Genome-wide association study identifies a novel breast cancer susceptibility locus at 6q25.1

Wei Zheng[1], **Jirong Long**[1], **Yu-Tang Gao**[2], **Chun Li**[3,6], **Ying Zheng**[4], **Yong-Bin Xiang**[2], **Wanqing Wen**[1], **Shawn Levy**[5], **Sandra L. Deming**[1], **Jonathan L. Haines**[6], **Kai Gu**[4], **Alecia Malin Fair**[7], **Qiuyin Cai**[1], **Wei Lu**[4], and **Xiao-Ou Shu**[1]

[1]Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, Tennessee

[2]Department of Epidemiology, Shanghai Cancer Institute, Shanghai, China

[3]Department of Biostatistics, Vanderbilt University School of Medicine, Nashville, Tennessee

[4]Shanghai Center for Disease Prevention and Control, Shanghai Institute of Preventive Medicine, Shanghai, China

[5]Department of Biomedical Informatics, Vanderbilt University School of Medicine, Nashville, Tennessee

[6]Center for Human Genetics Research, Vanderbilt University School of Medicine, Nashville, Tennessee

[7]Meharry Medical College, School of Medicine, Nashville, Tennessee

Abstract

A genome-wide association study was conducted among Chinese women to identify risk variants for breast cancer. By analyzing 607,728 SNPs in 1505 cases and 1522 controls, we selected 29 promising SNPs for a fast-track replication in an independent set of 1554 cases and 1576 controls. Four replicated loci were further investigated in a third set of samples including 3472 cases and 900 controls. SNP rs2046210 at 6q25.1, located upstream of the estrogen receptor 1 gene (*ESR1*), exhibited strong and consistent association with breast cancer across all three stages. Adjusted odds ratio (95% CI) were 1.36 (1.24–1.49) and 1.59 (1.40–1.82), respectively, for genotypes A/G and A/A versus G/G (P for trend, 2.0×10^{-15}) in the pooled analysis of samples from all three stages. A similar, although weaker, association was also found in an independent study including

1591 cases and 1466 controls of European ancestry (P_{trend}, 0.01). These results provide strong evidence implicating 6q25.1 as a susceptibility locus for breast cancer.

Breast cancer, a complex multifactorial disease, is one of the most common malignancies among women in the world. Genetic factors play an important role in the pathogenesis of both sporadic and familial breast cancer[1–3]. However, only a small fraction of breast cancer cases can be explained by the breast cancer susceptibility genes identified thus far, such as the *BRCA1* and *BRCA2* genes[4–6]. Family-based linkage studies have been successful in mapping genes associated with Mendelian disorders[1–3, 7]. However, this approach has had limited success in identifying common genetic variants that confer small to moderate risk of disease susceptibility. Over the past 15 years, a large number of association studies have evaluated genetic variants in many candidate genes in relation to breast cancer risk [1–3, 7–9]. Although numerous genetic variants have been implicated, only a few of them have been replicated in subsequent studies[10, 11]. Four recent GWA studies have identified several novel risk alleles for breast cancer[12–15]. All these studies, however, are conducted in women of European decent who differ from women of other ethnic groups in certain genetic architecture. Therefore, additional GWA studies, particularly those conducted in non-European decent population, are needed to fully uncover genetic basis for breast cancer susceptibility.

Since 1996 we have initiated multiple population-based epidemiologic studies of cancer in Shanghai, China, including the Shanghai Breast Cancer Study (SBCS) (see Supplementary Methods). Included in Stages I to III of the current GWA study were genomic DNA samples from 6,531 incident breast cancer cases and 3,998 community controls who participated in these studies (Table 1). The pilot phase of the GWA study was initiated in 2005 with 150 cases and 150 controls genotyped using the Affymetrix GeneChip Human Mapping 500K Array Set that contains approximately 500,000 SNPs. An additional 1,374 cases and 1,402 controls were genotyped in 2008 using the Affymetrix Genome-Wide Human SNP Array 6.0 that contains 906,602 SNPs. Cases and controls were matched on age. Included in the current analysis were 607,728 SNPs in the Affymetrix SNP array 6.0 and 330,885 SNPs in the Affymetrix 500K array set that met the following criteria: 1) 5% minor allele frequency (MAF), 2) 95% call rate, 3) 95% genotyping concordance rate in quality control (QC) samples, and additionally for SNPs in the Affymetrix 500K arrays only those that are also present in the Affymetrix 6.0 array (Supplementary Table S1). Of the initial 3,076 samples included in the GWA scan, 49 samples were excluded, due to < 95% call rate (n = 4), or sample duplication or contamination (n = 45). A total of 1,505 cases and 1,522 controls remained for the GWA analyses. Multidimensional scaling analyses based on pairwise identity-by-state showed no evidence of apparent genetic admixture in this study population (Supplementary Figure S1).

