An Admixture Scan in 1,484 African American Women with Breast Cancer

Laura Fejerman,[1] Christopher A. Haiman,[2] David Reich,[3,5] Arti Tandon,[3,5] Rahul C. Deo,[3,4] Esther M. John,[6,7] Sue A. Ingles,[2] Christine B. Ambrosone,[9] Dana Howard Bovbjerg,[10] Lina H. Jandorf,[11] Warren Davis,[9] Gregory Ciupak,[9] Alice S. Whittemore,[7] Michael F. Press,[8] Giske Ursin,[2,13] Leslie Bernstein,[12] Scott Huntsman,[1] Brian E. Henderson,[2] Elad Ziv,[1] and Matthew L. Freedman[5,14]

[1]Division of General Internal Medicine, Department of Medicine, Institute for Human Genetics and Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, California; [2]Department of Preventive Medicine, Keck School of Medicine, University of Southern California/Norris Comprehensive Cancer Center, Los Angeles, California; [3]Department of Genetics, Harvard Medical School, Boston, Massachusetts; [4]Cardiology Division, Department of Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts; [5]Program in Medical and Population Genetics, Broad Institute of Harvard and Massachusetts Institute of Technology, Boston, Massachusetts; [6]Northern California Cancer Center, Fremont, California; [7]Department of Health Research and Policy, Stanford University School of Medicine, Stanford, California; [8]Department of Pathology, Keck School of Medicine, University of Southern California/Norris Comprehensive Cancer Center, Los Angeles, California; [9]Department of Cancer Prevention and Control, Roswell Park Cancer Institute, Buffalo, New York; [10]University of Pittsburgh Cancer Institute, Pittsburgh, Pennsylvania; [11]Department of Oncological Sciences, Mount Sinai School of Medicine, New York, New York; [12]Beckman Research Institute, City of Hope, Duarte, California; [13]Department of Nutrition, University of Oslo, Oslo, Norway; and [14]Department of Medicine, Harvard Medical School and Dana-Farber Harvard Cancer Center, Boston, Massachusetts

Abstract

African American women with breast cancer present more commonly with aggressive tumors that do not express the estrogen receptor (ER) and progesterone receptor (PR) compared with European American women. Whether this disparity is the result of inherited factors has not been established. We did an admixture-based genome-wide scan to search for risk alleles for breast cancer that are highly differentiated in frequency between African American and European American women, and may contribute to specific breast cancer phenotypes, such as ER-negative (ER–) disease. African American women with invasive breast cancer (n = 1,484) were pooled from six population-based studies and typed at ~1,500 ancestry-informative markers. We investigated global genetic ancestry and did a whole genome admixture scan searching for breast cancer–predisposing loci in association with disease phenotypes. We found a significant difference in ancestry between ER+PR+ and ER–PR– women, with higher European ancestry among ER+PR+ individuals, after controlling for possible confounders (odds ratios for a 0 to 1 change in European ancestry proportion, 2.84; 95% confidence interval, 1.13-7.14; P = 0.026). Women with localized tumors had higher European ancestry than women with non–localized tumors (odds ratios, 2.65; 95% confidence interval, 1.11-6.35; P = 0.029). No genome-wide statistically significant associations were observed between European or African ancestry at any specific locus and breast cancer, or in analyses stratified by ER/PR status, stage, or grade. In summary, in African American women, genetic ancestry is associated with ER/PR status and disease stage. However, we found little evidence that genetic ancestry at any one region contributes significantly to breast cancer risk or hormone receptor status. (Cancer Epidemiol Biomarkers Prev 2009;18(11):3110–7)

Introduction

Breast cancer incidence and mortality varies widely among women of different population groups in the United States. African American women have lower age-adjusted incidence of breast cancer compared with European Americans (1). However, breast cancer incidence is higher in African Americans who are 35 years of age or younger (2). African American women are also diagnosed, on average, with later stage of disease, larger

Received 5/19/09; revised 8/12/09; accepted 9/1/09; published OnlineFirst 10/20/09.

Grant support: Postdoctoral fellowship from the Prevent Cancer Foundation (L. Fejerman). The Multiethnic Cohort Study was supported by National Cancer Institute grants CA63464 and CA54281. Komen Foundation Investigator-Initiated grant KG080165 (E. Ziv). The Women's Circle of Health Study is supported by grants NCI R01 CA100598 and DAMD17-01-1-0334. The San Francisco Bay Area Breast Cancer Study was funded by grants from the National Cancer Institute (R01 CA63446 and R01 CA77305) and the U.S. Army Medical Research Program (DAMD17-96-6071; E.M. John). Richard and Susan Smith Family Foundation, Chestnut Hill, MA (M. Freedman). Burroughs Wellcome Career Development Award in the Biomedical Sciences (D. Reich). The Los Angeles Component of the Women's Contraceptive and Reproductive Experiences Study was supported by a contract from the National Institute of Child Health and Human Development (N01-HD-3-3175; L. Bernstein). The Breast Cancer Family Registry (BCFR) is funded by the National Cancer Institute, NIH under RFA-CA-06-503 and through cooperative agreements with members of the BCFR and Principal Investigators. The registry contributing data to this analysis was supported by U01 CA69417 (Northern California Cancer Center). The content of this manuscript does not necessarily reflect the views or policies of the National Cancer Institute or any of the collaborating centers in the BCFR, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government or the BCFR.

Note: Supplementary data for this article are available at Cancer Epidemiology, Biomarkers & Prevention Online (http://cebp.aacrjournals.org/).

L. Fejerman and C.A. Haiman contributed equally to this work.

Requests for reprints: Elad Ziv, Division of General Internal Medicine, University of California, San Francisco, Box 1732, San Francisco, CA 94143. Phone: 415-353-7981; Fax: 415-514-4982. E-mail: elad.ziv@ucsf.edu and Matthew L. Freedman, Dana-Farber Cancer Institute, Medical Oncology, Boston, MA 02115. E-mail: freedman@broad.mit.edu

doi:10.1158/1055-9965.EPI-09-0464

tumors, and are more likely to present with lymph node metastases at the time of diagnosis (2, 3). Thus, despite the lower lifetime incidence of breast cancer among African American women compared with European American women, their breast cancer mortality rates are higher (4-6), particularly among younger women (7).

