Fine mapping of breast cancer genome-wide association studies loci in women of African ancestry identifies novel susceptibility markers

Yonglan Zheng, Temidayo O.Ogundiran1, Adeyinka G.Falusii2, Katherine L.Nathanson3, Esther M.John4,5, Anselm J.M.Hennis6,7, Stefan Amb8, Susan M.Domchek3, Timothy R.Rebbeck9, Michael S.Simon10, Barbara Nemesure7, Sub-Yuh Wu7, Maria Cristina Leske7, Abayomi Odetunde2, Qun Niu, Jing Zhang, Chibuzor Afolabi1, Eric R.Gamazon11, Nancy J.Cox11, Christopher O.Olopade, Olufunmilayo I.Olopade and Dezheng Huo12,8

Department of Medicine, Center for Clinical Cancer Genetics and Global Health, The University of Chicago, Chicago, IL 60637, USA, 1Department of Surgery and 2Institute for Medical Research and Training, College of Medicine, University of Ibadan, Ibadan, Oyo 23402, Nigeria, 3Department of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA, 4Department of Epidemiology, Cancer Prevention Institute of California, Fremont, CA 94538, USA, 5Department of Health Research & Policy, and Stanford Cancer Institute, Division of Epidemiology, Stanford University School of Medicine, Stanford, CA 94305, USA, 6Chronic Disease Research Centre, Tropical Medicine Research Institute, University of the West Indies, Bridgetown, Saint Michael BB11000, Barbados, 7Department of Preventive Medicine, Stony Brook University Medical Center, Stony Brook, NY 11794, USA, 8Laboratory of Human Carcinogenesis, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA, 9Department of Statistics and Epidemiology, University of Pennsylvania, Philadelphia, PA 19104, USA, 10Department of Oncology, Karmanos Cancer Institute at Wayne State University, Detroit, MI 48201, USA and 11Department of Medicine, Section of Genetic Medicine and 12Department of Health Studies, The University of Chicago, Chicago, IL 60637, USA

*To whom correspondence should be addressed. Tel: +1 773-834-0843; Fax: +1 773-702-1979;
Email: dhuo@uchicago.edu

Numerous single nucleotide polymorphisms (SNPs) associated with breast cancer susceptibility have been identified by genome-wide association studies (GWAS). However, these SNPs were primarily discovered and validated in women of European and Asian ancestry. Because linkage disequilibrium is ancestry-dependent and heterogeneous among racial/ethnic populations, we evaluated common genetic variants at 22 GWAS-identified breast cancer susceptibility loci in a pooled sample of 1502 breast cancer cases and 1378 controls of African ancestry. None of the 22 GWAS index SNPs could be validated, challenging the direct generalizability of breast cancer risk variants identified in Caucasians or Asians to other populations. Novel breast cancer risk variants for women of African ancestry were identified in regions including 5p12 (odds ratio [OR] = 1.40, 95% confidence interval [CI] = 1.11–1.76; P = 0.004), 5q11.2 (OR = 1.22, 95% CI = 1.09–1.36; P = 0.00085) and 10p15.1 (OR = 1.22, 95% CI = 1.06–1.38; P = 0.0015). We also found positive association signals in the regions 6q14.1 (OR = 1.28, 95% CI = 1.14–1.42) and 2q34 (OR = 1.23, 95% CI = 1.14–1.32) previously confirmed by fine mapping in women of African ancestry. In addition, polygenic model indicated that eight best markers in this study, compared with 22 GWAS-identified SNPs, could better predict breast cancer risk in women of African ancestry (per-allele OR = 1.21, 95% CI = 1.16–1.27; P = 9.7 × 10^-16). Our results demonstrate that fine mapping is a powerful approach to better characterize the breast cancer risk alleles in diverse populations. Future studies and new GWAS in women of African ancestry hold promise to discover additional variants for breast cancer susceptibility with clinical implications throughout the African diaspora.

Introduction

Globally, breast cancer affects more than 1.1 million women each year and is the most common cancer diagnosis among women in the USA. Breast cancer is a genetically heterogeneous disease characterized by different penetrance, complex phenotypes and probably a polygenic pattern of inheritance. In addition to the two breast cancer hereditary high-penetrance genes, *BRCA1* and *BRCA2*, other genes such as *ATM*, *CHEK2*, *BRIP1* and *PALB2* were found to confer a moderate disease risk. Nonetheless, these genes only account for a small proportion of breast cancer risk, whereas major genetic component of breast cancer risk remains uncharacterized, especially in the development of sporadic breast cancer (1). Hence, it has been proposed that breast cancer risk could be partly attributable to common low-risk alleles of genetic variants in the human genome that are usually associated with odds ratios (ORs) ranging from 1.1 to 1.3.

