Genetic Ancestry and Risk Factors for Breast Cancer among Latinas in the San Francisco Bay Area

Elad Ziv,[1,2,3,5] Esther M. John,[6] Shweta Choudhry,[2,5] Jennifer Kho,[2] Wendy Lorizio,[1,2] Eliseo J. Perez-Stable,[1,2,5] and Esteban Gonzalez Burchard[2,3,4]

[1]Division of General Internal Medicine, Departments of [2]Medicine and [3]Biopharmaceutical Sciences, [4]Center for Human Genetics, and [5]Comprehensive Cancer Center, University of California San Francisco, San Francisco, California and [6]Northern California Cancer Center, Fremont, California

Abstract

Background: **Genetic association studies using case-control designs are susceptible to false-positive and false-negative results if there are differences in genetic ancestry between cases and controls. We measured genetic ancestry among Latinas in a population-based case-control study of breast cancer and tested the association between ancestry and known breast cancer risk factors. We reasoned that if genetic ancestry is associated with known breast cancer risk factors, then the results of genetic association studies would be confounded.**

Methods: **We used 44 ancestry informative markers to estimate individuals' genetic ancestry in 563 Latina participants. To test whether ancestry is a predictor of hormone therapy use, parity, and body mass index (BMI), we used multivariate logistic regression models to estimate odds ratios (OR) and 95% confidence intervals (95% CI) associated with a 25% increase in Indigenous American ancestry, adjusting for age, education, and the participant's and grandparents' place of birth.**

Results: **Hormone therapy use was significantly less common among women with higher Indigenous American ancestry (OR, 0.78; 95% CI, 0.63-0.96). Higher Indigenous American ancestry was also significantly associated with overweight (BMI, 25-29.9 versus <25) and obesity (BMI, ≥30 versus <25), but only among foreign-born Latina women (OR, 3.44; 95% CI, 1.97-5.99 and OR, 1.95; 95% CI, 1.24-3.06, respectively).**

Conclusion: **Some breast cancer risk factors are associated with genetic ancestry among Latinas in the San Francisco Bay Area. Therefore, case-control genetic association studies for breast cancer should directly measure genetic ancestry to avoid potential confounding.** (Cancer Epidemiol Biomarkers Prev 2006;15(10):1878–85)

Introduction

Breast cancer incidence and mortality rates vary substantially among different racial and ethnic groups in the United States (1) In the San Francisco Bay Area, Latina women have incidence rates that are ~35% lower than the rates of Caucasian women (2). Latinos are known to be an admixed population with genetic ancestry from Europeans, Indigenous Americans, and Africans (3-7). The proportion of these ancestral contributions vary depending on the country and region of origin of individuals (8). In the San Francisco Bay Area, most Latinas are of Mexican or Central American descent. These women are, in turn, known to be of mostly European and Indigenous American ancestry.

In genetic association studies of cases (individuals with the disease of interest) and controls (individuals without the disease), admixture may lead to false-positive or false-negative results if cases and controls differ in their genetic ancestry (7, 9, 10). The degree to which such confounding would occur depends on whether genetic ancestry is associated with the disease under study (10, 11). If genetic ancestry is associated with disease, because of either genetic or environmental differences between ancestral groups in the admixed population, the likelihood of both false-positive results and false-negative results will be increased in case-control studies (11). It is important to note that the association between genetic ancestry and a trait may be due to non-genetic risk factors. For example, if certain environmental risk factors are more common in a population with one ancestry, then case-control association studies of genetic variants would still be confounded.

There has been considerable controversy about the degree to which population stratification may affect case-control studies of cancer (12-14), but there is little data on how stratification may affect cancer studies among Latinos. Because they are genetically admixed and have significantly lower breast cancer incidence rates than Whites or African Americans, Latinas offer an opportunity to address this question. We used a series of ancestry informative markers to estimate the genetic ancestry of 241 Latina breast cancer cases and 333 age-matched Latina population controls to determine the distribution of genetic ancestry and evidence for substructure and recent admixture among Latinas and to test the association of genetic ancestry with breast cancer risk factors. We reasoned that if risk factors for breast cancer are associated with genetic ancestry, then genetic association studies in this population are likely to be confounded by ancestry.

Materials and Methods

Population. The participants in this study are from a multiethnic population-based case-control study of breast cancer in the San Francisco Bay Area described elsewhere (15, 16). Briefly, Latina, African American, and White women with invasive breast cancer diagnosed from 1995 to 2002 were identified through the Greater Bay Area Cancer Registry, and controls were identified through random-digit dialing and frequency matched to cases on race/ethnicity and 5-year age group. This analysis is based on a subgroup of cases diagnosed

Received 2/8/06; revised 7/28/06; accepted 8/14/06.

Grant support: National Cancer Institute grant K22CA109351, Department of Defense Breast Cancer Research Program grant BC030551, American Cancer Society grant CCTG02-084-01-CCE, and Sandler Family Foundation.

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.

Note: A preliminary version of some of the results in this article was presented at the American Society of Human Genetics Meeting, Salt Lake City, Utah, 2005.

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: 415-353-7932. E-mail: elad.ziv@ucsf.edu

doi:10.1158/1055-9965.EPI-06-0092

from 1997 and 1999 and matched controls who provided a blood sample.

Cases. Of 4,842 cases identified through the cancer registry, 618 (13%) could not be contacted (168 deceased, 71 physician refusal, 379 moved or lost). A brief telephone screening interview that assessed self-identified race/ethnicity and study eligibility was completed by 90% of cases. Of 357 cases who self-identified as Hispanic or Latina, 324 (91%) completed the in-person interview, and 241 (68%) provided a blood sample.

