



Published in final edited form as:

Cancer Res. 2011 February 15; 71(4): 1344–1355. doi:10.1158/0008-5472.CAN-10-2733.

Replication and functional genomic analyses of the breast cancer susceptibility locus at 6q25.1 generalize its importance in women of Chinese, Japanese and European ancestry

Qiuyin Cai¹, Wanqing Wen¹, Shimian Qu¹, Guoliang Li¹, Kathleen M. Egan², Kexin Chen³, Sandra L Deming¹, Hongbing Shen⁴, Chen-Yang Shen⁵, Marilie D. Gammon⁶, William J. Blot^{1,7}, Keitaro Matsuo⁸, Christopher A. Haiman⁹, Ui Soon Khoo¹⁰, Motoki Iwasaki¹¹, Regina M. Santella¹², Lina Zhang³, Alecia Malin Fair¹³, Zhibin Hu⁴, Pei-Ei Wu⁵, Lisa B. Signorello^{1,7}, Linda Titus-Ernstoff¹⁴, Kazuo Tajima⁸, Brian E. Henderson⁹, Kelvin Y.K. Chan¹⁰, Yoshio Kasuga¹⁵, Polly A. Newcomb^{16,17}, Hong Zheng³, Yong Cui¹³, Furu Wang⁴, Ya-Lan Shieh⁵, Hiroji Iwata¹⁸, Loic Le Marchand¹⁹, Sum Yin Chan¹⁰, Martha J. Shrubsole¹, Amy Trentham-Dietz^{16,20}, Shoichiro Tsugane¹¹, Montserrat Garcia-Closas²¹, Jirong Long¹, Chun Li¹, Jiajun Shi¹, Bo Huang¹, Yong-Bing Xiang²², Yu-Tang Gao²², Wei Lu²³, Xiao-Ou Shu¹, and Wei Zheng¹

¹Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, TN 37203

²Division of Cancer Prevention and Control, H. Lee Moffitt Cancer Center and Research Institute, Medical Research Center, Tampa, FL 33612

³Department of Epidemiology and Biostatistics, Tianjin Medical University Cancer Institute and Hospital, Tianjin, 300060, China

⁴Department of Epidemiology and Biostatistics, Nanjing Medical University, Nanjing 210029, China

⁵Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan

⁶Department of Epidemiology, School of Public Health, University of North Carolina, Chapel Hill, NC 27599

⁷International Epidemiology Institute, Rockville, MD 20850

⁸Division of Epidemiology and Prevention, Aichi Cancer Center Research Institute, Nagoya, Japan

⁹Department of Preventive Medicine, Keck School of Medicine, University of Southern California/Norris Comprehensive Cancer Center, Los Angeles, CA 90033

¹⁰Department of Pathology, Li Ka Shing Faculty of Medicine, University of Hong Kong, Hong Kong SAR, China

¹¹Epidemiology and Prevention Division, Research Center for Cancer Prevention and Screening, National Cancer Center, Tokyo, Japan

¹²Department of Environmental Health Sciences, Mailman School of Public Health, Columbia University, New York, NY 10032

Corresponding author contact information: Qiuyin Cai, MD, PhD Vanderbilt Epidemiology Center and Vanderbilt-Ingram Cancer Center Vanderbilt University School of Medicine 1161 21st Avenue South, MCN B-2104 Nashville, TN 37232-2400 Phone: (615) 936-1351 Fax: (615) 322-1754 qiuyin.cai@vanderbilt.edu.

Conflicts of interest: No conflicts of interest to declare.

¹³Meharry Medical College, Nashville, TN

¹⁴Dartmouth Medical School, Norris Cotton Cancer Center and Hood Center for Children and Families, Lebanon, NH 03756

¹⁵Department of Surgery, Nagano Matsushiro General Hospital, Nagano, Japan

¹⁶Paul P. Carbone Comprehensive Cancer Center, University of Wisconsin, Madison, WI 53726

¹⁷Cancer Prevention Program, Fred Hutchinson Cancer Research Center, Seattle, WA 98104

¹⁸Department of Breast Oncology, Aichi Cancer Center Central Hospital, Nagoya, Japan

¹⁹Epidemiology Program, Cancer Research Center, University of Hawaii, Honolulu, HI 96813

²⁰Department of Population Health Sciences, University of Wisconsin-Madison, Madison, WI 53726

²¹Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD 20852

²²Department of Epidemiology, Shanghai Cancer Institute, Shanghai, China

²³Shanghai Center for Disease Control and Prevention, Shanghai, China

Abstract

We evaluated the generalizability of a single nucleotide polymorphism (SNP), rs2046210 (A/G allele), associated with breast cancer risk that was initially identified at 6q25.1 in a genome-wide association study conducted among Chinese women. In a pooled analysis of over 31,000 women of East-Asian, European, and African ancestry, we found a positive association for rs2046210 and breast cancer risk in Chinese women [ORs (95% CI)=1.30(1.22–1.38) and 1.64(1.50–1.80) for the AG and AA genotypes, respectively, P for trend = 1.54×10^{-30}], Japanese women [ORs (95% CI)=1.31(1.13–1.52) and 1.37(1.06–1.76), P for trend = 2.51×10^{-4}], and European-ancestry American women [ORs (95% CI)=1.07(0.99–1.16) and 1.18(1.04–1.34), P for trend = 0.0069]. No association with this SNP, however, was observed in African American women [ORs (95% CI)=0.81(0.63–1.06) and 0.85(0.65–1.11) for the AG and AA genotypes, respectively, P for trend = 0.4027]. *In vitro* functional genomic studies identified a putative functional variant, rs6913578. This SNP is 1,440 bp downstream of rs2046210 and is in high LD with rs2046210 in Chinese ($r^2=0.91$) and European-ancestry ($r^2=0.83$) populations, but not in Africans ($r^2=0.57$). SNP rs6913578 was found to be associated with breast cancer risk in Chinese and European-ancestry American women. After adjusting for rs2046210, the association of rs6913578 with breast cancer risk in African Americans approached borderline significance. Results from this large consortium study confirmed the association of rs2046210 with breast cancer risk among women of Chinese, Japanese, and European ancestry. This association may be explained in part by a putatively functional variant (rs6913578) identified in the region.

Introduction

Breast cancer, one of the most common malignancies among women worldwide, is a complex polygenic disorder for which genetic factors play a significant role in disease etiology (1;2). We recently identified a novel genetic susceptibility locus at 6q25.1 for breast cancer risk in a genome-wide association study (GWAS) conducted among Chinese women living in Shanghai (3). A nearly 60% elevated risk for breast cancer was found among women homozygous for the variant A allele in rs2046210, a single nucleotide polymorphism (SNP) located approximately 29 kb upstream of the *ESR1* gene. It has yet to be determined whether this SNP is associated with breast cancer risk in other populations. Investigation of the association in other racial and ethnic groups is needed to determine the generalizability

of this finding and to identify causal variants for the association. In this paper, we report a pooled analysis of the association between rs2046210 and breast cancer risk in a consortium of 14 studies including over 31,000 women of East-Asian, European, and African ancestry. We also performed functional genomic studies to identify possible causal variants at this locus.