Multiple genomic locations were revealed as potentially related to breast cancer risk (Figure 1), and the observed number of SNPs with a small P-value is larger than that expected by chance (Supplementary Figure S2). Similar results were obtained after excluding the 292 subjects genotyped by Affymetrix 500K Array Set from the analyses (data not shown). P-values presented in Figure 1 are derived from trend tests using logistic regression (df = 1)

after adjusting for age. Six of the 11 SNPs identified from published GWA studies[12–15] are included in the Affymetrix 6.0 array, and four of them showed an association with breast cancer risk consistent with that reported previously (Supplementary Table S2). Specifically, elevated risk of breast cancer was found to be associated with the minor allele of rs1219648 (*FGFR2*, P_{trend}=0.0025), rs2981582 (*FGFR2*, P_{trend}=0.001), rs3803662 (*TNRC9*, P_{trend}=0.012), and rs8051542 (*TNRC9*, P_{trend}=0.098). No apparent association, however, was found for rs3817198 (*LSP1*, P_{trend}=0.75), and the association with rs2180341 (6q22.33, P_{trend}=0.068) was in the opposite direction of the one reported initially in a study based on the Ashkenazi Jewish population[15].

For our fast track replication, 29 most promising SNPs were genotyped in an independent set of 1,554 cases and 1,576 controls recruited in the SBCS. These SNPs were selected from those that had 1) MAF 10%; 2) very clear genotyping clusters; 3) not yet confirmed previously as a genetic risk variant for breast cancer; and 4) P 1×10^{-4} for all samples along with a consistent association at P 0.05 in samples analyzed in the first batch (754 cases/741 controls) and the second batch (751 cases/781 controls) or P 5×10^{-4} for all subjects and consistent association at P 0.01 in both batches.

Of the 29 SNPs included in fast-track replication (Supplementary Table S3), four SNPs in stage II showed a significant association with breast cancer risk at P 0.05 and the fifth one had a P-value of 0.077 (Table 2). A highly significant association with breast cancer risk was identified for rs2046210 (P = 3.9×10^{-5}) and rs10872676 (P = 1.6×10^{-3}). Both SNPs are located at 6q25.1, approximately 4.4 kb apart, showing a high degree of LD (r^2 = 0.69). Therefore, rs2046210 was selected for further validation, as rs10872676 showed a weaker association with breast cancer risk than rs2046210, and its association was not statistically significant after adjusting for rs2046210.

Four SNPs were further evaluated in Stage III (Table 2), which included 3,472 cases who were recruited during 2002 and 2006 as part of the Shanghai Breast Cancer Survivor Study (SBCSS), along with 900 healthy women recruited from the same source population as the control group for a population-based endometrial cancer study that was conducted in parallel with the SBCS. Again, rs2046210 was associated with breast cancer risk (P = 3.3×10^{-7}) (Table 2), and the P-value reached 2.0×10^{-15} in the pooled analysis of samples from all three stages (Table 3). This p-value is substantially lower than the genome-wide significance level based on conservative Bonferroni adjustment of multiple comparisons at a level α = 0.05, providing unequivocal evidence for an association of this SNP with breast cancer risk. This SNP was associated with a population attributable risk (PAR) of 18.9% and an estimated 2.1% excess familial risk of breast cancer. The positive association of this SNP with breast cancer risk was found for both pre- and post-menopausal women, and the association was stronger for ER negative cancer than ER positive cancer (P=0.02) (Table 3). None of the other three SNPs, however, were replicated in Stage III (Table 2).

Figure 2 shows the 6q25.1 locus where rs2046210 is located. A cluster of SNPs that are in strong LD with rs2046210 all showed a significant association with breast cancer risk with P 0.001 in Stage I. Using data from Stage I, haplotype analyses of a haplotype block including rs2046210 and other 7 SNPs as defined by the method of Gabriel[16] or a larger

block including 7 additional SNPs failed to identify any particular SNP that may explain the observed association in this locus (Supplementary Table S4).

We also evaluated SNP rs2046210 in association with breast cancer risk among 1,590 cases and 1,466 controls of European ancestry, recruited as part of the Nashville Breast Health Study (NBHS), a population based case-control study conducted in Tennessee, USA (Table 1). Consistent with the findings from the Shanghai studies, the variant allele of this SNP was associated with an elevated risk of breast cancer, and the association was stronger in post- than pre-menopausal women (Table 4).

