



# HHS Public Access

Author manuscript

*Cancer Causes Control*. Author manuscript; available in PMC 2015 November 01.

Published in final edited form as:

*Cancer Causes Control*. 2014 November ; 25(11): 1461–1471. doi:10.1007/s10552-014-0448-5.

## Associations between *CYP19A1* polymorphisms, Native American ancestry, and breast cancer risk and mortality: the Breast Cancer Health Disparities Study

**Stephanie D. Boone,**

Department of Epidemiology and Population Health, James Graham Brown Cancer Center, School of Public Health and Information Sciences, University of Louisville, 485 E. Gray St., Louisville, KY 40202, USA

**Kathy B. Baumgartner,**

Department of Epidemiology and Population Health, James Graham Brown Cancer Center, School of Public Health and Information Sciences, University of Louisville, 485 E. Gray St., Louisville, KY 40202, USA

**Richard N. Baumgartner,**

Department of Epidemiology and Population Health, James Graham Brown Cancer Center, School of Public Health and Information Sciences, University of Louisville, 485 E. Gray St., Louisville, KY 40202, USA

**Avonne E. Connor,**

Department of Epidemiology and Population Health, James Graham Brown Cancer Center, School of Public Health and Information Sciences, University of Louisville, 485 E. Gray St., Louisville, KY 40202, USA

**Christina M. Pinkston,**

Department of Bioinformatics and Biostatistics, James Graham Brown Cancer Center, School of Public Health and Information Sciences, University of Louisville, Louisville, KY, USA

**Shesh N. Rai,**

Department of Bioinformatics and Biostatistics, James Graham Brown Cancer Center, School of Public Health and Information Sciences, University of Louisville, Louisville, KY, USA

**Elizabeth C. Riley,**

Department of Medicine-Oncology and Hematology, James Graham Brown Cancer Center, School of Medicine, University of Louisville, Louisville, KY, USA

**Lisa M. Hines,**

Department of Biology, University of Colorado, Colorado Springs, Colorado Springs, CO, USA

**Anna R. Giuliano,**

H. Lee Moffit Cancer Center and Research Institute, Tampa, FL, USA

---

© Springer International Publishing Switzerland 2014

Correspondence to: Stephanie D. Boone, [stephanie.boone@louisville.edu](mailto:stephanie.boone@louisville.edu).

Electronic supplementary material The online version of this article (doi: 10.1007/s10552-014-0448-5) contains supplementary material, which is available to authorized users.

**Esther M. John,**

Cancer Prevention Institute of California, Fremont, CA, USA. Division of Epidemiology, Department of Health Research and Policy, Stanford Cancer Institute, Stanford University School of Medicine, Stanford, CA, USA

**Mariana C. Stern,**

Department of Preventive Medicine, Norris Comprehensive Cancer Center, Keck School of Medicine of USC, University of Southern California, Los Angeles, CA, USA

**Gabriela Torres-Mejía,**

Centro de Investigación Salud Poblacional, Instituto Nacional de Salud Pública, Cuernavaca, Morelos, Mexico

**Roger K. Wolff, and**

Department of Internal Medicine, University of Utah, Salt Lake City, UT, USA

**Martha L. Slattery**

Department of Internal Medicine, University of Utah, Salt Lake City, UT, USA

Stephanie D. Boone: stephanie.boone@louisville.edu

**Abstract**

The cytochrome p450 family 19 gene (*CYP19A1*) encodes for aromatase, which catalyzes the final step in estrogen biosynthesis and conversion of androgens to estrogens. Genetic variation in *CYP19A1* is linked to higher circulating estrogen levels and increased aromatase expression. Using data from the Breast Cancer Health Disparities Study, a consortium of three population-based case-control studies in the United States ( $n = 3,030$  non-Hispanic Whites;  $n = 2,893$  Hispanic/Native Americans (H/NA) and Mexico ( $n = 1,810$ ), we examined influence of 25 *CYP19A1* tagging single-nucleotide polymorphisms (SNPs) on breast cancer risk and mortality, considering NA ancestry. Odds ratios (ORs) and 95 % confidence intervals (CIs) and hazard ratios estimated breast cancer risk and mortality. After multiple comparison adjustment, none of the SNPs were significantly associated with breast cancer risk or mortality. Two SNPs remained significantly associated with increased breast cancer risk in women of moderate to high NA ancestry (29 %): rs700518,  $OR_{GG} 1.36$ , 95 % CI 1.11–1.67 and rs11856927,  $OR_{GG} 1.35$ , 95 % CI 1.05–1.72. A significant interaction was observed for rs2470144 and menopausal status ( $p_{adj} = 0.03$ ); risk was increased in postmenopausal ( $OR_{AA} 1.22$ , 95 % CI 1.05–1.14), but not premenopausal ( $OR_{AA} 0.78$ , 95 % CI 0.64–0.95) women. The absence of an overall association with *CYP19A1* and breast cancer risk is similar to previous literature. However, this analysis provides support that variation in *CYP19A1* may influence breast cancer risk differently in women with moderate to high NA ancestry. Additional research is warranted to investigate the how variation in an estrogen-regulating gene contributes to racial/ethnic disparities in breast cancer.

**Keywords**

*CYP19A1* polymorphisms; Native American ancestry; Breast cancer risk; Breast cancer mortality

## Introduction

The cytochrome p450 family 19 gene (*CYP19A1*), located on chromosome 15, encodes for the enzyme aromatase, which catalyzes the final step in estrogen biosynthesis and metabolism converting androgens to estrogens, in the ovaries of premenopausal women and primarily in the adipose tissue of postmenopausal women [1, 2]. Genetic variation in the aromatase gene makes an attractive candidate to evaluate as a breast cancer risk factor due to this gene's involvement in the regulation of endogenous estrogen and to the recognized association between estrogen exposure levels in blood and urine and breast cancer risk [3–5]. During breast carcinogenesis, estrogen interacts with the cell cycle promoting cellular division and proliferation [6]. Estrogen may also act locally in surrounding adipose tissue to promote the growth of breast cancer [7]. Additionally, while clinical data are in agreement with an anti-invasive effect of estrogens, disruptions in the estrogen receptor-signaling pathway could lead to metastasis [8, 9].

Single-nucleotide polymorphisms (SNPs) in *CYP19A1* have been shown to be associated with circulating estrogen levels; specifically rs727479 (T), rs10046 (T), and rs4646 (G), while rs700518 (G/A at Val<sup>80</sup>, located in exon 3) was found to be associated with increased aromatase expression [10–14]. Despite this relationship, epidemiological studies have shown mixed results for an association between SNPs in *CYP19A1* and breast cancer risk [10, 11, 13, 15–20]. Likewise, epidemiological studies to date show inconsistent associations between SNPs in *CYP19A1* and breast cancer mortality [21–26]. Moreover, none of these studies have included Hispanic women; a genetically admixed population with European, Native American (NA), and African ancestry [27, 28]. In the USA and Mexico, Hispanic women with a higher proportion of European ancestry were found to have a higher risk of breast cancer [28, 29]. Additionally, previous findings suggest that the proportion of NA ancestry modifies the direction and strength of associations between reported genetic variants involved in hormones and breast cancer risk and mortality, and may contribute to the observed racial differences in breast cancer incidence and survival [28, 30]. It is plausible that gene–environment interactions involving hormonal pathways may also play a role in these differences, particularly given the observed variation by race in distributions of hormone receptor status, hormone-related risk factors, and genetic factors [31–36].