Materials and Methods

Study population

Fourteen studies contributing a total of 17,188 breast cancer cases and 14,660 controls participated in this consortium. Detailed descriptions of participating studies are included in the Supplement. Briefly, the consortium included 18,414 Chinese women from seven studies conducted in Shanghai [n=10,373; Shanghai Breast Cancer Study (SBCS)-I (3;4), SBCS-II (3), and Shanghai Breast Cancer Survival Study (SBCSS)/Shanghai Endometrial Cancer Study (SECS) (3)], Tianjin [n=3,115; Tianjin Study (5)], Nanjing [n=2,084; Nanjing Study (6;7)], Taiwan [n=2,014; Taiwan Study (8;9)], and Hong Kong [n=828; Hong Kong Study (10)]; 3,142 Japanese women from three studies conducted in Nagoya [n=1,288; Hospital-based Epidemiologic Research Program at Aichi Cancer Center (HERPACC-II (11)), Hawaii [n=1,048; Multiethnic Cohort Study (MEC) (12;13)], and Nagano [n=806; Nagano Breast Cancer Study (14)]; 8,258 European-ancestry Americans from three studies conducted in Wisconsin/Massachusetts/New Hampshire [n=3,266; Collaborative Breast Cancer Study (CBCS) (15;16)], Tennessee [n=3,060; Nashville Breast Health Study (NBHS) (3)], and New York [n=1,932; Long Island Breast Cancer Study Project (LIBCSP) (17)]; and 2,034 African Americans from two studies conducted in 12 southern U.S. states [n=1,568; Southern Community Cohort Study (SCCS) (18)] and Tennessee [n=466; NBHS] (Table 1).

Genotyping

Genotyping assays were performed at six different centers. The genotyping assay protocol was developed and validated at the Vanderbilt Molecular Epidemiology Laboratory, and TaqMan genotyping assay reagents were provided to investigators of the Tianjin study (Tianjin Cancer Institute and Hospital), Nanjing study (Nanjing Medical University), LIBCSP (Columbia University), MEC (University of Southern California), and Nagano Breast Cancer study (Japan National Cancer Center), who conducted the genotyping assays at their own laboratories. Samples from the other eight studies were genotyped at Vanderbilt using TaqMan and Affymetrix SNP arrays or at Proactive Genomics using the iPLEX™ Sequenom MassArray® platform. The Shanghai study samples were genotyped with Affymetrix Genome-Wide Human SNP Array 5.0 or 6.0 (Stage 1 of the initial GWAS) and Sequenom (Stages 2 & 3) as described previously (3). The SCCS samples were genotyped with Sequenom. All other samples were genotyped with the TaqMan assay.

Genotyping quality controls

Quality control (QC) procedures for samples from the Shanghai studies have been described previously (3). The consistency rate was 99.7% based on 2,572 comparisons with blinded QC samples and 99.2% based on 1,751 comparisons with HapMap DNA samples. For the SCCS samples genotyped with the Sequenom platform, two negative controls, two blinded duplicates, and two samples from the HapMap project were included in each 96-well plate. The QC consistency rate was 100% for blinded duplicates and 100% for the HapMap samples comparing genotyping data obtained from the current study with data obtained from the HapMap project. For TaqMan genotyping assays conducted at the Vanderbilt Molecular Epidemiology Laboratory, two negative controls and two blinded duplicates were included in each 96-well plate, along with 30 unrelated European and 45 Chinese samples from the

HapMap project for QC purposes. The consistency rate was 98.8% for the blinded duplicates and 100% for the HapMap samples comparing genotyping data obtained from the current study with data obtained from the HapMap project. Each of the non-Vanderbilt laboratories was asked to genotype a trial plate containing DNA from 46 unrelated European-ancestry and 70 Chinese-ancestry samples before the main study genotyping was conducted. The consistency rate across all centers for these trial samples was 100% compared with genotypes previously determined at Vanderbilt. In addition, replicate samples comparing 3–7% of all study samples were dispersed among the genotyping plates at all centers. The genotype distribution for rs2046210 was in Hardy-Weinberg equilibrium among controls for all participating studies with the exception of control samples from the Taiwan study ($P=0.003$). The genotype distributions for rs6929137, rs3734804, rs6913578, and rs7763637 were in Hardy-Weinberg equilibrium among controls for all participating studies.

Imputation

In order to evaluate the association of breast cancer risk with SNPs that were not directly genotyped in the initial GWA scan, we imputed the genotypes of these SNPs using the program MACH (19). MACH determines the probability distribution of missing genotypes conditional on a set of known haplotypes, while simultaneously estimating the fine-scale recombination map. For the Shanghai studies, the imputation was based on 660,118 autosomal SNPs genotyped using Affymetrix Genome-Wide Human SNP Array 6.0 with a minor allele frequency (MAF) $>1\%$ that passed the QC check and using phased HCB/JPT data from HapMap Phase II (release 22). For the National Cancer Institute Cancer Genetic Markers of Susceptibility (CGEMS) study (20), genotypes were imputed based on 513,602 autosomal SNPs genotyped using Illumina HumanHap550 BeadChip with a MAF $>1\%$ and phased CEU data from HapMap Phase II (release 22). Logistic regression was used to estimate the association of imputed SNPs of interest with breast cancer risk taking into account the degree of uncertainty of genotype imputation.

Plasmid constructs and luciferase assays

DNA fragments carrying the minor alleles of study SNPs were amplified by using PCR and cloned upstream of a luciferase reporter vector, pGL3 promoter or pGL3 basic (Promega, WI). The major alleles were generated by using a QuickChange Site-Directed Mutagenesis Kit (Stratagene, La Jolla, CA). Details on PCR primers and site-specific mutagenesis oligonucleotides are provided in the Supplement. All DNA constructs were verified by sequencing analysis. Enhancer and promoter activities were determined by transient transfection followed by an *in vitro* luciferase assay in HEK293 cells. Transfection was performed with the use of FuGene 6 Transfection Reagent (Roche Diagnostics, Indianapolis, IN) in triplicate for each of the constructs. Briefly, 2×10^5 cells were seeded in 24-well plates and co-transfected with pGL4.73, a Renilla expressing vector, which served as a reference for transfection efficiency. Thirty-six to 48 hours later the cells were lysed with Passive Lysis Buffer and luminescence (relative light units) was measured using the Dual-Luciferase Assay System (Promega, WI). Regulatory activity was measured as a ratio of firefly luciferase activity to *Renilla* luciferase activity, and the mean from at least three independent experiments are presented.

Electrophoretic mobility shift assay

Biotin-labeled, double-stranded oligonucleotide probes (details in Supplement) containing either the major or minor allele sequence were synthesized. The probes were incubated with nuclear protein extracts from HEK293 and MCF7 cells, in the presence or absence of competitors, i.e. unlabelled probes. Protein-DNA complexes were resolved by polyacrylamide gel electrophoresis and detected using a LightShift Chemiluminescent EMSA kit (Pierce Biotechnology, Rockford, IL).

Statistical analysis

Individual data were obtained from each study for a pooled analysis. Case-control differences in selected demographic characteristics and major risk factors were evaluated using t-tests (for continuous variables) and Chi-square tests (for categorical variables). Associations between SNPs and breast cancer risk were determined using odds ratios (ORs) and 95% confidence intervals (CIs) derived from logistic regression models. ORs were estimated for heterozygotes and homozygotes for the variant A allele compared with homozygotes for the common G allele. ORs were also estimated for the variant allele based on a log-additive model and adjusted for age, study site, and ethnicity, when appropriate. Adjusting for non-genetic risk factors, including age at first live birth, age at menarche, age at menopause, body mass index, participation in exercise, family history of breast cancer history, and history of benign breast diseases, did not alter the observed association, and thus only age- and study site-adjusted results are presented. Heterogeneity across studies and between ethnicities was assessed with likelihood ratio tests. Stratified analyses by ethnicity, menopausal status, and estrogen receptor (ER) status were carried out.