In the past 5 years, a number of genome-wide association studies (GWAS) have identified breast cancer susceptibility alleles among multiple racial/ethnic populations (2–23). However, most of the susceptibility single nucleotide polymorphisms (SNPs) were only discovered and validated in Caucasian women. Few follow-up replication studies have been conducted in women of African ancestry to confirm findings from the original GWAS (24–31). However, it is well known that linkage disequilibrium (LD) patterns vary across different populations, with the small and weak LD blocks observed in populations of African descent. This difference can be explored in fine-mapping studies to identify novel disease risk variants and has been applied to African Americans for the characterization of GWAS-defined breast cancer risk loci (31–35), including one large study that examined multiple regions (36).

In this study, we applied a fine-mapping strategy in the evaluation of common genetic variants at 22 GWAS-defined breast cancer susceptibility regions using 1502 breast cancer cases and 1378 controls of African ancestry assembled from six studies conducted in Nigeria, Barbados and the USA.

Materials and methods

Ethics statement

The six epidemiologic studies have been approved by institutional review boards at the participating institutions. Written informed consent was obtained from all participants.

Study subjects

The study populations have been described previously (30,37–42). Briefly, a total of 2880 subjects of African ancestry (1502 breast cancer cases and 1378 controls) were included in this study (Figure 1). The average age (mean ± standard deviation) was 48.1 ± 12.1 years in cases (age at diagnosis) and 47.2 ± 13.2 years in controls (age at interview). Ascertainment of cases and controls occurred in Ibadan, Nigeria (675 cases and 280 controls), Barbados (93 cases and 244 controls) and four sites in the USA (734 cases and 854 controls). The description of the study samples for each site can be found in Supplementary Material and Table S1, available at *Carcinogenesis* Online.

GWAS regions and SNP selection

Twenty-two genomic regions that showed strong associations with breast cancer in previous GWAS were selected for fine mapping: 1p11.2, 2q35, 3p24.1, 5p12, 5q11.2, 6q22.33, 6q25.1, 8q24.21, 9p21.3, 10p15.1, 10q21.2, 10q22.3, 10q26.13, 11p15.5, 11q13.3, 12p11.22, 12q24.21, 14q24.1, 16q12.1–q12.2, 17q22, 19p13.11 and 21q11.2–q21.1 (Table 1 and Supplementary Table S2, available at *Carcinogenesis* Online). Several other regions (such as 2q34, 3q25.1, 6q14.1, 9q31.2, 14q31.3 and 20q11.22; GWAS catalog: http://www.genome.gov/gwastudies/) were not included because they were not identified at the time when this study was designed.

Abbreviations: AIMs, ancestry informative markers; CEPH, Centre d’Etude du Polymorphisme Humain; CEU, Utah residents with Northern and Western European ancestry from the CEPH cell collection; CI, confidence interval; ER, estrogen receptor; GWAS, genome-wide association studies; LD, linkage disequilibrium; MAF, minor allele frequency; OR, odds ratio; PR, progesterone receptor; SNP, single nucleotide polymorphism; YRI, Yoruba in Ibadan.
Different SNP selection criteria were used to determine the suitability of SNPs across different chromosomal regions for genotyping (Supplementary Table S2, available at Carcinogenesis Online). (i) For six GWAS loci, 2q35, 3p11.2, 6q22.33, 10q26.13, 10q12.1-q12.2 and 19p13.11, a popular tagging-SNP method was applied in order to have a small set of SNPs capturing tagging all other SNPs with a minor allele frequency (MAF) of 0.05 in the entire regions and maintaining an acceptable coverage (≥80%). In addition, genotyping data for Yoruba in Ibadan (YRI, Nigeria) from HapMap (http://hapmap.ncbi.nlm.nih.gov/) Phase II Release #28 and The 1000 Genome Project (http://www.1000genomes.org/) 2010 March Release were merged to evaluate the tagging SNPs in LD block(s) using Haplovew v4.2 (http://www.broadinstitute.org/scientific-community/science/programs/medical-and-population-genetics/haplovew/haplovew). (ii) For 15 loci, 1p11.2, 3p24.1, 6q25.1, 8q24.21, 9p21.3, 10p15.1, 10q21.2, 10q22.3, 11p15.5, 11q13.3, 12p11.22, 12q24.21, 14q24.1, 17q22 and 21q11.2-q21.1, a correlated SNP method was used. SNPs that are within ±2 Mb distance and have LD $r^2 > 0.3$ (or 0.2 with the original GWAS index SNPs) were chosen in the CEU population (Utah residents with Northern and Western European ancestry from the Centre d’Etude du Polymorphisme Humain (CEPH) collection). Similarly, genotyping data of CEU from HapMap Phase II Release #28 and The 1000 Genome Project 2010 March Release were combined and then Haplovew v4.2 was employed to calculate the LD $r^2$-values between SNP pairs. One exception was the 6q25.1 locus, for which SNPs correlated with two index SNPs (rs2046210 and rs9397435) at $r^2 > 0.3$ in Han Chinese in Beijing (7) or SNPs correlated with rs9397435 at $r^2 > 0.3$ in CEU or YRI (35) were selected. (iii) For the remaining locus, 5p12, we applied both tagging and correlated SNPs, using the YRI combined data set (HapMap Phase II Release #28 and The 1000 Genome Project 2010 March Release).

