HHS Public Access

Author manuscript

Endocr Relat Cancer. Author manuscript; available in PMC 2015 December 01.

Published in final edited form as:

Endocr Relat Cancer. 2014 ; 21(6): 853–864. doi:10.1530/ERC-14-0250.

Variants in estrogen-related genes and breast cancer risk in European and African American women

Lei Quan[1,2], **Chi-Chen Hong**[1], **Gary Zirpoli**[1], **Michelle R. Roberts**[1], **Thaer Khoury**[1], **Lara E. Sucheston-Campbell**[1], **Dana H. Bovbjerg**[2], **Lina Jandorf**[3], **Karen Pawlish**[4], **Gregory Ciupak**[1], **Warren Davis**[1], **Elisa V. Bandera**[5], **Christine B. Ambrosone**[1], and **Song Yao**[1]

[1]Roswell Park Cancer Institute, Buffalo, NY

[3]Icahn School of Medicine at Mount Sinai, New York, NY

[4]New Jersey Department of Health, Trenton, NJ

[5]Rutgers Cancer Institute of New Jersey, New Brunswick, NJ

Abstract

Compared to women of European ancestry (EA), those of African ancestry (AA) are more likely to develop estrogen receptor (ER) negative breast cancer, although the mechanisms have not been elucidated. We tested the associations between breast cancer risk and a targeted set of 20 genes known to be involved in estrogen synthesis, metabolism and response and potential gene-environment interactions using data and samples from 1307 EA (658 cases) and 1365 AA (621 cases) participants from the Women's Circle of Health Study (WCHS). Multivariable logistic regression found evidence of associations with single nucleotide polymorphisms (SNPs) in the *ESR1* gene in EA women (rs1801132, OR=1.47, 95%CI=1.20–1.80, p=0.0002; rs2046210, OR=1.24, 95%CI=1.04–1.47, p=0.02; and rs3020314, OR=1.43, 95%CI=1.19–1.70, p=0.00009); but not in AA women. The only other gene associated with breast cancer risk was *CYP1A2* in AA women (rs2470893, OR=1.42, 95%CI=1.00–2.02, p=0.05), but not in EA women. When stratified by ER status, *ESR1* rs1801132, rs2046210 and rs3020314 showed stronger associations in ER positive than in ER negative breast cancer in only EA women. Associations with the *ESR1* SNPs in EA women also appeared to be stronger with longer endogenous estrogen exposure or hormonal replacement therapy (HRT) use. Our results indicate that there may be differential genetic influences on breast cancer risk in EA compared to AA women and these differences may be modified by tumor subtype and estrogen exposures. Future studies with a larger sample size may determine the full contribution of estrogen-related genes to racial/ethnic differences of breast cancer.

Address correspondence to: Song Yao, PhD, Roswell Park Cancer Institute, Elm and Carlton Streets, Buffalo, NY 14263, song.yao@roswellpark.org, Phone: 716-845-4968.

[2]Current affiliation: University of Pittsburgh Cancer Institute, Pittsburgh, PA;

Declaration of Interests

The authors have no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

Keywords

Breast Cancer; Disparity; African-American; Estrogen Receptor; Estrogen Synthesis; Estrogen Metabolism; Estrogen Response; ESR1

Introduction

Breast cancer is a heterogeneous disease. Distribution of tumor characteristics, such as the expression of estrogen receptor (ER), varies by age and ethnic/racial background (Ambrosone, et al. 2014a; Millikan, et al. 2008). Compared to American women of European ancestry (EA), those of African ancestry (AA) are more likely to be diagnosed with breast cancer before age 50, and to have tumors with more aggressive features, such as ER negative status (Amend, et al. 2006). These tumors do not respond to anti-estrogen therapy and typically have poorer clinical outcomes (Chlebowski, et al. 2005). Although the mechanisms underlying such disparities are not yet known, it has been reported that breast cancer racial differences persist after adjustment for socioeconomic status and lifestyle factors (Amend et al. 2006). A high proportion of ER negative breast cancer has also been shown in indigenous African women (Fregene and Newman 2005; Huo, et al. 2009), suggesting that genetic factors related to African ancestry may, in part, account for the higher proportion of more aggressive breast cancer in AA women. Because populations of African ancestry typically exhibit shorter linkage disequilibrium (LD) blocks across the genome compared to populations of European ancestry (Haiman and Stram 2010; Hinch, et al. 2011), genetic studies in AAs may not only illustrate mechanisms underlying racial disparities of breast cancer, but also facilitate the identification of casual genetic loci (O'Brien, et al. 2014).

High levels of circulating estrogens and their metabolites, which have recently been shown to correlate with those in the local breast tissue (Loud, et al. 2014), are associated with increased risk of breast cancer among postmenopausal women (Fuhrman, et al. 2012; Key, et al. 2002), likely due to the dual effects of estrogens to stimulate cell proliferation and gene expression through binding with estrogen receptors and to cause DNA damage via mutagenic estrogen metabolites such as 2-OH and 4-OH catechol estrogens (Liehr 1997; Pike, et al. 1993). Because AA women have elevated serum concentrations of estrogens (Pinheiro, et al. 2005; Setiawan, et al. 2006), as well as different profiles of hormone-related factors, such as age of menopause, age of menarche, breast feeding, parity and HRT use, as compared to EA women (Brett and Madans 1997; Millikan et al. 2008), it is postulated that breast cancer racial disparities may be attributed, in part, to estrogens and related hormonal factor. This postulation is supported in several recent studies, including those from our group (Ambrosone et al. 2014a; Ambrosone, et al. 2014b; Millikan et al. 2008; Palmer, et al. 2011).

Genetic variations involved in estrogen biosynthesis, metabolism and response pathways may contribute to a woman's lifetime estrogen exposure, and thus influence the risk of breast cancer (Thompson and Ambrosone 2000). A number of candidate association studies have examined variants in estrogen-related genes with breast cancer mostly in EAs, while

similar studies in AAs are still sparse (Kato, et al. 2009; Rebbeck, et al. 2007; Taioli, et al. 1995; Van Emburgh, et al. 2008). A SNP in estrogen receptor alpha (*ESR1*) gene, rs2046210, was associated with breast cancer risk in genome-wide association studies (GWAS) first in an Asian population (Zheng, et al. 2009b) and subsequently in populations of the European descendent (Michailidou, et al. 2013). However, it remains unclear whether *ESR1* rs2046210 is also associated with risk of breast cancer in AAs (Hutter, et al. 2011; Long, et al. 2013; O'Brien et al. 2014). Furthermore, it is possible that gene-environment interactions within the hormone-related pathways may contribute to racial disparities of breast cancer; nevertheless, even fewer studies have examined these interactions in AAs (Reding, et al. 2012).

In a large case-control study with a similar number of AA and EA women, we systematically examined associations between selected genetic variants in estrogen biosynthesis, metabolism and response pathways and risk of breast cancer, overall and by ER status, and their potential interactions with estrogen-related factors.

Materials and Methods

Study Participants

The Women's Circle of Health Study (WCHS) is a case-control study designed to evaluate risk factors for aggressive breast cancer in AA women. The study was conducted in the metropolitan area of New York City between 2003 and 2008 and in New Jersey (NJ) between 2003 and 2012, and has been previously described in detail (Ambrosone, et al. 2009; Yao, et al. 2012). Eligible participants included English-speaking AA and EA women age 20 to 75 years, diagnosed with incident, primary, histologically confirmed breast cancer and had no previous history of cancer other than non-melanoma skin cancer. Controls without a history of any cancer diagnosis other than non-melanoma skin cancer were identified by random-digit dialing (RDD) and matched to cases on race and 5-year age group. Enrollment of AA controls in NJ was supplemented with community-based recruitment, as a combination of RDD and community controls was shown to be more representative of the general population in NJ (Bandera, et al. 2013). Controls were recruited and interviewed using the same standardized method and during the same time period as the cases at both sites.

Overall, the participation rate for those who were contacted and eligible was 79% and 49% in EA cases and controls, respectively, and 79% and 48% in AA cases and controls, respectively. A total of 658 EA cases, 715 EA controls, 621 AA cases, and 744 AA controls from the WCHS were available at the time of genotyping and thus included in the study. This study was approved by the Institutional Review Boards at Roswell Park Cancer Institute (RPCI), Rutgers Cancer Institute of New Jersey (CINJ), the Icahn School of Medicine at Mount Sinai, and participating hospitals in New York City, functioning according to the 3rd edition of the Guidelines on the Practice of Ethical Committees in Medical Research issued by the Royal College of Physicians of London.