Results

The distributions of age and menopausal status for participating studies are shown in Table 1. Higher risk of breast cancer was consistently observed for all known major breast cancer risk factors, including a family history of breast cancer, a prior history of benign breast disease, physical inactivity, early onset of menarche, late onset of menopause, and late age at first live birth (data not shown). Except for the CBCS and SCCS, data on ER status were available from all studies.

Generalizability of the association of rs2046210 with breast cancer risk

Table 2 presents associations between rs2046210 genotypes and breast cancer risk by study site and ethnicity. The variant A allele, which was the minor allele in all groups except African Americans, was associated with increased breast cancer risk in all Chinese studies. Pooled analyses of samples from all studies conducted among Chinese women (SBCS-I, SBCS-II, SBCS/SECS, Tianjin, Nanjing, Taiwan, and Hong Kong) produced ORs of 1.30 (95% CI: 1.22–1.38) and 1.64 (95% CI: 1.50–1.80) for the AG and AA genotypes, respectively (P for trend = 1.54×10^{-30}). After excluding from the analysis the Shanghai data from which the original association was derived, the association with breast cancer was stronger; ORs were 1.26 (95% CI: 1.14–1.39) and 1.77 (95% CI: 1.55–2.02), respectively, for the AG and AA genotypes (P for trend = 2.82×10^{-17}). SNP rs2046210 was also associated with increased breast cancer risk in all three studies conducted among Japanese women (Nagoya, MEC, and Nagano), with pooled ORs of 1.31 (95% CI: 1.13–1.52) and 1.37 (95% CI: 1.06–1.76) for the AG and AA genotypes, respectively (P for trend = 2.51×10^{-4}). The heterogeneity test for results between the Chinese and Japanese studies was not statistically significant (p=0.42), therefore, all studies conducted among Chinese and Japanese women were combined into an “East-Asians” group for subsequent pooled analyses.

Among women of European ancestry, a positive association between the A allele of the rs2046210 variant and breast cancer risk was found in all three studies (NBHS, CBCS, and LIBCSP) with directly genotyped data, although the trend test was statistically significant only in the NBHS (Table 2). SNP rs2046210 was not directly genotyped in the CGEMS study. Genotype data for this SNP among 1,145 breast cancer cases and 1,142 controls were imputed (MACH score = 1.00). An association with breast cancer risk was found with ORs of 1.10 (95% CI: 0.92–1.31) and 1.26 (95% CI: 0.96–1.66) for the AG and AA genotypes, respectively, which is consistent with the data from the three studies conducted among

women of European-ancestry included in the current analysis. In pooled analyses of all samples (5,518 cases/5,027 controls) from women of European-ancestry (NBHS-White, CBCS, LIBCSP, and CGEMS), ORs were 1.07 (95% CI: 0.99–1.16) and 1.18 (95% CI: 1.04–1.34) for the AG and AA genotypes, respectively (P for trend = 0.0069) (Table 2).

SNP rs2046210 was not associated with breast cancer risk among African Americans (Table 2). In the SCCS analysis of prevalent breast cancer cases, the case and control distributions of alleles were nearly identical, while among NBHS African Americans the ORs for the AG and AA genotypes were below 1.0. In pooled analyses of African American samples (812 cases/1,222 controls) from the two studies (SCCS and NBHS-Black), ORs were 0.81 (95% CI: 0.63–1.06) and 0.85 (95% CI: 0.65–1.11) for the AG and AA genotypes, respectively (P for trend = 0.40). The sample size for African Americans included in this study, however, was small and the frequency of the A allele in the African American population (62.0%) was considerably higher than that in the East-Asian (34.8%) and European-ancestry (35.5%) populations. Fig. 1 presents a forest plot summarizing the results of these studies. We also performed analyses stratified by menopausal and estrogen receptor status and found that the association with rs2046210 is more evident for ER(–) breast cancer compared with ER(+) breast cancer (P=0.0004) in East-Asian women but not in women of European ancestry (Table 3).

Functional genomic studies of the chr 6q25.1 locus

There are two non-synonymous SNPs (rs6929137 and rs3734804) in the *C6orf97* gene, which is in the 6q25.1 locus. These two SNPs are in strong linkage disequilibrium (LD) with rs2046210 ($r^2=0.91$ in Chinese, 0.87 in Europeans, and 0.001 in Africans for rs6929137; $r^2=0.91$ in Chinese, 0.56 in Europeans, and 0.42 in Africans for rs3734804). In an attempt to identify SNPs that may be more strongly associated with breast cancer risk in women of European ancestry than the originally-reported SNP (rs2046210), we genotyped these two SNPs in 1,592 European-ancestry American cases and 1,468 controls from the NBHS (NBHS-White). The variant alleles of the two SNPs were also associated with breast cancer risk (per variant allele OR = 1.11 (95% CI: 0.99–1.24) for rs6929137 and 1.12 (95% CI: 1.01–1.24) for rs3734804). The associations, however, were not stronger than the initially-reported SNP rs2046210 in the NBHS-White group (OR per variant allele = 1.15, 95% CI: 1.04–1.28). These two SNPs are not included in Affymetrix Genome-Wide Human SNP Array 6.0 and thus we imputed genotype data for these two SNPs. Again, these SNPs showed a significant association with breast cancer risk in the Shanghai samples (2,073 cases and 2,084 controls) [ORs per variant allele, 1.26 (95% CI: 1.15–1.39) for rs6929137 and 1.27 (95% CI: 1.16–1.39) for rs3734804]. The associations with these two SNPs were slightly stronger than with the initially-reported SNP rs2046210 identified in the GWAS (OR per variant allele = 1.25 (95% CI: 1.14–1.36). However, rs6929137 was not associated with breast cancer risk in African American (21). Thus, further evaluations of these two SNPs were not conducted.

To evaluate whether SNP rs2046210 has any regulatory function, we conducted luciferase reporter assays. The reporter construct containing the major G allele and the construct containing the minor A allele produced similar levels of luciferase activity. The results of a search for transcription factor binding sites (the “TFBS Conserved” track of the UCSC Genome Browser (22)) showed that rs2046210 does not alter putative transcription factor binding..

