Fine-Scale Mapping of the 4q24 Locus Identifies Two Independent Loci Associated with Breast Cancer Risk


Abstract

**Background:** A recent association study identified a common variant (rs9790517) at 4q24 to be associated with breast cancer risk. Independent association signals and potential functional variants in this locus have not been explored.

**Methods:** We conducted a fine-mapping analysis in 55,540 breast cancer cases and 51,168 controls from the Breast Cancer Association Consortium.

**Results:** Conditional analyses identified two independent association signals among women of European ancestry, represented by rs9790517 (conditional $P = 2.51 \times 10^{-4}$; OR, 1.04; 95% confidence interval (CI), 1.02–1.07) and rs77928427 ($P = 1.86 \times 10^{-4}$; OR, 1.04; 95% CI, 1.02–1.07). Functional annotation using data from the Encyclopedia of DNA Elements (ENCODE) project revealed two putative functional variants, rs62331150 and rs73838678 in linkage disequilibrium (LD) with rs9790517 ($r^2 \geq 0.90$) residing in the active promoter or enhancer, respectively, of the nearest gene, TET2. Both variants are located in DNase I hypersensitivity and transcription factor–binding sites. Using data from both The Cancer Genome Atlas (TCGA) and Molecular Taxonomy of Breast Cancer International Consortium (METABRIC), we showed that rs62331150 was associated with level of expression of TET2 in breast normal and tumor tissue.

**Conclusion:** Our study identified two independent association signals at 4q24 in relation to breast cancer risk and suggested that observed association in this locus may be mediated through the regulation of TET2.

**Impact:** Fine-mapping study with large sample size warranted for identification of independent loci for breast cancer risk.
Introduction

A common genetic variant at 4q24, rs9790517, was recently identified to be associated with breast cancer risk, through a combined analysis of genome-wide association studies (GWAS) together with data from a large association study using a custom array, iCOGS (1, 2). This risk variant, termed subsequently as the index SNP in this article, is located in intron 11 of TET2, a chromatin-remodeling gene that functions as a tumor suppressor. TET2 has been found to be frequently somatically mutated in multiple cancers, including breast cancer (3–9). However, the index SNP is located in a region with no evidence of functional significance. The initial GWAS reported only the most strongly statistically associated SNP in this region, although many other SNPs at the same locus also may be associated with breast cancer risk, one or more of which are causally related to breast cancer risk. Comprehensive fine-scale mapping may help to identify the variants most likely to be functionally related to risk and may enable the identification of additional independent signals.

Dense fine-scale mapping of GWAS-identified loci has successfully identified novel putative causative variants for several common diseases, including breast cancer (10–17). For example, previous fine-mapping studies of 5p15, 2q416, 2q35, 5q11, and 11q13 have identified multiple independent risk signals as well as potential causative variants in each region, using data from the Breast Cancer Association Consortium (BCAC), refs. 12, 13, 16, 18–20. The index SNP (rs9790517) at 4q24 is close to another SNP, rs7679673 (r2 = 0.42, 23 kb apart), which has been associated with prostate cancer (21). In this fine-mapping project, a dense set of SNPs in this 4q24 region was genotyped in gDNA samples obtained from 106,708 participants included in the BCAC. We then analyzed data from 3,912 genotyped and imputed SNPs in this region in an attempt to identify potential functional variants that may explain the observed association of genetic variants in this locus with breast cancer risk.

Materials and Methods

Study populations

The study included 55,540 breast cancer cases and 51,168 controls from 50 studies participating in the BCAC. Details of the studies, sample selection, and genotypes are described elsewhere (1). The dataset included 39 studies from European-ancestry populations (48,155 cases and 43,612 controls), nine from Asian populations (6,269 cases and 6,624 controls), and two from populations of African ancestry (1,116 cases and 932 controls).