Controls. From the pool of eligible women identified through random-digit dialing (81% response to household enumeration), 1,479 were selected as controls. Of these, 103 (7%) could not be contacted (9 deceased, 94 moved or lost), and of the remaining controls, 93% completed the screening interview. Of 479 controls who self-identified as Hispanic or Latina, 421 (88%) completed the interview, and 333 (70%) provided a blood sample.

The analysis was based on 563 Latinas, after excluding 10 women who reported being born in or having all four grandparents from Spain or the Philippines and one woman who did not give any information about her own and her grandparents' place of birth.

Among foreign-born women, 175 were from Mexico, 91 from Central America (44 from El Salvador, 27 from Nicaragua, 12 from Guatemala, 4 from Costa Rica, 3 from Panama, and 1 from Honduras), 30 from South America (8 from Columbia, 8 from Peru, 4 from Argentina, 4 from Ecuador/Galapagos, 3 from Chile, 1 from Bolivia, 1 from Brazil, and 1 from Uruguay), and 10 from the Caribbean (5 from Puerto Rico, 4 from Cuba, 1 from the Dominican Republic). U.S.-born women were categorized further based their report of grandparents' place of birth, including Mexico ($n = 100$), Central America ($n = 2$), and Caribbean ($n = 8$), and those who reported grandparents from more than one region were classified as ''mixed origin'' ($n = 116$). Women with all four grandparents born in the United States were grouped separately ($n = 31$).

Institutional review boards at all participating institutions approved the study and all of the participants gave written informed consent.

Data Collection. An extensive structured questionnaire was given in participants' homes by bilingual and bicultural professional interviewers in English or Spanish to collect information on parents' and grandparents' country of birth, residential history, family history of breast cancer, menstrual and reproductive history, hormone therapy use, and other lifestyle factors. Body mass index (BMI) was calculated as weight (kg) divided by height (m) squared as measured by the interviewer at the time of the interview.

Marker Selection. Forty-four markers were preselected to be informative for either Indigenous American-European ancestry differences, European-African ancestry differences, or Indigenous American-African ancestry differences. Markers were identified either based on previous reports in the literature or differences in allele frequency in existing databases as previously described (6). For each marker, the allele frequency in ancestral populations were confirmed in samples from European populations ($n = 243$): Ireland, England, Germany, and Spain; Indigenous American populations from the western United States, Mexico, and Central America ($n = 148$): Maya, Pima, Cheyenne, and Pueblo; and sub-Saharan African populations ($n = 481$): Nigeria, Central African Republic, and Sierra Leone. The mean allele frequency difference for all 44 markers is 0.30 between the European and Indigenous American populations, 0.42 between the European and African populations, and 0.42 between the African American and Indigenous American populations.

Genotyping. Genotyping was done using single base extension, and detection of specific alleles was done by fluorescence polarization. A complete list of primers and conditions used for these primers is available in Salari et al. (6).

Statistical Analysis. We used χ^2 tests to test for deviations from Hardy-Weinberg equilibrium. We also examined the proportion of heterozygotes in the expected minus the observed data as a means of detecting the direction of deviation from Hardy-Weinberg equilibrium because population substructure leads to excess homozygosity (17). We also tested for allelic association among all pairs of markers on different chromosomes using Δ_{AB}, as described by Weir et al.(18), which tests for correlation among genotypes. Allelic association among markers that are physically unlinked implies the presence of population substructure (i.e., nonrandom mating) and/or recent admixture (19). Population substructure and recent admixture without substructure would both lead to variation in individual ancestry and the possibility of association between ancestry and the phenotype of interest.

We estimated individual genetic ancestry by a maximum likelihood approach (20, 21). Briefly, for each genotype, an expression of the likelihood of origin from each of three populations is derived, based on the allele frequencies in the ancestral populations. The sum of the log-likelihoods for all genotypes for an individual is maximized over the range of possible values of ancestry. A program for maximum likelihood estimation of ancestry in JAVA is available from the authors upon request. We also used *structure*, a population genetics program that implements a Bayesian approach to infer individual ancestry (22), as a complimentary method of analysis. For *structure*, we inputted the ancestral population data as part of the genotype file but did not use population labels. Thus, the inference about population membership for both ancestral populations and for the cases and controls was based on the *structure* inference alone. To compare ancestry among subgroups of Latinas, we used rank sum tests. We compared the results of the structure analysis and the maximum likelihood analysis and found very strong correlations between these two approaches for % European ancestry ($r = 0.95$), Indigenous American ancestry ($r = 0.95$), and African ancestry ($r = 0.94$). Given the high correlation, we present results from the maximum likelihood analysis only.

We used ANOVA to test the association between ancestry and various breast cancer risk factors, including BMI, parity (number of full-term pregnancies), age at first full-term pregnancy, history of hormone therapy use (yes/no), age at menarche, and age at menopause (surgical or natural). Because the purpose of the current study is to examine the association between risk factors and genetic ancestry, we included both breast cancer cases and controls in the ANOVA, adjusting for case/control status. We used logistic regression models to further explore associations of ancestry with overweight (BMI 25-29.9 versus <25), obesity (BMI ≥30 versus <25), parity (<2 versus ≥2), and history of hormone therapy use (yes versus no), adjusting for age, education (some high school or less, high school graduation, some college, and college graduation or higher), and grandparents' country of birth (Mexico, Central America, South America, Caribbean, United States, or mixed). In addition, logistic regression models that included women born outside of the United States adjusted for age at migration to the United States, and models that included all women adjusted for place of birth (foreign born versus U.S. born).