Several genes are located in the 1 Mb region centered on SNP rs2046210 including *PLEKHG1, MTHFD1L, AKAP12, ZBTB2, RMND1, C6orf211, C6orf97, ESR1, C6orf98, SYNE1*, and *NANOGP11*. Of them, the *ESR1* gene is perhaps of particular interest to breast carcinogenesis. The *ESR1* gene encodes estrogen receptor α (ERα) that regulates signal transduction of estrogen, a sex hormone that plays a central role in the etiology of breast cancer. Elevated estrogen levels have been shown to be associated with an increased risk of breast cancer in multiple prospective studies[17]. Since biological effects of estrogen are mediated primarily through high-affinity binding to ERs, genetic variants in ER genes, including *ESR1* and *ESR2* have been the focus of multiple previous epidemiologic studies[18–21]. The identified SNP (rs2046210) associated with breast cancer risk is located 29 kb upstream of the first untranslated exon and 180 kb upstream of the transcription start site of exon 1 of the *ESR1* gene[22]. None of the SNPs at this locus has been previously reported to be associated with breast cancer, nor in LD with two of the most widely studied polymorphisms in *ESR1*: rs2234693 and rs9340799 ($r^2<0.05$ in both HapMap Asian and women of European decent samples). SNP rs2234693 was genotyped in Stage I of the study and carried an OR (95% CI) of 0.95 (0.80–1.12) for C/T and 0.79 (0.63–1.00) for T/T genotype in relation to breast cancer risk. Because of the relatively close location to the *ESR1* gene and the biological function of ERα, it is possible that rs2046210 or SNPs in LD with it may alter *ESR1* gene expression and affect susceptibility to breast cancer. It is noteworthy that a recent GWA study has found that the 6q25.1 locus is associated with bone mineral density[23], a phenotype that is affected by estrogen.

SNP rs2046210 is located 6 kb downstream of *C6orf97*, the chromosome 6 open reading frame 97. The function of *C6orf97* is unknown. The LD block that includes rs2046210 spans a region of about 41 kb (151,971,942 to 152,013,380) which contains part of *C6orf97*. By running BLAST with C6orf97 coding peptide as the query sequence, a structural maintenance of chromosomes (SMC) domain was found in the C-terminal of the *C6orf97* protein. SMC proteins appear to play an important role in chromosome dynamics[24]. Further research to the functionality of *C6orf97* and its potential association with breast cancer may be warranted.

METHODS

Study population

For detailed descriptions of the component studies see Supplementary Materials. The study protocol was approved by the Institutional Review Boards of the Vanderbilt University

Medical Center, the Shanghai Cancer Institute, the Shanghai Center for Disease Prevention and Control, and Meharry Medical College. Informed consent was obtained from all participants.

Genotyping and quality control procedures

For detailed descriptions of the genotyping and quality control procedures see Supplementary Materials. Briefly, in Stage I, the initial 300 subjects were genotyped using the Affymetrix GeneChip Mapping 500K Array Set and the remaining 2,776 samples were genotyped using the Affymetrix Genome-Wide Human SNP Array 6.0. In each of the 96-well plates genotyped using the Affymetrix SNP 6.0 array, three positive QC samples (NA15510, NA10851, and NA18505) purchased from Coriell Cell Repositories (http://ccr.coreill.org/), and a negative QC sample (water) were included. SNP data obtained from the positive QC samples showed a very high concordance rate of called genotypes based on 79,764,872 comparisons (mean, 99.87%; median, 100%). In addition, 742 SNPs were genotyped using the Affymetrix Target Genotyping System, TaqMan or Sequenom for a subset of subjects included in previous studies. A high concordance rate was also observed between genotypes determined using these platforms and Affymetrix SNP Array 6.0 based on 1,478,383 comparisons (mean, 99.1%; median, 99.8%). Samples with genotyping call rates less than 95% were excluded, and the remaining samples were recalled by using Birdseed v2. The gender of all study samples was confirmed to be female. The identity-by-descent analysis based on identity-by-state was performed to detect first degree cryptic relationships using the PLINK 1.04 25 (http://pngu.mgh.harvard.edu/purcell/plink/). A final data set included 3,027 individuals and 607,728 markers in the Affymetrix SNP Array 6.0 and 330,885 in the Affymetrix 500K Array Set that met the following criteria: (1) genotype call rate 95%; (2) MAF 5%; and (3) genotyping concordance rate in QC samples 95%. Also excluded from the analyses were 21,223 SNPs that are on Affymetrix 500K Array Set but not on the SNP Array 6.0.

