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

Lei Quan1,2, Chi-Chen Hong1, Gary Zirpoli1, Michelle R Roberts1, Thaer Khoury1, Lara E Sucheston-Campbell1, Dana H Bovbjerg2, Lina Jandorf3, Karen Pawlish4, Gregory Ciupak1, Warren Davis1, Elisa V Bandera5, Christine B Ambrosone1 and Song Yao1

1Roswell Park Cancer Institute, Elm and Carlton Streets, Buffalo, New York 14263, USA
2University of Pittsburgh Cancer Institute, Pittsburgh, Pennsylvania, USA
3Icahn School of Medicine at Mount Sinai, New York, New York, USA
4New Jersey Department of Health, Trenton, New Jersey, USA
5Rutgers Cancer Institute of New Jersey, New Brunswick, New Jersey, USA

Correspondence should be addressed to S Yao
Email song.yao@roswellpark.org

Abstract

It has been observed previously that compared with 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 \textit{ESR1} gene in EA women (rs1801132, odds ratio (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 \textit{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, \textit{ESR1} rs1801132, rs2046210, and rs3020314 showed stronger associations in ER-positive than in ER-negative breast cancer in only EA women. Associations with the \textit{ESR1} SNPs in EA women also appeared to be stronger with longer endogenous estrogen exposure or hormonal replacement therapy use. Our results indicate that there may be differential genetic influences on breast cancer risk in EA compared with AA women and that 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 in breast cancer.

Key Words

- breast cancer
- disparity
- African–American
- estrogen receptor
- estrogen synthesis
- estrogen metabolism
- estrogen response
- \textit{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 (Millikan et al. 2008, Ambrosone et al. 2014a). Compared with American women of European ancestry (EA), those of African ancestry (AA) are more likely to be diagnosed with
Endocrine-related cancer, race and breast cancer

L Quan et al.

Research

Breast cancer before the age of 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 & Newman 2005, Huo et al. 2009), indicating that genetic factors related to AA may, in part, account for the higher proportion of more aggressive breast cancer in AA women. As the populations of AA typically exhibit shorter linkage disequilibrium (LD) blocks across the genome compared with populations of EA (Hainan & 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 causal 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 (Key et al. 2002, Fuhrman et al. 2012), probably due to the dual abilities of estrogens to stimulate cell proliferation and gene expression through binding with ERs and to cause DNA damage via mutagenic estrogen metabolites such as 2-OH and 4-OH catechol estrogens (Pike et al. 1993, Liehr 1997). As 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 hormonal replacement therapy (HRT) use, as compared with EA women (Brett & Madans 1997, Millikan et al. 2008), it is postulated that racial disparities in breast cancer may be attributed, in part, to estrogens, and related hormonal factors. This postulate is supported by results from several recent studies, including those from our group (Millikan et al. 2008, Palmer et al. 2011, Ambrosone et al. 2014a,b).

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 & Ambrosone 2000). A number of candidate association studies have examined the association of variants in estrogen-related genes with breast cancer mostly in EAs, while similar studies in AAs are still sparse (Taioli et al. 1995, Rebbeck et al. 2007, Van Emburgh et al. 2008, Kato et al. 2009). A single-nucleotide polymorphism (SNP) in the ER alpha (ESR1) gene, rs2046210, has been shown to be associated with breast cancer risk in genome-wide association studies (GWAS) carried out first in an Asian population (Zheng et al. 2009b) and subsequently in populations of European descent (Michailidou et al. 2013). However, it remains unclear whether ESR1 rs2046210 is also associated with the 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 in 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 the associations between selected genetic variants in genes related to estrogen biosynthesis, metabolism, and response pathways and the 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 the 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 of age 20–75 years, diagnosed with incident, primary, histologically confirmed breast cancer, and with no previous history of cancer other than non-melanoma skin cancer. The controls without a history of any cancer diagnosis other than non-melanoma skin cancer were identified by random-digit dialing (RDD) and matched with cases on the basis of race and 5-year age group. The enrollment of AA controls in NJ was supplemented with community-based recruitment, as a combination of RDD and community controls has been shown to be more representative of the general population in NJ (Bandera et al. 2013). The 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% for EA cases and controls, respectively, and 79 and 48% for AA cases and controls, respectively, and 79 and 49% for EA cases and controls, respectively, and 79 and 48% for AA cases and controls, respectively.
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, the Icahn School of Medicine at Mount Sinai, and participating hospitals in New York City, functioning according to the third 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, and 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.) and from saliva using the Oragene protocol (DNA Genotek, Inc., Ottawa, ON, Canada). The quality and quantity of purified DNA were evaluated using a Nanodrop UV-spectrometer (Thermo Fisher Scientific, Inc., Wilmington, DE, USA) and PicoGreen-based fluorometric assays (Invitrogen, Inc.). The DNA samples were stored at −80 °C until analysis.