Data Collection

In-depth in-person interviews were conducted to collect information on a variety of factors known or suspected to affect breast cancer risk, including demographics, reproductive factors, medical history, family history of cancer, lifestyle factors, as well as biospecimens. Anthropometric measures were also collected during the interview. Pathology reports were obtained from hospitals at which patients were diagnosed. Information on ER status was available for 468 EA cases (82 ER negative) and 473 AA cases (150 ER negative). Informed consent, including permission to obtain pathology data and tumor tissue blocks, was obtained from each participant.

Sample Collection and Genotyping

Initially, blood samples were collected from study participants. We later transitioned to non-invasive collection of saliva for DNA extraction. Genomic DNA was extracted in batches from whole blood using the FlexiGene DNA protocol (Qiagen Inc, Valencia, CA, US) and from saliva using the Oragene protocol (DNA Genotek Inc., Ottawa, ON, Canada). Quality and quantity of purified DNA were evaluated using Nanodrop UV-spectrometer (Thermo Fisher Scientific Inc., Wilmington, DE, US) and PicoGreen–based fluorometric assays (Invitrogen Inc., Carslbad, CA, US). DNA samples were stored at −80°C until analysis.

We included in our analysis all major genes involved in the estrogen biosynthesis, metabolism and response pathways, including cytochrome-dependent monooxygenase (CYP) genes, hydroxysteroid dehydrogenase (HSD) genes, Catechol-O-methyltransferase (COMT), UDP-glucuronosyltransferase (UGT) genes, and also estrogen receptor (ESR) genes (Suppl. Figure 1) (Germain 2011; Liang and Shang 2013; Yager and Davidson 2006). We then surveyed the Human Genome Epidemiology (HuGE) Navigator for the selected genes to identify SNPs within these genes that were previously associated with risk of any cancer or cancer outcome, with a focus on SNPs that were previously shown to be functional (Yu, et al. 2008). Genomic DNA was plated and genotyped at the Genomics Core Facility at RPCI using MassARRAY technology and iPLEX Gold Assay (Sequenom Inc., San Diego, CA, US). Five percent duplicates and two sets of in-house trio samples of European and African ancestry were included for quality control purposes. The concordance among blind duplicate pairs was >99.9%. The average successful genotyping rate for each sample and SNP was 95.9%. Samples or SNPs with call rate <90% were excluded, as were monomorphic SNPs or SNPs with an MAF<5% in both AA and EA controls. For each SNP, Hardy-Weinberg equilibrium was assessed among controls. SNPs deviating from Hardy-Weinberg equilibrium (HWE) in both EA and AA controls were excluded. Clustering plots of SNPs that were significant in the statistical analysis were manually re-inspected *post-hoc* to ensure that the calls for genotype of each sample were robust. We selected 62 SNPs for the assay design: 5 of them failed in the assay design and validation process, 3 were excluded due to low call rate, 3 were monomorphic or rare and excluded, and another 2 were excluded for violating HWE, leaving a total of 49 SNPs in 20 genes included in our analyses (Suppl. Table1). The major allele of each SNP in EA controls is defined as the reference allele and the minor allele as the coded allele. To account for the potential inaccuracy of self-reported race/ethnicity and to assess ancestry quantitatively, all DNA samples were also

genotyped for a panel of 100 ancestry informative markers (AIMs) (Ruiz-Narvaez, et al. 2011) using the Illumina GoldenGate targeted multiplex assay.

Statistical Analysis

All analyses were conducted using SAS 9.4 (SAS Institute, Cary CA) separately for EA and AA women, according to self-reported race. Descriptive variables were compared between cases and controls using Chi-square tests for categorical variables and Wilcoxon rank-sum test for continuous variables. Proportions of European and African ancestry in individual EA and AA woman were estimated quantitatively based on AIM genotypes using the Bayesian Markov Chain Monte Carlo clustering algorithm implemented in STRUCTURE 2.3 (Pritchard, et al. 2000). Since the sum of two ancestral proportions in each individual is always one, we used only the proportion of European Ancestry in all analyses. Genotype frequencies of each SNP were compared between EA and AA controls using Chi-square test or Fisher's Exact test where appropriate. To compare allele frequencies obtained from our study to those previously reported, frequencies for Caucasians (CEU), African Americans of the American Southwest (ASW) and Yoruban in Ibadan, Nigeria (YRI) were obtained from HapMap release #28 (phase 1, 2 and 3 merged).

Odds ratios (ORs) and 95% confidence intervals (CIs) for each SNP were derived from multivariable logistic regression models, assuming codominant or additive genetic models, with adjustment for known risk factors of breast cancer, overall and by ER subtypes or estrogen exposure status (Althuis, et al. 2004; Ambrosone et al. 2014b; Gordon 1995; Hwang, et al. 2005; Krieger, et al. 2008): including age at diagnosis, education, family history of breast cancer, history of benign breast disease, menopausal status, number of full-term pregnancies, breast feeding, hormone replacement therapy (HRT), body mass index (BMI), proportion of European ancestry and "estrogen months". The variable "estrogen months", reflecting total months of exposure to endogenous estrogen, is defined as following: "months between age at enrollment and age at menarche" ("months between age of menopause and age at menarche" in postmenopausal women) minus "months of pregnancy", then minus "months of breastfeeding". *P* for linear trend was calculated by coding SNPs as 0, 1, 2 and testing whether there was a linear dose-response effect of the variant allele as an ordinal variable. *P* values were adjusted for multiple comparisons using a modified false discovery rate (FDR) method, in which the critical value is determined by $\alpha/\sum_{i=1}^{k}(1/i)$ (Narum 2006).

Results

Participant Characteristics

Participant characteristics are shown in Table 1. Among EA women, cases were significantly more likely than controls to have a family history of breast cancer in a first-degree relative, less likely to have breastfed or to have attended college or graduate school. AA cases were less highly educated than controls, although differences were not statistically significant. Among both EA and AA women, cases were more likely to have a history of benign breast disease and to have longer exposure to endogenous estrogen before

menopause. Cases were also slightly older than controls but within 5-year age range in both EA and AA women. Cases did not differ significantly from controls in body mass index (BMI), HRT use in postmenopausal women, or number of full-term pregnancy in either group.

Breast Cancer Risk in EA and AA Women

For 41 of the 49 SNPs analyzed (84%), genotype frequencies differed significantly between AA and EA controls (Supplementary Table 1, $p < 0.05$ after correction for multiple testing). Sixteen SNPs had 'flipped' minor alleles, where the minor allele among EA controls was the major allele among AA controls, and vice versa (Supplementary Table 1, rs2046210 and rs3020314 in Table 2 for example). Genotype frequencies obtained from the HapMap database for each ancestry were very similar to those in our study (data not shown). Because of the notable differences in allele distributions between EA and AA women, all analyses were stratified by self-reported race.

ORs and 95% CIs for associations between overall breast cancer risk and all 49 SNPs analyzed are shown in Supplementary Table 2. Significant associations in either EA or AA women with a *p* for trend < 0.05 before correction for multiple testing are shown in Table 2. We found different associations in EA and AA women. Three SNPs in the *ESR1* gene, i.e. rs1801132, rs2046210 and rs3020314, were associated with an increased risk of breast cancer in EA women (Table 2, OR=1.47, 95% CI= 1.20–1.80, p=0.0002, p_c=0.001; OR=1.24, 95% CI= 1.04–1.47, p=0.02; and OR=1.43, 95% CI= 1.19–1.70, p=0.00009, p_c=0.0004; respectively). The three SNPs were not in linkage delirium (LD) with each other in the EA control subjects. In addition, SNP rs2470893 in *CYP1A2* was the only one marginally associated with an increased breast cancer risk in AA women (Table 2, OR=1.42, 95% CI= 1.00–2.02, p=0.05).

