



Published in final edited form as:

Cancer Epidemiol Biomarkers Prev. 2011 September ; 20(9): 1950–1959. doi:
10.1158/1055-9965.EPI-11-0524.

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

¹Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA

²Department of Social and Preventive Medicine, School of Public Health and Health Professions, University at Buffalo, Buffalo, NY

³Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, NY

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 (SNPs) 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* (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-value =0.027) and three regions (10p15.1, 10q26.13/*FGFR2* and 16q12.2/*TOX3*) showed a trend towards 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.

Keywords

Breast Cancer; Genome-wide Association Study; African-American; Epidemiology; Genetic Risk Factors

Introduction

African American women have a lower age-adjusted incidence of breast cancer than white women in the U.S. Age-adjusted annual incidence rates for 2002–2006 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 2002–2006 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) (2) and the differences in incidence between African American and white women do not appear to be fully explained by differences in established risk factors. Variation in inherited genetic risk factors, additional lifestyle and behavioral risk factors, screening and treatment patterns may also influence disparities between these two groups (2).

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 to the general population, even after controlling for common environmental exposures (3). Genetic susceptibility to breast cancer stems from three general classes of alleles (4) : very rare high-penetrance alleles (such as those in *BRCA1* and *BRCA2*), rare moderate-penetrance alleles (such as *ATM* and *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 SNPs 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 paper 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.

Using GWAS data from the Women’s Health Initiative we sought to replicate known GWAS findings for breast cancer in a cohort of post-menopausal 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.

Methods

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 50–79 yrs old were recruited from 40 clinical centers in the US between 1993 and 1998. WHI consists of an observational study, two clinical trials of postmenopausal hormone therapy (estrogen alone and estrogen plus progestin), 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 8515 self-identified African American women from WHI who provided consent for DNA analysis. We excluded subjects based on genotyping failure and quality control (n=94), relatedness (n=209) and genetic ancestry (described below; n=57), as well as subjects with non-invasive 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.

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 <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 <95%, or had concordance rates for duplicates <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 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 >1% and estimated R-squared >0.3.

Genetic ancestry was calculated using EIGENSTRAT (30). Specifically we obtained principal components 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 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 using the NHGRI catalog (5, 32) using a p-value cut-off of 5×10^{-7} and a requirement that the initial GWAS have a minimum of 100 cases and controls, with report of independent replication. We last accessed the catalog 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 catalog, we included three SNPs (rs4973768 at 3p24.1, rs10941679 at 5p12 and rs6504950 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 $< 5 \times 10^{-7}$ (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 $< 2\%$ in HapMap YRI and HapMap African ancestry in Southwest USA (ASW), so presumably was excluded from our sample due to 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 where 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 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 250kb 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 two 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 using the Genome Variation Server in batch mode (33). Because the regions are defined based on LD patterns, some regions contain more than one index SNP. Further, 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 since enrollment as our time axis, adjusting for age, region, and the first four 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: 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. We then 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 $-\log_{10}(p\text{-value})$ and LD with the index SNP by chromosomal location for regions of interest. For these plots LD was examined based on HapMap CEU and YRI populations. We did not calculate LD for imputed SNPs as it is not straightforward to obtain unbiased estimates of LD based on 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 relatives with breast cancer: 22.8% in cases and 15.3% in controls).

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 based on 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, four regions did not have any SNPs with $r^2 > 0.8$ in the CEU HapMap sample, leaving 14 regions for analysis (Table 2). Results for all 839 SNPs in each region are shown in Supplemental Table. Eight of fourteen regions had at least one SNP with p-value < 0.05 , with five regions having $> 10\%$ of the SNPs in the region with a p-value < 0.05 . When we examined the permutation results, performed 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-value =0.027) and three 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 four 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 if SNPs found to be associated with breast cancer in GWAS of European and Asian decent women replicate and generalize to a population of post-menopausal African American women. For the previously reported index SNP we found evidence for an association for 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 (Figure 2) (12, 20, 22, 24, 37). The meta-analysis showed that risk estimates for African American women, while 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 Multiethnic Cohort Study both also found a non-significant 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 Multiethnic 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 to 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 four other SNPs in the neighboring *LOC643714* gene at 16q12 (rs3104746, rs3112562, rs3104793 and rs8046994) (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 where 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 perform 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 based on 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 where 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, where the strongest association is for a SNP in high LD with the index SNP in CEU, but not in high LD in YRI (Figure 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 10q26.13 region are suggestive of the association being localized to the region depicted on the right side of the plot. However, this is not always the case, as exemplified in our results for 3p24 (Figure 1). The lack of replication for the index SNP coupled with observed associations for other SNPs in the regional results could also reflect allelic heterogeneity (i.e., different underlying causal variants) between ancestral groups. Larger sample sizes and functional follow-up studies would be needed to fully distinguish between these different possibilities.