The expression of steroid hormone receptors (estrogen and progesterone receptors) in breast cancer tumors also varies substantially by population. African American women are diagnosed more frequently with estrogen receptor–negative (ER–) and progesterone receptor–negative (PR–) breast cancer compared with European American women (5, 7-9). In the Women's Health Initiative, 32% of breast cancers among postmenopausal African American women were ER– with poor/anaplastic grade in comparison to only 10% among European American women, a difference which remained after adjustment for multiple potentially confounding factors, including differential access to health care (10). Given the greater incidence of hormone receptor–negative, high-grade disease among African Americans, we hypothesized that there may be one or more genetic variants with increased frequency in populations of African origin, which predispose women to this more aggressive form of breast cancer.

Admixture mapping is a powerful approach for identifying genetic variants for common phenotypes that have large allele frequency differences between ancestral populations (11-14). Admixed populations are defined as populations in which two or more ancestral groups have been mixing over several generations. Recently admixed populations show extended linkage disequilibrium between markers that have a large difference in allele frequency between ancestral populations and are, therefore, informative about ancestry (ancestry-informative markers or "AIM"; refs. 13, 15). The principle of admixture mapping is to identify regions of the genome with greater estimated ancestry from one of the ancestral populations than the chromosomal average in individuals from an admixed group. These regions may highlight candidate risk loci that are associated with complex phenotypes. We have previously used this approach to identify risk variants for prostate cancer at 8q24 that are common in African American men and contribute to their increased disease incidence (16).

Here, we did an admixture-based genome-wide scan in 1,484 African American women with invasive breast cancer pooled from six population-based studies. Samples were typed at ~1,500 AIMs to search for loci that might harbor predisposing variants for breast cancer, and more specifically, loci that may contribute to specific breast cancer phenotypes, such as ER– disease, a trait which is more common in African American women.

Materials and Methods

Samples. This analysis includes samples from six population-based breast cancer studies described in brief below.

The Multiethnic Cohort Study. This study is a prospective cohort that includes >215,000 individuals from Hawaii and California (primarily Los Angeles) that was assembled between 1993 and 1996 (17, 18). The cohort is comprised predominantly of African Americans, Native Hawaiians, Japanese, Latinos, and European Americans. Beginning in 1994, blood samples were collected from incident breast cancer cases identified by cohort linkage to Surveillance, Epidemiology and End Results (SEER) registries, as well as a random sample of Multiethnic Cohort participants to serve as controls for genetic analyses. The present study includes 423 invasive African American breast cancer cases from the Multiethnic Cohort, ages 45 to 82 y at diagnosis.

The Los Angeles Component of the Women's Contraceptive and Reproductive Experiences Study. A population-based case control study that included African American and Caucasian women with invasive breast cancer and control subjects, ages 35 to 64 y (19). Incident cases diagnosed between 1994 and 1998 were identified by the Los Angeles SEER registry. This study contributed 384 invasive African American breast cancer cases to the scan.

The Learning the Influence of Family and the Environment Study. This study included invasive African American breast cancer cases from Los Angeles county, ages 20 to 49 y (20). Incident cases diagnosed between 2000 and 2003 were identified from the Los Angeles SEER registry. In the current study, we used DNA samples obtained from 140 invasive cases.

The Women's Circle of Health Study. This study included African American women, 20 to 65 y of age, newly diagnosed with a first primary, histologically confirmed breast cancer. Cases were identified from major metropolitan hospitals in New York City serving a large minority population, and from the eight counties in New Jersey bordering the Hudson River. The present study includes 194 invasive breast cancer cases.

The San Francisco Bay Area Breast Cancer Study. A population-based case-control study of breast cancer in Hispanic, African American, and non-Hispanic white women (21, 22). Incident cases of invasive breast cancer ages 35 to 79 y were identified through the Greater Bay Area Cancer Registry. The present analysis includes 191 African American breast cancer cases diagnosed between 1997 and 1999.

Northern California Site of the Breast Cancer Family Registry. The Breast Cancer Family Registry is an international collaboration of six academic and research institutions, established in 1995 with support from the U.S. National Cancer Institute to serve as a resource for genetic studies of breast cancer (23). The California site enrolled newly diagnosed breast cancer cases ages <65 y that were identified through the Greater Bay Area Cancer Registry. The present study includes 314 unrelated African American breast cancer cases diagnosed between 1995 and 2003.

Genotyping. Invasive breast cancer cases in these six studies (1,646) were genotyped for two AIM panels using the Illumina GoldenGate assay (each panel consisting of 1,536 AIMs). The Women's Circle of Health Study, San Francisco Bay Area Breast Cancer Study, and the Breast Cancer Family Registry samples (set 1, $n = 699$) were genotyped at the University of California, San Francisco with a phase 2 panel, which was first published by Reich et al. (24). From this panel, 196 markers were dropped because of failure and replaced with 196 additional markers (phase 2 panel version b; Supplementary Table S1, 196 new SNPs are highlighted). A set of markers was

selected based on allele frequency differences in West Africans from London and Europeans from Centre d'Etude du Polymorphisme Humain and they were scored by the Illumina snp_score, which predicts how well the markers will be genotyped. Fst and δ values (two measures of allele frequency difference between populations) were calculated for the markers. A total of 196 evenly spaced markers with the top scores for the Illumina snp_score and with the highest Fst (>0.4) and δ values (>0.6) were selected to include in the new phase 2 panel version b. The Multiethnic Cohort, Women's Contraceptive and Reproductive Experiences (CARE), and Learning the Influence of Family and the Environment (LIFE) studies (set 2, $n = 947$) were genotyped at the University of Southern California Genomics Core Laboratory with a phase 3 AIM panel.[15]