Published OnlineFirst September 9, 2015; DOI: 10.1158/1055-9965.EPI-15-0363

Cancer Epidemiol Biomarkers Prev; 24(11) November 2015

1Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, Tennessee. 2Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Copenhagen, Copenhagen, Denmark. 3Department of Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, United Kingdom. 4Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, United Kingdom. 5Cancer Epidemiology Centre, Cancer Council Victoria, Melbourne, Victoria, Australia. 6Centre for Epidemiology and Biostatistics, School of Population and Global Health, The University of Melbourne, Melbourne, Victoria, Australia. 7Department of Genetics, QIMR Berghofer Medical Research Institute, Brisbane, Australia. 8Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital, Toronto, Canada. 9Department of Molecular Genetics, University of Toronto, Toronto, Canada. 10Department of Epidemiology, University of California Irvine, Irvine, California. 11Division of Clinical Epidemiology and Aging Research, German Cancer Research Center, Heidelberg, Germany. 12Department of Gynaecology and Obstetrics, University Hospital Erlangen, Friedrich-Alexander University Erlangen-Nuremberg, Erlangen, Germany. 13Human Cancer Genetics Program, Spanish National Cancer Research Centre, Madrid, Spain. 14Centro de Investigación en Red de Enfermedades Raras, Valencia, Spain. 15International Epidemiology Institute, Rockville, Maryland. 16Department of Radiation Oncology, Hannover Medical School, Hannover, Germany. 17Copenhagen General Population Cohort, Herlev Hospital, Herlev, Denmark. 18Department of Clinical Biochemistry, Herlev Hospital, Copenhagen University Hospital, Herlev, Denmark. 19Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. 20Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Stuttgart, Germany. 21University of Tubingen, Tubingen, Germany. 22German Cancer Consortium, German Cancer Research Center, Heidelberg, Germany. 23Division of Preventive Oncology, German Cancer Research Center (DKFZ), Heidelberg, Germany. 24Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, Maryland. 25Netherlands Cancer Institute, Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands. 26Institute for Prevention and Occupational Medicine of the German Social Accident Insurance, Bochum, Germany. 27Division of Molecular Genetic Epidemiology, German Cancer Research Center, Heidelberg, Germany. 28Molecular Epidemiology Group, German Cancer Research Center, Heidelberg, Germany. 29Netherlands Cancer Institute, Antoni van Leeuwenhoek Hospital, Amsterdam, the Netherlands. 30Department of Cancer Epidemiology, German Cancer Research Center, Heidelberg, Germany. 31Department of Biomedical Sciences, Seoul National University College of Medicine, Seoul, Korea. 32Cancer Research Institute, Seoul National University College of Medicine, Seoul, Korea. 33Division of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, Minnesota. 34Sheffield Cancer Research Centre, Department of Oncology, University of Sheffield, Sheffield, United Kingdom. 35Academic Unit of Pathology, Department of Neuroscience, University of Sheffield, Sheffield, United Kingdom. 36Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. 37Department of Pathology, Leiden University Medical Center, Leiden, The Netherlands. 38Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands. 39Centre Hospitalier Universitaire de Québec Research Center, Laval University, Quebec, Canada. 40Gynaecology Research Unit, Hannover Medical School, Hannover, Germany. 41David Geffen School of Medicine, Department of Medicine, Division of Hematology and Oncology, University of California at Los Angeles, Los Angeles, California. 42Division of Cancer Studies, Breakthrough Breast Cancer Research Centre, Institute of Cancer Research, London, United Kingdom. 43Department of Breast Surgery, Herlev Hospital, Copenhagen University Hospital, Herlev, Denmark. 44Molecular Diagnostics Laboratory, IRRP, National Centre for Scientific Research "Demokritos", Athens, Greece. 45International Agency for Research on Cancer, Lyon, France. 46Division of Genetics and Epidemiology, Institute of Cancer Research, London, United Kingdom. 47Department of Surgery, Oulu University Hospital and University of Oulu, Oulu, Finland. 48Environmental Epidemiology of Cancer, Center for Research in Epidemiology and Population Health, INSERM, Villejuif, France. 49University Paris-Sud, Villejuif, France. 50Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California. 51Department of Genetics, Keck School of Medicine, University of Southern California, Los Angeles, California. 52Molecular Genetics of Breast Cancer, German Cancer Research Center, Heidelberg, Germany. 53Saw Swee Hock School of Public Health, National University of Singapore, Singapore, Singapore. 54Department of Surgery, National University Health System, Singapore. 55Department of Medical Oncology, Erasmus University Medical Center, Rotterdam, The Netherlands. 56Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan. 57Peter MacCallum Cancer Centre, East Melbourne, Victoria, Australia. 58Division of Epidemiology and Prevention, Aichi Cancer Center Research Institute, Aichi, Japan. 59Division of Genetics and Pathology, Pomeranian Medical University, Szczecin, Poland. 60Department of Preventive Medicine, Seoul National University College of Medicine, Seoul, Korea. 61Department of Obstetrics and Gynecology, Helsinki University Central Hospital, University of Helsinki, Helsinki, Finland. 62Prosserman Centre for Health Research, Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, Ontario, Canada. 63Division of Epidemiology, Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada. 64School of Medicine, Institute of Clinical Medicine, Pathology and Forensic Medicine and Cancer Center of Eastern Finland, Finland.
Genotyping of 4q24