We also examined whether the association between individual ancestry informative markers and breast cancer risk is modified by adjustment for individual ancestry. We tested the association between each marker and breast cancer risk using logistic regression models and entering each marker into the model using an additive model. We then tested the association

Table 1. Characteristics of study population

Characteristic	Controls (n = 329)	Cases (n = 234)	P*
Age, mean (range)	53.7 (35-79)	55.3 (35-80)	0.08
Foreign born, % (n)	60.8 (200)	45.3 (106)	<0.001
Grandparents' place of birth, % (n)			
Mexico	51.4 (169)	45.3 (106)	
Central America	18.8 (62)	13.3 (31)	0.37
South America	4.6 (15)	6.4 (15)	0.22
Caribbean	3.7 (12)	2.7 (6)	0.66
United States	5.2 (17)	6.0 (14)	0.47
Mixed origin†	16.4 (54)	26.5 (62)	0.007
Education, % (n)			
Some high school or less	51.4 (169)	34.2 (80)	
High school graduation	17.3 (57)	20.9 (49)	0.011
Some college	16.1 (53)	29.9 (70)	<0.001
College graduation or higher	15.2 (50)	14.5 (35)	0.13
No. full-term pregnancies, mean ± SD	3.4 ± 2.4	2.7 ± 1.8	0.0012
Age at first full-term pregnancy, mean ± SD	22.9 ± 5.5	23.7 ± 5.3	0.15
History of hormone therapy use, % (n)	38.2 (124)	45.0 (103)	0.11
Age at menarche, mean ± SD	12.7 ± 1.6	12.5 ± 1.8	0.06
Postmenopausal status, % (n)	55.9 (184)	57.7 (135)	0.34
Age at natural menopause, mean ± SD	47.4 ± 5.3	48.6 ± 5.0	0.11

*Ps for comparison of proportions were obtained using χ^2 tests. Ps for comparison of continuous variables were obtained using t tests. Ps for comparison of ordinal variables (number of full-term pregnancies) were obtained using rank sum tests.
† Women with grandparents from more than one of the five regions (Mexico, Central America, South America, Caribbean, United States).

between each marker and breast cancer risk, adjusting for individual ancestry, using both African and Native American ancestry in the model. (Entering all three ancestral components into the model is not possible because knowledge of two of the ancestral components perfectly determines the third component). For each marker, we determined the negative log of the P value as a summary estimate of the strength of the association before and after adjustment for ancestry. We carried out the same analysis for BMI using logistic regression models, which compared women in the normal weight group (BMI < 25) with obese women (BMI > 30).

All statistical tests were two sided and done using Stata (version 8.0).

Results

The majority of the women were of Mexican and Central American descent, and approximately half were foreign born (Table 1). The largest contributions to ancestry were European and Indigenous American, with a small contribution of African ancestry (Table 2). Compared with the ancestry of women born in Mexico, the largest group, Indigenous American ancestry was significantly lower in women from the Caribbean and U.S.-born women with grandparents born in the United States or of mixed origin. Conversely, these groups, as well as U.S.-born women with Mexico-born grandparents, had significantly higher European ancestry. African ancestry was significantly higher in women from Central America and the Caribbean. We also found substantial variation in individual ancestry within each of the groups, with most of the variation between European and Indigenous American ancestry (Fig. 1).

Individual marker analysis suggested that there is significant population substructure and recent admixture in this population. Populations with substructure are expected to have excess homozygosity in comparison with expectations under Hardy-Weinberg equilibrium. Fourteen of the 44 markers tested were significantly ($P < 0.05$) out of Hardy-Weinberg equilibrium; of these, 13 showed excess homozygosity (Table 3). In addition, of the 30 markers not significantly out of Hardy-Weinberg equilibrium, 23 also showed excess homozygosity. When testing for allelic association (linkage disequilibrium) between the markers, we found that 179 (18%) of the 946 possible pairs of markers showed statistically significant ($P < 0.05$) linkage disequilibrium.

Indigenous American ancestry was associated with several breast cancer risk factors (Table 4). Indigenous American ancestry was higher in women with low education, high parity, and high BMI and in women without a history of hormone therapy use. We also did analyses between ancestry and these risk factors separately for cases and controls, although such analyses have considerably lower sample size. In analyses including only controls, we also found significant associations of higher Indigenous American ancestry with low education ($P = 0.006$), high parity (≥2 full-term pregnancies; $P = 0.04$), and high BMI (BMI ≥25; $P = 0.01$) and a nonsignificant association with no history of hormone therapy use ($P = 0.2$). In analysis of cases only, we found associations in the same direction: significant associations with high BMI ($P = 0.01$) and no history of hormone therapy use ($P = 0.02$) and nonsignificant associations with low education ($P = 0.1$) and high parity ($P = 0.1$). We found no evidence for a statistical interaction among genetic ancestry, case/control status, and any of these risk factors, although tests for interaction in this data set are likely to be underpowered.

The inverse association between Indigenous American ancestry and history of hormone therapy use was only significant among women who were born in the United States (Table 5) and was not attenuated when adjusting for education and grandparents' country of birth. Duration of hormone therapy use did not vary by genetic ancestry (data not shown). The association between parity and Indigenous American ancestry was only significant among women born outside of the United States and was attenuated by adjustment for education and grandparents' country of birth (Table 5).

