

## Research Article

## European Ancestry Is Positively Associated with Breast Cancer Risk in Mexican Women

Laura Fejerman<sup>1,3</sup>, Isabelle Romieu<sup>6,7</sup>, Esther M. John<sup>8,9</sup>, Eduardo Lazcano-Ponce<sup>6</sup>, Scott Huntsman<sup>1,3</sup>, Kenneth B. Beckman<sup>10</sup>, Eliseo J. Pérez-Stable<sup>2,3</sup>, Esteban González Burchard<sup>5</sup>, Elad Ziv<sup>1,3,4</sup>, and Gabriela Torres-Mejía<sup>6</sup>

## Abstract

The incidence of breast cancer is 35% lower in Hispanic women living in the San Francisco Bay Area than in non-Hispanic White women. We have previously described a significant association between genetic ancestry and risk for breast cancer in a sample of U.S. Hispanics/Latinas. We retested the association in women residing in Mexico because of the possibility that the original finding may be confounded by U.S. specific unmeasured environmental exposures. We genotyped a set of 106 ancestry informative markers in 846 Mexican women with breast cancer and 1,035 unaffected controls and estimated genetic ancestry using a maximum likelihood method. Odds ratios and 95% confidence intervals (95% CI) for ancestry modeled as a categorical and continuous variable were estimated using logistic regression and adjusted for reproductive and other known risk factors. Greater European ancestry was associated with increased breast cancer risk in this new and independent sample of Mexican women residing in Mexico. Compared with women with 0% to 25% European ancestry, the risk was increased for women with 51% to 75% and 76% to 100% European ancestry [odds ratios, 1.35 (95% CI, 0.96-1.91) and 2.44 (95% CI, 0.94-6.35), respectively; *P* for trend = 0.044]. For every 25% increase in European ancestry (modeled as a continuous variable), there was a 20% increase in risk for breast cancer (95% CI, 1.03-1.41; *P* = 0.019). These results suggest that nongenetic factors play a crucial role in explaining the difference in breast cancer incidence between Latinas and non-Latina White women, and it also points out to the possibility of a genetic component to this difference. *Cancer Epidemiol Biomarkers Prev*; 19(4); 1074–82. ©2010 AACR.

## Introduction

Breast cancer incidence varies substantially worldwide. Europe, Australia, North America, Argentina, and Uruguay have age adjusted breast cancer incidence rates of up to 101 per 100,000, whereas most African, Asian, and Latin American countries have incidences of <52 per 100,000 (1). This variation might be the result of differences in reproductive, hormonal, lifestyle, and pos-

sibly genetic factors between populations (2-4). Differences in breast cancer incidence have been observed not only between countries but also between populations within countries. For example, Hispanic women in the San Francisco Bay Area had a breast cancer incidence that was 35% lower than that for non-Hispanic White women for the period 1998 to 2002 (5). A large proportion of the difference in incidence disappears in second and third generation immigrants, suggesting a major environmental influence on breast cancer risk (2). The degree to which genetic factors play a role is unknown.

With respect to breast cancer susceptibility, U.S. Latinas are a unique group to study because of the diversity of their environmental and cultural exposures and their genetic ancestry. The term Hispanic or Latino describes a heterogeneous population with a shared history of colonization and a common language but does not refer to a fixed biological entity or a single common ancestry. Latinas, in terms of their genetic background, represent a complex mix of indigenous American, European, and African ancestries (6). Understanding cancer susceptibility in this population is crucial because Latinas account for 14% of the total population of the nation and are predicted to comprise 25% of the population by the year 2050 (7).

Unknown environmental factors might be important in explaining some of the differences in breast cancer risk among subsets of Latina women. Among U.S. Latinas,

**Authors' Affiliations:** <sup>1</sup>Division of General Internal Medicine, Department of Medicine, Institute for Human Genetics; <sup>2</sup>Division of General Internal Medicine, Department of Medicine, Medical Effectiveness Research Center for Diverse Populations; <sup>3</sup>Helen Diller Family Comprehensive Cancer Center; <sup>4</sup>Department of Epidemiology and Biostatistics; <sup>5</sup>Departments of Medicine and Bioengineering and Therapeutic Sciences, Lung Biology Center, Institute for Human Genetics, University of California, San Francisco, California; <sup>6</sup>Instituto Nacional de Salud Pública, Cuernavaca, Morelos, México; <sup>7</sup>Institut National de la Santé et de la Recherche Médicale (INSERM/ERI), ERI 20, EA 4045, Institut Gustave Roussy, Villejuif F-94805, France 20, Villejuif, France; <sup>8</sup>Northern California Cancer Center, Fremont, California; <sup>9</sup>Division of Epidemiology, Department of Health Research and Policy, Stanford Cancer Center, Stanford University School of Medicine, Stanford, California; <sup>10</sup>Children's Hospital Oakland Research Institute, Oakland, California

