Award Number: W81XWH-08-1-0383

TITLE: A Genome-wide Breast Cancer Scan in African Americans

PRINCIPAL INVESTIGATOR: Christopher A. Haiman

CONTRACTING ORGANIZATION: University of Southern California, Los Angeles, CA 90033

REPORT DATE: June 2010

TYPE OF REPORT: Annual

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

DISTRIBUTION STATEMENT: (Check one)

X Approved for public release; distribution unlimited

□ Distribution limited to U.S. Government agencies only; report contains proprietary information

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

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

1. REPORT DATE *(DD-MM-YYYY)* 1 June 2010	**2. REPORT TYPE** Annual	**3. DATES COVERED** *(From - To)* 1 June 2009 – 31 May 2010

4. TITLE AND SUBTITLE
A Genome-wide Breast Cancer Scan in African Americans

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-08-1-0383

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Christopher A. Haiman

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

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

University of Southern
California
Los Angeles, California
90033

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Material 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

This genome-wide association of breast cancer includes over 3,000 African American women with invasive disease and over 3,000 age-matched African American controls from many existing case-control studies of breast cancer in the U.S. In the first year of this project all DNA samples from these studies were sent to Dr. Haiman's laboratory at the University of Southern California, quantitated and arrayed for genotyping. We have also assembled the covariate file that contains established breast cancer risk factor data collected from these studies, and the data have been standardized for the analysis. As of April 1, 2009, we have genome-wide scanned 2,200 cases and 2,200 controls from a number of the participating studies using the Illumina 1M Beadchip. We have assessed the blinded quality control replicates and overall call rates by SNP and sample and the data appear to be of very high quality. We have conducted a very preliminary statistical analysis of the data and there are a number of promising signals including what we believe is a novel risk locus for breast cancer that may be of particular importance for estrogen receptor negative breast cancer. The analysis will continue as additional samples/studies are scanned over the next year.

15. SUBJECT TERMS
Breast cancer, estrogen receptor negative, genome-wide association study

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** USAMRMC
a. REPORT U	**b. ABSTRACT** U	**c. THIS PAGE** U	UU	15	**19b. TELEPHONE NUMBER** *(include area code)*

Table of Contents

BODY

The Specific Aim of this application is to identify genetic risk alleles for breast cancer among African American women by performing a well-powered genome-wide association study (GWAS). For this project, I have established a network of leaders in the breast cancer research community with long-standing interests in breast cancer research in African Americans, all of whom have existing case-control studies of breast cancer in the U.S. Funding for the genotyping of samples from the MEC, CARE, WCHS, SFBC and BCFR studies is covered by this DOD-BCRP grant. The genotyping of the other studies has been provided by a number of other sources. All together these studies include over 5,000 African American women with invasive breast cancer and over 5,000 age-matched African American controls with available DNA.

Stage 1 of the GWAS included 9 epidemiological studies of invasive breast cancer among African American women, which comprise a total of 3,153 cases and 2,831 controls. Replication testing of the top associations is being conducted (using other sources) in an independent sample of >2000 African American breast cancer cases and >2000 controls from three additional studies of breast cancer (discussed below). Below is a brief description of these 12 studies.

Stage 1 Studies:

The Multiethnic Cohort Study (MEC): The MEC is a prospective cohort study of 215,000 men and women in Hawaii and Los Angeles between the ages of 45 and 75 years at baseline (1993-1996). Through December, 31 2007, a nested breast cancer case-control study in the MEC included 556 African American cases (544 invasive and 12 in situ) and 1,003 African American controls. An additional 178 African American breast cancer cases (ages: 50-84) diagnosed between June 1, 2006 and December 31, 2007 in Los Angeles County (but outside of the MEC) were combined with the MEC samples in the analysis.

The Los Angeles component of The Women's Contraceptive and Reproductive Experiences (CARE) Study: The NICHD Women's CARE Study is a large multi-center population-based case-control study that was designed to examine the effects of oral contraceptive (OC) use on invasive breast cancer risk among African American women and white women ages 35-64 years in five U.S. locations. Cases in Los Angeles County were diagnosed from July 1, 1994 through April 30, 1998, and controls were sampled by random-digit dialing (RDD) from the same population and time period; 380 African American cases and 224 African American controls were included in stage 1 of the scan.