Genotyping and quality control

A total of 955 tagging SNPs and 30 ancestry informative markers (AIMs) were genotyped using the Illumina GoldenGate Genotyping platform (Illumina, San Diego, CA) as part of a panel of 1536 SNPs (Figure 1). Genotyping and the postgenotyping quality control steps have been described previously (30). In brief, the average call rate was 99.96% for successfully genotyped SNPs. Cases and controls were randomly placed in 96-well plates and distributed roughly equal in each plate, along with negative controls and intra- and inter-plate blind duplicate samples. The reproducibility rate for duplicate samples was 99.95%. We excluded 122 tagging SNPs and one AIM from further statistical analyses using the following filters: well-distinguishable genotyping clustering, MAF ≥ 0.01 and no deviation of Hardy-Weinberg Equilibrium ($P ≥ 0.00001$).

Individual African ancestry estimation

The African ancestry proportion of each individual was calculated by Structure v2.3.3 (http://pritch.bsd.uchicago.edu/structure.html), based on data from 29 genotyped AIMs. AIM selection criteria and evaluation have been described elsewhere (30). Briefly, a set of 1373 markers with maximum allele frequency differences between European and African descendants has been evaluated, and the results showed that ancestry estimates via these 30 AIMs were highly correlated ($r = 0.89$) with estimates using the entire set (43). The mean (median) proportion of African ancestry in this study was close to 1 in Nigerians (0.857 (0.882) in Barbadians and 0.792 (0.808) in African Americans. No significant difference was observed in the proportions of European admixture between cases and controls (30).

SNP imputation

MACH v1.0 (http://www.sph.umich.edu/csg/abecasis/MACH/index.html) was used to impute unobserved SNPs in the same regions. Phased YRI and CEU haplotype data from HapMap Phased II Release #82 and phased African haplotype data from The 1000 Genome Project November 2010 release served as the reference panels. In an attempt to ensure the imputation quality, only imputed SNPs with imputation $r^2 ≥ 0.6$ were eligible for further statistical analyses. Imputation quality scores and $r^2$-values from HapMap and The 1000 Genome imputation data sets were compared and found to be quite similar. For SNPs where the imputation $r^2$-values were different, the better ones were extracted as the final outputs. In addition, imputation was operated in slightly broader regions, compared with the ones determined for SNP selection, in order to accommodate any neighbor boundary regions. Some regions were trimmed down for plotting by removing unreliable boundary segments that were poorly imputed, if any (Supplementary Table S3, available at Carcinogenesis Online).

Statistical analysis

Association tests. Within each of the 22 loci, we identified the most significant markers by testing the association of each SNP with breast cancer risk using unconditional logistic regression model. OR and 95% confidence interval (CI) were calculated after the adjustment for African ancestry proportion, study site and age at diagnosis (or age at interview for controls) in logistic regressions as estimates for allele dosage effect. Because sister controls were used in one study (Northern California, 14% samples), a conditional regression model was used first, followed by meta-analysis technique to combine the estimated site-specific OR and 95% CI. The two methods gave very similar results and thus we presented the results from unconditional logistic regression to have a marginal or population average interpretation. We examined LD ($r^2$ and D’) in three populations (African, European and Asian) between the most significant markers in this fine-mapping study and index signals described by previous GWAS or fine-mapping studies. The LD analysis could reveal if our fine-mapping study identifies independent signals. As an exploratory analysis, we also examined whether SNPs are associated with breast cancer estrogen receptor (ER) or progesterone receptor (PR) status: ER-negative, ER-positive, PR-negative or PR-positive.