**Corresponding Author:** Gabriela Torres-Mejía, Instituto Nacional de Salud Pública, Avenida Universidad 655, Colonia Santa María Ahuacatitlán, Cuernavaca, Morelos. Código Postal 62100 México. Phone: 52-777-3112343; Fax: 52-777-3111148. E-mail: gtorres@correo.insp.mx

doi: 10.1158/1055-9965.EPI-09-1193

©2010 American Association for Cancer Research.

those born in the United States have a higher risk for postmenopausal breast cancer compared with foreign-born Latinas, even after controlling for reproductive and other risk factors (2). Genetic factors might also contribute to differences in breast cancer risk among subgroups of Latinas. We recently showed that European genetic ancestry was associated with increased risk for breast cancer among U.S. Latinas (8). We compared the mean genetic ancestry among U.S. Latina breast cancer cases and U.S. Latina controls from the San Francisco Bay Area and found that cases had significantly greater European ancestry and significantly lower indigenous American ancestry, even after adjusting for known reproductive and other environmental risk factors, including place of birth and age at migration.

We sought to retest the association in women residing in Mexico because of the possibility that the original findings between genetic ancestry and breast cancer risk in U.S. Latinas might be confounded by unmeasured environmental exposures. Specifically, we investigated the relationship between breast cancer risk and genetic ancestry using a new and independent sample of 1,881 women in Mexico. We hypothesized that, if there is a United States-specific environmental component that increases breast cancer risk and that is associated with European genetic ancestry in U.S. Latinas, then there should be no association between genetic ancestry and breast cancer risk in the Mexican women.

## Materials and Methods

### Source of cases and controls

**Mexican samples.** Analyses were done using DNA and epidemiological data from a multicenter case-control study designed to examine predictors of breast cancer risk among Mexican women with the age of 35 to 69 y who resided for at least 5 y in Mexico City, Monterrey, or Veracruz.

Newly diagnosed cases were identified at 12 hospitals from the major healthcare systems in Mexico: the Mexican Institute of Social Security (IMSS; six hospitals), the Social Security system for state workers (ISSSTE; two hospitals), and the Ministry of Health (SS; four hospitals), which provides health care to those who do not belong to any of the health care systems stated above. Inclusion criteria for cases were (a) histologically confirmed new diagnosis of breast cancer between the years 2004 and 2007, including invasive and *in situ* tumors; (b) no previous treatment such as radiotherapy, chemotherapy, or antiestrogens in the last 6 mo; and (c) no present treatment with exemestane, letrozole, anastrozole, or megestrol. Pregnant women or cases known to be HIV infected were excluded from the study. Only two cases were using antiestrogens and therefore excluded because we were interested in measuring breast density and antiestrogens could have modified it. Only one case known to be HIV positive was excluded from the study. In this study, cases and controls were derived from the same source populations.

Controls were selected based on a probabilistic multi-stage sampling design. One or more geostatistical areas (from Spanish; Área Geoestadística Básica) considered in the catchment area of the participating hospital were randomly selected. Women were randomly sampled to obtain specific numbers of women in each 5-y age category (35-39, 40-44, 45-49, 50-54, 55-59, 60-64, and 65-69 y) based on the age distribution of cases reported by the Mexican Tumor Registry in 2002. If more than one woman with the age of 35 to 69 y was present in the household, only one was selected (25 controls and 1 case within our sample had a sister of similar age living in the same household; ref. 9). Trained personnel visited the selected households and determined willingness to participate in the study.

Data collection included the administration of a structured questionnaire at the participant's home and collection of anthropometric measurements and a blood sample at the hospital. In addition, a mammogram was taken for controls. The response rate of the cases was 95.5% for Mexico City, 94.4% for Monterrey, and 97.4% for Veracruz. The response rate of the controls was 87.4% for Mexico City, 90.1% for Monterrey, and 97.6% for Veracruz. The complete sample included 1,000 cases and 1,074 controls. DNA for the present study was available for a total of 1,938 subjects (880 cases and 1,058 controls). More detailed information about the description of the multicenter population-based case-control study has been recently published (9).

All participants provided a written informed consent before the in-person interview. The study was approved by the Institutional Review Board at each institution participating in this collaborative study.

**Data collection.** The health questionnaire collected information on sociodemographic characteristics; reproductive factors; use of oral contraceptives and hormone replacement therapy; family and personal history of chronic diseases; personal history of transmitted sexual diseases; histories of body size, smoking, and alcohol consumption; and history of medical X-rays and mammograms. Information on usual dietary intake in the past year was collected using a food frequency questionnaire adapted from the Willet questionnaire (10) and validated in a sample of Mexican women (11).

Study participants were asked about usual physical activity during a 1-wk period that reflected the activity done during the last 12 mo before any breast cancer symptoms were perceived in cases and during the last 12 mo before recruitment in the controls (12). Three different categories of physical activity were defined: strenuous, moderate, and light (13).

Standing height, weight, and hip and waist circumferences were measured by the interviewer at the hospital. Body mass index (BMI) was calculated as measured weight (kilogram) divided by measured height (meter) squared.

