Fine-mapping of the 6q25 locus identifies a novel SNP associated with breast cancer risk in African–American women

Edward A. Ruiz-Narváez1,2,*, Lynn Rosenberg1,2, Song Yao3, Charles N. Rotimi4, Adrienne L. Cupples5, Elisa V. Bandera6, Christine B. Ambrosone7, Lucile L. Adams-Campbell1 and Julie R. Palmer1,2

1Slone Epidemiology Center, Boston University, Boston, MA 02215, USA, 2Department of Epidemiology, Boston University School of Public Health, Boston, MA 02118, USA, 3Department of Cancer Prevention and Control, Roswell Park Cancer Institute, Buffalo, NY 14263, USA, 4Center for Research on Genomics and Global Health, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA, 5Department of Biostatistics, Boston University School of Public Health, Boston, MA 02218, USA, 6The Cancer Institute of New Jersey/Robert Wood Johnson Medical School, New Brunswick, NJ 08903, USA and 7Georgetown University Medical Center, Washington, DC 20057, USA

*To whom correspondence should be addressed. Slone Epidemiology Center, Boston University, Boston, MA 02215, USA. Tel: +1-617-734-6006; Fax: +1-617-738-5119; Email: eruiznar@bu.edu

The rs2046210 single nucleotide polymorphism (SNP) in the 6q25.1 region was identified in a breast cancer genome-wide association study of Chinese women. The SNP has been replicated in European ancestry populations, but replication efforts have failed in African ancestry populations. We evaluated a total of 13 tagging SNPs in the linkage disequilibrium block around rs2046210 in a case-control study of breast cancer nested within the Black Women’s Health Study, which included 1191 cases and 1941 controls. Replication of initial significant findings was carried out in 665 cases and 821 controls of African ancestry from the Women’s Circle of Health Study (WCHS). No significant associations were found for rs2046210 in univariate analysis. A new SNP, rs2046211, was significantly associated with reduced risk of breast cancer and was replicated in data from WCHS. In joint analyses that included both SNPs, the rs2046210-A allele was associated with increased risk of breast cancer [odds ratio (OR) = 1.14; 95% confidence interval (CI) = 1.02–1.28], and the rs2046211-G allele was associated with reduced risk (OR = 0.80; 95% CI = 0.67–0.95). Haplotype analysis confirmed these results and showed that the rs2046210-A allele is present in high-risk (rs2046211-C/rs2046210-A) and low-risk (rs2046211-G/rs2046210-A) haplotypes. Our results confirm the importance of 6q25.1 as a breast cancer susceptibility region. We replicated the rs2046210 association, after accounting for the haplotype background that included rs2046211 in African–American women, and we report the presence of a novel signal that is tagged by rs2046211.

Introduction

A genome-wide association study (GWAS) of breast cancer conducted among Chinese women living in Shanghai identified a genetic susceptibility locus at 6q25.1 (1). The A-allele in rs2046210, a single nucleotide polymorphism (SNP) located upstream of the estrogen receptor (ESR1) gene, was associated with a 29% increase in risk of breast cancer. The association was significantly stronger for estrogen receptor negative (ER−) breast cancer than for estrogen receptor positive (ER+) breast cancer (1,2). The SNP has been associated with breast cancer overall in several European ancestry populations (1–4), although with a weaker association and no apparent difference by molecular subtype (2). To date, studies of African ancestry populations have failed to replicate the association with rs2046210 (2,3,5–7).

The greater genetic variation [lower levels of linkage disequilibrium (LD), more haplotypes, more divergent patterns of LD and more complex patterns of population substructure] present in African ancestry populations relative to European or Asian ancestry populations make African ancestry populations especially valuable for genetic research. For example, the weaker LD seen in African ancestry populations has been successfully used to fine-map and better localize loci initially identified from European and east Asian populations for multiple complex traits including type 2 diabetes (8), fasting blood sugar (9), serum uric acid (10) and bilirubin levels (11). The weaker LD in the genome of African ancestry populations makes it more probably that variants found to be associated with disease risk will occupy smaller regions, thus facilitating a more efficient search for the underlying causal variant(s). Thus, fine-mapping in African ancestry populations can be a productive approach to identify causal variants (12). In addition, the long evolutionary history of African populations and the resulting complex genetic architecture may lead to a causal variant being on different haplotype background compared with those of European and east Asian populations.

