Replication of Breast Cancer GWAS Susceptibility Loci in the Women's Health Initiative African American SHARE Study

Carolyn M. Hutter1, Alicia M. Young1, Heather M. Ochs-Balcom2, Cara L. Carty1, Tao Wang3, Christina T.L. Chen1, Thomas E. Rohan3, Charles Kooperberg1, and Ulrike Peters1

Abstract

**Background:** Genome-wide association studies (GWAS) have identified loci associated with risk of breast cancer. These studies have primarily been conducted in populations of European descent. To fully understand the impact of these loci, it is important to study groups with other genetic ancestries, including African American women.

**Methods:** We examined 22 single-nucleotide polymorphisms (SNP), previously identified in GWAS of breast cancer risk in European and Asian descent women (index SNPs), and SNPs in the surrounding regions in a study of 7,800 African American women (including 316 women with incident invasive breast cancer) from the Women’s Health Initiative SNP Health Association Resource.

**Results:** Two index SNPs were associated with breast cancer: rs3803662 at 16q12.2/TOX3 (Hazard ratio [HR] for the T allele = 0.79, 95% CI: 0.67–0.92, P = 0.003) and rs10941679 at 5p12 (HR for the G allele = 1.31, 95% CI: 1.06–1.63, P = 0.014). When we expanded to regions, the 3p24.1 region showed an association with breast cancer risk (permutation based P = 0.027) and three regions (10p15.1, 10q26.13/FGFR2, and 16q12.2/TOX3) showed a trend toward association.

**Conclusion:** Our findings provide evidence that some breast cancer GWAS regions may be associated with breast cancer in African American women. Larger, more comprehensive studies are needed to fully assess generalizability of published GWAS findings and to identify potential novel associations in African American populations.

**Impact:** Both replication and lack of replication of published GWAS findings in other ancestral groups provides important information of the genetic etiology of this disease and may impact translation of GWAS findings to clinical and public health settings. Cancer Epidemiol Biomarkers Prev; 20(9); 1950–9. ©2011 AACR.

Introduction

African American women have a lower age-adjusted incidence of breast cancer than white women in the United States. Age-adjusted annual incidence rates for 2002 to 2006 were 123.5 cases per 100,000 for white women and 113.0 cases per 100,000 for African American women (1). However, African American women are more likely to be diagnosed with breast cancer at a more advanced stage and have higher breast cancer mortality rates than white women. Age-adjusted mortality rates for 2002 to 2006 were 23.9 per 100,000 for white women and 33.0 per 100,000 for African American women. The role of environmental risk factors in explaining these disparities was investigated within the Women’s Health Initiative (WHI; ref. 2), and the differences in incidence between African American and white women do not seem to be fully explained by differences in established risk factors.

Understanding genetic risk factors in relation to breast cancer is important because identifying such factors might be useful for risk prediction, development of chemopreventive agents, and other preventive measures. First-degree relatives of women with breast cancer have approximately twice the risk of developing breast cancer compared with the general population, even after controlling for common environmental exposures (3). Genetic susceptibility to breast cancer stems from 3 general classes of alleles (4): very rare high-penetrance alleles (such as...
those in \textit{BRCA1} and \textit{BRCA2}, rare moderate-penetrance alleles (such as \textit{ATM} and \textit{CHK2}), and common low-penetrance alleles. The latter category are the types of alleles identified in genome-wide association studies (GWAS); specifically, this includes alleles with population frequencies above 5\% and relative risks of 1.05–1.3 (5, 6). GWAS published to date have successfully identified over 20 single-nucleotide polymorphisms (SNP) showing genome-wide significant associations with breast cancer risk (7–16). These studies were primarily conducted in populations of European descent, although two focused on populations of Asian descent (11, 16). Some of these variants have previously been examined in Chinese (17), Hispanic (18), and African American populations (19–24); however, the results are inconsistent and not all loci have been examined. For these reasons, additional replication is merited.

In the context of this article, we refer to the variants identified in the initial GWAS as “index SNPs.” These GWAS SNPs were identified because they showed a strong statistical association with disease risk in the discovery population. However, such SNPs are often not known to be the functional variants underlying the disease. Instead, the index SNPs are in linkage disequilibrium (LD) with other variants and can be thought of as “tagging” or identifying particular chromosomal regions of interest, with the functional variant potentially being located somewhere in that region. Because of differences in LD patterns according to genetic ancestry, an index SNP identified in studies including individuals of European descent may not be in high LD with the functional variant in other populations (e.g., African Americans). In such cases, the specific index SNP may not show evidence for replication in African Americans; however, other SNPs in the region may be in LD with the functional variant and, hence, further characterize associations with particular genomic regions. Therefore, a full exploration of potential replication/generalizability of GWAS findings in other racial/ethnic groups requires looking not only at the index SNP but also examining, if possible, the entire region tagged by the index SNP.