In this study, we conducted a comprehensive evaluation of associations between *CYP19A1* polymorphisms and breast cancer risk and breast cancer mortality using data from the Breast Cancer Health Disparities Study (BCHDS), a multicenter population-based case–control study including US non-Hispanic White (NHW), U.S. Hispanic/NA, and Mexican women. We considered potential interactions between *CYP19A1* genetic variants and variables thought to influence the estrogen pathway [hormone therapy (HT) use, oral contraceptive (OC) use, menopausal status, estrogen receptor (ER), and progesterone receptor (PR) status]. Additionally, we considered possible heterogeneity of the association between *CYP19A1* genetic variants and breast cancer risk and mortality by race/ethnicity and NA ancestry.

## Methods

### Study population

The BCHD study [28] is a consortium of three previously conducted population-based case-control studies, the 4-Corners Breast Cancer Study (4-CBCS) [37], the San Francisco Bay Area Breast Cancer Study (SFBCS) [38, 39], and the Mexico Breast Cancer Study (MBCS) [40]. In-person interviews were conducted and blood or saliva samples were collected. The 4-CBCS participants were non-Hispanic White, Hispanic, or NA, between 25 and 79 years of age and included histologically confirmed breast cancer cases (in situ or invasive) with a first primary diagnosed between 10/1999 and 05/2004, and controls selected from the target populations of cases living in Arizona, Colorado, New Mexico, and Utah and were frequency-matched to cases on ethnicity and 5-year age distribution. SFBCS participants were aged 35–79 years from the San Francisco Bay Area diagnosed with a first primary histologically confirmed invasive breast cancer between 04/1995 and 04/2002 (Hispanics) or between 04/1995 and 04/1999 (NHWs); and controls were identified by random-digit dialing and frequency-matched to cases on ethnicity and 5-year age distribution. MBCS participants were between 28 and 74 years of age: Cases were women diagnosed with a new histologically confirmed breast cancer (in situ or invasive) between 01/2004 and 12/2007 at 12 participating hospitals from three main health care systems; controls were randomly selected from the catchment area of the 12 participating hospitals using a probabilistic multistage design. All participants signed informed written consent prior to interview, and all studies were approved by their corresponding Institutional Review Board for Human Subjects.

### Data harmonization

Data were harmonized across all study centers and questionnaires as previously described [28]. Women were classified as either premenopausal or postmenopausal based on responses to questions on menstrual history and hormone therapy (HT) use. Women who reported still having periods during the referent year were classified as premenopausal. Study-specific definitions were used to define postmenopausal women. Women were classified as postmenopausal if: (1) they reported a natural menopause; or (2) they reported taking HT and were still having periods and were at or above the 95th percentile of age for race/ethnicity of those who reported having a natural menopause (i.e., >12 months since their last period) within their study center. The age at which 95 % of women reported having a natural menopause was 58 for NHW and 56 for Hispanics from the 4-CBCS, 55 for NHW and 56 for Hispanics from the SFBCS, and 54 for the MBCS. Race/ethnicity was self-reported for 4-CBCS and SFBCS; all women from MBCS were considered Hispanic ethnicity.

### Genetic data

DNA was extracted from available whole-blood ( $n = 7,287$ ) or saliva ( $n = 634$ ) samples. Whole-genome amplification (WGA) was applied to the saliva-derived DNA samples prior to genotyping. A tagSNP approach was utilized to characterize variation across the candidate gene. TagSNPs were selected using the following parameters: (1) Linkage disequilibrium (LD) blocks were defined using a Caucasian LD map and an  $r^2 = 0.8$  [41];

(2) minor allele frequency (MAF)  $>0.1$ ; (3) range  $\pm 1,500$  base pairs from the initiation codon or termination codon; and (4) 1 SNP/LD bin [28]. Variants within 3 kb were included to capture the majority of the regulatory region of the *CYP19A1* gene. This approach is similar to one employed by Stram et al. [42], which was based on the genotyping of a high density of SNPs selected every 3–5 kb in and surrounding a candidate gene. Additionally, 104 ancestry informative markers (AIMs) were used to distinguish European and NA ancestry [28, 43]. All markers were genotyped using a multiplexed bead array assay format based on GoldenGate chemistry (Illumina, San Diego, California). A genotyping call rate of 99.93 % was attained (99.65 % for WGA samples). We included 132 blinded internal replicates representing 1.6 % of the sample set. The duplicate concordance rate was 99.996 % as determined by 193,297 matching genotypes among sample pairs [28]. We investigated associations with 25 *CYP19A1* tagSNPs: rs4275794, rs4646, rs2899472, rs700518, rs17703883, rs727479, rs10459592, rs12591359, rs12908960, rs7172156, rs11856927, rs2414099, rs17601876, rs2470158, rs730154, rs17523880, rs2470152, rs3751591, rs1902584, rs2445762, rs2470144, rs7174997, rs8025191, rs1961177, and rs6493497. We attempted to capture all of the variation across the *CYP19A1* gene using the 25 tagSNPs so that we have active surrogate markers, detecting all variants. These markers should be covering all of the unmeasured variants with 80 % correlation or better. These tag SNPs were chosen based on information obtained HapMap databases and were not supposed to be in high LD with one another. Supplemental Graph 1 includes an LD matrix stratified by ethnicity which shows that LD between SNPs is not high ( $r^2 < 0.80$ ) and is similar between ethnic groups for the most part. However, LD differs by ethnic groups for a few SNPs, for example: rs2899472 and rs2470158 are considered to be in high LD within NHW women ( $r^2 = 0.80$ ), but not within Hispanic/NA women ( $r^2 = 0.52$ ). Approximately four sets of SNPs (eight total SNPs) were considered to be in very high LD ( $r^2 > 0.80$ ) for both ethnic groups.

### Tumor characteristics and survival data

Data on tumor characteristics and survival were not available for cases from Mexico; therefore, evaluation of these variables was limited to data (available through Spring 2012) obtained from 4-CBCS and SFBCS. Statewide cancer registries in Arizona, Colorado, New Mexico, Utah, and California provided information on stage at diagnosis, ER and PR tumor status, months of survival after diagnosis (calculated as difference between diagnosis date and date of death or last follow-up), and primary cause of death (*International Classification of Diseases, 10th Revision*) [44].

### Statistical methods

The program STRUCTURE 2.0 was utilized to calculate individual proportion of genetic ancestry based on two founding populations (European and NA) [28, 45]. Participants were classified by level of percent NA ancestry, which was categorized based on the distribution of NA ancestry in the control population. Percent NA ancestry was evaluated by three groups (low  $<29$  %, moderate 29 to 70 %, and high  $>70$  %), to allow sufficient power to assess associations of ancestral groups with breast cancer risk and mortality as previously described [28]. Findings are presented by two NA ancestry groups ( $<29$  %, 29 %) because

most of the estimates for breast cancer risk and mortality are not divergent for moderate or high NA ancestry groups; thus increasing power for this group.

Descriptive statistics were calculated for all covariates by level of percent NA ancestry and case-control status, and chi-square ( $\chi^2$ ) tests were used to assess differences between groups. Minor allele frequencies (MAF) and Hardy-Weinberg equilibrium (HWE)  $p$  values were calculated based on the frequencies of alleles and genotypes in the control population. Unconditional multivariable logistic regression was used to estimate odds ratios (ORs) and 95 % confidence intervals (CIs) for breast cancer risk associations with SNPs. SNPs were further evaluated for their association with breast cancer risk by strata of race/ethnicity, NA ancestry, menopausal status, HT use (never, ever, former, current), and OC use (never, ever) for all women. Risk of developing a particular breast cancer subtype based on ER/PR tumor phenotype) was evaluated using multinomial logistic regression [46, 47] for the 4-CBCS and SFBCS.