In our analysis, we included all major genes involved in the estrogen biosynthesis, metabolism, and response pathways, including cytochrome-dependent monoxygenase (CYP) genes, hydroxysteroid dehydrogenase (HSD) genes, catechol-O-methyltransferase (COMT), UDP-glucuronosyltransferase (UGT) genes, and also ER (ESR) genes (Supplementary Fig. 1) (Yager & Davidson 2006, Germain 2011, Liang & Shang 2013). We then surveyed the Human Genome Epidemiology (HuGE) Navigator for the selected genes to identify SNPs within these genes that had been previously shown to be associated with the risk of any cancer or cancer outcome, with a focus on SNPs that had been 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, USA). Five percent duplicates and two sets of in-house trio samples of EA and AA 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%. The samples or SNPs with a call rate <90% were excluded, as were monomorphic SNPs or SNPs with an minor allele frequency <5% in both AA and EA controls. For each SNP, Hardy–Weinberg equilibrium (HWE) was assessed among controls. The SNPs deviating from HWE in both EA and AA controls were excluded. The clustering plots of SNPs that were significant in the statistical analysis were manually re-inspected post-hoc to ensure that the calls for the genotype of each sample were robust. We selected 62 SNPs for the assay design: five of them failed in the assay design and validation process, three were excluded due to low call rate, three were monomorphic or rare and excluded, and another two were excluded for violating HWE, leaving a total of 49 SNPs in 20 genes included in our analyses (Supplementary Table 1). The major allele of each SNP in EA controls was 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 analyses

All analyses were conducted using SAS 9.4 (SAS Institute, Cary, NC, USA) separately for EA and AA women, according to self-reported race. Descriptive variables were compared between cases and controls using χ² tests for categorical variables and Wilcoxon rank-sum test for continuous variables. The proportions of EA and AA in individual EA and AA women 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). As the sum of two ancestral proportions in each individual is always one, we used only the proportion of EA in all analyses. Genotype frequencies of each SNP were compared between EA and AA controls using χ² test or Fisher’s exact test where appropriate. To compare allele
frequencies obtained from our study with 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).

The odds ratios (ORs) and 95% 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 (Gordon 1995, Althuis et al. 2004, Hwang et al. 2005, Krieger et al. 2008, Ambrosone et al. 2014b), including age at diagnosis, education, family history of breast cancer, history of benign breast disease, menopausal status, number of full-term pregnancies, breast feeding, HRT, BMI, proportion of EA, 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, and 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 $a = \frac{\sum_{i=1}^{k} (1/i)}{Narum 2006}$.

Results

Participant characteristics

Participant characteristics are given 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. The cases were also slightly older than the controls, but within a 5-year age range in both EA and AA

Table 1  Participant characteristics in the Women’s Circle of Health Study (WCHS)

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

HRT, hormone replacement therapy.