Because multiple testing can generate false-positive findings, we performed permutation testing within each of the 22 loci to assess statistical
<table>
<thead>
<tr>
<th>Locus</th>
<th>RefSeq genes</th>
<th>GWAS index SNPs and fine-mapping SNPs: SNP1</th>
<th>Most significant SNPs in this study: SNP2</th>
<th>$r^2/D'$ between SNP1 and SNP2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1p1.2</td>
<td>EMBP1</td>
<td>GWAS rs11249433 (A/G)</td>
<td>rs11249433 (A/G)</td>
<td>NA</td>
</tr>
<tr>
<td>2q35</td>
<td>Intergenic</td>
<td>GWAS rs13387042 (G/A)</td>
<td>rs13387803 (G/A)</td>
<td>0.001/0.086</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fine mapping rs13000023 (A/G)</td>
<td>rs13000023 (A/G)</td>
<td>0.005/0.185</td>
</tr>
<tr>
<td>3p24.1</td>
<td>NEK10, SLC4A7</td>
<td>GWAS rs4973768 (C/T)</td>
<td>rs73821937 (A/C)</td>
<td>0.068/1</td>
</tr>
<tr>
<td>5p12</td>
<td>FGFR10, MRPS30, HCN1, MAP5K1, C5orf35, MIER3</td>
<td>GWAS rs4410584 (C/T)</td>
<td>rs114245766 (G/A)</td>
<td>0.088/0.807</td>
</tr>
<tr>
<td>5q1.2</td>
<td>Intergenic</td>
<td>GWAS rs889312 (A/C)</td>
<td>rs832539 (T/G)</td>
<td>0.036/0.205</td>
</tr>
<tr>
<td>6q22.33</td>
<td>RNF146, ECHDC1, C5orf174, C5orf97, ESR1</td>
<td>GWAS rs2046210 (A/G)</td>
<td>rs9383931 (G/C)</td>
<td>0.010/0.246</td>
</tr>
<tr>
<td>6q25.1</td>
<td>Intergenic</td>
<td>GWAS rs2046210 (A/G)</td>
<td>rs9383931 (G/C)</td>
<td>0.051/0.427</td>
</tr>
<tr>
<td>8q24.21</td>
<td>Intergenic</td>
<td>GWAS rs13281615 (A/G)</td>
<td>rs978683 (A/G)</td>
<td>0.270/0.677</td>
</tr>
<tr>
<td>9p21.3</td>
<td>CDKN2A and others</td>
<td>GWAS rs1011970 (G/T)</td>
<td>rs73652816 (G/A)</td>
<td>0.163/0.139</td>
</tr>
<tr>
<td>10p15.1</td>
<td>GDO2, ANKRD16, FBX018, ZNF365</td>
<td>GWAS rs2380205 (T/C)</td>
<td>rs1255602 (T/A)</td>
<td>0.250/0.811</td>
</tr>
<tr>
<td>10q21.2</td>
<td>Intergenic</td>
<td>GWAS rs1095190 (A/G)</td>
<td>rs57245264 (T/C)</td>
<td>0.055/0.302</td>
</tr>
<tr>
<td>10q22.3</td>
<td>LOC283030, ZMIZ1</td>
<td>GWAS rs704010 (C/T)</td>
<td>rs4980024 (G/A)</td>
<td>0.009/0.315</td>
</tr>
<tr>
<td>10q26.13</td>
<td>FGFR2</td>
<td>GWAS rs2981582 (A/G)</td>
<td>rs2917778 (A/G)</td>
<td>0.009/0.315</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fine mapping rs2981578 (A/G)</td>
<td>rs2917778 (A/G)</td>
<td>0.924/1</td>
</tr>
<tr>
<td>11p1.55</td>
<td>LSP1 and others</td>
<td>GWAS rs3817198 (T/C)</td>
<td>rs74047621 (C/T)</td>
<td>0.109/0.394</td>
</tr>
<tr>
<td>11q1.3</td>
<td>CCND1, ORAOV1, FGFR19, PTHLH</td>
<td>GWAS rs614367 (C/T)</td>
<td>rs679162 (T/C)</td>
<td>0.002/0.055</td>
</tr>
<tr>
<td>12p1.22</td>
<td>Intergenic</td>
<td>GWAS rs10771399 (G/A)</td>
<td>rs2737456 (G/A)</td>
<td>0.034/0.738</td>
</tr>
<tr>
<td>12q24.21</td>
<td>Intergenic</td>
<td>GWAS rs1292011 (G/A)</td>
<td>rs1391715 (T/G)</td>
<td>0.248/0.643</td>
</tr>
<tr>
<td>14q24.1</td>
<td>RAD51L1</td>
<td>GWAS rs999737 (T/C)</td>
<td>rs34882439 (G/A)</td>
<td>0.083/1</td>
</tr>
</tbody>
</table>

GW AS refers to the GWAS index SNPs and fine-mapping SNPs. $r^2/D'$ indicates the correlation coefficient between the two SNPs.
Table I. Continued