Initially, a co-dominant mode of inheritance was used to determine the relationship between number of alleles expressed and risk of breast cancer or breast cancer death. The assumed co-dominant mode of inheritance allowed us to evaluate whether there was a trend toward a dominant or recessive genetic model, and whether there was power to collapse genotype groups. SNPs were also assessed as continuous variables, and linear trend  $p$  values for the overall model (and within strata) were reported. All models were evaluated, and the most appropriate models for each SNP were selected for subsequent analyses. Confounding by categories of body mass index, first-degree family history of breast cancer, menopausal status, parity, OC use, HT use, alcohol consumption, and smoking status, continuous measures of physical activity and NA ancestry were evaluated. Covariates were considered confounders if the univariate  $p$  value was  $\leq 0.20$  and adjustment produced a change of  $\geq 10\%$  in the effect estimate for the overall association of the genotype with breast cancer risk [48]. We did not observe confounding by any factor assessed. However, all models were adjusted for matching variables of age and study center as well as by NA ancestry. For stratified analyses, tests for interactions were calculated using a one degree of freedom (1  $df$ ) likelihood ratio test for logistic regression models with and without an interaction term. All cases (in situ and invasive) and controls were included in the analysis of breast cancer risk ( $n = 7,733$ ). Sensitivity analysis restricted to only invasive cancers was not conducted because it would limit power as well as comparisons to the estimates including all women, because stage of disease is not available for MBCS participants.

For survival analyses, associations between SNPs and breast cancer mortality were evaluated using multivariable Cox proportional hazard models to obtain hazard ratios (HRs) and 95 % CIs adjusted for SEER summary stage at diagnosis (local, regional, and distant), age, NA ancestry, and study center. Participants were censored if lost to follow-up or died of cause other than breast cancer. Stratified analyses were conducted to determine the presence of effect modification by race/ethnicity, NA ancestry, and menopausal status and assessed using  $p$  values from 1  $df$  likelihood ratio tests. Invasive cases from 4-CBCS and SFBCS with available data for survival and stage were included in the survival analysis ( $n = 2,218$ ).



*p* values from linear trend tests and interactions in overall models and within stratum were adjusted for multiple comparisons (MCs) within each gene using a step-down Bonferroni correction method [49]. This method takes into account the correlation of the data using the SNP spectral decomposition method based on the eigenvalues of the correlation matrix among the SNPs for each gene as proposed by Nyholt [50] and modified by Li and Ji [51]. The effective number of independent SNPs is determined and used when adjusting for MC, rather than total SNP number. This method is conservative, especially when evaluating correlated variables such as SNPs within a gene. However, it is less conservative than the conventional Bonferroni correction because you have more opportunities to reject the null hypotheses, which results in an increase in statistical power [51]. An adjusted  $p = 0.05$  ( $\alpha = 0.05$ ) for main effects and interactions was considered statistically significant. All statistical analyses were performed using SAS version 9.3 (SAS Institute, Cary, NC).

## Results

Breast cancer cases included in this study were predominantly postmenopausal (59.7 %) and between the ages of 40–59 years of age (58.7 %) (Table 1). Compared to women with low NA ancestry (<29 %), cases and controls, respectively, with moderate to high NA ancestry (29 %) had a higher proportion of premenopausal status (40.3 and 39.9 vs. 33.6 and 31.5 %), a low proportion tended to report ever use of HT (27.5 and 25.6 vs. 53.4 and 54.0 %) and OC (54.1 and 52.8 vs. 69.6 and 64.9 %), and cases more frequently presented with ER–/PR– tumors (17.4 vs. 13.2 %) and regional stage of disease (38.9 vs. 29.8 %). Majority of cases and controls with low NA ancestry were NHW ethnicity (83.7 and 85.0 %, respectively), whereas cases and controls with moderate to high NA ancestry were predominantly Hispanic/NA ethnicity (99.4 and 99.7 %, respectively). A similar proportion of breast cancer cases were deceased in low NA ancestry (15.8 %) and moderate to high NA ancestry (16.3 %), and majority of deaths were due to breast cancer (56.6 and 62 %, respectively) (data not shown in table). Table 2 provides a description of the selected SNPs. *CYP19A1* (rs700518) is located in exon 3; both rs4646 and rs4275794 are located in the 3′-untranslated region (c).

### Associations with breast cancer risk

When considering all women combined, we did not observe any significant associations between the 25 tag-SNPs on *CYP19A1* and breast cancer risk after MC adjustment. Prior to MC adjustment, one SNP (rs4646) was inversely associated with breast cancer risk for all women ( $OR_{AA} 0.87$ , 95 % CI 0.77–0.98) (Table 3). Results stratified by NA ancestry were divergent between women with low (<29 %) or moderate to high (29 %) NA ancestry for several SNPs. We observed inverse associations ( $OR 0.64–0.82$ ) for the variant alleles in three SNPs (rs10459592, rs1961177, and rs6493497) and positive associations ( $OR 1.20–1.67$ ) for variant alleles in five SNPs (rs4275794, rs2899472, rs700518, rs12908960, and rs11856927) for women with 29 % NA ancestry, compared to women with <29 % NA ancestry (Table 3). After MC adjustment, two SNPs remained significantly associated with increased breast cancer risk in women with 29 % NA ancestry (rs700518,  $OR_{GG} 1.36$ , 95 % CI 1.11–1.67) and (rs11856927,  $OR_{TG} 1.20$ , 95 % CI 1.05–1.36) and ( $OR_{GG} 1.35$ , 95 % CI 1.05–1.72). We observed several interactions with NA ancestry; however, none of these

associations remained statistically significant after MC adjustment (Table 3). Results were similar when stratified by race/ethnicity for all women (NHW vs. Hispanic/NA) as well as by NA ancestry within Hispanic/NA only (data not shown). There was one significant interaction observed with menopausal status after MC adjustment: rs2470144 ( $p_{\text{adj}} = 0.03$ ); breast cancer risk was increased among postmenopausal women ( $OR_{AA} = 1.22$ , 95 % CI 1.05–1.14); however, an inverse relationship was seen among premenopausal women ( $OR_{AA} = 0.78$ , 95 % CI 0.64–0.95) (data not shown in table).

### **CYP19A1 gene–environment interactions and breast cancer risk**

We considered possible interactions between *CYP19A1* SNPs and HT use and OC use. None of the observed interactions remained statistically significant after MC adjustment (data not shown). None of the SNPs were considered significantly associated with breast cancer risk when considering breast cancer subtypes defined by ER/PR tumor phenotypes.

