

# Genetic Ancestry and Risk of Breast Cancer among U.S. Latinas

Laura Fejerman,<sup>1,4</sup> Esther M. John,<sup>6,7</sup> Scott Huntsman,<sup>1,4</sup> Kenny Beckman,<sup>8</sup> Shweta Choudhry,<sup>2,3</sup> Eliseo Perez-Stable,<sup>1,4,5</sup> Esteban González Burchard,<sup>2,3</sup> and Elad Ziv<sup>1,4</sup>

<sup>1</sup>Division of General Internal Medicine and <sup>2</sup>Lung Biology Center, Department of Medicine; <sup>3</sup>Department of Biopharmaceutical Sciences; <sup>4</sup>Helen Diller Family Comprehensive Cancer Center; and <sup>5</sup>Medical Effectiveness Research Center for Diverse Populations, University of California, San Francisco, San Francisco, California; <sup>6</sup>Northern California Cancer Center, Fremont, California; <sup>7</sup>Stanford School of Medicine and Stanford Cancer Center, Stanford, California; and <sup>8</sup>Children's Hospital Research Institute, Oakland, California

## Abstract

**U.S. Latinas have a lower incidence of breast cancer compared with non-Latina White women. This difference is partially explained by differences in the prevalence of known risk factors. Genetic factors may also contribute to this difference in incidence. Latinas are an admixed population with most of their genetic ancestry from Europeans and Indigenous Americans. We used genetic markers to estimate the ancestry of Latina breast cancer cases and controls and assessed the association with genetic ancestry, adjusting for reproductive and other risk factors. We typed a set of 106 ancestry informative markers in 440 Latina women with breast cancer and 597 Latina controls from the San Francisco Bay area and estimated genetic ancestry using a maximum likelihood method. Odds ratios (OR) and 95% confidence intervals (95% CI) for ancestry modeled as a continuous variable were estimated using logistic regression with known risk factors included as covariates. Higher European ancestry was associated with increased breast cancer risk. The OR for a 25% increase in European ancestry was 1.79 (95% CI, 1.28–2.79;  $P < 0.001$ ). When known risk factors and place of birth were adjusted for, the association with European ancestry was attenuated but remained statistically significant (OR, 1.39; 95% CI, 1.06–2.11;  $P = 0.013$ ). Further work is needed to determine if the association is due to genetic differences between populations or possibly due to environmental factors not measured. [Cancer Res 2008;68(23):9723–8]**

## Introduction

Breast cancer incidence varies across populations in the United States. Data from the Surveillance, Epidemiology, and End Results program show that the age-adjusted incidence (per 100,000) of breast cancer (from 1998 to 2002) is highest in White women (141.0), followed by African American (119.4), Asian American (96.6), and Latina (89.9) women, with the lowest incidence in Indigenous American women (54.8; ref. 1). Variation in exposure to known risk factors may explain some (2–5), but not all (6–8), of these differences in incidence. The residual difference among populations may be due to incomplete assessment of known risk factors or to risk factors not yet identified. It could also be partly due to differences between populations in the allele frequencies of predisposing genetic variants.

Women of mixed descent, like U.S. Latinas, present both a challenge and a unique opportunity in genetic association studies (9–11). On one hand, studies in Latinos may be confounded due to the potentially underlying dissimilarity between cases and controls in terms of genetic ancestry (12, 13). On the other hand, populations of mixed ancestry provide an opportunity for examining the role of genetic and environmental factors in explaining observed differences in incidence between populations and, eventually, for locating alleles that contribute to dissimilarities in disease risk. This can be achieved by means of admixture mapping, an approach that is based on the idea that if a marker increases the risk of disease and is found at a much higher frequency in one population, then that marker will also be found more commonly among cases and will be strongly associated with other ancestry specific markers across large stretches of the genome (14). Breast cancer among Latinas presents a particularly interesting case because the main ancestral components of the Latino population (European and Indigenous American) have the highest and lowest breast cancer incidence (1).

We have previously investigated the association between genetic ancestry and breast cancer risk factors among Latinas in the San Francisco Bay Area using 44 ancestry informative markers (AIM; ref. 7). Here we use DNA samples from our previous study (167 cases and 286 controls) and DNA samples for an additional 273 cases and 311 controls to test the association between breast cancer risk and genetic ancestry among Latinas. We used 106 AIMs to determine the genetic ancestry in all of the women and compared ancestry between cases and controls, adjusting for known breast cancer risk factors in an effort to identify a genetic ancestry component to breast cancer risk. We also investigated the use of genetic ancestry as a covariate in genetic association studies for breast cancer among Latinas.