We carried out fine-mapping of the 6q25.1 region in the Black Women’s Health Study (BWHS), a prospective cohort study of 59 000 African–American women, to narrow the position of the potential causal variant for breast cancer as well as to identify any novel genetic signal associated with the disease. Replication was carried out among African–American cases of breast cancer and controls who participated in the Women’s Circle of Health Study (WCHS).

Materials and methods

Study subjects

The black women’s health study. At baseline in 1995, approximately 59 000 African–American women from all regions of the USA enrolled in the BWHS by completing mailed questionnaires that included comprehensive questions on medical history, use of medications, demographic factors, body size, reproductive history, family history of breast cancer and behavioral factors. Follow-up is by biennial mailed questionnaires and annual searches of state cancer registries and the National Death Index. Participants are asked about new diagnoses of cancer on each questionnaire, and pathology reports and/or state cancer registry data are obtained to confirm self-reports of breast cancer and to obtain data on tumor characteristics. DNA samples were obtained from BWHS participants by the mouthwash-swish method (13) with all samples stored in freezers at −80°C. Saliva samples were provided by approximately 50% of BWHS participants (26 800 women). Women who provided a sample were slightly older than women who did not, but the two groups were similar with regard to geographic region of residence, educational level, body mass index and family history of breast cancer.

Cases for this study were all participants who had been diagnosed with breast cancer as of 2010 and had provided a DNA sample. Data on ER status were available from medical records for 54% of cases; 63% were classified as ER+, and 37% as ER−.

Controls were selected from among BWHS participants with DNA samples who were free of breast cancer through 2010. Approximately, two controls were selected for each case, matched on year of birth (±1 year) and geographic region of residence (northeast, south, midwest and west). The study protocol was approved by the institutional review board of Boston University.

The WCHS. This study was designed to examine the role of genetic and non-genetic factors in relation to risk of breast cancer in African–American and European American women. The study design, enrollment criteria and collection of bio-specimens and questionnaire data of the WCHS have been described in detail (14). Briefly, women diagnosed with incident breast cancer were identified from targeted hospitals with large referral patterns for African-Americans in four boroughs of the metropolitan New York City area and from population-based rapid case ascertainment.
in seven counties in New Jersey through the New Jersey State Cancer Registry. Cases were 20–75 years of age at diagnosis, with no history of cancer other than non-melanoma skin cancer, and recently diagnosed with primary, histologically confirmed breast cancer. Controls without a history of any cancer diagnosis other than non-melanoma skin cancer living in the same area as cases were identified through random digit dialing of residential telephone and cell phone numbers and community recruitment, and were matched to cases by self-reported race and 5-year age categories. Saliva samples were collected using Oragene™ kits (DNA Genotek, Kanata, Ontario, Canada). The study protocol was approved by the institutional review boards of Roswell Park Cancer Institute, the Cancer Institute of New Jersey, Mount Sinai School of Medicine and the participating hospitals in New York City.

Selection of tag SNPs and ancestral informative markers

To select SNPs for fine-mapping, we first identified a LD block of 38 kb around the rs2046210 SNP in the HapMap release 27 of the merged phases 2 and 3 (http://hapmap.ncbi.nlm.nih.gov/) of Han Chinese in Beijing (CHB) population. The LD block was identified using the confidence interval (CI) D' algorithm developed by Gabriel et al. (15) as implemented in the Haplovew software version 4.2 (http://www.broadinstitute.org/scientific-community/science/programs/medical-and-population-genetics/haplovew/haplovew). We then downloaded SNPs covering the entire 38 kb LD block from the HapMap Yoruba (YRI) database from the Haplovew release 27 of the merged phases 2 and 3. We used the Tagger software implemented in Haplovew version 4.2 to select all tagging SNPs with a minor allele frequency ≥ 5% and r² ≥ 0.9. The index rs2046210 SNP was forced into the set. A total of 14 SNPs were selected for genotyping.