We genotyped the 1,646 samples for a total of 2,427 AIMs. For each set, we removed samples and SNPs that did not pass our quality control criteria. We removed samples with missing histology (set 1, $n = 88$; set 2, $n = 0$) and those with low call rates (defined as <85%) or that showed genotypes that are not consistent with the expectation based on the estimated global European ancestry (ref. 25; set 1, $n = 13$; set 2, $n = 56$). We removed five samples because of overlap between studies. Overall, we removed 106 samples from set 1 and 56 samples from set 2. We also removed 187 AIMs that either had low call rates (<85%) or did not pass the different filters we applied to the data before analysis, which include a test of plausibility of parental allele frequencies, a measure of Hardy-Weinberg equilibrium with special attention to excess heterozygosity, and a linkage disequilibrium test (25). For quality controls, eight duplicate pairs were analyzed in set 1, and eight duplicate pairs plus eight CEU HapMap trios were analyzed in set 2. The overall quality control concordance rate was >99.9% for both SNP panels. The final data set consisted of 1,484 invasive breast cancer cases (593 from set 1 and 891 from set 2) and 2,240 AIMs, with 645 SNPs overlapping between the two sets. The final average number of AIMs per individual used in the analysis was 1,370.

Data Analysis

Ancestry Estimation. We used the ANCESTRYMAP software (26) as the central engine of the analysis. ANCESTRYMAP calculates the percentage of ancestry for each individual in the study. These estimates are reported in Supplementary Table S2 along with the standard deviations.

Association between Global Ancestry and Tumor Characteristics. We tested the association between proportion of global individual European ancestry (values range from 0 to 1) and ER, ER/PR status, stage [localized versus non-localized (non–localized tumors includes those with regional extension only, regional nodes only, regional extension and nodes, and remote)], and grade (1 and 2 versus 3) using logistic regression models run with the STATA statistical package. Reported odds ratios (OR) refer to the difference in risk associated with a change in European ancestry proportion from 0 to 1. Age at diagnosis and study were included in the basic models as covariates. The adjusted models also included the following covariates: age at first full-term pregnancy and number of full-term pregnancies (0, no pregnancies; 1, one or two children at age less than 21; 2, one or two children at age 21 or older; 3, three or more children at age less than 21; and 4, three or more children at age 21 or older—categorical), age at menarche (1, ≤12; 2, 13-14; 3, ≥15—categorical), body mass index (BMI; continuous), family history of breast cancer in first-degree relative (0, no; 1, yes—categorical), hormone replacement therapy, and menopausal status (0, premenopausal and no current hormone replacement therapy; 1, postmenopausal and no current hormone replacement therapy; 2, postmenopausal and current hormone replacement therapy—categorical).

Association between Locus-Specific Ancestry and Breast Cancer or Tumor Characteristics. The Logarithm (base 10) of the odds score for association is defined as the log of the likelihood ratio of the data under a disease locus model versus a no-disease locus model. The ANCESTRYMAP software uses Bayesian statistics and thus requires specification of a prior distribution on risk models before carrying out the analysis. We carried out the analysis assuming a prior distribution for ancestry risk that tested both for loci associated with increased risk due to European ancestry, and increased risk due to African ancestry. For all phenotypes (all cases, ER status, ER/PR combined status, ER/grade combined status, ER/age combined status), and stages (localized versus non-localized), we ran a prior distribution considering equally likely models of 0.5, 0.6, 0.7, 0.8, 1.3, 1.5, 1.7, and 2.0-fold increased risk for European ancestry. The ANCESTRYMAP program calculates a log factor for association at equally spaced points in the genome. A local score of 5, for example, means that the data at that locus are $10^5 = 100{,}000$ times more likely under an appropriately weighted average of the disease models, than under the null model. We followed the criteria used by Deo et al. (24) of a high threshold of >5 to be considered genome-wide significant. The frequencies of the typed SNPs in the ancestral populations were estimated based on data from European Americans and West African controls from previous studies (16, 24, 27).

Construction of Exclusion Map. To obtain credible intervals for increased risk due to African or European ancestry across the genome, we modified the procedure described elsewhere (24). ANCESTRYMAP was run for each of the three case definitions (ER+ only, ER– only, and all cases) using 85 independent disease risk models (0.30, 0.32, 0.34, 0.36, …, 1.94, 1.96, and 1.98-fold increased risk due to one European allele). We evaluated LOD scores at equally spaced points across the genome and searched for the maximum likelihood risk model at each of these points. This allowed the computation of 99.99% credible intervals for increased risk due to African (or European) ancestry by a likelihood ratio test, with the interval including all risk models for which the $\log_{10}$ of the likelihood of the disease model was within 3.275 of the maximum. Assuming 500 independent loci in the genome, these correspond to 95% genome-wide credible intervals by the Sidak correction for multiple hypothesis testing.

Results

Descriptive and tumor characteristics for cases in each of the six studies as well as for the combined sample of 1,484

[15] http://www.illumina.com/downloads/AfricanAmericanAdmixture_DataSheet.pdf

women are summarized in Table 1. The mean age at diagnosis of all cases was 54 years (range, 22-83). The average percentage of European ancestry over all cases was 23% (range, 1-98) and was relatively homogeneous among studies. The Women's Circle of Health Study had the lowest average percentage of European ancestry (19%) and the Multiethnic Cohort had the highest (25%). We observed 31% of individuals with ER– tumors, 53% with ER+ tumors, and 16% with missing status. ER– tumors were overrepresented among younger cases as noted in the LIFE study (42%) and the Los Angeles component of the Women's CARE study (35%), which is consistent with previous reports (28-30). Regarding tumor stage, 53% of the individuals had localized tumors, 34% were non-localized and 13% had missing data. In the LIFE and CARE studies, which included higher proportions of younger cases, only ~50% of the tumors were localized. For tumor grade, we observed a similar pattern, with a smaller proportion of lower grade tumors (grades 1 and 2) in the two studies that targeted younger women compared with the other studies. The percentage of European ancestry was significantly higher among individuals with hormone receptor–positive tumors compared with hormone receptor–negative tumors and women with localized disease compared with women with non-localized disease (Tables 1 and 2). We also observed a significantly higher percentage of European ancestry in women who were never pregnant compared with women who had one or more full-term pregnancies (Table 1).