To identify potential causal SNPs, we performed a series of heterologous promoter and enhancer assays, focusing on the 36 kb region between the *C6orf97* and *ESR1* genes. We divided the 36 kb region (chromosome 6:151,983,304–152,019,420) into 4 parts and used long-range PCR to amplify 4 DNA fragments [a, 8.6 kb (harboring rs2046210 polymorphic

site); b, 9.2 kb; c, 9.3 kb; and d, 8.9 kb], then cloned the 4 fragments into both pGL3 basic and pGL3 promoter vectors (Fig. 2A). The templates for PCR were DNA carrying the minor or major alleles of SNP rs2046210. The SNP rs2046210A construct carried the minor alleles of SNP rs2046210 and other SNPs in close proximity and strong LD with rs2046210, while the rs2046210G construct carried the major alleles of rs2046210 and other SNPs in close proximity and strong LD with rs2046210. Luciferase activity derived from the fragment “a” construct in the pGL3 promoter vector with rs2046210A was significantly different from the fragment “a” construct with rs2046210G (data not shown). To refine the location of potential causal SNPs, we subdivided fragment “a” into 3 smaller DNA fragments [e, 2.2 kb; f, 4.1 kb; and g, 2.3 kb] containing either rs2046210A or rs2046210G into the pGL3 promoter vector and carried out luciferase assays. Luciferase activity derived from fragment “g” with rs2046210A was significantly different from that of fragment “g” with rs2046210G (data not shown). Six SNPs, including rs2046210, in fragment “g” were associated with breast cancer risk in Stage 1 of the initial SBCS GWAS. After excluding SNPs that showed no evidence of alteration of putative transcription factor binding in the database search, we found 3 candidate SNPs (rs7740686, rs7763637, and rs6913578) in this region (Fig 2A). We then generated major allele constructs for each of these 3 SNPs by using site-directed mutagenesis by using the 2.3 kb fragment “g” with the rs2046210A construct as the template. Luciferase activity was significantly higher in constructs harboring the major alleles of rs6913578 (rs6913578-A in Fig. 2B) or rs7763637 (rs7763537-G in Fig. 2B) compared with the corresponding minor alleles (minor alleles in Fig. 2B).

To investigate whether the DNA sequences containing rs6913578 or rs7763637 interact with nuclear proteins and, if so, whether these SNP alter protein-DNA interactions, we performed electrophoretic mobility shift assays. We found that the minor allele (C) of rs6913578 significantly altered DNA-protein complex (II) intensity in both HEK293 and MCF7 cells (Fig. 2C), whereas there was no detectable interaction of rs7763637 with nuclear proteins (data not shown).

Evaluation of putative functional variants with breast cancer risk

SNP rs6913578 is located 1,440 bp downstream of rs2046210. SNP rs2046210 is in strong LD with rs6913578 and rs7763637 in Chinese populations ($r^2 = 0.91$ and 0.901 , respectively) and European-ancestry populations ($r^2 = 0.83$ and 0.87 , respectively), but is not in African populations. Both rs6913578 and rs7763637 are associated with breast cancer risk in Chinese women and European-ancestry Americans, and the association was stronger than with rs2046210 (Table 4). The positive associations of these SNPs with breast cancer risk diminished (Table 4) after adjusting for rs2046210, which is not surprising given the high LD with rs2046210. These two SNPs, rs6913578 and rs7763637, showed weak associations with breast cancer risk in African Americans (Table 4). After adjusting for rs2046210, however, the associations in African Americans approached borderline significance (P for trend = 0.096 and 0.077 , respectively).

Discussion

In this pooled analysis of 17,188 cases and 14,660 controls, we confirmed the association of rs2046210 at 6q25.1 with breast cancer risk among women with Chinese, Japanese, and European ancestry. *In vitro* functional genomic studies identified a putatively functional variant, rs6913578, a SNP 1,440 bp downstream of rs2046210, which is in high LD with rs2046210 in Chinese and European-ancestry populations, but is not in Africans. SNP rs6913578 had a stronger association with breast cancer risk in European-ancestry Americans than rs2046210, the SNP originally associated with breast cancer risk in a GWAS conducted in a Chinese population. In African Americans, the association of

rs6913578 with breast cancer risk approached borderline significance after adjusting for rs2046210.

Genes that are located in the 1 Mb region centered around rs2046210 include PLEKHG1, MTHFD1L, AKAP12, ZBTB2, RMND1, C6orf211, C6orf97, ESR1, C6orf98, SYNE1, and *NANOGP11*. SNP rs2046210 is located 29 kb upstream of the first untranslated region of the *ESR1* gene, 180 kb upstream of the transcription start site of its first exon (3;23), and 6 kb downstream of the *C6orf97* gene. Because of its relative proximity to the *ESR1* gene and the biological function of ER- α , it is possible that SNP rs2046210, or SNPs in LD with it, may alter *ESR1* gene expression and thereby affect susceptibility to breast cancer. A search of predicted transcription factor binding sites using the “TFBS Conserved” track of the UCSC Genome Browser (22) indicated that there is no transcription factor binding site on this SNP. We further scanned for non-coding RNA (EvoFold), as well as miRNA/snoRNA/scaRNA (sno/miRNA), in this region by using the UCSC Genome Browser and found that this SNP is not in the coding region for any non-coding RNA or miRNA/snoRNA/scaRNA. Our functional genomic analyses also provided no support for the potential functionality of rs2046210.

Our *in vitro* functional genomic experiments indicated that the location of the potential functional SNPs may be in a 2.3 kb region. Specifically, SNPs rs6913578, which is 1,440 bp downstream of rs2046210, and rs7763637, which is 947 bp downstream of rs2046210, altered luciferase reporter activity. These results suggest that these two common SNPs may influence DNA binding protein interactions and affect the expression of neighboring genes. We conducted electrophoretic mobility shift assays to examine this hypothesis and confirmed that the C allele of rs6913578 significantly altered DNA-nuclear protein interaction. Thus, it is possible that nuclear protein(s) selectively and differently bind to specific alleles of the rs6913578 polymorphic site resulting in modification of the transcription of neighboring genes. However, there has been no confirmation to date that the putative transcription factors or their associated proteins are involved in the regulation of *ESR1*, *C6orf97*, or nearby genes. Interestingly, both rs6913578 and rs7763637 were associated with breast cancer risk among Chinese women and European-ancestry Americans (and the associations were stronger than rs2046210 in European-ancestry Americans), but not among African American women. However, after adjusting for rs2046210, the association of rs6913578 and rs7763637 with breast cancer risk in African Americans approached a borderline significance level. Our data demonstrate and highlight the importance of conducting inter-racial genetic association studies in populations with different LD structures to identify potential causal genetic variants for breast cancer and other complex diseases. Further studies will be required to determine causal SNPs related to breast cancer risk at the 6q25.1 locus.

In a recent study, Stacey et al. (24) reported an association of rs9397435 at the 6q25.1 locus with breast cancer risk in European, Chinese, and African populations. This SNP is located 2,854 bp downstream of rs2046210 and 1,414 bp downstream of rs6913578 and is only weakly correlated with rs2046210 in Europeans ($r^2 = 0.087$) and African ($r^2 = 0.039$) populations. The risk allele frequency of this SNP in Asians is approximately 32%, comparable to that found for rs2046210, but it is very low, only about 6.3% among the European and African populations. Very recently, Turnbull et al. (25) evaluated the 6q25.1 locus with breast cancer risk in a GWAS conducted among 3,659 European-ancestry cases and 4,897 similar controls and identified SNP rs3757318 (MAF = 7%) to have the most significant association with breast cancer risk. This SNP is located ~200 kb upstream of *ESR1* in an intron of the *C6orf97* gene and 34,253 bp upstream of rs2046210. SNP rs3757318 is only weakly correlated with rs2046210 in Europeans ($r^2 = 0.088$), while the correlation is stronger in Chinese populations ($r^2 = 0.48$). Similarly, SNP rs3757318 is only

weakly correlated with rs6913578 in European ($r^2 = 0.038$) and Chinese populations ($r^2 = 0.181$). Using imputed data from our GWAS, we showed that rs3757318 was associated with breast cancer risk with a per variant allele OR of 1.21 (P for trend = 5.4×10^{-4}), an association that is not as strong as rs2046210. It is unclear, however, whether this SNP is functional or is related to breast cancer risk in women of African ancestry.