### Marker selection and ancestral populations

A set of 106 single nucleotide polymorphisms (SNPs) that can separate indigenous American, African, and

European ancestry was used to estimate the proportion of genetic ancestry in the sample of Mexican women with breast cancer and unaffected controls. Simulation studies have shown that ~100 ancestry informative markers with allele frequency differences similar to the ones we used are required to achieve an  $r$  of >0.9 with true ancestry (14). The ancestry informative markers used in this study were biallelic SNPs selected from the Affymetrix 100K SNP chip (Affymetrix). Ancestry informative marker 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 ancestry informative markers are widely spaced throughout the genome and have a well-balanced distribution across all 22 autosomal chromosomes. The average distance between markers is about  $2.4 \times 10^7$  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; refs. 8, 15).

### Genotyping

Genotyping of the 106 ancestry informative markers for the Mexican women was done at the Children's Hospital Oakland Research Institute. Genotyping was done using a multiplex PCR coupled with single base extension methodology with allele calls using a Sequenom analyzer. Primers and reaction conditions have been previously published (8). Samples were genotyped without knowledge about case/control status by the laboratory personnel.

A total of 1,938 Mexican cases and controls were newly genotyped for this study (880 cases and 1,058 controls). The average call rate for the SNPs was 98.9%. After we removed three SNPs with a call rate <90%, the average call rate for the SNPs was 99.2%. The average sample call rate was 98.7%. After we removed 57 samples (34 cases, 23 controls) with a call rate <85%, the average call rate for the samples was 99.6% (the excluded samples had no significant differences for the variables analyzed). We genotyped 66 duplicate pairs, and of these, three pairs were excluded from the mismatch analysis because the call rate for one of the duplicates was low (7%, 26%, and 25%) compared with the high call rate of most samples in the study. The overall error rate without including the duplicate pairs with low call rates was 0.02%. All the ancestry informative markers were in Hardy-Weinberg equilibrium.

Genotypes and phenotype information was available for a total of 1,881 women from Mexico (846 cases and 1,035 controls).

### Statistical analysis

We used a maximum likelihood approach to estimate genetic ancestry at the individual level (16, 17). To compare the characteristics of cases and controls for Mexican

women, we used  $t$  tests for continuous variables and Fisher's exact tests for categorical variables. Mean genetic ancestry was estimated as the average of the individual genetic ancestry estimates within a group.

To assess the association between breast cancer risk and genetic ancestry in the sample of Mexican women residing in Mexico, we used unadjusted and adjusted logistic regression models. European or indigenous American genetic ancestry were modeled as categorical variables (0-25%, 1; 26-50%, 2; 51-75%, 3; 76-100%, 4). We also evaluated the models with genetic ancestry represented as a continuous variable (percent genetic ancestry). The unadjusted and adjusted models included as covariates the recruitment site (Mexico City, Veracruz, or Monterrey) and the type of health insurance to which the individuals belong (SSA, ISSSTE, and IMSS). We stratified the unadjusted analysis by recruitment site to evaluate if the estimation between genetic ancestry and breast cancer risk was consistent across the three sites. The multivariate models adjusted for European ancestry, age (continuous), family history of breast cancer in first-degree relatives (yes, no), personal history of benign breast disease (yes, no), age at menarche (continuous), number of full-term pregnancies (continuous 0-6,  $\geq 7$ ), age at first full-term pregnancy (1,  $\leq 20$  y; 2, between 21 and 30 y; 3, >30 y; 4, no full-term pregnancies), breastfeeding (ever, never), history of hormone replacement therapy use/menopausal status (0, postmenopausal/ever use of hormone replacement therapy; 1, premenopausal/ever use of hormone replacement therapy; 2, postmenopausal/no hormone replacement therapy; 3, premenopausal/no hormone replacement therapy), alcohol intake during the reference year (defined as 12 mo previous to diagnosis in cases and 12 mo previous to recruitment in controls; one or more drinks a month in a year or longer, 1; otherwise, 0), daily caloric intake during the reference year (continuous), education (none, 0; some elementary school, 1; completed elementary school, 2; high school, 3; college, 4; postgraduate studies, 5), moderate physical activity (hours per week; continuous), and socioeconomic status (low, medium, high). For the construction of the socioeconomic status variable, we combined information about different belongings (that is, gas or electric stove, water heating system, radio or cassette recorder, television, videocassette recorder, CD player, refrigerator, washing machine, microwave oven, blender, vacuum cleaner, water pump, motorcycle, car or van, fixed phone, cellular phone, computer, and dish antenna). The polychoric correlation was used for the construction of a socioeconomic status index (18) categorized into tertiles among controls (low, medium, high).

Individuals with missing data were dropped from the multivariate analysis (161 cases and 180 controls). We evaluated models including European and African ancestry. We present results based on a model that included European ancestry as the predictor. Native American ancestry is the counterpart of European ancestry, and therefore, the results can also be interpreted in terms of the former.

We evaluated possible interactions between genetic ancestry and other risk factors (e.g., hormone replacement therapy, BMI, parity, age at first full-term pregnancy, breast-feeding, and menopausal status).

All statistical analyses were done using the program STATA (19), and all tests are two sided.