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 minor allele frequencies. 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 using information on LD from the HapMap 2 populations. We may have been too stringent in our LD cut-off, or we may have mis-estimated the extent of LD because HapMap does not contain data on all SNPs. We initially considered using less stringent cut-offs 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 where variants in the region are associated with risk in African Americans, but those variants were simply not well 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 seven 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 perform stratified analysis based on 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, while 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Acknowledgements and Grant Support

The WHI program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services through contracts N01WH22110, 24152, 32100–2, 32105–6, 32108–9, 32111–13, 32115, 32118–32119, 32122, 42107–26, 42129–32, and 44221. This manuscript 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_2005-2010.pdf. The datasets used for the analyses described in this manuscript were obtained from dbGaP at <http://www.ncbi.nlm.nih.gov/sites/entrez?db=gap> through dbGaP accession phs000200.v3.p1. CMH was funded in part by R25CA094880.

Reference List

1. Jemal A, Siegel R, Xu JQ, Ward E. Cancer Statistics, 2010. *Ca-A Cancer Journal for Clinicians*. 2010; 60(5):277–300. [PubMed: 20610543]
2. Chlebowski RT, Chen Z, Anderson GL, Rohan T, Aragaki A, Lane D, et al. Ethnicity and breast cancer: Factors influencing differences in incidence and outcome. *Journal of the National Cancer Institute*. 2005; 97(6):439–448. [PubMed: 15770008]
3. Collaborative Group on Hormonal Factors in Breast Cancer. Familial breast cancer: collaborative reanalysis of individual data from 52 epidemiological studies including 58 209 women with breast cancer and 101 986 women without the disease. *Lancet*. 2001; 358(9291):1389–1399. [PubMed: 11705483]
4. Stratton MR, Rahman N. The emerging landscape of breast cancer susceptibility. *Nat Genet*. 2008; 40(1):17–22. [PubMed: 18163131]