Compared to women with ER– tumors, women with ER+ tumors had higher European ancestry [OR, 2.35; 95% confidence interval (CI), 1.06-5.20; Table 2]. This

Table 1. Sample and tumor characteristics for 1,484 African American women with breast cancer

	SFBABCS	BCFR	CARE	LIFE	MEC	WCHS	Total	EA % (SD)	*P**
n	185	304	372	110	409	104	1,484	23 (15)	
Age mean (SD)	55.2 (11.7)	50.4 (9.3)	48.8 (7.9)	42.3 (5.3)	65.8 (9.0)	50.0 (9.5)	54.2 (11.9)		
BMI mean kg/m^2 (SD)	30.4 (5.9)	30.3 (6.7)	27.6 (6.1)	29.0 (6.9)	29.1 (6.1)	30.4 (6.8)	29.2 (6.4)		
FHBC[†]									
Percent with FHBC	15	31	11	15	20	17	19	24 (17)	0.57
Percent without FHBC	85	69	84	79	72	83	77	23 (15)	
Age at first full-term pregnancy									
Percent no pregnancies	22	25	13	25	15	8	18	25 (17)	0.03[‡]
Percent <20	41	37	47	41	37	43	41	22 (14)	
Percent 20-30	29	34	33	27	38	30	33	24 (16)	
Percent >30	8	4	7	6	6	11	6	20 (14)	
Age at menarche									
Percent ≤12	52	47	57	54	51	47	52	23 (16)	0.18
Percent 13-14	33	39	33	37	37	40	36	23 (15)	
Percent 15 or more	14	12	10	9	10	13	11	23 (16)	
No. of full-term pregnancies									
Percent 0	22	23	13	24	15	8	17	25 (17)	<0.01
Percent 1-2	37	44	47	41	38	51	42	23 (15)	
Percent 3-5	33	29	34	31	35	27	32	23 (15)	
Percent 6 or more	8	3	6	3	8	7	6	18 (11)	
HRT/menopause status									
Percent pre-no HRT	31	58	47	81	11	35	39	23 (15)	0.13
Percent post-no HRT	45	23	23	16	61	32	36	23 (16)	
Percent post-yes HRT	16	9	14	2	16	0	12	25 (15)	
Percent estimated EA (SD)	22 (15)	23 (15)	23 (15)	23 (15)	25 (16)	19 (18)	23 (15)		<0.01[§]
ER status, *n* (%)									
ER+	96 (52)	158 (52)	192 (52)	45 (41)	237 (58)	57 (55)	785 (53)	24 (16)	0.04
ER–	52 (28)	92 (30)	131 (35)	46 (42)	94 (23)	41 (39)	456 (31)	22 (14)	
PR status, *n* (%)									
PR+	86 (46)	144 (47)	153 (41)	42 (38)	168 (41)	45 (43)	638 (43)	24 (16)	<0.01
PR–	61 (33)	104 (34)	121 (33)	44 (40)	111 (27)	53 (51)	494 (33)	22 (14)	
ER/PR status, *n* (%)									
ER+PR+	78 (42)	131 (43)	128 (34)	38 (35)	152 (37)	45 (43)	572 (39)	24 (17)	<0.01
ER–PR–	44 (24)	78 (26)	92 (25)	43 (39)	77 (19)	41 (39)	375 (25)	22 (14)	
ER+PR–	17 (9)	26 (9)	28 (8)	1 (1)	34 (8)	12 (12)	118 (8)	21 (14)	
ER–PR+	8 (4)	13 (4)	23 (6)	3 (3)	15 (4)	0 (0)	62 (4)	24 (12)	
Stage, *n* (%)									
Localized	118 (64)	143 (47)	186 (50)	58 (53)	281 (69)	0 (0)	786 (53)	25 (16)	<0.01
Non-localized	61 (33)	85 (28)	183 (49)	50 (45)	124 (30)	0 (0)	503 (34)	22 (14)	
Grade, *n* (%)									
1	20 (11)	34 (11)	42 (11)	8 (7)	67 (16)	12 (11)	183 (12)	24 (15)	0.27
2	63 (34)	81 (27)	98 (26)	32 (29)	132 (32)	34 (33)	440 (30)	24 (16)	
3	71 (38)	121 (40)	194 (53)	61 (56)	143 (36)	51 (49)	641 (43)	23 (15)	

NOTE: Percentages within the table did not add up to 100 because of missing data.
Abbreviations: SFBABCS, San Francisco Bay Area Breast Cancer Study; BCFR, Breast Cancer Family Registry; CARE, Contraceptive and Reproductive Experiences study; LIFE, Learning the Influence of Family and the Environment study; MEC, The Multiethnic Cohort study; WCHS, Women's Circle of Health Study; HR, hormone receptor; EA, European ancestry; FHBC, family history of breast cancer.
*P value of ANOVA (variables are unadjusted), evaluating if there is a significant difference in the percentage of European ancestry between different groups within variables. European genetic ancestry was log-transformed to approximate normality.
[†]In first-degree relatives.
[‡]For this particular test, which compared mean genetic ancestry for the different age groups at first full-term pregnancy, we restricted the analysis to women who had at least one full-term pregnancy.
[§]P value for the comparison of European genetic ancestry between studies.

