epithelial cells, with the assumption that atypia identified by NDL confers the same relative risk for breast carcinoma as atypia identified by NAF, FNAB, or surgical biopsy.

DOI 10.1002/cncr.20884
Published online 31 January 2005 in Wiley InterScience (www.interscience.wiley.com).

Although this has not been confirmed in a prospective trial to date, there is justifiable interest in developing and validating a minimally invasive procedure for the detection of atypical hyperplasia.

The National Surgical Adjuvant Breast Project (NSABP)-sponsored Breast Cancer Prevention Trial (BCPT-P1) demonstrated that 5 years of tamoxifen reduced the risk of breast carcinoma by ≈ 50% in increased risk women,[5] but most eligible women refuse to take tamoxifen.[6] Because women with atypical hyperplasia are at significantly increased risk for breast carcinoma and experience the greatest risk reduction with tamoxifen (86%), a test for atypia, such as NDL, may help eligible women decide to accept tamoxifen.

Although clinical guidelines for NDL were published previously.[7,8] many of the basic performance characteristics of the procedure remain unknown. A multiinstitutional study comparing NDL with NAF found that the insufficient sample rate was much lower for NDL than for NAF (22% vs. 73%) and that the atypia rate was much higher (24% vs. 10%).[9] However, currently, it is unknown whether atypia diagnosed by NDL predicts an increased risk for breast carcinoma. Some atypical lavages may reflect underlying atypical hyperplasia, but others are likely to reflect reversible physiologic changes related to the hormonal milieu, benign intermediate-risk lesions (such as intraductal papilloma or papillomatosis), or fully developed ductal carcinoma in situ (DCIS). Lavage atypia that is not reproducible may be related to reversible physiologic changes in the breast epithelium, whereas atypia that is reproduced may be related to fixed, underlying pathologic alterations. We performed repeat NDL and magnetic resonance imaging (MRI) scans in women with lavage atypia to estimate the prevalence of persistent lavage atypia and the prevalence of mammographically occult DCIS or invasive carcinoma when atypical cells are identified.

MATERIALS AND METHODS

Eligibility Criteria

The Institutional Review Board at the University of Texas Southwestern Medical Center at Dallas approved the protocol for this study, and written informed consent was documented for all participants. Patients from the Mary L. Brown Cancer Genetics and Risk Assessment Clinic at the University of Texas Southwestern Center for Breast Care were recruited for enrollment. A comprehensive risk assessment was performed for each patient. Patients with incident breast carcinoma and unaffected women age ≥ 18 years who presented for breast carcinoma risk assessment were offered duct lavage regardless of their calculated risk level. Exclusion criteria included the presence of an undefined palpable or mammographic breast lesion suspicious for malignancy; bilateral prophylactic mastectomy; any prior breast irradiation; any systemic chemotherapy in the past; a performance status that restricted normal activity for a significant portion of the day; current use of androgens, luteinizing hormone-releasing hormone analogs, prolactin inhibitors, antiandrogens, or corticosteroids (women were eligible if these drugs were discontinued 3 months prior to lavage); ever use of tamoxifen, raloxifene, or other selective estrogen receptor modulator (SERM) therapy; or pregnancy or lactation within 6 months.

NDL Procedure

Local anesthetic cream (EMLA; AstraZeneca, London, United Kingdom) was applied to the nipple and then covered with an occlusive patch 1–2 hours prior to the procedure. At the start of the procedure, the patient performed a self-breast massage, after which, the nipple was dekeratinized with a mild abrasive gel (Nuprep; D. O. Weaver and Company, Aurora, CO). The operator then continued the breast massage in an effort to express NAF. If no NAF was elicited manually, then a nipple aspirator (FirstCyte; Cytyc Health Corporation) was used. Fluid-producing ducts initially were cannulated with a tapered dilator coated with 2% lidocaine jelly, after which, a ductal lavage microcatheter (FirstCyte Microcatheter; Cytyc Health Corporation) was inserted. Saline (10 mL) was infused into the duct in 0.5-mL increments, and the effluent fluid was aspirated. An attempt was made to lavage all fluid-producing ducts and at least one nonfluid-producing duct from each breast. The location of each cannulated duct orifice was recorded on a circular grid with 45 cells, so that the orifice of any duct that yielded atypical cells could be identified in the future. Repeat lavage was offered to women whose initial lavage returned atypical cells. At the time of repeat lavage, every effort was made to recannulate the same ducts that were cannulated at the initial lavage. Breast MRI was recommended for all women whose initial lavage returned atypical cells but was performed only if third-party payor approval could be obtained.

Cytologic Evaluation

The lavage effluents from each duct were collected separately in 30 mL of CytoLyt solution (Cytyc Health Corporation). Cytology slides were prepared using the ThinPrep method and were stained using the Papanicolaou technique. All slides were evaluated by the same breast cytopathologist (R.A.), who classified each sample according to the most severe alterations iden-

tified: insufficient for diagnosis, normal epithelium or apocrine metaplasia only, typical epithelial hyperplasia, mild atypia, or marked atypia. Cytologic interpretation was performed according to the guidelines published by the Cytyc Health Corporation (http://www.ductallavage.com/professionals/cytologyTraining.cfm). Briefly, mild atypia was defined as clusters of crowded, overlapping cells with slight nuclear enlargement, mild anisonucleosis, prominent nucleoli, occasional myoepithelial cells, and granular, evenly distributed chromatin. Marked atypia was diagnosed when these same features were more pronounced and included marked anisonucleosis, significantly increased nuclear:cytoplasmic ratios, and irregular, clumping chromatin.

Statistical Analysis

Proportions and atypia prevalence rates were compared using the Fisher exact test. The α value was set at 0.05.

RESULTS

Demographics

Ductal lavage was performed in 377 ducts from 208 breasts in 108 female patients. On average, 1.8 ducts were lavaged per breast, and 3.5 ducts were lavaged per patient. There were 41 women with incident breast carcinoma and 67 women who were unaffected with breast carcinoma but who had completed a comprehensive breast carcinoma risk assessment. Among the women who were unaffected with breast carcinoma, 52% had a 5-year Gail risk $< 1.7\%$, and 48% had a 5-year Gail risk $\geq 1.7\%$. The mean patient age was 46.3 years (range, 30–82 years), and 42% of patients were postmenopausal. Most of the patents were Caucasian (81%), and 35% of patients were using oral contraceptives or hormone replacement therapy at the time of initial sampling. NAF was expressible from 86% of the patients (Table 1).

TABLE 1
Characteristics of the Study Sample

Characteristic	No. of patients (%)
Total patients	108 (100.0)
Age (yrs)	
Mean	46.3
Range	30.0–81.5
Ethnicity	
Caucasian	87 (80.6)
African American	16 (14.8)
Hispanic	4 (3.7)
Asian	1 (0.9)
Expressible nipple aspirate fluid	93 (86.1)
Menopausal status	
Premenopausal	56 (51.9)
Perimenopausal	7 (6.5)
Postmenopausal	45 (41.7)
Oral contraceptive use (premenopausal)	16/56 (28.6)
Hormone replacement (perimenopausal and postmenopausal)	22/52 (42.3)
Risk groups	
Breasts ipsilateral to a breast carcinoma	36 (100.0)
DCIS only	3/36 (8.3)
Infiltrating ductal carcinoma	29/36 (80.6)
Infiltrating lobular carcinoma	3/36 (8.3)
Medullary carcinoma	1/36 (2.8)
Any associated DCIS	29/36 (80.6)
Breasts contralateral to a breast carcinoma	38
Unaffected risk assessed patients	67/108 (62.0)
History of ADH	1/67 (1.5)
BRCA gene mutation	3/67 (4.5)
5-Yr Gail risk	
0.01–0.85	20/67 (29.9)
0.86–1.69	15/67 (22.4)
1.70–2.54	17/67 (25.4)
> 2.54	15/67 (22.4)

DCIS: ductal carcinoma in situ; ADH: atypical ductal hyperplasia.