Risk for Breast Cancer by ER Status

As expected, ER negative breast cancer was more common among AA (31.7%) than among EA women (17.5%). We next examined whether the associations between estrogen-related genes and breast cancer risk by ER status also differ in EA and AA women. Number of subjects, ORs and 95% CIs of ER positive or ER negative breast cancer for all 49 SNPs are shown in Supplementary Table 3. Significant associations in either EA or AA women with a *p* for trend <0.05 before correction for multiple testing are shown in Table 3. All four significant SNPs identified in the overall analysis remained significant. The three SNPs in *ESR1* gene, i.e. rs1801132, rs2046210 and rs3020314, were associated with an increased risk of both ER positive and ER negative breast cancer in EA women, although the OR estimates for ER negative breast cancer were slightly lower and only significant for rs3020314 (Table 3). SNP rs2470893 in *CYP1A2* was primarily associated with an increased risk of ER positive breast cancer in AA women (Table 3). We also found three new associations: *CYP17A1* rs12413409 was associated with an increased risk of ER negative breast cancer in EA women (Table 3, OR=1.92, 95% CI=1.15–3.21, p=0.01); *HSD17B2* rs4445895 was associated with a decreased risk of ER negative breast cancer in EA women (Table 3, OR=0.62, 95% CI=0.43–0.90, p=0.01), but an increased risk of ER positive breast cancer in AA women (Table 3, OR=1.24, 95% CI=1.01–1.52, p=0.04); *UGT1A9* rs6714486

was marginally associated with a decreased risk of ER positive breast cancer in AA women (Table 3, OR=0.75, 95% CI=0.57–0.98, p=0.04).

Stratification by Estrogen Months and HRT Use

It is plausible that gene-environment interactions within estrogen-related pathways may also contribute to the ethnic/racial differences in breast cancer characteristics. We therefore stratified the data by length of lifetime exposure to endogenous estrogen (longer or shorter than 384 estrogen months - the median in both AA and EA controls) or usage of HRT (yes or no). Number of subjects, ORs and 95% CIs of ER positive or ER negative breast cancer for all 49 SNPs are shown in Supplementary Tables 4 & 5. Estrogen month and usage of HRT were not independently associated with breast cancer risk in either EA or AA women (data not shown). However, the three SNPs in the *ESR1* gene, i.e. rs1801132, rs2046210 and rs3020314, were associated with an increased risk of breast cancer in EA women with longer estrogen exposure, with smaller effects in EA women with shorter estrogen exposure (Table 4). Similar results were found when using race-specific medians as cutoffs (Supplementary Table 6). When stratified by HRT status in postmenopausal women, rs1801132 and rs3020314 in *ESR1* were associated with an increased risk of breast cancers in EA women with HRT use (Table 5). *ESR*1 rs2046210, however, was associated with an increased risk of breast cancer only in non-user postmenopausal EA women (Table 5).

Discussion

Disparities in breast cancer biology are evident between EA and AA women and may be due, in part, to differences in genetic background (Quan, et al. 2014; Yao et al. 2012). Estrogens are synthesized from cholesterol through a cascade of enzymatic reactions (Tang, et al. 2011; Thompson and Ambrosone 2000). The parent estrogens, i.e. estrone and estradiol, can be irreversibly hydroxylated at the 2, 4 and 16 positions of the steroid ring to produce estrogen metabolites that have different affinity for the estrogen receptors, which may lead to differential risk of breast cancer (Zhu, et al. 2006). Moreover, both parent estrogens and estrogen metabolites can be modified by a series of enzymes that influences their bioavailability to breast tissue (Raftogianis, et al. 2000). In this study, we systematically examined associations between genes involved in estrogen biosynthesis, metabolism and response pathways (Suppl. Figure 1) and overall breast cancer risk, as well as by ER status and by estrogen exposure, in EA and AA women.

The primary study findings revealed differences between EA and AA women in the genetic architecture of selected genes in estrogen biosynthesis, metabolism and response pathways. Among the 49 SNPs in 20 genes examined, the allele frequencies of 41 SNPs (84%) differed significantly between EA and AA controls (Supplementary Table 1). Subsequently, we found differential relationships between estrogen-related genes and breast cancer risk in EA and AA women, in overall analysis, ER subtype analyses, and stratified analyses by estrogen exposures.

We found strong associations between three SNPs in the *ESR1* gene and overall breast cancer risk in only EA women. The human *ESR1* gene has eight exons that span 300kb on chromosome 6q25.1 and encodes the alpha form of the estrogen receptor (ERα). The three

SNPs, i.e. rs1801132, rs2046210 and rs3020314, were not in linkage disequilibrium (LD) with each other. Rs1801132 is a synonymous SNP at codon 325, which may influence mRNA stability and translation efficiency (Li, et al. 2010; Sauna, et al. 2007). The G allele of rs1801132 was associated with a decreased risk of breast cancer in a meta-analysis of 5649 cases and 6856 controls, the majority of whom were Hispanic and Caucasians (CG/GG vs CC, OR=0.92, 95% CI=0.85–0.99) (Li et al. 2010). Rs2046210 is a SNP located upstream of *ESR1* that was reported to be associated with breast cancer risk in a three-stage genome-wide association study (GWAS) in Asians (AA vs GG, OR=1.59, 95% CI=1.40–1.82) (Zheng et al. 2009b) and subsequently in the European population (Michailidou et al. 2013). SNP rs3020314 tags a highly conserved region of *ESR1* intron 4 and may change the ratio of two mRNA splice forms (Dunning, et al. 2009; Fuqua and Wolf 1995). The G allele of rs3020314 was associated with an increased risk of breast cancer in a recent GWAS (OR=1.05, 95% CI=1.02–1.09) (Dunning et al. 2009). We found associations between breast cancer risk and all three SNPs in the same directions as previous reports in EA women, but no associations in AA women (Table 2), even after stratifications by ER status or by estrogen exposure status (Tables 3, 4 & 5). Previously, efforts to evaluate breast cancer GWAS variants in AA women have failed to replicate most variants discovered in European and Asian populations, including *ESR1* rs2046210 (Campa, et al. 2011; Chen, et al. 2011; Hutter et al. 2011; Long et al. 2013; Stacey, et al. 2010; Zheng, et al. 2009a). However, a recently study using a Bayesian Approach replicated the association between breast cancer risk and *ESR1* rs2046210 in AA women but not EA women (Hutter et al. 2011; Long et al. 2013; O'Brien et al. 2014). Future studies are required to understand the detailed function of this SNP in different populations.

Estrogens are metabolized by a number cytochrome P450 enzymes. Previously, CYP enzymes such as CYP1A1 and CYP1B1 have been associated with breast cancer risk among AA women (Kato et al. 2009; Reding et al. 2012). As far as we know, we are the first to report a positive association between the A allele of *CYP1A2* rs2470893 and breast cancer risk, which was only observed in AA women (Table 2). *CYP1A2* rs2470893 is located in the bidirectional promoter region of the *CYP1A1-CYP1A2* locus on chromosome 15q24. In addition to estrogen, the CYP1A2 enzyme metabolizes many exogenous compounds such as caffeine, insulin, and blood lipids, and may have complex effects on the risk of breast cancer (Hong, et al. 2004).

The associations between estrogen-related genes and risk of breast cancers by ER status have not been widely investigated. Among the SNPs associated with overall breast cancer risk, *ESR1* rs1801132, rs2046210 and rs3020314 were associated with an increased risk of both ER positive and ER negative breast cancer in EA women, with stronger effects on ER positive tumors (Table 3). Such results indicated a more general effect of *ESR1* regarding ER status of breast cancer. Indeed, a recent genome-wide association study reported that *ESR1* rs2046210 are associated with risk of ER negative and triple negative (TN) breast cancers (Purrington, et al. 2014). In our ER subtype analyses, we also identified three other associations that were not observed for overall risk (Table 3). *CYP17A1* rs12413409 was with the strongest effect, associated with an almost two-fold increase in the risk of ER negative breast cancer in EA women (Table 3). CYP17 converts pregnenolone into

dehydroepiandrosterone and progesterone into 17α-hydroxyprogesterone, both are essential precursors of estrogens (Feigelson, et al. 1998). Previously, *CYP17A1* have been associated with breast cancer risk among younger women (Bergman-Jungestrom, et al. 1999; Spurdle, et al. 2000) and among subgroup of cases with more advanced breast cancer (Feigelson, et al. 1997). However, other studies have found null association between breast cancer risk and *CYP17A1* (Miyoshi and Noguchi 2003). The association between *CYP17A1* rs12413409 and ER negative breast cancer in EA women detected in the current study warrants future validation (Table 3).