### **Associations with breast cancer mortality**

None of the *CYP19A1* SNPs were associated with risk of breast cancer mortality, for all cases combined after MC adjustment. Only rs2470152 was associated with an increased risk ( $HR_{CC} = 1.40$ , 95 % CI 1.03–1.91,  $p_{\text{adj}} = 0.48$ ) prior to MC adjustment (Table 4). Prior to MC adjustment, we observed an interaction between rs12591359 and NA ancestry ( $p = 0.03$ ): a modest increase in risk for those with <29 % NA ancestry ( $HR_{AA} = 1.70$ , 95 % CI 1.00–2.62) and a non-significant reduced risk for those with ≥29 % NA ancestry ( $HR_{AA} = 0.77$ , 95 % CI 0.45–1.34). Additionally, three SNPs (rs2414099, rs2445762, and rs8025191) showed inverse associations for the variant allele(s) within the <29 % NA ancestry strata ( $HR = 0.59–0.65$ ) (Table 4). However, tests of interaction were not statistically significant for any of these SNPs after MC adjustment. Analyses stratifying by menopausal status showed no statistically significant associations after MC adjustment.

## **Discussion**

We examined the associations between 25 tagSNPs in *CYP19A1*, a gene involved in the biosynthesis of estrogen, and breast cancer risk and mortality among Hispanic/NA and NHW women taking into account the proportion of NA ancestry. We also evaluated associations taking into account breast cancer subtypes defined by ER/PR status, menopausal status, and possible gene by environment interactions considering HT and OC use. None of the SNPs remained significantly associated with breast cancer risk after MC adjustment. However, many associations tended to differ by NA ancestry groups and two SNPs (rs700518 and rs11856927) remained associated with a slightly increased risk of breast cancer among women with moderate to high NA ancestry; these associations withstood MC adjustments. There was one significant interaction with menopausal status (rs2470144) in which risk was increased for postmenopausal women only. Generally, there were no distinct associations when considering breast cancer subtypes defined by ER/PR, or interactions with HT or OC use. None of the SNPs remained significantly associated with breast cancer mortality overall or when stratified by NA ancestry after MC adjustment.



The *CYP19A1* gene encodes for aromatase, a 503-amino acid protein containing nine coding exons, is found to be expressed in breast tissue, and is also the source of local estrogen production in breast tumors [52]. The expression of *CYP19A1* is regulated by tissue-specific promoters which differ between normal breast tissue and breast cancer tissue. The switch in promoters alters plasma estrogen levels that influence estrogen levels in normal breast tissue and breast cancer tissue and are related to increased risk of breast cancer [7, 10]. Due to the fact that the majority of breast cancers are hormone receptor positive and driven by the ER signaling pathway, genetic variation of genes involved in regulation of hormone production is hypothesized to be a plausible factor in breast cancer susceptibility and survival.

Generally, epidemiology studies including predominantly NHW populations have found that genetic variants in *CYP19A1* do not have a strong effect on breast cancer risk, regardless of study design (case-control vs. cohort, tagSNP vs. haplotype approach) [10, 11, 13, 15, 17–20, 53–59]. Some studies found associations among subgroups, for example: NHW premenopausal women and ER– tumor status [53], and Japanese women with ER+ premenopausal breast cancer [59]. One study found a significant increase in endogenous estrogen levels associated with several SNPs in NHW postmenopausal women [13]. Diergaard et al. [60] evaluated an interaction between several *CYP19A1* polymorphisms and HT use and did not find any significant associations in a NHW population, similar to our findings in this genetically admixed population. Furthermore, several studies [21–26] have examined genetic variation on *CYP19A1* and risk of breast cancer mortality, although findings are mixed. For example, Long et al. [22] found an increased risk of breast cancer mortality for five *CYP19A1* SNPs in a Chinese population, while Fasching et al. [21] found an inverse relationship for one SNP (rs10046) and Udler et al. [23] did not find associations with four *CYP19A1* SNPs in a NHW population. This study does not provide conclusive evidence that genetic variation in *CYP19A1* is a prognostic factor for breast cancer mortality, or that NA ancestry modifies the association.

To our knowledge, the three tagSNPs we found associated with breast cancer risk in moderate to high NA ancestry (rs700518 and rs11856927) and postmenopausal women (rs2470144) have not been previously evaluated in epidemiology studies. We utilized the program SNAP (SNP Annotation and Proxy Search, Broad Institute) [61] to evaluate whether these three SNPs were in high LD with SNPs reported in previous literature. Pair-wise LD is pre-calculated based on phased genotype data from the International HapMap Project [62]. *CYP19A1* rs11856927 or rs2470144 was not found to be in high LD with previously reported SNPs. *CYP19A1* rs700518 and rs10046 are in high LD ( $r^2 = 0.87$ ) in the CEU population (European ancestry); however, we do not know if it is in high LD in the Mexican ancestry population (MEX). *CYP19A1* rs700518 and rs4775936 are also found to be in high LD in both CEU and MEX populations ( $r^2 = 0.81$  and  $r^2 = 0.87$ , respectively). *CYP19A1* rs10046 (T) and the rs4775936 (A) are reportedly associated with higher levels of circulating estrogen or estrogen to androgen ratios in several studies [10–13, 15]. Despite this relationship, studies have not consistently found an association between these two SNPs and breast cancer risk [15, 63]. One study found that rs10046 is associated with breast cancer risk in Chinese women [64]. In addition to being in LD with these two SNPs that influence estrogen levels, *CYP19A1* rs700518 is located in a coding region (exon 3) and was

found to be an ‘allelic expression marker’; part of a haplotype that is associated with expression of this aromatase gene [14]. Moreover, *CYP19A1* rs700518 has been reported as a predictor of endocrine therapy effectiveness in ER+ breast cancer [65].

Findings from our study indicate that NA ancestry may modify the association between genetic variation in *CYP19A1* and breast cancer risk. Two possible explanations for the observed results include: (1) Selected genetic variants in *CYP19A1* may be susceptibility loci or surrogates for unmeasured functional variants that are more frequent in women with higher NA ancestry and influence differences in breast cancer risk; and (2) estimated NA ancestry may serve as a genetic component to the difference in breast cancer risk. Additionally, we cannot rule out the possibility of three-way gene–environment interactions with *CYP19A1* variants, NA ancestry, and environmental factors, which future research could explain. Further evaluation of genetic variants associated with breast cancer risk that varies in frequency and LD patterns across racial/ethnic groups is needed and may help to identify regions of the genome associated with breast cancer by ethnic groups [36]. For example, Fejerman et al. used an admixture mapping approach and confirmed that the observed association between genetic ancestry and breast cancer risk and is in part due to admixture signals in genomic regions (6q25 and 11p15) that differ between indigenous European and NA ancestral groups [30].

This study has several strengths and limitations. This is the first study to evaluate genetic variation on the *CYP19A1* gene in Hispanic/NA women and NA ancestry groups. The large sample size allowed us to evaluate different modes of inheritance and to conduct stratified analyses. Nevertheless, power was limited when assessing breast cancer survival and ER/PR status, partly due to lack of data from the MBCS. A strength of the survival analyses conducted is the long follow-up period for 4-CBCS and SFBCS, approximately 8 and 10 years, respectively. When stratifying by NA ancestry, we reported estimates including both Hispanic/NA and NHW in the analyses of breast cancer risk and mortality. We were able to rely on genetic markers to determine ancestry and not assume ancestry based on self-reported race. Moreover, women from Mexico were not asked their race/ethnicity and were all considered Hispanic. Interestingly, that study population includes a small proportion of women with low NA ancestry. Some Hispanics/NA in 4-CBCS and SFBCS also have low NA ancestry. Additionally, utilizing a tagSNP approach we were able to cover variation across the entire gene; however, this approach resulted in assessment of many SNPs making it difficult to observe significant findings after adjustment for MC, which was conservative and may result in false-negative associations. Lastly, after FDR adjustment, one SNP (rs4646) was found to violate HWE and was kept in analysis because these results indicate that there is a difference in observed genotypes by ethnic groups, which we hypothesize, may be due to biological differences in ancestry and may be informative for future studies including women of NA ancestry.

