Fine-scale mapping of the 4q24 locus identifies two independent loci associated with breast cancer risk

A full list of authors and affiliations appears at the end of the article.

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% 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 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.

Keywords

Breast Cancer; Genetics; GWAS; 4q24; TET2

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,

*Correspondence to: Wei Zheng, MD, PhD, Vanderbilt Epidemiology Center, Vanderbilt University School of Medicine, 2525 West End Avenue, 8th Floor, Nashville, TN 37203-1738, USA, Phone: (615) 936-0682; Fax: (615) 936-8241, wei.zheng@vanderbilt.edu.

2). This risk variant, termed subsequently as the index SNP in this paper, 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, 20q16, 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) (12, 13,16, 18–20). The index SNP (rs9790517) at 4q24 is close to another SNP, rs7679673 (r^2 = 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 genomic DNA 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 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).

Genotyping of 4q24

A dense set of SNPs at 4q24 were selected for genotyping on iCOGS based on evidence of a prostate cancer associated SNP, rs7679673 (17), since at the time of the assay design this region had not yet been linked to breast cancer risk. An interval of 596kb (positions in chr4, 105932103 – 106528262 from hg19) was identified based on all SNPs with r^2 > 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)(23), together with HapMap 3, and we selected SNPs for genotyping which had an MAF > 2% in Europeans and an Illumina Design score > 0.8. From this set, all SNPs with r^2 > 0.1 with SNP rs7679673 were selected, together with an additional set of SNPs to tag the remaining SNPs at r^2 > 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 $r^2 > 0.3$ 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). Odds ratios (ORs) and 95% confidence intervals (CIs) 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 $p < 1 \times 10^{-4}$ 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. Since no SNPs showed $p < 1 \times 10^{-4}$ 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. Haplotype-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, HSMM, 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, HSMM, 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, 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 TSS from two previous studies including Hnisz et al (32) and Andersson et al (FANTOM5 project) (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.

TCGA data resource and eQTL analysis

We downloaded RNA-Seq V2 data (level 3) of 1,006 breast cancer tumor tissues from the 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 CbioPortal (35). We analyzed a total of 645 breast tumor tissues in Caucasian population including matched CNV, 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 et al (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 based on 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 et al (37).

We also extracted matched genotypes and gene expression levels as described above in a total of 135 tumor-adjacent normal breast tissues in European ancestry individuals from the METABRIC project (38). Gene expression profiling was generated on the Illumina HT12 v3 microarray platform and probe-level measurements were used. Genotyping was performed on the Affymetrix SNP 6.0 with genotypes being imputed using the 1000 Genomes March 2012 CEU reference panel. Matrix eQTL was performed for evaluating the association between genotypes and gene expression levels (39).

Results

Association Analyses

We evaluated associations for 490 genotyped and 3,422 well-imputed SNPs at 4q24 spanning 596 kb (positions in chr4: 105932103 – 106528262 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 $p < 1 \times 10^{-4}$ (Figure 1, Supplementary Table 1). Of these, 15 variants were directly genotyped and 14 were imputed with $r^2 > 0.9$. All risk-associated variants had minor allele frequencies (MAF) > 0.05. The index SNP, rs9790517, showed strong evidence of a significant association with breast cancer risk (OR = 1.05; 95% CI 1.03–1.08; $p = 5.44 \times 10^{-6}$), 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 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 rs77928427 (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 LD with the SNP rs77928427 in the second risk signal ($r^2 = 0.04$). These two SNPs are more than 300kb 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 ER-negative (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 based on the top SNPs from the two signals: rs9790517 and rs77928427 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; $p = 2.31 \times 10^{-8}$) (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.

We compared the average age among those cases carrying risk and non-risk alleles of rs9790517. Interestingly, we observed that the cases carrying risk alleles were slightly younger than those carrying non-risk alleles (average age: 57.54, 57.62 and 57.64 respectively for patients carrying alleles TT, TC and CC of rs9790517; $p < 2 \times 10^{-16}$). No such pattern was observed for rs77928427.

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 two 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 in order to identify potentially causative variants in this region (Supplementary Table 1). 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 were 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 ($r^2 = 0.98$ and $r^2 = 0.09$, 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 (Human Mammary Epithelial Cell) but not for other SNPs in either signal 1 or 2 (Figure 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) (Figure 2A, 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 transcription start sites (TSS) and TSS of *TET2* in multiple cells (Figure 2A). We also examined potential functional chromatin interactions between distal and proximal regulatory transcription-factor 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 (Figure 2A).

A search of RegulomeDB indicated that rs62331150 and rs73838678 were annotated to lie in the breast cancer related transcription factor (TF) SP1 (Specificity Protein 1) and PR (progesterone receptor) (40, 41) predicted binding motifs, respectively (Figure 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 (Figure 2B). For variant rs73838678, no significant allelic 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 (Figure 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.

Gene expression analyses

We used both The Cancer Genome Atlas (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 *PPA2*, *ARHGEF38*, *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* (Figure 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).

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 others. Of those SNPs, rs62331150 resides in the promoter 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 *TERT2*, 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 methylcytosine 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 TFs sites. In particular, it lies in the ERα (Estrogen Receptor-α) 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Authors