Frequency of Atypia

Table 2 summarizes the atypia rates for 36 breasts with an incident breast carcinoma, 38 breasts contralateral to an incident breast carcinoma, and 134 breasts from women who were unaffected by breast carcinoma. Results are reported separately for the right and left breasts of the unaffected women to permit comparisons between patients with breast carcinoma and unaffected patients on a per-breast basis. The insufficient sample rate was higher for ducts from breasts with an incident carcinoma (40%) than for ducts from breasts that were unaffected with breast carcinoma (27%; $P = 0.06$). Atypia of any degree was diagnosed in 36% of breasts with an incident breast carcinoma and in 24% of breasts that were unaffected with breast carcinoma ($P = 0.19$), but marked atypia was diagnosed more frequently in breasts with an incident breast carcinoma (22%) than in unaffected breasts (7%; $P = 0.01$). Among breasts that were unaffected with breast carcinoma, we diagnosed cytologic atypia in 18% of ducts, 24% of breasts, and 30% of patients. There were no trends in the insufficient sample rate or in the frequency of diagnosis of mild or marked atypia for initial lavages over time (Fig. 1).

Reproducibility of Atypia

Among the 32 patients who had an atypical lavage from a breast that was unaffected with breast carcinoma, repeat lavage was performed for 23 patients. Four patients with breast carcinoma received chemotherapy after the initial lavage, which rendered them ineligible for repeat lavage of the contralateral breast; two women without breast carcinoma moved out of

TABLE 2
Frequency of Atypia by Sampling Group

	No. of patients (%)							
	Ducts				Breasts			
Variable	ICMD	Mild atypia	Marked atypia	Any atypia	ICMD	Mild atypia	Marked atypia	Any atypia
Cancerous breast	24/60 (40.0)	7/60 (11.7)	8/60 (13.3)	15/60 (25.0)	11/36 (30.6)	5/36 (13.9)	8/36 (22.2)	13/36 (36.1)
Contralateral breast	19/57 (33.3)	6/57 (10.5)	3/57 (5.3)	9/57 (15.8)	10/38 (26.3)	5/38 (13.2)	3/38 (7.9)	8/38 (21.1)
Unaffected right breast	35/133 (26.3)	15/133 (11.3)	7/133 (5.3)	22/133 (16.5)	10/67 (14.9)	12/67 (17.9)	5/67 (7.5)	17/67 (25.4)
Unaffected left breast	32/127 (25.2)	19/127 (15.0)	6/127 (4.7)	25/127 (21.3)	13/67 (19.4)	12/67 (17.9)	4/67 (6.0)	16/67 (23.9)

ICM: insufficient cellular material for diagnosis.

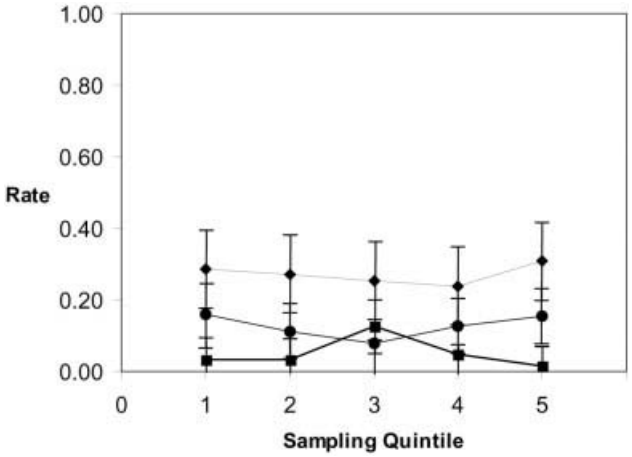


FIGURE 1. Insufficient cellular material for diagnosis (ICMD) rate and frequency of mild and marked atypia for initial lavages over time. The entire study sample was divided into quintiles to detect trends in the insufficient sample rate or in the frequency of diagnosis of mild or marked atypia that might suggest systematic changes in the performance or interpretation of nipple duct lavage over time. No such trends were identified. Error bars bracket the 95% confidence intervals. ♦: ICMD; ●: Mild atypia; ■: Marked atypia.

the area; two women declined repeat lavage; and one woman underwent mastectomy without repeating the lavage after an MRI was interpreted as highly suspicious (Fig. 2). Repeat lavage was performed 2.3–14.3 months (median, 8.3 months) after the initial lavage. Because every effort was made to relavage all of the ducts that had been lavaged initially in the patients who had at least 1 duct diagnosed as atypical, a total of 78 ducts from 45 breasts were relavaged in these 23 patients. If any duct was classified as atypical on the repeat lavage, then the atypia was scored as "reproducible" for that patient. The repeat lavage was classified as atypical for 11 of 23 patients (48%), 11 of 32 breasts (34%), and 8 of 42 ducts (19%) that were diagnosed initially as atypical (Table 3). Failure to reproduce the atypia was due to an insufficient sample on the second lavage in 13% of patients and due to a diagnosis of normal epithelium or typical hyperplasia only in 39% of patients. Marked atypia on the initial lavage was no more predictive of an atypical second lavage than mild atypia (44% vs. 50%; $P = 1.0$).

Among the patients who produced NAF, atypia was diagnosed on the second lavage in 55%, compared with 0% for the 3 patients who did not produce NAF ($P = 0.25$). Among the patients who had a 5-year Gail risk $\geq 1.7\%$, atypia was diagnosed on the second lavage in 22% of patients, compared with 70% of the patients who had a 5-year Gail risk $< 1.7\%$ ($P = 0.10$). Reproducibility rates were similar for premenopausal women compared with perimenopausal or postmenopausal women, for women who were taking hormonal medications compared with women who were not taking these medications, and for women who underwent repeat lavage < 8.3 months after the initial lavage (the median interval for this series) compared with women who underwent repeat lavage ≥ 8.3 months after the initial lavage.

MRI Findings

MRI was performed in 24 breasts from 17 women whose initial lavage was interpreted as atypical. The atypia was marked in 9 breasts, and the MRI was abnormal in 1 of those breasts (11%). Total mastectomy revealed 10 cm of DCIS in this patient (Fig. 2). Repeat lavage was not performed prior to the MRI and subsequent surgery. The initial lavage was interpreted as mildly atypical in 15 breasts; and of those, the MRI was abnormal in 2 breasts (13%). The MRI was interpreted as borderline suspicious in both of those breasts. Repeat MRI in one patient demonstrated resolution of the region of abnormal enhancement, and repeat lavage in this patient was classified as typical epithelial hyperplasia only. In the second patient with