The Women's Circle of Health Study (WCHS): The WCHS is an ongoing case-control study of breast cancer among European women and African American women in the New York City boroughs (Manhattan, the Bronx, Brooklyn and Queens) and in seven counties in New Jersey (Bergen, Essex, Hudson, Mercer, Middlesex, Passaic, and Union). Eligible cases included women with invasive breast cancer between 20 and 74 years of age; controls were identified through RDD. The WCHS contributed 272 invasive African American cases and 240 African American controls to stage 1 of the GWAS.

The San Francisco Bay Area Breast Cancer Study (SFBC): The SFBC is a population-based case-control study of invasive breast cancer in Hispanic, African American and non-Hispanic White women conducted between 1995 and 2003 in the San Francisco Bay Area. African American cases, ages 35-79 years, were diagnosed between April 1, 1995 and April 30, 1999, with controls identified through RDD. Stage 1 included 172 invasive African American cases and 231 African American controls from SFBC.

The Northern California Breast Cancer Family Registry (NC-BCFR): The NC-BCFR is an on-going population-based family study conducted in the Greater San Francisco Bay Area, and is one of 6 sites collaborating in the Breast Cancer Family Registry (BCFR), an international consortium funded by NCI. African American breast cancer cases in NC-BCFR were diagnosed after January 1, 1995 and between the ages of 18 and 64 years; population controls were identified through RDD. Stage 1 genotyping was conducted for 440 invasive African American cases and 53 African American controls.

The Carolina Breast Cancer Study (CBCS): The CBCS is a population-based case-control study conducted between 1993 and 2001 in 24 counties of central and eastern North Carolina. Cases were identified by rapid case ascertainment system in cooperation with the North Carolina Central Cancer Registry and controls were selected from the North Carolina Division of Motor Vehicle and United States Health Care Financing Administration beneficiary lists. Participants' ages ranged from 20 to 74 years. For stage 1, DNA samples were provided from 656 African American cases with invasive breast cancer and 608 African American controls.

The Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial (PLCO) Cohort: PLCO, coordinated by the U.S. National Cancer Institute (NCI) in 10 U.S. centers, enrolled during 1993 - 2001 approximately 155,000 men and women, aged 55-74 years, in a randomized, two-arm trial to evaluate the efficacy of screening for these four cancers. A total of 64 African American invasive breast cancer cases and 133 African American controls contributed to stage 1 of the GWAS.

The Nashville Breast Health Study (NBHS): The NBHS is a population-based case-control study of incident breast cancer conducted in Tennessee. The study was initiated in 2001 to recruit patients with invasive breast cancer or ductal carcinoma in situ, and controls, recruited through RDD between the ages of 25 and 75 years. NBHS contributed 310 African American cases (57 in situ), and 186 African American controls to stage 1 of the GWAS.

Wake Forest University Breast Cancer Study (WFBC): African American breast cancer cases and controls in WFBC were recruited at Wake Forest University Health Sciences from November 1998 through December 2008. Controls were recruited from the patient population receiving routine mammography at the Breast Screening and Diagnostic Center. Age range of participants was 30-86 years. WFBC contributed 125 cases (116 invasive and 9 in situ) and 153 controls to the stage 1 analysis.

Replication Studies:

The Black Women's Health Study (BWHS): The BWHS is a prospective cohort of 59,000 African-American women 21-69 years of age at baseline (1995) from across the U.S. The present analysis includes invasive breast cancer cases diagnosed before 2008 and age-matched controls who provided a mouthwash sample. A total of 790 cases and 1,119 controls are contributing to the replication analysis.

The Women's Insights and Shared Experiences study (WISE): The WISE study is a population-based case-control study in three counties, Philadelphia, PA, Delaware, DE and Camden, NJ. Potentially eligible cases and controls were Caucasian and African American women who ranged in age from 50-79 years. Invasive breast cancer cases were diagnosed between July 1, 1999 and June 30, 2002. The data available to the replication analysis included 145 African American breast cancer cases and 367 age-group matched RDD African American controls.