## Results

Characteristics of breast cancer cases and controls from Mexico are presented in Table 1. Mexican cases were significantly older at diagnosis, had fewer full-term pregnancies, were less likely to breast-feed, were more likely to report a personal history of benign breast disease, a family history of breast cancer, or history of hormone replacement therapy use, and had higher alcohol intake and higher daily caloric intake. Cases also reported a significantly higher level of education and socioeconomic status and lower BMI than controls (the

relationship between BMI and breast cancer risk was only observed among premenopausal women). Finally, women with breast cancer had more European and less indigenous American ancestry than controls. There were no significant differences between cases and controls in age at menarche or African ancestry. The proportion of European and indigenous American ancestry in Mexican controls differed by recruitment site. Monterrey had the largest proportion of European ancestry (40%; SD = 16) compared with Mexico City (28%; SD = 19) and Veracruz (30%; SD = 18). The proportion of indigenous American ancestry was estimated to be 54% (SD = 15) in Monterrey, 69% (SD = 20) in Mexico City, and 64% (SD = 20) in Veracruz. The proportion of African ancestry was estimated to be 6% (SD = 5) in Monterrey, 3% (SD = 4) in Mexico City, and 6% (SD = 8) in Veracruz.

We investigated the association between the different reproductive, demographic, lifestyle characteristics, and genetic ancestry among the controls (Table 2) and

**Table 1. Characteristics of Mexican women by case-control status, 2004 to 2007**

	Cases (n = 846)	Controls (n = 1035)	P*
Age (y) <sup>†</sup>	52.0 (10)	51.1 (9)	0.02
BMI <sup>†</sup> premenopausal women	28.2 (5)	30.1 (5)	<0.01
BMI <sup>†</sup> postmenopausal women	30.2 (6)	30.8 (5)	0.06
Age at menarche (y) <sup>†</sup>	12.8 (2)	12.8 (2)	0.21
No. of full-term pregnancies <sup>†</sup>	3.0 (2)	3.7 (2)	<0.01
Ever breast-fed (%) <sup>†</sup>	86.0	89.5	0.03
Family history of breast cancer (%)	6.6	3.8	0.004
Benign breast disease (%)	14.7	7.51	<0.01
Postmenopausal HRT use (%) <sup>§</sup>	21.9	14.5	<0.01
Postmenopausal (%)	57.6	55.5	0.37
Alcohol consumption (ever in past year/never; %)	19.3	10.9	<0.01
Socioeconomic status (%)			<0.01
Low	25.9	32.6	
Middle	24.4	34.0	
High	49.7	33.4	
Education (%)			<0.01
None	6.6	8.5	
Some elementary school	19.5	26.3	
Completed elementary school	30.4	29.6	
High school graduate	25.8	25.5	
College graduate	8.6	5.9	
Postgraduate studies	9.0	4.2	
Caloric intake (kcal/d) <sup>†</sup>	2,225.4 (789)	1,935.4 (673)	<0.01
Percent European ancestry <sup>†</sup>	35.9 (19)	31.8 (18)	<0.01
Percent indigenous American ancestry <sup>†</sup>	59.5 (20)	64.1 (20)	<0.01
Percent African ancestry <sup>†</sup>	4.5 (6)	4.1 (5)	0.11

Abbreviation: HRT, hormone replacement therapy.

\*P values are for bivariate Fisher's exact tests in 2 × 2 tables (for categorical variables) and for t tests comparing mean values between cases and controls (for continuous variables).

<sup>†</sup>Mean (SD).

<sup>‡</sup>Women who had at least one full-term pregnancy.

<sup>§</sup>Proportion of menopausal women that use hormone replacement therapy.

**Table 2.** Characteristics of Mexican controls by European genetic ancestry, 2004 to 2007

	European genetic ancestry				P*
	<25	26-50	51-75	76-100	
Age (y) <sup>†</sup>	51.9 (9)	51.3 (9)	50.9 (9)	51.3 (10)	0.954
BMI premenopausal women <sup>†</sup>	29.7 (5)	30.3 (5)	30.7 (6)	30.8 (5)	0.592
BMI postmenopausal women <sup>†</sup>	30.4 (5)	31.3 (5)	30.2 (5)	34.7 (8)	0.070
Age at menarche (y) <sup>†</sup>	12.9 (2)	12.8 (2)	12.9 (2)	13.2 (2)	0.332
No. of full-term pregnancies <sup>†</sup>	3.6 (2)	3.9 (3)	3.4 (2)	4.1 (3)	0.073
Ever breast-fed (%) <sup>‡</sup>	90.7	89.7	85.0	100.0	0.278
Family history of breast cancer (%)	2.4	3.5	7.4	10.0	0.034
Benign breast disease (%)	6.8	8.0	7.4	10.0	0.768
Postmenopausal HRT use (%) <sup>§</sup>	14.7	12.7	21.1	16.7	0.266
Postmenopausal (%)	54.7	56.5	53.3	60.0	0.891
Alcohol consumption (ever in past year/never; %)	11.7	10.1	12.2	11.1	0.761
Socioeconomic status (%)					<0.001
Low	42.3	28.0	25.0	0.0	
Middle	31.5	36.9	29.7	33.3	
High	26.2	34.9	45.3	66.7	
Education (%)					<0.001
None	11.8	7.3	4.4	0.0	
Some elementary school	32.4	24.6	17.8	0.0	
Completed elementary school	24.9	29.6	40.7	50.0	
High school graduate	23.3	27.8	22.2	30.0	
College graduate	5.1	5.8	8.9	0.0	
Postgraduate studies	2.4	4.8	5.9	20.0	
Caloric intake (kcal/d) <sup>†</sup>	1819.4 (584)	1955.9 (682)	2132.1 (761)	2413.4 (1070)	<0.001

\*P values are for bivariate Fisher's exact tests in 2 × 2 tables (for categorical variables) and for ANOVA tests (for continuous variables).