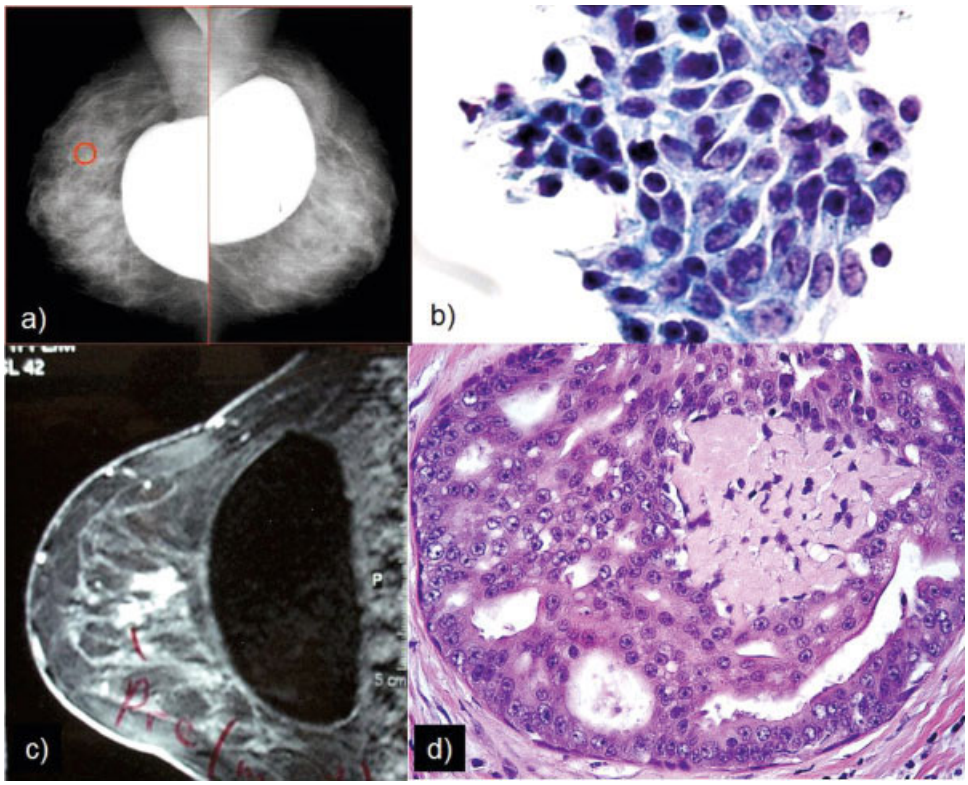


FIGURE 2. An interesting patient. (a) A screening mammogram revealed a suspicious cluster of microcalcifications in the right breast. The left breast was interpreted as normal. (b) Nipple duct lavage of the left breast at the time of right mastectomy returned markedly atypical cells. (c) This left breast magnetic resonance image demonstrates three areas of abnormal enhancement. (d) A subsequent left total mastectomy revealed 10 cm of ductal carcinoma in situ.

TABLE 3
Classification of Repeat Lavage Results According to Initial Lavage Results

	Initial lavage: No. of patients (%)										
	Mild atypia					Marked atypia					
Repeat lavage	Marked	Mild	Normal or EH	ICMD	Any atypia	Marked	Mild	Normal or EH	ICMD	Any atypia	Any atypia
By patient	3/14 (0.21)	4/14 (0.29)	5/14 (0.36)	2/14 (0.14)	7/14 (0.50)	2/9 (0.22)	2/9 (0.22)	4/9 (0.44)	1/9 (0.11)	4/9 (0.44)	11/23 (0.48)
By breast	4/22 (0.18)	5/22 (0.23)	9/22 (0.41)	4/22 (0.18)	9/22 (0.41)	1/10 (0.10)	1/10 (0.10)	6/10 (0.60)	2/10 (0.20)	2/10 (0.20)	11/32 (0.34)
By duct	3/30 (0.10)	4/30 (0.13)	13/30 (0.43)	10/30 (0.33)	7/30 (0.23)	0/12 (0.00)	1/12 (0.08)	9/12 (0.75)	2/12 (0.17)	1/12 (0.08)	8/42 (0.19)

EH: typical epithelial hyperplasia; ICMD: insufficient cellular material for diagnosis.

mild atypia, a repeat lavage was classified as marked atypia, and a targeted ultrasound examination was unremarkable. Repeat MRI in this patient was interpreted as entirely normal.

DISCUSSION

NDL has been proposed as a secondary risk-stratification procedure for women who are determined to be at increased risk for breast carcinoma based on epidemiologic models. It is believed that the atypical cells identified by NDL confer the same breast carcinoma risk as atypia found in nipple aspirates, needle biopsies, and surgical biopsies; however, currently, there are no data to confirm this, and many of the essential performance characteristics of NDL are yet to be elucidated.

Although NDL currently is recommended for risk stratification, and not for early detection of breast carcinoma, it is reasonable to suppose that, if lavage atypia is a strong predictor of breast carcinoma risk,

TABLE 4
Comparison of Insufficient Cellular Material for Diagnosis and Atypia Rates in the Current Series with Previously Reported Rates

	No. of patients (%)	
Results	**Dooley et al., 2001[9]**	**Current series**
By patient		
ICMD	84/383 (22)	17/105 (16)
Mild atypia	66/383 (17)	19/105 (18)
Marked atypia	26/383 (7)	13/105 (12)
By duct		
ICMD	173/591 (29)	86/317 (27)
Mild atypia	77/591 (13)	40/317 (13)
Marked atypia	28/591 (5)	16/317 (5)

ICMD: insufficient cellular material for diagnosis.

then it would occur at a high frequency in breasts with an incident breast carcinoma. Marked atypia, in fact, was more common in breasts with an incident breast carcinoma (22%) than in unaffected breasts (6–8%). This is a higher marked atypia rate than that reported for a series of 28 mastectomy patients.[10] In that study, atypia rates were reported per lavage sample (ducts) rather than by breast, and marked atypia was identified in only 4 of 29 (14%) adequate samples from patients with an incident breast carcinoma. The exclusion of patients who had insufficient samples from the current series resulted in a marked atypia rate of 8 of 25 (32%) for breasts with an incident carcinoma compared with 12 of 139 (9%) for unaffected breasts ($P = 0.003$). In the prior study, the extent of carcinoma in situ appeared to correlate with the degree of atypia identified in the lavage samples, but DCIS was identified in 23 of 28 patients (82%), and it was not possible to determine whether or not lavage atypia was associated primarily with DCIS. Similarly, 29 of 36 patients (81%) with incident breast carcinoma in the current series had a DCIS component, but marked atypia was identified in 3 of 7 patients (42%) who had no DCIS component. It is clear that the marked atypia rate for breasts with an incident breast carcinoma is considerably higher than that previously reported and is considerably higher than that observed for breasts that are unaffected by breast carcinoma.

The prevalence of lavage atypia measured in our series of patients, who were unselected for breast carcinoma risk, was nearly identical to that reported in the first large validation series, a series that was limited to women who were at increased risk for breast carcinoma[9] (Table 4). To make this comparison, we have considered only breasts that were unaffected by breast carcinoma (38 breasts contralateral to a breast carcinoma and 134 breasts from women who were unaffected with breast carcinoma). Results for 3 women with breast carcinoma who did not undergo lavage of the contralateral breast were excluded, leaving a total of 105 patients for comparison.