Genotyping for the replication sets was completed using the iPLEX[tm] Sequenom MassArray® platform. Included in each 96-well plate as QC samples were two negative controls (water), two blinded duplicates, and two samples included in the HapMap project. The mean concordance rate was 99.7% for blind duplicates (2,572 comparisons) and 99.2% for HapMap samples (1,751 comparisons).

Statistical analyses

The PLINK version 1.04 was used to analyze genome-wide data obtained in Stage I. A set of 4,305 SNPs with MAF 0.35 and 100 kb between two adjacent SNPs was selected to evaluate the population structure. The inflation factor λ was estimated to be 1.024, suggesting that population substructure, if present, should not have any appreciable effect on the results. Odds ratios (OR) and 95% confidence intervals (CI) were estimated from logistic regression analysis. Age was adjusted in the GWA analyses of Stage I data to select promising SNPs. Additional adjustment was made for education in the analyses of Stages II and III data. Finally, multivariate analyses including age, education, study stage, body mass index, age at menarche, and age at first live birth were performed in the pooled analyses of

data from all stages. None of the 29 markers genotyped in Stage II showed deviations from Hardy-Weinberg equilibrium ($P > 0.05$). P-values based on 2-tailed tests are presented.

Haplotype analyses methods: SNPs genotyped in Stage I in a region of ~100kb containing rs2046210 were used to reconstruct haplotypes. LD between SNPs was assessed by Haploview26. Haplotype blocks were defined using the methods of Gabriel et al16 and 9 blocks were observed (Supplementary Figure S3). Block 5 includes rs2046210 and seven other SNPs, rs7740686, rs7763637, rs9397436, rs9397437, rs6908732, rs852004, and rs865898 (Supplementary Table S4). Associations of haplotypes with breast cancer risk were analyzed with HAPSTAT software27 based on additive models. Additional analyses were performed for haplotypes defined using SNPs included in blocks 4 and 5 with a total of 15 SNPs (Supplementary Figure S3, Supplementary Table S4).

Imputation analysis: We used PLINK to impute genotypes for autosomal SNPs that were present in HapMap Phase II release 23a but not genotyped in our GWA scan. HapMap genotype data from the 90 Asian HapMap subjects that have MAF 1% and genotyping rate 95% were used as reference. During the imputation, an information score was generated for each imputed SNP that reflected how confidently genotypes were inferred. Values below 0.80 were taken to be indicative of poor quality and were not analyzed. Besides the 0.61 million observed SNPs, an additional 1.14 million imputed SNPs were tested for association with breast cancer.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgement

Reference List

1. Nathanson KL, Wooster R, Weber BL. Breast cancer genetics: what we know and what we need. Nat. Med. 2001; 7:552–556. [PubMed: 11329055]

2. Balmain A, Gray J, Ponder B. The genetics and genomics of cancer. Nat. Genet. 2003; 33(Suppl): 238–244. [PubMed: 12610533]

3. Houlston RS, Peto J. The search for low-penetrance cancer susceptibility alleles. Oncogene. 2004; 23:6471–6476. [PubMed: 15322517]

4. Walsh T, et al. Spectrum of mutations in BRCA1, BRCA2, CHEK2, and TP53 in families at high risk of breast cancer. JAMA. 2006; 295:1379–1388. [PubMed: 16551709]

5. Miki Y, et al. A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. Science. 1994; 266:66–71. [PubMed: 7545954]

6. Wooster R, et al. Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12–13. Science. 1994; 265:2088–2090. [PubMed: 8091231]

7. Botstein D, Risch N. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. Nat. Genet. 2003; 33(Suppl):228–237. [PubMed: 12610532]

8. Pharoah PD, Tyrer J, Dunning AM, Easton DF, Ponder BA. Association between common variation in 120 candidate genes and breast cancer risk. PLoS. Genet. 2007; 3:e42. [PubMed: 17367212]

9. Pharoah PD, Dunning AM, Ponder BA, Easton DF. Association studies for finding cancer-susceptibility genetic variants. Nat. Rev. Cancer. 2004; 4:850–860. [PubMed: 15516958]

10. Cox A, et al. A common coding variant in CASP8 is associated with breast cancer risk. Nat. Genet. 2007; 39:352–358. [PubMed: 17293864]

11. Frank B, et al. Association of a common AKAP9 variant with breast cancer risk: a collaborative analysis. J Natl. Cancer Inst. 2008; 100:437–442. [PubMed: 18334708]