A dense set of SNPs at 4q24 was selected for genotyping on iCOGS based on evidence of a prostate cancer–associated SNP, rs7679673 (17), as at the time of the assay design this region had not yet been linked to breast cancer risk. An interval of 596 kb (positions in chr4, 105932103–106528262 from hg19) was identified on the basis of all SNPs with r2 > 0.1 with the SNP rs7679673 based on HapMap 2 CEU (22). All SNPs in the interval were then identified from the 1000 Genomes Project CEU (April 2010; ref. 23), together with HapMap 3, and we selected SNPs for genotyping which had a minor allelic frequency (MAF) > 2% in Europeans and an Illumina Design score > 0.8. From this set, all SNPs with r2 > 0.1 with SNP rs7679673 were selected, together with an additional set of SNPs to tag the remaining SNPs at r2 > 0.9. In total, 490 SNPs were successfully genotyped and passed quality control. We imputed genotypes for the remaining SNPs using the program IMPUTE2 (24) and the March 2012 release of the 1000 Genomes Project as a reference. Those imputed SNPs with common SNPs (MAF > 0.02) and imputation r2 > 0.3 were included in the current analysis.

Statistical analyses

For each genotyped and imputed SNP, we evaluated its association with breast cancer risk using a logistic regression model with adjustment for age, study site, and principal components to correct for potential population stratification (the first six principal components, plus one additional principal component for the LMBC in analyses of the European ancestry data, or the first two principal components in the analyses of the Asian and African ancestry data), as previously described (1). ORs and 95% confidence intervals (CI) were estimated under a log-additive model.

We conducted separate analyses within European, Asian, and African American populations.

To identify independent association signals, we performed stepwise forward logistic regression analyses for the associated SNPs with an MAF > 0.02 showing association at P < 1 × 10−4 in the single marker SNP analysis. We used the Step function implemented in the R package (25) with the penalty K = 10 for inclusion of additional SNPs in the model. Because no SNPs showed P < 1 × 10−4 in the Asian or African populations, this analysis was performed only in the European population. The model was adjusted for the same factors as in the single SNP analysis. To define potentially causative variants, we computed a likelihood ratio for each SNP relative to the best associated SNP in each signal and excluded SNPs with a likelihood ratio < 1/100. Haplootype-specific ORs were estimated using haplo.stats in R, including age, study site, and the first six principal components, plus one additional principal component for the LMBC study.

Functional annotation

We annotated 29 candidate causative variants for potential functional significance using chromHMM annotation across nine ENCODE (26) cell lines: HMEC, GM12878, H1-hESC, K562, HepG2, HSM, HUVEC, NHEK, and NHLF (27). For each variant, we investigated whether it is mapped to functional regions (i.e., promoter and enhancer) through chromatin states annotation from the UCSC Genome Browser (28). The epigenetic landscape of histone markers H3K4Me1, H3K4Me3, and H3K27Ac was also examined through layered histone tracks on seven ENCODE cell lines including GM12878, H1-hESC, K562, HSM, HUVEC, NHEK, and NHLF from the UCSC Genome Browser.
DNase I hypersensitive and TF ChIP-Seq datasets were investigated in all available ENCODE cell lines, including breast normal cell line, human mammary epithelial cell (HMEC), and breast cancer cell lines, T-47D and MCF-7. Two publicly available tools, RegulomeDB (29) and HaploReg v2 (30), were also used to evaluate those likely functional variants (9, 31). In addition, we also investigated whether each variant is overlapped with regulatory elements of enhancers and transcription start sites (TSS) from two previous studies including Hnisz and colleagues (32) and Andersson and colleagues (FANTOM5 project; ref. 33). Chromatin Interaction Analysis by Paired End Tag (ChiA-PET; mediated by RNA polymerase 2) data from MCF7 cell were downloaded from GEO (GSE39495), and the ggbio R package was used to represent the interactions between cell enhancers (containing a strongly associated variant) and a predicted gene promoter.

The Cancer Genome Atlas data resource and eQTL analysis

We downloaded RNA-Seq V2 data (level 3) of 1,006 breast cancer tumor tissues from The Cancer Genome Atlas (TCGA) data portal (34). DNA methylation data measured by the Illumina HumanMethylation450 BeadChip were also retrieved from TCGA level 3 data. We also downloaded level 3 SNP data genotyped using the Affymetrix SNP 6.0 array. Copy number alteration (CNA) data for genes PPA2, ARHGEF38, INTS12, GSTCD, and TET2 at 4q24 for TCGA samples were collected from the ChioPortal (35). We analyzed a total of 645 breast tumor tissues in Caucasian population including matched copy number variation, genotype, and expression data.

