Evaluation of potential regulatory function of breast cancer risk locus at 6q25.1

Yaqiong Sun¹², Chuanzhong Ye¹, Xingyi Guo¹, Wanqing Wen¹, Jirong Long¹, Yu-Tang Gao³, Xiao Ou Shu¹, Wei Zheng³ and Qiuyin Cai¹,*

¹Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, TN, 37203, USA, ²Shanghai Municipal Center for Disease Control and Prevention, Shanghai, China and ³Department of Epidemiology, Shanghai Cancer Institute, Shanghai, China

*To whom correspondence should be addressed. Tel: +1 615 9361351; Fax: +1 615 9368291; Email: qiuyin.cai@vanderbilt.edu

Abstract

In a genome-wide association study conducted among Chinese women, we identified the single nucleotide polymorphism (SNP) rs2046210 at 6q25.1 for breast cancer risk. To explore a potential regulatory role for this risk locus, we measured expression levels of nine genes at the locus in breast cancer tissue and adjacent normal tissue samples obtained from 67 patients recruited in the Shanghai Breast Cancer Study. We found that rs2046210 had a statistically significant association with the expression levels of the AKAP12 and ESR1 genes in adjacent normal breast tissues. Women who carry the AA/AG risk genotypes had higher expressions of these two genes compared to those who carry G/G genotypes (P = 0.02 and 0.04 for the AKAP12 and ESR1, respectively). However, no significant differences of SNP rs2046210 with gene expression levels were found in tumor tissues. In The Cancer Genome Atlas samples, the AA/AG risk genotypes of SNP rs2046210 were associated with a significantly higher expression level of the AKAP12 gene and a lower level of the ESR1 gene in tumor tissue. Functional analysis using ENCODE data revealed that SNP rs7763637, which is in strong linkage disequilibrium with SNP rs2046210, is likely a potential functional variant, regulating the AKAP12 gene. Taken together, these results from our study suggest that the association between the 6q25.1 locus and breast cancer risk may be mediated through SNPs that regulate expressions of the AKAP12 gene.

Introduction

Genome-wide association studies have been instrumental in identifying common genetic variants associated with breast cancer risk. To date, nearly 100 novel genetic loci for breast cancer have been identified (1,2). We have previously identified several novel loci in Asian women (3–10), including a novel genetic susceptibility variant (rs2046210) at 6q25.1 (5). A nearly 60% elevated risk for breast cancer was found among women homozygous for the AA allele of rs2046210 (5). A similar association was found in other East Asian women including Japanese (10,11) and Koreans (12,13), while a weaker association was also found in women of European ancestry (5,10,14), and none for women of African ancestry (10,15,16). Single nucleotide polymorphism (SNP) rs2046210 is located in a non-coding region ~29 kb upstream of the ESR1 gene and 6 kb downstream of the CCDC170 (C6orf97) gene (5,10). Several previous studies have been conducted to identify potential causal and/or functional variants for this locus (10,17,18). However, the results are inconsistent. SNP rs2046210 and/or SNP(s) in strong linkage disequilibrium (LD) with SNP rs2046210 may affect breast cancer risk by regulating expressions of local genes in the locus. In this study, we investigated the expression levels of nine genes flanking a 1Mb region around SNP rs2046210 in breast cancer tissue and adjacent normal tissues from 67 patients, as well as data from The Cancer Genome Atlas (TCGA). We performed analyses to evaluate the association between expression of these genes and SNP rs2046210 and SNPs in strong LD with SNP rs2046210. Additionally, we performed functional annotation on local SNPs in strong LD
Material and methods

Study subjects and tissues samples
The subjects in this study were a subset of patients of the Shanghai Breast Cancer Study (SBCS), a population-based case control study. Details of the study have been described elsewhere (19). The subjects in the current study were 67 randomly selected breast cancer patients who provided both tumor and adjacent non-tumor tissue samples (Table 1). All patients provided written, informed consent to participate in the study, and the study protocols were approved by the Institutional Review Boards of all institutions involved in the study. Tumor tissue samples were removed during surgery from the center of the lesion, whereas adjacent non-tumor tissue samples were obtained from the distal edge of the resection. These samples were snap-frozen in liquid nitrogen as soon as possible, typically within 10 minutes. Samples were stored at −80°C until the relevant assays were done. All patients were interviewed at the time of recruitment. Two senior pathologists reviewed all tissue slides to confirm the diagnosis.