12. Easton DF, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. Nature. 2007; 447:1087–1093. [PubMed: 17529967]

13. Hunter DJ, et al. A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. Nat. Genet. 2007; 39:870–874. [PubMed: 17529973]

14. Stacey SN, et al. Common variants on chromosome 5p12 confer susceptibility to estrogen receptor-positive breast cancer. Nat. Genet. 2008; 40:703–706. [PubMed: 18438407]

15. Gold B, et al. Genome-wide association study provides evidence for a breast cancer risk locus at 6q22.33. Proc. Natl. Acad. Sci. U. S. A. 2008; 105:4340–4345. [PubMed: 18326623]

16. Gabriel SB, et al. The structure of haplotype blocks in the human genome. Science. 2002; 296:2225–2229. [PubMed: 12029063]

17. Key T, Appleby P, Barnes I, Reeves G. Endogenous sex hormones and breast cancer in postmenopausal women: reanalysis of nine prospective studies. J Natl. Cancer Inst. 2002; 94:606–616. [PubMed: 11959894]

18. Mitrunen K, Hirvonen A. Molecular epidemiology of sporadic breast cancer. The role of polymorphic genes involved in oestrogen biosynthesis and metabolism. Mutat. Res. 2003; 544:9–41. [PubMed: 12888106]

19. Cai Q, et al. Association of breast cancer risk with a GT dinucleotide repeat polymorphism upstream of the estrogen receptor-alpha gene. Cancer Res. 2003; 63:5727–5730. [PubMed: 14522892]

20. Cai Q, et al. Genetic polymorphisms in the estrogen receptor alpha gene and risk of breast cancer: results from the Shanghai Breast Cancer Study. Cancer Epidemiol Biomarkers Prev. 2003; 12:853–859. [PubMed: 14504194]

21. Zheng SL, et al. Joint effect of estrogen receptor beta sequence variants and endogenous estrogen exposure on breast cancer risk in Chinese women. Cancer Res. 2003; 63:7624–7629. [PubMed: 14633679]

22. Kos M, Reid G, Denger S, Gannon F. Minireview: genomic organization of the human ERalpha gene promoter region. Mol. Endocrinol. 2001; 15:2057–2063. [PubMed: 11731608]

23. Styrkarsdottir U, et al. Multiple genetic loci for bone mineral density and fractures. N. Engl. J Med. 2008; 358:2355–2365. [PubMed: 18445777]

24. Hirano T. At the heart of the chromosome: SMC proteins in action. Nat. Rev. Mol. Cell Biol. 2006; 7:311–322. [PubMed: 16633335]

25. Purcell S, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum. Genet. 2007; 81:559–575. [PubMed: 17701901]

26. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics. 2005; 21:263–265. [PubMed: 15297300]