5. Hindorf LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proceedings of the National Academy of Sciences of the United States of America*. 2009; 106(23): 9362–9367. [PubMed: 19474294]
6. Manolio TA. Genomewide Association Studies and Assessment of the Risk of Disease. *New England Journal of Medicine*. 2010; 363(2):166–176. [PubMed: 20647212]
7. Ahmed S, Thomas G, Ghossaini M, Healey CS, Humphreys MK, Platte R, et al. Newly discovered breast cancer susceptibility loci on 3p24 and 17q23.2. *Nature Genetics*. 2009; 41(5):585–590. [PubMed: 19330027]
8. Easton DF, Pooley KA, Dunning AM, Pharoah PDP, Thompson D, Ballinger DG, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature*. 2007; 447(7148) 1087-U7.
9. Gold B, Kirchoff T, Stefanov S, Lautenberger J, Viale A, Garber J, et al. Genome-wide association study provides evidence for a breast cancer risk locus at 6q22-33. *Proceedings of the National Academy of Sciences of the United States of America*. 2008; 105(11):4340–4345. [PubMed: 18326623]
10. Hunter DJ, Kraft P, Jacobs KB, Cox DG, Yeager M, Hankinson SE, et al. A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nature Genetics*. 2007; 39(7):870–874. [PubMed: 17529973]
11. Long JR, Cai QY, Shu XO, Qu SM, Li C, Zheng Y, et al. Identification of a Functional Genetic Variant at 16q12.1 for Breast Cancer Risk: Results from the Asia Breast Cancer Consortium. *Plos Genetics*. 2010; 6(6)
12. Stacey SN, Manolescu A, Sulem P, Rafnar T, Gudmundsson J, Gudjonsson SA, et al. Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. *Nature Genetics*. 2007; 39(7):865–869. [PubMed: 17529974]
13. Stacey SN, Manolescu A, Sulem P, Thorlacius S, Gudjonsson SA, Jonsson GF, et al. Common variants on chromosome 5p12 confer susceptibility to estrogen receptor-positive breast cancer. *Nature Genetics*. 2008; 40(6):703–706. [PubMed: 18438407]
14. Thomas G, Jacobs KB, Kraft P, Yeager M, Wacholder S, Cox DG, et al. A multistage genome-wide association study in breast cancer identifies two new risk alleles at 1p11.2 and 14q24.1 (RAD51L1). *Nature Genetics*. 2009; 41(5):579–584. [PubMed: 19330030]
15. Turnbull C, Ahmed S, Morrison J, Pernet D, Renwick A, Maranian M, et al. Genome-wide association study identifies five new breast cancer susceptibility loci. *Nature Genetics*. 2010; 42(6) 504-U47.
16. Zheng W, Long JR, Gao YT, Li C, Zheng Y, Xiang YB, et al. Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. *Nature Genetics*. 2009; 41(3):324–328. [PubMed: 19219042]
17. Long JR, Shu XO, Cai QY, Gao YT, Zheng Y, Li GL, et al. Evaluation of Breast Cancer Susceptibility Loci in Chinese Women. *Cancer Epidemiology Biomarkers & Prevention*. 2010; 19(9):2357–2365.
18. Slattery ML, Baumgartner KB, Giuliano AR, Byers T, Herrick JS, Wolff RK. Replication of five GWAS-identified loci and breast cancer risk among Hispanic and non-Hispanic white women living in the Southwestern United States. *Breast Cancer Research and Treatment*. 2011 Apr 8. [Epub ahead of print].
19. Barnholtz-Sloan JS, Shetty PB, Guan XW, Nyante SJ, Luo JC, Brennan DJ, et al. FGFR2 and other loci identified in genome-wide association studies are associated with breast cancer in African-American and younger women. *Carcinogenesis*. 2010; 31(8):1417–1423. [PubMed: 20554749]
20. Chen F, Stram DO, Le Marchand L, Monroe KR, Kolonel LN, Henderson BE, et al. Caution in generalizing known genetic risk markers for breast cancer across all ethnic/racial populations. *European Journal of Human Genetics*. 2011; 19:243–245. [PubMed: 21102626]
21. Ruiz-Narvaez EA, Rosenberg L, Rotimi CN, Cupples LA, Boggs DA, Adeyemo A, et al. Genetic variants on chromosome 5p12 are associated with risk of breast cancer in African American women: the Black Women's Health Study. *Breast Cancer Research and Treatment*. 2010; 123(2): 525–530. [PubMed: 20140701]