Laboratory assays
Total RNA was extracted from tissue specimens by homogenization in TRIzol solution (Invitrogen, Carlsbad, CA), phase separation, precipitation and washing following the manufacturer’s instructions. The quantity and quality of RNA was measured by spectrophotometric analysis. RNA was reverse-transcribed (RT) using a High Capacity cDNA Archive Kit (Applied Biosystems, Foster City, CA) in a final volume of 15 μl containing 0.15 μg RNA and 1× RT polymerase chain reaction (RT-PCR) buffer, 5.5 mM MgCl₂, 500 μM each dNTP, 2.5 μM Random Hexamers, 0.40 U/μl RNase inhibitor and 3.125 U/μl MultiScribe reverse transcriptase. The mixture was incubated at 25°C for 10 min, 37°C for 120 min and 95°C for 5 min.

We measured the mRNA expression levels of nine genes (PLEKHG1, MTHFD1L, AKAP12, ZBTB2, RMDN1, ARMT1, CCDC170 (C6orf97), ESR1 and SYNE1) located within the 1 Mb region centered on SNP rs2046210. Real-time quantitative PCR reaction was carried out in an iCycler detection system (Bio-Rad, Hercules, CA) in a 25 μl volume. The reaction mixture consisted of 12.5 μl iQSYBR Green Supermix, 200 nM of each primer and 1 μl of cDNA template. Reactions were performed for 45 cycles (95°C for 15 s, 60°C for 30 s and 72°C for 30 s) after initial 3-min incubation at 95°C. Primers for the different genes amplified are presented in Supplementary Table 1, available at Carcinogenesis Online. All reactions were done in duplicates. The threshold cycle (Ct) was obtained directly by subtracting Ct for the technical markers H3K4Me1, H3K4Me3 and H3K27Ac in all cell lines analyzed by ENCODE, including breast normal cell line, HMEC (HMEC-1; H1-hESC, K562, HepG2, HSMM, HUVEC, NHEK and NHLE. We also evaluated DNase I hypersensitive and transcription factor binding sites (TFBs) in all cell lines analyzed by ENCODE, including breast normal cell line, HMEC and breast cancer cell lines, T-47D and MCF-7. We assessed the histone modification markers H3K4Me1, H3K4Me3 and H3K27Ac in all cell lines analyzed by ENCODE through the layered histone tracks from the UCSC Genome Browser. Two publicly-available tools, RegulomeDB (20) (see URLs) and HaploReg V2 (21) (see URLs), were also used to evaluate candidate functional variants.

Results

Differential gene expression analysis
Geometric means and 95% confidence intervals (CI) for the expression levels of the nine genes in 67 SBCS breast cancer tumor and normal tissues are shown in Table 2. The RMDN1 gene exhibited the highest expression level among nine studied genes, whereas the AKAP12, PLEKHG1 and ZBTB2 genes exhibited the lowest expression levels. SNP rs2046210 was significantly associated with expression levels of the AKAP12 and ESR1 genes in the adjacent normal tissues. Women who carry the rs2046210 AA/AG risk genotypes had substantially higher AKAP12 gene expression (3.35 vs. 1.36, P = 0.02) than those who carry the rs2046210 GG genotypes (Table 2; Figure 1A). This was also true for the expression level of the ESR1 gene (13.90 vs. 5.26, P = 0.04) (Table 2). The associations were seen in both estrogen receptor (ER)-positive and ER-negative samples (data not available).
shown), although the associations were not statistically significant due to a smaller sample size. SNP rs2046210 was associated with MTHFD1L gene expression (4.37 vs. 2.11) with borderline significance \((P = 0.08)\). No association between SNP rs2046210 and the other six genes was observed in the adjacent normal tissue. In the tumor tissues, however, expression levels of these nine genes did not differ significantly by rs2046210 genotype status (Table 2).

eQTL analysis using TCGA data

We also conducted an eQTL analysis to evaluate the association of SNP rs2046210 with expression of the AKAP12 and ESR1 genes in tumor tissue using data from 1,006 samples included in TCGA. SNP rs2046210 was found to be correlated with AKAP12 gene expression. The women who carry the AA/AG risk genotypes had a higher AKAP12 gene expression level than women who carry the GG genotypes (Figure 1B) \((P < 0.05)\). The result was consistent with the observation in the adjacent normal tissue from the association analysis on 67 SBCS patients. We observed that SNP rs2046210 was significantly correlated with ESR1 gene expression level. Women who carry the AA/AG risk genotypes had a decreased expression level of ESR1 gene compared to women who carry the GG genotypes \((P < 0.01)\). We also found a decreased expression in the SBCS tumor tissue samples among women who carry the AA/AG risk genotypes compared to women who carry the GG genotype, although the difference was not statistically significant (Table 2). However, this result was inconsistent with the observation in the adjacent normal tissue from the association analysis of the 67 SBCS patients. Nevertheless, our results suggest that the association between SNP rs2046210 and breast cancer risk may be mediated through its role in the regulating expressions of the AKAP12 and ESR1 genes.