Xingyi Guo[1], Jirong Long[1], Chenjie Zeng[1], Kyriaki Michailidou[2], Maya Ghoussaini[3], Manjeet K. Bolla[2], Qin Wang[2], Roger L. Milne[4,5], Xiao-Ou Shu[1], Qiuyin Cai[1], Jonathan Beesley[6], Siddhartha P. Kar[3], Irene L. Andrulis[7,8], Hoda Anton-Culver[9], Volker Arndt[10], Matthias W. Beckmann[11], Alicia Beeghly-Fadiel[1], Javier Benitez[12,13], William Blot[1,14], Natalia Bogdanova[15], Stig E. Bojesen[16,17,18], Hiltrud Brauch[19,20,21], Hermann Brenner[10,21,22], Louise Brinton[23], Annegien Broeks[24], Thomas Brüning[25], Barbara Burwinkel[26,27], Hui Cai[1], Sander Canisius[28], Jenny Chang-Claude[29], Ji-Yeob Choi[30,31], Fergus J. Couch[32], Angela Cox[33], Simon S. Cross[34], Kamila Czene[35], Hatef Darabi[35], Peter Devilee[36,37], Arnaud Droit[38], Thilo Dörk[39], Peter A. Fasching[11,40], Olivia Fletcher[41], Henrik Flyger[42], Florentia Fostira[43], Valerie Gaborieau[44], Montserrat García-Closas[45,41], Graham G. Giles[4,5], Mervi Grip[46], Pascal Guénel[47,48], Christopher A. Haiman[49], Ute Hamann[50], Mikael Hartman[51,52], Antoinette Hollestelle[53], John L. Hopper[5], Chia-Ni Hsiung[54], kConFab Investigators[55], Hidemi Ito[56], Anna Jakubowska[57], Nichola Johnson[41], Maria Kabisch[50], Daehee Kang[58,30,31], Sofia Khan[59], Julia A. Knight[60,61], Veli-Matti Kosma[62], Diether Lambrechts[63,64], Loic Le Marchand[65], Jingmei Li[35], Annika Lindblom[66], Artitaya Lophatananon[67], Jan Lubinski[57], Arto Mannermaa[62],

Siranoush Manoukian[68], Sara Margolin[69], Frederik Marme[70,71], Keitaro Matsuo[72], Catriona A. McLean[73], Alfons Meindl[74], Kenneth Muir[68,75], Susan L. Neuhausen[76], Heli Nevanlinna[59], Silje Nord[77,78], Janet E. Olson[79], Nick Orr[80], Paolo Peterlongo[81], Thomas Choudary Putti[82], Anja Rudolph[29], Suleeporn Sangrajrang[83], Elinor J. Sawyer[84], Marjanka K. Schmidt[24], Rita K. Schmutzler[85,86,87,88], Chen-Yang Shen[89,90], Jiajun Shi[1], Martha J Shrubsole[1], Melissa C. Southey[91], Anthony Swerdlow[92], Soo Hwang Teo[93,94], Bernard Thienpont[63,64], Amanda Ewart Toland[95], Robert A.E.M. Tollenaar[96], Ian P.M. Tomlinson[97], Thérèse Truong[47,48], Chiu-chen Tseng[49], Ans van den Ouweland[98], Wanqing Wen[1], Robert Winqvist[99,100], Anna Wu[49], Cheng Har Yip[94], M. Pilar Zamora[101], Ying Zheng[102], Per Hall[35], Paul D.P. Pharoah[2,3], Jacques Simard[38], Georgia Chenevix-Trench[6], Alison M. Dunning[3], Douglas F. Easton[2,3], and Wei Zheng[1,*]

Affiliations

[1]Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, TN, USA [2]Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK [3]Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK [4]Cancer Epidemiology Centre, Cancer Council Victoria, Melbourne, Victoria, Australia [5]Centre for Epidemiology and Biostatistics, School of Population and Global health, The University of Melbourne, Melbourne, Victoria, Australia [6]Department of Genetics, QIMR Berghofer Medical Research Institute, Brisbane, Australia [7]Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital, Toronto, Canada [8]Department of Molecular Genetics, University of Toronto, Toronto, Canada [9]Department of Epidemiology, University of California Irvine, Irvine, CA, USA [10]Division of Clinical Epidemiology and Aging Research, German Cancer Research Center, Heidelberg, Germany [11]Department of Gynaecology and Obstetrics, University Hospital Erlangen, Friedrich-Alexander University Erlangen-Nuremberg, Erlangen, Germany [12]Human Cancer Genetics Program, Spanish National Cancer Research Centre, Madrid, Spain [13]Centro de Investigación en Red de Enfermedades Raras, Valencia, Spain [14]International Epidemiology Institute, Rockville, MD, USA [15]Department of Radiation Oncology, Hannover Medical School, Hannover, Germany [16]Copenhagen General Population Study, Herlev Hospital, Herlev, Denmark [17]Department of Clinical Biochemistry, Herlev Hospital, Copenhagen University Hospital, Herlev, Denmark [18]Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark [19]Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Stuttgart, Germany [20]University of Tübingen, Tübingen, Germany [21]German Cancer Consortium, German Cancer Research Center, Heidelberg, Germany [22]Division of Preventive Oncology, German Cancer Research Center (DKFZ), Heidelberg, Germany [23]Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD, USA [24]Netherlands Cancer Institute, Antoni van Leeuwenhoek hospital, Amsterdam, The Netherlands [25]Institute for Prevention and Occupational Medicine of the German Social Accident Insurance, Bochum, Germany [26]Division of Molecular Genetic