22. Ruiz-Narvaez EA, Rosenberg L, Cozier YC, Cupples LA, Adams-Campbell LL, Palmer JR. Polymorphisms in the TOX3/LOC643714 Locus and Risk of Breast Cancer in African-American Women. *Cancer Epidemiology Biomarkers & Prevention*. 2010; 19(5):1320–1327.
23. Udler MS, Meyer KB, Pooley KA, Karlins E, Struewing JP, Zhang J, et al. FGFR2 variants and breast cancer risk: fine-scale mapping using African American studies and analysis of chromatin conformation. *Human Molecular Genetics*. 2009; 18(9):1692–1703. [PubMed: 19223389]
24. Zheng W, Cai QY, Signorello LB, Long JR, Hargreaves MK, Deming SL, et al. Evaluation of 11 Breast Cancer Susceptibility Loci in African-American Women. *Cancer Epidemiology Biomarkers & Prevention*. 2009; 18(10):2761–2764.
25. Prentice RL, Anderson G, Cummings S, Freedman LS, Furberg C, Henderson M, et al. Design of the Women's Health Initiative Clinical Trial and Observational Study. *Controlled Clinical Trials*. 1998; 19(1):61–109. [PubMed: 9492970]
26. Hays J, Hunt JR, Hubbell FA, Anderson GL, Limacher M, Allen C, et al. The Women's Health Initiative recruitment methods and results. *Annals of Epidemiology*. 2003; 13(9):S18–S77. [PubMed: 14575939]
27. Chlebowski RT, Hendrix SL, Langer RD, Stefanick ML, Gass M, Lane D, et al. Influence of estrogen plus progestin on breast, cancer and mammography in healthy postmenopausal women - The Women's Health Initiative Randomized trial. *Jama-Journal of the American Medical Association*. 2003; 289(24):3243–3253.
28. Curb JD, McTiernan A, Heckbert SR, Kooperberg C, Stanford J, Nevitt M, et al. Outcomes ascertainment and adjudication methods in the Women's Health Initiative. *Annals of Epidemiology*. 2003; 13(9):S122–S128. [PubMed: 14575944]
29. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet Epidemiol*. 2010; 34(8):816–834. [PubMed: 21058334]
30. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*. 2006; 38(8):904–909. [PubMed: 16862161]
31. Tang H, Peng J, Wang P, Risch NJ. Estimation of individual admixture: Analytical and study design considerations. *Genetic Epidemiology*. 2005; 28(4):289–301. [PubMed: 15712363]
32. Hindorf, LA.; Junkins, HA.; Hall, PN.; Mehta, JP.; Manolio, TA. [2011 February 1] A Catalog of Published Genome-Wide Association Studies. Available at: www.genome.gov/gwastudies.
33. Genome Variation Server. 2010. Available at <http://gvsbatch.gs.washington.edu/GVSBatch/>.
34. Saxena R, Voight BF, Lyssenko V, Burtt NP, de Bakker PIW, Chen H, et al. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science*. 2007; 316(5829):1331–1336. [PubMed: 17463246]
35. Yuan SH, Qiu ZL, Ghosh A. TOX3 regulates calcium-dependent transcription in neurons. *Proceedings of the National Academy of Sciences of the United States of America*. 2009; 106(8):2909–2914. [PubMed: 19196971]
36. Dittmer S, Kovacs Z, Yuan SH, Siszler G, Kogl M, Summer H, et al. TOX3 is a neuronal survival factor that induces transcription depending on the presence of CITED1 or phosphorylated CREB in the transcriptionally active complex. *Journal of Cell Science*. 2011; 124(2):252–260. [PubMed: 21172805]
37. Udler MS, Ahmed S, Healey CS, Meyer K, Struewing J, Maranian M, et al. Fine scale mapping of the breast cancer 16q12 locus. *Human Molecular Genetics*. 2010; 19(12):2507–2515. [PubMed: 20332101]
38. Bryc K, Auton A, Nelson MR, Oksenberg JR, Hauser SL, Williams S, et al. Genome-wide patterns of population structure and admixture in West Africans and African Americans. *Proceedings of the National Academy of Sciences of the United States of America*. 2010; 107(2):786–791. [PubMed: 20080753]
39. Tishkoff SA, Reed FA, Friedlaender FR, Ehret C, Ranciaro A, Froment A, et al. The Genetic Structure and History of Africans and African Americans. *Science*. 2009; 324(5930):1035–1044. [PubMed: 19407144]