The Southern Community Cohort (SCCS): The SCCS is a prospective cohort investigation initiated in 2001 enrolling residents aged 40-79 years across 12 southern states. Both African and non-African Americans are included to enable a direct comparison of information obtained in identical ways for different racial/ethnic groups, with more than twice as many African Americans enrolled to help address under-representation of blacks in previous epidemiologic studies of cancer. While recruitment is still ongoing, it is projected that the final cohort size will reach approximately 90,000 in 2009. Each study participant has (or will have) completed a detailed baseline questionnaire, and to date nearly 90% have provided a biologic specimen (approximately 45% a blood sample and 45% buccal cells). Follow-up of the cohort is conducted by linkage to national mortality registers and with linkage to state cancer registries. Prevalent African American breast cancer cases (>500) and African American women without breast cancer (>1000) will contribute to the replication phase.

Genotyping and Quality Control

Genotyping in stage 1 was conducted using the Illumina Human1M-Duo BeadChip. Of the 5,984 samples from these studies (3,153 cases and 2,831 controls), we attempted genotyping of 5,932, removing samples (n=52) with DNA concentrations <20 ng/ul. Following genotyping, we removed samples based on the following exclusion criteria: 1) unknown replicates (≥98.9% genetically identical) that we were able to confirm (only one of each duplicate was removed, n=15); 2) unknown replicates that we were not able to confirm through discussions with study investigators (pair or triplicate removed, n=14); 3) samples with call rates <95% after a second attempt (n=100); 4) samples with ≤ 5% African ancestry (n=36); and, 5) samples with <15% mean heterozygosity of SNPs in the X chromosome and/or similar mean allele intensities of SNPs on the X and Y chromosomes (n=6) (these are likely to be males). In the analysis, we removed SNPs with <95% call rate (n=21,732) or minor allele frequencies (MAFs) <1% (n=80,193). To assess genotyping reproducibility we included 138 replicate samples; the average concordance rate was 99.95% (>99.93% for all pairs). We also eliminated SNPs with genotyping concordance rates <98% based on the replicates (n=11,701). The final analysis dataset included 1,043,036 SNPs genotyped on 3,016 cases (1,520 ER+, 988 ER- and the remaining 508 cases with unknown ER status) and 2,745 controls, with an average SNP call rate of 99.7% and average sample call rate of 99.8%.

Statistical Analysis of the Stage 1 Data

In stage 1, we utilized STRUCTURE (1) to infer percent African ancestry on an individual level. A total of 2,546 ancestry-informative SNPs from the Illumina array were selected based on low inter-marker correlation and ability to differentiate between samples of African and European descent (Nick Patterson, personal communication). In evaluating the distribution of the fraction of African ancestry across the nine study populations (range across studies: 73%-82% African ancestry), statistically significant differences (ANOVA $p<10^{-16}$) were noted. We also applied principal components analysis (PCA) (2) to estimate axes of variation among the 5,761 individuals using the same 2,546 ancestry informative markers. The first eigenvector accounted for 10.1% of the variation between subjects, and subsequent eigenvectors accounted for no more than 0.5%. Using input genotypes from the HapMap populations, CEU (CEPH Utah), YRI (Yoruba), and JPT (Japanese), we could demonstrate that the first eigenvector captures most of the genetic variation in the stage 1 sample as well as that between Europeans and West Africans in the HapMap samples. In the single SNP analyses, we examined the observed versus the expected distribution of the Chi squared test statistics from the 1 degree of freedom (df) trend test, comparing genotype counts in all cases and controls, and stratified by the ER status of the tumor. To improve coverage, we augmented the set of SNPs tested for association through imputation using MACH (3). Phased haplotypes from the African American (ASW) population of HapMap Phase 3 were used to infer all Phase 3 genotypes available in this population which corresponds to the SNPs on either of the Illumina 1M or Affymetrix 6.0 arrays. For each imputed SNP, we opted to make use of genotype dosage scores, which are coded as continuous variables, thus implicitly incorporating uncertainty to help improve statistical power, control bias in risk estimation, and improve confidence interval estimation (relative to using the "most likely" genotype). Odds ratios (OR) and 95% confidence intervals (CI) for each SNP were estimated using unconditional logistic regression, adjusting for age at diagnosis, the first eigenvector and study. For each SNP we tested significance through a one degree of freedom (df) Wald chi-square test.