By using GWAS data from the WHI, we sought to replicate known GWAS findings for breast cancer in a cohort of postmenopausal African American women. Because of differences in LD patterns based on genetic ancestry, we examined associations for the index SNPs reported in the original GWAS and also for SNPs in regions defined by LD around the index SNPs.

\textbf{Methods}

\textbf{Study population}

The WHI is a long-term national health study that focuses on understanding risk factors for common diseases such as heart disease, cancer, and fracture in postmenopausal women. A total of 161,838 women aged 50 to 79 years old were recruited from 40 clinical centers in the United States between 1993 and 1998. WHI consists of an observational study, 2 clinical trials of postmenopausal hormone therapy (estrogen alone and estrogen plus progesterone), a calcium and vitamin D supplement trial, and a dietary modification trial (25). Study recruitment and exclusion criteria have been described previously (26). Study protocols and consent forms were approved by the Institutional Review Boards at all participating institutions.

Medical history was updated annually (for women in the observational study) or semiannually (for women in the clinical trials) by mail and/or telephone questionnaires. Breast cancers were verified by medical record and pathology report review by centrally trained WHI physician adjudicators, as described previously (27, 28).

The WHI SNP Health Association Resource (SHARe) includes 8,515 self-identified African American women from WHI who provided consent for DNA analysis. We excluded subjects on the basis of the genotype failure and quality control (\(n = 94\), relatedness (\(n = 209\)), and genetic ancestry (described below; \(n = 57\)), as well as subjects with noninvasive breast cancer (\(n = 91\)), and subjects with report of prevalent breast cancer at baseline (\(n = 264\)). Breast cancer cases were defined as cases with incident invasive breast cancer, confirmed by central adjudication. Our final sample size was 7,800 women, 316 of whom had incident invasive breast cancer.

\textbf{Genotyping and QC}

DNA was extracted from blood specimens collected at time of WHI enrollment. All samples, plus 2\% blinded duplicates, were genotyped at Affymetrix Inc on the Genome-wide Human SNP Array 6.0 (909,622 SNPs). Approximately 1\% of samples failed genotyping; we further excluded samples with call rate less than 95\%, unexpected duplicates, and samples with genotype calls on the Y chromosome. We used concordance information to identify relatives (parent-offspring, twins, siblings, and half-siblings) and only included the sample with the highest call rate for each identified family set (\(n = 266\) exclusions). SNPs were excluded if they were located on the Y chromosome, were Affymetrix QC probes (total \(n = 3280\)), had a call rate less than 95\%, or had concordance rates for duplicates less than 98\%. The average concordance for blinded duplicate samples was 99.8\% and the average sample call rate after SNP exclusions was 99.8\%.

Imputation for African Americans was carried out by using MACH (29). After filtering, 829,370 genotyped SNPs were used for imputation. We used 2,203,609 SNPs in HapMap 2, release 22, from 240 phased haplotypes for the HapMap Yoruba in Ibadan, Nigeria (YRI) population and the HapMap Utah residents with Northern and Western European ancestry (CEPH) collection (CEU) populations as the reference panel. We estimated parameters on a subset of 200 WHI subjects and then imputed all African American subjects. For 2,190,779 SNPs, we obtained imputations with minor allele frequency (MAF) greater than 1\% and estimated R-squared more than 0.3.
Genetic ancestry was calculated by using EIGENSTRAT (30). Specifically, we obtained principal components by using 178,101 SNP markers that were common between our samples and our reference panels, comprising 475 publically available samples from the YRI population, the CEU population, the Human Genome Diversity Project (HGDP) East Asian population, and the HGDP Native American populations. These same samples were used to determine ancestral percentages by using Frappe (31). We excluded 57 samples that were outliers in the Frappe analysis.