40. Eswarakumar VP, Lax I, Schlessinger J. Cellular signaling by fibroblast growth factor receptors. *Cytokine & Growth Factor Reviews*. 2005; 16(2):139–149. [PubMed: 15863030]
41. Lu PF, Ewald AJ, Martin GR, Werb Z. Genetic mosaic analysis reveals FGF receptor 2 function in terminal end buds during mammary gland branching morphogenesis. *Developmental Biology*. 2008; 321(1):77–87. [PubMed: 18585375]
42. Parsa S, Ramasamy SK, De Langhe S, Gupte VV, Haigh JJ, Medina D, et al. Terminal end bud maintenance in mammary gland is dependent upon FGFR2b signaling. *Developmental Biology*. 2008; 317(1):121–131. [PubMed: 18381212]
43. Martin AJ, Grant A, Ashfield AM, Palmer CN, Baker L, Quinlan PR, et al. FGFR2 protein expression in breast cancer: nuclear localisation and correlation with patient genotype. *BMC Research Notes*. 2011; 4(72)
44. Travis RC, Reeves GK, Green J, Bull D, Tipper SJ, Baker K, et al. Gene-environment interactions in 7610 women with breast cancer: prospective evidence from the Million Women Study. *Lancet*. 2010; 375(9732):2143–2151. [PubMed: 20605201]
45. Antoniou AC, Wang XS, Fredericksen ZS, McGuffog L, Tarrell R, Sinilnikova OM, et al. A locus on 19p13 modifies risk of breast cancer in BRCA1 mutation carriers and is associated with hormone receptor-negative breast cancer in the general population. *Nature Genetics*. 2010; 42(10):885–+
46. Li C, Malone KE, Daling JR. Differences in breast cancer hormone receptor status and histology by race and ethnicity among women 50 years of age and older. *Cancer Epidemiology Biomarkers & Prevention*. 2002; 11:301–307.
47. Hao K, Chudin E, McElwee J, Schadt EE. Accuracy of genome-wide imputation of untyped markers and impacts on statistical power for association studies. *Bmc Genetics*. 2009; 10
48. Price AL, Zaitlen NA, Reich D, Patterson NJ. New approaches to population stratification in genome-wide association studies. *Nature Reviews Genetics*. 2010; 11:459–463.
49. Chang BL, Spangler E, Gallagher S, Haiman CA, Henderson BE, Isaacs W, et al. Validation of Genome-Wide Prostate Cancer Associations in Men of African Descent. *Cancer Epidemiology Biomarkers & Prevention*. 2011; 20(1):23–32.
50. He J, Wilkens LR, Stram DO, Kolonel LK, Henderson BE, Wu AH, et al. Generalizability and Epidemiologic Characterization of Eleven Colorectal Cancer GWAS Hits in Multiple Populations. *Cancer Epidemiology Biomarkers & Prevention*. 2011; 20(1):70–81.