Epidemiology, German Cancer Research Center, Heidelberg, Germany [27]Molecular Epidemiology Group, German Cancer Research Center, Heidelberg, Germany [28]Netherlands Cancer Institute, Antoni van Leeuwenhoek Hospital, Amsterdam, the Netherlands [29]Division of Cancer Epidemiology, German Cancer Research Center, Heidelberg, Germany [30]Department of Biomedical Sciences, Seoul National University College of Medicine, Seoul, Korea [31]Cancer Research Institute, Seoul National University College of Medicine, Seoul, Korea [32]Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN 55905, USA [33]Sheffield Cancer Research Centre, Department of Oncology, University of Sheffield, Sheffield, UK [34]Academic Unit of Pathology, Department of Neuroscience, University of Sheffield, Sheffield, UK [35]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden [36]Department of Pathology, Leiden University Medical Center, Leiden, The Netherlands [37]Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands [38]Centre Hospitalier Universitaire de Québec Research Center, Laval University, Québec, Canada [39]Gynaecology Research Unit, Hannover Medical School, Hannover, Germany [40]David Geffen School of Medicine, Department of Medicine Division of Hematology and Oncology, University of California at Los Angeles, Los Angeles, CA, USA [41]Division of Cancer Studies, Breakthrough Breast Cancer Research Centre, Institute of Cancer Research, London, UK [42]Department of Breast Surgery, Herlev Hospital, Copenhagen University Hospital, Herlev, Denmark [43]Molecular Diagnostics Laboratory, IRRP, National Centre for Scientific Research "Demokritos", Athens, Greece [44]International Agency for Research on Cancer, Lyon, France [45]Division of Genetics and Epidemiology, Institute of Cancer Research, London, UK [46]Department of Surgery, Oulu University Hospital and University of Oulu, Oulu, Finland [47]Environmental Epidemiology of Cancer, Center for Research in Epidemiology and Population Health, INSERM, Villejuif, France [48]University Paris-Sud, Villejuif, France [49]Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA [50]Molecular Genetics of Breast Cancer, German Cancer Research Center, Heidelberg, Germany [51]Saw Swee Hock School of Public Health, National University of Singapore, Singapore, Singapore [52]Department of Surgery, National University Health System, Singapore [53]Department of Medical Oncology, Erasmus University Medical Center, Rotterdam, The Netherlands [54]Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan [55]Peter MacCallum Cancer Centre, East Melbourne, VIC, Australia [56]Division of Epidemiology and Prevention, Aichi Cancer Center Research Institute, Aichi, Japan [57]Department of Genetics and Pathology, Pomeranian Medical University, Szczecin, Poland [58]Department of Preventive Medicine, Seoul National University College of Medicine, Seoul, Korea [59]Department of Obstetrics and Gynecology, Helsinki University Central Hospital, University of Helsinki, Helsinki, FI-00029 HUS, Finland [60]Prosserman Centre for Health Research, Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital, Toronto, ON, M5G 1X5, Canada [61]Division of Epidemiology, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, M5S 1A8, Canada

[62]School of Medicine, Institute of Clinical Medicine, Pathology and Forensic Medicine and Cancer Center of Eastern Finland, University of Eastern Finland, Kuopio, Finland; Imaging Center, Department of Clinical Pathology, Kuopio University Hospital, Kuopio, Finland [63]Vesalius Research Center, Leuven, Belgium [64]Laboratory for Translational Genetics, Department of Oncology, University of Leuven, Leuven, Belgium [65]University of Hawaii Cancer Center, Honolulu, HI, USA [66]Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm SE, Sweden [67]Division of Health Sciences, Warwick Medical School, Warwick University, Coventry, UK [68]Unit of Medical Genetics, Department of Preventive and Predictive Medicine, Fondazione IRCCS Istituto Nazionale dei Tumori (INT), Milan, Italy [69]Department of Oncology - Pathology, Karolinska Institutet, Stockholm SE, Sweden [70]National Center for Tumor Diseases, University of Heidelberg, Heidelberg, Germany [71]Department of Obstetrics and Gynecology, University of Heidelberg, Heidelberg, Germany [72]Department of Preventive Medicine, Kyushu University Faculty of Medical Sciences, Fukuoka, Japan [73]Anatomical Pathology, The Alfred Hospital, Melbourne, Victoria, Australia [74]Division of Gynaecology and Obstetrics, Technische Universität München, Munich, Germany [75]Institute of Population Health, University of Manchester, Manchester, UK [76]Beckman Research Institute of City of Hope, Duarte, CA, USA [77]Department of Genetics, Institute for Cancer Research, Oslo University Hospital, Radiumhospitalet, Ullernchausseen, Oslo, Norway [78]K.G. Jebsen Center for Breast Cancer Research, Institute for Clinical Medicine, Faculty of Medicine, University of Oslo, Kirkeveien, Oslo, Norway [79]Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA [80]Division of Breast Cancer Research, Institute of Cancer Research, London, UK; Cancer Research, Institute of Cancer Research, London, UK [81]IFOM, the FIRC Institute of Molecular Oncology, Milan, Italy [82]Department of Pathology, National University Health System, Singapore [83]National Cancer Institute, Bangkok, Thailand [84]Research Oncology, Guy's Hospital, King's College London, London, UK [85]Division of Molecular Gyneco-Oncology, Department of Gynaecology and Obstetrics, University Hospital of Cologne, Cologne, Germany [86]Center for Integrated Oncology, University Hospital of Cologne, Cologne, Germany [87]Center for Molecular Medicine, University Hospital of Cologne, Cologne, Germany [88]Center of Familial Breast and Ovarian Cancer, University Hospital of Cologne, Cologne, Germany [89]School of Public Health, China Medical University, Taichung, Taiwan [90]Taiwan Biobank, Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan [91]Department of Pathology, The University of Melbourne, Melbourne, Victoria, Australia [92]Division of Genetics and Epidemiology and Division of Breast Cancer Research, Institute of Cancer Research, London, UK [93]Cancer Research Initiatives Foundation, Subang Jaya, Selangor, Malaysia [94]Breast Cancer Research Unit, Cancer Research Institute, University Malaya Medical Centre, Kuala Lumpur, Malaysia [95]Department of Molecular Virology, Immunology and Medical Genetics, Comprehensive Cancer Center, The Ohio State University, Columbus, OH, USA [96]Department of Surgical Oncology, Leiden University Medical Center, Leiden, The Netherlands [97]Wellcome Trust Centre for Human Genetics and Oxford Biomedical

Research Centre, University of Oxford, Oxford, UK [98]Department of Clinical Genetics, Erasmus University Medical Center, Rotterdam, The Netherlands [99]Laboratory of Cancer Genetics and Tumor Biology, Department of Clinical Chemistry, University of Oulu, Oulu, Finland [100]Laboratory of Cancer Genetics and Tumor Biology, Northern Finland Laboratory Centre NordLab, Oulu, Finland [101]Servicio de Oncología Médica, Hospital Universitario La Paz, Madrid, Spain [102]Shanghai Municipal Center for Disease Control and Prevention, Shanghai, PR China