Selection of SNPs and regions for replication of previous findings

Breast cancer loci from previous GWAS, which we term as “index SNPs,” were identified by using the NHGRI catalog (5, 32) using a P value cutoff of $5 \times 10^{-8}$ and a requirement that the initial GWAS have a minimum of 100 cases and controls, with report of independent replication. We last accessed the catalogue on March 1, 2011. We did not include SNPs identified in GWAS restricted to BRCA1 or BRCA2 carriers. In addition to SNPs identified through the NHGRI catalogue, we included 3 SNPs (rs4973768 at 3p24.1, rs10941679 at 5p12, and rs6049590 at 17q23.2) as index SNPs because these SNPs fulfilled our criteria (identification through GWAS and combined GWAS and large-scale replication resulted in P value of less than $5 \times 10^{-8}$; ref. 7, 13). At this stage, we did not screen SNPs based on LD, so some of these index SNPs are in high LD with one another. All SNPs were either genotyped directly or imputed in our data except for rs999737. This SNP has a MAF of $<2\%$ in HapMap YRI and HapMap African ancestry in South-west USA (ASW), so presumably was excluded from our sample because of low MAF. We used a second SNP in high LD in the CEU population with rs999737 (rs10483813) as a substitute for that SNP. The rs999737 and rs10483813 SNPs are 3,398 base pairs apart with a pairwise $r^2 = 1$ in the CEU population.

The index SNPs tag a region defined by LD in the population used in the initial GWAS. Because we are studying a population with a different genetic ancestry, and because groups with different ancestries may have different haplotype patterns, we chose to examine both the index SNP and SNPs in the surrounding region. Specifically, we considered the situation in which the underlying causal variant is in the region defined by high LD, with the index SNP in the discovery population, but is not in high LD with the index SNP in African American women. In these situations, we would not see replication of the index SNP but, potentially, we might expect to see association for other SNPs in the region. Therefore, we defined regions for the index SNP by using LD information in HapMap. Specifically, we used HapMap data to find the most distant SNP upstream and downstream with an LD $r^2 > 0.8$ within a maximum distance of 250 kb in either direction. We defined the “region” to include all genotyped and imputed SNPs between these boundaries, regardless of their LD with the index SNP. Regions were defined using CEU for all SNPs, except rs2046210 and rs4784227. These 2 SNPs were initially discovered in samples of Asian, rather than European descent, so we used the HapMap Han Chinese in Beijing, China (CHB) population to define regions for those SNPs. LD information was obtained by using the Genome Variation Server in batch mode (33). Because the regions are defined on the basis of LD patterns, some regions contain more than one index SNP. Furthermore, some index SNPs had no SNPs with $r^2 > 0.8$ in the HapMap population and, hence, are not included in the regional analysis. The final regional analysis used 839 SNPs in 14 regions (median of 34.5 SNPs per region; range 13–188).

Statistical methods

Cox proportional hazards models were used to assess associations between each SNP and breast cancer with time because enrollment as our time axis, adjusting for age, region, and the first 4 principal components representing global ancestry. As a sensitivity analysis, we further adjusted for randomization assignment within the WHI trial arms, including an indicator for the observational study participants. We used a log-additive genetic model: for directly genotyped SNPs, we used the SNP data coded 0/1/2, and for the imputed SNPs, we used the dosage data from MACH. For all SNPs, the major allele was used as the reference. We first examined the index SNP and then the region around each SNP as described above. Within each region, we report on the following: (a) the total number of SNPs in the region; (b) the number of SNPs in the region with $P < 0.05$; (c) the HR, 95% CI, and $P$ value for the SNP with the lowest $P$ value in the region; and (d) a permutation $P$ value for the region. Permutation $P$ values were calculated by 10,000 permutations per region. In each iteration, we permuted the outcome and ran the adjusted Cox model to obtain the $P$ value for each SNP in the region. Then we obtained the minimum $P$ value among all SNPs within the region for each permutation, counted the number of times the minimum $P$ value for the region was less than the observed minimum $P$ value for the region in our analysis, and divided that count by 10,000.

We created regional association plots (34) to visually display the $-\log_{10} (P)$ and LD with the index SNP by chromosomesal location for regions of interest. For these plots, LD was examined on the basis of HapMap CEU and YRI populations. We did not calculate LD for imputed SNPs, as it is not straightforward to obtain unbiased estimates of LD on the basis of imputed data.

Results

The median follow-up time in the cohort was 7.94 years. As expected, cases were slightly younger than the controls (61 vs. 62 years) and more likely to have a positive family history of breast cancer (first-degree...
The results for the 22 index SNPs identified in previous GWAS are shown in Table 1. These 22 SNPs are in 18 independent genomic regions, with independent defined on the basis of LD in the CEU population. The strongest evidence for an association in African Americans was for SNP rs3803662 at 16q12.2/TOX3 (HR for the T allele = 0.79, 95% CI: 0.67–0.92, P = 0.003). A second SNP rs10941679 at 5p12/MRPS30 was also significant at P = 0.05 (HR for the G allele = 1.31, 95% CI: 1.06–1.63, P = 0.014) and rs1219648 at 10q26.13/FGFR2 showed marginal significance (HR for the G allele 1.17, 95% CI: 1.00–1.37, P = 0.051). No index SNP was significant after a Bonferroni correction for multiple testing. Additional adjustment for randomization assignment had little to no effect on the risk estimates (correlation of HR with and without this additional adjustment = 0.998).