<sup>†</sup>Mean (SD).

<sup>‡</sup>Women who had at least one full-term pregnancy.

<sup>§</sup>Proportion of menopausal women that use hormone replacement therapy.

observed that socioeconomic status, education, daily kilocalorie intake, and family history of breast cancer significantly differ by ancestry category, with family history being more common among women with higher European ancestry and socioeconomic status, education, and daily kilocalorie intake being higher in women with higher European ancestry. We also explored the relationship between the different characteristics of the controls and socioeconomic status (Table 3). These results show a very strong relationship between socioeconomic status and genetic ancestry, all reproductive variables, education, alcohol consumption, and daily kilocalorie intake. Women with higher socioeconomic status tend to have less full-term pregnancies, higher European ancestry, breast-feed less, consume more alcohol and kilocalories, and have more years of education than women with lower socioeconomic status.

The association between genetic ancestry and breast cancer risk in the sample of Mexican women is shown in Table 4. The unadjusted model showed a strong association with European genetic ancestry. Using 0% to 25%

European ancestry as the reference, odds ratios were 1.16 [95% confidence interval (95% CI), 0.94-1.43;  $P = 0.171$ ] for the 26% to 50% European ancestry, 1.80 (95% CI, 1.35-2.39;  $P < 0.001$ ) for the 51% to 75% category, and 2.22 (95% CI, 1.46-7.11;  $P = 0.004$ ) for the 76% to 100% category. When known risk factors were adjusted for (Table 4), the association with European ancestry was attenuated but the trend remained statistically significant (26-50% odds ratio, 1.01; CI, 0.78-1.30;  $P = 0.962$ ; 51-75% odds ratio, 1.35; 95% CI, 0.96-1.91;  $P = 0.087$ ; 76-100% odds ratio, 2.44; 95% CI, 0.94-6.35;  $P = 0.067$ ;  $P$  for trend = 0.044). We chose to represent ancestry as a categorical variable because it allowed us to compare the effect size for the extremes of the ancestry distribution. However, the effect is also seen when ancestry is entered as a continuous variable in the model. For the model that included genetic ancestry as a continuous variable, the odds ratio for every 25% increase in European ancestry was 1.20 (95% CI, 1.03-1.41;  $P = 0.019$ ). To ensure that there was no confounding due to regional differences between cases and controls, the unadjusted model was stratified by

**Table 3.** Characteristics of Mexican controls by socioeconomic status, 2004 to 2007

	Socioeconomic status			P*
	Low	Medium	High	
European ancestry (%)				<0.001
0-25	47.4	33.8	28.7	
26-50	42.7	54.0	51.9	
51-75	9.9	11.3	17.5	
76-100	0.0	0.9	1.8	
Age (y) <sup>†</sup>	51.1 (10)	51.5 (9)	50.5 (8)	0.391
BMI <sup>†</sup> premenopausal women	30.5 (6)	30.4 (5)	29.3 (5)	0.111
BMI <sup>†</sup> postmenopausal women	30.6 (5)	31.0 (5)	30.5 (5)	0.616
Age at menarche (y) <sup>†</sup>	12.9 (2)	12.9 (2)	12.7 (2)	0.040
No. of full-term pregnancies <sup>†</sup>	3.5 (2)	3.2 (2)	2.6 (2)	<0.001
Ever breast-fed (%) <sup>‡</sup>	88.5	93.2	85.3	0.005
Family history of breast cancer (%)	2.2	4.7	4.5	0.142
Benign breast disease (%)	4.4	7.8	10.1	0.020
Postmenopausal HRT use (%) <sup>§</sup>	9.9	15.3	15.4	0.222
Postmenopausal (%)	53.6	60.2	51.1	0.047
Alcohol consumption (ever in past year/never; %)	7.5	9.8	13.8	0.03
Education (%)				<0.001
None	12.1	10.7	3.6	
Some elementary school	36.2	28.8	13.9	
Completed elementary school	28.8	30.9	28.4	
High school graduate	20.7	23.7	32.3	
College graduate	2.2	3.9	11.2	
Postgraduate studies	0.0	2.1	10.6	
Caloric intake (kcal/d) <sup>†</sup>	1773.9 (558)	1909.6 (630)	2034.0 (711)	<0.001

\*P values are for bivariate Fisher's exact tests in 2 × 2 tables (for categorical variables) and for ANOVA tests (for continuous variables).