Stage 1 Results

As presented in the previous progress report (2009), the quantile-quantile plot demonstrates only a very slight excess of associations at $p<0.05$ than what would be expected based on chance, especially in the 10^{-3} to 10^{-5} range. No SNP association was observed at the genome-wide level of significance ($p<10^{-7}$). We are currently following-up the 100 top associations in the replication studies (in addition to the chromosome 5q31 locus described in the 2009 progress report). We expect the genotyping to be complete in the Fall of 2010.

Testing of Known Breast Cancer Risk Variants

Using the African American cases and controls from stage 1, we also examined associations with 11 loci identified in previous GWAS in populations of European ancestry (4-8) and 1 in Chinese (9). We had good statistical power (range: >99% - 69%) to detect nominally significant associations ($p<0.05$) for eight of the twelve established risk variants. Ten of the markers were on the Illumina Human1M-Duo BeadChip, whereas close proxies were identified for the other two loci (6q25 and 17q22). Only two variants (Table 1) were found to be nominally associated with risk among African Americans (rs13387042 at 2q35: OR=1.12, $p=8.8x10^{-3}$; and, rs10736303 (*FGFR2)* on 10q26, OR=1.15; $p=8.3x10^{-3}$). These data suggest that most of the index signals at the established breast cancer GWAS loci may be poor risk markers for African Americans.

Meta-Analyses with other Breast Cancer GWAS

We have also initiated meta-analyses with other GWAS including a GWAS of triple negative breast cancer (PI, Fergus Couch), a GWAS of bilateral breast cancer (Julian Peto), and a GWAS of estrogen receptor negative breast from the NCI Breast and Prostate Cancer Cohort Consortium (BPC3). SNPs from our scan in African American women that replicate in these additional studies among women of European ancestry will be targeted for follow-up in other studies (such as the African American replication studies mentioned above). In addition, we have provided the top 5000 SNPs from the African American breast cancer scan to COGs, a European collaborative study which is has designed a SNP array with that will be genotyped in 10,000's of breast cancer cases and 10,000's of controls. This collaboration will allow for very large-scale replication of findings from our African American breast cancer scan.

Table 2. Associations with known risk variants for breast cancer in African Americans.

Chr. Index SNP (position)[a]	Risk Allele, Freq (EA/CH,AA)[b]	OR per allele[c] (95% CI)	P_{1df}
1p11			
rs11249433 (120982136)	G (0.43,0.13)	0.98(0.87-1.10)	0.75
2q35			
rs13387042 (217614077)	A (0.57,0.73)	1.12(1.03-1.22)	$8.8x10^{-3}$
3p24			
rs4973768 (27391017)	T (0.44,0.37)	1.04(0.96-1.12)	0.29
5p12			
rs4415084 (44698272)	T (0.38,0.63)	1.03(0.95-1.11)	0.43
5q11			
rs16886165 (56058840)	G (0.16,0.33)	1.15(1.06-1.24)	$7.6x10^{-4}$
6q25			
rs2046210[d] (151990059)	A (0.37,0.63)	1.01(0.93-1.09)	0.86
8q24			
rs13281615 (128424800)	G (0.46,0.44)	1.05(0.97-1.13)	0.24
10q26			
rs10736303[e] (123324447)	G (0.55,0.86)	1.25(1.11-1.41)	$3.4x10^{-4}$
11p15			
rs3817198 (1865582)	C (0.33,0.17)	0.97(0.88-1.08)	0.61
14q24			
rs999737 (68104435)	T (0.24,0.95)	1.01(0.85-1.21)	0.89
16q12			
rs3803662 (51143842)	A (0.30,0.51)	0.99(0.92-1.07)	0.82
17q22			
rs7222197[f] (50422498)	G (0.68,0.66)	1.06(0.98-1.15)	0.12

[a] Based on NCBI build 36. [b] Estimated for European Ancestry (EA) in CEU HapMap samples and among African American (AA) controls in stage 1, except for rs2046210 which is reported for Chinese (CH) in Zheng et al. (9). [c] Adjusted for age, study and the first principle component. [d] Imputed. [e] Imputed surrogate for rs2981578; r^2=1.0 in African Americans. [f] Surrogate for rs6504950; r^2=0.99 in ASW HapMap sample.