Table 2. Association between tumor characteristics and proportion of global European genetic ancestry (values of European ancestry range from 0 to 1)

	OR (95% CI)	*P*
ER+ vs. ER− status (n = 1,241)*,†	2.35 (1.06-5.20)	0.034
ER+ vs. ER− status adjusted‡	2.06 (0.90-4.71)	0.087
ER+PR+ vs. ER−PR− status (n = 947)†	4.73 (1.56-14.33)	0.006
ER+PR+ vs. ER−PR− status adjusted‡	2.84 (1.13-7.14)	0.026
Stage (localized vs. non-localized, n = 1,289)†	2.89 (1.22-6.81)	0.015
Stage adjusted§	2.65 (1.11-6.35)	0.029
Grade (1 and 2 vs. 3, n = 1,264)†	1.60 (0.77-3.32)	0.205
Grade adjusted§	1.21 (0.48-3.08)	0.687

*ER+ coded as 1 and ER− coded as 0.
†Adjusted for age and study.
‡Adjusted for number of full-term pregnancies, age at first full-term pregnancy, hormone replacement therapy use, menopausal status, BMI, age, study, age at menarche, and family history of breast cancer.
§Adjusted for number of full-term pregnancies, age at first full-term pregnancy, hormone replacement therapy use, menopausal status, BMI, age, study, age at menarche, family history of breast cancer, and estrogen receptor status.

trend was observed both in cases with localized and non–localized tumors (localized: OR, 1.60; P = 0.39, n = 641; non-localized: OR, 2.08; P = 0.30, n = 429). For ER+PR+ (versus ER−PR−) tumors the association between European ancestry and positive receptor status became stronger (OR, 4.73; 95% CI, 1.56-14.33). We adjusted the models to include factors that have been found to correlate with hormone receptor status (i.e., number of full-term pregnancies, age at first full-term pregnancy, hormone replacement therapy, menopausal status, age at menarche, BMI, and family history of breast cancer). In the adjusted model, ER status alone was no longer significantly associated with ancestry (OR, 2.06; 95% CI, 0.90-4.71). The ER+PR+ versus ER−PR− analysis showed a significant ancestry effect. The OR of the unadjusted model was 4.73 (95% CI, 1.56-14.33; $P < 0.01$). After we adjusted for potential confounders, the effect of ancestry was reduced but remained statistically significant (OR, 2.84; 95% CI, 1.13-7.14; P = 0.026). Among the factors included in the adjusted model, the number of full-term pregnancies had the strongest effect, with nulliparous women being more likely to have ER+PR+ tumors compared with women who have one or more children (OR for being ER+PR+ if woman has one or more children: 0.40; 95% CI, 0.26-0.60; $P < 0.01$). We observed an association between European ancestry and disease stage (localized versus non-localized), with higher European ancestry among women with localized tumors (multivariate adjusted OR, 2.65; 95% CI, 1.11-6.35) compared with women with non-localized disease. We did not find a significant relationship between tumor grade and European ancestry (Table 2).

Admixture Mapping Does Not Show Significant or Suggestive Results Either for Breast Cancer Risk or for Tumor Characteristics. We next conducted a series of genome-wide admixture scans evaluating a number of breast cancer phenotypes (as described in Materials and Methods) among 1,484 African American women with breast cancer and 1,370 AIMs per subject, on average. The data were analyzed using an affected-only statistic, which calculates the likelihood of association based on an estimate of the ancestry at a particular location relative to the overall average ancestry of the individual's genome.

No genome-wide statistically significant association was observed between European or African ancestry and breast cancer at any specific locus (Table 3). The largest LOD score genome-wide was 2.9 (we set a threshold of >5 for significance; ref. 24) on chromosome X and 2.4 on chromosome 10 (in both cases, the African allele was associated with increased risk).

A series of analyses looking at hormone receptor status and at hormone receptor status and grade combined (Table 3) were not significant. Stratifying the analyses by age did not significantly alter the results. Case to case analyses were done comparing women with tumors that were hormone receptor–negative to those with hormone receptor–positive tumors as well as women with localized tumors versus non–localized tumors. The differences in locus-specific ancestry were not significant.

Analysis of Known Breast Cancer Risk Loci. We also searched for ancestry associations within regions that have previously been reported to be associated with breast cancer risk in other populations. Four genome-wide scans have been reported to date; all of them have been conducted in populations of European or Asian ancestry. The different regions that were found to be associated with risk were 4p14, 6q22, 7q22, 10q26, 5q11, 16q12, 11p15, 8q24, 2p24, 5p12, and 2q35 (31-36). Many of these regions have also been more strongly associated with ER+ status in Europeans (37). We found a weak deviation towards higher African ancestry within the 10q26 region compared with the rest of the chromosome; this region includes the *FGFR2* gene. The *FGFR2* gene has been repeatedly identified as a breast cancer susceptibility locus by genome-wide association studies (31, 33-35), and has also recently been fine-mapped to identify specific variants (38).

Table 3. Admixture mapping whole genome scan LOD scores for 1,484 African American women with breast cancer

	Cases	ER+	ER−	ER+PR+	ER−PR−	ER+ (grade 1 and 2)	ER− (grade 3)	ER+ PR+ (grade 1 and 2)	ER−PR− (grade 3)	RA
n	1,484	785	456	572	375	462	334	331	286	
Ch 3p24	0.73	**2.86**	0.15	**2.18**	0.1	1.35	0.23	0.89	−0.12	A
Ch 5p15	0.43	0.95	1.02	0.72	**1.5**	**1.37**	1.24	0.84	**1.65**	A
Ch 10q26	**2.39**	**2.41**	**1.86**	1.56	1.06	0.83	1.11	1.09	1.15	A
Ch 18q21	−0.77	1.38	−0.34	**2.22**	−0.33	0.34	0.22	1.29	−0.19	E
Ch Xp22	**2.94**	**2.57**	0.73	1.66	0.89	0.93	**1.69**	**1.54**	0.54	A

NOTE: The best LOD scores, or scores higher than 2, for the different admixture mapping whole genome scans are in boldface. Results are presented only for chromosomes that included the highest scores in a particular scan.
Abbreviations: RA, risk allele; A, African; E, European.