$^a$P values were calculated by Wilcoxon rank-sum test for continuous variables and by $\chi^2$ 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’.
women. The cases did not differ significantly from controls with respect to BMI, HRT use in postmenopausal women, or number of full-term pregnancies 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, see section on supplementary data at the end of this article, rs2046210 and rs3020314 in Supplementary Table 2 given at the end of this article 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 given 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 given in Table 2. We found different associations between 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, Pc=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=0.0009, Pc=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 of breast cancer with respect to 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 according to 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, see section on supplementary data given at the end of this article. Significant associations in either EA or AA women with a P for trend <0.05 before correction for multiple testing are

Table 2 Significant SNPs associated with breast cancer risk among European American or African American women in WCHS

<table>
<thead>
<tr>
<th>Gene</th>
<th>SNP</th>
<th>Coded/ reference allele</th>
<th>Allele frequency in controls</th>
<th>Allele frequency in cases</th>
<th>OR for trend</th>
<th>P for trend</th>
<th>OR for case</th>
<th>P for case</th>
</tr>
</thead>
<tbody>
<tr>
<td>CYP1A2</td>
<td>rs2470893</td>
<td>A/G</td>
<td>0.25</td>
<td>0.24</td>
<td>1.04</td>
<td>0.06</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>ESR1</td>
<td>rs1801132</td>
<td>C/G</td>
<td>0.26</td>
<td>0.26</td>
<td>0.95</td>
<td>0.86</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>ESR1</td>
<td>rs2046210</td>
<td>A/G</td>
<td>0.41</td>
<td>0.40</td>
<td>1.24</td>
<td>0.005</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>ESR1</td>
<td>rs3020314</td>
<td>G/A</td>
<td>0.36</td>
<td>0.30</td>
<td>1.47</td>
<td>0.001</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

**Note:** OR, odds ratio; EA, European American; AA, African American; coded allele/reference allele, the minor allele of each SNP in EA controls is defined as the coded allele, whereas the other allele is defined as the reference allele; allele frequency, allele frequencies of the coded alleles; OR for trend, OR value for linear trend; NS, not significant after correction for multiple testing. Detailed information on statistical methods is given in the section on supplementary data at the end of this article.

Table 3 Significant SNPs associated with breast cancer risk among European American or African American women in WCHS

<table>
<thead>
<tr>
<th>Gene</th>
<th>SNP</th>
<th>Coded/ reference allele</th>
<th>Allele frequency in controls</th>
<th>Allele frequency in cases</th>
<th>OR for trend</th>
<th>P for trend</th>
<th>OR for case</th>
<th>P for case</th>
</tr>
</thead>
<tbody>
<tr>
<td>CYP1A2</td>
<td>rs2470893</td>
<td>A/G</td>
<td>0.25</td>
<td>0.24</td>
<td>1.04</td>
<td>0.06</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>ESR1</td>
<td>rs1801132</td>
<td>C/G</td>
<td>0.26</td>
<td>0.26</td>
<td>0.95</td>
<td>0.86</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>ESR1</td>
<td>rs2046210</td>
<td>A/G</td>
<td>0.41</td>
<td>0.40</td>
<td>1.24</td>
<td>0.005</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>ESR1</td>
<td>rs3020314</td>
<td>G/A</td>
<td>0.36</td>
<td>0.30</td>
<td>1.47</td>
<td>0.001</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>
## Table 3