The origins of ER negative breast cancer have been under debate (Santen and Allred 2007). One hypothesis suggests that ER negative breast cancer evolves from ER negative precursor cells. This view is supported by the fact that tamoxifen prevents only ER positive breast cancer in women of high risk (Sims, et al. 2007). In addition, there is evidence that ER negative stem cells are essential for normal breast development and gives rise to more differentiated ER-positive progenitor cells (Asselin-Labat, et al. 2006). An alternative hypothesis suggests that there are multiple mechanisms for the development of ER negative breast cancer, including that many may rise from ER positive precursors. This concept is supported by the fact that increased estrogen exposure is associated with an increased incidence of all breast cancers and that the proportion of ER negative breast cancer increases along with time of tumor progression (Santen and Allred 2007). Our findings that SNPs in the *ESR1* gene were associated with an increased risk of both ER positive and ER negative breast cancers in EA women (Table 3), may provide further information on this issue.

Genotype alone is an incomplete surrogate for gene functions, which may interacts with other endogenous or exogenous factors (Thompson and Ambrosone 2000). We further stratified our analyses by the length of endogenous estrogen exposure or usage of HRT. We found associations with the three *ESR1* SNPs, which is plausible since the *ESR1* gene encodes effector molecules for estrogen and their metabolites. The associations appeared to be more pronounced in women with longer estrogen months or with HRT use, with the exception for rs2046210 among HRT non-users, and were observed only among EA women (Table 4 and 5).

Our study has several strengths. We conducted in-person interviews to assure data accuracy. In-depth information on medical history, family history of cancer, hormone-related, and lifestyle factors were collected, allowing us to adjust for potential confounders. We included in the study a large number of cases and controls, both EA and AA, enabling us to stratify by ancestry group and address racial disparities. To account for the high levels of genetic diversity within self-reported race group, we also estimated the percentage of European ancestry (or African ancestry) using a set of AIMs (Ruiz-Narvaez et al. 2011). Data on ER status was available for a large proportion of cases, so we were able to identify SNPs specifically associated with ER positive or ER negative breast cancers.

We are aware that SNPs identified as related to breast cancer in our study may not be causal. It is possible that they are in LD with other SNPs that are in a causal pathway but were not tested. We showed all SNPs that were significant before correction for multiple comparisons because our primary goal was to identify promising and biologically relevant associations,

which may motivate future studies for replication and extension. P values adjusted for multiple comparisons were also provided in the tables. The response rates among EA and AA control subjects were relatively low (49% and 48.2%, respectively) in our study and could lead to potential selection bias. However, an approximately 50% response rate was not uncommon in population studies (Hartge 2006). ER status was missing for 190 cases (28.9%) among EA women and 148 cases (23.8%) among AA women in our study. The missing rate in our study is comparable or lower than in other similar studies (Long et al. 2013; Palmer, et al. 2013). We are aware of potential impact of this substantial data missing on our analysis. It is possible that ER status may not be missing completely at random. Some studies have suggested ER status is more frequently missing among women of color or of lower socioeconomic status, likely due to inadequacies of medical care (Krieger N. et al, 2008). However, in our study, ER status was missing at a similar rate between EA and AA women; and there were no obvious differences in missing ER status crossing various levels of education among either EA or AA women in our study (data not shown). It is thus likely the missing is non-differential and would have limited impact on the analysis besides lowering statistical power.

In conclusion, the present study revealed genetic variants and potential gene-environment interactions within the estrogen biosynthesis, metabolism and response pathways associated with breast cancer risk in EA and AA women. The associations appeared to differ between EA and AA women, and this is true after stratification by ER status or estrogen exposure. Additional research is now warranted to validate these initial findings in a larger sample size and to assess their utility in predicting breast cancer risk or treatment response for anti-estrogen prevention in women at high risk.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding

This work was supported by grants from the U.S. Army Medical Research and Material Command (USARMMC) (DAMD-17-01-1-0334 and W81XWH-08-1-0379), the National Cancer Institute (P01 CA151135, R01 CA100598, R01 CA136483, and R25 CA114101), the Breast Cancer Research Foundation (BCRF), and a gift from Philip L. Hubbell family. Samples were stored and managed by the RPCI DataBank and BioRepository (DBBR) and genotyping was performed in the RPCI Genomics Core Facility, both CCSG shared resources, supported by P30 CA016056-32. The New Jersey State Cancer Registry (NJSCR) is a participant in the Centers for Disease Control and Prevention's National Program of Cancer Registries (NPCR) and is a National Cancer Institute SEER Expansion Registry. The NJSCR is supported by the Centers for Disease Control and Prevention under cooperative agreement 1U58DP00039311-01 awarded to the New Jersey Department of Health. The collection of State of New Jersey cancer incidence data is also supported by the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) Program under contract N01-PC-2010-00027 and the State of New Jersey. The funding agents played no role in design, in the collection, analysis, and interpretation of data, in the writing of the manuscript, or in the decision to submit the manuscript for publication.

References

Althuis MD, Fergenbaum JH, Garcia-Closas M, Brinton LA, Madigan MP, Sherman ME. Etiology of hormone receptor-defined breast cancer: a systematic review of the literature. Cancer Epidemiol Biomarkers Prev. 2004; 13:1558–1568. [PubMed: 15466970]

Ambrosone CB, Ciupak GL, Bandera EV, Jandorf L, Bovbjerg DH, Zirpoli G, Pawlish K, Godbold J, Furberg H, Fatone A, et al. Conducting Molecular Epidemiological Research in the Age of HIPAA: A Multi-Institutional Case-Control Study of Breast Cancer in African-American and European-American Women. J Oncol. 2009; 2009:871250. [PubMed: 19865486]

Ambrosone CB, Zirpoli G, Ruszczyk M, Shankar J, Hong CC, McIlwain D, Roberts M, Yao S, McCann SE, Ciupak G, et al. Parity and breastfeeding among African-American women: differential effects on breast cancer risk by estrogen receptor status in the Women's Circle of Health Study. Cancer Causes Control. 2014a; 25:259–265. [PubMed: 24249438]

Ambrosone CB, Zirpoli GR, Bovbjerg DH, Shankar J, Hong CC, McCann SE, Ruszczyk M, Khoury T, Yao S, Ciupak GL, et al. Associations between estrogen receptor-negative breast cancer and timing of reproductive events differ between African American and European American women. Cancer Epidemiol Biomarkers Prev. 2014b; 23:1115–1120. [PubMed: 24718280]

Amend K, Hicks D, Ambrosone CB. Breast cancer in African-American women: differences in tumor biology from European-American women. Cancer Res. 2006; 66:8327–8330. [PubMed: 16951137]

Asselin-Labat ML, Shackleton M, Stingl J, Vaillant F, Forrest NC, Eaves CJ, Visvader JE, Lindeman GJ. Steroid hormone receptor status of mouse mammary stem cells. J Natl Cancer Inst. 2006; 98:1011–1014. [PubMed: 16849684]

Bandera EV, Chandran U, Zirpoli G, McCann SE, Ciupak G, Ambrosone CB. Rethinking sources of representative controls for the conduct of case-control studies in minority populations. BMC Med Res Methodol. 2013; 13:71. [PubMed: 23721229]

Bergman-Jungestrom M, Gentile M, Lundin AC, Wingren S. Association between CYP17 gene polymorphism and risk of breast cancer in young women. Int J Cancer. 1999; 84:350–353. [PubMed: 10404084]

Brett KM, Madans JH. Use of postmenopausal hormone replacement therapy: estimates from a nationally representative cohort study. Am J Epidemiol. 1997; 145:536–545. [PubMed: 9063344]

Campa D, Kaaks R, Le Marchand L, Haiman CA, Travis RC, Berg CD, Buring JE, Chanock SJ, Diver WR, Dostal L, et al. Interactions between genetic variants and breast cancer risk factors in the breast and prostate cancer cohort consortium. J Natl Cancer Inst. 2011; 103:1252–1263. [PubMed: 21791674]

Chen F, Chen GK, Millikan RC, John EM, Ambrosone CB, Bernstein L, Zheng W, Hu JJ, Ziegler RG, Deming SL, et al. Fine-mapping of breast cancer susceptibility loci characterizes genetic risk in African Americans. Hum Mol Genet. 2011; 20:4491–4503. [PubMed: 21852243]

Chlebowski RT, Chen Z, Anderson GL, Rohan T, Aragaki A, Lane D, Dolan NC, Paskett ED, McTiernan A, Hubbell FA, et al. Ethnicity and breast cancer: factors influencing differences in incidence and outcome. J Natl Cancer Inst. 2005; 97:439–448. [PubMed: 15770008]