We performed eQTL analysis in TCGA tumor tissues described above. We applied several steps to reduce the batch or other technical effects on gene expressions following the approach described by Pickrell and colleagues (36). First, the RNA-Seq by expectation–maximization value of each gene was log2-transformed and those genes with a median expression level of 0 across tissues were removed. We then performed the principal component correction on gene expression to remove potential batch effects. A linear regression of expression values on the first five principal components was constructed and the residuals were used to replace the expression values of each gene among tissues. To make the data better conform to the linear model for the eQTL analysis, we further transformed the gene expression levels to fit quantiles of N(0,1) distribution on the basis of the ranks of the expression values to their respective quantiles. Residual linear regression models were constructed to detect eQTLs, while adjusting for methylation and CNA, according to the approach used by Li and colleagues (37).

We also extracted matched genotypes and gene expression levels as described above in a total of 135 tumor-adjacent normal tissues from the TCGA tumor cohort.
Results

Association analyses

We evaluated associations for 490 genotyped and 3,422 well-imputed SNPs at 4q24 spanning 596 kb (positions in chr4: 105,932,103–106,528,262 from hg19) in 48,155 cases and 43,612 controls of European descent. A total of 29 variants were significantly associated with breast cancer risk at \( P < 1 \times 10^{-4} \) (Fig. 1; Supplementary Table S1). Of these, 15 variants were directly genotyped and 14 were imputed with \( r^2 > 0.9 \). All risk-associated variants had MAF > 0.05. The index SNP, rs9790517, showed strong evidence of a significant association with breast cancer risk (OR, 1.05; 95% confidence interval (CI), 1.03–1.08; \( P = 5.44 \times 10^{-4} \)), which was consistent with the report from the original study (1). The strongest association was, however, found for an imputed SNP rs73838678 (OR, 1.12; 95% CI, 1.07–1.17; \( P = 1.29 \times 10^{-6} \)).

To identify potential independent association signals, we carried out forward stepwise logistic regression analysis on two independent association signals for overall breast cancer risk among women of European ancestry

<table>
<thead>
<tr>
<th>Signal</th>
<th>SNPs</th>
<th>Position (hg 19)</th>
<th>Alleles</th>
<th>RAF</th>
<th>LD (( r^2 ))</th>
<th>OR (95% CI)</th>
<th>( P_{\text{adj}} )</th>
<th>OR (95% CI)</th>
<th>( P_{\text{adj}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>rs9790517</td>
<td>106,084,778</td>
<td>T/C</td>
<td>0.23</td>
<td>—</td>
<td>1.05 (1.03–1.08)</td>
<td>4.07 \times 10^{-6}</td>
<td>1.04 (1.02–1.07)</td>
<td>1.86 \times 10^{-4}</td>
</tr>
<tr>
<td>2</td>
<td>rs77928427</td>
<td>106,356,761</td>
<td>A/C</td>
<td>0.24</td>
<td>0.04</td>
<td>1.05 (1.03–1.08)</td>
<td>2.51 \times 10^{-4}</td>
<td>1.04 (1.02–1.07)</td>
<td>1.04 \times 10^{-4}</td>
</tr>
</tbody>
</table>

SNPs associated with breast cancer at \( P < 1 \times 10^{-4} \). Two independent association signals were revealed: index SNP rs9790517 (conditional \( P = 2.51 \times 10^{-4} \)), after adjustment for the SNP in the second signal) and SNP rs779284277 (conditional \( P = 1.86 \times 10^{-4} \), after adjusting for the index SNP; Table 1). The index SNP rs9790517 in signal 1 was in weak linkage disequilibrium (LD) with the SNP rs77928427 in the second risk signal (\( r^2 = 0.04 \)). These two SNPs are more than 300 kb apart from each other.

We performed similar analyses, restricting to cases with estrogen receptor–positive (ER+ cancer) and identified 17 variants associated with ER+ breast cancer risk at \( P < 1 \times 10^{-4} \) in women of European ancestry. No SNP was found to be associated with ERnegative (ER-) disease at \( P < 1 \times 10^{-4} \). However, the per-allele ORs for the two SNPs independently associated with overall breast cancer risk were similar for ER+ and ER- disease (Table 1; all tests of heterogeneity by ER status: \( P > 0.10 \)). Conditional analysis yielded similar associations for ER+ breast cancer to those for overall breast cancer for the two independently associated SNPs.