Atypia was reproduced on repeat lavage in only 8 of 42 ducts (19%), 11 of 32 breasts (34%), and 11 of 23 patients (48%). A recent series of duct lavages in 38 high-risk women reported an atypia prevalence of 23% for those with an adequate sample and reproducibility of the atypia in only 1 of the 4 patients who underwent repeat lavage.[11] The reproducibility of any cytology-based screening test will be related to the physiologic factors that affect the cytologic features of the cells collected, variability in the sampling procedure, and variability in the cytologic interpretation. The breast is an exquisitely hormone-responsive organ, and fluctuations in the hormonal milieu may affect cytology-based screening tests. Exogenous estrogens have been associated with atypical hyperplasia in humans,[12] monkeys,[13] and rodents,[14] but it is not known whether these lesions are reversible in humans when the hormones are withdrawn. It is noteworthy that nine of our patients who underwent repeat lavage were using hormonal medications at the time of the initial lavage. Atypia was reproduced in 2 of the 4 patients (50%) who discontinued these medications between the first and second lavages. Endogenous estrogens also may influence the cytologic appearance of breast epithelial cells, although Mitchell et al. found no significant changes in breast epithelial cells recovered from weekly NAF samples that were collected over two menstrual cycles.[15]

Although every effort was made to recannulate the same ducts that were diagnosed as atypical on the initial lavage, technical problems with resampling may have contributed to the low reproducibility. The shear numbers of duct orifices clustered near the center of the papilla (11–48 orifices; median, 27 orifices)[16] presents a challenge for recannulation of a specific duct. This may have been compounded by our study design, which permitted cannulation of duct orifices that were not producing NAF, because it is likely that NAF production, in conjunction with location information recorded on a grid, provides valuable visual cues for reidentifying a specific duct orifice. The insufficient sample rate on relavage for ducts that initially returned atypical cells was 12 of 42 samples (29%), essentially identical to the insufficient sample rate for the initial series of lavages. Possible reasons for the failure to obtain an adequate sample from ducts initially yielding atypical cells include inadvertent cannulation of a different duct orifice, regression of an atypical proliferative lesion either as a consequence of the initial lavage procedure or for reasons

unrelated to the initial lavage, and ductal injury at the time of the initial lavage that precluded satisfactory recannulation and lavage. The same operator performed all of the lavage procedures, and insufficient sample rates were stable over time (Fig. 1), excluding differences in operator experience or technique as factors in the assessment of reproducibility. It has been suggested that the use of normal saline for NDL can induce artifactual atypia and that plasmolyte is a superior lavage solution. Because all initial and repeat lavages were performed using normal saline, this is unlikely to have influenced reproducibility rates.

Finally, interobserver and intraobserver variation in cytology scoring can impact the reproducibility of cytologic screening tests. Interobserver variability was excluded by having the same cytopathologist evaluate all of the samples from this study. It is possible that atypia was over-called in the earliest period of the study, resulting in lower atypia rates for the repeat lavages, but this is unlikely, because the frequency of diagnosis of mild or marked atypia for the initial lavages was stable over time (Fig. 1).

It is likely that a combination of physiologic and technical factors accounted for the low reproducibility of repeat lavage measured in this series. Neither marked atypia on the initial lavage nor a 5-year Gail risk $\geq$ 1.7% predicted atypia on repeat lavage, but there was a trend toward improved reproducibility when the atypia initially was diagnosed in a NAF-producing duct. We previously reported that the frequency of lavage atypia is similar for patients with a 5-year Gail risk $<$ 1.7% and $\geq$ 1.7% and for ducts that produce NAF compared with ducts that do not produce NAF.[17] The high prevalence of lavage atypia noted in this and prior studies, combined with a low reproducibility, makes it unlikely that a single NDL demonstrating either mildly or markedly atypical cells will predict a high risk for breast carcinoma. Adjunctive tests, such as tumor suppressor gene methylation status,[18] loss of heterozygosity analysis,[19] or chromosome copy number determination,[20,21] are feasible for NDL samples and may improve the predictive value of NDL cytology.

Lavage atypia was associated with significant MRI abnormalities in only 1 of 24 breasts. The atypia in this breast was classified as marked and ultimately was diagnosed as ductal carcinoma in situ. In 23 breasts with atypical lavage results, there were no reproducible MRI findings. This is in marked contrast to a recent series that identified MRI abnormalities in six of seven breasts with atypical lavages.[11] Only one of those breasts was biopsied, and the diagnosis was atypical ductal hyperplasia. We performed MRI only in women whose third-party payors agreed to reimburse for the test (17 of 32 women). This is likely to have biased our results; however, given the low frequency of MRI findings, we currently are performing MRI only if marked atypia is confirmed on repeat lavage.

Ductal lavage is an excellent tool for retrieving breast epithelial cells, but the reproducibility of serial sampling is poor. In addition, lavage atypia is associated only infrequently with MRI findings but may represent mammographically occult DCIS. Until prospective studies validate lavage atypia as a marker for breast carcinoma risk, it is best to use it in the context of clinical trials.

REFERENCES

1. Wrensch MR, Petrakis NL, King EB, et al. Breast cancer incidence in women with abnormal cytology in nipple aspirates of breast fluid. *Am J Epidemiol.* 1992;135:130–141.
2. Wrensch MR, Petrakis NL, Miike R, et al. Breast cancer risk in women with abnormal cytology in nipple aspirates of breast fluid. *J Natl Cancer Inst.* 2001;93:1791–1798.
3. Fabian CJ, Kimler BF, Zalles CM, et al. Short-term breast cancer prediction by random periareolar fine-needle aspiration cytology and the Gail risk model. *J Natl Cancer Inst.* 2000;92:1217–1227.
4. Dupont WD, Parl FF, Hartman WH, et al. Breast Cancer risk associated with proliferative breast disease and atypical hyperplasia. *Cancer.* 1993;71:1258–1265.
5. Fisher B, Constantino 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;90:1371–1388.
6. Port ER, Montgomery LL, Heerdt AS, Borgen PI. Patient reluctance toward tamoxifen use for breast cancer primary prevention. *Ann Surg Oncol.* 2001;8:580–585.
7. Morrow M, Vogel V, Ljung BM, O'Shaughnessy JA. Evaluation and management of the woman with an abnormal ductal lavage. *J Am Coll Surg.* 2002;194:648–656.
8. O'Shaughnessy JA, Ljung BM, Dooley WC, et al. Ductal lavage and the clinical management of women at high risk for breast carcinoma. *Cancer.* 2002;94:292–298.
9. Dooley WC, Ljung B, Veronesi U, et al. Ductal lavage for the detection of cellular atypia in women at high risk for breast cancer. *J Natl Cancer Inst.* 2001;93:1624–1632.
10. Brogi E, Robson M, Panageas KS, Casadio C, Ljung B-M, Montgomery L. Ductal lavage in patients undergoing mastectomy for mammary carcinoma. A correlative study. *Cancer.* 2003;98:2170–2176.
11. Hartman A-R, Daniel BL, Kurian AW, et al. Breast magnetic resonance image screening and ductal lavage in women at high genetic risk for breast carcinoma. *Cancer.* 2004;100:479–489.
12. Zera RT, Danielson D, Van Camp JM, et al. Atypical hyperplasia, proliferative fibrocystic change, and exogenous hormone use. *Surgery.* 2001;130:732–737.
13. Tavassoli FA. The influence of endogenous and exogenous reproductive hormones on the mammary glands with emphasis on experimental studies in rhesus monkeys. *Verh Dtsch Ges Pathol.* 1997;81:514–520.