## Materials and Methods

### Source of Cases and Controls

Analyses were done using DNA and data from two population-based studies conducted in the San Francisco Bay Area: a case-control study of breast cancer and a family registry for breast cancer.

The San Francisco Bay Area Breast Cancer Study, described elsewhere (8, 15), is a multiethnic population-based case-control study of breast cancer initiated in 1995, and with biospecimen collection added for cases diagnosed between April 1, 1997 and April 30, 2002 and matching controls. Depending on the study protocol, study participants were invited to provide a blood or buccal sample. Women ages 35 to 79 y; residing in San Francisco, San Mateo, Alameda, Contra Costa, or Santa Clara counties; and newly diagnosed with a first primary invasive breast cancer were identified through the Greater Bay Area Cancer Registry, which ascertains all incident cancers as part of the Surveillance, Epidemiology, and End Results program and the California Cancer Registry. A brief telephone screening interview that assessed study eligibility and self-reported race/ethnicity (89% response among those contacted) identified 873 eligible Latina cases. Of

**Requests for reprints:** Laura Fejerman, University of California, San Francisco, 1701 Divisadero Street, 5th Floor, San Francisco, CA 94143-1732. Phone: 415-885-7504; Fax: 415-353-7932; E-mail: laura.fejerman@ucsf.edu.

©2008 American Association for Cancer Research.  
doi:10.1158/0008-5472.CAN-08-2039

these, 798 (91%) completed an in-person interview and 747 (86%) provided a biospecimen sample. Control women, ages 35 to 79 y and residing in the same five Bay Area counties, were ascertained by random digit dialing. They were frequency matched to cases by race/ethnicity and expected 5-y age group. The telephone screening interview, completed by 93% of women selected as controls, identified 1,126 eligible Latina controls without a personal history of breast cancer. Of these, 999 (89%) completed the in-person interview and 911 (81%) provided a biospecimen sample.

The present analysis includes only cases and controls who donated a blood sample. Sixty-three of the cases that participated in the current case-control study also participated in the Northern California site of the Breast Cancer Family Registry (16) and donated a blood sample as part of that study, which was obtained for this analysis.

The total number of blood samples available for the study was 503 cases and 679 controls. Individuals who did not provide information about country of birth ( $n = 9$ ) or who were born in Europe ( $n = 6$ ), Hawaii ( $n = 2$ ), Philippines ( $n = 1$ ), or in a country that was represented only by one individual (Brazil, Dominican Republic) were excluded from the present analysis (11 cases and 9 controls). The final number of samples genotyped was 492 cases and 670 controls.

All participants provided written informed consent and the research protocols were approved by the respective Institutional Review Boards at University of California, San Francisco and the Northern California Cancer Center.

## Measures

**Survey data.** Data on age, demographic background (education in years, country of birth, age at migration if not U.S. born, and country of birth of parents and grandparents), and known or suspected breast cancer risk factors (age at menarche, parity, age at first full-term pregnancy, breastfeeding, use of oral contraceptives, use of hormone replacement therapy, daily alcohol intake, family history of breast cancer, and benign breast disease) were collected by in-person interview using a structured questionnaire (7). Dietary intake during the reference year (defined as the year before diagnosis for cases and the year before selection into the study for controls) was assessed using a modified version of the Block Food Frequency Questionnaire. Standing height and weight were measured by the interviewers. Body mass index (BMI) was calculated as measured weight (kg) divided by measured height (m) squared. For participants (13 cases and 21 controls) who declined the measurements, the BMI was based on self-reported height and weight during the reference year.

Tumor grade, stage, histologic type, and hormone receptor status were obtained from the Surveillance, Epidemiology, and End Results Cancer Registry records. Estrogen and progesterone receptor status were dichotomized (positive, negative) based on categories reported in pathology records. Information on human epidermal growth factor receptor 2 (Her2) status was not routinely obtained by the cancer registry for cases diagnosed before 2002. Therefore, we did not include Her2 status in the present analysis.