LD based on YRI

LD based on CEU

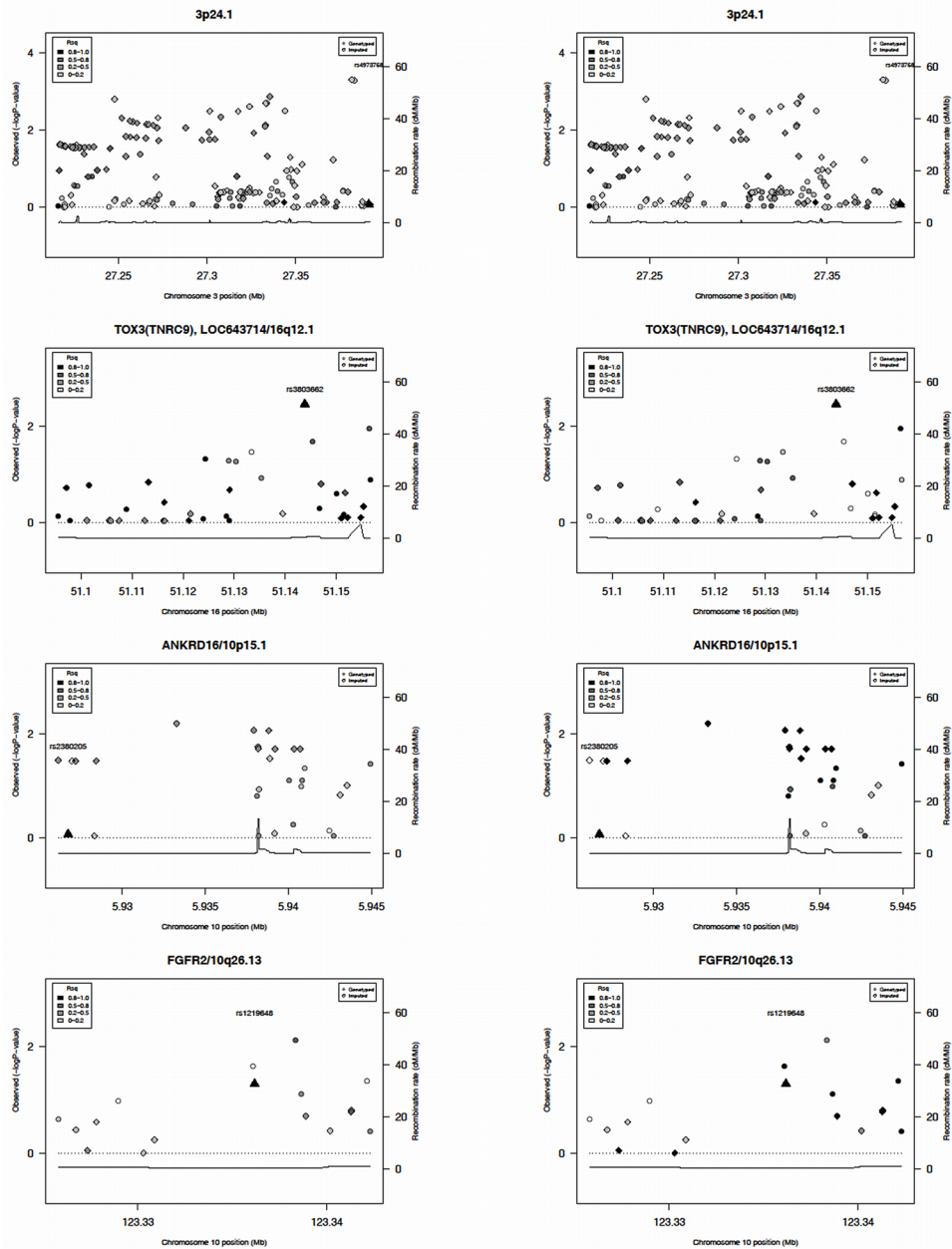


Figure 1. Regional association plots

The y axis is the $-\log_{10}$ of the p-value for the association between SNPs and breast cancer from our analysis using Cox proportional hazard regression for a log-additive model in the WHI SHARe data. The x axis is chromosomal position. The index SNP is represented as a triangle. For other SNPs, imputed SNPs are diamonds, directly genotyped SNPs are circles and points are shaded in a gray scale based on linkage disequilibrium (R_{sq}) with the index SNP in two different HapMap populations (CEU and YRI). Darker grey indicates higher LD. Circles or diamonds filled with white shading do not have LD information in the reference panel indexed (either due to low allele frequency/monomorphic or due to not

being genotyped in the HapMap panel). Plots were created by modifying existing R code available at <http://www.broadinstitute.org/diabetes/scandinav/figures.html>.