Results reported to date from GWAS have clearly shown that GWAS results cannot be applied uniformly across all ethnic groups. Several SNPs identified in GWAS conducted among women of European ancestry could not be replicated in Asian-ancestry women (26–30). In our study, we have shown that the strength of the association with rs2046210 varies considerably across ethnic groups. This is not surprising given that most, if not all SNPs identified in GWAS are tagging SNPs, and there exists considerable differences in genetic architecture across ethnic groups. Fine-mapping studies are needed to identify additional genetic risk variants and/or causal variants for breast cancer.

Major strengths of our study are its large sample size and its ability to evaluate the consistency of the findings across multiple studies conducted in different locations and in populations with different ethnic ancestry. In addition, we conducted functional genomic studies of this locus to identify possible functional variants. Ancestry informative markers, however, were not adjusted for in this study. In addition, our *in vitro* functional experiments were conducted only on a 36 Kb region and then were narrowed down to 4 common polymorphisms in a 2.3 Kb region. It is possible that other functional SNPs, both common and rare, exist at this locus.

In summary, results from this large consortium study confirmed the association of rs2046210 with breast cancer risk among Chinese women, Japanese women, and European-ancestry Americans. SNP rs6913578 may be a functional SNP responsible for the observed association with breast cancer risk of SNPs at the 6q25.1 locus. Additional fine-scale mapping studies are needed to identify causal variants at this locus.

Précis

Results from this large consortium study broaden the significance of a chromosomal marker associated with breast cancer risk in women of diverse ancestry.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors wish to thank study participants and research staff for their contributions and commitment to this project. This research was supported by US NIH grant R01CA124558. The genotyping assays conducted at Vanderbilt University were performed at the Survey and Biospecimen Core, which is supported in part by the Vanderbilt-Ingram Cancer Center (P30 CA68485). Participating studies (Principal Investigator, grant support) of the consortium are as follows: the Shanghai Breast Cancer Study (W. Zheng, R01CA64277), the Nashville Breast Health Study (W. Zheng, R01CA100374), the Shanghai Breast Cancer Survival Study (X.O. Shu, R01CA118229), the Shanghai Endometrial Cancer Study (X.O. Shu, R01CA92585, contributed only controls to the consortium), the Tianjin Study (K. Chen, the National Natural Science Foundation of China Grant No. 30771844), the Nanjing Study (H. Shen, IRT0631, China), the Taiwan Biobank Study (C.Y. Shen, DOH97-01), the Hong Kong Study (U.S. Khoo, Research Grant Council, Hong Kong SAR, China, HKU 7520/05M and 76730M), the Nagoya study (K. Tajima, Grants-in-Aid for Scientific Research on Priority Areas (17015052) from the Ministry of Education, Culture, Sports, Science, and Technology of Japan; H. Tanaka, Grants-in-Aid for the Third Term Comprehensive Ten-Year Strategy for Cancer Control from the Ministry of Health, Labor and Welfare of Japan, H20-002), the Multiethnic Cohort Study (B. Henderson, CA63464; L. Kolonel, CA54281; and C. Haiman, CA132839), the Nagano Breast Cancer Study [S. Tsugane, Grants-in-Aid for the Third Term Comprehensive Ten-Year Strategy for

Cancer Control from the Ministry of Health, Labor and Welfare of Japan, and for Scientific Research on Priority Areas (17015049) from the Ministry of Education, Culture, Sports, Science, and Technology of Japan], the Collaborative Breast Cancer Study including Massachusetts (K. Egan, R01 CA47305), Wisconsin (P. Newcomb, R01 CA47147) and New Hampshire (L. Titus-Ernstoff, R01 CA69664) centers, the Long Island Breast Cancer Study Project (M.D. Gammon, U01 CA/ES66572; R.M. Santella, P30ES009089; J.A. Swenberg, P30ES010126), and the Southern Community Cohort Study (W.J. Blot, R01CA092447). The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding agencies.

Grant support: This work was supported primarily by a grant from the National Institutes of Health (R01CA124558).

Reference List

- (1). Nathanson KL, Wooster R, Weber BL. Breast cancer genetics: what we know and what we need. *Nat Med.* May; 2001 7(5):552–6. [PubMed: 11329055]
- (2). Balmain A, Gray J, Ponder B. The genetics and genomics of cancer. *Nat Genet.* Mar; 2003 33(Suppl):238–44. [PubMed: 12610533]
- (3). Zheng W, Long J, Gao YT, Li C, Zheng Y, Xiang YB, et al. Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. *Nat Genet.* Mar; 2009 41(3):324–8. [PubMed: 19219042]
- (4). Gao YT, Shu XO, Dai Q, Potter JD, Brinton LA, Wen W, et al. Association of menstrual and reproductive factors with breast cancer risk: results from the Shanghai Breast Cancer Study. *Int J Cancer.* 2000; 87(2):295–300. [PubMed: 10861490]
- (5). Zhang L, Gu L, Qian B, Hao X, Zhang W, Wei Q, et al. Association of genetic polymorphisms of ER-alpha and the estradiol-synthesizing enzyme genes CYP17 and CYP19 with breast cancer risk in Chinese women. *Breast Cancer Res Treat.* Mar; 2009 114(2):327–38. [PubMed: 18629629]
- (6). Liang J, Chen P, Hu Z, Zhou X, Chen L, Li M, et al. Genetic variants in fibroblast growth factor receptor 2 (FGFR2) contribute to susceptibility of breast cancer in Chinese women. *Carcinogenesis.* Dec; 2008 29(12):2341–6. [PubMed: 18845558]
- (7). Wang Y, Hu Z, Liang J, Wang Z, Tang J, Wang S, et al. A tandem repeat of human telomerase reverse transcriptase (hTERT) and risk of breast cancer development and metastasis in Chinese women. *Carcinogenesis.* Jun; 2008 29(6):1197–201. [PubMed: 18413362]
- (8). Ding SL, Yu JC, Chen ST, Hsu GC, Kuo SJ, Lin YH, et al. Genetic variants of BLM interact with RAD51 to increase breast cancer susceptibility. *Carcinogenesis.* Jan; 2009 30(1):43–9. [PubMed: 18974064]
- (9). Hsu HM, Wang HC, Chen ST, Hsu GC, Shen CY, Yu JC. Breast cancer risk is associated with the genes encoding the DNA double-strand break repair Mre11/Rad50/Nbs1 complex. *Cancer Epidemiol Biomarkers Prev.* Oct; 2007 16(10):2024–32. [PubMed: 17932350]
- (10). Chan KY, Liu W, Long JR, Yip SP, Chan SY, Shu XO, et al. Functional polymorphisms in the BRCA1 promoter influence transcription and are associated with decreased risk for breast cancer in Chinese women. *J Med Genet.* Jan; 2009 46(1):32–9. [PubMed: 18782836]
- (11). Hamajima N, Matsuo K, Saito T, Hirose K, Inoue M, Takezaki T, et al. Gene-environment Interactions and Polymorphism Studies of Cancer Risk in the Hospital-based Epidemiologic Research Program at Aichi Cancer Center II (HERPACC-II). *Asian Pac J Cancer Prev.* 2001; 2(2):99–107. [PubMed: 12718640]
- (12). Haiman CA, Garcia RR, Hsu C, Xia L, Ha H, Sheng X, et al. Screening and association testing of common coding variation in steroid hormone receptor co-activator and co-repressor genes in relation to breast cancer risk: the Multiethnic Cohort. *BMC Cancer.* 2009; 9:43. [PubMed: 19183483]
- (13). Kolonel LN, Henderson BE, Hankin JH, Nomura AM, Wilkens LR, Pike MC, et al. A multiethnic cohort in Hawaii and Los Angeles: baseline characteristics. *Am J Epidemiol.* Feb 15; 2000 151(4):346–57. [PubMed: 10695593]
- (14). Itoh H, Iwasaki M, Hanaoka T, Kasuga Y, Yokoyama S, Onuma H, et al. Serum organochlorines and breast cancer risk in Japanese women: a case-control study. *Cancer Causes Control.* Jul 1; 2009 20(5):567–80. [PubMed: 19031103]