Table 4. Proportion of genome excluded as contributing to differential risk for all affected individuals and for ER+ and ER– phenotypes, comparing African and European ancestries

African*	Percentage of genome excluded†			European*	Percentage of genome excluded†		
	ER+	ER–	All		ER+	ER–	All
1.0‡	0.01	0.01	0.01	1.0	0.01	0.01	0.01
1.1	1	0.01	2	1.1	3	0.2	5
1.2	8	1	28	1.2	13	3	32
1.3	31	9	64	1.3	39	11	73
1.4	57	26	85	1.4	70	28	98
1.5	78	48	96	1.5	92	55	100
1.6	89	66	99	1.6	98	79	100
1.7	95	78	100	1.7	100	92	100
1.8	98	87	100	1.8	100	99	100
1.9	100	92	100	1.9	100	100	100
2.0	100	95	100	2.0	100	100	100
2.1	100	97	100	2.1	100	100	100
2.2	100	99	100	2.2	100	100	100
2.3	100	99	100	2.3	100	100	100
2.4	100	100	100	2.4	100	100	100

*Factor by which African (European) ancestry increases risk at this locus compared with European (African) ancestry.
†Percentage of genome excluded as having this risk or more at $P < 0.05$ genome-wide.
‡The percentage of the genome in which the null hypothesis (relative risk due to ancestry = 1) is excluded was ~0.01% for all scenarios, as expected using a $P < 0.0001$ significance cutoff, which is the corrected 5% cutoff for genome-wide significance (assuming 500 independent loci).

Exclusion Map. We prepared an exclusion map for the three case definitions with the largest sample sizes: all cases, ER+, and ER– cases. At least 98% of the genome can be excluded as having a European effect on risk of 1.4 or more, and at least 96% can be excluded as having an African effect on risk of 1.5 or more (Table 4). The power of the ER status analysis is less than that for all cases because of the smaller sample size. In the case of ER– disease, we can exclude 87% of the genome as having an increased risk of 1.8 or higher due to African ancestry and 92% as having an increased risk of 1.7 or higher due to European ancestry. In the case of ER+ disease, we can exclude 89% of the genome as having an increased risk of 1.6 or higher associated with African ancestry and 92% as having an increased risk of 1.5 or higher associated with European ancestry (Table 4).

Discussion

The present study represents the first genome-wide admixture scan conducted in African American women with breast cancer. In this study, we did not find an association between breast cancer risk and African or European ancestry at any specific loci among all cases or within subtypes of breast cancer, at genome-wide levels of significance. We detected European ancestry to be overrepresented among women with ER+ tumors. However, adjustment for known breast cancer risk factors could explain this association. A significant association remained for ER+PR+ tumors following adjustment, which could be due to misclassification of these risk factors, other risk factors which we did not consider (e.g., alcohol consumption), or that we do not know about that do correlate with ancestry and influence tumor characteristics. At the same time, it is possible that this association is due to genetic risk factors that correlate with ancestry. We observed that nulliparity was associated with both ER+PR+ disease as well as European ancestry. The association between number of full-term pregnancies and hormone receptors status has been reported previously in African Americans and white women (29), and our data replicates these results. The association between nulliparity and ER+PR+ disease could be the result of an underlying biological mechanism or could be due to the correlation between this risk factor and other known or unknown risk factors that we did not account for. The association between European ancestry and nulliparity was also significant ($P = 0.01$) but could not completely explain the association that we observed between ancestry and ER/PR status. We also detected European ancestry to be significantly overrepresented among women with localized tumors compared with women with non–localized tumors (OR, 2.65; 95% CI, 1.11-6.35; $P = 0.029$). This association could not be explained by the known breast cancer risk factors.

The exclusion map shows that for the analysis of the ER– cases, we had reasonable power to detect an increased risk due to an African allele of 1.8 and above and an increased risk due to a European allele of 1.6 and above. Therefore, the fact that our scan did not detect any significant signal does not discard the possibility that ancestry effects of 1.7 or lower are present. The observed association between ancestry and ER/PR status supports this possibility and suggests that further analyses are needed with adequate power to detect ancestry effects on risk of 1.7 or less.

We detected a nonsignificant deviation towards higher African ancestry on chromosome 10q26 compared with the chromosomal average. This region includes the *FGFR2* gene and a common variant that is associated with increased risk of breast cancer in Asian and European populations (33, 34, 38). A recently published study investigated FGFR2 variants in African Americans, Asians, and Europeans to search for causative variants and to evaluate if the same variants were associated with risk of breast cancer in the different racial/ethnic groups (38). Based on association results, and an analysis of DNase I hypersensitive sites looking at chromatin accessibility, the conclusion was reached that two variants, rs2981578 and rs10736303, are the most likely to be causal variants. The frequency of these two variants is different in African populations compared with Europeans or Asians. The

frequency of the risk allele for the variant rs2981578 is 0.93 in the HapMap African sample and 0.46 in the HapMap European samples. A similar difference was observed for rs10736303, with the risk allele having a frequency of 0.92 in Africans and 0.60 in Europeans.[16] The increase in African ancestry that we observed in the admixture mapping analysis within the 10q26 region could potentially be explained by the higher frequency of causal risk alleles in this region, which are likely to be more common in African than European populations.

There was no apparent deviation from the average chromosomal ancestry for any other region of the genome previously reported to have a risk variant. Different studies have reported associations between variants in the *FGFR2* gene and breast cancer risk, with per allele ORs that varied between 1.20 and 1.30 (33, 34, 38-40). The reported ORs for the *FGFR2* gene are among the higher reported ORs compared with those of other risk variants discovered through whole genome association studies (~1.25 compared with <1.20; ref. 39). Adding to this, the candidate variants within the *FGFR2* gene show a large allele frequency difference between Europeans and Africans. Therefore, it is likely that we did not observe any other ancestry deviations because of lack of power (we had power >80% to detect risk variants with an allele effect of 1.5 or larger; if the allele effect was ~1.2, then the allele frequency difference between the ancestral populations needed to be larger than 0.7 to achieve a power above 40%).

One limitation of this study is the sample size. Although the study included >1,400 women, ER–PR– cases are still a minority of cases, even among African Americans, and thus, we had limited power to assess associations for the different breast cancer phenotypes.