## Marker Selection and Ancestral Populations

A set of 106 single nucleotide polymorphisms (SNP) that can separate Indigenous American, African, and European ancestry was used to estimate proportion of genetic ancestry in the sample of U.S. Latinas. Simulation studies have shown that ~100 AIMs with allele frequency differences similar to the ones we used are required to achieve a correlation coefficient of >0.9 with true ancestry (13); thus, we genotyped 112 markers with the goal of successfully typing >100 markers. The AIMs used in this study were biallelic SNPs selected from the Affymetrix 100K SNP chip. AIM selection was based on calculations of allele frequency differences between Europeans, West Africans, and Indigenous Americans. The SNPs chosen maximize information for more than one ancestral population pairing, with a large difference in allele frequency between ancestral populations (>0.5). The AIMs are widely spaced throughout the genome and have a well-balanced distribution across all 22 autosomal chromosomes. The average distance between markers is  $\sim 2.4 \times 10^7$  bp. The parental population samples that were genotyped on the Affymetrix 100K SNP chip included 42 Europeans (Coriell's North American Caucasian panel), 37 West Africans

(nonadmixed Africans living in London, United Kingdom and South Carolina) and 30 Indigenous Americans (15 Mayans and 15 Nahuas). (More detailed information on the AIMs is available from the authors on request).

## Genotyping

Genotyping of the 106 AIMs was done by Dr. Kenneth Beckman at the Children's Hospital Oakland Research Institute. Quality control was done on all DNA using a two-part procedure. Quantitative quality control (part 1) involved nonallelic quantitative real-time PCR using a single TaqMan probe to ensure amplifiability of DNA samples. Qualitative quality control (part 2) involved genotyping using a balanced polymorphism present in most human populations (rs3818) to ensure that cross-contamination of samples has not occurred. Genotyping was done using iPLEX reagents and protocols for multiplex PCR, single-base primer extension, and generation of mass spectra, as per manufacturer's instructions (for complete details, see iPLEX Application Note, Sequenom). It involved four multiplexed assays containing 29, 29, 28, and 26 SNPs, respectively, for a total of 112 candidate AIMs. Of these 112 markers, 106 robustly generated call rates at 90% of samples or higher, with typical call rates in excess of 99% of samples. Only those 106 markers were used in the study. Multiplexed PCR was done in 5- $\mu$ L reactions on 384-well plates containing 5 ng of genomic DNA. Reactions contained 0.5 unit HotStarTaq polymerase (Qiagen), 100 nmol/L primers, 1.25 $\times$  HotStarTaq buffer, 1.625 mmol/L MgCl<sub>2</sub>, and 500  $\mu$ mol/L deoxy-nucleotide triphosphates (dNTP). Following enzyme activation at 94°C for 15 min, DNA was amplified with 45 cycles of 94°C  $\times$  20 s, 56°C  $\times$  30 s, 72°C  $\times$  1 min, followed by a 3-min extension at 72°C. Unincorporated dNTPs were removed using shrimp alkaline phosphatase (0.3 unit; Sequenom). Single-base extension was carried out by addition of single-base primers at concentrations from 0.625  $\mu$ mol/L (low molecular weight primers) to 1.25  $\mu$ mol/L (high molecular weight primers) using iPLEX enzyme and buffers (Sequenom) in 9- $\mu$ L reactions. Reactions were desalted and single-base primer products measured using the MassARRAY Compact system, and mass spectra were analyzed using TYPER software (Sequenom) to generate genotype calls and allele frequencies.

There was insufficient DNA available from 574 individuals in the study. Therefore, DNA from these samples was amplified using a commercially available whole genome amplification kit (Qiagen REPLI-g Midi Kit). From the original set of samples that went through amplification, 92 yielded low-quality DNA and were excluded from the genotyping phase. A total of 1,070 samples (462 cases and 608 controls) were genotyped. Quality control measures were high for the whole genome amplification samples and the nonamplified ones. For whole genome amplification samples, the average AIM success rate was 98.5%, compared with 99% for the nonamplified samples. The average sample call rate was 95.6% for the whole genome amplification samples and 97.4% for the nonamplified samples. Samples with call rate smaller than 75% were excluded from the analysis (22 cases and 11 controls).