We performed haplotype analysis on the basis of the top SNPs from the two signals: rs9790517 and rs779284277 in European descendants. Three major haplotypes were observed. Compared with the most common haplotype carrying the common allele at both SNPs, haplotype TA carrying two risk alleles showed the strongest association with breast cancer risk (OR, 1.11; 95% CI, 1.07–1.15; \( P = 2.31 \times 10^{-6} \); Table 2). The frequency of this haplotype was 9.4%. Haplotypes CA and TC, carrying the risk allele in either signal 1 or 2, also were associated with elevated risk of breast cancer, although the association was only marginally significant. Thus, the haplotype analyses were consistent with the hypothesis that there are two independently associated variants in the region.

To identify potential independent association signals, we carried out forward stepwise logistic regression analysis on
We carried out association analysis for all SNPs with breast cancer in subjects of Asian and African descent. None of the SNPs identified in women of European ancestry as associated at \( P < 10^{-4} \) showed a significant association in either Asians or African women at \( P < 0.05 \) (Table 3). However, the 95\% CI for the OR estimates in Asians and Africans included the point estimate in Europeans for both of the top independent SNPs. We found one SNP associated with breast cancer risk in Asians and three in Africans, at \( P < 0.01 \) (strongest signal rs1116764: OR, 1.10; 95\% CI, 1.04–1.16; \( P = 4.21 \times 10^{-4} \)), none of these SNPs were in LD with the two independent association signals identified in European women (Table 3).

Functional annotation

We used a likelihood ratio > 1:100 relative to the best associated SNP in each signal to select candidate variants for functional annotation to identify potentially causative variants in this region (Supplementary Table S1). In total, 29 SNPs were identified including 24 for signal 1 and 5 for signal 2. Of these, 17 SNPs in signal 1 were strongly correlated with the original index SNP rs9790517, and the remainder was more weakly correlated. All SNPs were evaluated using DNase-Seq and ChIP-Seq data from the ENCODE project. The most promising evidence for functionality was found for SNPs rs62331150 and rs73838678, both in LD with rs9790517 (\( r^2 \approx 0.98 \) and \( r^2 \approx 0.09 \), respectively) in signal 1. The annotation from chromatin states (27) revealed that rs62331150 resides an active promoter region, and rs73838678 in a strong enhancer region, on several ENCODE cell lines including HMEC but not for other SNPs in either signal 1 or 2 (Fig. 2A). The active promoter-associated histone marks (H3K4me3 and H3K27Ac) and enhancer-associated histone marker H3K27Ac were enriched in the intervals containing rs62331150 and rs73838678, respectively, in several ENCODE cells, and both SNPs were also found to be located in or near a DNase I hypersensitive site (DHS; Fig. 2A and B). In addition, both variants were found to overlap with predicted enhancer regions of TET2 in multiple cells including HMEC as reported in a recent study (32). None of the other SNPs in signal 1, and none of the 5 SNPs in signal 2 fell into a strong annotated promoter or enhancer region in those cells.

To identify putative gene targets, we examined the annotation of TSS and TSS-associated enhancers using Cap Analysis of Gene Expression (CAGE) from the FANTOM5 project (23). We found that rs62331150 and rs73838678 reside in regulatory elements of enhancers associated with TSS and TSS of TET2 in multiple cells (Fig. 2A). We also examined potential functional chromatin interactions between distal and proximal regulatory transcription factor (TF)-binding sites and the promoters at the risk regions using ChIA-PET data. ChIA-PET data for Pol2 in MCF-7 breast tumor-derived cells showed multiple chromosomal interactions across the entire region, but these interactions were particularly dense in the vicinity of the TET2 promoter region, encompassing the strongest candidate causal variant rs62331150 and rs73838678 (Fig. 2A).

A search of RegulomeDB indicated that rs62331150 and rs73838678 were annotated to lie in the breast cancer related TF SP1 (specificity protein 1) and PR (progesterone receptor; refs. 40, 41) predicted binding motifs, respectively (Fig. 2B). We observed that the G nucleotide was more frequently found in the SP1 motif than the T nucleotide, indicating that the SP1 may preferentially bind to the reference G allele (Fig. 2B). For variant rs73838678, no significant allele frequency difference in the PR motif was observed. Using ChIP-Seq data from a total of 161 TFs from the ENCODE project (ChIP-Seq V3), we found that both variants are located in multiple TF-binding sites (Fig. 2B). As an example, ChIP-Seq binding peaks of breast cancer-related TFs, EGR1 and NIFC, harbor the variant rs62331150 and rs73838678, respectively (42, 43). In particular, we observed that P300, marking the active enhancer, was found to bind close to both variants in multiple ENCODE cell lines, suggesting that the variant in the region may lead to TET2 transcriptional activation.