Of the 18 potentially independent regions determined by LD patterns around the index SNPs in the GWAS discovery population, 4 regions did not have any SNPs with r² > 0.8 in the CEU HapMap sample, leaving 14 regions for analysis (Table 2). Results for all 839 SNPs in each region are shown in Supplementary Table. Eight of 14 regions had at least 1 SNP with P < 0.05, with 5 regions having more than 10% of the SNPs in the region with a P < 0.05. When we examined the permutation results, done to account for multiple testing in this region-wide approach, we found that the 3p24.1 region showed a significant positive association with breast cancer risk (permutation based P = 0.027) and 3 regions (10p15.1, 10q26.13/FGFR2, and 16q12.2/TOX3) showed suggestive associations (defined as P < 0.1). Plots of regional LD and strength of association by chromosomal position for these 4 regions of interest are shown in Figure 1.

Discussion

We used GWAS data from almost 8,000 AA women of the WHI SHARE project to investigate whether SNPs found to be associated with breast cancer in GWAS of European and Asian decent women replicate and generalize to a population of postmenopausal African American women. For the previously reported index SNP, we found evidence for an association with rs3803662 at 16q12.2/TOX3 and rs10941679 at 5p12/MRPS30 in the WHI African American sample; however, these findings were not significant after Bonferroni correction. When we expanded to LD regions around the index SNPs, variants in the 3p24.1 region showed a significant association with breast cancer risk and the 10p15.1, 10q26.13/FGFR2, and 16q12.2/TOX3 regions showed suggestive associations.

Our findings contribute to a growing body of literature examining known GWAS loci in African American women (19–24). Focusing on index SNPs, our most statistically significant finding was for rs3803662 at 16q12.2/TOX3 gene, with the T allele associated with a decreased breast cancer risk in African Americans. This is in contrast to the initial GWAS finding of increased risk for the T allele in European decent women. This region had a potential functional link to breast cancer. TOX3 is a calcium-dependent transcription factor (35). This protein may play a role in estrogen-dependent signal transduction and enhance survival of breast cancer cells (36). To summarize results for African American, we conducted a meta-analysis of previously published studies, including our results (Fig. 2; refs. 12, 20, 22, 24, 37). The meta-analysis showed that risk estimates for African American women, although suggestive of a decreased risk for the T allele, are heterogeneous and not statistically significant in the random effect meta-analysis (OR = 0.92; 95% CI: 0.82–1.03). A potential explanation for the heterogeneity between studies is the genetic variation among African Americans (38, 39). Differences in the underlying population composition of the different studies may impact the observed effect estimates for this SNP. Potential explanations for the difference between the findings in European and African American populations include potential effect modification by other genetic or environmental risk factors, differences in haplotype tagging patterns, or chance.

For the index SNP rs10941679 in the 5p12 region, we found evidence for an increased risk for the G allele, consistent with the original GWAS finding in European populations (13). Analysis of the Black Woman’s Health Study and the African American sample from the Multi-ethnic Cohort Study both also found a nonsignificant trend for increased risk with the G allele at this SNP (13, 20, 21). Similarly, our marginal finding for a trend for an increased risk for the G allele at rs1219648 at 10q26.1/FGFR2 is consistent in direction, both with the original GWAS finding (10) and with combined results for African Americans from the Southern Community Cohort Study and the Nashville Breast Health Study (24). As a receptor tyrosine kinase, the FGFR2 protein is involved in cell signaling pathways (40). This protein is known to have a role in breast tissue development (41, 42) and has been shown to have nuclear localization in breast normal and tumor tissue (43).

Comparing our regional findings to results from previous studies, for the 3p24.1 region, we did not find evidence for an association for the index SNP rs4973768, a finding that is consistent with a recent report on African Americans in the Multi-ethnic Cohort Study (20); however, that study did not look at other SNPs in the region, so we cannot compare our finding of a significant association in the region with their data. To date, no other study has reported on other SNPs in the 3p24.1 region in African American women. Our results for the 10q26.1/FGFR2 region are consistent with those of several other studies, which have found that additional SNPs in this region are associated with breast cancer risk in African American women (19, 23). For the 16q12.2/TOX3 region, the Black Women’s Health Study found evidence for an association with breast cancer risk for the index SNP and also for 4 other SNPs in the neighboring LOC643714 gene.
Table 1. Association of susceptibility loci from previous GWAS with breast cancer risk in African American women in WHI