Three of the AIMs deviated significantly from Hardy-Weinberg equilibrium ( $P < 0.0005$ ), all of them showing excess homozygosity, which is expected in the presence of population substructure (17).

Genotype and phenotype information was available for a total of 1,037 individuals (440 cases and 597 controls).

## Statistical Analysis

Estimates of each individual's genetic ancestry were derived using a maximum likelihood approach (18, 19). The maximum likelihood model infers ancestry of each individual as a function of the probability of the genotypes observed at each locus based on the ancestral allele frequencies (Java script available from the authors on request). We used  $t$  tests (for continuous variables) and Fisher's exact tests for two by two frequency tables (for categorical variables) to determine if there were significant differences in characteristics between cases and controls. Mean genetic ancestry was estimated as the average of the individual genetic ancestry estimates within a group.

Associations between breast cancer risk and genetic ancestry were assessed using logistic regression models. Genetic ancestry was modeled as a continuous variable (with each unit change representing a 25% increase in European or African ancestry). The multivariate adjusted models included

European ancestry, age (continuous), family history of breast cancer in first-degree relatives (yes, no), place of birth (U.S. born, foreign born), personal history of benign breast disease (yes, no), age at menarche, number of full-term pregnancies, months of breast-feeding per child, use of hormone replacement therapy (yes, no), daily alcohol intake ( $\leq 10$  versus  $> 10$  g), daily calorie intake (log transformed) during the reference year, and education (elementary school, middle school, high school, and college). Individuals with missing data were dropped from the multivariate analysis (32 cases and 25 controls). We evaluated models including both European and African ancestry (continuous) and using parent/grandparent European origin instead of genetic ancestry. The association with each AIM was evaluated with a logistic regression model with and without inclusion of genetic ancestry as a covariate to compare the distribution of  $z$  statistics before and after correction for population substructure.

All statistical tests were done using the programs STATA (20) and R (21), and all tests are two-sided.

## Results

The characteristics of breast cancer cases and controls are presented in Table 1. Cases had a mean age of 55 years at diagnosis, which was not significantly different from that of controls. In

bivariate analyses, cases had significantly more full-term pregnancies than controls; were less likely to breast-feed; and were more likely to report a personal history of benign breast disease, a family history of breast cancer, earlier menarche, higher alcohol intake, and higher daily calorie intake. Cases also reported a significantly higher level of education and were more likely to have been born in the United States. They had more European and less Indigenous American ancestry than controls. There were no significant differences between cases and controls in use of hormone replacement therapy or oral contraceptives, age at first full-term pregnancy, and BMI.

In unadjusted models, we found a strong association between genetic ancestry (continuous) and breast cancer risk. Higher European ancestry was associated with increased risk, with an odds ratio (OR) of 1.79 [95% confidence intervals (95% CI), 1.28–2.79;  $P < 0.001$ ] for every 25% increase in European ancestry. When known risk factors and place of birth were adjusted for (Table 2), the association with European ancestry was somewhat attenuated but remained statistically significant (OR, 1.39; 95% CI, 1.06–2.11;  $P = 0.013$ ). When African ancestry was included in the adjusted

**Table 1.** Characteristics of cases and controls

Variable	Controls, $n = 597$	Cases, $n = 440$	$P^*$
	Mean (SD)		
<b>Demographic information</b>			
Age	54.3 (11.18)	55.1 (10.62)	0.21
Foreign born (%)	63	45	<0.001
Age at migration (if foreign born)	28.0 (12.2)	25.9 (14.2)	0.07
Education (%)			<0.001
Elementary	31	17	
Middle school	13	11	
High school	33	41	
University	23	31	
<b>Breast cancer risk factors</b>			
BMI	30.72 (5.65)	30.32 (6.13)	0.29
Daily calorie intake (kcal)	2,223 (1,006)	2,377 (1,213.04)	0.027
Daily alcohol intake (g; %)			<0.001
0	68	64	
$\leq 10$	27	24	
$> 10$	5	12	
Age at menarche	12.7 (1.77)	12.4 (1.73)	0.002
Age at first full-term pregnancy	23.1 (5.24)	23.44 (5.28)	0.33
No. full-term pregnancies (%)			<0.001
0	6	13	
1–2	29	36	
$> 2$	65	51	
History of breast-feeding (%)	77	59	<0.001
Hormone replacement therapy use (%)	40	42	0.32
Oral contraceptive use (%)	63	66	0.23
Postmenopausal status (%)	64	64	0.89
Family history of breast cancer (%)	9	15	0.006
Prior benign breast disease (%)	14	18	0.042
<b>Estimated ancestry</b>			
African ancestry	0.07 (0.09)	0.07 (0.08)	0.30
European ancestry	0.53 (0.18)	0.58 (0.19)	<0.001
Indigenous American ancestry	0.40 (0.18)	0.35 (0.18)	<0.001