27. Lin DY, Zeng D, Millikan R. Maximum likelihood estimation of haplotype effects and haplotype-environment interactions in association studies. Genet. Epidemiol. 2005; 29:299–312. [PubMed: 16240443]
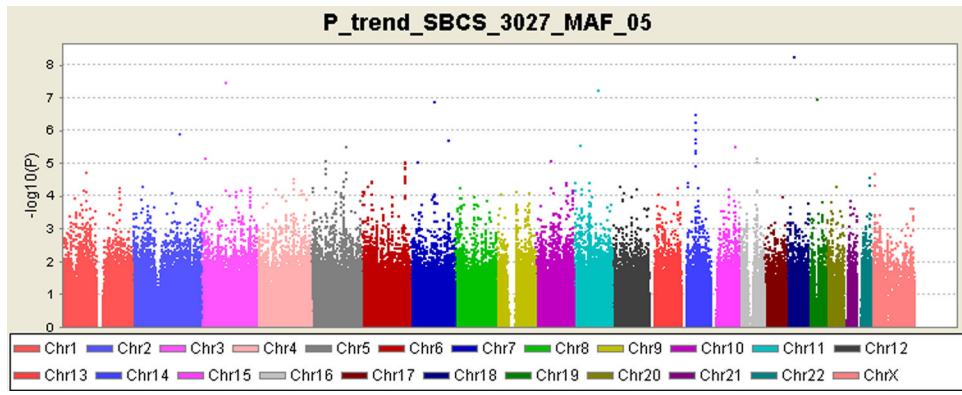
Figure 1.
Genome-wide association results in the Shanghai Breast Cancer Study, Scatter plot of P-values in log-scale from the trend test for 607,728 genotyped SNPs comparing 1,505 cases and 1,522 controls
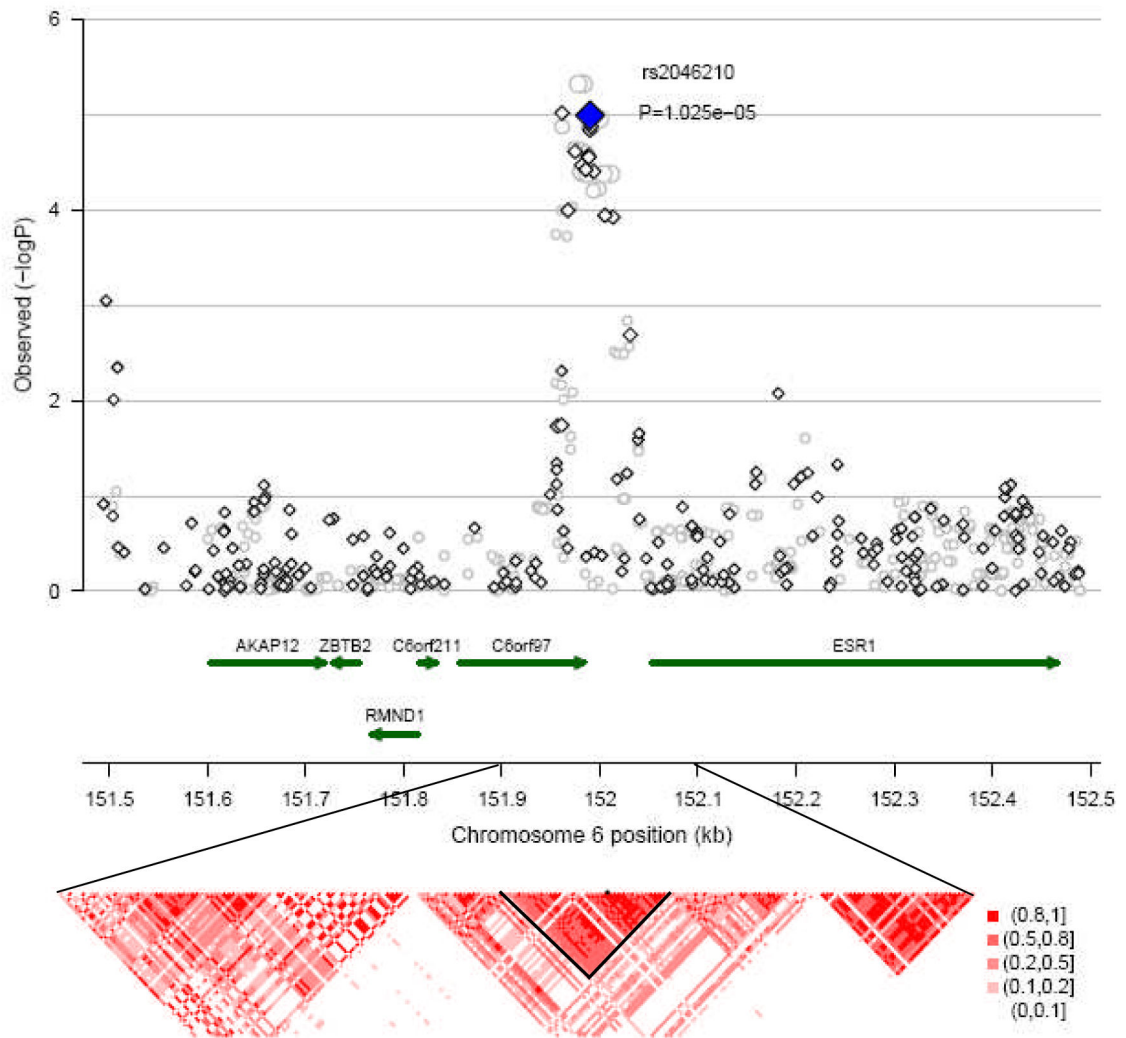
Figure 2.
Regional plot of chromosome 6q25.1 locus. Results (-logP) are shown for directly
genotyped (diamonds) and imputed (circles) SNPs for the region of 151.5–152.5 Mb,
flanking 500kb of SNP rs2046210. SNP rs2046210 is shown in blue. Gene locations are
from the March 2006 UCSC genome browser assembly. The lower panel shows estimates of
pairwise r^2 for common SNPs (with MAF   5%) from the HapMap release 23a for the Asian
population for the region of 151.9–152.1 Mb, flanking 100 kb of SNP rs2046210.