<table>
<thead>
<tr>
<th>Chromosomal Position</th>
<th>GWAS reference</th>
<th>SNP Position</th>
<th>Alleles (^a)</th>
<th>MAF</th>
<th>Imputed</th>
<th>Imputation (r^2) (^b)</th>
<th>Discovery GWAS OR (^c)</th>
<th>HR (95% CI)</th>
<th>(P^d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1p11.2</td>
<td>14</td>
<td>rs11249433</td>
<td>120982136</td>
<td>A/G</td>
<td>0.16</td>
<td>Yes</td>
<td>0.6841</td>
<td>1.16</td>
<td>1.08 (0.83-1.41)</td>
</tr>
<tr>
<td>2q35</td>
<td>12, 14, 15</td>
<td>rs13387042</td>
<td>217614077</td>
<td>A/G</td>
<td>0.27</td>
<td>Yes</td>
<td>0.9925</td>
<td>0.80</td>
<td>1.07 (0.89-1.27)</td>
</tr>
<tr>
<td>3p24.1</td>
<td>7</td>
<td>rs4973768</td>
<td>27391017</td>
<td>C/T</td>
<td>0.21</td>
<td>Yes</td>
<td>0.9445</td>
<td>1.16</td>
<td>1.02 (0.86-1.20)</td>
</tr>
<tr>
<td>5p12</td>
<td>13</td>
<td>rs10941679</td>
<td>44742255</td>
<td>A/G</td>
<td>0.37</td>
<td>Yes</td>
<td>0.9562</td>
<td>1.19</td>
<td>1.31 (1.06-1.63)</td>
</tr>
<tr>
<td>5q11.2</td>
<td>8, 15</td>
<td>rs16886165</td>
<td>56058840</td>
<td>T/G</td>
<td>0.33</td>
<td>Yes</td>
<td>0.9688</td>
<td>1.23</td>
<td>1.05 (0.89-1.25)</td>
</tr>
<tr>
<td>5q11.2</td>
<td>14</td>
<td>rs890312</td>
<td>5607641</td>
<td>A/C</td>
<td>0.32</td>
<td>Yes</td>
<td>0.96</td>
<td>1.22</td>
<td>0.98 (0.83-1.46)</td>
</tr>
<tr>
<td>6q22.33</td>
<td>9</td>
<td>rs2180341</td>
<td>127642323</td>
<td>A/G</td>
<td>0.31</td>
<td>Yes</td>
<td>0.9989</td>
<td>1.41</td>
<td>0.98 (0.83-1.47)</td>
</tr>
<tr>
<td>6q25.1</td>
<td>16</td>
<td>rs2046210</td>
<td>151990059</td>
<td>A/G</td>
<td>0.48</td>
<td>No</td>
<td>Genotyped</td>
<td>0.77</td>
<td>1.00 (0.85-1.18)</td>
</tr>
<tr>
<td>8q24.21</td>
<td>8</td>
<td>rs13281615</td>
<td>128424800</td>
<td>G/A</td>
<td>0.44</td>
<td>Yes</td>
<td>0.9176</td>
<td>1.08</td>
<td>1.16 (0.99-1.37)</td>
</tr>
<tr>
<td>9p21.3</td>
<td>15</td>
<td>rs1011970</td>
<td>22052134</td>
<td>G/T</td>
<td>0.32</td>
<td>Yes</td>
<td>0.9398</td>
<td>1.09</td>
<td>1.07 (0.90-1.28)</td>
</tr>
<tr>
<td>10p15.1</td>
<td>15</td>
<td>rs2380205</td>