Functional annotation

We further evaluated and annotated the functional significance of a total of 20 SNPs in strong LD with rs2046210 in an Asian population \((r^2 > 0.8)\) using ChiP-Seq and DNase-Seq data from the ENCODE Project. In line with previous findings, no epigenetic signal was observed for rs2046210, indicating it may be a non-functional SNP (Supplementary Table 2, available at Carcinogenesis Online).

Interestingly, SNP rs7763637, located 946bp away and in strong LD with SNP rs2046210 \((r^2 = 0.92)\), was found to be potentially functional. The epigenetic signals of active promoter associated histone mark (H3K4Me3), enhancer associated histone marks (H3K4Me1 and H3K27Ac) and chromatin accessibility were enriched in the intervals containing SNP rs7763637 in multiple ENCODE cell lines. Functional annotation using chromatin states from nine ENCODE cell lines showed that this SNP is located in the active promoter region of the CCDC170 gene in two cell lines, including normal breast cell line HMEC. Using transcription factor (TF) ChiP-Seq data revealed that the variant is located in multiple TF binding sites, including several breast cancer-related TFs such as ZNF217, FOS, KAP1, JUND, FOSL2, JUN and MYC (Figure 2). In particular, TF FOSL2 and JUN were observed to harbor the variant in breast cancer cell line MCF-7 (Figure 2). Consistent with this observation, functional annotation using two publicly-available tools, RegulomeDB and HaploReg v2, revealed that this SNP exhibited potential functional significance (Supplementary Table 2, available at Carcinogenesis Online).

SNP rs7763637 was associated with breast cancer risk among 2867 breast cancer cases and 2,285 controls \((P = 1.90 \times 10^{-7})\), similar to that of rs2046210 \((P = 2.54 \times 10^{-7})\) in our Genome-wide association study samples. SNP rs7763637 was significantly associated with expression levels of the AKAP12 gene and borderline significant with the ESR1 gene in the adjacent normal tissues of SBCS patients (Supplementary Table 3, available at Carcinogenesis Online). Women who carry the rs7763637 AA/AG risk genotypes had substantially higher AKAP12 gene expression \((3.62 vs. 1.26, P = 0.046)\) than those who carry the rs7763637 GG genotypes. This was also true for the expression level of the ESR1 gene \((13.49 vs. 4.72, P = 0.06)\). No association between SNP rs7763637 and the other 7 genes was observed.
in the adjacent normal tissue (Supplementary Table 3, available at Carcinogenesis Online). Using the TCGA data, we found that SNP rs7763637 was associated with AKAP12 gene expression level. Women who carry the AA/AG genotypes had a higher AKAP12 gene expression level than women carrying the GG genotypes ($P = 0.02$). In addition, women who carry the AA/AG genotypes had lower expression levels of the ESR1 ($P = 0.01$), RMND1 ($P = 0.003$) and ZBTB2 ($P = 0.03$) genes than women carrying the GG genotypes in the TCGA tumor tissue samples. Taken together, these findings indicate that SNP rs7763637 is highly likely to be a functional variant at the 6q25.1 locus, and thus may affect the expression of target genes.

**Discussion**

SNP rs2046210, located upstream of the ESR1 gene, is associated with breast cancer risk among women of both European
and East Asian ancestry (5,10). In this study, we investigated the mRNA expression levels of genes in the 1Mb region of SNP rs2046210 in breast tumor tissues and their adjacent normal tissues. We found that women who carry rs2046210 AA/AG risk genotypes had higher AKAP12 and ESR1 expression levels compared with women carrying G/G genotypes in the adjacent normal tissue. eQTL analyses using data from TCGA also showed that rs2046210 AA/AG risk genotypes were associated with higher level of AKAP12 gene expression, but associated with lower level of ESR1 gene expression in tumor tissue. Functional analysis revealed that SNP rs7763637, which is in strong LD with SNP rs2046210, is a potential functional variant at the 6q25.1 locus and may be associated with breast cancer risk mediated through the regulation of the AKAP12 gene.