- (15). Garcia-Closas M, Egan KM, Newcomb PA, Brinton LA, Titus-Ernstoff L, Chanock S, et al. Polymorphisms in DNA double-strand break repair genes and risk of breast cancer: two population-based studies in USA and Poland, and meta-analyses. *Hum Genet.* May; 2006 119(4): 376–88. [PubMed: 16485136]
- (16). Sprague BL, Trentham-Dietz A, Newcomb PA, Titus-Ernstoff L, Hampton JM, Egan KM. Lifetime recreational and occupational physical activity and risk of in situ and invasive breast cancer. *Cancer Epidemiol Biomarkers Prev.* Feb; 2007 16(2):236–43. [PubMed: 17301255]
- (17). Gammon MD, Neugut AI, Santella RM, Teitelbaum SL, Britton JA, Terry MB, et al. The Long Island Breast Cancer Study Project: description of a multi-institutional collaboration to identify environmental risk factors for breast cancer. *Breast Cancer Res Treat.* Jun; 2002 74(3):235–54. [PubMed: 12206514]
- (18). Signorello LB, Hargreaves MK, Steinwandel MD, Zheng W, Cai Q, Schlundt DG, et al. Southern community cohort study: establishing a cohort to investigate health disparities. *J Natl Med Assoc.* 2005; 97(7):972–9. [PubMed: 16080667]
- (19). Center for Statistical Genetics. MACH 1.0. [Accessed: July 1, 2009]. www.sph.umich.edu/csg/abecasis/mach.
- (20). National Cancer Institute. Cancer Genetic Markers of Susceptibility (CGWMS). [Accessed: July 1, 2009]. <http://cgems.cancer.gov/>.
- (21). Zheng W, Cai Q, Signorello LB, Long J, Hargreaves MK, Deming SL, et al. Evaluation of 11 breast cancer susceptibility loci in African-American women. *Cancer Epidemiol Biomarkers Prev.* Oct; 2009 18(10):2761–4. [PubMed: 19789366]
- (22). UCSC Genome Browser. [Accessed: July 1, 2009]. <http://genome.ucsc.edu>.
- (23). Kos M, Reid G, Denger S, Gannon F. Minireview: genomic organization of the human ERalpha gene promoter region. *Mol Endocrinol.* Dec; 2001 15(12):2057–63. [PubMed: 11731608]
- (24). Stacey SN, Sulem P, Zanon C, Gudjonsson SA, Thorleifsson G, Helgason A, et al. Ancestry-shift refinement mapping of the C6orf97-ESR1 breast cancer susceptibility locus. *PLoS Genet.* 2010; 6(7):e1001029. [PubMed: 20661439]
- (25). 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. *Nat Genet.* Jun; 2010 42(6): 504–7. [PubMed: 20453838]
- (26). Ahmed S, Thomas G, Ghoussaini M, Healey CS, Humphreys MK, Platte R, et al. Newly discovered breast cancer susceptibility loci on 3p24 and 17q23.2. *Nat Genet.* May; 2009 41(5): 585–90. [PubMed: 19330027]
- (27). Easton DF, Pooley KA, Dunning AM, Pharoah PD, Thompson D, Ballinger DG, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature.* Jun 28; 2007 447(7148):1087–93. [PubMed: 17529967]
- (28). 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. *Nat Genet.* Jul; 2007 39(7):865–9. [PubMed: 17529974]
- (29). Long J, Shu XO, Cai Q, Gao YT, Zheng Y, Li G, et al. Evaluation of breast cancer susceptibility Loci in Chinese women. *Cancer Epidemiol Biomarkers Prev.* Sep; 2010 19(9):2357–65. [PubMed: 20699374]
- (30). Zheng W, Wen W, Gao YT, Shyr Y, Zheng Y, Long J, et al. Genetic and clinical predictors for breast cancer risk assessment and stratification among Chinese women. *J Natl Cancer Inst.* Jul 7; 2010 102(13):972–81. [PubMed: 20484103]

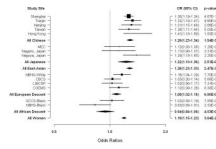


Figure 1. ORs (95%) per risk allele for breast cancer by study site and ethnicity. The size of the boxes is proportional to the sample size of each study. The width of the diamonds represents the range of confident intervals of combined ORs derived from meta-analyses.

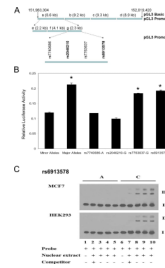


Figure 2.

In vitro functional characterization of SNP rs2046210 and other potential functional SNPs in 6q25.1. **A.** Diagram of cloning strategy. A 36 kb region (chromosome 6:151,983,304–152,019,420) between the *C6orf97* and *ESR1* genes was divided into 4 DNA fragments (a–d), which were separately cloned into pGL3 basic and pGL3 promoter vectors. The 8.6 kb “a” fragment was further divided to three DNA fragments (e, f, and g) and subcloned into a pGL3 promoter vector. The “g” fragment harbored 4 SNPs (rs7740686, rs2046210, rs7763637, and rs6913578). **B.** Luciferase reporter activity assays: HEK293 cells were transiently transfected with pGL3 promoter/luciferase reporter constructs containing the 2.3 Kb “g” fragment. 1. Minor Alleles construct: contained the minor alleles for all four SNPs (rs7740686-T, rs2046210-A, rs7763637-A, and rs6913578 C); 2. Major Alleles construct: contained the major alleles for all four SNPs (rs7740686-A, rs2046210-G, rs7763637-G, and rs6913578-A); 3. rs7740686-A construct: contained the rs7740686 major allele A and the minor alleles for the other three SNPs; 4. rs2046210-G construct: contained the rs2046210 major allele G and the minor alleles for the other three SNPs; 5. rs7763637-G construct: contained rs7763637 major allele G and the minor alleles for the other three SNPs; 6. rs6913578-A construct: contained rs6913578 major allele A and the minor alleles for the other three SNPs. Relative luciferase activity is shown as the mean ± SD of three experiments conducted in triplicate (relative to the Minor Allele construct). Statistical analysis was conducted by using Student's t test to compare the minor and major alleles (* $P < 0.01$ when compared with the minor alleles, $n = 9$). Relative luciferase activity is shown as the mean ± SD of three experiments conducted in triplicate (relative to the Minor Alleles construct). Statistical analysis was conducted using Student's t test to compare the minor and major alleles (* $P < 0.01$ when compared with the minor alleles, $n = 9$). **C.** Electrophoretic mobility shift assays. Nuclear protein extracts from MCF-7 (top panel) and HEK293 (bottom panel) cells were incubated with biotin-labeled probes corresponding to reference allele (lanes 1–5) or the risk allele (lanes 6–10) in the absence or presence of competitors. Lanes 1 and 6, no nuclear extracts; lanes 2 and 7, unlabeled competitor in 200-fold molar excess; lanes 3 and 8 (5 mM MgCl₂), lanes 4 and 9 (2.5 mM MgCl₂), and lanes 5 and 10 (1.25 mM MgCl₂), no competitor. I: free biotin-labeled probes. II: specific DNA-protein complex bands.