<td>5926740</td>
<td>C/T</td>
<td>0.49</td>
<td>No</td>
<td>Genotyped</td>
<td>0.94</td>
<td>0.98 (0.84-1.16)</td>
</tr>
<tr>
<td>10q21.2</td>
<td>15</td>
<td>rs10995190</td>
<td>63948688</td>
<td>G/A</td>
<td>0.16</td>
<td>Yes</td>
<td>0.9754</td>
<td>0.86</td>
<td>1.12 (0.89-1.41)</td>
</tr>
<tr>
<td>10q26.13</td>
<td>10</td>
<td>rs704010</td>
<td>80511154</td>
<td>A/G</td>
<td>0.19</td>
<td>No</td>
<td>Genotyped</td>
<td>0.93</td>
<td>1.11 (0.84-1.47)</td>
</tr>
<tr>
<td>10q26.13</td>
<td>14, 15</td>
<td>rs2981579</td>
<td>123327325</td>
<td>G/A</td>
<td>0.45</td>
<td>Yes</td>
<td>0.8214</td>
<td>1.17</td>
<td>0.99 (0.83-1.17)</td>
</tr>
<tr>
<td>10q26.13</td>
<td>8</td>
<td>rs1219648</td>
<td>123336180</td>
<td>A/G</td>
<td>0.41</td>
<td>No</td>
<td>Genotyped</td>
<td>1.20</td>
<td>1.17 (1.00-1.37)</td>
</tr>
<tr>
<td>10p22.3</td>
<td>15</td>
<td>rs2981582</td>
<td>123342307</td>
<td>A/G</td>
<td>0.45</td>
<td>No</td>
<td>Genotyped</td>
<td>1.26</td>
<td>0.93 (0.79-1.09)</td>
</tr>
<tr>
<td>11q13.1</td>
<td>8</td>
<td>rs3817198</td>
<td>1865582</td>
<td>A/G</td>
<td>0.19</td>
<td>No</td>
<td>Genotyped</td>
<td>1.07</td>
<td>1.10 (0.91-1.34)</td>
</tr>
<tr>
<td>11p15.5</td>
<td>8</td>
<td>rs614367</td>
<td>69037945</td>
<td>A/G</td>
<td>0.12</td>
<td>No</td>
<td>Genotyped</td>
<td>1.07</td>
<td>1.13 (0.89-1.44)</td>
</tr>
<tr>
<td>14q24.1</td>
<td>14</td>
<td>rs10483819</td>
<td>68100137</td>
<td>T/A</td>
<td>0.10</td>
<td>Yes</td>
<td>0.7915</td>
<td>1.06</td>
<td>1.05 (0.78-1.42)</td>
</tr>
<tr>
<td>16q12.1</td>
<td>8, 12, 14, 15</td>
<td>rs3803662</td>
<td>51143842</td>
<td>C/T</td>
<td>0.47</td>
<td>No</td>
<td>Genotyped</td>
<td>1.30</td>
<td>0.79 (0.67-0.92)</td>
</tr>
<tr>
<td>16q12.1</td>
<td>11</td>
<td>rs4784227</td>
<td>51156689</td>
<td>C/T</td>
<td>0.15</td>
<td>No</td>
<td>Genotyped</td>
<td>1.24</td>
<td>1.23 (0.94-1.61)</td>
</tr>
<tr>
<td>17q23.2</td>
<td>7</td>
<td>rs6504950</td>
<td>50411470</td>
<td>G/A</td>
<td>0.35</td>
<td>Yes</td>
<td>0.9930</td>
<td>0.95</td>
<td>1.06 (0.90-1.28)</td>
</tr>
</tbody>
</table>