<table>
<thead>
<tr>
<th>Locus</th>
<th>RefSeq genes</th>
<th>GWAS/fine mapping</th>
<th>rs# (ref/risk allele)</th>
<th>RAF (cases/controls)</th>
<th>Per risk allele OR (95% CI)</th>
<th>P-value</th>
<th>rs# (ref/risk allele)</th>
<th>RAF (cases/controls)</th>
<th>Per risk allele OR (95% CI)</th>
<th>P-value</th>
<th>(African)</th>
<th>(European)</th>
<th>(Asian)</th>
</tr>
</thead>
<tbody>
<tr>
<td>16q12.1–q12.2</td>
<td>TOX3, LOC643714</td>
<td>GWAS</td>
<td>rs3803662 (G/A)</td>
<td>0.51/0.516</td>
<td>0.96 (0.86–1.07)</td>
<td>0.41</td>
<td>rs3104788 (T/C)</td>
<td>0.450/0.407</td>
<td>1.19 (1.07–1.32)</td>
<td>0.014</td>
<td>0.254</td>
<td>0.269</td>
<td>0.294</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fine mapping</td>
<td>rs3104793 (T/C)</td>
<td>0.625/0.584</td>
<td>1.12 (1.00–1.25)</td>
<td>0.044</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.584</td>
<td>0.984</td>
<td>0.989</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fine mapping</td>
<td>rs3112572 (G/A)</td>
<td>0.248/0.229</td>
<td>1.05 (0.92–1.19)</td>
<td>0.47</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.150</td>
<td>0.002</td>
<td>0.045</td>
</tr>
<tr>
<td>17q22</td>
<td>TOM1L1, COX11, STXBP4</td>
<td>GWAS</td>
<td>rs6504950 (A/G)</td>
<td>0.635/0.654</td>
<td>0.90 (0.80–1.01)</td>
<td>0.068</td>
<td>rs2958915 (G/T)</td>
<td>0.579/0.556</td>
<td>1.12 (1.00–1.25)</td>
<td>0.056</td>
<td>0.088</td>
<td>0.399</td>
<td>0.010</td>
</tr>
<tr>
<td>19p13.11</td>
<td>C19orf92, ANKLE1, and others</td>
<td>GWAS</td>
<td>rs2363956 (G/T)</td>
<td>0.515/0.501</td>
<td>1.08 (0.97–1.20)</td>
<td>0.18</td>
<td>rs76900423 (T/A)</td>
<td>0.865/0.853</td>
<td>1.17 (1.00–1.36)</td>
<td>0.048</td>
<td>0.025</td>
<td>0.011</td>
<td>0.286</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fine mapping</td>
<td>rs3745185 (A/G)</td>
<td>0.808/0.786</td>
<td>1.10 (0.96–1.26)</td>
<td>0.18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.051</td>
<td>0.102</td>
<td>0.024</td>
</tr>
<tr>
<td>21q11.2–q21.1</td>
<td>NRIP1</td>
<td>GWAS</td>
<td>rs2823093 (A/G)</td>
<td>0.550/0.544</td>
<td>1.09 (0.98–1.22)</td>
<td>0.12</td>
<td>rs2823093 (A/G)</td>
<td>0.550/0.544</td>
<td>1.09 (0.98–1.22)</td>
<td>0.12</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

NA, Not applicable.

*Alleles on plus strand, risk alleles are from previous studies.

*RAF, risk allele frequency.

*Per risk allele ORs are adjusted for age, study site and individual African ancestry estimate.

*Alleles on plus strand, risk alleles are determined in this study.

*Imputed SNP.

*Eight best SNPs for polygenic risk model.

*Imputed SNPs are presented rather than genotyped SNPs when imputation $r^2 > 0.9$ and stronger signal (>1 order of magnitude change in the P-value) were found.
significance of a novel, independent signal at each locus. One thousand thousand permutations were conducted using PLINK v1.07 (http://pngu.mgh.harvard.edu/~purcell/plink/). A corrected P-value of ≤ 0.05 was considered statistically significant and a corrected P-value of 0.05–0.1 was considered marginally significant.

**Risk modeling.** Two composite risk models were built; one used the 22 risk variants reported in previous GWAS and the other used the 8 fine-mapping risk variants identified in this study. The composite risk score was calculated as the count of risk alleles and only one SNP per genomic locus was chosen. For individuals with missing data, the score was calculated as the average risk allele count multiplied by the number of total SNPs. As missing data were infrequent (~0.2%), they had no material impact on the composite score. Both continuous and categorical risk scores (grouped by quartiles) were examined in relation to breast cancer risk using logistic regression, adjusted for age, study site and African ancestry proportion.

**Results**

**Replication of GWAS index SNPs and fine-mapping SNPs**

Among 22 breast cancer GWAS regions, we examined 22 GWAS index SNPs and 10 SNPs from previous fine-mapping studies conducted in women of African ancestry (Table I).