\* $P$  values are for bivariate Fisher's exact tests in two by two tables (for categorical variables) and for  $t$  tests comparing mean values between cases and controls (for continuous variables).

model, the association with European ancestry became stronger [OR for European ancestry, 1.54 (95% CI, 1.11–2.52;  $P = 0.004$ ), and OR for African ancestry, 2.05 (95% CI, 1.00–7.56;  $P = 0.055$ )]. In all models, the associations between breast cancer and alcohol consumption, parity, family history, age at menarche, and history of breast-feeding were in the expected direction (Table 2). To ensure that there was no confounding due to differences in place of birth between cases and controls, the same analysis was stratified by place of birth (United States, Mexico, South America, and Central America) with all results showing the same trend as the global analysis (OR for the association with ancestry varied from 1.10 to 1.82; data not shown). We observed a significant association between the number of European-born parents/grandparents and breast cancer risk, with higher number of European ancestors being associated with increased risk (OR, 1.21; 95% CI, 1.02–1.44;  $P = 0.025$ , adjusted model).

We found no evidence that associations with genetic ancestry differed by tumor characteristics such as hormone receptor status, stage, or grade (Table 3). However, there were interesting trends. For example, there was a trend toward higher Indigenous American ancestry for cases with mucinous adenocarcinoma and a trend toward higher European ancestry for cases with mixed ductal/lobular histology, compared with the estimated mean ancestry for cases. Cases diagnosed at a more advanced stage had a trend toward higher Indigenous American ancestry.

We examined the effect of adjustment for genetic ancestry on the association between risk of breast cancer and each of the 106 AIMs. Without adjustment for ancestry, 20 of 106 markers were nominally associated with breast cancer risk. After adjustment for ancestry, only 4 markers had  $P < 0.05$ , which were no longer

significant after adjustment for multiple testing (rs1398829,  $P = 0.005$ ; rs10498919,  $P = 0.018$ ; rs7535375,  $P = 0.018$ ; rs1470524,  $P = 0.018$ ).

Adjustment for place of birth (U.S. born versus foreign born) and number of European-born ancestors was not as effective as genetic ancestry in eliminating the excess number of AIMs associated with risk of breast cancer. In models that included these factors but did not include genetic ancestry, 13 of 106 markers were nominally associated with breast cancer.

We estimated individual ancestry with and without the three AIMs that were not in Hardy-Weinberg equilibrium. Estimates were very similar and the associations remained significant.

## Discussion

The incidence of breast cancer among Latinas is up to 40% lower than the incidence among European American women. Genetic factors may contribute to this difference. We have investigated the association between genetic ancestry and breast cancer risk among Latina women. In analyses not adjusted for known risk factors, such as reproductive and lifestyle factors, we found a strong association between European genetic ancestry and breast cancer risk. This association was somewhat attenuated after adjustment for known risk factors as expected (7), but it remained significant. When African ancestry was included in the model, the effect of European ancestry was enhanced possibly due to the concomitant decrease in Indigenous American ancestry.

The association between European genetic ancestry and breast cancer needs to be interpreted with caution. There may be unmeasured or unknown risk factors for breast cancer that underlie the association that we observed. The present and previous studies (6, 8) found that breast cancer risk is higher among U.S. born Latinas, which suggests the influence of important unmeasured confounders. For example, place of birth (U.S. born versus foreign born) is significantly associated with breast cancer risk in our multivariate model and is likely to be a marker of some other more proximate risk factor. Similarly, genetic ancestry may be associated with other unmeasured, nongenetic factors that underlie breast cancer risk. Alternatively, our results suggest that there might be genetic variants with different frequencies in Indigenous American and European populations that influence risk for breast cancer. The only way to directly test this is to identify the genetic factors that underlie breast cancer susceptibility among Latinas. Such work is currently under way in a larger Latina population.