*aListed as reference allele/tested allele.

*bImputation \(r^2\) is a measure of imputation quality from MACH.

*cORs for discovery GWAS in this column are presented to have the same reference allele/tested allele as HR in this analysis (i.e., they are expressed in terms of the risk for the test allele compared with the reference allele. In some cases, this required changing the reference allele from what was presented in the original presentation. All discovery GWAS had \(P < 5 \times 10^{-7}\).

*d\(P\) from test of trend for each additional copy of the test allele.
Table 2. Association of SNPs in LD region around index SNP with breast cancer risk in African American women in WHI

| Region  | Region boundaries | Total # SNPs in region | # SNPs in region with \( P < 0.05 \) | SNPs in region with \( P < 0.05 \), % | SNP with lowest \( P \) in region | Permutation \( P \)
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2q35</td>
<td>217614024–217629014</td>
<td>21</td>
<td>0</td>
<td>0.0</td>
<td>rs10490444 0.90 (0.76–1.05) 0.18</td>
<td>0.87</td>
</tr>
<tr>
<td>3p24.1</td>
<td>27216223–27392003</td>
<td>165</td>
<td>66</td>
<td>40.0</td>
<td>rs9843350 1.58 (1.22–2.05) 0.0005</td>
<td>0.027</td>
</tr>
<tr>
<td>5q11.2</td>
<td>50644769–50660000</td>
<td>15</td>
<td>2</td>
<td>13.3</td>
<td>rs6862199 1.21 (1.02–1.43) 0.032</td>
<td>0.31</td>
</tr>
<tr>
<td>6q22.33</td>
<td>127587185–127752866</td>
<td>110</td>
<td>0</td>
<td>0.0</td>
<td>rs11759744 1.27 (0.99–1.62) 0.061</td>
<td>0.73</td>
</tr>
<tr>
<td>6q25.1</td>
<td>151978370–152013380</td>
<td>47</td>
<td>0</td>
<td>0.0</td>
<td>rs6930633 1.20 (1.00–1.48) 0.050</td>
<td>0.71</td>
</tr>
<tr>
<td>8q24.21</td>
<td>128413061–128448145</td>
<td>66</td>
<td>5</td>
<td>7.6</td>
<td>rs13262406 1.27 (1.06–1.51) 0.0091</td>
<td>0.18</td>
</tr>
<tr>
<td>10p15.1</td>
<td>5926155–5944924</td>
<td>29</td>
<td>15</td>
<td>51.7</td>
<td>rs7894083 1.29 (1.07–1.55) 0.0063</td>
<td>0.079</td>
</tr>
<tr>
<td>10q21.2</td>
<td>63945267–63958395</td>
<td>17</td>
<td>0</td>
<td>0.0</td>
<td>rs7082733 1.11 (0.94–1.29) 0.21</td>
<td>0.73</td>
</tr>
<tr>
<td>10q26.13</td>
<td>123325790–123342307</td>
<td>17</td>
<td>3</td>
<td>17.6</td>
<td>rs1219643 0.79 (0.67–0.94) 0.0078</td>
<td>0.084</td>
</tr>
<tr>
<td>10p22.3</td>
<td>80515417–80525429</td>
<td>27</td>
<td>0</td>
<td>0.0</td>
<td>rs1250002 1.17 (1.00–1.37) 0.058</td>
<td>0.50</td>
</tr>
<tr>
<td>11q13.1</td>
<td>69037693–69041851</td>
<td>13</td>
<td>0</td>
<td>0.0</td>
<td>rs634300 0.87 (0.71–1.06) 0.17</td>
<td>0.74</td>
</tr>
<tr>
<td>14q24.1</td>
<td>68049588–68101073</td>
<td>84</td>
<td>2</td>
<td>2.4</td>
<td>rs2842324 1.47 (1.11–1.94) 0.0079</td>
<td>0.29</td>
</tr>
<tr>
<td>16q12.1</td>
<td>51095541–51156689</td>
<td>40</td>
<td>5</td>
<td>12.5</td>
<td>rs3006366 0.79 (0.67–0.92) 0.0034</td>
<td>0.074</td>
</tr>
<tr>
<td>17q23.2</td>
<td>50383225–50604678</td>
<td>188</td>
<td>10</td>
<td>5.3</td>
<td>rs10515089 1.42 (1.06–1.91) 0.019</td>
<td>0.57</td>
</tr>
</tbody>
</table>

*Regions 1p11.2, 5p12, 9p21.3 and 11p15.5 are not included, because there were no SNPs with \( r^2 > 0.8 \) for those SNPs in reference HapMap population.

*Regions defined on the basis of LD in HapMap CEU or CHB.

*\( P \) from test of trend for each additional copy of the test allele.

*\( P \) for region on the basis of 10,000 permutations of phenotype.
at 16q12 (rs3104746, rs3112562, rs3104793, and rs8046994; ref. 22). These SNPs were outside of our defined regions of interest. However, we were able to examine the SNPs from our GWAS data and found marginal evidence for an association for rs3112562 (HR for the C allele: 1.15; 95% CI: 0.98–1.35; \( P = 0.076 \)). Our final interesting region, 10p15.1, has an index SNP that was identified in a more recent meta-analysis (15) and has not been included in other studies of African American women published to date. Our findings indicate that future studies should examine not just the index SNP but also additional SNPs in the region if attempting to replicate 10p15.1 in African American women.

As discussed above, we have several examples in which we did not observe statistical evidence for replication for the index SNP but did observe statistical evidence...
for association for other SNPs in the region. This may be a chance finding, although we did carry out permutation tests to account for the multiple testing involved in looking at additional SNPs within each region. It is possible that the findings reflect difference in LD patterns on the basis of genetic ancestry (i.e., the underlying causal variant is the same for European and African ancestry, but different SNPs tag the variant in different groups). This is the situation in which the index risk variant may be in high LD, with the functional variant in the GWAS discovery population (European ancestry), but not in high LD in the African American population used in this study. For example, this may explain our findings for the 10p15.1 region, in which the strongest association is for a SNP in high LD with the index SNP in CEU but not in high LD in YRI (Fig. 1). Because LD regions are typically smaller in African American populations, this type of analysis may help narrow the region of interest. For example, our results for the 10p15.1 region, in which the strongest association is for a SNP in high LD with the index SNP in CEU but not in high LD in YRI (Fig. 1). Because LD regions are typically smaller in African American populations, this type of analysis may help narrow the region of interest. For example, our results for the 10p15.1 region, in which the strongest association is for a SNP in high LD with the index SNP in CEU but not in high LD in YRI (Fig. 1). Because LD regions are typically smaller in African American populations, this type of analysis may help narrow the region of interest. For example, our results for the 10p15.1 region, in which the strongest association is for a SNP in high LD with the index SNP in CEU but not in high LD in YRI (Fig. 1). Because LD regions are typically smaller in African American populations, this type of analysis may help narrow the region of interest.