Dunning AM, Healey CS, Baynes C, Maia AT, Scollen S, Vega A, Rodriguez R, Barbosa-Morais NL, Ponder BA, Search, et al. Association of ESR1 gene tagging SNPs with breast cancer risk. Hum Mol Genet. 2009; 18:1131–1139. [PubMed: 19126777]

Feigelson HS, Coetzee GA, Kolonel LN, Ross RK, Henderson BE. A polymorphism in the CYP17 gene increases the risk of breast cancer. Cancer Res. 1997; 57:1063–1065. [PubMed: 9067272]

Feigelson HS, Shames LS, Pike MC, Coetzee GA, Stanczyk FZ, Henderson BE. Cytochrome P450c17alpha gene (CYP17) polymorphism is associated with serum estrogen and progesterone concentrations. Cancer Res. 1998; 58:585–587. [PubMed: 9485002]

Fregene A, Newman LA. Breast cancer in sub-Saharan Africa: how does it relate to breast cancer in African-American women? Cancer. 2005; 103:1540–1550. [PubMed: 15768434]

Fuhrman BJ, Schairer C, Gail MH, Boyd-Morin J, Xu X, Sue LY, Buys SS, Isaacs C, Keefer LK, Veenstra TD, et al. Estrogen metabolism and risk of breast cancer in postmenopausal women. J Natl Cancer Inst. 2012; 104:326–339. [PubMed: 22232133]

Fuqua SA, Wolf DM. Molecular aspects of estrogen receptor variants in breast cancer. Breast Cancer Res Treat. 1995; 35:233–241. [PubMed: 7579493]

Germain D. Estrogen carcinogenesis in breast cancer. Endocrinol Metab Clin North Am. 2011; 40:473–484. vii. [PubMed: 21889715]

Gordon NH. Association of education and income with estrogen receptor status in primary breast cancer. Am J Epidemiol. 1995; 142:796–803. [PubMed: 7572955]

Haiman CA, Stram DO. Exploring genetic susceptibility to cancer in diverse populations. Curr Opin Genet Dev. 2010; 20:330–335. [PubMed: 20359883]

Hartge P. Participation in population studies. Epidemiology. 2006; 17:252–254. [PubMed: 16617271]

Hinch AG, Tandon A, Patterson N, Song Y, Rohland N, Palmer CD, Chen GK, Wang K, Buxbaum SG, Akylbekova EL, et al. The landscape of recombination in African Americans. Nature. 2011; 476:170–175. [PubMed: 21775986]

Hong CC, Tang BK, Hammond GL, Tritchler D, Yaffe M, Boyd NF. Cytochrome P450 1A2 (CYP1A2) activity and risk factors for breast cancer: a cross-sectional study. Breast Cancer Res. 2004; 6:R352–365. [PubMed: 15217502]

Huo D, Ikpatt F, Khramtsov A, Dangou JM, Nanda R, Dignam J, Zhang B, Grushko T, Zhang C, Oluwasola O, et al. Population differences in breast cancer: survey in indigenous African women reveals over-representation of triple-negative breast cancer. J Clin Oncol. 2009; 27:4515–4521. [PubMed: 19704069]

Hutter CM, Young AM, Ochs-Balcom HM, Carty CL, Wang T, Chen CT, Rohan TE, Kooperberg C, Peters U. Replication of breast cancer GWAS susceptibility loci in the Women's Health Initiative African American SHARe Study. Cancer Epidemiol Biomarkers Prev. 2011; 20:1950–1959. [PubMed: 21795501]

Hwang ES, Chew T, Shiboski S, Farren G, Benz CC, Wrensch M. Risk factors for estrogen receptor-positive breast cancer. Arch Surg. 2005; 140:58–62. [PubMed: 15655207]

Kato I, Cichon M, Yee CL, Land S, Korczak JF. African American-preponderant single nucleotide polymorphisms (SNPs) and risk of breast cancer. Cancer Epidemiol. 2009; 33:24–30. [PubMed: 19679043]

Key T, Appleby P, Barnes I, Reeves G, Endogenous H. Breast Cancer Collaborative G. Endogenous sex hormones and breast cancer in postmenopausal women: reanalysis of nine prospective studies. J Natl Cancer Inst. 2002; 94:606–616. [PubMed: 11959894]

Krieger N, Chen JT, Ware JH, Kaddour A. Race/ethnicity and breast cancer estrogen receptor status: impact of class, missing data, and modeling assumptions. Cancer Causes Control. 2008; 19:1305–1318. [PubMed: 18704721]

Li N, Dong J, Hu Z, Shen H, Dai M. Potentially functional polymorphisms in ESR1 and breast cancer risk: a meta-analysis. Breast Cancer Res Treat. 2010; 121:177–184. [PubMed: 19760036]

Liang J, Shang Y. Estrogen and cancer. Annu Rev Physiol. 2013; 75:225–240. [PubMed: 23043248]

Liehr JG. Dual role of oestrogens as hormones and pro-carcinogens: tumour initiation by metabolic activation of oestrogens. Eur J Cancer Prev. 1997; 6:3–10. [PubMed: 9161806]

Long J, Zhang B, Signorello LB, Cai Q, Deming-Halverson S, Shrubsole MJ, Sanderson M, Dennis J, Michailiou K, Easton DF, et al. Evaluating genome-wide association study-identified breast cancer risk variants in African-American women. PLoS One. 2013; 8:e58350. [PubMed: 23593120]

Loud JT, Gierach GL, Veenstra TD, Falk RT, Nichols K, Guttmann A, Xu X, Greene MH, Gail MH. Circulating estrogens and estrogens within the breast among postmenopausal BRCA1/2 mutation carriers. Breast Cancer Res Treat. 2014; 143:517–529. [PubMed: 24442642]

Michailidou K, Hall P, Gonzalez-Neira A, Ghoussaini M, Dennis J, Milne RL, Schmidt MK, Chang-Claude J, Bojesen SE, Bolla MK, et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. Nat Genet. 2013; 45:353–361. 361e351–352. [PubMed: 23535729]

Millikan RC, Newman B, Tse CK, Moorman PG, Conway K, Dressler LG, Smith LV, Labbok MH, Geradts J, Bensen JT, et al. Epidemiology of basal-like breast cancer. Breast Cancer Res Treat. 2008; 109:123–139. [PubMed: 17578664]

Miyoshi Y, Noguchi S. Polymorphisms of estrogen synthesizing and metabolizing genes and breast cancer risk in Japanese women. Biomed Pharmacother. 2003; 57:471–481. [PubMed: 14637391]

Narum SR. Beyond Bonferroni: Less conservative analyses for conservation genetics. Conservation Genetics. 2006; 7:783–787.

O'Brien KM, Cole SR, Poole C, Bensen JT, Herring AH, Engel LS, Millikan RC. Replication of breast cancer susceptibility loci in whites and African Americans using a Bayesian approach. Am J Epidemiol. 2014; 179:382–394. [PubMed: 24218030]

Palmer JR, Boggs DA, Wise LA, Ambrosone CB, Adams-Campbell LL, Rosenberg L. Parity and lactation in relation to estrogen receptor negative breast cancer in African American women. Cancer Epidemiol Biomarkers Prev. 2011; 20:1883–1891. [PubMed: 21846820]

Palmer JR, Ruiz-Narvaez EA, Rotimi CN, Cupples LA, Cozier YC, Adams-Campbell LL, Rosenberg L. Genetic susceptibility loci for subtypes of breast cancer in an African American population. Cancer Epidemiol Biomarkers Prev. 2013; 22:127–134. [PubMed: 23136140]

Pike MC, Spicer DV, Dahmoush L, Press MF. Estrogens, progestogens, normal breast cell proliferation, and breast cancer risk. Epidemiol Rev. 1993; 15:17–35. [PubMed: 8405201]

Pinheiro SP, Holmes MD, Pollak MN, Barbieri RL, Hankinson SE. Racial differences in premenopausal endogenous hormones. Cancer Epidemiol Biomarkers Prev. 2005; 14:2147–2153. [PubMed: 16172224]

Pritchard JK, Stephens M, Rosenberg NA, Donnelly P. Association mapping in structured populations. Am J Hum Genet. 2000; 67:170–181. [PubMed: 10827107]

Purrington KS, Slager S, Eccles D, Yannoukakos D, Fasching PA, Miron P, Carpenter J, Chang-Claude J, Martin NG, Montgomery GW, et al. Genome-wide association study identifies 25 known breast cancer susceptibility loci as risk factors for triple-negative breast cancer. Carcinogenesis. 2014; 35:1012–1019. [PubMed: 24325915]