An important caveat in interpreting our results is that Indigenous American populations in the United States are diverse and may have some systematic genetic (as well as obvious nongenetic) differences compared with Indigenous American populations in Mexico, Central America, and South America. Wang and colleagues (22) recently explored the population genetics in Amerindian populations from North, Central America, and South America. They found substantial genetic differences among populations in the Americas compared with the differences among Asian or European populations. This may be due to repeated founder effects that occurred during the settlement of the Americas. Thus, even if the association we found is due to genetic factors, it may not be applicable to all indigenous populations in the Americas.

We found no evidence that associations with genetic ancestry differed by tumor characteristics such as hormone receptor status,

**Table 2.** Multivariate logistic regression model of association between genetic ancestry and breast cancer risk ( $n = 975$ )

	OR (95% CI)	$P >  z $
Univariate analysis		
European ancestry*	1.79 (1.28–2.79)	<0.001
Multivariate analysis		
European ancestry	1.39 (1.06–2.11)	0.013
Age at diagnosis	1.02 (1.01–1.04)	0.013
Foreign born	0.73 (0.54–0.99)	0.046
Family history of breast cancer	1.34 (0.88–2.04)	0.160
Benign breast disease	1.12 (0.77–1.59)	0.580
Age at menarche	0.93 (0.86–1.01)	0.074
Hormone replacement therapy use	0.92 (0.68–1.24)	0.570
Daily alcohol intake <sup>†</sup>	1.98 (1.21–3.24)	0.006
Ln daily kilocalorie intake <sup>‡</sup>	1.78 (1.24–2.42)	0.001
Parity	0.86 (0.80–0.94)	<0.001
Breast-feeding per child	0.97 (0.95–1.00)	0.070
Education level	1.11 (0.96–1.28)	0.131

NOTE: Thirty-two cases and 25 controls were excluded from the analysis because of missing data.

\*OR is for every 25% increase in European ancestry.

<sup>†</sup>Daily intake of >10 versus ≤10 g.

<sup>‡</sup>Individuals with daily kilocalorie intake of <600 or >5,000 were excluded from the analysis. Daily kilocalorie intake was log transformed for analysis.

**Table 3.** Tumor characteristics and genetic ancestry for 440 Latinas with breast cancer

Tumor characteristics	<i>n</i> (%)	% Eur (SD)	<i>P</i> *	% Ind (SD)	<i>P</i> *	% Afr (SD)	<i>P</i> *
Estrogen receptor status							
Positive	294 (67)	57 (18)	0.709	35 (17)	0.908	8 (8)	0.239
Negative	83 (19)	58 (20)		36 (19)		6 (7)	
Missing	63 (14)						
Progesterone receptor status							
Positive	246 (56)	56 (18)	0.496	36 (17)	0.982	8 (7)	0.072
Negative	127 (29)	58 (19)		36 (18)		6 (7)	
Missing	67 (15)						
Stage							
Local	267 (61)	58 (19)	0.886	35 (18)	0.961	7 (8)	0.797
Regional extension	11 (2.5)	58 (19)		34 (18)		8 (9)	
Regional nodes	120 (27)	57 (17)		35 (17)		8 (8)	
Regional extension and nodes	14 (3)	55 (14)	0.476	39 (15)	0.343	6 (5)	0.365
Remote	2 (0.5)	55 (9)		41 (14)		4 (5)	
Missing	26 (6)						
Grade							
1	84 (19)	56 (21)	0.503	36 (19)	0.633	8 (9)	0.572
2	163 (37)	59 (19)		33 (18)		8 (8)	
3	142 (32.5)	56 (16)		37 (17)		7 (7)	
4	6 (1.5)	72 (26)	0.255	20 (22)	0.149	8 (7)	0.579
Missing	45 (10)						
Histologic type							
IDC	341 (77.5)	58 (19)	0.683	35 (18)	0.729	7 (7)	0.832
LC	32 (7)	57 (17)		32 (17)		11 (12)	
IDLC	24 (5.5)	63 (20)		29 (17)		8 (9)	
MA	15 (3.5)	51 (16)	0.123	41 (15)	0.136	7 (7)	0.817
Other	28 (6.5)	59 (18)		34 (16)		7 (6)	

Abbreviations: Eur, European ancestry; Ind, Indigenous American ancestry; Afr, African ancestry; IDC, intraductal carcinoma; LC, lobular carcinoma; IDLC, intraductal and lobular carcinoma; MA, mucinous adenocarcinoma.