Indigenous American ancestry varied significantly by BMI, with the lowest proportion found in women with normal

Table 2. Mean and SDs of individual ancestry estimates by participant's and grandparents' place of birth

Place of birth of participant and grandparents	n	African Ancestry (%)	P*	European Ancestry (%)	P*	Indigenous American Ancestry (%)	P*
Mexican born	175	3.7 ± 5.5		52.0 ± 20.1		44.3 ± 20.1	
U.S. born, grandparents Mexican born	100	4.8 ± 6.4	0.16	47.3 ± 18.7	0.047	47.8 ± 18.8	0.095
Central American born and U.S. born, grandparents born in Central America	93	10.2 ± 13.4	0.0001	47.5 ± 22.1	0.094	42.3 ± 22.3	0.49
Caribbean born	10	20.2 ± 26.2	0.005	65.9 ± 24.7	0.06	13.9 ± 13.6	0.0001
U.S. born, grandparents Caribbean born	8	11.6 ± 10.3	0.03	67.0 ± 18.5	0.05	21.4 ± 23.5	0.006
South American born	30	3.2 ± 5.1	0.57	55.9 ± 26.1	0.53	40.9 ± 24.5	0.66
U.S. born, grandparents U.S. born	31	4.2 ± 5.4	0.59	60.5 ± 22.1	0.039	35.4 ± 21.1	0.028
U.S. born, grandparents mixed origin†	116	4.8 ± 8.6	0.30	58.0 ± 23.5	0.029	37.3 ± 22.8	0.007

*Ps are for comparison with Mexican-born Latinas.
† Women with grandparents from more than one of the five regions (Mexico, Central America, South America, Caribbean, United States).

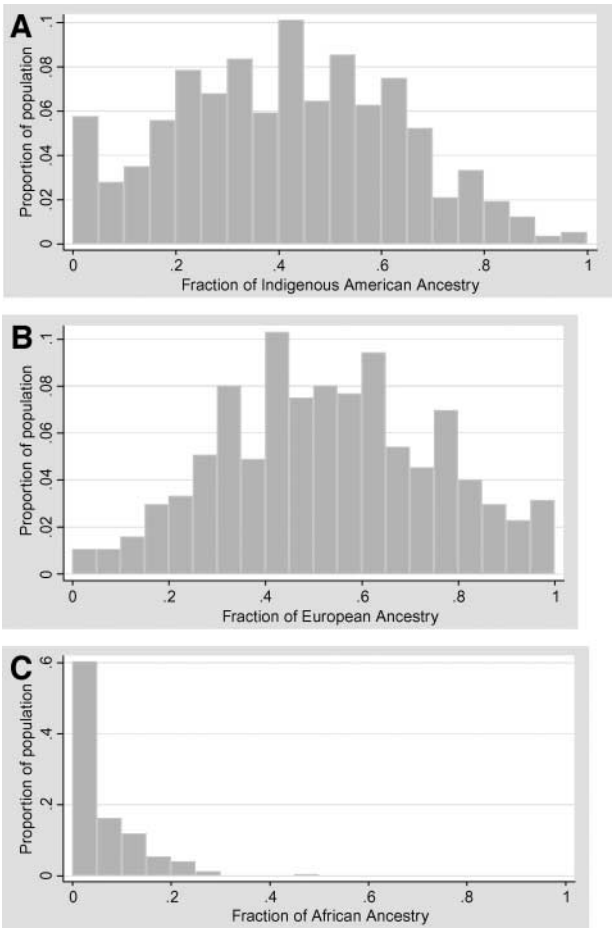

Figure 1. Distribution of individual ancestry estimates. Each graph represents the distribution of ancestry for Indigenous American, European, or African ancestry.

weight (BMI < 25). However, obese women did not have higher Indigenous American ancestry than overweight women (Table 4). Furthermore, the association between BMI and Indigenous American ancestry was significant only in foreign-born Latinas (Table 5; $P_{interaction} = 0.02$).

Because Indigenous American and European ancestry are inversely related (see Fig. 1), the risk factors associated positively with Indigenous American ancestry were inversely associated with European ancestry. We found no significant associations between African ancestry and breast cancer risk factors.

We tested the association for each ancestry informative marker and breast cancer risk before and after adjustment for ancestry. To compare the association before and after adjustment, we plotted the negative log *P*s for association with each marker. Figure 2 represents the results of this analysis, with each dot on Fig. 2 representing one marker's negative log *P* for association before (on the *x*-axis) and after (on the *y*-axis) adjustment. A negative log $P > 1.3$ is equivalent to $P < 0.05$. Only 2 of the 44 markers were significantly associated with breast cancer risk ($P < 0.05$). After adjustment, only one of these markers remained significant. Overall adding ancestry to the models had little effect on the overall distribution of associations with these markers (Fig. 2A).

We also tested the association of each ancestry informative marker with obesity (comparing women with BMI < 25 with women with BMI ≥ 30) before and after adjustment for ancestry. Five of the 44 markers were significantly associated with obesity before adjustment. However, after adjustment, only 3 of 44 markers remained significant. Furthermore, the overall distribution of associations seems substantially more attenuated after adjustment for ancestry in the models (Fig. 2B).

Discussion

We analyzed genetic admixture in Latina women living in the San Francisco Bay Area with the purpose of determining whether case-control association studies could be confounded by genetic ancestry in this population. The present analysis of population substructure reveals substantial excess homozygosity among many of the 44 markers tested. Deviations from Hardy-Weinberg equilibrium may be due to genotyping errors, chance, or non-randomly mating subgroups within a population. The large number of markers with a significant deviation towards excess homozygosity and a trend in the same direction for the majority of other markers strongly suggests that non-randomly mating subgroups are present. The finding of allelic association (linkage disequilibrium) between the ancestry informative markers is also consistent with population substructure and/or recent admixture. The analysis of admixture reveals differences in individual ancestry between immigrants from different regions in Latin America and the Caribbean. These differences are consistent with the ancestral proportions seen in prior studies. In particular, Indigenous American ancestry was lower, and African ancestry was higher among Caribbean populations compared with Mexican and Central American populations, as has been reported by others (3-7, 23-26). In addition, we observed higher African ancestry and lower European ancestry in Central American 1st-generation immigrants compared with Mexican 1st-generation immigrants. This difference may either be due to true differences in the proportions of ancestry between Mexican and Central American populations, or due to differences in migration patterns in that individuals from particular ancestral backgrounds may be more likely to immigrate to the United States. Among Mexican Americans, the largest group in our study, we detected a significantly higher proportion of European ancestry among 1st-generation immigrants compared with Mexican Americans born in the United States. This suggests that different subgroups of Mexicans may have been migrating over different generations. This could either be due to migrations from different geographic regions, different socioeconomic groups, or both.