We selected 30 ancestral informative markers (AIMs) to estimate and control for population stratification due to European admixture in the BWHS. The 30 AIMs were selected from a list of validated SNPs in which the top 30 AIMs had allele frequency differences between Africans and Europeans of at least 0.75. We used a Bayesian approach, as implemented in the Admixture software (16,17) to estimate individual admixture proportions. We have shown previously that estimation of percentage European versus African ancestry with 30 AIMs correlates well with the ancestry proportion estimate obtained with a full admixture panel of approximately 1500 AIMs (18).

Genotyping and quality control

DNA from BWHS samples was isolated from the mouthwash samples at the Boston University Molecular Core Genetics Laboratory using the QIAAMP DNA Mini Kit (Qiagen). Whole genome amplification was done with the Qiagen RePLI-g Kits using the method of multiple displacement amplification with input of 50 ng of genomic DNA per reaction. Amplified samples underwent purification and PicoGreen quantification at the Broad Institute Center for Genotyping and Analysis (Cambridge, MA) before being plated for genotyping. Genotyping was also carried out at the Broad Institute Center for Genotyping and Analysis, using the Sequenom MassArray iPLEX technology. Two percent of samples were blinded duplicates included to assess reproducibility of genotypes. An average reproducibility of 99% was obtained. All SNPs with a call rate of <90% or a deviation from Hardy-Weinberg equilibrium of P < 0.001 in the control sample were excluded. We also excluded samples with call rates of <80%. The final analysis included 13 tagging SNPs in 3132 samples (1191 breast cancer cases and 1941 controls). The analysis included 409 cases classified as ER+, and 243 cases classified as ER−. Mean call rate in the final data set was 98% for SNPs and 98% for controls.

Genotyping in the WCHS was carried out at the Genomics Core Facility at Roswell Park Cancer Institute using Sequenom MassArray iPLEX technology. A panel of 98 ancestry informative markers was genotyped to ascertain genetic ancestry and control for population stratification due to genetic admixture (19). For rs2046211, genotype data were available from 665 African-American breast cancer cases and 821 African-American controls, and were analyzed in WCHS for this replication effort. About 70% of the cases had data on ER status; 69% were classified as ER+ (320 cases) and 31% as ER− (146 cases).

Data analysis

We used PLINK (20) version 1.06 to calculate summary statistics for the genotype data. We tested for association with breast cancer using the Cochran–Armitage trend test of an additive genetic model with 10 000 permutations to calculate empirical P values. We used PROC LOGISTIC of the SAS statistical software version 9.1.3 (SAS Institute, Cary, NC) to estimate odds ratios (OR) and 95% CI for the SNPs significant at the nominal values (P = 0.05). We adjusted the ORs for year of birth, geographic region of residence (northeast, south, midwest, west), place of birth (USA, foreign country) and European admixture proportion.

ORs of haplotypes of significant SNPs were estimated using an expectation substitution approach (21,22) that estimates the probabilities of all possible haplotype configurations of each individual in the sample, conditional on her genotype and case-control status. Haplotype with an estimated frequency of <1% were pooled in one single group.

Results

The A-allele of the index SNP, rs2046410, showed a breast cancer OR in the same direction in the BWHS as observed in the original genome-wide association study (OR = 1.10 in BWHS and OR = 1.29 in Zheng et al. (1)), but the BWHS association was not statistically significant (Table I; P = 0.10). However, we found that the minor allele G rs2046211, located just 82 base pairs away from the index SNP, was significantly associated with a reduced risk of breast cancer (OR = 0.83; 95% CI = 0.70–0.99; P for trend = 0.03). The association was even stronger in ER− breast cancer (OR = 0.66; 95% CI = 0.42–0.88; P for trend = 0.008).