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

<table>
<thead>
<tr>
<th>Gene</th>
<th>SNP</th>
<th>Coded/ reference allele</th>
<th>Race</th>
<th>Estrogen receptor status</th>
<th>Allele frequency in cases</th>
<th>Allele frequency in controls</th>
<th>Heterozygotes OR (95% CI)*</th>
<th>Homozygotes OR (95% CI)*</th>
<th>Per allele OR (95% CI)*</th>
<th>P for trend</th>
<th>P_c^</th>
</tr>
</thead>
<tbody>
<tr>
<td>CYP17A1</td>
<td>rs12413409</td>
<td>A/G</td>
<td>EA</td>
<td>+</td>
<td>0.10</td>
<td>0.10</td>
<td>0.95 (0.66–1.36)</td>
<td>1.89 (0.54–6.61)</td>
<td>1.04 (0.75–1.43)</td>
<td>0.83</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>−</td>
<td>0.15</td>
<td>0.10</td>
<td>2.01 (1.15–3.54)</td>
<td>2.55 (0.28–23.31)</td>
<td>1.92 (1.15–3.21)</td>
<td>0.01</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>AA</td>
<td>0.05</td>
<td>0.05</td>
<td>1.01 (0.64–1.60)</td>
<td>NA^b</td>
<td>NA^b</td>
<td>NA^b</td>
<td>NA^b</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>−</td>
<td>0.05</td>
<td>0.05</td>
<td>0.90 (0.48–1.71)</td>
<td>NA^b</td>
<td>NA^b</td>
<td>NA^b</td>
<td>NA^b</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>AA</td>
<td>0.21</td>
<td>0.26</td>
<td>1.03 (0.78–1.37)</td>
<td>0.61 (0.33–1.13)</td>
<td>0.91 (0.73–1.14)</td>
<td>0.40</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>−</td>
<td>0.08</td>
<td>0.05</td>
<td>0.89 (0.53–1.50)</td>
<td>0.33 (0.07–1.45)</td>
<td>0.76 (0.49–1.16)</td>
<td>0.20</td>
<td>0.90</td>
</tr>
<tr>
<td>ESR1</td>
<td>rs1801132</td>
<td>C/G</td>
<td>EA</td>
<td>+</td>
<td>0.24</td>
<td>0.17</td>
<td>1.34 (1.00–1.80)</td>
<td>2.69 (1.38–5.23)</td>
<td>1.46 (1.16–1.85)</td>
<td>0.002</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>−</td>
<td>0.23</td>
<td>0.17</td>
<td>1.25 (0.73–2.14)</td>
<td>2.69 (0.92–7.89)</td>
<td>1.42 (0.84–2.16)</td>
<td>0.10</td>
<td>0.45</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>AA</td>
<td>0.11</td>
<td>0.11</td>
<td>0.96 (0.68–1.35)</td>
<td>0.85 (0.20–3.68)</td>
<td>0.95 (0.69–1.31)</td>
<td>0.76</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>−</td>
<td>0.12</td>
<td>0.11</td>
<td>0.99 (0.61–1.58)</td>
<td>2.62 (0.61–11.29)</td>
<td>1.11 (0.73–1.69)</td>
<td>0.61</td>
<td>1.00</td>
</tr>
<tr>
<td>ESR1</td>
<td>rs2046210</td>
<td>A/G</td>
<td>EA</td>
<td>+</td>
<td>0.41</td>
<td>0.35</td>
<td>1.11 (0.82–1.49)</td>
<td>1.74 (1.14–2.67)</td>
<td>1.26 (1.03–1.55)</td>
<td>0.02</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>−</td>
<td>0.40</td>
<td>0.35</td>
<td>0.82 (0.48–1.42)</td>
<td>1.62 (0.79–3.32)</td>
<td>1.16 (0.80–1.67)</td>
<td>0.44</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>AA</td>
<td>0.63</td>
<td>0.64</td>
<td>1.49 (0.96–2.30)</td>
<td>1.08 (0.69–1.70)</td>
<td>0.95 (0.77–1.16)</td>
<td>0.59</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>−</td>
<td>0.64</td>
<td>0.64</td>
<td>0.75 (0.44–1.30)</td>
<td>0.84 (0.49–1.45)</td>
<td>0.96 (0.74–1.25)</td>
<td>0.75</td>
<td>1.00</td>
</tr>
<tr>
<td>ESR1</td>
<td>rs3020314</td>
<td>G/A</td>
<td>EA</td>
<td>+</td>
<td>0.36</td>
<td>0.28</td>
<td>1.46 (1.10–1.95)</td>
<td>1.88 (1.18–3.00)</td>
<td>1.40 (1.14–1.72)</td>
<td>0.001</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>−</td>
<td>0.36</td>
<td>0.28</td>
<td>2.01 (1.19–3.39)</td>
<td>1.78 (0.71–4.48)</td>
<td>1.55 (0.70–2.25)</td>
<td>0.02</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>AA</td>
<td>0.72</td>
<td>0.72</td>
<td>0.98 (0.57–1.69)</td>
<td>1.02 (0.59–1.74)</td>
<td>1.02 (0.82–1.28)</td>
<td>0.84</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>−</td>
<td>0.71</td>
<td>0.72</td>
<td>0.76 (0.38–1.52)</td>
<td>0.77 (0.39–1.52)</td>
<td>0.93 (0.70–1.25)</td>
<td>0.63</td>
<td>1.00</td>
</tr>
<tr>
<td>HSD17B2</td>
<td>rs4445895</td>
<td>A/G</td>
<td>EA</td>
<td>+</td>
<td>0.41</td>
<td>0.41</td>
<td>1.03 (0.76–1.39)</td>
<td>1.00 (0.67–1.50)</td>
<td>1.01 (0.83–1.22)</td>
<td>0.96</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>−</td>
<td>0.29</td>
<td>0.41</td>
<td>0.41 (0.24–0.71)</td>
<td>0.54 (0.26–1.12)</td>
<td>0.62 (0.43–0.90)</td>
<td>0.01</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>AA</td>
<td>0.36</td>
<td>0.31</td>
<td>1.23 (0.92–1.66)</td>
<td>1.54 (0.99–2.40)</td>
<td>1.24 (1.01–1.52)</td>
<td>0.04</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>−</td>
<td>0.29</td>
<td>0.31</td>
<td>1.08 (0.74–1.59)</td>
<td>0.71 (0.35–1.47)</td>
<td>0.94 (0.71–1.25)</td>
<td>0.67</td>
<td>1.00</td>
</tr>
<tr>
<td>UGT1A9</td>
<td>rs6714486</td>
<td>A/T</td>
<td>EA</td>
<td>+</td>
<td>0.06</td>
<td>0.05</td>
<td>1.35 (0.87–2.1)</td>
<td>NA^b</td>
<td>NA^b</td>
<td>NA^b</td>
<td>NA^b</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>−</td>
<td>0.03</td>
<td>0.05</td>
<td>0.47 (0.16–141)</td>
<td>NA^b</td>
<td>NA^b</td>
<td>NA^b</td>
<td>NA^b</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>AA</td>
<td>0.13</td>
<td>0.17</td>
<td>0.70 (0.50–0.98)</td>
<td>0.70 (0.31–1.56)</td>
<td>0.75 (0.57–0.98)</td>
<td>0.04</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>−</td>
<td>0.16</td>
<td>0.17</td>
<td>1.08 (0.71–1.63)</td>
<td>0.14 (0.02–1.14)</td>
<td>0.85 (0.60–1.22)</td>
<td>0.39</td>
<td>1.00</td>
</tr>
</tbody>
</table>