<sup>†</sup>Mean (SD).

<sup>‡</sup>Women who had at least one full-term pregnancy.

<sup>§</sup>Proportion of menopausal women that use hormone replacement therapy.

recruitment site (Mexico City, Monterrey, Veracruz), with all results showing the same trend as the global analysis (Table 4), with the effect weakening as sample size decreases.

In the adjusted model, the associations between breast cancer and alcohol consumption, parity, family history, age at menarche, benign breast disease, kilocalorie intake, and moderate physical activity were in the expected direction (Table 4). When the model testing the association between genetic ancestry and breast cancer risk included socioeconomic status and education as the only covariates, the two covariates showed an effect on breast cancer risk. However, when other covariates were added to the model, the effect of socioeconomic status and education became nonsignificant (Table 4). Number of full-term pregnancies, daily kilocalorie intake, and benign breast disease were the variables that, when added to the model, absorbed most of the effect of socioeconomic status and education.

We found no evidence of significant interaction between genetic ancestry and hormone replacement therapy

use, BMI, parity, age at first full-term pregnancy, breast-feeding, and menopausal status.

## Discussion

We found an association between European genetic ancestry and breast cancer risk among Mexican women that reside in Mexico. Women with greater European ancestry had an increased risk for breast cancer compared with women with lower European ancestry, which could also be interpreted as indigenous American ancestry being protective against the development of breast cancer. We have previously reported a significant association between genetic ancestry and breast cancer risk in U.S. Latinas, with a 39% increase in breast cancer risk with every 25% increase in European ancestry (8). If the observed association in the U.S. Latinas were solely due to confounding by a specific environmental risk factor associated with genetic ancestry in the United States, then we would expect not to observe an association in the Mexican

**Table 4.** Association between genetic ancestry and breast cancer risk in women residing in Mexico, 2004 to 2007

	Odds ratio	95% CI	P
Unadjusted model (n = 1,881; 846 cases and 1,035 controls)			
European ancestry (%)*			<0.001 Trend
26-50	1.16	0.94-1.43	0.171
51-75	1.80	1.35-2.39	<0.001
76-100	3.22	1.46-7.11	0.004
Continuous†	1.37	1.21-1.55	<0.001
Continuous† Mexico city (n = 1,128)	1.45	1.24-1.69	<0.001
Continuous† Monterrey (n = 452)	1.27	0.96-1.70	0.096
Continuous† Veracruz (n = 301)	1.16	0.82-1.62	0.395
Unadjusted model for individuals with no missing data (n = 1540; 685 cases and 855 controls)			
European ancestry (%)*			<0.001 Trend
26-50	1.10	0.87-1.40	0.409
51-75	1.79	1.35-2.39	<0.001
76-100	3.34	1.39-8.03	0.007
Continuous†	1.36	1.19-1.57	<0.001
Adjusted by socioeconomic status (n = 1,787; 796 cases and 991 controls)			
European ancestry (%)*			<0.001 Trend
26-50	1.13	0.91-1.41	0.275
51-75	1.76	1.31-2.36	<0.001
76-100	2.89	1.25-6.67	0.013
Continuous†	1.34	1.17-1.53	<0.001
Socioeconomic status	1.29	1.13-1.46	<0.001
Adjusted by socioeconomic status and education (n = 1,786; 795 cases and 991 controls)			
European ancestry (%)*			<0.001 Trend
26-50	1.09	0.87-1.36	0.438
51-75	1.67	1.24-2.26	0.001
76-100	2.68	1.16-6.20	0.021
Continuous†	1.30	1.13-1.48	<0.001
Socioeconomic status	1.18	1.05-1.37	0.008
Education	1.16	1.07-1.26	<0.001
Fully adjusted model (n = 1,540; 685 cases and 855 controls)			
European ancestry (%)*			0.044 Trend
26-50	1.01	0.78-1.30	0.962
51-75	1.35	0.96-1.91	0.087
76-100	2.44	0.94-6.35	0.067
Continuous†	1.20	1.03-1.41	0.019
Age at diagnosis (y)	1.04	1.02-1.07	<0.001
BMI (kg/m <sup>2</sup> )	0.95	0.93-0.97	<0.001
No. of full-term pregnancies (0 to ≥7)	0.85	0.78-0.92	<0.001
Breast-feeding (ever/never)	0.96	0.65-1.42	0.853
Benign breast disease	2.11	1.45-3.08	<0.001
Family history of breast cancer	1.70	1.03-2.79	0.036
Moderate physical activity (h/wk)	0.97	0.97-0.98	<0.001
Education	1.05	0.93-1.18	0.438
Socioeconomic status	1.01	0.86-1.18	0.919
Alcohol consumption (ever in past year)	1.78	1.29-2.45	0.001
Daily caloric intake (per 500 kcal unit)	1.40	1.28-1.52	<0.001
HRT/menopause‡			
Premenopausal/HRT	1.02	0.41-2.53	0.964
Postmenopausal/no HRT	0.80	0.54-1.20	0.281
Premenopausal/no HRT	0.93	0.58-1.47	0.740

(Continued on the following page)

**Table 4.** Association between genetic ancestry and breast cancer risk in women residing in Mexico, 2004 to 2007 (Cont'd)

	Odds ratio	95% CI	P
Age at first full-term pregnancy (y) <sup>§</sup>			
21-30	0.98	0.76-1.28	0.908
≥31	1.56	0.91-2.66	0.103
No full-term pregnancy	1.00	0.53-1.91	0.991
Age at menarche (y)	0.93	0.87-1.00	0.043

\*European ancestry between 0% and 25% is the reference.