Table 1

Characteristics of studies participating in the breast cancer consortium

Study (reference)	Ethnicity	Study design ^d	Study period	N ^b	Age (mean)	Menopause (%)	ER (+) (%) ^c
SBCS-I ^d (3;4)	Chinese	Population	1996–1998	1105/1213	47.5/47.3	32.7/36.3	63.8
SBCS-II ^d (3)	Chinese	Population	2002–2005	1915/1836	50.9/51.7 ^e	43.6/49.4 ^e	63.6
SBCSS/SECS ^d (3)	Chinese	Population	2002–2006	3405/899	55.0/54.9	53.7/61.7 ^e	64.5
Tianjin (5)	Chinese	Hospital	2004–2008	1532/1583	51.7/51.9	51.7/55.4 ^f	44.3
Nanjing (6;7)	Chinese	Hospital	2004–2008	1050/1034	51.6/52.0	53.7/55.2	55.4
Taiwan (8;9)	Chinese	Hospital	2004–2007	1001/1013	51.6/47.4 ^e	52.7/39.8 ^e	65.9
Hong Kong (10)	Chinese	Hospital	2003–2004	407/421	45.5/45.4	52.2/41.4 ^e	72.0
Nagoya, Japan ^d (11)	Japanese	Hospital	2000–2005	644/644	51.4/51.1	48.5/48.5	72.8
MEC ^d (12;13)	Japanese	Population	1993–2008	541/507	65.1/60.3 ^e	86.4/83.3	86.2
Nagano, Japan ^d (14)	Japanese	Hospital	2001–2005	403/403	53.7/53.9	54.6/65.0 ^e	74.8
NBHS-White ^d (3)	European	Population	2001–2008	1592/1468	54.8/52.2 ^e	65.5/58.9 ^e	74.9
CBCC ^d (15;16)	European	Population	1998–2001	1828/1438	53.7/53.4	59.0/61.0	NA ^f
LBCSP ^d (17)	European	Population	1996–1997	953/979	58.8/56.7 ^e	67.6/66.5	76.7
CGEMS	European	Population	NA ^g	1145/1142	NA ^g	NA ^g	NA
SCCS-Black ^d (18)	African	Population	2002–2008	522/1046	48.1/56.6 ^e	59.5/76.7 ^e	NA
NBHS-Black (3)	African	Population	2001–2008	290/176	54.5/52.1 ^f	70.7/61.9	NA

^aWith the exception of the MEC and SCCS, all other studies used the case-control study design using either a population-based or hospital-based approach.^bCases/controls.^cAmong cases with estrogen receptor (ER) data^dSBCS-I: Shanghai Breast Cancer Study-I; SBCS-II: Shanghai Breast Cancer Study-II; SBCSS: Shanghai Breast Cancer Survival Study; SECS: Shanghai Endometrial Cancer Study; Nagoya, Japan: Hospital-based Epidemiologic Research Program at Aichi Cancer Center; Nagano, Japan: Nagano Breast Cancer study; MEC: Multiethnic Cohort Study; LBCSP: Long Island Breast Cancer Study Project; CBCCS: Collaborative Breast Cancer Study; NBHS: Nashville Breast Health Study; SCCS: Southern Community Cohort Study.^eSignificant at $\alpha=0.01$ level (t-test for continuous variables, Chi-square test for categorical variables).^fSignificant at $\alpha=0.05$ level (t-test for continuous variables, Chi-square test for categorical variables).

⁸Data not available.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 2

Association of SNP rs2046210 with breast cancer risk by study site and ethnicity

	Frequency of the A allele (%)				GG		AG		AA		P for trend
	N ^a	OR (95% CI)	N ^a	OR (95% CI)	N ^a	OR (95% CI)	N ^a	OR (95% CI)	N ^a	OR (95% CI)	
By study site^b											
Shanghai	41.9/36.4	2144/1611	1.00 ^d	1.33(1.22–1.45)	3183/1802	1.33(1.22–1.45)	1098/535	1.54(1.37–1.74)	4.67 × 10 ⁻¹⁵		
Tianjin	42.1/35.5	512/655	1.00	1.31(1.12–1.53)	750/732	1.31(1.12–1.53)	270/196	1.76(1.42–2.19)	9.60 × 10 ⁻⁸		
Nanjing	43.2/36.8	341/415	1.00	1.30(1.08–1.57)	510/477	1.30(1.08–1.57)	199/142	1.71(1.32–2.21)	2.66 × 10 ⁻⁵		
Taiwan	42.0/36.7	334/384	1.00	1.11(0.91–1.34)	494/514	1.11(0.91–1.34)	173/115	1.73(1.31–2.28)	5.02 × 10 ⁻⁴		
Hong Kong	45.0/36.4	129/256	1.00	1.36(1.02–1.80)	190/278	1.36(1.02–1.80)	88/87	2.01(1.40–2.89)	1.63 × 10 ⁻⁴		
Nagoya, Japan	34.0/26.9	273/349	1.00	1.60(1.27–2.02)	295/236	1.60(1.27–2.02)	69/54	1.63(1.11–2.41)	1.28 × 10 ⁻⁴		
MEC-Japanese	28.7/26.3	280/277	1.00	1.08(0.84–1.40)	211/193	1.08(0.84–1.40)	50/37	1.34(0.85–2.11)	0.2251		
Nagano, Japan	30.3/28.0	195/214	1.00	1.24(0.93–1.66)	172/152	1.24(0.93–1.66)	36/37	1.07(0.65–1.76)	0.3310		
NBHS-White	37.6/34.4	613/618	1.00	1.11(0.95–1.29)	761/691	1.11(0.95–1.29)	218/159	1.38(1.10–1.75)	0.0077		
CBCS-White	37.3/36.6	706/567	1.00	1.03(0.89–1.19)	882/690	1.03(0.89–1.19)	240/181	1.07(0.85–1.33)	0.5684		
LIBCSP-White	37.4/36.8	370/391	1.00	1.05(0.87–1.28)	454/455	1.05(0.87–1.28)	129/133	1.02(0.77–1.36)	0.7309		
CGEMS-White	36.9/34.5	446/478	1.00	1.10(0.92–1.31)	554/541	1.10(0.92–1.31)	145/123	1.26(0.96–1.66)	0.0838		
SCCS-Black	62.6/62.1	74/149	1.00	0.99(0.72–1.35)	242/494	0.99(0.72–1.35)	206/403	1.03(0.74–1.43)	0.7848		
NBHS-Black	57.2/61.4	62/26	1.00	0.62(0.36–1.06)	124/84	0.62(0.36–1.06)	104/66	0.66(0.38–1.15)	0.2332		
By ethnic group^c											
Chinese	42.2/36.3	3460/3321	1.00	1.30(1.22–1.38)	5127/3803	1.30(1.22–1.38)	1828/1075	1.64(1.50–1.80)	1.54 × 10 ⁻³⁰		
Chinese (excl. Shanghai)	42.7/36.2	1316/1710	1.00	1.26(1.14–1.39)	1944/2001	1.26(1.14–1.39)	730/540	1.77(1.55–2.02)	2.82 × 10 ⁻¹⁷		
Japanese	31.3/27.0	748/840	1.00	1.31(1.13–1.52)	678/581	1.31(1.13–1.52)	155/128	1.37(1.06–1.76)	2.51 × 10 ⁻⁴		
East-Asians	40.7/34.8	4208/4161	1.00	1.30(1.22–1.38)	5805/4384	1.30(1.22–1.38)	1983/1203	1.61(1.48–1.76)	2.47 × 10 ⁻³³		
European ancestry	37.3/35.5	2135/2054	1.00	1.07(0.99–1.16)	2651/2377	1.07(0.99–1.16)	732/596	1.18(1.04–1.34)	0.0069		
African Americans	60.7/62.0	136/175	1.00	0.81(0.63–1.06)	366/578	0.81(0.63–1.06)	310/469	0.85(0.65–1.11)	0.4027		
All women^c	40.6/38.0	6479/6469	1.00	1.20(1.15–1.26)	8822/7605	1.20(1.15–1.26)	3025/2482	1.41(1.32–1.50)	3.64 × 10 ⁻²⁷		