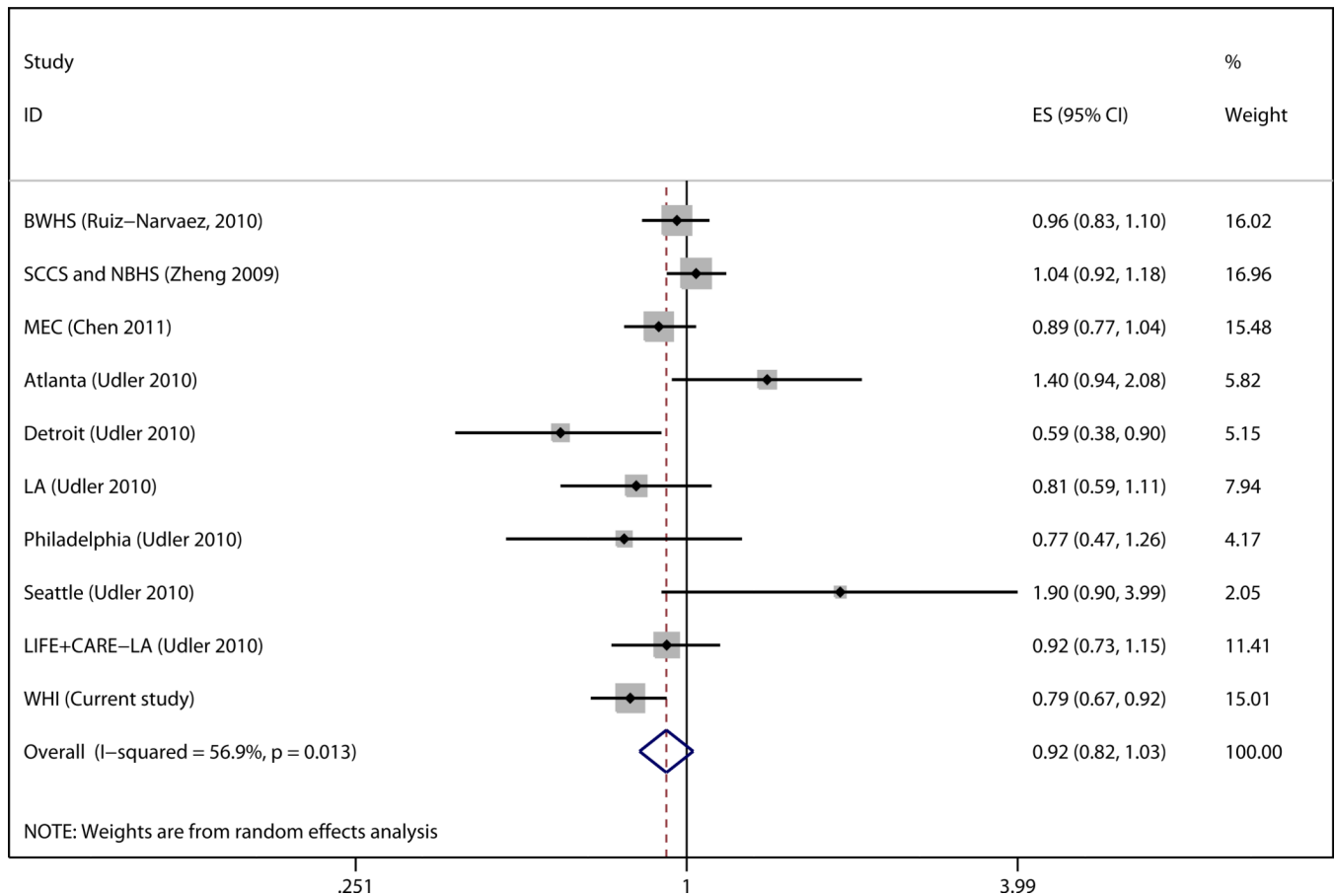


Figure 2. Meta-analysis of the association between breast cancer and the 16q12.2/TOX3 SNP rs3803662

Forest plot of effect sizes and inverse-variance weighted random-effects meta-analysis of the association between breast cancer risk and the 16q12.2/TOX3 index SNP rs3803662 in African American women. Studies are presented by Author with study population in parentheses. The symbol size indicates the weight for each study, and lines indicate the confidence intervals. BWHS=Black Women's Health Study. MEC=Multi-Ethnic Cohort; SCCS=Southern Community Cohort Study; NBHS=Nashville Breast Health Study; LA=Los Angeles; CARE= Women's Contraceptive and Reproductive Experiences Study; WHI=Women's Health Initiative. MEC results were included in several studies, we present results only from the most recent/largest report by Chen et al.

Table 1
Association of susceptibility loci from previous GWAS with breast cancer risk in African American women in WHI

Chromosomal Position	GWAS reference	SNP	Position	Alleles ¹	MAF ²	Imputed	Imputation r^2 ³	Discovery GWAS OR ⁴	HR (95% CI)	P-value ⁵
1p11.2	(14)	rs11249433	120982136	A/G	0.16	Yes	0.6841	1.16	1.08 (0.83,1.41)	0.56
2q35	(12, 14, 15)	rs13387042	217614077	A/G	0.27	Yes	0.9925	0.80	1.07 (0.89,1.27)	0.48
3p24.1	(7)	rs4973768	27391017	C/T	0.21	Yes	0.9445	1.16	1.02 (0.86, 1.20)	0.84
5p12	(13)	rs10941679	44742255	A/G	0.37	Yes	0.9562	1.19	1.31 (1.06, 1.63)	0.014
5q11.2	(8, 15)	rs16886165	56058840	T/G	0.33	Yes	0.9688	1.23	1.05 (0.89,1.25)	0.57
5q11.2	(14)	rs889312	56067641	A/C	0.32	Yes	0.96	1.22	0.98 (0.83,1.16)	0.79
6q22.33	(9)	rs2180341	127642323	A/G	0.31	Yes	0.9989	1.41	0.98 (0.83,1.17)	0.86
6q25.1	(16)	rs2046210	151990059	A/G	0.48	No	Genotyped	0.77	1.00 (0.85,1.18)	1.00
8q24.21	(8)	rs13281615	128424800	G/A	0.44	Yes	0.9176	1.08	1.16 (0.99,1.37)	0.072
9p21.3	(15)	rs1011970	22052134	G/T	0.32	Yes	0.9398	1.09	1.07 (0.90,1.28)	0.45
10p15.1	(15)	rs2380205	5926740	C/T	0.49	No	Genotyped	0.94	0.98 (0.84,1.16)	0.85
10q21.2	(15)	rs10995190	63948688	G/A	0.16	Yes	0.9754	0.86	1.12 (0.89,1.41)	0.33
10q26.13	(10)	rs704010	80511154	A/G	0.19	No	Genotyped	0.93	1.11 (0.84,1.47)	0.47
10q26.13	(14, 15)	rs2981579	123327325	G/A	0.45	Yes	0.8214	1.17	0.99 (0.83,1.17)	0.87
10q26.13	(8)	rs1219648	123336180	A/G	0.41	No	Genotyped	1.20	1.17 (1.00,1.37)	0.051
10p22.3	(15)	rs2981582	123342307	A/G	0.45	No	Genotyped	1.26	0.93 (0.79,1.09)	0.38
11q13.1	(15)	rs3817198	1865582	A/G	0.19	No	Genotyped	1.07	1.10 (0.91,1.34)	0.32
11p15.5	(8)	rs614367	69037945	A/G	0.12	No	Genotyped	0.87	1.13 (0.89,1.44)	0.30
14q24.1	(14)	rs10483813	68101037	T/A	0.10	Yes	0.7915	1.06	1.05 (0.78, 1.42)	0.76
16q12.1	(8, 12, 14, 15)	rs3803662	51143842	C/T	0.47	No	Genotyped	1.30	0.79 (0.67,0.92)	0.0034
16q12.1	(11)	rs4784227	51156689	C/T	0.15	No	Genotyped	1.24	1.23 (0.94,1.61)	0.13
17q23.2	(7)	rs6504950	50411470	G/A	0.35	Yes	0.9930	0.95	1.06 (0.90, 1.26)	0.47