Acknowledgment

References

1. Michailidou K, Hall P, Gonzalez-Neira A, Ghoussaini M, Dennis J, Milne RL, et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. Nature genetics. 2013; 45:353–361. 61e1-2. [PubMed: 23535729]

2. iCOGs [Internet]. Cambridge (UK): Center for Cancer Genetic Epidemiology, Department of Public Health and Primary Care/Department of Oncology, University of Cambridge; Available from: http://ccge.medschl.cam.ac.uk/research/consortia/icogs/

3. Tefferi A, Lim KH, Levine R. Mutation in TET2 in myeloid cancers. The New England journal of medicine. 2009; 361:1117. author reply -8. [PubMed: 19741235]

4. Delhommeau F, Dupont S, Della Valle V, James C, Trannoy S, Masse A, et al. Mutation in TET2 in myeloid cancers. The New England journal of medicine. 2009; 360:2289–2301. [PubMed: 19474426]

5. Cancer Genome Atlas Research N. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. The New England journal of medicine. 2013; 368:2059–2074. [PubMed: 23634996]

6. Figueroa ME, Abdel-Wahab O, Lu C, Ward PS, Patel J, Shih A, et al. Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. Cancer cell. 2010; 18:553–567. [PubMed: 21130701]

7. Seshagiri S, Stawiski EW, Durinck S, Modrusan Z, Storm EE, Conboy CB, et al. Recurrent R-spondin fusions in colon cancer. Nature. 2012; 488:660–664. [PubMed: 22895193]

8. Stephens PJ, Tarpey PS, Davies H, Van Loo P, Greenman C, Wedge DC, et al. The landscape of cancer genes and mutational processes in breast cancer. Nature. 2012; 486:400–404. [PubMed: 22722201]

9. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, et al. Annotation of functional variation in personal genomes using RegulomeDB. Genome research. 2012; 22:1790–1797. [PubMed: 22955989]

10. Liu JZ, Almarri MA, Gaffney DJ, Mells GF, Jostins L, Cordell HJ, et al. Dense fine-mapping study identifies new susceptibility loci for primary biliary cirrhosis. Nature genetics. 2012; 44:1137–1141. [PubMed: 22961000]

11. Trynka G, Hunt KA, Bockett NA, Romanos J, Mistry V, Szperl A, et al. Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. Nature genetics. 2011; 43:1193–1201. [PubMed: 22057235]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

12. Bojesen SE, Pooley KA, Johnatty SE, Beesley J, Michailidou K, Tyrer JP, et al. Multiple independent variants at the TERT locus are associated with telomere length and risks of breast and ovarian cancer. Nature genetics. 2013; 45:371–384. 84e1-2. [PubMed: 23535731]

13. Meyer KB, O'Reilly M, Michailidou K, Carlebur S, Edwards SL, French JD, et al. Fine-scale mapping of the FGFR2 breast cancer risk locus: putative functional variants differentially bind FOXA1 and E2F1. American journal of human genetics. 2013; 93:1046–1060. [PubMed: 24290378]

14. Kote-Jarai Z, Saunders EJ, Leongamornlert DA, Tymrakiewicz M, Dadaev T, Jugurnauth-Little S, et al. Fine-mapping identifies multiple prostate cancer risk loci at 5p15, one of which associates with TERT expression. Human molecular genetics. 2013; 22:2520–2528. [PubMed: 23535824]

15. Gong J, Schumacher F, Lim U, Hindorff LA, Haessler J, Buyske S, et al. Fine Mapping and Identification of BMI Loci in African Americans. American journal of human genetics. 2013; 93:661–671. [PubMed: 24094743]

16. French JD, Ghoussaini M, Edwards SL, Meyer KB, Michailidou K, Ahmed S, et al. Functional variants at the 11q13 risk locus for breast cancer regulate cyclin D1 expression through long-range enhancers. American journal of human genetics. 2013; 92:489–503. [PubMed: 23540573]

17. Hughes T, Coit P, Adler A, Yilmaz V, Aksu K, Duzgun N, et al. Identification of multiple independent susceptibility loci in the HLA region in Behcet's disease. Nature genetics. 2013; 45:319–324. [PubMed: 23396137]

18. Glubb DM, Maranian MJ, Michailidou K, Pooley KA, Meyer KB, Kar S, et al. Fine-Scale Mapping of the 5q11.2 Breast Cancer Locus Reveals at Least Three Independent Risk Variants Regulating MAP3K1. American journal of human genetics. 2015; 96:5–20. [PubMed: 25529635]

19. Ghoussaini M, Edwards SL, Michailidou K, Nord S, Cowper-Sal Lari R, Desai K, et al. Evidence that breast cancer risk at the 2q35 locus is mediated through IGFBP5 regulation. Nature communications. 2014; 4:4999.

20. Breast Cancer Association Consortium (BCAC) [Internet]. Cambridge (UK): Genetic Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, UK; Available from: http://www.srl.cam.ac.uk/consortia/bcac/.

21. Eeles RA, Kote-Jarai Z, Al Olama AA, Giles GG, Guy M, Severi G, et al. Identification of seven new prostate cancer susceptibility loci through a genome-wide association study. Nature genetics. 2009; 41:1116–1121. [PubMed: 19767753]

22. HapMap project [Internet]. Bethesda (MD): National Institutes of Health, National Library of Medicine, National Center for Biotechnology Information; Available from: http://hapmap.ncbi.nlm.nih.gov/

23. 1000 Genomes [Internet]. Bethesda (MD): National Institutes of Health, National Library of Medicine, National Center for Biotechnology Information; c2008–2012 Available from: http://browser.1000genomes.org/

24. IMPUTE v.2.2 [Internet]. Oxford (UK): Oxford University; Available from: https://mathgen.stats.ox.ac.uk/impute/impute_v2.html