This analysis also identified extensive diversity in genetic ancestry within immigrants from each region. Such diversity creates the possibility of false-positive and false-negative association results in case-control studies. However, for false-positive and false-negative results to occur, there also needs to be an association between breast cancer risk and ancestry (either due to genetic or environmental differences between ancestral groups in the admixed population).

We found significant associations between several risk factors for breast cancer and genetic ancestry among Latinas. Hormone therapy use was lower among Latinas with higher Indigenous American ancestry or lower European ancestry. Even among Latinas born in the United States, and after adjustment for education and grandparents' country of birth, there remained an association between higher Indigenous American ancestry and lower hormone therapy use. Thus, there are other factors, possibly either related to cultural factors or differential access to care, that are associated with lower likelihood of hormone therapy use among U.S. born Latinas with higher Indigenous American ancestry. We also found an association between parity and higher Indigenous American ancestry. However, this association was largely accounted for by differences in education and grandparents' country of birth.

Clearly, the associations of less hormone therapy use and higher parity with Indigenous American ancestry are due to non-genetic differences among Latinas of different ancestral backgrounds. However, these associations may also confound genetic association studies. For example, Latinas with higher Indigenous American ancestry who are less likely to have used hormone therapy and more likely to have had higher parity may be at lower risk of breast cancer. Thus, any allele that is at higher frequency in the European ancestral population may be associated with breast cancer due to a confounding effect from differences in non-genetic factors.

Adjusting for known non-genetic risk factors should eliminate confounding by such factors in genetic association studies. However, because not all risk factors are known for breast cancer, adjusting for the currently known risk factors may not completely eliminate confounding. Among Latinas, unknown risk factors that correlate with country of birth, socioeconomic status, and acculturation may be important determinants of breast cancer risk. For example, Latinas who are primarily Spanish speaking are at lower risk of breast cancer compared with Latinas who are English speaking, even after adjustment for all known risk factors (16). Thus, other unknown environmental risk factors for breast cancer may exist in this population, and these factors may also be associated with cultural practices and genetic ancestry. Therefore, our results imply that genetic association studies in unrelated individuals from Latina populations should include measurement of genetic ancestry to avoid confounding.

A stronger case for confounding in genetic association studies among Latinas can be made if markers at candidate genes are associated with breast cancer, and it can be shown that these associations are attenuated after adjustment for genetic ancestry. The present results show a substantial effect of admixture on the association between ancestry informative markers and obesity; the magnitude of association was diminished for most markers. Furthermore, because our ancestry estimates is free of error, which tends to diminish the effect of the adjustment, it is possible that we would have had even greater adjustment had we used more markers for a more precise estimate of ancestry. Our study did not find any substantial effect of adjustment for ancestry on the association between breast cancer and these ancestry informative markers. It is possible, however, that more modest effects of adjustment for ancestry would have been detected with a larger sample size.

Table 3. List of markers and deviations from Hardy-Weinberg equilibrium

dbSNP accession	Location	African	European	Indigenous American	*P*-value for deviation from Hardy-Weinberg equilibrium	Excess homozygotes
rs1042602	11q21	0	0.47	0.05	0.95	0.64
rs1079598	11q23.1	0.06	0.14	0.63	0.59	5.39
rs140864	1p34.3	0.11	0.01	0.55	0.004	24.16
rs146026	13q13.1	0.26	0.92	0.83	0.77	−2.01
rs16383	22q13.2	0.27	0.8	0.11	0.10	18.89
rs17203	3p12.3	0.81	0.15	0.76	0.12	17.72
rs1800404	15q13.1	0.14	0.72	0.48	0.90	1.35
rs1800498	11q23.1	0.14	0.65	0.09	0.50	7.55
rs1985080	7p14.3	0.1	0.64	0.97	0.87	1.39
rs1989486	19q13.42	0.04	0.58	0.4	0.90	1.43
rs203096	17q21.33	0.65	0.72	0.28	0.94	−0.92
rs2065160	1q32.1	0.5	0.92	0.17	0.21	14.69
rs2077863	18p11	0.51	0.93	0.93	0.17	4.82
rs2161	7q22.1	0.44	0.3	0.62	<0.001	78.49
rs2228478	16q24.3	0.51	0.14	0.04	0.92	0.49
rs223830	16q13	0.03	0.19	0.64	0.03	23.96
rs235936	21q21.3	0.18	0.49	0.37	0.40	9.83
rs2695	9q21.31	0.81	0.86	0.22	0.003	34.80
rs2763	7p22.3	0.14	0.16	0.52	0.009	27.92
rs2814778	1q23.2	0	0.99	0.99	0.008	6.59
rs2816	17p12	0	0.49	0.08	0.65	4.36
rs285	8p21.3	0.97	0.52	0.45	0.27	−12.84
rs2862	15q14	0.38	0.17	0.69	0.02	26.98
rs2891	17p13.2	0.02	0.51	0.43	0.35	10.94
rs3188519	1p34.1	0.76	0.37	0.32	0.94	0.84
rs3188520	20q11.22	0.83	0.35	0.44	0.69	4.31
rs326946	11q23.1	0.61	0.17	0.07	<0.001	20.72
rs3287	2p16.1	0.73	0.2	0.21	0.83	1.73
rs3309	5q11.2	0.4	0.28	0.69	0.97	0.39
rs3317	5q23.1	0.05	0.59	0.73	0.10	18.51
rs3340	5q33.2	0.06	0.19	0.65	0.005	32.65
rs4646	15q21.2	0.32	0.29	0.72	0.23	13.77
rs4884	19q13.32	0.15	0.29	0.86	0.41	9.77
rs518116	9q33.3	0.13	0.67	0.58	0.87	−1.78
rs584059	3q22.3	0.49	0.14	0.47	0.45	6.95
rs594689	11q11	0.09	0.47	0.13	0.012	−26.01
rs6003	1q31.3	0.7	0.08	0.03	0.95	−0.26
rs7041	4q13.3	0.93	0.41	0.45	0.03	25.41
rs722098	21q21.1	0.9	0.18	0.72	0.02	25.91
rs723632	1q32.3	0.1	0.92	0.67	0.35	6.54
rs7349	10p11.22	0.04	0.87	0.96	<0.001	15.36
rs736394	14q32.12	0.52	0.74	0.99	0.02	12.71
Rs930072	5p13.2	0.96	0.1	0.45	0.87	1.67
Rs994174	10q23.1	0.76	0.25	0.26	0.73	3.78