Her2 status was not available for the majority of cases because most of the cases in the different studies were recruited at a time when Her2 status was not routinely assayed for clinical testing. Therefore, we were unable to analyze ER–PR–Her2-negative breast cancer cases (i.e., "triple negatives"), an aggressive subset of tumors that has been estimated to be more common in African Americans than in European Americans (28-30, 41). Much larger studies in African populations, with available tumor specimen resources for tumor phenotyping, will be needed to evaluate the genetic contribution to the various breast cancer subtypes.

Information about ER and PR status, grade, and stage, comes from pathology reports or from the cancer registry, depending on the study. Therefore, it is likely that there were differences in how the tumors were classified. This potential misclassification could have contributed to the negative results observed. However, the frequency of the different tumor characteristics in the six studies are similar and when they differ, they do it in the expected direction given the age distribution of the women in the studies. This suggests that misclassification might not be a serious problem for these data, although caution must be taken in the interpretation of the results. Future studies involving centralized tumor marker data collection will be necessary to avoid the potential effect of misclassification in genetic epidemiology studies with multiple data sources.

The AIMs selected to infer genetic ancestry are assumed to have homogenous frequency within the African continent. Given that African Americans are likely to have a mixed ancestry from different regions of Western Africa (42), which might not share the same allele frequencies for the markers used in the present study, results must be interpreted with caution.

The clinical implications of the differences in tumor presentation of African American women with breast cancer compared with European American patients are substantial. Although the overall incidence of breast cancer is lower in African American women, the mortality rate is higher in African American women than in European American women (43). This may be in part be due to higher rates of ER– disease because hormonal treatment, either with selective estrogen receptor modifiers (tamoxifen or raloxifene) or with aromatase inhibitors, is highly effective for ER+ disease only (44). Furthermore, ER– disease often occurs in younger women who have never had screening because they are younger than the standard screening age and because screening with mammography is less sensitive among younger women (45). The high rates of ER– disease among African Americans may also have implications for breast cancer prevention. Tamoxifen and raloxifene have been shown to prevent ER+ breast cancer in primary prevention studies, and some have advocated that the medications be used in women at high risk (44, 46). In addition, aromatase inhibitors may also be useful in the prevention of breast cancer (47). However, there is no clear preventive strategy for ER– breast cancers. Identifying the causal factors that explain the difference in incidence of hormone receptor–negative tumors between European American and African American women should be a high priority.

The present admixture mapping scan in 1,484 African American women with breast cancer suggests that the difference in breast cancer risk between Europeans and African Americans is unlikely to be due to an effect of a European or African allele on risk larger than 1.7. It also excludes an effect on risk for ER+ status larger than 1.9 and for ER– status larger than 2.4. Global ancestry association results, however, show a positive association of European ancestry with stage of disease, and with ER+PR+ disease. These associations could result from population differences in nongenetic risk factors or from the effect of multiple genetic variants each with a relatively moderate contribution to the ancestry-related risk difference.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

We thank Loreall Pooler, David Wong, and David Van Den Berg for their laboratory assistance, and Dr. Kristine Monroe and Hank Huang for their technical support. We also want to thank the study participants.