In conclusion, our unique findings suggest there is a positive association with genetic variants in *CYP19A1* and breast cancer risk in women of moderate to high NA ancestry, which was not observed in women with low NA ancestry. The plausible functionality of *CYP19A1* rs700518 could justify the association observed in women with moderate to high NA ancestry; however, this finding needs to be replicated. These results strengthen the

hypothesis that the proportion of estimated NA ancestry may influence biological pathways, increasing susceptibility to breast cancer.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

We would also like to acknowledge the contributions of the following individuals to the study: Sandra Edwards for data harmonization oversight; Erica Wolff and Michael Hoffman for laboratory support; Carolina Ortega for her assistance with data management for the Mexico Breast Cancer Study, Jocelyn Koo for data management for the San Francisco Bay Area Breast Cancer Study, Dr. Tim Byers for his contribution to the 4-Corner's Breast Cancer Study, and Dr. Josh Galanter for assistance in selection of AIMS markers. The Breast Cancer Health Disparities Study was funded by Grant CA14002 from the National Cancer Institute to Dr. Slattery. The San Francisco Bay Area Breast Cancer Study was supported by Grants CA63446 and CA77305 from the National Cancer Institute, Grant DAMD17-96-1-6071 from the U.S. Department of Defense and Grant 7PB-0068 from the California Breast Cancer Research Program. The collection of cancer incidence data used in this study was supported by the California Department of Public Health as part of the statewide cancer reporting program mandated by California Health and Safety Code Section 103885; the National Cancer Institute's Surveillance, Epidemiology, and End Results Program under contract HHSN261201000036C awarded to the Cancer Prevention Institute of California; and the Centers for Disease Control and Prevention's National Program of Cancer Registries, under agreement #1U58 DP000807-01 awarded to the Public Health Institute. The 4-Corners Breast Cancer Study was funded by Grants CA078682, CA078762, CA078552, and CA078802 from the National Cancer Institute. The research also was supported by the Utah Cancer Registry, which is funded by contract N01-PC-67000 from the National Cancer Institute, with additional support from the State of Utah Department of Health, the New Mexico Tumor Registry, and the Arizona and Colorado cancer registries, funded by the Centers for Disease Control and Prevention National Program of Cancer Registries and additional state support. The contents of this manuscript are solely the responsibility of the authors and do not necessarily represent the official view of the National Cancer Institute or endorsement by the State of California Department of Public Health, the National Cancer Institute, and the Centers for Disease Control and Prevention or their Contractors and Subcontractors. The Mexico Breast Cancer Study was funded by Consejo Nacional de Ciencia y Tecnología (CONACyT) (SALUD-2002-C01-7462).

## References

1. Simpson ER, et al. Aromatase cytochrome P450, the enzyme responsible for estrogen biosynthesis. *Endocr Rev.* 1994; 15(3):342–355. [PubMed: 8076586]
2. Mahendroo MS, et al. Tissue-specific expression of human P-450AROM. The promoter responsible for expression in adipose tissue is different from that utilized in placenta. *J Biol Chem.* 1991; 266(17):11276–11281. [PubMed: 2040633]
3. Hulka BS, Liu ET, Lininger RA. Steroid hormones and risk of breast cancer. *Cancer.* 1994; 74(3 Suppl):1111–1124. [PubMed: 8039146]
4. Eliassen AH, et al. Urinary estrogens and estrogen metabolites and subsequent risk of breast cancer among pre-menopausal women. *Cancer Res.* 2012; 72(3):696–706. [PubMed: 22144471]
5. Key T, et al. Endogenous sex hormones and breast cancer in postmenopausal women: reanalysis of nine prospective studies. *J Natl Cancer Inst.* 2002; 94(8):606–616. [PubMed: 11959894]
6. Henderson, BE.; Ponder, B.; Ross, R. *Hormones, genes, and cancer.* Oxford University Press; New York: 2003. p. 450
7. Bulun SE, et al. A link between breast cancer and local estrogen biosynthesis suggested by quantification of breast adipose tissue aromatase cytochrome P450 transcripts using competitive polymerase chain reaction after reverse transcription. *J Clin Endocrinol Metab.* 1993; 77(6):1622–1628. [PubMed: 8117355]
8. Platet N, et al. Estrogens and their receptors in breast cancer progression: a dual role in cancer proliferation and invasion. *Crit Rev Oncol Hematol.* 2004; 51(1):55–67. [PubMed: 15207254]
9. Saha Roy S, Vadlamudi RK. Role of estrogen receptor signaling in breast cancer metastasis. *Int J Breast Cancer.* 2012; 2012:654698. [PubMed: 22295247]

10. Dunning AM, et al. Polymorphisms associated with circulating sex hormone levels in postmenopausal women. *J Natl Cancer Inst.* 2004; 96(12):936–945. [PubMed: 15199113]
11. Beckmann L, et al. Comprehensive analysis of hormone and genetic variation in 36 genes related to steroid hormone metabolism in pre- and postmenopausal women from the breast and prostate cancer cohort consortium (BPC3). *J Clin Endocrinol Metab.* 2011; 96(2):E360–E367. [PubMed: 21177793]
12. Lundin E, et al. Selected polymorphisms in sex hormone-related genes, circulating sex hormones and risk of endometrial cancer. *Cancer Epidemiol.* 2012; 36(5):445–452. [PubMed: 22633539]
13. Haiman CA, et al. Genetic variation at the CYP19A1 locus predicts circulating estrogen levels but not breast cancer risk in postmenopausal women. *Cancer Res.* 2007; 67(5):1893–1897. [PubMed: 17325027]
14. Riancho JA, et al. Identification of an aromatase haplotype that is associated with gene expression and postmenopausal osteoporosis. *J Clin Endocrinol Metab.* 2007; 92(2):660–665. [PubMed: 17118999]
15. Clendenen T, et al. Genetic variants in hormone-related genes and risk of breast cancer. *PLoS ONE.* 2013; 8(7):e69367. [PubMed: 23935996]
16. Zhang B, et al. Genetic variants associated with breast-cancer risk: comprehensive research synopsis, meta-analysis, and epidemiological evidence. *Lancet Oncol.* 2011; 12(5):477–488. [PubMed: 21514219]
17. Canzian F, et al. Comprehensive analysis of common genetic variation in 61 genes related to steroid hormone and insulin-like growth factor-I metabolism and breast cancer risk in the NCI breast and prostate cancer cohort consortium. *Hum Mol Genet.* 2010; 19(19):3873–3884. [PubMed: 20634197]
18. Chen C, et al. Genetic variation in CYP19A1 and risk of breast cancer and fibrocystic breast conditions among women in Shanghai China. *Cancer Epidemiol Biomarkers Prev.* 2008; 17(12):3457–3466. [PubMed: 19064562]
19. Kristensen VN, et al. Genetic variants of CYP19 (aromatase) and breast cancer risk. *Oncogene.* 2000; 19(10):1329–1333. [PubMed: 10713674]
20. Olson JE, et al. A comprehensive examination of CYP19 variation and risk of breast cancer using two haplotype-tagging approaches. *Breast Cancer Res Treat.* 2007; 102(2):237–247. [PubMed: 17004113]
21. Fasching PA, et al. Single nucleotide polymorphisms of the aromatase gene (CYP19A1), HER2/neu status, and prognosis in breast cancer patients. *Breast Cancer Res Treat.* 2008; 112(1):89–98. [PubMed: 18049890]
22. Long JR, et al. Genetic polymorphisms of the CYP19A1 gene and breast cancer survival. *Cancer Epidemiol Biomarkers Prev.* 2006; 15(11):2115–2122. [PubMed: 17119036]
23. Udler MS, et al. Common germline polymorphisms in COMT, CYP19A1, ESR1, PGR, SULT1E1 and STS and survival after a diagnosis of breast cancer. *Int J Cancer.* 2009; 125(11):2687–2696. [PubMed: 19551860]
24. Goode EL, et al. Effect of germ-line genetic variation on breast cancer survival in a population-based study. *Cancer Res.* 2002; 62(11):3052–3057. [PubMed: 12036913]
25. Huang CS, et al. The CYP19 TTTA repeat polymorphism is related to the prognosis of premenopausal stage I–II and operable stage III breast cancers. *Oncologist.* 2008; 13(7):751–760. [PubMed: 18614591]
26. Kuo SH, et al. CYP19 genetic polymorphism haplotype AASA is associated with a poor prognosis in premenopausal women with lymph node-negative, hormone receptor-positive breast cancer. *Biomed Res Int.* 2013; 2013:562197. [PubMed: 24324964]
27. Price AL, et al. A genomewide admixture map for Latino populations. *Am J Hum Genet.* 2007; 80(6):1024–1036. [PubMed: 17503322]
28. Slattery M, et al. Genetic variation in genes involved in hormones, inflammation, and energetic factors and breast cancer risk in an admixed population. *Carcinogenesis.* 2012; 33(8):1512–1521. [PubMed: 22562547]
29. Fejerman L, et al. European ancestry is positively associated with breast cancer risk in Mexican women. *Cancer Epidemiol Biomarkers Prev.* 2010; 19(4):1074–1082. [PubMed: 20332279]