Multiethnic Fine-Mapping

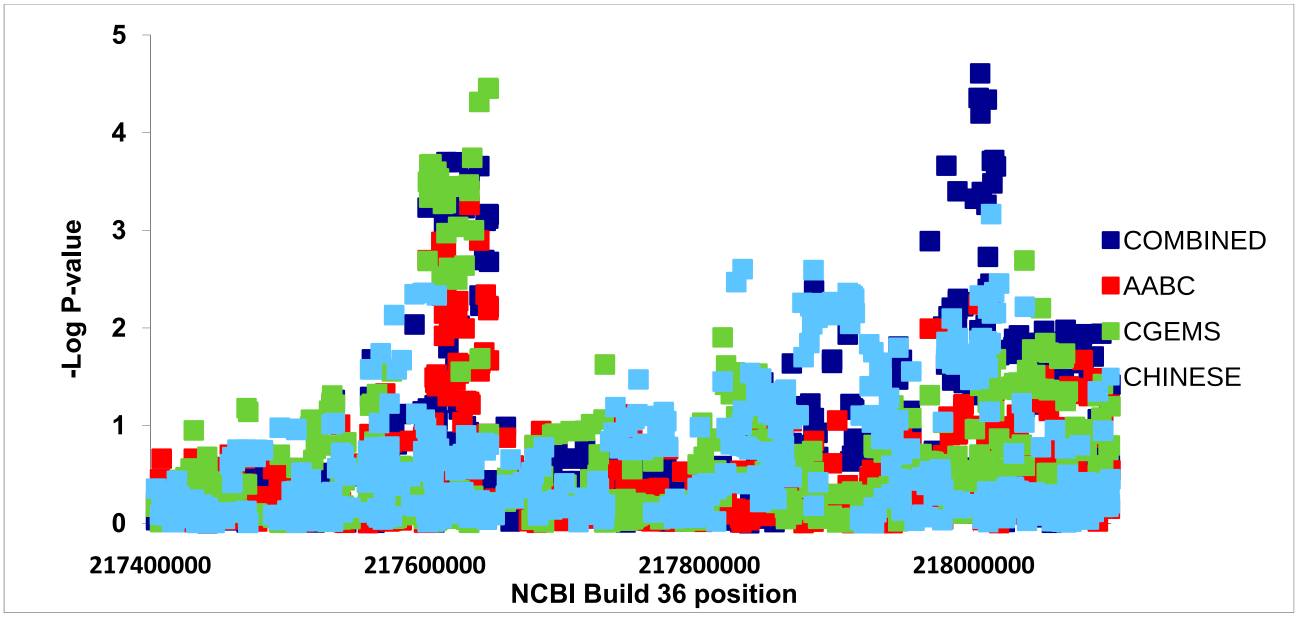
Genome-wide association studies (GWAS) have been conducted primarily in single racial/ethnic populations. These discovery efforts have focused on discrete, rather than multiple populations of different ancestries, to avoid false-positive signals and the potential masking of real associations resulting from population differences in allele frequencies, patterns of linkage disequilibrium (LD) and genetic and environmental modifiers. GWAS and fine-mapping studies that include multiple racial and ethnic populations allow for the assessment of a much wider spectrum of genetic variation and offer great power to detect risk variants that are common globally. We have combined our GWAS data in African Americans with existing breast cancer GWAS data among Chinese and European American women and conducted multiethnic mapping to search for novel risk variants at the known breast cancer susceptibility loci. Below are our results for the risk locus on chromosome 2q35.