Quan L, Gong Z, Yao S, Bandera EV, Zirpoli G, Hwang H, Roberts M, Ciupak G, Davis W, Sucheston L, et al. Cytokine and cytokine receptor genes of the adaptive immune response are differentially associated with breast cancer risk in American women of African and European ancestry. Int J Cancer. 2014; 134:1408–1421. [PubMed: 23996684]

Raftogianis R, Creveling C, Weinshilboum R, Weisz J. Estrogen metabolism by conjugation. J Natl Cancer Inst Monogr. 2000:113–124. [PubMed: 10963623]

Rebbeck TR, Troxel AB, Walker AH, Panossian S, Gallagher S, Shatalova EG, Blanchard R, Norman S, Bunin G, DeMichele A, et al. Pairwise combinations of estrogen metabolism genotypes in postmenopausal breast cancer etiology. Cancer Epidemiol Biomarkers Prev. 2007; 16:444–450. [PubMed: 17372239]

Reding KW, Chen C, Lowe K, Doody DR, Carlson CS, Chen CT, Houck J, Weiss LK, Marchbanks PA, Bernstein L, et al. Estrogen-related genes and their contribution to racial differences in breast cancer risk. Cancer Causes Control. 2012; 23:671–681. [PubMed: 22418777]

Ruiz-Narvaez EA, Rosenberg L, Wise LA, Reich D, Palmer JR. Validation of a small set of ancestral informative markers for control of population admixture in African Americans. Am J Epidemiol. 2011; 173:587–592. [PubMed: 21262910]

Santen RJ, Allred DC. The estrogen paradox. Nat Clin Pract Endocrinol Metab. 2007; 3:496–497. [PubMed: 17519914]

Sauna ZE, Kimchi-Sarfaty C, Ambudkar SV, Gottesman MM. Silent polymorphisms speak: how they affect pharmacogenomics and the treatment of cancer. Cancer Res. 2007; 67:9609–9612. [PubMed: 17942888]

Setiawan VW, Cheng I, Stram DO, Giorgi E, Pike MC, Van Den Berg D, Pooler L, Burtt NP, Le Marchand L, Altshuler D, et al. A systematic assessment of common genetic variation in CYP11A and risk of breast cancer. Cancer Res. 2006; 66:12019–12025. [PubMed: 17178901]

Sims AH, Howell A, Howell SJ, Clarke RB. Origins of breast cancer subtypes and therapeutic implications. Nat Clin Pract Oncol. 2007; 4:516–525. [PubMed: 17728710]

Spurdle AB, Hopper JL, Dite GS, Chen X, Cui J, McCredie MR, Giles GG, Southey MC, Venter DJ, Easton DF, et al. CYP17 promoter polymorphism and breast cancer in Australian women under age forty years. J Natl Cancer Inst. 2000; 92:1674–1681. [PubMed: 11036113]

Stacey SN, Sulem P, Zanon C, Gudjonsson SA, Thorleifsson G, Helgason A, Jonasdottir A, Besenbacher S, Kostic JP, Fackenthal JD, et al. Ancestry-shift refinement mapping of the C6orf97-ESR1 breast cancer susceptibility locus. PLoS Genet. 2010; 6:e1001029. [PubMed: 20661439]

Taioli E, Trachman J, Chen X, Toniolo P, Garte SJ. A CYP1A1 restriction fragment length polymorphism is associated with breast cancer in African-American women. Cancer Res. 1995; 55:3757–3758. [PubMed: 7641189]

Tang L, Yao S, Till C, Goodman PJ, Tangen CM, Wu Y, Kristal AR, Platz EA, Neuhouser ML, Stanczyk FZ, et al. Repeat polymorphisms in estrogen metabolism genes and prostate cancer risk: results from the Prostate Cancer Prevention Trial. Carcinogenesis. 2011; 32:1500–1506. [PubMed: 21771722]

Thompson PA, Ambrosone C. Molecular epidemiology of genetic polymorphisms in estrogen metabolizing enzymes in human breast cancer. J Natl Cancer Inst Monogr. 2000:125–134. [PubMed: 10963624]

Van Emburgh BO, Hu JJ, Levine EA, Mosley LJ, Perrier ND, Freimanis RI, Allen GO, Rubin P, Sherrill GB, Shaw CS, et al. Polymorphisms in CYP1B1, GSTM1, GSTT1 and GSTP1, and susceptibility to breast cancer. Oncol Rep. 2008; 19:1311–1321. [PubMed: 18425393]

Yager JD, Davidson NE. Estrogen carcinogenesis in breast cancer. N Engl J Med. 2006; 354:270–282. [PubMed: 16421368]

Yao S, Zirpoli G, Bovbjerg DH, Jandorf L, Hong CC, Zhao H, Sucheston LE, Tang L, Roberts M, Ciupak G, et al. Variants in the vitamin D pathway, serum levels of vitamin D, and estrogen receptor negative breast cancer among African-American women: a case-control study. Breast Cancer Res. 2012; 14:R58. [PubMed: 22480149]

Yu W, Gwinn M, Clyne M, Yesupriya A, Khoury MJ. A navigator for human genome epidemiology. Nat Genet. 2008; 40:124–125. [PubMed: 18227866]

Zheng W, Cai Q, Signorello LB, Long J, Hargreaves MK, Deming SL, Li G, Li C, Cui Y, Blot WJ. Evaluation of 11 breast cancer susceptibility loci in African-American women. Cancer Epidemiol Biomarkers Prev. 2009a; 18:2761–2764. [PubMed: 19789366]

Zheng W, Long J, Gao YT, Li C, Zheng Y, Xiang YB, Wen W, Levy S, Deming SL, Haines JL, et al. Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. Nat Genet. 2009b; 41:324–328. [PubMed: 19219042]

Zhu BT, Han GZ, Shim JY, Wen Y, Jiang XR. Quantitative structure-activity relationship of various endogenous estrogen metabolites for human estrogen receptor alpha and beta subtypes: Insights into the structural determinants favoring a differential subtype binding. Endocrinology. 2006; 147:4132–4150. [PubMed: 16728493]

Endocr Relat Cancer. Author manuscript; available in PMC 2015 December 01.

Table 1

Participant characteristics in the Women's Circle of Health Study (WCHS).

Characteristics	European American			African American		
	Cases (n=658)	Controls (n=649)	P[a]	Cases (n=621)	Controls (n=744)	P[a]
Age (y), mean (SD)	52.1 (10.1)	49.7 (8.7)	0.0002	51.5 (10.4)	48.7 (9.4)	0.0001
Education, n (%)						
Less than high school	20 (3.0)	6 (0.9)	0.0001	84 (13.5)	103 (13.8)	0.06
High school	115 (17.5)	67 (10.3)		196 (31.6)	192 (25.8)	
College and graduate school	523 (79.5)	576 (88.8)		341 (54.9)	449 (60.4)	
Family history, n (%)						
Yes	161 (24.5)	109 (16.8)	0.0006	89 (14.3)	87 (11.7)	0.15
No	497 (75.5)	540 (83.2)		532 (85.7)	657 (88.3)	
History of benign breast disease, n (%)						
Yes	269 (41.3)	210 (32.5)	0.0009	193 (31.2)	157 (21.1)	0.0001
No	382 (58.7)	437 (67.5)		426 (68.8)	586 (78.9)	
Menopausal status, n (%)						
Premenopausal	343 (52.1)	356 (54.9)	0.32	309 (49.8)	412 (55.4)	0.04
Postmenopausal	315 (47.9)	293 (45.1)		312 (50.2)	332 (44.6)	
Number of full-term pregnancy, mean (SD)	1.55 (1.4)	1.57 (1.5)	0.90	2.17 (1.7)	2.19 (1.7)	0.78
Breast feeding						
Yes	282 (42.8)	330 (50.9)	0.01	260 (41.9)	307 (41.3)	0.70
No	169 (25.7)	134 (20.6)		260 (41.9)	325 (43.7)	
Nulliparous	207 (31.5)	185 (28.5)		101 (16.2)	112 (15.0)	
HRT, n (%)[b]						
Yes	140 (44.4)	134 (45.7)	0.75	78 (25.2)	64 (19.3)	0.07
No	175 (55.6)	159 (54.3)		231 (74.8)	267 (80.7)	
BMI, mean (SD)	27.2 (6.6)	27.4 (7.2)	0.97	31.1 (6.72)	32.0 (7.8)	0.16
Proportion of European ancestry, mean (SD)	0.97 (0.08)	0.99 (0.04)	0.0005	0.14 (0.16)	0.14 (0.14)	0.07
Estrogen months, mean (SD)[c]	395.8 (77.8)	381.4 (82.4)	0.004	381.0 (83.3)	364.2 (87.3)	0.0002

Abbreviations: BMI, body mass index; HRT, hormone replacement therapy.