14. Harvell DM, Strecker TE, Tochacek M, et al. Rat strain-specific actions of 17beta-estradiol in the mammary gland: correlation between estrogen-induced lobuloalveolar hyperplasia and susceptibility to estrogen-induced mammary cancers. *Proc Natl Acad Sci USA.* 2000;97:2779–2784.
15. Mitchell G, Trott PA, Morris L, Coleman N, Sauter E, Eeles RA. Cellular characteristics of nipple aspiration fluid during the menstrual cycle in healthy premenopausal women. *Cytopathology.* 2001;12:184–196.
16. Going JJ, Moffat DF. Escaping from Flatland: clinical and biological aspects of human mammary duct anatomy in three dimensions. *J Pathol.* 2004;203:538–544.
17. Johnson-Maddux A, Ashfaq R, Naftalis EZ, Leitch AM, Hoover S, Euhus DM. Patient and duct selection for nipple duct lavage. *Am J Surg.* 2004;188:390–394.
18. Evron E, Dooley WC, Umbricht CB, et al. Detection of breast cancer cells in ductal lavage fluid by methylation-specific PCR. *Lancet.* 2001;357:1335–1336.
19. Isaacs C, Cavalli LR, Cohen Y, et al. Detection of LOH and mitochondrial DNA alterations in ductal lavage and nipple aspirate fluids from high-risk patients. *Breast Cancer Res Treat.* 2004;84:99–105.
20. Yamamoto D, Senzaki H, Nakagawa H, Okugawa H, Gondo H, Tanaka K. Detection of chromosomal aneusomy by fluorescence in situ hybridization for patients with nipple discharge. *Cancer.* 2003;97:690–694.
21. King BL, Tsai SC, Gryga ME, et al. Detection of chromosomal instability in paired breast surgery and ductal lavage specimens by interphase fluorescence in situ hybridization. *Clin Cancer Res.* 2003;9:1509–1516.

ELSEVIER

Molecular and Cellular Probes 20 (2006) 191–196

Molecular and Cellular Probes

www.elsevier.com/locate/ymcpr

A comparison of five methods for extracting DNA from paucicellular clinical samples

Leslie Cler, Dawei Bu, Cheryl Lewis, David Euhus *

Hamon Center for Therapeutic Oncology Research, UT Southwestern Medical Center at Dallas, 5323 Harry Hines Boulevard, Dallas, TX 75390-9155, USA

Received 23 February 2005; accepted for publication 20 December 2005
Available online 3 March 2006

Abstract

Translational protocols in cancer and carcinogenesis often require isolation of genomic DNA from paucicellular clinical samples. DNA extraction methods for PCR-based applications should optimize the recovery of amplifiable DNA. We compared five methods for DNA extraction in paucicellular epithelial and lymphocyte samples using proportion of extractions producing amplifiable DNA and mean real-time PCR Ct values for GAPDH as the endpoint measures. The methods included solid-phase DNA adsorption (QIAamp), sequential protein and DNA precipitation (Puregene), magnetic bead adsorption (Dynabeads), phenol–chloroform extraction, and single-step proteinase K digestion. In general, the performance of the three commercial kits was superior to either phenol–chloroform extraction or single-step proteinase K digestion. However, QIAamp and Puregene produced amplifiable DNA more frequently than Dynabeads for starting cell numbers $<50{,}000$. GAPDH Ct values for QIAamp extractions showed the greatest dynamic range and the best linearity across the range of starting cell numbers, but QIAamp was not statistically significantly superior to Puregene. Of the three commercial kits, Puregene is the least expensive. QIAamp and Puregene DNA extraction methods are well-suited for the preparation of paucicellular clinical samples for PCR-based assays.

Keywords: DNA extraction; Paucicellular; Real-time PCR; Methods; Cost analysis

1. Introduction

Translational protocols in cancer and carcinogenesis often require isolation of genomic DNA from paucicellular clinical samples such as fine needle aspirates [1], nipple fluid aspirates [2,3], sputum [4], buccal swabs [5], or urine [6]. If high quality amplifiable DNA can be extracted from these samples, they are often suitable for LOH analysis [7], gene copy number determinations [8], genotyping [9], mutation analysis [10,11], or promoter methylation studies [12]. The classical approach to DNA extraction employs organic solvents to dissolve DNA after which it is precipitated in absolute alcohol. Though suitable for highly cellular samples, this approach requires multiple centrifugation steps and often results in poor yields of amplifiable DNA when the starting material is limited. Newer approaches for DNA extraction include single-step proteinase K digestion (without the use of organic solvents), adsorption of DNA on to silica gel membranes or magnetic beads, or simplified approaches for sequentially precipitating proteins and then DNA.

The DNA content of an extract is often estimated by measuring the absorbance of light at 260 nm (A260) and the purity of the DNA by calculating the A260/A280 ratio. A260 values between 0.1 and 1.0 correlate with DNA content in a linear fashion, but values this high are rarely achieved when the starting material is limited. Real-time PCR provides a semi-quantitative approach for estimating the content of amplifiable DNA in extracts from paucicellular samples. The Ct value determined by this method is the number of PCR cycles required to generate a specified quantity of product. Ct values correlate inversely with the quantity of starting template (e.g. lower Ct values indicate greater quantities of starting template).

We compared five DNA extraction methods using paucicellular samples from various aneuploid and diploid cells. The selected extraction methods are representative of the diverse approaches that are commonly employed in modern laboratories: organic solvent extraction, non-solvent-based

* Corresponding author. Address: Division of Surgical Oncology, Hamon Center for Therapeutic Oncology Research, UT Southwestern Medical Center at Dallas, E6.222, 5323 Harry Hines Boulevard, Dallas, TX 75390-9155, USA. Tel.: +1 214 648 6467; fax: +1 214 648 7965.

E-mail addresses: leslie.cler@utsouthwestern.edu (L. Cler), dawei.bu@utsouthwestern.edu (D. Bu), cheryl.lewis@utsouthwestern.edu (C. Lewis), david.euhus@utsouthwestern.edu (D. Euhus).

0890-8508/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.mcp.2005.12.003

enzymatic digestion, solid-phase adsorption, sequential protein and DNA precipitation, and magnetic bead adsorption. The endpoints compared were proportion of extractions producing amplifiable DNA and Ct values for GAPDH as measured by real-time PCR.

2. Materials and methods

2.1. Cells

DNA was extracted from two aneuploid tumor cells lines (cervical cancer cell line HeLa [13] and breast cancer cell line HCC1806 [14]), two diploid Human Mammary Epithelial Cell (HMEC) cultures (UTSW991 and UTSW1004), and freshly isolated lymphocytes from two donors. The lymphocytes were isolated from whole blood using Vacutainer CPT tubes (Becton Dickinson and Company, Franklin Lakes, NJ) according to the manufacturer's instructions. Based on hemocytometer counts 50, 500, 5000, or 50,000 cells were pelleted for DNA extraction.

2.2. DNA extraction

The salient features of each of the five DNA extraction methods are summarized in Table 1. Separate DNA extractions were performed in triplicate for each of the six cell types, each of the four starting cell counts and each of the five methods (360 extractions). Three methods employed commercially available kits (QIAamp, Puregene, and Dynabeads, respectively) in which case extractions were performed according to the manufacturer's instructions. A standard phenol/chloroform extraction method was tested as well. Briefly, the cells were suspended in 350 μl TE buffer with 20 μl of 10% SDS. Ten microliter of proteinase K at 10 μg/μl was added and the tubes incubated at 37 °C for 36 h. The contents of the tubes were mixed with 20 μl of 5 M NaCl after which 400 μl of phenol:chloroform:isoamyl alcohol (25:24:1) at pH 8.2 was added. The tubes were then centrifuged at 16,000*g* for 10 min and the supernatant transferred to a fresh tube. The phenol:chloroform:isoamyl alcohol step was repeated twice more and then the DNA was precipitated with 800 μl of cold absolute ethanol. For the single-step proteinase K method, the cell pellet was suspended in 50 μl of extraction solution prepared by combining 0.5 M EDTA pH 8.0 (20 μl), 1 M Tris pH 8.0 (200 μl), Tween-20 (50 μl), Proteinase K at 20 mg/ml (100 μl), and ultrapure water (9.63 ml). The tubes were incubated at 37 °C for 36 h after which the proteinase K was inactivated by heating to 95 °C for 10 min.