Table 1

Distribution of demographic characteristics and known breast cancer risk factors for cases and controls included in the study

	Stage I		Stage II		Stage III		NBHS	
	Cases (n=1,505)	Controls (n=1,522)	Cases (n=1,554)	Controls (n=1,576)	Cases (n=3,472)	Controls (n=900)	Cases (n=1,591)	Controls (n=1,466)
Demographic Factors[a]								
Age (yr)	46.9 ± 7.2	47.0 ± 7.5[b]	52.4 ± 8.5	52.8 ± 9.0	54.9 ± 10.5	54.9 ± 8.6	54.8 ± 10.2	52.2 ± 11.0[b]
Education ≥ high school (%)	47.5	47.8	56.9	43.5[b]	52.2	39.9[b]	94.8	96.5
Reproductive Risk Factors[a]								
Age at menarche (yr)	14.4 ± 1.6	14.7 ± 1.7[b]	14.4 ± 1.7	14.7 ± 1.8[b]	14.4 ± 1.7	14.8 ± 1.8[b]	12.6 ± 2.6	12.8 ± 4.2
Postmenopausal (%)	26.5	30.4[b]	52.2	57.7[b]	53.6	61.8[b]	65.3	59.0[b]
Age at menopause[c]	47.9 ± 4.7	47.1 ± 5.0[b]	48.7 ± 4.3	48.5 ± 4.5	49.1 ± 4.3	49.1 ± 3.6	46.2 ± 8.5	45.3 ± 8.7[b]
Number of live births	1.4 ± 0.7	1.4 ± 0.8	1.4 ± 0.7	1.5 ± 0.9[b]	1.6 ± 1.0	1.8 ± 1.1[b]	2.7 ± 0.7	2.7 ± 0.7
Age at first live birth (yr)[d]	26.8 ± 3.8	26.3 ± 3.6[b]	26.1 ± 3.7	25.5 ± 4.0[b]	26.9 ± 4.0	25.3 ± 4.1[b]	24.3 ± 5.3	24.6 ± 5.2
Used HRT (%)[e]	2.2	2.2	4.1	2.9	3.8	4.6	51.1	56.5[b]
Other Risk factors[a]								
Had breast cancer family history (%)	5.4	3.2[b]	4.0	2.5[b]	5.7	2.3[b]	19.7	14.1[b]
Prior BBD diagnosis (%)[f]	10.3	5.5[b]	9.4	5.3[b]	N/A	N/A	49.3	33.4[b]
Body mass index	23.6 ± 3.4	23.2 ± 3.3[b]	23.8 ± 3.2	23.5 ± 3.2[b]	24.2 ± 3.5	23.8 ± 3.5[b]	27.2 ± 7.1	26.7 ± 6.9
Waist-to-hip ratio	0.81 ± 0.05	0.80 ± 0.06[b]	0.83 ± 0.05	0.82 ± 0.06[b]	0.84 ± 0.05	0.82 ± 0.06[b]	NA	NA

[a] Unless otherwise specified, mean (sd) are presented.

[b] P ≤ 0.05 for case-control comparisons.

[c] Among postmenopausal women.

[d] Among parous women.

[e] HRT: Hormone Replacement Therapy

[f] BBD defined as breast fibroadenoma in Stages I and II and breast cysts or lumps in the NBHS.

Table 2

Summary results for five SNPs showing a promising association with breast cancer risk in Stage II

SNP (allele)[a]	Position[b] (chromosome)	Stage	MAF[c] (%)	OR$_{het}$ (95% CI)[d]	OR$_{homo}$ (95% CI)[d]	P for trend[d]	P for homogeneity
rs6767967 (A/G)	24683878 (3p24.2)	I	25.2	1.34 (1.16–1.56)	1.31 (0.99–1.73)	3.6×10^{-4}	
		II	26.4	1.26 (1.09–1.47)	1.11 (0.83–1.48)	0.020	0.018
		III	26.4	0.96 (0.82–1.12)	1.09 (0.80–1.49)	0.984	
rs10872676[e] (T/C)	151985670 (6q25.1)	I	42.1	1.25 (1.05–1.49)	1.55 (1.25–1.93)	5.7×10^{-5}	
		II	43.1	1.17 (0.99–1.39)	1.40 (1.13–1.72)	1.6×10^{-3}	0.745
rs2046210 (A/G)	151990059 (6q25.1)	I	36.5	1.30 (1.11–1.52)	1.56 (1.26–1.94)	1.4×10^{-5}	
		II	37.1	1.22 (1.04–1.42)	1.57 (1.26–1.96)	3.9×10^{-5}	0.131
		III	34.9	1.59 (1.35–1.87)	1.56 (1.25–1.96)	3.3×10^{-7}	
rs10781090 (C/T)	74029509 (9q21.13)	I	42.7	1.36 (1.15–1.61)	1.42 (1.15–1.75)	3.4×10^{-4}	
		II	43.7	1.13 (0.95–1.33)	1.20 (0.97–1.47)	0.077	0.033
		III	44.0	0.91 (0.76–1.07)	1.06 (0.85–1.32)	0.793	
rs11609565 (A/G)	130658977 (12q24.33)	I	34.2	1.20 (1.02–1.41)	1.54 (1.21–1.96)	2.8×10^{-4}	
		II	34.5	1.14 (0.98–1.33)	1.22 (0.97–1.54)	0.047	0.022
		III	37.0	0.90 (0.77–1.06)	0.93 (0.74–1.18)	0.334	