[a] *P* values were calculated by Wilcoxon rank-sum test for continuous variables and by Chi-square test for categorical variables.

[b] HRT use in postmenopausal women.

[c] In premenopausal women: "months between age at enrollment and age at menarche" minus "months of pregnancy", then minus "months of breastfeeding"; in postmenopausal women: "months between age at menopause and age at menarche" minus "months of pregnancy", then minus "months of breastfeeding".

Table 2

Significant SNPs associated with breast cancer risk among European American or African American women in WCHS.

Gene	SNP	Coded/ reference allele	Race	Allele frequency in cases	Allele frequency in controls	Heterozygotes OR(95%CI)[a]	Homozygotes OR(95%CI)[a]	Per allele OR(95%CI)[a]	P for trend	Pc[b]
CYP1A2	rs2470893	A/G	EA	0.24	0.26	1.04 (0.82–1.33)	0.76 (0.46–1.25)	0.96 (0.79–1.16)	0.63	1.00
			AA	0.06	0.05	1.37 (0.93–2.01)	2.78 (0.60–12.80)	1.42 (1.00–2.02)	0.05	0.22
ESR1	rs1801132	C/G	EA	0.24	0.17	1.34 (1.04–1.73)	2.74 (1.53–4.92)	1.47 (1.20–1.80)	0.0002	0.001
			AA	0.11	0.11	1.06 (0.80–1.40)	1.06 (0.80–1.40)	1.07 (0.83–1.37)	0.63	1.00
ESR1	rs2046210	A/G	EA	0.40	0.35	1.02 (0.79–1.31)	1.75 (1.21–2.54)	1.24 (1.04–1.47)	0.02	0.09
			AA	0.63	0.64	1.13 (0.8–1.58)	0.97 (0.68–1.37)	0.95 (0.81–1.12)	0.54	1.00
ESR1	rs3020314	G/A	EA	0.36	0.28	1.47 (1.15–1.88)	1.95 (1.30–2.94)	1.43 (1.19–1.70)	0.00009	0.0004
			AA	0.71	0.72	0.82 (0.53–1.26)	0.82 (0.53–1.25)	0.94 (0.79–1.13)	0.51	1.00

Abbreviations: OR, odds ratio; 95%CI, 95% confidence interval; EA, European American; AA, African American; Code allele/reference allele, the minor allele of each SNP in EA controls is defined as coded allele, whereas the other allele is defined as reference allele; Allele frequency, allele frequencies of the coded alleles; P for trend, *p* value for linear trend; NS, not significant after correction for multiple testing. Detailed information on number of subjects, ORs and 95% CIs for all 49 SNPs analyzed are shown in Supplementary Table 2.

[a] Adjusted for age at diagnosis (continuous), education (less than high school, high school, college and graduate school), family history of breast cancer (yes, no), history of benign breast disease (yes, no), menopausal status (premenopausal, postmenopausal), number of full pregnancies (continuous), breast feeding (yes, no, nulliparous), Hormone Replacement Therapy (HRT, yes, no), body mass index (continuous), proportion of European ancestry (continuous) and “estrogen months” (continuous).

[b] *P* values corrected for multiple testing using a modified FDR method.

Table 3

Associations between SNPs and risk of breast cancer by ER status among European American or African American women.

Gene	SNP	Coded/ reference allele	Race	Estrogen receptor status	Allele frequency in cases	Allele frequency in controls	Heterozygotes OR(95%CI)[a]	Homozygotes OR(95%CI)[a]	Per allele OR(95%CI)[a]	P for trend	Pc[c]
CYP17A1	rs12413409	A/G	EA	+	0.10	0.10	0.95 (0.66–1.36)	1.89 (0.54–6.61)	1.04 (0.75–1.43)	0.83	1.00
				−	0.15	0.10	2.01 (1.15–3.54)	2.55 (0.28–23.31)	1.92 (1.15–3.21)	0.01	0.05
			AA	+	0.05	0.05	1.01 (0.64–1.60)	N/A[b]	0.99 (0.63–1.55)	N/A[b]	N/A[b]
				−	0.05	0.05	0.90 (0.48–1.71)	N/A[b]	0.89 (0.47–1.67)	N/A[b]	N/A[b]
CYP1A2	rs2470893	A/G	EA	+	0.25	0.26	1.03 (0.78–1.37)	0.61 (0.33–1.13)	0.91 (0.73–1.14)	0.40	1.00
				−	0.21	0.26	0.89 (0.53–1.50)	0.33 (0.07–1.45)	0.76 (0.49–1.16)	0.20	0.90
			AA	+	0.08	0.05	1.78 (1.14–2.79)	2.80 (0.52–14.99)	1.76 (1.17–2.64)	0.006	0.03
				−	0.05	0.05	0.72 (0.33–1.59)	N/A[b]	0.67 (0.31–1.45)	N/A[b]	N/A[b]
ESR1	rs1801132	C/G	EA	+	0.24	0.17	1.34 (1.00–1.80)	2.69 (1.38–5.23)	1.46 (1.16–1.85)	0.002	0.01
				−	0.23	0.17	1.25 (0.73–2.14)	2.69 (0.92–7.89)	1.42 (0.84–2.16)	0.10	0.45
			AA	+	0.11	0.11	0.96 (0.68–1.35)	0.85 (0.20–3.68)	0.95 (0.69–1.31)	0.76	1.00
				−	0.12	0.11	0.99 (0.61–1.58)	2.62 (0.61–11.29)	1.11 (0.73–1.69)	0.61	1.00
ESR1	rs2046210	A/G	EA	+	0.41	0.35	1.11 (0.82–1.49)	1.74 (1.14–2.67)	1.26 (1.03–1.55)	0.02	0.09
				−	0.40	0.35	0.82 (0.48–1.42)	1.62 (0.79–3.32)	1.16 (0.80–1.67)	0.44	1.00
			AA	+	0.63	0.64	1.49 (0.96–2.30)	1.08 (0.69–1.70)	0.95 (0.77–1.16)	0.59	1.00
				−	0.64	0.64	0.75 (0.44–1.30)	0.84 (0.49–1.45)	0.96 (0.74–1.25)	0.75	1.00
ESR1	rs3020314	G/A	EA	+	0.36	0.28	1.46 (1.10–1.95)	1.88 (1.18–3.00)	1.40 (1.14–1.72)	0.001	0.005
				−	0.36	0.28	2.01 (1.19–3.39)	1.78 (0.71–4.48)	1.55 (1.07–2.25)	0.02	0.09
			AA	+	0.72	0.72	0.98 (0.57–1.69)	1.02 (0.59–1.74)	1.02 (0.82–1.28)	0.84	1.00
				−	0.71	0.72	0.76 (0.38–1.52)	0.77 (0.39–1.52)	0.93 (0.70–1.25)	0.63	1.00
HSD17B2	rs4445895	A/G	EA	+	0.41	0.41	1.03 (0.76–1.39)	1.00 (0.67–1.50)	1.01 (0.83–1.22)	0.96	1.00
				−	0.29	0.41	0.41 (0.24–0.71)	0.54 (0.26–1.12)	0.62 (0.43–0.90)	0.01	0.05
			AA	+	0.36	0.31	1.23 (0.92–1.66)	1.54 (0.99–2.40)	1.24 (1.01–1.52)	0.04	0.18
				−	0.29	0.31	1.08 (0.74–1.59)	0.71 (0.35–1.47)	0.94 (0.71–1.25)	0.67	1.00
UGT1A9	rs6714486	A/T	EA	+	0.06	0.05	1.35 (0.87–2.1)	N/A[b]	1.24 (0.81–1.89)	N/A[b]	N/A[b]
				−	0.03	0.05	0.47 (0.16–1.41)	N/A[b]	0.45 (0.16–1.33)	N/A[b]	N/A[b]

Gene	SNP	Coded/ reference allele	Race	Estrogen receptor status	Allele frequency in cases	Allele frequency in controls	Heterozygotes OR(95%CI)[a]	Homozygotes OR(95%CI)[a]	Per allele OR(95%CI)[a]	P for trend	Pc[c]
			AA	+	0.13	0.17	0.70 (0.50–0.98)	0.70 (0.31–1.56)	0.75 (0.57–0.98)	0.04	0.18
				−	0.16	0.17	1.08 (0.71–1.63)	0.14 (0.02–1.14)	0.85 (0.60–1.22)	0.39	1.00

Abbreviations: OR, odds ratio; 95%CI, 95% confidence interval; EA, European American; AA, African American; Code allele/reference allele, the minor allele of each SNP in EA controls is defined as coded allele, whereas the other allele is defined as reference allele; Allele frequency, allele frequencies of the coded alleles; P for trend, *p* value for linear trend. Detailed information on number of subjects, ORs and 95% CIs for all 49 SNPs analyzed are shown in Supplementary Table 3.