<table>
<thead>
<tr>
<th>Top SNPs</th>
<th>Alleles</th>
<th>RAF</th>
<th>LD (( r^2 ))</th>
<th>OR (95% CI)</th>
<th>( P_{\text{meta}} )</th>
<th>LD (( r^2 ))</th>
<th>OR (95% CI)</th>
<th>( P_{\text{meta}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Signal 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs9790517*</td>
<td>T/C</td>
<td>0.60</td>
<td>—</td>
<td>1.00 (0.95–1.06)</td>
<td>0.93</td>
<td>0.06</td>
<td>—</td>
<td>1.21 (0.98–1.55)</td>
</tr>
<tr>
<td>Signal 2</td>
<td>rs79219151</td>
<td>A/C</td>
<td>0.05</td>
<td>0.01</td>
<td>1.02 (0.91–1.12)</td>
<td>0.50</td>
<td>0.16</td>
<td>0</td>
</tr>
<tr>
<td><strong>Signal 2 rs77928427</strong></td>
<td>G/A</td>
<td>0.66</td>
<td>0.13</td>
<td>1.10 (1.04–1.16)</td>
<td>4.21 ( \times 10^{-4} )</td>
<td>0.89</td>
<td>0</td>
<td>1.02 (0.81–1.23)</td>
</tr>
<tr>
<td><strong>Signal 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs79219151</td>
<td>C/T</td>
<td>NA</td>
<td></td>
<td></td>
<td></td>
<td>0.95</td>
<td>0</td>
<td>1.63 (1.13–2.31)</td>
</tr>
<tr>
<td>rs11095278</td>
<td>C/T</td>
<td>NA</td>
<td></td>
<td></td>
<td></td>
<td>0.95</td>
<td>0</td>
<td>1.65 (1.16–2.14)</td>
</tr>
<tr>
<td>rs144956461</td>
<td>A/T</td>
<td>NA</td>
<td></td>
<td></td>
<td></td>
<td>0.93</td>
<td>0</td>
<td>1.56 (1.12–2.10)</td>
</tr>
</tbody>
</table>

Abbreviation: RAF, risk allele frequency.

*Index SNP

**Risk/reference allele; risk alleles are shown in bold.

\( r^2 \) for LD with the index SNP rs9790517 in Asians and Africans, respectively.

*Adjusted for age, study, and the first six PC and an additional PC for LMBC study.

**Table 3. Association of lead SNPs identified in women of European and non-European descent with breast cancer risk among women of Asian (6,269 cases and 6,624 controls) and African ancestry (116 cases and 932 controls).**

We used both TCGA and Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) data to examine the association of the putative functional SNP rs62331150 and rs73838678 with expression of TET2 and several other neighboring genes, including PP2A, ARHGFB38, INTS12, and GSTCD, in breast cancer tissues. No significant correlations with any genes were observed for variant rs73838678. Variant rs62331150 was weakly correlated with TET2 expression in both datasets \( (P = 0.039 \) and \( P = 0.025 \), respectively, for TCGA and METABRIC), the reference allele G being associated with increased expression relative to the risk allele T (Fig. 3).
The result was consistent with the observation from our functional annotation that SP1 may preferentially bind to the reference G allele, leading to a significant increase in TET2 transcription activation. No correlation between rs62331150 and the expression of any other gene in the region was found in either dataset. Overall, our findings supported a hypothesis that TET2 is the target gene for the signal 1 association and that the association with breast cancer risk may be mediated through regulation of TET2 gene expression. The result is also in line with previous findings that TET2 functions as a tumor suppressor and its high expression level may reduce breast cancer risk (44, 45).
Fine Mapping of 4q24 Locus for Breast Cancer Risk

Discussion

In this study, we identified two independent association signals at 4q24 in women of European ancestry. Statistical analyses reduced the set of likely causative variants to 29. Using functional genomic data, we provided strong evidence for two variants as functional variants. Our study suggests that the breast cancer risk may be mediated through their regulation of TET2 gene expression.