25. R version 2.13.0 [Internet]. Vienna (Austria): The R Foundation; Available from: http://www.r-project.org/

26. Encyclopedia of DNA Elements at UCSC (ENCODE) [Internet]. Santa Cruz (CA): Genome Bioinformatics Group, Center for Biomolecular Science and Engineering at the University of California Santa Cruz; Available from: http://genome.ucsc.edu/ENCODE/

27. Ernst J, Kellis M. ChromHMM: automating chromatin-state discovery and characterization. Nature methods. 2012; 9:215–216. [PubMed: 22373907]

28. UCSC Genome Browser [Internet]. Santa Cruz (CA): Genome Bioinformatics Group, Center for Biomolecular Science and Engineering at the University of California Santa Cruz; Available from: http://genome.ucsc.edu

29. RegulomeDB [Internet]. Stanford (CA): Center for Genomics and Personalized Medicine at Stanford University; Available from: http://regulome.stanford.edu/

30. HaploReg v2 [Internet]. Cambridge (MA): he Broad Institute of MIT and Harvard; Available from: http://www.broadinstitute.org/mammals/haploreg/haploreg.php

31. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic acids research. 2012; 40:D930–D934. [PubMed: 22064851]

32. Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-Andre V, Sigova AA, et al. Super-enhancers in the control of cell identity and disease. Cell. 2013; 155:934–947. [PubMed: 24119843]

33. Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, et al. An atlas of active enhancers across human cell types and tissues. Nature. 2014; 507:455–461. [PubMed: 24670763]

34. The Cancer Genome Atlas (TCGA) [Internet]. Bethesda (MD): US Department of Health and Human Services, National Institutes of Health, National Cancer Institute, National Human Genome Research Institute; Available at: http://cancergenome.nih.gov/

35. CbioPortal [Internet]. New York (NY): Memorial Sloan Kettering Cancer Center, cBioPortal for Cancer Genomics; Available from: http://www.cbioportal.org/public-portal/

36. Pickrell JK, Marioni JC, Pai AA, Degner JF, Engelhardt BE, Nkadori E, et al. Understanding mechanisms underlying human gene expression variation with RNA sequencing. Nature. 2010; 464:768–772. [PubMed: 20220758]

37. Li Q, Seo JH, Stranger B, McKenna A, Pe'er I, Laframboise T, et al. Integrative eQTL-based analyses reveal the biology of breast cancer risk loci. Cell. 2013; 152:633–641. [PubMed: 23374354]

38. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature. 2012; 486:346–352. [PubMed: 22522925]

39. Shabalin AA. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. Bioinformatics. 2012; 28:1353–1358. [PubMed: 22492648]

40. Wei M, Liu B, Gu Q, Su L, Yu Y, Zhu Z. Stat6 cooperates with Sp1 in controlling breast cancer cell proliferation by modulating the expression of p21(Cip1/WAF1) and p27 (Kip1). Cellular oncology. 2013; 36:79–93.

41. Wang XB, Peng WQ, Yi ZJ, Zhu SL, Gan QH. Expression and prognostic value of transcriptional factor sp1 in breast cancer. Ai zheng = Aizheng = Chinese journal of cancer. 2007; 26:996–1000. [PubMed: 17927860]

42. Mitchell A, Dass CR, Sun LQ, Khachigian LM. Inhibition of human breast carcinoma proliferation, migration, chemoinvasion and solid tumour growth by DNAzymes targeting the zinc finger transcription factor EGR-1. Nucleic acids research. 2004; 32:3065–3069. [PubMed: 15181171]

43. Eeckhoute J, Carroll JS, Geistlinger TR, Torres-Arzayus MI, Brown M. A cell-type-specific transcriptional network required for estrogen regulation of cyclin D1 and cell cycle progression in breast cancer. Genes & development. 2006; 20:2513–2526. [PubMed: 16980581]

44. Ko M, Bandukwala HS, An J, Lamperti ED, Thompson EC, Hastie R, et al. Ten-Eleven-Translocation 2 (TET2) negatively regulates homeostasis and differentiation of hematopoietic stem cells in mice. Proceedings of the National Academy of Sciences of the United States of America. 2011; 108:14566–14571. [PubMed: 21873190]

45. Song SJ, Ito K, Ala U, Kats L, Webster K, Sun SM, et al. The oncogenic microRNA miR-22 targets the TET2 tumor suppressor to promote hematopoietic stem cell self-renewal and transformation. Cell stem cell. 2013; 13:87–101. [PubMed: 23827711]

46. Ito S, D'Alessio AC, Taranova OV, Hong K, Sowers LC, Zhang Y. Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. Nature. 2010; 466:1129–1133. [PubMed: 20639862]

47. Ko M, Huang Y, Jankowska AM, Pape UJ, Tahiliani M, Bandukwala HS, et al. Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. Nature. 2010; 468:839–843. [PubMed: 21057493]

48. Koh KP, Yabuuchi A, Rao S, Huang Y, Cunniff K, Nardone J, et al. Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. Cell stem cell. 2011; 8:200–213. [PubMed: 21295276]

49. Schoofs T, Berdel WE, Muller-Tidow C. Origins of aberrant DNA methylation in acute myeloid leukemia. Leukemia. 2014; 28:1–14. [PubMed: 23958917]

50. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics. 2010; 26:2336–2337. [PubMed: 20634204]
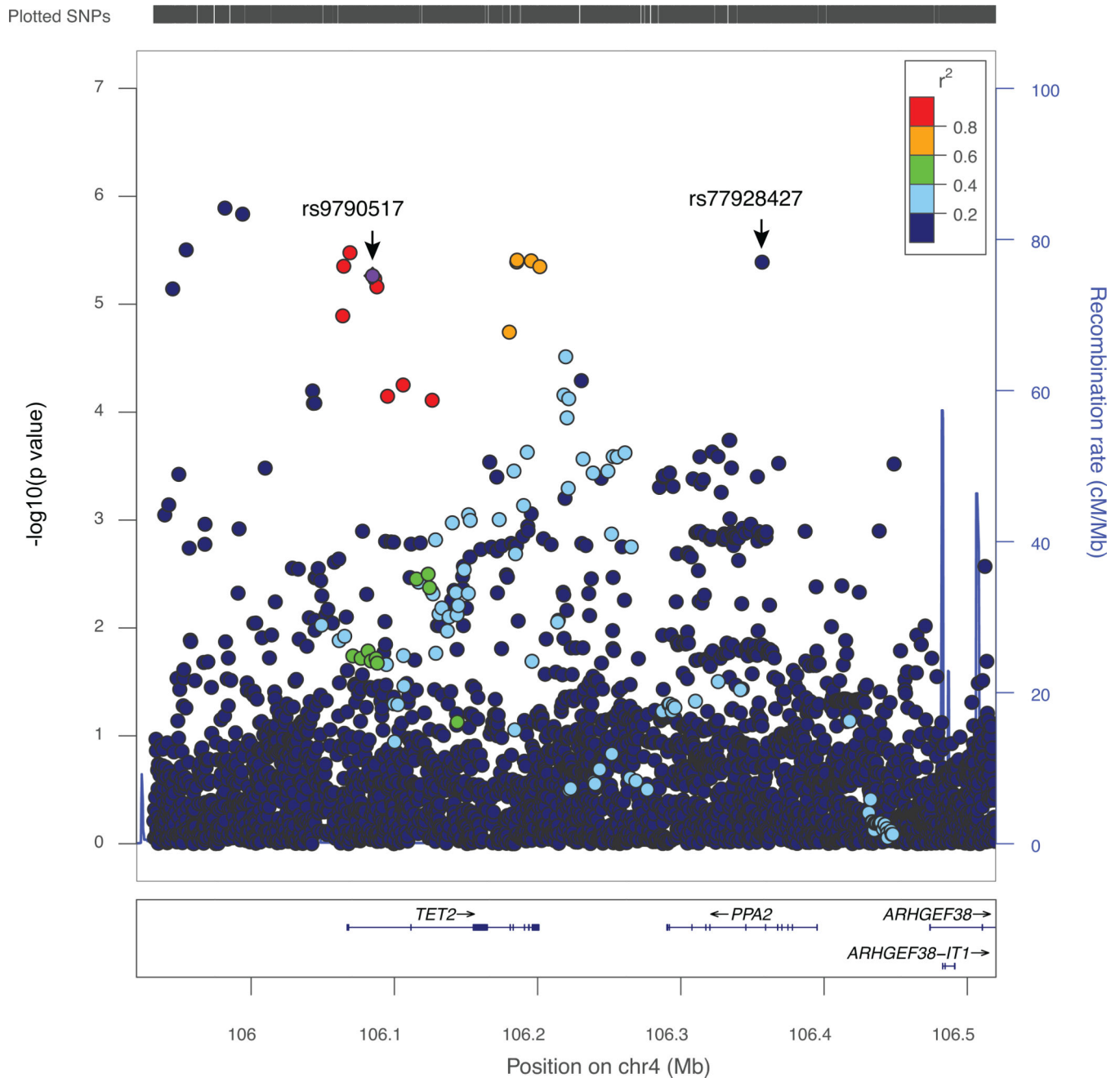
Figure 1. Regional plot of genetic variants associated with breast cancer risk at 4q24

A) The index SNP rs9790517 is plotted in diamond purple. The LD (r^2) for the index SNP with each SNP was computed based on European ancestry subjects included in the 1000 Genome Mar 2012 EUR. P values were from the single-marker analysis based on logistic regression models after adjusted for age, study sites and the first six principal components plus one additional principal component for the LMBC in analyses of data from European descendants. The plot was generated using LocusZoom (50).
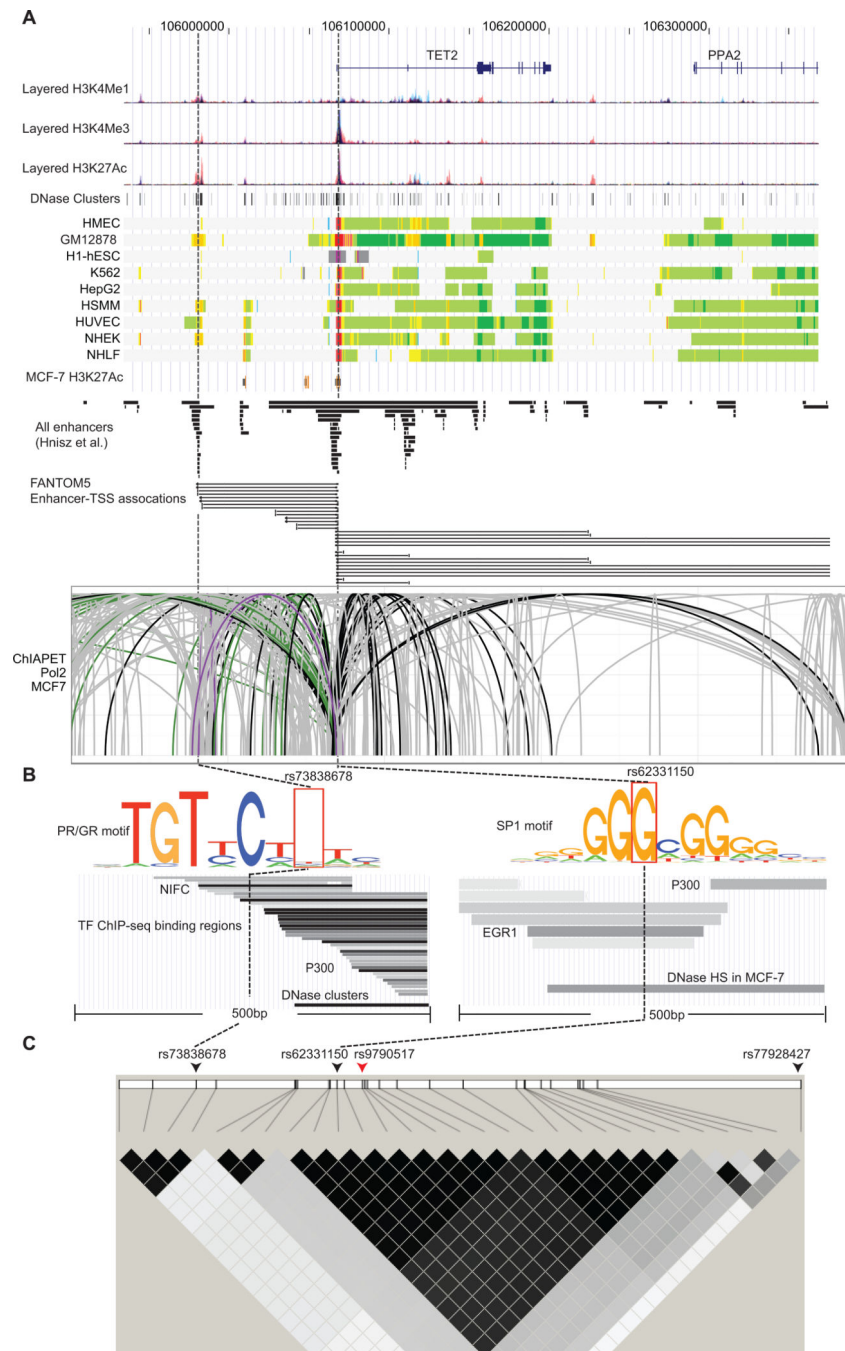
Figure 2. Functional annotation of SNPs association with breast cancer risk at 4q24