30. Fejerman L, et al. Admixture mapping identifies a locus on 6q25 associated with breast cancer risk in US Latinas. *Hum Mol Genet.* 2012; 21(8):1907–1917. [PubMed: 22228098]
31. Li CI, Malone KE, Daling JR. Differences in breast cancer stage, treatment, and survival by race and ethnicity. *Arch Intern Med.* 2003; 163:49–56. [PubMed: 12523916]
32. Hines LM, et al. Comparative analysis of breast cancer risk factors among Hispanic and non-Hispanic white women. *Cancer.* 2010; 116(13):3215–3223. [PubMed: 20564638]
33. Hines LM, et al. Ethnic disparities in breast tumor phenotypic subtypes in Hispanic and non-Hispanic white women. *J Womens Health (Larchmt).* 2011; 20(10):1543–1550. [PubMed: 21721934]
34. Boone SD, et al. The joint contribution of tumor phenotype and education to breast cancer survival disparity between Hispanic and non-Hispanic white women. *Cancer Causes Control.* 2014; 25(3): 273–282. [PubMed: 24337810]
35. Li R, et al. Hormone replacement therapy and breast carcinoma risk in Hispanic and non-Hispanic women. *Cancer.* 2002; 95(5):960–968. [PubMed: 12209678]
36. Gonzalez Burchard E. Latino populations: a unique opportunity for the study of race, genetics, and social environment in epidemiological research. *Am J Public Health.* 2005; 95(12):2161–2168. [PubMed: 16257940]
37. Slattery ML, et al. Body size, weight change, fat distribution and breast cancer risk in Hispanic and non-Hispanic white women. *Breast Cancer Res Treat.* 2007; 102(1):85–101. [PubMed: 17080310]
38. 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(11 Pt 1):1143–1152. [PubMed: 14652273]
39. John EM, et al. Migration history, acculturation, and breast cancer risk in Hispanic women. *Cancer Epidemiol Biomarkers Prev.* 2005; 14(12):2905–2913. [PubMed: 16365008]
40. Angeles-Llerenas A, et al. Moderate physical activity and breast cancer risk: the effect of menopausal status. *Cancer Causes Control.* 2010; 21(4):577–586. [PubMed: 20084545]
41. Carlson CS, et al. Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. *Am J Hum Genet.* 2004; 74(1):106–120. [PubMed: 14681826]
42. Stram DO, et al. Choosing haplotype-tagging SNPS based on unphased genotype data using a preliminary sample of unrelated subjects with an example from the multiethnic cohort study. *Hum Hered.* 2003; 55(1):27–36. [PubMed: 12890923]
43. Collins-Schramm HE, 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(3):263–271. [PubMed: 14628215]
44. WHO. International statistical classification of diseases and related health problems (10th revision). World Health Organization; Geneva: 1992.
45. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics.* 2000; 155(2):945–959. [PubMed: 10835412]
46. Dubin N, Pasternack BS. Risk assessment for case-control subgroups by polychotomous logistic regression. *Am J Epidemiol.* 1986; 123(6):1101–1117. [PubMed: 3706280]
47. Hosmer, D., Jr; Lemeshow, S. Probability and mathematical statistics. Wiley; New York: 1989. Applied logistic regression.
48. Bursac Z, et al. Purposeful selection of variables in logistic regression. *Source Code Biol Med.* 2008; 3:17. [PubMed: 19087314]
49. Holm S. A simple sequentially rejective multiple test procedure. *Scand J Stat.* 1979; 6:65–70.
50. Nyholt DR. A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. *Am J Hum Genet.* 2004; 74(4):765–769. [PubMed: 14997420]
51. Li J, Ji L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity (Edinb).* 2005; 95(3):221–227. [PubMed: 16077740]
52. Yue W, et al. In situ aromatization enhances breast tumor estradiol levels and cellular proliferation. *Cancer Res.* 1998; 58(5):927–932. [PubMed: 9500452]

53. Talbott KE, et al. A CYP19 (aromatase) polymorphism is associated with increased premenopausal breast cancer risk. *Breast Cancer Res Treat.* 2008; 111(3):481–487. [PubMed: 17975727]
54. Zins K, et al. Analysis of the rs10046 polymorphism of aromatase (CYP19) in premenopausal onset of human breast cancer. *Int J Mol Sci.* 2014; 15(1):712–724. [PubMed: 24402127]
55. Haiman CA, et al. No association between a single nucleotide polymorphism in CYP19 and breast cancer risk. *Cancer Epidemiol Biomarkers Prev.* 2002; 11(2):215–216. [PubMed: 11867511]
56. Haiman CA, et al. A tetranucleotide repeat polymorphism in CYP19 and breast cancer risk. *Int J Cancer.* 2000; 87(2):204–210. [PubMed: 10861475]
57. Haiman CA, et al. A comprehensive haplotype analysis of CYP19 and breast cancer risk: the multiethnic cohort. *Hum Mol Genet.* 2003; 12(20):2679–2692. [PubMed: 12944421]
58. Reding KW, et al. Estrogen-related genes and their contribution to racial differences in breast cancer risk. *Cancer Causes Control.* 2012; 23(5):671–681. [PubMed: 22418777]
59. Yoshimoto N, et al. Genetic and environmental predictors, endogenous hormones and growth factors, and risk of estrogen receptor-positive breast cancer in Japanese women. *Cancer Sci.* 2011; 102(11):2065–2072. [PubMed: 21790896]
60. Diergaarde B, et al. Polymorphisms in genes involved in sex hormone metabolism, estrogen plus progestin hormone therapy use, and risk of postmenopausal breast cancer. *Cancer Epidemiol Biomarkers Prev.* 2008; 17(7):1751–1759. [PubMed: 18628428]
61. Johnson AD, et al. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics.* 2008; 24(24):2938–2939. [PubMed: 18974171]
62. International HapMap C. The International HapMap Project. *Nature.* 2003; 426(6968):789–796. [PubMed: 14685227]
63. Pineda B, et al. Associations between aromatase CYP19 rs10046 polymorphism and breast cancer risk: from a case–control to a meta-analysis of 20,098 subjects. *PLoS One.* 2013; 8(1):e53902. [PubMed: 23342035]
64. Zhang L, et al. Association of genetic polymorphisms of ER-alpha and the estradiol-synthesizing enzyme genes CYP17 and CYP19 with breast cancer risk in Chinese women. *Breast Cancer Res Treat.* 2009; 114(2):327–338. [PubMed: 18629629]
65. Miron L, et al. Research on aromatase gene (CYP19A1) polymorphisms as a predictor of endocrine therapy effectiveness in breast cancer. *Rev Med Chir Soc Med Nat Iasi.* 2012; 116(4): 997–1004. [PubMed: 23700878]