**GWAS index SNPs.** We successfully genotyped 19 and imputed 3 GWAS index SNPs (rs13387042 at 2q35, rs4415084 at 5p12 and rs2380205 at 10p15.1) (Table I and Supplementary Figure S1, available at Carcinogenesis Online). All 22 SNPs were common variants in women of African ancestry, with MAF > 0.05. The per-allele ORs for the risk alleles were >1.0 for nine of the GWAS index SNPs, however, none of the variants was statistically significantly associated with breast cancer (P ≥ 0.05). ORs for the remaining 13 SNPs were ≤1.0, thus showing a suggestive inverse association with breast cancer risk when compared with previous GWAS findings.

**Fine-mapping SNPs.** We also examined 10 SNPs located in GWAS-defined breast cancer risk loci that were previously identified in fine-mapping studies and were found to be associated with breast cancer in women of African ancestry (31,32,34–36). Five of 10 SNPs, were directly genotyped, four variants were successfully imputed, whereas one (rs609275s at 11q13.3) could not be reliably imputed (r² = 0.02). The nine analyzable variants were found to be common in Black women (MAF ≥ 0.05). We obtained positive associations with each of these nine SNPs with ORs >1.0, ranging from 1.05 to 1.33 (Table I and Supplementary Table S1, available at Carcinogenesis Online). In addition to those three (rs9397435 at 6q25.1, rs2981578 at 10q26.13 and rs3104793 at 16q12.1–q12.2) fine-mapping variants consistently replicated to confer increased breast cancer risk in women of African ancestry (30,32,34,35), we found two additional SNPs that were also associated with breast cancer risk: rs16886165 at 5q11.2 (9,36) (OR = 1.12, 95% CI = 1.00–1.26; P = 0.053) and rs2046211 at 6q25.1 (31) (OR = 1.24, 95% CI = 1.03–1.47; P = 0.02). However, the associations were not significant after permutation testing (Supplementary Table S4, available at Carcinogenesis Online). No significant P-values were found for the four remaining SNPs (rs13000023, rs3112572, rs12355688 and rs3745185) (36), but the direction of association remained the same (Table I).

**Newly identified breast cancer risk markers in women of African ancestry**

5q11.2. This region (FGF10-MRPS30-HCN1) has been reported to be a breast cancer predisposition locus by GWAS (2,14) (Figure 2). Our fine-mapping analysis revealed a significant association between SNP rs114245766 and breast cancer (OR = 1.40, 95% CI = 1.11–1.76; P = 0.004). However, this association did not remain significant after permutation testing (Table S4). This SNP is located 260 kb centromeric of the initial GWAS signal, and it is polymorphic in Africans but monomorphic in Europeans and Asians. The LD between rs114245766 and the GWAS index SNP rs4415084 is low (r² = 0.088; Table I), implying that rs114245766 is a novel candidate breast cancer susceptibility variant in women of African ancestry. Notably, several SNPs in moderate or high LD with rs114245766 are located in a region spanning a small (6.6 kb in size) gene, MRPS30 (Figure 2).
Fine-mapping breast cancer loci in Blacks

11q13.3, 12p11.22, 14q24.1, 17q22, 19p13.11 and 21q11.2–q21.1 did not provide evidence of improved fine-mapping signals that could better define breast cancer risk in women of African ancestry (Supplementary Figure S2, available at Carcinogenesis Online).

Lastly, we conducted exploratory association analyses after stratification by breast cancer subtype because the association of genetic susceptibility factors with breast cancer could be subtype-dependent. Our preliminary findings are reported after stratification by estrogen and progesterone receptor status and can be found in Supplementary Materials (Supplementary Table S7 and Figures S3–S6, available at Carcinogenesis Online).

**Polygenic risk model**

We constructed a composite risk score from unweighted risk allele counts using the 22 index SNPs reported from previous GWAS, as well as a composite risk score using the eight best genetic markers (5p12, 5q11.2, 6q25.1, 9p21.3, 10p15.1, 10q26.13, 12q24.21 and 16q12.1–q12.2) identified in women of African ancestry (Table II and Supplementary Tables S5 and S6, available at Carcinogenesis Online). The composite risk score from the 22 index SNPs was not associated with breast cancer risk in women of African ancestry (per-allele OR = 0.99, 95% CI = 0.96–1.02; P = 0.46). In contrast, the composite risk score from the eight best markers in women of...
African ancestry was significantly associated with breast cancer risk in women of African ancestry (per allele OR = 1.21, 95% CI = 1.16–1.27; $P = 9.7 \times 10^{-16}$). Women with seven or more risk alleles had a 2.2-fold increased risk compared with women with fewer than five risk alleles. This composite score was significantly associated with both ER-positive (per allele OR = 1.23, 95% CI = 1.14–1.33; $P = 4.7 \times 10^{-5}$) and ER-negative breast cancer (per allele OR = 1.29, 95% CI = 1.19–1.39; $P = 3.0 \times 10^{-5}$).