In our initial single marker analysis, we observed that the majority of variants, including the index SNP, were located in or near the TET2 gene region. Through eQTL analysis based on TCGA data, we found that multiple SNPs in signal 1 were correlated with TET2 expression, which was expected given their strong LD with each other. Of those SNPs, rs62331150 resides in the promoter region of TET2. Although eQTL analysis is helpful to identify potential target genes, it is difficult to use eQTL results to pinpoint the causal variant particularly when multiple SNPs are in strong LD. In addition to residing in the promoter region of the TET2 gene, the variant rs62331150 was also found to be located in the binding sites of multiple TFs including the breast cancer–related TF EGR1, potentially affecting the binding affinities of specific TFs. Interestingly, the putative functional SNP rs62331150 is close to SNP rs7679673 that has been associated with prostate cancer risk (21), indicating that TET2 gene may also be involved in prostate cancer risk. In comparison to rs62331150, rs73838678 in signal 1 was not found to have a significant association with TET2 and any other nearby genes. One possible reason is that the statistical power is low for rs73838678 due to its relative low allele frequency (MAF, 0.049). We also could not exclude the other possible target genes for rs73838678. Future studies using in vitro and in vivo assays are warranted to verify this conclusion. Cumulative evidence shows that TET2 has an important function in tumor suppression. This gene can alter the epigenetic status of DNA base methylation to 5-hydroxymethylcytosine and therefore, have a genome-wide scale of influence on gene expression (46–48). Accordingly, TET2 gene dysregulation could cause aberrant DNA methylations and consequently contribute to cancer development (3–6, 45, 49). Here, we reported TET2 as a candidate susceptibility gene for both ER+ and ER−breast cancer types. Although the associations for the top SNPs, rs9790517 and rs77928427, with breast cancer risk in Asian- and African-ancestry populations were not statistically significant, likely due to a small sample size, the direction of the associations was mostly consistent in all population, suggesting that the TET2 gene play a similar role in the etiology of breast cancer in all three populations.

Although our fine-mapping analysis represents the most comprehensive analysis of variants at 4q24 thus far, many SNPs, particularly rare variants, cannot be imputed. Deep sequencing of this region may reveal additional risk variants for breast cancer. For example, rs76682196, located 884 bp upstream of rs62331150, was found to be potentially functional using the ENCODE data. The variant is present in DHS and TF sites. In particular, it lies in the ERE-predicted binding motif and ChIP-Seq peak in breast cancer cell line T-47D. However, this variant was not included in the study due to its low frequency (MAF < 0.01) in populations from all three ethnic groups.

In conclusion, this dense fine-mapping study identified two independent association signals with breast cancer risk at 4q24, increasing the estimated familial relative risk of breast cancer explained by this locus from the original 0.07% to 0.15% among women of European descent. Functional analyses revealed one potentially functional variant, rs62331150. The risk allele is associated with lower expression of TET2, consistent with previous findings that this gene acts as a tumor suppressor.

Disclosure of Potential Conflicts of Interest

P.A. Fasching reports receiving commercial research grants from Amgen and Novartis, and has received speakers bureau honoraria from Amgen, Celgene, GSK, Novartis, Pfizer, Roche, and Teva. No potential conflicts of interest were disclosed by the other authors.

Disclaimer

The content of this manuscript does not necessarily reflect the views or policies of the National Cancer Institute or any of the collaborating centers in the Breast Cancer Family Registry (BCFR), nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government or the BCFR.