**Table 1**

Descriptive characteristics by Native American ancestry and case–control status: the Breast Cancer Health Disparities Study ( $n = 7,733$ )

Characteristic	<29 % Native American ancestry <sup>a</sup>		29 % Native American ancestry <sup>b</sup>	
	Cases <i>n</i> (%)	Controls <i>n</i> (%)	Cases <i>n</i> (%)	Controls <i>n</i> (%)
<i>Total subjects<sup>c</sup></i>	1,696 (21.9)	1,871 (24.2)	1,828 (23.6)	2,338 (30.2)
Study site				
4-CBCS	1,267 (74.7)	1,462 (78.1)	489 (26.8)	609 (26.1)
MBCS	26 (1.5)	11 (0.6)	790 (43.2)	983 (42.0)
SFBCS	403 (23.8)	398 (21.3)	549 (30.0)	746 (31.9)
Race/ethnicity				
Non-Hispanic White	1,420 (83.7)	1,591 (85.0)	11 (0.6)	8 (0.3)
Hispanic	276 (16.3)	280 (15.0)	1,817 (99.4)	2,330 (99.7)
Age (years)				
<40	109 (6.4)	149 (8.0)	176 (9.6)	281 (12.0)
40–49	483 (28.5)	485 (25.9)	626 (34.3)	758 (32.4)
50–59	482 (28.4)	493 (26.4)	535 (29.3)	675 (28.9)
60–69	398 (23.5)	405 (21.7)	367 (20.1)	481 (20.6)
70	224 (13.2)	339 (18.1)	124 (6.8)	143 (6.1)
Menopausal status				
Premenopausal	570 (33.6)	589 (31.5)	736 (40.3)	932 (39.9)
Postmenopausal	1,070 (63.1)	1,236 (66.1)	1,035 (56.6)	1,338 (57.2)
Missing	56 (3.3)	46 (2.5)	57 (3.1)	68 (2.9)
Hormone therapy use				
Never	481 (28.4)	529 (28.3)	1,155 (63.2)	1,580 (67.6)
Ever	906 (53.4)	1,010 (54.0)	503 (27.5)	598 (25.6)
Missing	309 (18.2)	332 (17.7)	170 (9.3)	160 (6.8)
Oral contraceptive use				
Never	484 (28.5)	629 (33.6)	813 (44.5)	1,077 (46.1)
Ever	1,181 (69.6)	1,214 (64.9)	989 (54.1)	1,234 (52.8)
Missing	31 (1.8)	28 (1.5)	26 (1.4)	27 (1.2)
Estrogen/progesterone receptor status <sup>d</sup>				
ER+/PR+	786 (47.1)	–	482 (46.4)	–
ER+/PR–	138 (8.3)	–	92 (8.9)	–
ER–/PR+	19 (1.1)	–	24 (2.3)	–
ER–/PR–	221 (13.2)	–	181 (17.4)	–
Missing	506 (30.3)	–	259 (24.9)	–
SEER summary stage <sup>d</sup>				
Local	928 (69.1)	–	526 (60.2)	–
Regional	401 (29.8)	–	340 (38.9)	–
Distant	15 (1.1)	–	8 (0.9)	–

4-CBCS 4-Corners Breast Cancer Study, MBCS Mexico Breast Cancer Study, SFBCS San Francisco Bay Area Breast Cancer Study. Column percentages may not add up to 100 % due to rounding

<sup>a</sup> Statistically significant differences ( $p < 0.05$ ) between cases and controls in women with <29 % Native American ancestry for study and oral contraceptive use

<sup>b</sup> Statistically significant differences ( $p < 0.05$ ) between cases and controls in women with ≥ 29 % Native American ancestry for hormone therapy use

<sup>c</sup> Statistically significant differences ( $p < 0.05$ ) between Native American ancestry groups, regardless of case-control status, were observed for all variables

<sup>d</sup> Data on ER/PR status and stage are for primary invasive cancers only and are not available for the Mexico site

**Table 2**

Description of all tagSNPs in *CYP19A1* gene genotyped

SNP (rs)	Chromosome		Alleles		MAF <sup>a</sup>		HWE <sup>b</sup>		Proportion missing
	Region	Position	Major	Minor	NHW	Hispanic/NA	NHW	Hispanic/NA	
rs4275794	3' UTR	51,501,117	T	C	0.2025	0.1725	0.9649	0.7076	0.0002
rs4646	3' UTR	51,502,844	C	A	0.2647	0.4820	0.9729	0.0216	0.0002
rs2899472	INTRON	51,516,055	C	A	0.2638	0.1309	0.9783	0.8558	0.0002
rs700518	Exon 3	51,529,112	A	G	0.4919	0.3103	0.9649	0.9152	0.0002
rs17703883	INTRON	51,530,097	T	C	0.2509	0.3441	0.9649	0.3225	0.0000
rs727479	INTRON	51,534,547	T	G	0.3508	0.4020	0.9649	0.0742	0.0019
rs10459592	INTERGENIC	51,536,141	G	T	0.4334	0.4828	0.9635	0.2682	0.0005
rs12591359	INTERGENIC	51,539,368	G	A	0.4106	0.4716	0.9834	0.1758	0.0005
rs12908960	INTERGENIC	51,545,860	G	A	0.4581	0.2648	0.9961	0.4525	0.0000
rs7172156	INTERGENIC	51,546,298	A	G	0.395	0.4138	0.9635	0.3029	0.0000
rs11856927	INTERGENIC	51,548,705	T	G	0.4115	0.2536	0.9851	0.3950	0.0010
rs2414099	INTERGENIC	51,548,782	T	C	0.1644	0.2468	0.9649	0.4325	0.0002
rs17601876	INTERGENIC	51,553,909	A	G	0.4759	0.3611	0.9649	0.3161	0.0002
rs2470158	INTERGENIC	51,588,395	C	T	0.0991	0.0546	0.9649	0.9740	0.0000
rs730154	INTERGENIC	51,591,204	A	G	0.1457	0.1791	0.9343	0.9293	0.0007
rs17523880	INTERGENIC	51,592,543	C	A	0.1419	0.0812	0.8892	0.6817	0.0000
rs2470152	INTERGENIC	51,594,972	T	C	0.4941	0.4211	0.9961	0.4631	0.0000
rs3751591	INTERGENIC	51,606,710	T	C	0.1644	0.199	0.9649	0.5091	0.0000
rs1902584	INTERGENIC	51,611,654	A	T	0.0809	0.1115	0.9649	0.5892	0.0000
rs2445762	INTRON	51,617,708	T	C	0.2856	0.4117	0.9649	0.8268	0.0019
rs2470144	INTRON	51,621,725	G	A	0.4975	0.3789	0.9649	0.9462	0.0000
rs7174997	INTRON	51,622,128	G	T	0.1801	0.1461	0.6760	0.4968	0.0007
rs8025191	INTRON	51,623,785	A	G	0.2166	0.2782	0.9649	0.1786	0.0000
rs1961177	INTRON	51,625,078	C	T	0.1181	0.1651	0.9552	0.4051	0.0000
rs6493497	INTERGENIC	51,630,835	G	A	0.1166	0.1557	0.9649	0.5028	0.0000