Table 1
Summary of the five DNA extraction methods

Method	Name	Vendor	Catalog no.	Salient features
1	QIAamp blood Mini-kit	Qiagen	51104	Silica-gel membrane binds DNA, contaminants pass through spin column
2	Puregene	Gentra	D-5500A	Protein precipitation followed by DNA precipitation
3	Dynabeads DNA DIRECT universal	Dynal	630.06	DNA adsorbed onto magnetic beads
4	Phenol/ chloroform	N/A	N/A	Proteinase K, followed by organic extraction, followed by DNA precipitation
5	Proteinase K	N/A	N/A	Proteinase K without DNA precipitation

N/A, not applicable.

Each extract was brought to a final volume of 20 μl in ultrapure water and the A260 and A280 measured spectrophotometrically after diluting 2 μl of the extract in 98 μl of water (Beckman DU-64, Beckman Instruments, Inc.). Samples were stored at 4 °C until analysis (a maximum of 3 weeks).

2.3. Real-time PCR

Genomic DNA for GAPDH was amplified by real-time PCR after combining 20 μl SYBR Green Jumpstart Taq ReadyMix (Sigma, 20 mM Tris–HCl at pH 8.3, 100 mM KCl, 7 mM $MgCL_2$, dNTP's at 0.4 mM each, 0.05 units/μl Taq DNA Polymerase, JumpStart Taq antibody, and SYBR Green I), 0.4 μl internal reference dye, 18 μl of water and 2 μl of the DNA extract. GAPDH primers were used at a final concentration of 0.2 μM; fwd: GCCTGCTTCAC-CACCTTCTTG, rev: GTCCACTGGCGTCTTCACCAC. The PCR was run on an ABI GeneAmp 5700 Sequence Detection System (Perkins–Elmer Applied Biosystems) as follows: after a 1 min preincubation at 95 °C, amplification cycles of 95 °C for 15 s and 60 °C for 1 min were repeated 50 times. DNA extractions were scored as successful if the GAPDH Ct was less than that of the water blank.

2.4. Cost

The cost per extraction was calculated for the commercially available kits by dividing the cost of the kit by the number of extractions that could be performed with the kit. The costs of phenol/chloroform and single-step proteinase K extractions were calculated based on the cost of the consumable supplies required for each extraction. The per extraction cost of each method was compared relative to the cost of phenol/chloroform alcohol extraction (i.e. per extraction cost for Method X divided by the per extraction cost for phenol/chloroform extraction).

2.5. Validation using alcohol-fixed clinical samples

Six alcohol-fixed (CytoLyt, Cytyc Corporation, Marlborough, MA) nipple duct lavage (NDL) samples containing <1000 cells and four NDL samples reported as acellular were extracted using the Puregene method in order to determine whether this method was suitable for alcohol-fixed samples. A260 and A280 values were measured, and real-time PCR performed as described in Section 1. Each extract was diluted

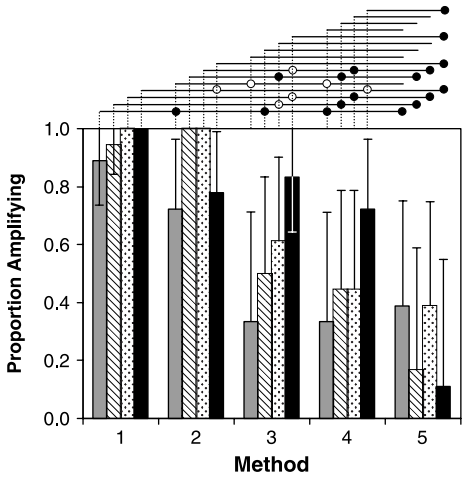


Fig. 1. Proportion of extractions producing amplifiable DNA defined as a GAPDH Ct less than that of water blank. (1) QIAamp, (2) Puregene, (3) Dynabeads, (4) phenol/chloroform, (5) single-step proteinase K. ▫ 50 cells, ▨ 500 cells, ⚃ 5000 cells, ■ 50,000 cells. Error bars show the 95% confidence interval for the means. *P*-values for comparisons between methods for the same starting cell numbers are represented as ○ $0.05 > P \geq 0.005$ and ● $P < 0.005$.

twofold from 1:2 to 1:16 and the real-time PCR was run in triplicate for each dilution using 1 μl of template.

2.6. Statistical analysis

The proportion of extractions producing amplifiable DNA was compared using Fisher's exact test. Mean Ct values are reported with 95% confidence intervals and compared using Student's *t*-tests. All statistical comparisons are two-tailed. No adjustments were made for the multiple comparisons.

3. Results

3.1. DNA yield for different cell types

In general, amplifiable DNA yield, as measured by Ct for GAPDH, was best for the cancer cells intermediate for HMEC and worst for lymphocytes for all methods and for all starting cell numbers. Analyzing the data for lymphocytes and epithelial cells separately did not affect the conclusions concerning the relative efficiencies of the five methods, so data for the six cell sources were combined.

3.2. Number of successful extractions

The QIAamp and Puregene extractions produced amplifiable DNA more frequently than the other methods when the starting cell counts were <50,000 (Fig. 1). The proportion of extractions producing amplifiable DNA was similar for QIAamp and Puregene across all starting cell numbers. For starting cell numbers of 50,000, the proportion of extractions producing amplifiable DNA was similar for all methods except for the single-step proteinase K extraction which performed poorly for all starting cell numbers.

3.3. Ct values by real-time PCR

Mean GAPDH Ct values, 95% confidence intervals, and statistical comparisons for the five methods and four starting cell numbers are depicted in Table 2. Of note, mean GAPDH Ct value for the 50-cell extractions was lowest for Puregene, but this result was not statistically significant. GAPDH Ct values for QIAamp extractions showed the greatest dynamic range and the best linearity ($R^2 = 0.919$) across the range of starting cell numbers (Fig. 2).

3.4. Absorbance at 260 nm

A260 values ranged from 0.029 to 0.034 for QIAamp, 0.024 to 0.060 for Puregene, 0.017 to 0.023 for Dynabeads, 0.180 to 0.663 for phenol/chloroform, and 0.341 to 0.542 for the single-step proteinase K extractions. GAPDH Ct did not correlate with A260 for any method. The A260/A280 ratios ranged from 2.82 to 4.83 for QIAamp, 1.81 to 2.26 for Puregene, 1.21 to 1.65 for Dynabeads, 1.41 to 1.46 for phenol/chloroform, and 0.64 to 0.70 for single-step proteinase K extractions.

3.5. Relative costs

Compared to phenol/chloroform extraction (relative cost = 1.0), the relative supply costs for the other methods were 6.2 for QIAamp, 0.3 for Puregene, 1.9 for Dynabeads, and 0.6 for single-step proteinase K. On the average, QIAamp and Puregene required 2 h of technician time from start to finish, Dynabeads 2.5 h, phenol/chloroform 3 h and single-step proteinase K 30 min.