\**P* value for *t* test comparing mean ancestry of different tumor subtypes to mean ancestry of all cases.

stage, or grade. However, because sample sizes for most of the tumor subtypes were small, further work will be needed to explore the observed trends.

A related question that our study addresses is whether the variation in genetic ancestry among Latina women acts as a confounding factor in genetic association studies of breast cancer. Our results show that such studies may be confounded by genetic ancestry. Without adjustment for genetic ancestry, there was a dramatic deviation from the null hypothesis when testing the association between specific AIMs and breast cancer risk. However, there was no deviation after adjusting for ancestry differences, as expected based on theoretical results (23–29) and previous empirical studies (11–13, 28, 30–32). It is important to note that the AIMs we tested are among the markers that are most likely to be falsely associated with disease precisely because they are strongly correlated with genetic ancestry. However, the bias due to stratification may affect even less informative markers as the sample size increases (27).

We observed a strong association between the number of European-born parents and grandparents and breast cancer risk. This implies that the information provided by Latina women about place of birth of parents and grandparents could be an adequate approximation to genetic ancestry for risk assessment purposes. However, using the number of European parents and grandparents

to adjust the association of individual markers with breast cancer risk, 13 of 106 markers were left significant at  $P < 0.05$ , compared with 4 of 106 markers when genetic ancestry was adjusted for. Thus, use of genetic ancestry in recently admixed populations may provide information above that of grandparents' origin. The four SNPs that had  $P < 0.05$  after adjustment for ancestry are likely to be false positives because they did not achieve significance when we corrected the significant *P* value for multiple testing.

In summary, European genetic ancestry in U.S. Latinas residing in the San Francisco Bay area was associated with increased breast cancer risk after adjustment for known risk factors. Further work is needed to evaluate if the observed association is solely due to differences in nongenetic risk factors not included in the model or to genetic differences between populations.

## Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

## Acknowledgments

Received 5/30/2008; revised 8/11/2008; accepted 9/3/2008.

**Grant support:** Department of Defense Breast Cancer Research Program grants BC030551 (E. Ziv) and DAMD17-96-6071 (E.M. John); National Cancer Institute (NCI) grants K22 CA10935 and R01 CA120120 (E. Ziv) and grants R01 CA63446 and R01

CA77305 (E.M. John); University of California, San Francisco Clinical and Translational Sciences Institute, Career Development Award (L. Fejerman); Postdoctoral Fellowship from Prevent Cancer Foundation (L. Fejerman); California Breast Cancer Research Program grant 7PB-0068 (E.M. John); NIH grant RO1 HL078885; Tobacco-Related Disease Research Program New Investigator Award 15KT-0008 (S. Choudhry); NCI Redes En Acción grant U01-CA86117; and NCI Cooperative agreement no. U01CA069417, RFA #CA-95-011 (to The Northern California site of the Breast Cancer Family Registry).

The content of this article does not necessarily reflect the views or policies of the NCI or any of the collaborating centers in the Breast Cancer Family Registry, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government or the Breast Cancer Family Registry.

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 all the study participants.