[a]Minor/major alleles.

[b]From NCBI genome build 36.

[c]Minor allele frequency in controls.

[d]Adjusted for age and education.

[e]SNP rs10872676 was not selected for Stage III validation because of its high degree of LD with rs2046210, and its association with breast cancer risk was not statistically significant (P for trend, 0.658) after adjusting for rs2046210 in a pooled analysis of Stages I and II data.

Table 3

Association of rs2046210 with breast cancer risk among Chinese women in the pooled analysis of cases and controls included in all three stages of the Shanghai Studies.

Genotypes	No of cases	No of controls	OR (95% CI)[a]	P for trend
All women (6,472 cases, 3,962 controls)				
GG	2162	1614	1.00 (reference)	
AG	3208	1812	1.36 (1.24–1.49)	
AA	1102	536	1.59 (1.40–1.81)	
Per A allele			1.29 (1.21–1.37)	2.0×10^{-15}
Premenopausal women (3,430 cases, 2,051 controls)				
GG	1135	815	1.00 (reference)	
AG	1714	953	1.30 (1.14–1.47)	
AA	581	283	1.54 (1.29–1.84)	
Per A allele			1.26 (1.15–1.37)	1.6×10^{-7}
Postmenopausal women (3,038 cases, 1,906 controls)				
GG	1025	798	1.00 (reference)	
AG	1492	855	1.44 (1.26–1.65)	
AA	521	253	1.67 (1.38–2.01)	
Per A allele			1.33 (1.21–1.45)	1.1×10^{-9}
All women (3,654 ER (+) cases, 3,962 controls)[b]				
GG	1263	1614	1.00 (reference)	
AG	1795	1812	1.33 (1.19–1.48)	
AA	596	536	1.52 (1.31–1.35)	
Per A allele			1.26 (1.17–1.35)	7.1×10^{-10}
All women (2,038 ER (−) cases, 3,962 controls)[b]				
GG	648	1614	1.00 (reference)	
AG	1022	1812	1.45 (1.28–1.65)	
AA	368	536	1.75 (1.47–2.08)	
Per A allele			1.35 (1.24–1.46)	4.2×10^{-12}

[a] Adjusted for age, education, study stage, body mass index, age at menarche, and age at first birth.

[b] Compared to cases with ER (+) cancer, adjusted ORs (95% CI) for ER (−) cancer were 1.00 (reference), 1.10 (0.998–1.25), and 1.20 (1.02–1.42), respectively for genotypes GG, AG, and AA (p for trend, 0.02).

Table 4

Association of rs2046210 with breast cancer risk among women of European ancestry in the Nashville Breast Health Study.

Genotypes	No of cases	No of controls	ORs (95% CI)	P for trend
All Women (1,591 cases, 1,466 controls)				
G/G	614	617	1.00 (reference)	
G/A	761	690	1.11 (0.95–1.29)	
A/A	216	159	1.35 (1.06–1.71)	
Per A allele			1.15 (1.03–1.28)	0.01
Pre/peri-menopausal (626 cases, 683 controls)				
G/G	246	290	1.00 (reference)	
G/A	301	314	1.12 (0.89–1.43)	
A/A	79	79	1.13 (0.79–1.63)	
Per A allele			1.08 (0.92–1.23)	0.36
Post-menopausal women (965 cases, 783 controls)				
G/G	368	327	1.00 (reference)	
G/A	460	376	1.09 (0.89–1.34)	
A/A	137	80	1.55 (1.13–2.12)	
Per A allele			1.19 (1.03–1.38)	0.02

[a] Adjusted for age, education, age at menarche, age at first live birth, and body mass index