A) Epigenetic landscape at 4q24 risk locus for breast cancer. From top to bottom, RefSeq genes (*TET2* and *PPA2*), layered H3K4Me1, H3K4Me3 and H3K27Ac histone modifications, DNase clusters, annotation using chromatin states on the ENCODE cell lines, and H3K27Ac histone modification in MCF-7, predicted enhancers reported in the Hnisz et al. study, regulatory elements of enhancers associated with TSS and TSSs from the FANTOM5 project and ChIA-PET interactions in MCF-7 cell (mediated by RNA polymerase 2) between enhancers and *TET2* promoter are shown. The signals of different

layered histone modifications from the same ENCODE cell line are shown in the same color (The detailed color scheme for each ENCODE cell line described in the UCSC genome browser). The red and orange colors in chromatin states refer to active promoter and strong enhancer regions, respectively (The detailed color scheme of the chromatin states described in the previous study (27)). For ChIA-PET track, black lines represented interactions with the promoter region (−1500/+500) of *TET2*, and gray lines represent chromatin interactions that do not involve the *TET2* promoter region. Purple and green lines represent interactions within +/− 500pb of rs73838678 and rs62331150 variants, respectively. B) Epigenetic signals of two potential functional variants rs73838678 and rs62331150. From top to bottom, lanes showing that the variant mapped to TF predicted binding motifs, TF ChIP-Seq binding peaks and DNase I hypersensitivity sites. The corresponding location of the variant is indicated by dashed line. C) LD plot for breast cancer risk associated SNPs at 4q24. In the top lane, two SNPs representing independent association signals are indicated by the black arrows. The index SNP is indicated by the red arrow. In the bottom lane, two LD SNP blocks were shown based on r^2 values, which were computed based on the genotype data from the BCAC.
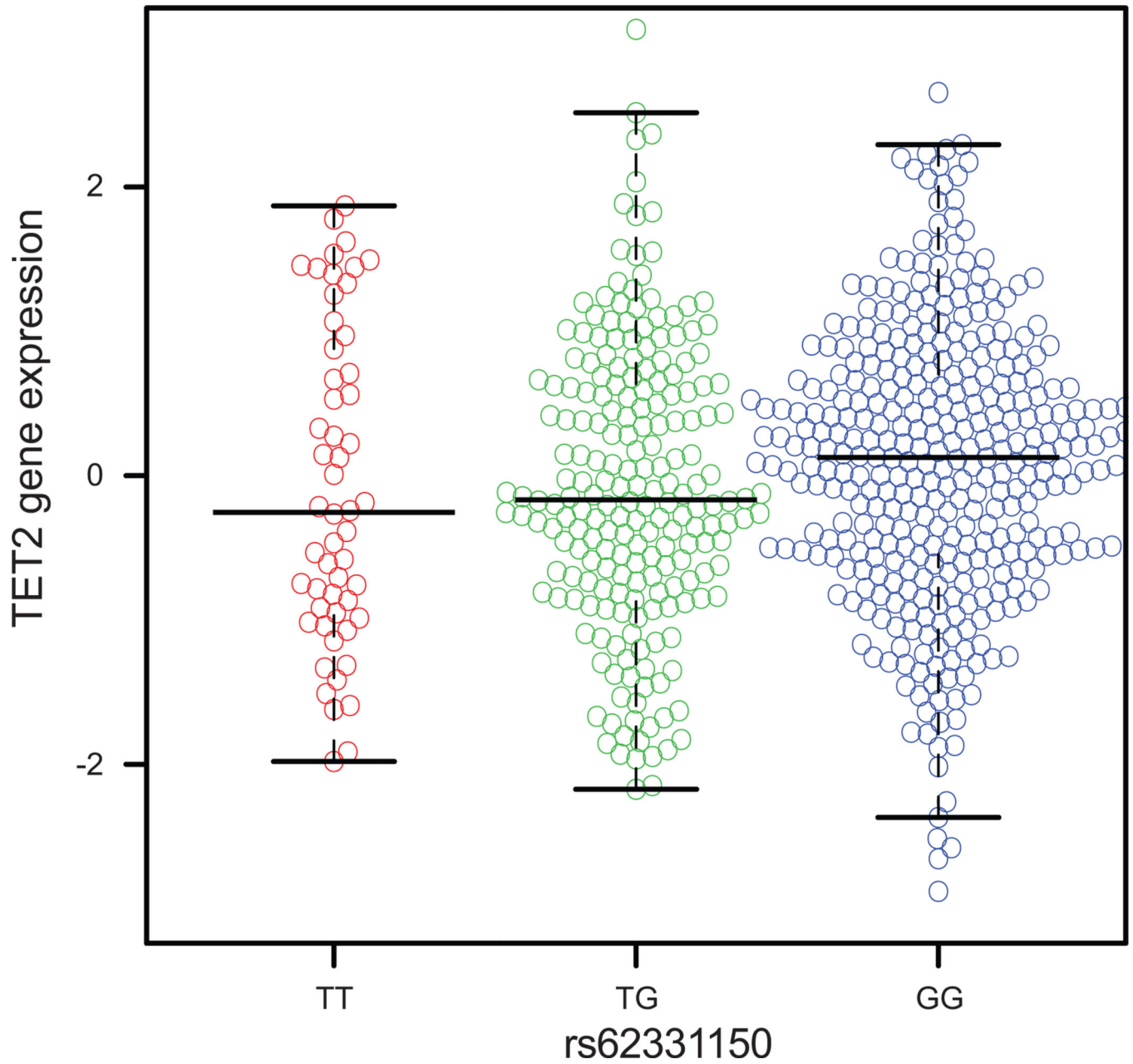
Figure 3. The association between SNP rs62331150 and *TET2* expression in breast cancer tissues from TCGA