OR, odds ratio; EA, European American; AA, African American; coded allele/reference allele, the minor allele of each SNP in EA controls is defined as the coded allele, whereas the other allele is defined as the 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 given in Supplementary Table 3, see section on supplementary data at the end of this article.

*Adjusted for age at diagnosis (continuous), education (less than high school, high school, college, and graduate school), family history of breast cancer (yes and no), history of benign breast disease (yes and no), menopausal status (premenopausal, postmenopausal), number of full pregnancies (continuous), breast feeding (yes, no, and nulliparous), hormone replacement therapy (HRT, yes, and no), BMI (continuous), proportion of European ancestry (continuous), and ‘estrogen months’ (continuous). P values corrected for multiple comparisons using a modified FDR method.

^bOR, 95% CI, and corresponding P for trend were not calculated due to the small number of breast cancer cases in the category.

^cP values corrected for multiple testing using a modified FDR method.
given 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 the 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 and 5, see section on supplementary data at the end of this article. 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, see section on supplementary data at the end of this article). 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). ESR1 rs2046210, however, was associated with an increased risk of breast cancer only in non-user postmenopausal EA women (Table 5).
Disparities in breast cancer biology are evident between EA and AA women and may be due, in part, to differences in genetic background (Yao et al. 2012, Quan et al. 2014). The estrogens are synthesized from cholesterol through a cascade of enzymatic reactions (Thompson & Ambrosone 2000, Tang et al. 2011). 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 affinities for the ERs, which may lead to differences in risk of breast cancer (Zhu et al. 2006).