NOTE: The 3rd through 5th columns give the frequency of a reference allele frequency in each of the population (reference alleles are as described in Salari et al. (6)). The last column represents the difference between the number of observed individuals with homozygous genotypes subtracted from the number of individuals expected to have homozygous genotypes under Hardy-Weinberg equilibrium.

Table 4. Association between Indigenous American ancestry and demographic variables and reproductive risk factors for breast cancer

	% Indigenous American ancestry (*n*)			*P**
	Controls	Cases	All	
Place of birth				
Foreign born	42.6 (200)	42.6 (106)	42.6	0.24
U.S. born	40.1 (129)	40.8 (128)	40.4	
Education				
Some high school or less	45.3 (169)	45.9 (80)	45.5	0.001
High school graduation	40.9 (57)	38.0 (49)	39.6	
Some college	35.7 (53)	37.7 (70)	36.8	
College graduation or higher	36.2 (50)	44.5 (35)	39.6	
Full-term pregnancies				
>2	43.3 (209)	43.7 (124)	43.5	0.01
≤2	38.6 (120)	39.2 (110)	38.9	
Age at first full-term pregnancy				
<30	42.3 (269)	42.7 (173)	41.9	0.50
≥30	41.4 (60)	38.6 (61)	40.4	
Age at menarche				
>12	42.0 (179)	41.3 (101)	41.7	0.89
≥12	41.1 (150)	41.8 (133)	41.5	
BMI				
<25	34.5 (47)	33.2 (48)	33.8	0.0005
25-29.9	45.7 (121)	44.3 (71)	45.2	
30-35	41.7 (86)	42.6 (59)	42.1	
>35	39.3 (75)	44.3 (56)	41.4	
History of hormone therapy use				
Yes	39.4 (124)	37.9 (103)	38.7	0.01
No	42.9 (201)	44.4 (126)	43.5	
Age at menopause†				
<50	45.3 (99)	39.1 (65)	42.8	0.15
≥50	38.3 (55)	39.0 (46)	38.7	

*Ps reported are the significance levels for association between ancestry and the risk factor, using ANOVA and adjusting for case/control status. There were no significant interaction terms among case/control status, genetic ancestry, and any of these risk factors.

† Age at natural or surgical menopause.

Higher BMI has generally been associated with increased risk of breast cancer in postmenopausal women. This analysis identified an association between BMI and Indigenous American ancestry among foreign-born Latinas, with a higher proportion of Indigenous American ancestry found among overweight and obese women. This association remained significant when adjusting for country of origin. However, there may be non-genetic explanations for this association; unmeasured differences in dietary habits and physical activity associated with culture, education, and socioeconomic status may underlie these differences in part. In addition, our observation that the association with BMI and ancestry differs by place of birth suggests that even if this is a genetic effect, it may be modified by environmental factors.

Several studies of cancer have shown that genetic ancestry could be a source of confounding. Kittles et al. showed that among African-American prostate cancer cases and controls, the distribution of ancestry informative markers was significantly different, suggesting that case-control studies of prostate cancer among African Americans may be confounded by genetic ancestry (27). Similarly, Freedman et al. showed confounding in a case-control study of prostate cancer among African Americans (28). Barnholtz-Sloan et al. showed significant confounding by genetic ancestry, even after accounting for self-reported race/ethnicity in a study of lung cancer (29). We have previously shown that studies of asthma susceptibility (7) and severity (6) are potentially confounded by genetic ancestry among Latinos.

Because the populations we studied were mainly of European and Indigenous American ancestry, our study has good statistical power to assess associations with differences between European and Indigenous American ancestry. Although we found no significant associations between African ancestry and breast cancer risk factors, the population in this study had relatively little African ancestry, which limited our ability to draw conclusions about associations with African ancestry.

This study was also limited by the use of only 44 markers to assess genetic ancestry. Measurement of genetic ancestry with markers is always associated with a certain amount of random error, due to the limited number of markers used and the imperfect information from each marker (11, 30). The greater the number of markers and the more informative each marker is for ancestry, the lower the error is. Simulation studies of a three-population model with a range of ancestry informative markers similar to the one used in this study suggest that the correlation coefficient between the ancestry estimate and the true ancestry is ~0.8 (31). Because the error in the estimate of ancestry is random with respect to the associations we tested, it should generally bias our results towards the null hypothesis. Thus, the associations we observed in this study are likely to represent conservative estimates.