A GWAS among Icelandic individuals conducted by deCODE Genetics (5) initially identified an association with variant rs13387042 (position 217,614,077 NCBI Build 36; OR=1.2 per copy of the A allele) on chromosome 2q35 that has since been replicated in a number of additional studies in diverse samples. The variant is located in an 85 kb LD block, with the nearest known genes being *TNP1*, *IGFBP2* and *IGFBP1* located between 181 and 376 kb centromeric to the index signal and *TNS1* 781 kb telomeric. We conducted a meta-analysis of genotype data spanning this risk locus from three existing breast cancer scans. These studies included our scan among African American women (AABC) of 3,153 invasive cases and 2,831 controls genotyped with the Illumina Human-1M Duo, a scan among Chinese women of 2,050 invasive cases and 2,065 controls genotyped with the Affymetrix Genome-Wide Human SNP Array 6.0 (9), and the NCI CGEMS scan among 1,145 invasive cases and 1,142 controls women of European ancestry which utilized the Illumina HumanHap550 array (7). In each study, imputation to Phase 2 HapMap was performed to create a dense map for testing common alleles that are shared between populations. The boundaries for association testing were arbitrarily defined as ~500 kb upstream and downstream of the index signal (217,115,543-218,113,654). This map included 854 SNPs, with 711 being common (MAF≥0.05) in all three populations (1 common SNP per 1.4 kb, on average). These SNPs capture (r^2≥0.80) most of the common alleles (≥5% frequency) in the Phase 2 HapMap (93% CEU, 92% in JPT/CHN, and 76% in YRI; mean r^2 all >0.99).

In the combined sample (6,151 cases and 5,903 controls) we observed evidence of association at the known risk region (region 1; LD block 217.565-217.650 Mb), with 15 SNPs with p-values at 10^{-4}, including the previously published variant (rs13387042, p=5.6x10^{-4}). The associations with these SNPs were much stronger in African American and Europeans, than in Chinese. The risk-associated markers were correlated (r^2≥0.25, CEU HapMap) and no secondary signals were noted with these 15 markers when conditioning on rs13387042.

We observed a second region of association (region 2) located ~400 kb centromeric to region 1 (Figure 1), with 13 SNPs having p-values between 10^{-4} and 10^{-5}. These SNPs reside in a separate LD block of ~160 kb that is located ~330 kb from the nearest gene, *TNS1*. The associations with these SNPs were generally consistent and nominally significant in most populations. The 13 SNPs were correlated (r^2≥0.35 in CEU HapMap) and not linked to the markers in region 1 (r^2≤0.1 in CEU HapMap). We are currently genotyping the most associated SNPs in region 2 in additional studies of invasive breast cancer in populations of European, Asian and African ancestry.

Figure 1. Fine-mapping of the 2q35 risk locus in multiple populations. Shown are -log p-values of SNPs examined by population and in the combined sample. The known risk locus (region 1) is located at 217.565-217.650 Mb and the novel risk region (region 2) at 217.948-218.106 Mb.



Other Ongoing Statistical Analyses

Enhancing the interpretation of GWAS results through local ancestry information

Software developed by our collaborators at the Broad Institute in Boston (Alkes Price) has allowed us to better understand the genetic basis behind racial disparities in the prevalence of ER/PR negative forms of breast cancer. The program *ancestrymap* was designed to detect regions of the genome that harbored a greater proportion of alleles from one ancestral population than the expected proportion (averaged across the genome). Although designed for coarse marker panels, we applied the method on our African American breast cancer dataset, revealing a wide peak of local ancestry deviation at Chromosome 10 in the region of FGFR2, a known breast cancer susceptibility locus. We have recently applied a newer program called *hapmix* that is optimized for use in high-density SNP panels (e.g. Illumina Infinium products) to the same dataset. We have narrowed the peak at Chromosome 10 and revealed other smaller peaks elsewhere. Currently, we are applying a method that is still under development at the Broad with the goal of combining signals from a case-only admixture mapping method (i.e., output from *hapmix*) with p-values from a GWAS that uses cases and controls. The method is expected to yield greater power over one that relies on conventional GWAS p-values since the statistic is based on a 1 df test that uses additional information about local ancestry.