[16] http://www.ncbi.nlm.nih.gov/sites/entrez

References

1. Ries LAG, Eisner MP, Kosary CL, et al. SEER cancer statistics review 1975–2000. Bethesda: National Cancer Institute 2003.
2. Althuis MD, Brogan DD, Coates RJ, et al. Breast cancers among very young premenopausal women (United States). Cancer Causes Control 2003;14:151–60.
3. Joslyn SA, West MM. Racial differences in breast carcinoma survival. Cancer 2000;88:114–23.
4. Fejerman L, Ziv E. Population differences in breast cancer severity. Pharmacogenomics 2008;9:323–33.
5. Chlebowski RT, Chen Z, Rohan T, et al. Ethnicity and breast cancer in the Women's Health Initiative: a unifying concept for unfavorable outcome in African American women [abstract 1008]. J Clin Oncol 2004;22.
6. Amend K, Hicks D, Ambrosone CB. Breast cancer in African-American women: differences in tumor biology from European-American women. Cancer Res 2006;66:8327–30.
7. Newman LA, Bunner S, Carolin K, et al. Ethnicity related differences in the survival of young breast carcinoma patients. Cancer 2002;95: 21–7.
8. Joslyn SA. Hormone receptors in breast cancer: racial differences in distribution and survival. Breast Cancer Res Treat 2002;73:45–59.
9. English WP, Cleveland KE, Barber WH. There is no difference in survival between African-American and white women with breast cancer. Am Surg 2002;68:594–7.
10. Chlebowski RT, Chen Z, Anderson GL, et al. Ethnicity and breast cancer: factors influencing differences in incidence and outcome. J Natl Cancer Inst 2005;97:439–48.
11. Rosenberg NA, Li LM, Ward R, Pritchard JK. Informativeness of genetic markers for inference of ancestry. Am J Hum Genet 2003;73: 1402–22.
12. McKeigue PM. Mapping genes that underlie ethnic differences in disease risk: methods for detecting linkage in admixed populations, by conditioning on parental admixture. Am J Hum Genet 1998;63:241–51.
13. Collins-Schramm HE, Phillips CM, Operario DJ, et al. Ethnic-difference markers for use in mapping by admixture linkage disequilibrium. Am J Hum Genet 2002;70:737–50.
14. Smith MW, O'Brien SJ. Mapping by admixture linkage disequilibrium: advances, limitations and guidelines. Nat Rev Genet 2005;6:623–32.
15. Collins-Schramm HE, Chima B, Operario DJ, Criswell LA, Seldin MF. Markers informative for ancestry demonstrate consistent megabase-length linkage disequilibrium in the African American population. Hum Genet 2003;113:211–9.
16. Freedman ML, Haiman CA, Patterson N, et al. Admixture mapping identifies 8q24 as a prostate cancer risk locus in African-American men. Proc Natl Acad Sci U S A 2006;103:14068–73.
17. Pike MC, Kolonel LN, Henderson BE, et al. Breast cancer in a multiethnic cohort in Hawaii and Los Angeles: risk factor-adjusted incidence in Japanese equals and in Hawaiians exceeds that in whites. Cancer Epidemiol Biomarkers Prev 2002;11:795–800.
18. Kolonel LN, Henderson BE, Hankin JH, et al. A multiethnic cohort in Hawaii and Los Angeles: baseline characteristics. Am J Epidemiol 2000;151:346–57.
19. Marchbanks PA, McDonald JA, Wilson HG, et al. The NICHD Women's Contraceptive and Reproductive Experiences Study: methods and operational results. Ann Epidemiol 2002;12:213–21.
20. Ma H, Bernstein L, Ross RK, Ursin G. Hormone-related risk factors for breast cancer in women under age 50 years by estrogen and progesterone receptor status: results from a case-control and a case-case comparison. Breast Cancer Res 2006;8:R39.
21. John EM, Schwartz GG, Koo J, Wang W, Ingles SA. Sun exposure, vitamin D receptor gene polymorphisms, and breast cancer risk in a multiethnic population. Am J Epidemiol 2007;166:1409–19.
22. John EM, Horn-Ross PL, Koo J. Lifetime physical activity and breast cancer risk in a multiethnic population: the San Francisco Bay Area Breast Cancer Study. Cancer Epidemiol Biomarkers Prev 2003;12: 1143–52.
23. John EM, Hopper JL, Beck JC, et al. The Breast Cancer Family Registry: an infrastructure for cooperative multinational, interdisciplinary and translational studies of the genetic epidemiology of breast cancer. Breast Cancer Res 2004;6:R375–89.
24. Deo RC, Patterson N, Tandon A, et al. A high-density admixture scan in 1,670 African Americans with hypertension. PLoS Genet 2007;3:e196.
25. Reich D, Patterson N, De Jager PL, et al. A whole-genome admixture scan finds a candidate locus for multiple sclerosis susceptibility. Nat Genet 2005;37:1113–8.
26. Patterson N, Hattangadi N, Lane B, et al. Methods for high-density admixture mapping of disease genes. Am J Hum Genet 2004; 74:979–1000.
27. Smith MW, Patterson N, Lautenberger JA, et al. A high-density admixture map for disease gene discovery in African Americans. Am J Hum Genet 2004;74:1001–13.
28. Carey LA, Perou CM, Livasy CA, et al. Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. JAMA 2006;295: 2492–502.
29. Millikan RC, Newman B, Tse CK, et al. Epidemiology of basal-like breast cancer. Breast Cancer Res Treat 2008;109:123–39.
30. Bauer KR, Brown M, Cress RD, Parise CA, Caggiano V. Descriptive analysis of estrogen receptor (ER)-negative, progesterone receptor (PR)-negative, and HER2-negative invasive breast cancer, the so-called triple-negative phenotype: a population-based study from the California Cancer Registry. Cancer 2007;109:1721–8.
31. Stacey SN, Manolescu A, Sulem P, et al. Common variants on chromosome 5p12 confer susceptibility to estrogen receptor-positive breast cancer. Nat Genet 2008;40:703–6.
32. Stacey SN, Manolescu A, Sulem P, et al. Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. Nat Genet 2007;39:865–9.
33. Hunter DJ, Kraft P, Jacobs KB, et al. A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. Nat Genet 2007;39:870–4.
34. Easton DF, Pooley KA, Dunning AM, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. Nature 2007; 447:1087–93.
35. Gold B, Kirchhoff T, Stefanov S, et al. Genome-wide association study provides evidence for a breast cancer risk locus at 6q22.33. Proc Natl Acad Sci U S A 2008;105:4340–5.
36. Ghoussaini M, Song H, Koessler T, et al. Multiple loci with different cancer specificities within the 8q24 gene desert. J Natl Cancer Inst 2008;100:962–6.
37. Garcia-Closas M, Hall P, Nevanlinna H, et al. Heterogeneity of breast cancer associations with five susceptibility loci by clinical and pathological characteristics. PLoS Genet 2008;4:e1000054.
38. Udler MS, Meyer KB, Pooley KA, et al. FGFR2 variants and breast cancer risk: fine-scale mapping using African American studies and analysis of chromatin conformation. Hum Mol Genet 2009; 18:1692–703.
39. Easton DF, Eeles RA. Genome-wide association studies in cancer. Hum Mol Genet 2008;17:R109–15.
40. Thomas G, Jacobs KB, Kraft P, 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;41:579–84.
41. Olopade OI, Ikpatt FO, Dignam JJ, et al. "Intrinsic Gene Expression" subtypes correlated with grade and morphometric parameters reveal a high proportion of aggressive basal-like tumors among black women of African ancestry [abstract 9509]. J Clin Oncol 2004;22.
42. Tishkoff SA, Reed FA, Friedlaender FR, et al. The genetic structure and history of Africans and African Americans. Science 2009;324:1035–44.
43. Ademuyiwa FO, Olopade OI. Racial differences in genetic factors associated with breast cancer. Cancer Metastasis Rev 2003;22:47–53.
44. Fisher B, Costantino JP, Wickerham DL, et al. Tamoxifen for prevention of breast cancer: report of the National Surgical Adjuvant Breast and Bowel Project P-1 Study. J Natl Cancer Inst 1998;90:1371–88.
45. Cunningham JE, Butler WM. Racial disparities in female breast cancer in South Carolina: clinical evidence for a biological basis. Breast Cancer Res Treat 2004;88:161–76.
46. Cummings SR, Eckert S, Krueger KA, et al. The effect of raloxifene on risk of breast cancer in postmenopausal women: results from the MORE randomized trial. Multiple Outcomes of Raloxifene Evaluation. JAMA 1999;281:2189–97.
47. Dunn BK, Wickerham DL, Ford L. Prevention of hormone-related cancers: breast cancer. J Clin Oncol 2005;23:357–67.