We then asked WCHS investigators to genotype rs2046211 in their African-American case and controls samples. Results from the WCHS were in the same direction as in BWHS (Table I). In combined analyses of BWHS and WCHS results, the G-allele of rs2046211 was significantly associated with an overall reduced risk of breast cancer (OR = 0.84; 95% CI = 0.73–0.97; P for trend = 0.02). The association was somewhat stronger in ER− breast cancer (OR = 0.71; 95% CI = 0.50–1.02) as compared with ER+ breast cancer (OR = 0.88; 95% CI = 0.72–1.09).

We then conducted a joint analysis (i.e. a logistic regression model with both SNPs included) of the index rs2046210 SNP and the newly found rs2046211 SNP in the BWHS. The rs2046210-A allele was associated with increased risk of breast cancer (OR = 1.14; 95% CI = 1.02–1.28; P for trend = 0.02), and the rs2046211-G allele was associated with reduced risk (OR = 0.80; 95% CI = 0.67–0.95; for trend = 0.01) (Table II). Haplotype analysis of both SNPs confirmed these results (Table II). Haplotype frequency distributions were significantly different between breast cancer cases and controls (P = 0.008). The rs2046211-C/rs2046210-A haplotype (CA haplotype) was more frequent in cases than controls (54.2% versus 50.3%, respectively; P = 0.003) and was associated with a 14% increase of

<table>
<thead>
<tr>
<th>SNP_allele</th>
<th>Allele frequency (%)</th>
<th>Per-allele, OR (95% CI)</th>
<th>P for trend</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs2046210_A</td>
<td>All cases</td>
<td>63.3 (1149)</td>
<td>61.0 (1841)</td>
</tr>
<tr>
<td></td>
<td>ER+</td>
<td>63.2 (409)</td>
<td>62.3 (235)</td>
</tr>
<tr>
<td></td>
<td>ER−</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs2046211_G</td>
<td>All cases</td>
<td>9.1 (1144)</td>
<td>10.8 (1842)</td>
</tr>
<tr>
<td></td>
<td>ER+</td>
<td>10.3 (408)</td>
<td>9.5 (234)</td>
</tr>
<tr>
<td></td>
<td>ER−</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Replication in WCHS</td>
<td>All cases</td>
<td>8.1 (665)</td>
<td>9.3 (821)</td>
</tr>
<tr>
<td></td>
<td>ER+</td>
<td>7.2 (320)</td>
<td>7.6 (146)</td>
</tr>
<tr>
<td></td>
<td>ER−</td>
<td>8.2 (146)</td>
<td></td>
</tr>
</tbody>
</table>

A meta-analysis under a random effects model.

Table I. OR* and 95% CIs for the previously reported rs2046210 SNP and the newly identified rs2046211 SNP in the 6q25.1 region

---

*Adjusted for year of birth, geographic region of residence, place of birth and European admixture.

Meta-analysis under a random effects model.
breast cancer risk. On the other hand, the rs2046211-G/rs2046210-A haplotype (GA haplotype) was less frequent in cases than controls (9.1% versus 10.7%, respectively; \( P = 0.04 \)) and was associated with a 10% reduction of breast cancer risk.