The present analysis of genetic ancestry and breast cancer risk factors combined cases and controls and adjusted for case/control status. However, if there were interactions among case/control status, genetic ancestry, and any of these risk factors, an adjusted analysis would be flawed. We found no significant interactions among any of these risk factors, genetic ancestry, and case/control status.

This study included Latina women living in the San Francisco Bay Area. Other regions of the United States have different distributions of Latinos from Mexico, Central America, South America, and the Caribbean. Therefore, the distributions of ancestry and the association with breast cancer risk factors we identified are likely to vary in

Table 5. Association between Indigenous American ancestry (per 25% increase) and breast cancer risk factors: hormone therapy use, lower parity, and BMI

	Odds ratio (95% confidence interval)	Adjusted odds ratio (95% confidence interval)
History of hormone therapy use (yes vs no)		
All Latinas	0.80 (0.66-0.97)	0.78 (0.63-0.96)*
U.S.-born Latinas	0.71 (0.54-0.96)	0.70 (0.51-0.95)†
Foreign-born Latinas	0.89 (0.68-1.16)	0.86 (0.64-1.16)‡
Low parity (<2 vs 2 full-term pregnancies)		
All Latinas	0.78 (0.65-0.95)	0.90 (0.72-1.12)*
U.S.-born Latinas	0.95 (0.72-1.26)	1.04 (0.76-1.44)†
Foreign-born Latinas	0.66 (0.51-0.87)	0.78 (0.55-1.04)‡
Overweight (BMI 25-29.9 vs <25)		
All Latinas	1.95 (1.42-2.67)	1.93 (1.38-2.69)*
U.S.-born Latinas	1.22 (0.80-1.85)	1.25 (0.79-1.96)†
Foreign-born Latinas	3.29 (1.97-5.48)	3.44 (1.97-5.99)‡
Obesity (BMI ≥30 vs <25)		
All Latinas	1.51 (1.15-1.99)	1.51 (1.12-2.04)*
U.S.-born Latinas	1.20 (0.83-1.75)	1.26 (0.83-1.92)†
Foreign-born Latinas	1.94 (1.29-2.93)	1.95 (1.24-3.06)‡

*Adjusted for age, case/control status, grandparents' place of birth (Mexico, Central America, South America, Caribbean, United States, mixed origin), age at migration, education, and birthplace (U.S. born vs foreign born).

† Adjusted for age, case/control status, grandparents' place of birth (Mexico, Central America, South America, Caribbean, United States, mixed origin), and education.

‡ Adjusted for age, case/control status, grandparents' place of birth (Mexico, Central America, South America, Caribbean, United States, mixed origin), age at migration, and education.

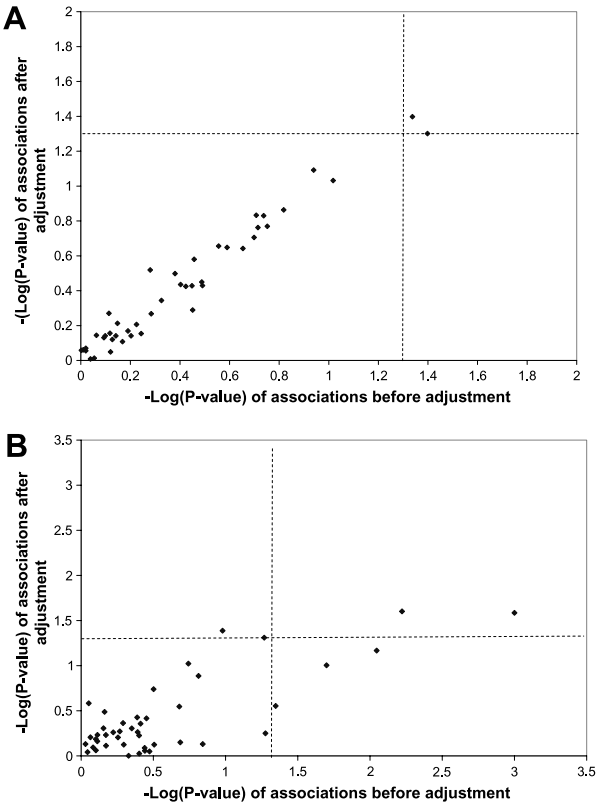

Figure 2. Association between markers and breast cancer (*A*) and between individual markers and obesity (*B*). Each point represents a single marker with association before ancestry adjustment (*x-axis*) and after ancestry adjustment (*y-axis*). Dotted lines are the threshold of nominal significance ($P < 0.05$).

different regions in the United States. In addition, the associations we identified are likely due to various degrees of acculturation, which may differ in different regions of the United States.

We used grandparents' country of birth to assess whether genetic ancestry would provide additional information. A further question, not directly assessed in this study, is whether self-report of grandparents' or parents' ethnicity (e.g., European, Indigenous, and African) is a reliable estimate of genetic ancestry. Williams et al. reported that among Pima Indians, self-reported ancestry based on grandparents' origin performs similarly to genetic ancestry based on 12 markers. We did not directly ask women about their parents' and grandparents' Indigenous American, European, or African ancestry. The degree to which this would be well known depends on how recent admixture has been in these populations. We are not aware of any empirical data comparing self-reported ethnicity/population ancestry with genetic assessment among Latinos.

The chance of a false-positive result due to differences in the genetic ancestry of cases and controls increases as the sample size increases (32). Because the goal of genetic association studies for complex traits, including breast cancer, is often to identify relatively modest effects, the sample sizes planned for such studies are often in the thousands (33). Even if there are subtle associations between ancestry and the phenotype of interest, the chance of false positives may be magnified by the large sample sizes required for such studies.