†Odds ratio is for every 25% increase in European ancestry.

‡Postmenopausal and hormone replacement therapy is the reference.

§First full-term pregnancy before the age of 20 years is the reference.

women. However, given that the association between breast cancer risk and genetic ancestry was also present in Latina women residing outside the United States, it is reasonable to conclude that, if there is an environmental component that is driving the association between genetic ancestry and breast cancer risk, it must be common to U.S. Latinas and Mexican women. The association is weaker in the Mexicans than in the U.S. Latinas, which could be due to differences in the available covariates. It may also be due to the possibility that unmeasured confounding was greater in the U.S. samples.

Among Mexican women, the association between genetic ancestry and breast cancer risk was attenuated after adjustment for known risk factors. Number of full-term pregnancies, daily caloric intake, benign breast disease, and socioeconomic status/education were the main variables responsible for this attenuation. It is interesting to notice that socioeconomic status/education, which were associated with genetic ancestry and with breast cancer risk in the univariate models, did not have a significant effect on breast cancer risk once other variables, such as number of full term pregnancies and kilocalorie intake, were included in the model.

The results of this study confirmed our previous finding in an independent sample of Mexican women living outside the United States. The effect of ancestry on breast cancer risk could be due to the effect of unmeasured environmental confounders present in Mexico and the United States. However, our results also provide support to the hypothesis of a genetic component underlying the association between genetic ancestry and breast cancer risk.

A limitation of the present study is the possibility that our measure of socioeconomic status might not be optimal. Our models included adjustment for socioeconomic status and education. However, our data included only three categories for socioeconomic status, and this may not capture the true variation in the population. The correlation between socioeconomic status and genetic ancestry in the Mexican sample is high ( $r = 0.23$ ). If the true socioeconomic status were more strongly correlated with

genetic ancestry, then it would be difficult to separate the effect of the two.

Another limitation of our analysis is that we treated each ancestral population, Europeans, indigenous Americans, and Africans, as one population despite the fact that numerous populations contributed to the gene pool of contemporary Latinos, and, in particular, there were multiple indigenous American populations who were genetically (20) and culturally diverse. Therefore, our results in Mexican women may not be generalizable to populations from different parts of the Americas.

In summary, we replicated our previous results, although with weaker effects, showing an association between European genetic ancestry and breast cancer risk in an independent sample of Latina women living in three different regions of Mexico. Future studies in other populations in the Americas will allow us to further explore the effect of genetic ancestry on breast cancer risk in diverse environments, populations with different breast cancer incidence, and in groups with different proportions of European, indigenous American, and African ancestry.

#### Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

#### Acknowledgments

We thank the study participants and the physicians responsible for the project in the different participating hospitals in Mexico: Dr. Germán Castelazo (IMSS, Hospital de la Raza, Ciudad de México, Distrito Federal (DF)), Dr. Sinhué Barroso Bravo (IMSS, Hospital Siglo XXI, Ciudad de México, DF), Dr. Fernando Mainero Ratchelous (IMSS, Hospital de Gineco-obstetricia No 4. "Luis Castelazo Ayala," Ciudad de México, DF), Dr. Hernando Miranda Hernández (SS, Hospital General de México, Ciudad de México, DF), Dr. Joaquín Zarco Méndez (ISSSTE, Hospital 20 de Noviembre, Ciudad de México, DF), Dr. Edelmiro Pérez Rodríguez (Hospital Universitario, Monterrey, Nuevo León), Dr. Jesús Pablo Esparza Cano (IMSS, Hospital No. 23 de Ginecología, Monterrey, Nuevo León), Dr. Heriberto Fabela (IMSS, Hospital No. 23 de Ginecología, Monterrey, Nuevo León), Dr. José Pulido Rodríguez (SS, Hospital Metropolitano Dr "Bernardo Sepulveda," Monterrey, Nuevo León), Dr. Manuel de Jesús García Solís (SS, Hospital Metropolitano Dr "Bernardo Sepulveda," Monterrey, Nuevo León), Dr. Fausto Hernández Morales (ISSSTE, Hospital General, Veracruz,



Veracruz), Dr. Pedro Coronel Brizio (SS, Centro Estatal de Cancerología "Dr. Miguel Dorantes Mesa," Xalapa, Veracruz), Dr. Vicente A. Saldaña Quiroz (IMSS, Hospital Gineco-Pediatría No 71, Veracruz, Veracruz), and Teresa Shama Levi (Instituto Nacional de Salud Pública, Cuernavaca, Morelos).