^a Cases/controls.^b Adjusted for age and study site.

^c Adjusted for age, study site, and ethnicity.

^d Reference group

^e P values derived from homogeneity tests are 0.7615 for seven studies conducted among Chinese women, 0.4222 between Chinese and Japanese studies, 0.3529 across four studies conducted among women of European ancestry, and $p=3.18 \times 10^{-6}$ between East-Asian and European-ancestry women.

Table 3

Association of SNP rs2046210 with breast cancer risk by ethnicity, menopausal status, and ER status^a

	East-Asians		European-ancestry Americans ^b	
	Cases	Controls	OR (95% CI)	OR (95% CI)
All women				
GG	4208	4161	1.00(reference)	1.00(reference)
AG	5805	4384	1.30(1.22–1.38)	1.06(0.97–1.17)
AA	1983	1203	1.61(1.48–1.76)	1.16(1.01–1.33)
Per A allele			1.28(1.23–1.33)	1.07(1.01–1.14)
p for trend			2.47 × 10 ⁻³³	0.0321
Pre-menopausal women				
GG	2040	1867	1.00(reference)	1.00(reference)
AG	2827	2075	1.23(1.13–1.34)	1.08(0.93–1.27)
AA	956	572	1.50(1.33–1.69)	1.10(0.87–1.39)
Per A allele			1.23(1.16–1.30)	1.16(0.95–1.18)
p for trend			4.35 × 10 ⁻¹²	0.2998
Post-menopausal women				
GG	2127	2126	1.00(reference)	1.00(reference)
AG	2910	2121	1.36(1.25–1.48)	1.02(0.90–1.15)
AA	1002	577	1.71(1.52–1.93)	1.17(0.97–1.40)
Per A allele			1.32(1.25–1.40)	1.06(0.98–1.16)
p for trend			4.48 × 10 ⁻²²	0.1562
			p for interaction with menopause: 0.0850	p for interaction with menopause: 0.9816
ER(+)^c				
GG	2418	4161	1.00(reference)	1.00(reference)
AG	3142	4384	1.24(1.16–1.33)	1.08(0.93–1.26)
AA	1043	1203	1.52(1.37–1.68)	1.07(0.85–1.36)
Per A allele			1.23(1.18–1.29)	1.05(0.94–1.17)
p for trend			1.06 × 10 ⁻¹⁸	0.3847
ER(-)^c				
GG	1295	4161	1.00(reference)	1.00(reference)

	East-Asians		European-ancestry Americans ^b	
	Cases	Controls	OR (95% CI)	OR (95% CI)
AG	1930	4384	1.36(1.25–1.48)	1.07(0.84–1.37)
AA	695	1203	1.77(1.58–1.98)	1.05(0.73–1.52)
Per A allele			1.34(1.27–1.41)	1.04(0.88–1.23)
p for trend			3.64×10^{-25}	0.6745
	P for association with ER status ^d : 0.0004		P for association with ER status: 0.9883	

^aAdjusted for age and study site.

^bIncludes NBHS-White, CBCS, and LIBCSP.

^cNo ER information was available in the CBCS, and thus this study was not included in the analysis.

^dDerived from the Chi-squared test to examine the association between ER status and rs2046210 genotypes in the case group only.

Table 4

Association of SNPs rs6913578 and rs7763637 with breast cancer risk

SNP	Study ^a	Allele ^b	Genotyped or imputed	Adjusted SNP	No of cases	No of controls	Frequency (%) ^c	OR (95% CI) ^d		P for trend
								Heterozygous	Homozygous	
rs6913578	Chinese	C/A	Imputed	None	2069	2080	39.6/34.4	1.26 (1.10–1.44)	1.54 (1.27–1.87)	1.44 × 10 ⁻⁶
	European-ancestry Americans		Imputed + genotyped	rs2046210	2064	2077	39.6/34.4	1.15 (0.85–1.55)	1.27 (0.72–2.26)	0.3852
	African Americans		Genotyped	rs2046210	2691	2571	33.7/31.4	1.06 (0.95–1.19)	1.31 (1.08–1.60)	0.0128
rs7763637	SBCS-GWAS	A/G	Genotyped	None	799	1,736	33.6/31.3	0.89 (0.68–1.17)	0.92 (0.57–1.47)	0.5349
	European-ancestry Americans		Imputed + genotyped	rs2046210	2653	2528	47.9/47.5	1.04 (0.84–1.28)	1.13 (0.88–1.45)	0.3363
	African Americans		Genotyped	rs2046210	795	1,727	48.1/47.5	1.16 (0.89–1.49)	1.38 (0.94–2.03)	0.0964
	SBCS-GWAS	A/G	Genotyped	None	1927	1936	39.5/33.9	1.28 (1.11–1.46)	1.62 (1.33–1.98)	3.49 × 10 ⁻⁷
	European-ancestry Americans		Imputed + genotyped	rs2046210	1924	1933	39.6/33.9	1.26 (0.92–1.72)	1.58 (0.87–2.86)	0.1294
	African Americans		Genotyped	rs2046210	2654	2528	34.8/32.5	1.08 (0.96–1.21)	1.29 (1.07–1.55)	0.0104
	SBCS-GWAS	A/G	Genotyped	None	790	1749	34.7/32.3	0.94 (0.71–1.24)	0.85 (0.53–1.36)	0.4950
	European-ancestry Americans		Imputed + genotyped	rs2046210	2654	2528	47.8/47.8	1.01 (0.82–1.25)	1.12 (0.87–1.43)	0.3974
	SBCS-GWAS	A/G	Genotyped	None	786	1741	48.0/47.8	1.15 (0.89–1.49)	1.42 (0.97–2.09)	0.0768
	European-ancestry Americans		Imputed + genotyped	rs2046210	2654	2528	48.0/47.8	1.15 (0.89–1.49)	1.42 (0.97–2.09)	0.0768

^a Chinese: SBCS-GWAS; European-ancestry Americans: NBHS-White and CGEMS; African Americans: SCCS and NBHS-black^b Risk allele/reference allele.^c Test allele frequency in cases/controls.^d Adjusted for age.