## References

- Smigal C, Jemal A, Ward E, et al. Trends in breast cancer by race and ethnicity: update 2006. *CA Cancer J Clin* 2006;56:168–83.
- 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.
- Haas JS, Kaplan CP, Gerstenberger EP, Kerlikowske K. Changes in the use of postmenopausal hormone therapy after the publication of clinical trial results. *Ann Intern Med* 2004;140:184–8.
- Newman LA, Griffith KA, Jatoi I, Simon MS, Crowe JP, Colditz GA. Meta-analysis of survival in African American and white American patients with breast cancer: ethnicity compared with socioeconomic status. *J Clin Oncol* 2006;24:1342–9.
- Newman LA, Mason J, Cote D, et al. African-American ethnicity, socioeconomic status, and breast cancer survival: a meta-analysis of 14 studies involving over 10,000 African-American and 40,000 White American patients with carcinoma of the breast. *Cancer* 2002;94:2844–54.
- 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.
- Ziv E, John EM, Choudhry S, et al. Genetic ancestry and risk factors for breast cancer among Latinas in the San Francisco Bay Area. *Cancer Epidemiol Biomarkers Prev* 2006;15:1878–85.
- John EM, Phipps AI, Davis A, Koo J. Migration history, acculturation, and breast cancer risk in Hispanic women. *Cancer Epidemiol Biomarkers Prev* 2005;14:2905–13.
- Bertoni B, Budowle B, Sans M, Barton SA, Chakraborty R. Admixture in Hispanics: distribution of ancestral population contributions in the Continental United States. *Hum Biol* 2003;75:1–11.
- Bonilla C, Parra EJ, Pfaff CL, et al. Admixture in the Hispanics of the San Luis Valley, Colorado, and its implications for complex trait gene mapping. *Ann Hum Genet* 2004;68:139–53.
- Salari K, Choudhry S, Tang H, et al. Genetic admixture and asthma-related phenotypes in Mexican American and Puerto Rican asthmatics. *Genet Epidemiol* 2005;29:76–86.
- Choudhry S, Coyle NE, Tang H, et al. Population stratification confounds genetic association studies among Latinos. *Hum Genet* 2006;118:652–64.
- Gonzalez Burchard E, Borrell LN, Choudhry S, et al. Latino populations: a unique opportunity for the study of race, genetics, and social environment in epidemiological research. *Am J Public Health* 2005;95:2161–8.
- Tang H, Jorgenson E, Gadde M, et al. Racial admixture and its impact on BMI and blood pressure in African and Mexican Americans. *Hum Genet* 2006;119:624–33.
- 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.
- 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.
- Wittke-Thompson JK, Pluzhnikov A, Cox NJ. Rational inferences about departures from Hardy-Weinberg equilibrium. *Am J Hum Genet* 2005;76:967–86.
- Chakraborty R, Weiss KM. Frequencies of complex diseases in hybrid populations. *Am J Phys Anthropol* 1986;70:489–503.
- Chakraborty R, Kambh MI, Ferrell RE. “Unique” alleles in admixed populations: a strategy for determining “hereditary” population differences of disease frequencies. *Ethn Dis* 1991;1:245–56.
- StataCorp. Stata Statistical Software. 8.2 ed. College Station (TX): StataCorp LP; 2003.
- Team RDC. R: A language and environment for statistical computing. 2.4.0 ed. Vienna, Austria: R Foundation for Statistical Computing; 2006.
- Wang S, Lewis CM, Jakobsson M, et al. Genetic variation and population structure in Native Americans. *PLoS Genet* 2007;3:e185.
- Reich DE, Goldstein DB. Detecting association in a case-control study while correcting for population stratification. *Genet Epidemiol* 2001;20:4–16.
- Pritchard JK, Rosenberg NA. Use of unlinked genetic markers to detect population stratification in association studies. *Am J Hum Genet* 1999;65:220–8.
- Heiman GA, Hodge SE, Gorroochurn P, Zhang J, Greenberg DA. Effect of population stratification on case-control association studies. I. Elevation in false positive rates and comparison to confounding risk ratios (a simulation study). *Hum Hered* 2004;58:30–9.
- Cardon LR, Palmer LJ. Population stratification and spurious allelic association. *Lancet* 2003;361:598–604.
- Marchini J, Cardon LR, Phillips MS, Donnelly P. The effects of human population structure on large genetic association studies. *Nat Genet* 2004;36:512–7.
- Hoggart CJ, Parra EJ, Shriver MD, et al. Control of confounding of genetic associations in stratified populations. *Am J Hum Genet* 2003;72:1492–504.
- Pritchard JK, Donnelly P. Case-control studies of association in structured or admixed populations. *Theor Popul Biol* 2001;60:227–37.
- Tsai HJ, Shaikh N, Kho JY, et al.  $\beta$ 2-adrenergic receptor polymorphisms: pharmacogenetic response to bronchodilator among African American asthmatics. *Hum Genet* 2006;119:547–57.
- Tsai HJ, Kho JY, Shaikh N, et al. Admixture-matched case-control study: a practical approach for genetic association studies in admixed populations. *Hum Genet* 2006;118:626–39.
- Satten GA, Flanders WD, Yang Q. Accounting for unmeasured population substructure in case-control studies of genetic association using a novel latent-class model. *Am J Hum Genet* 2001;68:466–77.