### Grant Support

National Cancer Institute (R01CA120120; E. Ziv); Consejo Nacional de Ciencia y Tecnología (CONACyT) (SALUD-2002-C01-7462; G. Torres-Mejía); University of California, San Francisco, Clinical and Translational Sciences Institute (Career Development Award) and Prevent Cancer Foundation (postdoctoral fellowship; L. Fejerman); NIH (HL078885 and

HL088133); Robert Wood Johnson Foundation (RWJF) Amos Medical Faculty Development Award, and The Sandler Family Supporting Foundation (E. Gonzalez Burchard); National Cancer Institute (Redes En Acción, U01-CA86117; E.J. Pérez-Stable); and National Cancer Institute (R01 CA63446 and R01 CA77305) and California Breast Cancer Research Program (7PB-0068; E.M. John).

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.

Received 11/23/2009; revised 02/01/2010; accepted 02/08/2010; published OnlineFirst 03/23/2010.

### References

- Curado MP, Edwards B, Shin HR, Ferlay J, Heanue M, Boyle P, Storm H. Cancer incidence in five continents, volume IX. Lyon: IARC Scientific Publication; 2008.
- 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.
- Kamangar F, Dores GM, Anderson WF. Patterns of cancer incidence, mortality, and prevalence across five continents: defining priorities to reduce cancer disparities in different geographic regions of the world. *J Clin Oncol* 2006;24:2137–50.
- Murtaugh MA, Sweeney C, Giuliano AR, et al. Diet patterns and breast cancer risk in Hispanic and non-Hispanic white women: the Four-Corners Breast Cancer Study. *Am J Clin Nutr* 2008;87:978–84.
- Le GM, Gomez SL, O'Malley CD. Cancer incidence and mortality in the Greater Bay Area 1998–2002. Fremont: Northern California Cancer Center; 2005.
- 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.
- Cardon LR, Palmer LJ. Population stratification and spurious allelic association. *Lancet* 2003;361:598–604.
- Fejerman L, John EM, Huntsman S, et al. Genetic ancestry and risk of breast cancer among U.S. Latinas. *Cancer Res* 2008;68:9723–8.
- Angeles-Llerenas A, Ortega-Olvera C, Perez-Rodriguez E, et al. Moderate physical activity and breast cancer risk: the effect of menopausal status. *Cancer Causes Control*.
- Willett W. Nutritional epidemiology. 2nd ed. New York: Oxford University Press; 1998.
- Hernandez-Avila M, Romieu I, Parra S, Hernandez-Avila J, Madrigal H, Willett W. Validity and reproducibility of a food frequency questionnaire to assess dietary intake of women living in Mexico City. *Salud Publica Mex* 1998;40:133–40.
- Ainsworth BE, Haskell WL, Whitt MC, et al. Compendium of physical activities: an update of activity codes and MET intensities. *Med Sci Sports Exerc* 2000;32:S498–504.
- Ortiz-Rodriguez SP, Torres-Mejia G, Mainero-Ratchelous F, et al. Physical activity and breast cancer risk in Mexican women. *Salud Publica Mex* 2008;50:126–35.
- 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.
- Choudhry S, Taub M, Mei R, et al. Genome-wide screen for asthma in Puerto Ricans: evidence for association with 5q23 region. *Hum Genet* 2008;123:455–68.
- Chakraborty R, Weiss KM. Frequencies of complex diseases in hybrid populations. *Am J Phys Anthropol* 1986;70:489–503.
- 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.
- Kolenikov S, Angeles G. The use of discrete data in PCA: theory, simulations, and applications to the socioeconomic indices. Chapel Hill: Chapel Hill and University of North Carolina; 2004.
- StataCorp. Stata statistical software. 8.2 ed. College Station, TX: StataCorp LP; 2003.
- Wang S, Lewis CM, Jakobsson M, et al. Genetic variation and population structure in Native Americans. *PLoS Genet* 2007;3:e185.

# Cancer Epidemiology, Biomarkers & Prevention

**AACR** American Association  
for Cancer Research

## European Ancestry Is Positively Associated with Breast Cancer Risk in Mexican Women

Laura Fejerman, Isabelle Romieu, Esther M. John, et al.

*Cancer Epidemiol Biomarkers Prev* 2010;19:1074-1082. Published OnlineFirst March 23, 2010.

**Updated version** Access the most recent version of this article at:  
doi:[10.1158/1055-9965.EPI-09-1193](https://doi.org/10.1158/1055-9965.EPI-09-1193)

**Cited articles** This article cites 14 articles, 4 of which you can access for free at:  
<http://cebp.aacrjournals.org/content/19/4/1074.full#ref-list-1>

**Citing articles** This article has been cited by 19 HighWire-hosted articles. Access the articles at:  
<http://cebp.aacrjournals.org/content/19/4/1074.full#related-urls>

**E-mail alerts** [Sign up to receive free email-alerts](#) related to this article or journal.

**Reprints and Subscriptions** To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at [pubs@aacr.org](mailto:pubs@aacr.org).

**Permissions** To request permission to re-use all or part of this article, use this link  
<http://cebp.aacrjournals.org/content/19/4/1074>.  
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.