Table 2
Mean GAPDH Ct values (95% confidence intervals) from 2 μl of extract

Method	Starting cell number			
	50	500	5000	50,000
1	30.7 (29.4–32.0)	29.4[a] (27.6–31.3)	24.3[a,b,c] (22.2–26.5)	23.4[a] (21.5–25.4)
2	29.0 (27.0–30.9)	26.4[a,b,c] (24.4–28.3)	25.1[d,e,f] (23.5–26.7)	24.9 (21.4–28.5)
3	31.3 (30.8–31.9)	28.3[d,e] (26.1–30.5)	29.2[a,d] (26.7–31.8)	23.8[b] (21.0–26.6)
4	30.8 (27.5–34.2)	31.0[b,d,f] (30.1–32.0)	29.7[b,e] (27.9–31.4)	28.0[a,b] (25.1–30.9)
5	30.1 (27.8–32.4)	32.8[c,e,f] (30.0–35.5)	30.5[c,f] (28.6–32.3)	27.8 (3.8–51.7)

Method 1, QIAamp; method 2, Puregene; method 3, Dynabeads; method 4, phenol/chloroform; method 5, single-step proteinase K. Superscript a–f denotes pairs where the difference in mean Ct was statistically significant ($P < 0.05$) by two-tailed *t*-test.

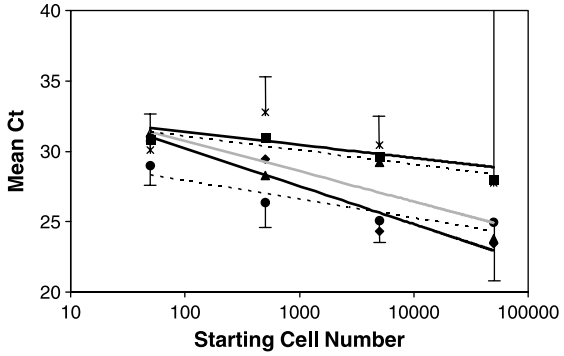


Fig. 2. Mean GAPDH Ct values by method and starting cell number. Error bars showing 95% confidence intervals are depicted for methods 2 and 5 only for clarity. All of the 95% confidence intervals are reported in Table 2. – ◆ – QIAamp ($R^2=0.919$), – - ● - – Puregene ($R^2=0.859$), – ▲ – Dynabeads ($R^2=0.775$), – - ■ - – phenol/chloroform ($R^2=0.837$), – * – single-step proteinase K ($R^2=0.344$).

3.6. Alcohol-fixed clinical samples

Performance of the Puregene method for alcohol-fixed NDL samples was similar to that observed for the unfixed test samples. Specifically, there was excellent linearity between GAPDH Ct and starting template quantity (Fig. 3). The A260 for these samples correlated well with GAPDH Ct ($R^2=0.436$, $P=0.03$).

4. Discussion

Modern clinical and translational research protocols often require PCR amplification of DNA obtained from paucicellular clinical samples. We used benign and malignant epithelial cells as well as lymphocytes for this comparative evaluation as these are representative of the samples that are frequently obtained for translational studies. The primary endpoint was amplifiable DNA as measured by real-time PCR. The Puregene method, which employs sequential protein and DNA precipitation steps, and the QIAamp method, which is based on solid-phase adsorption of DNA, outperformed the other methods for extraction of DNA from paucicellular samples. The Puregene method provided the additional advantage of lower cost. Our laboratory is primarily interested in DNA from alcohol-fixed breast epithelial cells obtained by nipple duct lavage or random periareolar fine needle aspiration biopsy. There are a variety of additional DNA extraction methods that we did not evaluate and a variety of modifications available for the assays we did evaluate. Our conclusions may not be generalizable to other applications, particularly those that use formalin-fixed tissues.

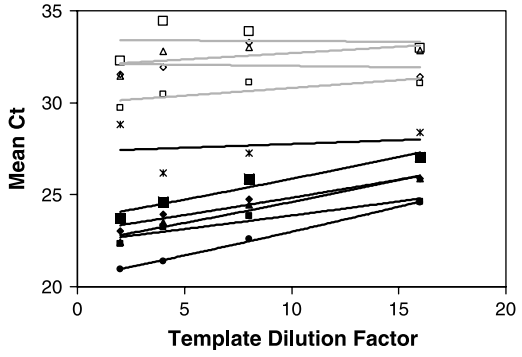


Fig. 3. Mean GAPDH Ct values for 10 nipple duct lavage samples extracted using the Puregene method. Template was diluted from 1:2 to 1:16 for each extraction. Closed symbols and solid black lines are for six samples with <1000 cells; open symbols and gray lines are for four samples reported as acellular.

Fifty cells was the lowest cell count we evaluated. Some clinical samples will contain fewer than 50 cells of interest. We are primarily interested in quantitative real-time PCR for gene copy number determinations and promoter region methylation studies and have found that the reproducibility of these assays declines precipitously when starting template is generated from fewer than 50 cells. Reproducible methods for DNA extraction and analysis of ultra-paucicellular clinical samples are needed.

4.1. Traditional approaches to DNA extraction

Phenol–chloroform–isoamyl alcohol extraction is the classical method for extracting DNA from clinical samples. A Proteinase K treatment step is often incorporated as the enzyme degrades proteins into sub-tetrameric fragments and has been shown to improve the efficiency of PCR-based applications by destroying DNases and RNases [15]. The multiple centrifugation steps required for phenol–chloroform–isoamyl alcohol extraction is cumbersome prompting some to evaluate proteinase K digestion with [16] or without [17,18] a single salting-out step rather than solvent extraction.

4.2. Previously published comparisons of DNA extraction methods

A study that compared methods for extracting DNA from bone marrow cells scraped from Giemsa-stained slides used amplification of β-globin and β-actin genes as the endpoint measure [16]. The PCR products were electrophoresed on 4% agarose gels and extraction scored as successful if a band could be detected by ethidium bromide staining. Bands were detected in 20 of 20 samples extracted using the QIAamp or proteinase K salting-out methods, and 19 of 20 samples using classical phenol–chloroform–isoamyl alcohol extraction (with proteinase K digestion). The three non-enzymatic boiling methods that were tested produced bands in only 35–80% of samples. The traditional phenol–chloroform–isoamyl alcohol extraction did not perform nearly as well in our hands. It is conceivable that the use of a DNA/RNA carrier such as glycogen and the use of phase-separating tubes would have improved the efficiency of this approach for paucicelluar samples. Nevertheless, the multiple pipetting and centrifugation steps required by this approach limits its utility for larger studies.

In another study, DNA suitable for PCR amplification was obtained from 13 of 14 fine needle aspiration samples of putative renal cell carcinoma metastases extracted using a single-step proteinase K method [17]. In this study, the extraction was scored as successful if PCR-amplified

microsatellite repeats resolved on a 6% acrylamide gel could be visualized by autoradiography. The single-step proteinase K method (without salting-out) was also found to be superior to phenol–chloroform extraction when applied to formalin-fixed paraffin-embedded thymoma tissue as determined by A260 values and intensity of SYBR Green I stained bands obtained by electrophoresing β-globin gene PCR products [19].

In contrast to these results are those obtained for DNA extracted from buccal cells using three different methods [20]. Amplification of the β-globin gene detected by ethidium bromide staining of electrophoresed PCR products was the endpoint measure in this study. Phenol–chloroform–isoamyl alcohol extraction (with proteinase K) was successful in 16/17 (94%) samples, QIAamp extraction in 12/16 (75%) and single-step proteinase K extraction in 2/16 (13%). A comparison of five commercially available solid-phase adsorption kits that used A260 as the endpoint concluded that the Genomic DNA Isolation Kit provided by Sigma produced the best DNA yields when the starting sample was whole blood, buccal swabs or muscle [21]. The QIAamp kit, which was included in this study, ranked in the middle for these five kits, but the analysis cannot be considered valid as A260 was the only measure of DNA yield employed.