We also assessed other SNPs in the 6q25.1 region that have been reported to be associated with breast cancer risk in African ancestry populations (2,3). None of those SNPs was associated with breast cancer risk in the BWHS (Table III). However, it is noteworthy that all the previously reported high-risk alleles were in complete or high LD with the high-risk CA haplotype that we are reporting in this study. For example, the high-risk G-allele of rs9397435 reported by Stacey et al. (3) appeared only in the presence of the high-risk CA haplotype in the BWHS (Table III). In addition, the G-allele of rs3757322 (i.e. a perfect proxy of the rs6913578 and rs7763637 reported by Cai et al. (2)) was in high LD with the high-risk CA haplotype (i.e. 97% of the rs3757322-G alleles appear in the presence of the CA haplotype). The G-allele of rs3757322 was only associated with a higher risk of breast cancer in presence of the high-risk CA haplotype (Table III). The high LD between rs9397435 and rs3757322 with the high-risk CA haplotype suggests that these two SNPs do not have independent effects on the risk of breast cancer. In conditional haplotype analysis, the association between the high-risk CA haplotype and breast cancer risk was not modified by either rs9397435 (\( P = 0.65 \)) or rs3757322 (\( P = 0.50 \)) (Table III).

**Discussion**

In this study, we report a novel SNP, rs2046211, that is associated with breast cancer risk in African–American women. The signal tagged with the high-risk CA haplotype (i.e. 97% of the rs3757322-G alleles appear in the presence of the CA haplotype). The G-allele of rs3757322 was only associated with a higher risk of breast cancer in presence of the high-risk CA haplotype (Table III). The high LD between rs9397435 and rs3757322 with the high-risk CA haplotype suggests that these two SNPs do not have independent effects on the risk of breast cancer. In conditional haplotype analysis, the association between the high-risk CA haplotype and breast cancer risk was not modified by either rs9397435 (\( P = 0.65 \)) or rs3757322 (\( P = 0.50 \)) (Table III).
by rs2046211 seems to be a different one from the signal tagged by rs2046210, which was initially identified in Chinese women (1). In our analyses, the association with rs2046210 was statistically significant only after adjusting for rs2046211, which may explain the previously failed efforts to replicate rs2046210 association with breast cancer risk in African Americans; however, the per-allele ORs approached statistical significance after control for rs2046210. Although we did not genotype either of these two SNPs, which are in perfect LD ($r^2 = 1.00$) in HapMap YRI population, we genotyped the proxy rs3757322 ($r^2 = 0.96$). Although rs3757322 was not significantly associated with breast cancer risk by itself, we found that in the presence of the high-risk CA haplotype, the G-allele of rs3757322 is indeed associated with higher risk of breast cancer. Our results show that the associations reported by Cai et al. (2) in African–American women are most probably due to the high-risk CA haplotype, as shown by the high LD between rs3757322 and the high-risk CA haplotype and the absence of independent effect of rs3757322 after conditioning for the high-risk CA haplotype.

In summary, the results provide evidence of a second signal, tagged by rs2046211, in the 6q25.1 region that is independent of rs2046210 previously reported in east Asian women. Haplotype analysis of rs2046211 and rs2046210 showed that the latter is indeed associated with breast cancer risk in African–American women after adjusting for the haplotype background that included rs2046211.

**Funding**

National Institutes of Health (R01 CA1058420 and R01 CA098663); and the Susan G. Komen for the Cure Foundation. The WCHS is supported by grants from National Institutes of Health (R01 CA100598); United States Army Medical Research and Material Command (DAMD17-01-1-0334). The research teams of both BWHS and WCHS are also supported by the National Cancer Institute (P01 CA151135). National Cancer Institute (K22 CA138563 to E.V.B.).

**Acknowledgements**

We thank the Black Women’s Health Study participants for their continuing participation in this research effort. The content is solely the responsibility of the authors and does not necessarily represent the official views of the United States Army, the National Cancer Institute or the National Institutes of Health. We thank the following state cancer registries for pathology data (AZ, CA, CO, CT, DE, DC, FL, GA, IL, IN, KY, LA, MD, MA, MI, NJ, NY, NC, OK, PA, SC, TN, TX, VA): results reported do not necessarily represent their views.

**Conflict of Interest Statement:** None declared.

**References**


Received July 19, 2012; revised October 5, 2012; accepted October 18, 2012.