The reference allele G of rs62331150 is significantly associated with the increased gene expression relative to the risk allele T.

Table 1

Identification of two independent association signals for overall breast cancer risk among women of European ancestry.

Signal	SNPs	Position (hg 19)	Alleles[b]	RAF	LD[c](r²)	Single Marker Analysis		Conditional Analysis	
						OR (95% CI) [d]	P trend[d]	OR (95% CI)[e]	P trend[e]
					All cases (48,155 cases and 43,612 controls)				
1[f]	**rs9790517**[a]	106084778	T/C	0.23	-	1.05(1.03–1.08)	5.44×10^{-6}	1.04(1.02–1.07)	2.51×10^{-4}
2[g]	rs77928427	106356761	A/C	0.24	0.04	1.05(1.03–1.08)	4.07×10^{-6}	1.04(1.02–1.07)	1.86×10^{-4}
					ER (+) (28,038 cases and 43,612 controls)				
1	**rs9790517**[a]	106084778	T/C	0.23	-	1.06(1.03–1.09)	1.20×10^{-5}	1.05(1.02–1.08)	2.49×10^{-4}
2	rs77928427	106356761	A/C	0.24	0.04	1.05(1.02–1.08)	1.40×10^{-4}	1.04(1.01–1.07)	3.07×10^{-3}
					ER (−) (7,786 cases and 43,612 controls)				
1	**rs9790517**[a]	106084778	T/C	0.22	-	1.04(0.99–1.08)	0.16	1.02(0.98–1.07)	0.3396
2	rs77928427	106356761	A/C	0.24	0.04	1.05(1.01–1.09)	0.03	1.04(1.00–1.09)	0.0508

Abbreviations: OR, odds ratio; CI, confidence interval; RAF, risk allele frequency

[a] Index SNP.

[b] Risk/reference allele; risk alleles are shown in bold.

[c] r² for linkage disequilibrium with the index SNP rs9790517.

[d] Adjusted for age, study and the first six and an additional PC for LMBC study.

[e] Included both top SNPs and adjusted for other top SNPs, age, study sites and the first six and an additional PC for LMBC study.

[f] A total of 23 SNPs cannot be excluded using LR < 1/100 as candidate causal variants (see Supplementary table 1).

[g] A total of 4 SNPs cannot be excluded using LR < 1/100 as candidate causal variants (see Supplementary table 1)

Table 2

Haplotype analyses of the lead SNPs in two independent signals in relation to breast cancer risk among women of European ancestry.

	rs9790517[a]	rs77928427	Signal		
			%[b]	OR(95% CI)[c]	P trend[c]
Reference	C	C	62.1	Reference (1.00)	
1	C	A	15.1	1.03(1.00–1.06)	0.06
2	T	C	13.4	1.03(1.00–1.06)	0.09
3	T	A	9.4	1.11(1.07–1.15)	2.31×10^{-8}

Abbreviations: OR, odds ratio; CI, confidence interval.

[a] Index SNP.

[b] Haplotype frequency.

[c] Adjusted for age, study and the first six PCs 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 (1,116 cases and 932 controls).

Top SNPs		Alleles[b]	Single Marker Analysis (Asian)				Single Marker Analysis (African)			
			RAF	LD (r²)[c]	OR (95% CI)[d]	P trend[d]	RAF	LD (r²)[c]	OR (95% CI)[d]	P trend[d]
Identified in women of European descent										
Signal 1	**rs9790517**[a]	T/C	0.60	-	1.00(0.95–1.06)	0.93	0.06	-	1.21(0.88–1.55)	0.28
Signal 2	rs77928427	A/C	0.06	0.01	1.02(0.91–1.12)	0.50	0.16	0	1.03(0.85–1.22)	0.86
Identified in women of non-European descent										
	rs1116764	G/A	0.66	0.13	1.10(1.04–1.16)	4.21×10^{-4}	0.89	0	1.02(0.81–1.23)	0.98
	rs79219151	C/T	NA				0.95	0	1.63(1.13–2.13)	7.44×10^{-3}
	rs112095278	C/T	NA				0.95	0	1.65(1.16–2.14)	4.13×10^{-3}
	rs144956461	A/T	NA				0.93	0	1.56(1.12–2.01)	6.73×10^{-3}

Abbreviations: OR, odds ratio; CI, confidence interval; RAF, risk allele frequency.

[a]Index SNP

[b]Risk/reference allele; risk alleles are shown in bold.

[c]r² for linkage disequilibrium with the index SNP rs9790517 in Asians and Africans, respectively.

[d]Adjusted for age, study and the first six PC and an additional PC for LMBC study.