4.3. A260 determinations as measure of DNA content

Quantities of nucleic acids in solution are often estimated based on the absorbance of light at a wavelength of 260 nm. An A260 of 1.0 correlates roughly with a double stranded DNA content of 50 μg/ml. A260 values between 0.10 and 1.00 are thought to correlate in a linear fashion with nucleic acid content. The A280 is traditionally taken as a measure of protein content in a solution (though nucleic acids absorb a considerable amount of light at 280 nm) and the A260/A280 ratio as a measure of the purity of the nucleic acid extract. A260/A280 ratios of 1.8–2.0 are generally considered relatively free of protein contamination, though in reality a solution with an A260/A280 of 1.8 may represent a 60/40 mixture of protein and nucleic acids [22]. A pure nucleic acid solution should have an A260/280 of 2.0.

QIAamp, Puregene and Dynabeads all produced A260 values <0.10, even with the 50,000 cell extractions. This is not unexpected when DNA is extracted from paucicellular samples using methods that effectively exclude protein carryover. Phenol–chloroform extraction and the single-step proteinase K method both produced A260 values well above 0.10 but these values must be viewed as artifactual based on the poor PCR performance that was observed when these methods were applied to paucicellular samples. It is likely that both methods were compromised by significant protein carryover or carry-over of factors that inhibit PCR [23].

4.4. Conclusions

Prior comparisons of DNA extraction methods have largely relied on the detection of PCR products on electrophoresis gels as the endpoint measure. Real-time PCR is a more sensitive and specific measure of amplifiable DNA. Each of the methods we tested yielded amplifiable DNA, but the QIAamp and Puregene methods were successful more frequently when starting cell numbers were low and were associated with the lowest Ct values by real-time PCR. These methods would seem best suited to paucicellular clinical samples.

References

[1] Bofin AM, Ytterhus B, Martin C, O'Leary JJ, Hagmar BM. Detection and quantification of HER-2 gene amplification and protein expression in breast carcinoma. Am J Clin Pathol 2004;122:110–9.
[2] Isaacs C, Cavalli LR, Cohen Y, Pennanen M, Shankar LK, Freedman M, et al. Detection of LOH and mitochondrial DNA alterations in ductal lavage and nipple aspirate fluids from high-risk patients. Breast Cancer Res Treat 2004;84:99–105.
[3] Phillips HA, Howard GC, Miller WR. Pilot studies on the p53 gene in nipple aspirate fluid from patients with breast cancer. Breast Cancer Res Treat 2000;61:139–43.
[4] Kersting M, Friedl C, Kraus A, Behn M, Pankow W, Schuermann M. Differential frequencies of p16(INK4a) promoter hypermethylation, p53 mutation, and K-ras mutation in exfoliative material mark the development of lung cancer in symptomatic chronic smokers. J Clin Oncol 2000;18:3221–9.
[5] Walker AH, Najarian D, White DL, Jaffe JF, Kanetsky PA, Rebbeck TR. Collection of genomic DNA by buccal swabs for polymerase chain reaction-based biomarker assays. Environ Health Perspect 1999;107:517–20.
[6] Seripa D, Parrella P, Gallucci M, Gravina C, Papa S, Fortunato P, et al. Sensitive detection of transitional cell carcinoma of the bladder by microsatellite analysis of cells exfoliated in urine. Int J Cancer 2001;95:364–9.
[7] Euhus DM, Cler L, Shivapurkar N, Milchgrub S, Peters GN, Leitch AM, et al. Loss of heterozygosity in benign breast epithelium in relation to breast cancer risk. J Natl Cancer Inst 2002;94:858–60.
[8] Bu D, Lewis C, Cler L, Elias K, Milchgrub S, Minna J, et al. Comparative real-time PCR for the detection of c-Myc, cyclin D1 and Her-2/neu gene amplification in breast fine needle aspirates. Am Assoc Cancer Res 2003.
[9] Wang SS, Thornton K, Kuhn AM, Nadeau JG, Hellyer TJ. Homogeneous real-time detection of single-nucleotide polymorphisms by strand displacement amplification on the BD ProbeTec ET system. Clin Chem 2003;49:1599–607.
[10] Zhang LF, Gao WM, Gealy R, Weissfeld J, Elder E, Whiteside TL, et al. Comparison of K-ras gene mutations in tumour and sputum DNA of patients with lung cancer. Biomarkers 2003;8:156–61.
[11] Phillips HA, Howard GC, Miller WR. p53 mutations as a marker of malignancy in bladder washing samples from patients with bladder cancer. Br J Cancer 2000;82:136–41.
[12] Lewis CM, Cler LR, Bu D-w, Zöchbauer-Müller S, Milchgrub S, Naftalis EZ, et al. Promoter hypermethylation in benign breast epithelium in relation to predicted breast cancer risk. Clin Cancer Res 2005;11:166–72.
[13] Hay R, Caputo J, Macy M, Mcclintock P, Reid Y. ATCC cell lines and hybridomas. American Type Culture Collection, Manassas, VA; 1994.
[14] Gazdar AF, Kurvari V, Virmani A, Gollahon L, Sakaguchi M, Westerfield M, et al. Characterization of paired tumor and non-tumor cell lines established from patients with breast cancer. Int J Cancer 1998;78:766–74.
[15] Wiegers U, Hilz H. A new method using 'proteinase K' to prevent mRNA degradation during isolation from HeLa cells. Biochem Biophys Res Commun 1971;44:513–9.
[16] Vince A, Poljak M, Seme K. DNA extraction from archival Giemsa-stained bone-marrow slides: comparison of six rapid methods. Br J Haematol 1998;101:349–51.

[17] Beaty MW, Zhuang Z, Park WS, Emmert-Buck MR, Linehan WM, Lubensky IA, et al. Fine-needle aspiration of metastatic clear cell carcinoma of the kidney: employment of microdissection and the polymerase chain reaction as a potential diagnostic tool. Cancer 1997;81:180–6.

[18] Banerjee SK, Makdisi WF, Weston AP, Mitchell SM, Campbell DR. Microwave-based DNA extraction from paraffin-embedded tissue for PCR amplification. Biotechniques 1995;18:768–73.

[19] Sato Y, Sugie R, Tsuchiya B, Kameya T, Natori M, Mukai K. Comparison of the DNA extraction methods for polymerase chain reaction amplification from formalin-fixed and paraffin-embedded tissues. Diagn Mol Pathol 2001;10:265–71.

[20] Cao W, Hashibe M, Rao J-Y, Morgenstern H, Zhang Z-F. Comparison of methods for DNA extraction from paraffin-embedded tissues and buccal cells. Cancer Detect Prev 2003;27:397–404.

[21] Pepinski W, Soltyszewski I, Janica J, Skawonska M, Koc-Zorawska E. Comparison of five commercial kits for DNA extraction from human blood, saliva and muscle samples. Rocz Akad Med Bialymst 2002;47: 270–5.

[22] Glasel JA. Validity of nucleic acid purities monitored by 260 nm/280 nm absorbance ratios. Biotechniques 1995;18:62–3.

[23] Wilson IG. Inhibition and facilitation of nucleic acid amplification. Appl Environ Microbiol 1997;63:3741–51.