<sup>a</sup> Minor allele frequency (MAF) is based on control population

<sup>b</sup> Hardy-Weinberg equilibrium (HWE) is based on the control population and is adjusted for false discovery rate (FDR)

Table 3

Overall associations between selected variants in *CYP19A1* gene with breast cancer risk, stratified by Native American ancestry, the Breast Cancer Health Disparities Study

Gene/Sup	Genotype	All women		Stratified by Native American ancestry		<i>p</i> <sup>c</sup>
		Cases <i>n</i> (%)	Controls <i>n</i> (%)	<29 % OR <sup>b</sup> (95 % CI)	29 % OR <sup>b</sup> (95 % CI)	
rs4275794	TT/TC	3,387 (45.40)	4,074 (54.60)	1.00	1.00	0.13 (0.90)
	CC	137 (50.55)	134 (49.45)	1.23 (0.96–1.56)	1.03 (0.74–1.45)	1.47 (1.03–2.10)
rs4646	CC/CA	2,984 (46.36)	3,452 (53.64)	1.00	1.00	0.31 (1.00)
	AA	540 (41.67)	756 (58.33)	0.87 (0.77–0.98)	0.94 (0.73–1.21)	0.82 (0.71–0.95)
rs2899472	CC	2,334 (45.10)	2,841 (54.90)	1.00	1.00	0.02 (0.28)
	CA	1,046 (46.43)	1,207 (53.57)	1.02 (0.92–1.13)	0.96 (0.84–1.11)	1.10 (0.95–1.28)
	AA	144 (47.37)	160 (52.63)	1.03 (0.81–1.30)	0.84 (0.64–1.12)	1.67 (1.07–2.61)
rs700518	AA/AG	2,893 (45.01)	3,534 (54.99)	1.00	1.00	0.01 (0.15)
	GG	631 (48.35)	674 (51.65)	1.10 (0.97–1.24)	0.98 (0.84–1.14)	<b>1.36 (1.11–1.67)</b>
rs10459592	GG	977 (46.86)	1,108 (53.14)	1.00	1.00	0.02 (0.28)
	GT	1,753 (45.32)	2,115 (54.68)	0.96 (0.86–1.06)	1.10 (0.94–1.28)	0.82 (0.71–0.96)
	TT	794 (44.66)	984 (55.34)	0.94 (0.83–1.07)	1.08 (0.89–1.30)	0.80 (0.67–0.95)
rs12908960	GG/GA	3,029 (45.17)	3,677 (54.83)	1.00	1.00	0.05 (0.49)
	AA	495 (48.20)	532 (51.80)	1.09 (0.95–1.24)	0.99 (0.84–1.17)	1.30 (1.04–1.64)
rs11856927	TT	1,600 (44.17)	2,022 (55.83)	1.00	1.00	0.01 (0.19)
	TG	1,517 (46.75)	1,728 (53.25)	1.08 (0.98–1.19)	0.95 (0.82–1.10)	<b>1.20 (1.05–1.36)</b>
	GG	406 (47.15)	455 (52.85)	1.08 (0.93–1.26)	0.91 (0.75–1.12)	<b>1.35 (1.05–1.72)</b>
rs1961177	CC/CT	3,448 (45.68)	4,100 (54.32)	1.00	1.00	0.05 (0.49)
	TT	76 (41.08)	109 (58.92)	0.83 (0.61–1.11)	1.17 (0.74–1.86)	0.64 (0.43–0.96)
rs6493497	GG/GA	3,454 (45.66)	4,111 (54.34)	1.00	1.00	0.07 (0.51)
	AA	69 (41.32)	98 (58.68)	0.83 (0.61–1.13)	1.15 (0.72–1.84)	0.64 (0.42–0.99)

Bold text designates significant trend associations at the 0.05 level after adjustment for multiple comparisons

<sup>a</sup>Odds ratios (ORs) adjusted for age, study center, and Native American ancestry

<sup>b</sup>Odds ratios adjusted for age and study center

<sup>c</sup>Wald  $p$  value for 1  $f$  test for interaction between SNP\*Native American ancestry;  $p$  value adjusted for multiple comparisons (Bonferroni–Holm) in parentheses

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 4

Overall associations between selected variants in *CYP19A1* gene with breast cancer mortality, stratified by Native American ancestry, the Breast Cancer Health Disparities Study

Gene/SNP	Genotype	All women		Stratified by Native American ancestry				<i>p</i> <sup>c</sup>			
		Death	PY	HR <sup>a</sup> (95% CI)	Death	PY	HR <sup>b</sup> (95% CI)		Death	PY	HR <sup>b</sup> (95% CI)
rs12591359	GG/GA	164	16,035	1.00	92	9,761	1.00	72	6,274	1.00	0.03 (0.46)
	AA	44	3,492	1.21 (0.86–1.69)	28	1,885	1.70 (1.11–2.62)	16	1,607	0.77 (0.45–1.34)	
rs2414099	TT	147	12,914	1.00	94	8,144	1.00	53	4,770	1.00	0.11 (1.00)
	TC/CC	61	6,604	0.81 (0.60–1.10)	26	3,502	0.63 (0.41–0.98)	35	3,102	1.07 (0.69–1.65)	
rs2470152	TT/TC	151	15,031	1.00	86	8,750	1.00	65	6,281	1.00	0.58 (1.00)
	CC	57	4,496	1.40 (1.03–1.91)	34	2,896	1.37 (0.91–2.04)	23	1,599	1.51 (0.93–2.45)	
rs2445762	TT	104	8,520	1.00	73	5,719	1.00	31	2,800	1.00	0.18 (1.00)
	TC/CC	102	10,958	0.70 (0.53–0.93)	45	5,878	0.59 (0.40–0.86)	57	5,080	0.87 (0.55–1.36)	
rs8025191	AA	130	11,093	1.00	86	7,019	1.00	44	4,074	1.00	0.14 (1.00)
	AG/GG	78	8,423	0.78 (0.58–1.04)	34	4,627	0.65 (0.43–0.97)	44	3,796	0.97 (0.63–1.50)	

<sup>a</sup> Hazard ratios (HRs) adjusted for age, study center, stage of disease, and Native American ancestry

<sup>b</sup> Hazard ratios adjusted for age, study center, and stage of disease

<sup>c</sup> Wald *p* value for 1 *df* test; *p* value adjusted for multiple comparisons (Bonferroni–Holm) in parentheses