Discussion

LD differences across populations enable us to refine causal variants for a complex disease like breast cancer to a restricted genomic region (44). Therefore, we conducted this fine-mapping study of GWAS-defined breast cancer susceptibility loci in women of African ancestry in the context of previous studies that only focused on women of European and Asian ancestry. Our approach yielded several independent signals for breast cancer risk and better genetic markers for the disease in women of African ancestry at genomic regions including 5p12, 5q11.2, 6q25.1, 10p15.1, 10q26.13 and 16q12.1–q12.2.

Evidence of association with breast cancer risk at 5p12 has been reported, with one ‘hit’ located between the FGF10 and MRPS30 genes (14) and another one within the HCN1 gene (2). A combined analysis of three GWAS suggested the location of a candidate-causative SNP between the MRPS30 and HCN1 genes (45). The best marker rs114245766 in this study at this locus as well as other SNPs correlated with it, spread in a 400 kb region harboring the MRPS30 gene that encodes a mitochondrial ribosomal protein S30. It remains to be addressed whether this gene plays a role in the etiology of breast cancer.

The GWAS index SNP rs889312 lies in a 300 kb LD block in women of African ancestry containing MAP3K1, C5orf35 and MIER3 genes (5q11.2). In line with two previous studies (25,36), we found no association with this SNP in women of African ancestry. Although a replication study in African Americans failed to validate the association with rs16886165, which was found to be associated with breast cancer risk in White women (9,29), a positive association was reported by a fine-mapping study in African Americans (36). We observed a $P$-value of 0.053 for this SNP but the association did not remain significant after controlling for multiple testing by permutation. In addition, we identified rs832559 to be strongly associated with breast cancer risk in women of African ancestry, suggesting that this SNP could serve as a better marker at this locus. This variant is located 3’ downstream of the MAP3K1 gene and 5’ upstream of the C5orf35 gene. Similarly, association signals at higher statistical significance level than that for rs16886165 were found to cluster in the 3’ end of MAP3K1 (36), in agreement with this study. The MAP3K1 gene encodes a serine/threonine kinase that is involved in the extracellular signal-regulated kinase and c-jun N-terminal kinase pathways implicated in control of cell proliferation and death. Somatic mutations in MAP3K1 have been detected in 2–6% of breast cancer, particularly in ER-positive tumors (46–50), and thus, it illuminates the vital function of MAP3K1 in breast carcinogenesis.

At 10p15.1, we identified a better marker, rs11255602, for breast cancer risk in Black women. rs11255602 and other SNPs highly correlated to it are located in a LD block harboring the ANKRD16 and FBXO18 genes whose biological roles in breast cancer development are still unclear. We were able to discover additional novel breast cancer risk variants by conducting a fine-mapping study in women of African ancestry, but, we acknowledge the importance of further validating the implications of the above markers in other populations, especially the same or similar ethnic populations.

Three breast cancer GWAS susceptibility loci, 6q25.1, 10q26.13 and 16q12.1–q12.2, have been studied in African and African American women (31,32,34–36). Our fine-mapping study again confirmed that these three regions are candidate breast cancer susceptibility regions in women of African ancestry. Our findings suggest that the susceptibility SNPs identified in this study (rs9383931, rs2912778 and rs3104788) could be novel candidate markers that define breast cancer susceptibility in women of African descent. Additionally, it emphasizes the importance of fine-scale approach in diverse populations with different LD structures.

As expected, each risk allele of the SNP marker only confers a modest relative risk of <1.30. When building a polygenic risk prediction model using the eight fine-mapping markers from eight susceptibility loci, we showed that women in the highest quartile have a 2.2-fold increased risk compared with women in the lowest quartile. This risk score is applicable for both ER-positive and ER-negative breast cancer. Although the magnitude of effect of the current model is similar to that from a previous study in African Americans (36), only three regions (5q11.2, 10q26.13 and 16q12.1–q12.2) are shared between the two models. Notably, both studies showed that population-specific models have a better prediction value than models directly from GWAS-identified variants. We believe that population-specific models are warranted until the causal alleles in each susceptibility region have been identified and their effects have been well characterized.

With multiple biomarkers being tested in this fine-mapping study, false-positive findings are a concern. The permutation test is an attempt to maintain the region-level type I error rate. As most of the novel fine-mapping SNPs remained significant or marginally significant after permutation testing, they are less likely to be false positive although further validation is necessary. On the other hand, our failure to replicate GWAS index signals may be partially due to limited statistical power; the study had <40% power in 11 of the 22 regions. In addition, as the same samples were used in both fine mapping and model building, our risk prediction model may overestimate the effects. Further validation is needed before the model can be used in clinical practice.