Protein kinase A anchor protein 12 gene (AKAP12) lies -26.8 kb upstream of SNP rs2046210. It is a scaffold protein that plays important roles in cell proliferation, migration, apoptosis and angiogenesis (22). Studies have shown that AKAP12 functions as a metastasis suppressor (23). Downregulation or loss of AKAP12 was associated with malignancy and metastasis in many cancer types including breast cancer (22,24–28). In contrast, other studies have found that the expression of AKAP12 was increased in cancer tissue or cell lines (22,29,30), indicating multiple functions of the AKAP12 gene in carcinogenesis. However, no evidence to date has shown that AKAP12 affects primary tumor growth (29). Our study found that SNP rs2046210 may increase AKAP12 expression in both normal tissues and tumor tissues. This suggests that genetic variations in this locus may play a role in multiple stages of breast cancer development, including initiation, progression, and metastasis.

The ESR1 gene encodes estrogen receptor α (ER-α), which regulates signal transduction of estrogen, a sex hormone that plays a central role in the etiology of breast cancer. SNP rs2046210 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 (31). The associations of SNP rs2046210 with ESR1 gene expression differ between tumor tissues and normal tissues. Women in the SBCS who carry the rs2046210 AA/AG risk genotypes had a higher ESR1 gene expression level in adjacent normal tissues. However, women from TCGA who carry the AA/AG risk genotypes had a decreased expression level of the ESR1 gene in tumor tissues. We also found that ESR1 gene expression levels were lower in the AA/AG risk genotypes compared to the G/G genotype group in the SBCS tumor tissue samples, although the difference was not statistically significant. The reason for the different effect of rs2046210 on ESR1 gene expression in tumor and adjacent normal tissue is not clear. Genetic variations in this locus may affect allelic expression imbalance and/or DNA methylation, which may in turn affect ESR1 gene expression. Using TCGA RNA-seq data from 177 individuals, Li et al. found that rs2046210 genotypes were correlated with an allelic expression imbalance of the ESR1 gene. Experimental evidence also shows physical interaction between risk locus 6q25 and the promoter of ESR1 gene (18). Therefore, it is possible that SNP rs2046210 plays a regulatory role in ESR1 gene transcription.

SNP rs7763637, which is in strong LD with rs2046210, is located in multiple transcriptional regulatory regions such as the binding site of transcription factors FOS, FOSL2, and JunD. FOS and FOSL2 belong to the FOS gene family which can dimerize with proteins of the JUN family (including c-Jun, JunB and JunD) to form the transcription factor complex AP-1 (32,33). Several studies have suggested that the expression of AP-1 protein is related to breast cancer growth, progression and metastasis (34,35). Thus, SNP rs7763637 could be a potentially functional variant in breast cancer risk locus 6q25.1 that regulates the transcription of AP-1 proteins.

There are some limitations of our study. The sample size of SBCS tissue sample is not large. However, the results of the association of SNPs rs2046210 and rs7763637 with AKAP12 gene expression are consistent between SBCS samples and TCGA samples. In addition, we did not adjust for multiple testing in our analyses. Further studies with larger sample sizes are warranted to confirm our findings.

In summary, SNP rs2046210 and SNPs in strong LD with it may influence the expression of the AKAP12 gene in breast tissues. Future research is needed to elucidate the potential mechanism by which the AKAP12 gene is related to breast cancer risk.

Supplementary material
Supplementary Tables 1-3 can be found at http://carcin.oxfordjournals.org/

URLs
The Cancer Genome Atlas (TCGA), http://cancergenome.nih.gov/
1000 Genomes, http://browser.1000genomes.org
Minimac, http://genome.sph.umich.edu/wiki/Minimac
ENCODE, http://genome.ucsc.edu/ENCODE/
UCSC Genome Browser, http://genome.ucsc.edu
RegulomeDB, http://regulome.stanford.edu/

Funding
National Institutes of Health (R01CA064277, R01CA090899 and R01CA122756); Fogarty International Center training grant (D43 TW08313).

Acknowledgements
The authors wish to thank the study participants and research staff for their contributions and commitment to this project. We thank Regina Courtney and Jie Wu for laboratory assistance, and Nan Kennedy for assistance with editing and manuscript preparation. The sample preparation and gene expression assays was performed at the Survey and Biospecimen Shared Resource, which is supported in part by the Vanderbilt-Ingram Cancer Center (P30CA068485).

Conflict of Interest Statement: None declared.

References