¹. Listed as reference allele/tested allele.

². Minor allele frequency

³. Imputation r^2 is a measure of imputation quality from MACH

⁴ORs 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 to 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-value $< 5 \times 10^{-7}$.

⁵P-value 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 ¹	Region Boundaries ²		Total # SNPs in region	# SNPs in region with p < 0.05	SNPs in region with p < 0.05	SNP with lowest p-value in region			Permutation p-value ⁴
	Lower bound	Upper bound				rnumber	HR (95% CI)	p-value ³	
2q35	217614024	217629014	21	0	0.0%	rs10490444	0.90 (0.76, 1.05)	0.18	0.87
3p24.1	27216223	27392003	165	66	40.0%	rs9843350	1.58 (1.22, 2.05)	0.0005	0.027
5q11.2	56044769	56065000	15	2	13.3%	rs6862199	1.21 (1.02, 1.43)	0.032	0.31
6q22.33	127587185	127752866	110	0	0.0%	rs11759744	1.27 (0.99, 1.62)	0.061	0.73
6q25.1	151978370	152013380	47	0	0.0%	rs6930633	1.20 (1.00, 1.38)	0.050	0.71
8q24.21	128413061	128448145	66	5	7.6%	rs13262406	1.27 (1.06, 1.51)	0.0091	0.18
10p15.1	5926155	5944924	29	15	51.7%	rs7894083	1.29 (1.07, 1.55)	0.0063	0.079
10q21.2	63945267	63958395	17	0	0.0%	rs7082733	1.11 (0.94, 1.29)	0.21	0.73
10q26.13	123325790	123342307	17	3	17.6%	rs1219643	0.79 (0.67, 0.94)	0.0078	0.084
10p22.3	80515417	80525429	27	0	0.0%	rs1250002	1.17 (1.00, 1.37)	0.058	0.50
11q13.1	69037693	69041851	13	0	0.0%	rs634300	0.87 (0.71, 1.06)	0.17	0.74
14q24.1	68049588	68101073	84	2	2.4%	rs2842324	1.47 (1.11, 1.94)	0.0079	0.29
16q12.1	51095541	51156689	40	5	12.5%	rs3803662	0.79 (0.67, 0.92)	0.0034	0.074
17q23.2	50385225	50604678	188	10	5.3%	rs10515089	1.42 (1.06, 1.91)	0.019	0.57

¹Regions 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.

²Regions defined based on LD in HapMap CEU or CHB.

³P-value from test of trend for each additional copy of the test allele

⁴P-value for region based on 10,000 permutations of phenotype.