Authors’ Contributions


www.aacjrournals.org
Acknowledgments

The authors thank all the individuals who took part in these studies and all the researchers, study staff, clinicians, and other healthcare providers, technicians and administrative staff who have enabled this work to be carried out. In particular, they thank Andrew Berchuck (OCAC), Rosanna et al, Ali Amin Al Olama, Tarek Kotb-Isara, Lai-Wei Tung, and the team at the Danish Breast Cancer Cooperative (DBCG). They acknowledge the support of the Breast Cancer Association (BCA) and the Breast Cancer Research Center (DKFZ) [HB] Department of Internal Medicine, Evangelische Kliniken Bonn gGmbH, Johanniter Krankenhau, Bonn, Germany [Yvon-Duchan Ko, Christian Baich]. Institute of Pathology, University of Bonn, Germany [Hans-Peter Fischer], Molecular Genetics of Breast Cancer, Deutsches Krebsforschungszentrum [DKFZ] Heidelberg, Germany [Heth Hamann], Institute for Prevention and Occupational Medicine of the German Social Accident Insurance, Institute of the Ruhr University Bochum (IAP), Germany [TB, Beatrice Schen, Sylvia Rabenstein, Anne Loes]. Institute of Occupational Medicine and Maritime Medicine, University Medical Center Hamburg-Eppendorf, Germany [Volker Hartl)] Tuomas Heikkinen, Irja Erkkila, Kirsini Aaltonen, Karl von Smitten, Natalia Antonenkovna, Peter Hillermann, Hans Christiansen, Elja Miyolainen, Helena Kemilainen, Heitali Thorne, Eveline Niemi, Peter Bugert, the Ingram Professorship and Anne Potter Wilson Chair. BCAC is funded by Cancer Research UK (C1287/A10118, C1287/A12014) and by the European Community’s Seventh Framework Programme under grant agreement n° 223175 (HEALTH-F2-2009-223175, COGS). Meetings of the BCAC have been funded by the European Cancer Research and Treatment Centre (EORTC), the Ingram Professorship and Anne Potter Wilson Chair. BCAC is funded by Cancer Research UK (C1287/A10118, C1287/A12014) and by the European Community’s Seventh Framework Programme under grant agreement n°223175 (HEALTH-F2-2009-223175, COGS). Meetings of the BCAC have been funded by the European Cancer Research and Treatment Centre, EORTC. The iCOGS array was funded by the European Union (EU).
financed by the Swedish Cancer Society (5128-B07-01PAF). The SEBCS was supported primarily by NIH grants R01CA64277, R01CA48667, and R37CA070876. Biological sample preparation was conducted the Survey and Biospecimen Shared Resource, which is supported by P30 CA68485. The SEBCS was supported by Yorkshire Cancer Research 5305PA, S299 and S295, and the Sheffield Experimental Cancer Medicine Centre. Funding for the SCCS was provided by NIH grant R01 CA092447. The Arkansas Central Cancer Registry is fully funded by a grant from National Program of Cancer Registries, Centers for Disease Control and Prevention (CDC). Data on SCCS cancer cases from Mississippi were collected by the Mississippi Cancer Registry, which participa-
tes in the National Program of Cancer Registries (NPCR) of the Centers for Disease Control and Prevention (CDC). The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of the CDC or the Mississippi Cancer Registry. SEARCH is funded by a programme grant from Cancer Research UK (C490/A10124) and sup-
ported by the UK National Institute for Health Research Biomedical Research Centre at the University of Cambridge. The SEBCS was supported by the BRL (Basic Research Laboratory) program through the National Research Foundation of Korea funded by the Ministry of Education, Science, and Technology (2012-00000347). SGCCS was funded by the NIH start-up Grant, NCIS Centre Grant, and NMRC Clinician Scientist Award. Additional controls were recruited by the Singapore Consortium of Cohort Studies-Multi-ethnic cohort (SCCS-MEC), which was funded by the Biomedical Research Council, grant number: 05/121/19/425. SKKDKEZTS is supported by the DKEZ. The SZBCS was supported by Grant PR2_XBN_122/P03/2004. K.J. is a fellow of International PhD program, Postgraduate School of Molecular Medicine, Warsaw Medical University, supported by the Polish Foundation of Science. The TNBCC was supported by the NIH grant (CA128978), Breast Cancer Medician Research Foundation, Komen Foundation for the Cure, the Ohio State University Comprehensive Cancer Center, the Stefanie Spielman Fund for Breast Cancer Research, and a generous gift from the David F. and Margaret T. Grohne Family Foundation and the Ting Tsung and Wei Fong Chiao Foundation. Part of the TNBCC (DEMOKRITOS) has been co-financed by the European Union (European Social Fund–ESF) and Greek National Funds through the Operational Program ‘Education and Lifelong Learning’ of the National Strategic Reference Framework (NSRF)—Research Funding Program of the General Secretariat for Research & Tech-
ology: ARISTEIA. The TBWCS is supported by the Institute of Biomedical Sciences, Academia Sinica and the National Science Council, Taiwan. The UKBCS is funded by Breakthrough Breast Cancer and the Institute of Cancer Research (ICR). ICR acknowledges NHS funding to the NIHBR Biomedical Research Centre.

Received April 9, 2015; revised July 14, 2015; accepted July 20, 2015; published OnlineFirst September 9, 2015.

References

13. HapMap project [Internet]. Bethesda (MD): National Institutes of Health, National Library of Medicine, National Center for Biotechnology Informa-
15. Guo et al.
28. UCSC Genome Browser [Internet]. Santa Cruz (CA): genome bioinformatics group, center for biomolecular science and engineering at the University of California Santa Cruz. [cited 2015 Sept. 21]. Available from: http://genome.ucsc.edu.


Fine-Scale Mapping of the 4q24 Locus Identifies Two Independent Loci Associated with Breast Cancer Risk


Cancer Epidemiol Biomarkers Prev 2015;24:1680-1691. Published OnlineFirst September 9, 2015.

Updated version
Access the most recent version of this article at:
doi:10.1158/1055-9965.EPI-15-0363

Supplementary Material
Access the most recent supplemental material at:
http://cebp.aacrjournals.org/content/suppl/2015/09/09/1055-9965.EPI-15-0363.DC1

Cited articles
This article cites 38 articles, 8 of which you can access for free at:
http://cebp.aacrjournals.org/content/24/11/1680.full.html#ref-list-1

Citing articles
This article has been cited by 1 HighWire-hosted articles. Access the articles at:
/content/24/11/1680.full.html#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.