For 14 of the index SNPs, we did not observe statistically significant evidence for replication for either the index SNP or for other SNPs in the region. This may be because we were underpowered to detect the association, especially for lower MAFs. Another possibility is that we did not consider a wide enough region around the index SNP. Our goal in setting boundaries for regions was to capture all SNPs that may have been tagged by the index SNP in the initial GWAS studies. We determined this by using information on LD from the HapMap 2 populations. We may have been too stringent in our LD cutoff, or we may have misestimated the extent of LD because HapMap does not contain data on all SNPs. We initially considered using less stringent cutoffs to define regions but opted not to because of the increased noise and increased multiple testing burden associated with boundaries that are too wide. The lack of replication may also reflect the situation in which variants in the region are associated with risk in African Americans, but those variants were simply not tagged or imputed in our dataset. It is also possible that the effects of the GWAS loci may have been modified by environmental, lifestyle, or other factors that differ among groups. It is worth noting that a recent study that examined potential gene–environment interactions for 7 of the loci examined in this study failed to yield significant evidence for effect modification with established risk factors for breast cancer (44).

A strength of this study is that it is a large cohort study with central adjudication of breast cancer that allowed us
to examine incident invasive breast cancer cases with minimal misclassification of outcome. We were able to leverage existing genome-wide data in this sample to not only look at index SNPs identified in previous GWAS but to also extend our analysis to large LD blocks of surrounding regions. However, even though WHI represents a large cohort of African American women with GWAS data, our study is still limited by the relatively small number of invasive incident breast cancer cases. Given our small sample size, we were not able to carry out stratified analysis on the basis of disease severity, hormone receptor status, or other patient characteristics. Estrogen receptor (ER) status may be particularly important given that some GWAS findings are specific to ER+ and ER− cancers (12, 13, 45) and because a higher proportion of African American are diagnosed with ER− cancers (2, 46), resulting in prognostic differences.

We were able to use imputation to the HapMap to study SNPs that were not directly genotyped on our platform. A key question in imputation for admixed populations is the selection of an appropriate reference panel. We used a combination of the CEU and YRI HapMap populations, which has been shown to be an appropriate approach to use for African American populations (47). The rs11249433 SNP did have a relatively low imputation quality score ($r^2 = 0.68$). Combined with the low MAF, we may have had a highly reduced power to detect that particular SNP. However, all the other imputed SNPs had very high imputation $r^2$ values (Table 1), indicating a high imputation quality. In addition to attention to admixture in the choice of our imputation reference panel, we also used Frappe (31) and EIGENSTRAT (30) to identify ethnic outliers and adjust for underlying population structure. This minimizes the chance that our results are strongly confounded by population stratification (48).

Overall, these results add to a growing body of work indicating that some genetic loci identified as risk factors for breast cancer (17–24) and other cancer phenotypes (49, 50) via GWAS in European populations are generalizable to other ethnic/racial groups, whereas other loci are not. A full understanding of these loci in relation to disease risk will require additional follow-up with detailed fine mapping data in large ancestrally diverse populations. A full characterization of the role of common genetic variants in African American populations will also require large, well-powered GWAS, with replication, to identify potentially novel loci.

### Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

### Acknowledgments

The manuscript for this article was prepared in collaboration with investigators of the WHI and has been approved by the Women’s Health Initiative (WHI). WHI investigators are listed at http://www.whiscience.org/publications/WHI_investigators_shortlist_2010-2010.pdf. The datasets used for the analyses described in this article were obtained from dbGaP at http://www.ncbi.nlm.nih.gov/sites/entrez?db=gap through dbGaP accession phs000200.v3.p1.

### Grant Support

The WHI program is funded by the National Heart, Lung, and Blood Institute, NIH, U.S. Department of Health and Human Services through contracts N01WH22110, 24152, 32109-2, 32105-6, 32109-4, 32111-13, 32115, 32118-32119, 32122, 42107-26, 42129-32, and 44221. C.M. Hutter was funded in part by R25CA094880.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received June 2, 2011; revised July 13, 2011; accepted July 21, 2011; published OnlineFirst July 27, 2011.

### References