Moreover, both parent estrogens and estrogen metabolites can be modified by a series of enzymes that influence their bioavailability to breast tissue (Raftogianis et al. 2000). In this study, we systematically examined the associations between genes involved in estrogen biosynthesis, metabolism, and response pathways (Supplementary Figure 1) and overall breast cancer risk, as well as by risk stratified 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 ERα gene and overall breast cancer risk in only EA women. The human ERα gene has eight exons that span 300 kb on chromosome 6q25.1 and encodes the α form of the ER (ERα). The three SNPs, i.e. rs1801132, rs2046210, and rs3020314, were not in LD with each other. Rs1801132 is a synonymous SNP at codon 325, which may influence mRNA stability and translation efficiency (Sauna et al. 2007, Li et al. 2010). 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 or Caucasian (CG/GG vs CC, OR = 0.92, 95% CI = 0.85–0.99) (Li et al. 2010). Rs2046210 is a SNP located upstream of ERα that has been reported to be associated with breast cancer risk in a three-stage 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 (Michaillidou et al. 2013). SNP rs3020314 tags a highly conserved region of ERα intron 4 and may change the ratio of two mRNA splice forms (Fuqua & Wolf 1995, Dunning et al. 2009).
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 reports in previous studies 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, and 5). Previously, efforts to evaluate breast stratifications by ER status or by estrogen exposure status (Table 2), even after directions as reports in previous studies in EA women, 1.02–1.09) (Dunning et al. 2009). 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). The associations between estrogen-related genes and risk of breast cancers stratified by ER status have not been widely investigated. Among the SNPs associated with overall breast cancer risk, ESR1 rs2046210 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 stratified 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, results from a recent GWAS indicated that ESR1 rs2046210 is associated with risk of ER-negative and triple-negative 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 associated with the strongest effect, associated with an almost twofold 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 variants have been associated with breast cancer risk among younger women (Bergman-Jungestrom et al. 1999, Spurdle et al. 2000) and among a 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 & 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 & 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 at high risk (Sims et al. 2007). In addition, there is evidence that ER-negative stem cells are essential for normal breast development and give 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 the possibility 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 & 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 & 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 (Tables 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, and 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 EA (or AA) 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 associated with 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 similar to or lower than those in other similar studies (Long et al. 2013, Palmer et al. 2013). We are aware of potential effect of this substantial missing data on our analysis. It is possible that ER status may not be missing completely at random. Results of some studies have indicated that ER status is more frequently missing among women of color or of lower socioeconomic status, probably due to inadequacies of medical care (Krieger et al. 2008). However, in our study, ER status was missing at a similar rate in EA and AA women; there were no obvious differences in missing ER status across various levels of education among either EA or AA women in our study (data not shown). It is thus likely that the missing is non-differential and would have limited effects on the results of the analysis besides lowering statistical power.

In conclusion, this 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 and to assess their utility in predicting breast cancer risk or treatment response for anti-estrogen prevention in women at high risk.

Supplementary data
This is linked to the online version of the paper at http://dx.doi.org/10.1530/ERC-14-0250.

Declaration of interest
The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

Funding
This work was supported by grants from the U.S. Army Medical Research and Material Command (USARMMC) (DAMD-1701-1-0334 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 the Philip L Hubbell family. The samples were stored and managed by the Roswell Park Cancer Institute (RPCI) DataBank and BioRepository (DBBR) and genotyping was performed in the RPCI Genomics Core Facility, both Cancer Center Support Grant 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 USBDP00039311-01 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
Amend K, Hicks D & Ambrosone CB 2006 Breast cancer in African–American women: differences in tumor biology from


Received in final form 29 August 2014
Accepted 16 September 2014
Made available online as an Accepted Preprint 16 September 2014