In summary, we identified substantial variation in individual ancestry among Latina women living in the San Francisco Bay Area and associations between genetic ancestry and hormone therapy use and BMI. These results suggest that population stratification may affect the results of genetic association studies for breast cancer among Latinas. Therefore, such studies should collect information about genetic ancestry to assess and adjust for differences in ancestry between cases and controls.

References

1. Ries L, Eisner M, Kosary C, et al. SEER cancer statistics review, 1975–2002. Bethesda (MD): National Cancer Institute; 2004.
2. Le GM, Gomez SL, O'Malley CD, et al. Cancer incidence and mortality in the Greater Bay Area 1998–2002. Fremont (CA): Northern California Cancer Center; 2005.
3. 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.
4. Hanis CL, Hewett-Emmett D, Bertin TK, Schull WJ. Origins of U.S. Hispanics. Implications for diabetes. Diabetes Care 1991;14:618–27.
5. 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.
6. 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.
7. Choudhry S, Coyle NE, Tang H, et al. Population stratification confounds genetic association studies among Latinos. Hum Genet 2006;118:652–64.
8. Burchard EG, 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.
9. Cardon LR, Palmer LJ. Population stratification and spurious allelic association. Lancet 2003;361:598–604.
10. Ziv E, Burchard EG. Human population structure and genetic association studies. Pharmacogenomics 2003;4:431–41.
11. Risch N, Burchard E, Ziv E, Tang H. Categorization of humans in biomedical research: genes, race and disease. Genome Biol 2002;3:comment2007.
12. Thomas DC, Witte JS. Point: population stratification: a problem for case-control studies of candidate-gene associations? Cancer Epidemiol Biomarkers Prev 2002;11:505–12.
13. Wacholder S, Rothman N, Caporaso N. Population stratification in epidemiologic studies of common genetic variants and cancer: quantification of bias. J Natl Cancer Inst 2000;92:1151–8.
14. Wacholder S, Rothman N, Caporaso N. Counterpoint: bias from population stratification is not a major threat to the validity of conclusions from epidemiological studies of common polymorphisms and cancer. Cancer Epidemiol Biomarkers Prev 2002;11:513–20.
15. 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.
16. 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.
17. Wittke-Thompson JK, Pluzhnikov A, Cox NJ. Rational inferences about departures from Hardy-Weinberg equilibrium. Am J Hum Genet 2005;76:967–86.
18. Weir BS, Hill WG, Cardon LR. Allelic association patterns for a dense SNP map. Genet Epidemiol 2004;27:442–50.
19. Chakraborty R, Weiss KM. Admixture as a tool for finding linked genes and detecting that difference from allelic association between loci. Proc Natl Acad Sci U S A 1988;85:9119–23.
20. Chakraborty R, Kamboh MI, Ferrell RE. ''Unique'' alleles in admixed populations: a strategy for determining ''hereditary'' population differences of disease frequencies. Ethn Dis 1991;1:245–56.
21. Chakraborty R, Weiss KM. Frequencies of complex diseases in hybrid populations. Am J Phys Anthropol 1986;70:489–503.
22. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics 2000;155:945–59.
23. Bonilla C, Shriver MD, Parra EJ, Jones A, Fernandez JR. Ancestral proportions and their association with skin pigmentation and bone mineral density in Puerto Rican women from New York city. Hum Genet 2004;115:57–68.
24. Cerda-Flores RM, Kshatriya GK, Bertin TK, Hewett-Emmett D, Hanis CL, Chakraborty R. Gene diversity and estimation of genetic admixture among Mexican-Americans of Starr County, Texas. Ann Hum Biol 1992;19:347–60.
25. Collins-Schramm HE, Chima B, Morii T, et al. Mexican American ancestry-informative markers: examination of population structure and marker characteristics in European Americans, Mexican Americans, Amerindians and Asians. Hum Genet 2004;114:263–71.
26. Long JC, Williams RC, McAuley JE, et al. Genetic variation in Arizona Mexican Americans: estimation and interpretation of admixture proportions. Am J Phys Anthropol 1991;84:141–57.
27. Kittles RA, Chen W, Panguluri RK, et al. CYP3A4-V and prostate cancer in

African Americans: causal or confounding association because of population stratification? Hum Genet 2002;110:553–60.
28. Freedman ML, Reich D, Penney KL, et al. Assessing the impact of population stratification on genetic association studies. Nat Genet 2004;36:388–93.
29. Barnholtz-Sloan JS, Chakraborty R, Sellers TA, Schwartz AG. Examining population stratification via individual ancestry estimates versus self-reported race. Cancer Epidemiol Biomarkers Prev 2005;14:1545–51.
30. Rosenberg NA, Li LM, Ward R, Pritchard JK. Informativeness of genetic markers for inference of ancestry. Am J Hum Genet 2003;73:1402–22.
31. Tsai HJ, Choudhry S, Naqvi M, Rodriguez-Cintron W, Burchard EG, Ziv E. Comparison of three methods to estimate genetic ancestry and control for stratification in genetic association studies among admixed populations. Hum Genet 2005;118:424–33.
32. 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.
33. Pharoah PD, Dunning AM, Ponder BA, Easton DF. The reliable identification of disease-gene associations. Cancer Epidemiol Biomarkers Prev 2005;14:1362.

Genetic Ancestry and Risk Factors for Breast Cancer among Latinas in the San Francisco Bay Area

Elad Ziv, Esther M. John, Shweta Choudhry, et al.

Cancer Epidemiol Biomarkers Prev 2006;15:1878-1885.