[a] Adjusted for age at diagnosis (continuous), education (less than high school, high school, college and graduate school), family history of breast cancer (yes, no), history of benign breast disease (yes, no), menopausal status (premenopausal, postmenopausal), number of full pregnancies (continuous), breast feeding (yes, no, nulliparous), Hormone Replacement Therapy (HRT, yes, no), body mass index (continuous), proportion of European ancestry (continuous) and “estrogen months” (continuous). *P* values corrected for multiple comparisons using a modified FDR method.

[b] OR, 95% CI and corresponding *P* for trend were not calculated due to the small number of breast cancer cases in the category.

[c] *P* values corrected for multiple testing using a modified FDR method.

Table 4

Association between SNPs and risk of breast cancer by length of estrogen exposure among European American or African American women in WCHS.

Gene	SNP	Coded/ reference allele	Race	Estrogen exposure status	Allele frequency in cases	Allele frequency in controls	Heterozygotes OR(95%CI)[a]	Homozygotes OR(95%CI)[a]	Per allele OR(95%CI)[a]	P for trend	Pc[c]
ESR1	rs1801132	C/G	EA	Long[b]	0.33	0.17	1.48 (1.06–2.07)	3.83 (1.57–9.36)	1.64 (1.24–2.16)	0.0005	0.002
				Short	0.24	0.19	1.22 (0.81–1.82)	2.31 (1.03–5.20)	1.36 (1.00–1.85)	0.05	0.22
			AA	Long	0.10	0.11	0.94 (0.63–1.43)	4.67 (0.49–44.84)	1.06 (0.72–1.55)	0.78	1.00
				Short	0.14	0.12	1.27 (0.86–1.87)	0.77 (0.17–3.48)	1.18 (0.83–1.68)	0.35	1.00
ESR1	rs2046210	A/G	EA	Long	0.40	0.34	1.09 (0.79–1.52)	1.88 (1.14–3.09)	1.28 (1.02–1.61)	0.03	0.13
				Short	0.40	0.38	0.93 (0.62–1.39)	1.63 (0.92–2.89)	1.19 (0.91–1.56)	0.21	0.94
			AA	Long	0.63	0.61	1.21 (0.76–1.94)	1.14 (0.70–1.84)	1.04 (0.82–1.37)	0.75	1.00
				Short	0.63	0.66	0.99 (0.60–1.64)	0.77 (0.46–1.30)	0.85 (0.67–1.08)	0.19	0.85
ESR1	rs3020314	G/A	EA	Long	0.36	0.27	1.54 (1.11–2.12)	2.25 (1.27–3.98)	1.52 (1.19–1.93)	0.0007	0.003
				Short	0.37	0.30	1.51 (1.01–2.24)	1.72 (0.94–3.15)	1.37 (1.04–1.80)	0.02	0.19
			AA	Long	0.72	0.71	1.19 (0.64–2.22)	1.18 (0.63–2.19)	1.04 (0.81–1.35)	0.75	1.00
				Short	0.70	0.73	0.58 (0.32–1.08)	0.62 (0.34–1.14)	0.89 (0.69–1.15)	0.38	1.00

Abbreviations: OR, odds ratio; 95%CI, 95% confidence interval; EA, European American; AA, African American; Code allele/reference allele, the minor allele of each SNP in EA controls is defined as coded allele, whereas the other allele is defined as reference allele; Allele frequency, allele frequencies of the coded alleles; P for trend, *p* value for linear trend; NS, not significant after correction for multiple testing. Detailed information on number of subjects, ORs and 95% CIs for all 49 SNPs analyzed are shown in Supplementary Table 4.

[a] Adjusted for age at diagnosis (continuous), education (less than high school, high school, college and graduate school), family history of breast cancer (yes, no), history of benign breast disease (yes, no), menopausal status (premenopausal, postmenopausal), number of full pregnancies (continuous), breast feeding (yes, no, nulliparous), Hormone Replacement Therapy (HRT, yes, no), body mass index (continuous) and proportion of European ancestry (continuous).

[b] Stratified by the length of endogenous estrogen exposure using median in both EA and AA controls as the cutoff (longer or shorter than 384 months). Results of stratification using race-specific medians as cutoffs are similar and shown in Supplementary Table 4.

[c] *P* values corrected for multiple testing using a modified FDR method.

Table 5

Association between SNPs and risk of breast cancer by hormone replacement therapy (HRT) status among postmenopausal European American or African American women in WCHS.

Gene	SNP	Coded/ reference allele	Race	HRT usage status	Allele frequency in cases	Allele frequency in controls	Heterozygotes OR(95%CI)[a]	Homozygotes OR(95%CI)[a]	Per allele OR(95%CI)[a]	P for trend	Pc[c]
ESR1	rs1801132	C/G	EA	Yes	0.27	0.18	1.38 (0.77–2.46)	5.25 (1.35–20.41)	1.73 (1.10–2.72)	0.02	0.19
				No	0.24	0.17	1.38 (0.81–2.36)	3.17 (0.69–14.46)	1.50 (0.95–2.36)	0.08	0.36
			AA	Yes	0.13	0.10	3.03 (1.07–8.59)	0.52 (0.04–7.11)	1.79 (0.77–4.20)	0.18	0.81
				No	0.12	0.11	0.98 (0.61–1.58)	N/A[b]	1.10 (0.70–1.72)	0.68	N/A[b]
ESR1	rs2046210	A/G	EA	Yes	0.38	0.38	0.76 (0.42–1.36)	0.91 (0.39–2.12)	0.90 (0.61–.1.34)	0.61	1.00
				No	0.37	0.32	1.16 (0.69–1.97)	3.69 (1.44–9.43)	1.58 (1.06–2.34)	0.02	0.19
			AA	Yes	0.66	0.56	1.32 (0.46–3.79)	2.03 (0.68–6.07)	1.45 (0.86–2.45)	0.17	0.76
				No	0.63	0.64	1.08 (0.61–1.92)	0.93 (0.52–1.67)	0.94 (0.71–1.24)	0.66	1.00
ESR1	rs3020314	G/A	EA	Yes	0.38	0.27	1.97 (1.12–3.48)	2.61 (1.00–7.17)	1.76 (1.15–2.68)	0.009	0.04
				No	0.33	0.29	0.94 (0.56–1.56)	2.53 (0.93–6.91)	1.24 (0.83–1.83)	0.29	1.00
			AA	Yes	0.67	0.70	1.61 (0.47–5.56)	1.17 (0.33–4.12)	0.95 (0.54–1.69)	0.87	1.00
				No	0.75	0.72	1.49 (0.63–3.50)	1.68 (0.72–3.94)	1.21 (0.87–1.67)	0.25	1.00

Abbreviations: OR, odds ratio; 95%CI, 95% confidence interval; EA, European American; AA, African American; Code allele/reference allele, the minor allele of each SNP in EA controls is defined as coded allele, whereas the other allele is defined as reference allele; Allele frequency, allele frequencies of the coded alleles; P for trend, *p* value for linear trend; NS, not significant after correction for multiple testing. Detailed information on number of subjects, ORs and 95% CIs for all 49 SNPs analyzed are shown in Supplementary Table 5.

[a] Adjusted for age at diagnosis (continuous), education (less than high school, high school, college and graduate school), family history of breast cancer (yes, no), history of benign breast disease (yes, no), number of full pregnancies (continuous), breast feeding (yes, no, nulliparous), body mass index (continuous), proportion of European ancestry (continuous) and “estrogen months” (continuous).

[b] OR, 95% CI and corresponding *P* for trend were not calculated due to the small number of breast cancer cases in the category.

[c] *P* values corrected for multiple testing using a modified FDR method.