In line with most findings from previous studies evaluating breast cancer GWAS index signals in African Americans (24–31), this study elaborated the unsuccessful replication of initial associations using indigenous Nigerians, Barbadians and African Americans. No

<table>
<thead>
<tr>
<th>Risk allele counts and breast cancer risk in women of African ancestry</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean number of risk alleles</td>
<td>22 GWAS index SNPs</td>
<td>Eight best markers in women of African ancestry</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Per allele</td>
<td>Case</td>
<td>Control</td>
<td>OR (95% CI)</td>
<td>$P$-value</td>
<td>Case</td>
<td>Control</td>
<td>OR (95% CI)</td>
</tr>
<tr>
<td>Number of subjects</td>
<td>1378</td>
<td>1502</td>
<td>0.99 (0.96–1.02)</td>
<td>0.46</td>
<td>1378</td>
<td>1502</td>
<td>1.21 (1.16–1.27)</td>
</tr>
<tr>
<td>Risk quartiles$^{a}$</td>
<td>Q1</td>
<td>340</td>
<td>365</td>
<td>1.00 (ref.)</td>
<td></td>
<td>319</td>
<td>243</td>
</tr>
<tr>
<td>Q2</td>
<td>356</td>
<td>425</td>
<td>1.08 (0.87–1.34)</td>
<td>0.42</td>
<td>333</td>
<td>277</td>
<td>1.11 (0.87–1.41)</td>
</tr>
<tr>
<td>Q3</td>
<td>362</td>
<td>373</td>
<td>0.91 (0.73–1.13)</td>
<td></td>
<td>330</td>
<td>343</td>
<td>1.37 (1.08–1.74)</td>
</tr>
<tr>
<td>Q4</td>
<td>320</td>
<td>339</td>
<td>0.95 (0.76–1.19)</td>
<td></td>
<td>396</td>
<td>639</td>
<td>2.26 (1.81–2.82)</td>
</tr>
</tbody>
</table>

$^{a}$ORs are adjusted for age, study site and individual African ancestry estimate.

$^{b}$Based on distribution in controls: categories for index markers are 12–19, 20–21, 22–23 and 24–31; categories for best markers are 1–4, 5, 6 and 7–12.
statistical significance was observed at any of the 22 breast cancer GWAS index SNPs, whereas associations in reverse direction were found for several SNPs (Table 1). This flip-flop association phenomenon has been commonly reported in different racial/ethnic populations, but its clinical implication cannot be over-emphasized (51). Non-replication could be attributable to different LD patterns among different populations, as well as etiological heterogeneity, population stratification and differences in environmental exposures (51). These inconsistent findings imply that breast cancer risk variants identified by previous GWAS in Caucasians and Asians are not generalizable in other populations such as women of African origin. Our findings highlight the importance and necessity of conducting refinement studies of original breast cancer GWAS hits in independent and diverse populations. In order to efficiently and reliably apply the current observations of genetic-phenotype association and the value of prediction models based on common low-penetrance alleles to the clinic, further validation and confirmation using larger sample collections are indispensable.

Supplementary material
Supplementary Material, Tables S1–S7, and Figures S1–S6 can be found at http://carcin.oxfordjournals.org/

Funding
National Cancer Institute (R01 CA89085, P50 ES012382, R01 CA141712, R01 CA142996, PO1 CA82707 and P50 CA125183); Noreen Fraser Foundation; Breast Cancer Research Foundation. The Breast Cancer Family Registry (BCFR) was supported by the NCI under RFA CA-06-503 and through cooperative agreements with members of the BCFR and Principal Investigators, including the Cancer Prevention Institute of California (U01 CA69417).

Acknowledgements
The authors thank all the women who participated in this research. The authors appreciate Drs Habibul Ahsan, James D.Fackenthal, Muhammad G.Kibriya and Jonathan Pritchard for their advice in study design, and Mr Toshio Y Yoshimatsu for Linux maintenance.

The content of this manuscript does not necessarily reflect the views or policies of the National Cancer Institute or any of the collaborating centers in the BCFR, nor does mention of trade names, commercial products or organizations imply endorsement by the USA government or the BCFR. Data were provided by the Georgetown University Medical Center Informatics Support Center, Georgetown University, Washington, DC, USA, the Cancer Prevention Institute of California, and the Cancer Prevention Institute of California Informatics Support Center, which are funded by the National Cancer Institute and the Cancer Prevention Institute of California (U01 CA69417). None declared.

Conflict of Interest Statement: None declared.

References

Received January 22, 2013; revised February 24, 2013; accepted March 5, 2013