Discovery of pathway-level gene or interaction (e.g. gene by gene, gene by environment) effects

Often, GWAS scans have generated lists of small p-values, many of which have failed to replicate. In some cases, SNPs that do replicate are ranked very low in the initial first stage scan, leading investigators to relax stringent alpha levels in order to reduce the chances of missing a true positive. We have recognized the need to develop creative ideas that can prioritize SNPs for further analysis. A reasonable approach is to use biological evidence to rank SNPs through an approach such as the weighted FDR (10), or though semi-empirical Bayes approaches (11, 12). We have developed this idea further by incorporating prior knowledge in a fully Bayesian variable selection method. This approach allows tests for association to be conducted in a multivariate analysis (so independent main effects can be discovered), with priors used to help steer the algorithm towards more biologically plausible models (so that not all possible interactions need to be tested). A full Bayesian treatment also addresses multiple comparisons issue gracefully. A grant application that we are currently developing will propose the use of gene-expression data towards defining pathway modules. Pathway modules will aid our novel algorithm when applied on our stage 1 African American GWAS dataset.

Key Research Accomplishments

- Completing the largest single stage 1 genome-wide association study in any racial/ethnic population using the Illumina 1M technology.
- Identifying a select set of top associations with breast cancer risk and collaborators willing to assist in replicating findings.
- Conducting a large-scale multiethnic fine-mapping analysis of known breast cancer loci.
- Contributing to the development and application of novel analytic methods for combining GWAS and admixture-scan data in the admixed African American population.

Reportable Outcomes

- The findings presented in this progress report were discussed at the American Association of Cancer Research annual conference in Denver, CO (April, 2009) as well as the Breast Cancer Association Consortium meeting in New York (October, 2009).

Conclusion

The stage 1 analyses have presented many exciting findings that we are actively pursuing through collaboration with new investigators with studies of breast cancer in African Americans. We are also working with many investigators leading GWAS of breast cancer in populations of European ancestry. It is clear that groups need to work together in the discovery phase as we are trying to identify risk variants with subtle effects on disease risk and very large sample sizes are needed. Together these large-scale meta-analyses and replication efforts have great potential to reveal novel risk loci for breast cancer.

A future direction of our work will be to combine the GWAS data for multiple minority populations (as we have done for the chromosome 2 locus except genome-wide) as part of the NCI-supported Post-GWAS effort. Given our large scan in African American women, we hope to be in a position to lead this effort.

Revealing the genetic causes of breast cancer *in each population* will in time translate into more targeted preventive measures and treatment strategies for those at risk of developing the breast cancer.

References

1. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics 2000;155(2):945-59.
2. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 2006;38(8):904-9.
3. Li Y, Abecasis GR. Mach 1.0: Rapid Haplotype Reconstruction and Missing Genotype Inference. Am J Hum Genet 2006;S79:2290.
4. Easton DF, Pooley KA, Dunning AM, Pharoah PD; et al. Genome-wide association study identifies novel breast cancer susceptibility loci. Nature. 2007; 447 (7148): 1087-93
5. Stacey SN, Manolescu A, Sulem P, Rafnar T; et al. Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. Nat. Genet. 2007; 39(7):865-9
6. Stacey SN, Manolescu A, Sulem P, Thorlacius S; et al. Common variants on chromosome 5p12 confer susceptibility to estrogen receptor-positive breast cancer. Nat. Genet. 2008; 40 (6): 703-6
7. Thomas G, Jacobs KB, Kraft P, Yeager M; et al. A multistage genome-wide association study in breast cancer identifies two new risk alleles at 1p11.2 and 14q24.1 (RAD51L1). Nat. Genet. 2009; 41(5):579-84
8. Ahmed S, Thomas G, Ghoussaini M, Healey CS; et al. Newly discovered breast cancer susceptibility loci on 3p24 and 17q23.2. Nat. Genet. 2009; 41(5):585-90
9. Zheng W, Long J, Gao YT, Li C; et al. Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. Nat. Genet. 2009;41(3):324-8
10. Roeder K, Bacanu SA, Wasserman L, DevlinB. Using linkage genome scans to improve power of association in genome scans. Am J Hum Genet. 2006;78 (2): 243-52
11. Lewinger JP, Conti DV, Baurley JW, Triche TJ, Thomas DC. Hierarchical Bayes prioritization of marker associations from a genome-wide association scan for further investigation. Genet Epidemiol. 2007; 31 (8):871-82
12. Chen GK, Witte Js. Enriching the analysis of genomewide association studies with hierarchical modeling. Am J Hum Genet. 2007; 81(2):397-404

Appendices

NA