

UC Office of the President

Recent Work

Title

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

Permalink

<https://escholarship.org/uc/item/8z2960zr>

Author

Guo, Xingyi

Publication Date

2015-11-01

DOI

10.1158/1055-9965.epi-15-0363

Peer reviewed

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

Xingyi Guo¹, Jirong Long¹, Chenjie Zeng¹, Kyriaki Michailidou², Maya Ghousaini³, Manjeet K. Bolla², Qin Wang², Roger L. Milne^{4,5}, Xiao-Ou Shu¹, Qiuyin Cai¹, Jonathan Beesley⁶, Siddhartha P. Kar³, Irene L. Andrulis^{7,8}, Hoda Anton-Culver⁹, Volker Arndt¹⁰, Matthias W. Beckmann¹¹, Alicia Beeghly-Fadiel¹, Javier Benitez^{12,13}, William Blot^{1,14}, Natalia Bogdanova¹⁵, Stig E. Bojesen^{16,17,18}, Hiltrud Brauch^{19,20,21}, Hermann Brenner^{10,21,22}, Louise Brinton²³, Annegien Broeks²⁴, Thomas Brüning²⁵, Barbara Burwinkel^{26,27}, Hui Cai¹, Sander Canisius²⁸, Jenny Chang-Claude²⁹, Ji-Yeob Choi^{30,31}, Fergus J. Couch³², Angela Cox³³, Simon S. Cross³⁴, Kamila Czene³⁵, Hatef Darabi³⁵, Peter Devilee^{36,37}, Arnaud Droit³⁸, Thilo Dörk³⁹, Peter A. Fasching^{11,40}, Olivia Fletcher⁴¹, Henrik Flyger⁴², Florentia Fostira⁴³, Valerie Gaborieau⁴⁴, Montserrat García-Closas^{41,45}, Graham G. Giles^{4,5}, Mervi Grip⁴⁶, Pascal Guénel^{47,48}, Christopher A. Haiman⁴⁹, Ute Hamann⁵⁰, Mikael Hartman^{51,52}, Antoinette Hollestelle⁵³, John L. Hopper⁵, Chia-Ni Hsiung⁵⁴, Hidemi Ito⁵⁶, Anna Jakubowska⁵⁷, Nichola Johnson⁴¹, Maria Kabisch⁵⁰, Daehee Kang^{30,31,58}, Sofia Khan⁵⁹, Julia A. Knight^{60,61}, Veli-Matti Kosma⁶², Diether Lambrechts^{63,64}, Loic Le Marchand⁶⁵, Jingmei Li³⁵, Annika Lindblom⁶⁶, Artitaya Lophatananon⁶⁷, Jan Lubinski⁵⁷, Arto Mannermaa⁶², Siranoush Manoukian⁶⁸, Sara Margolin⁶⁹, Frederik Marme^{70,71}, Keitaro Matsuo⁷², Catriona A. McLean⁷³, Alfons Meindl⁷⁴, Kenneth Muir^{68,75}, Susan L. Neuhausen⁷⁶, Heli Nevanlinna⁵⁹, Silje Nord^{77,78}, Janet E. Olson⁷⁹, Nick Orr⁸⁰, Paolo Peterlongo⁸¹, Thomas Choudary Putti⁸², Anja Rudolph²⁹, Suleeporn Sangrajrang⁸³, Elinor J. Sawyer⁸⁴, Marjanka K. Schmidt²⁴, Rita K. Schmutzler^{85,86,87,88}, Chen-Yang Shen^{89,90}, Jiajun Shi¹, Martha J. Shrubsole¹, Melissa C. Southey⁹¹, Anthony Swerdlow⁹², Soo Hwang Teo^{93,94}, Bernard Thienpont^{63,64}, Amanda Ewart Toland⁹⁵, Robert A.E.M. Tollenaar⁹⁶, Ian P.M. Tomlinson⁹⁷, Thérèse Truong^{47,48}, Chiu-chen Tseng⁴⁹, Ans van den Ouweland⁹⁸, Wanqing Wen¹, Robert Winqvist^{99,100}, Anna Wu⁴⁹, Cheng Har Yip⁹⁴, M. Pilar Zamora¹⁰¹, Ying Zheng¹⁰², Per Hall³⁵, Paul D.P. Pharoah^{2,3}, Jacques Simard³⁸, Georgia Chenevix-Trench⁶, on behalf of kConFab Investigators⁵⁵; Alison M. Dunning³, Douglas F. Easton^{2,3}, and Wei Zheng¹

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. *Cancer Epidemiol Biomarkers Prev*; 24(11); 1680–91. ©2015 AACR.

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, 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; refs. 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 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).

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 Public Health and Primary Care, University of Cambridge, Cambridge, United Kingdom. ³Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, United Kingdom. ⁴Cancer Epidemiology Centre, Cancer Council Victoria, Melbourne, Victoria, Australia. ⁵Centre for Epidemiology and Biostatistics, School of Population and Global Health, The University of Melbourne, Melbourne, Victoria, Australia. ⁶Department of Genetics, QIMR Berghofer Medical Research Institute, Brisbane, Australia. ⁷Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital, Toronto, Canada. ⁸Department of Molecular Genetics, University of Toronto, Toronto, Canada. ⁹Department of Epidemiology, University of California Irvine, Irvine, California. ¹⁰Division of Clinical Epidemiology and Aging Research, German Cancer Research Center, Heidelberg, Germany. ¹¹Department of Gynaecology and Obstetrics, University Hospital Erlangen, Friedrich-Alexander University Erlangen-Nuremberg, Erlangen, Germany. ¹²Human Cancer Genetics Program, Spanish National Cancer Research Centre, Madrid, Spain. ¹³Centro de Investigación en Red de Enfermedades Raras, Valencia, Spain. ¹⁴International Epidemiology Institute, Rockville, Maryland. ¹⁵Department of Radiation Oncology, Hannover Medical School, Hannover, Germany. ¹⁶Copenhagen General Population Study, Herlev Hospital, Herlev, Denmark. ¹⁷Department of Clinical Biochemistry, Herlev Hospital, Copenhagen University Hospital, Herlev, Denmark. ¹⁸Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. ¹⁹Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Stuttgart, Germany. ²⁰University of Tübingen, Tübingen, Germany. ²¹German Cancer Consortium, German Cancer Research Center, Heidelberg, Germany. ²²Division of Preventive Oncology, German Cancer Research Center (DKFZ), Heidelberg, Germany. ²³Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, Maryland. ²⁴Netherlands Cancer Institute, Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands. ²⁵Institute for Prevention and Occupational Medicine of the German Social Accident Insurance, Bochum, Germany. ²⁶Division of Molecular Genetic Epidemiology, German Cancer Research Center, Heidelberg, Germany. ²⁷Molecular Epidemiology Group, German Cancer Research Center, Heidelberg, Germany. ²⁸Netherlands Cancer Institute, Antoni van Leeuwenhoek Hospital, Amsterdam, the Netherlands. ²⁹Division of Cancer Epidemiology, German Cancer Research Center, Heidelberg, Germany. ³⁰Department of Biomedical Sciences, Seoul National University College of Medicine, Seoul, Korea. ³¹Cancer Research Institute, Seoul National University College of Medicine, Seoul, Korea. ³²Department of Labo-

ratory Medicine and Pathology, Mayo Clinic, Rochester, Minnesota. ³³Sheffield Cancer Research Centre, Department of Oncology, University of Sheffield, Sheffield, United Kingdom. ³⁴Academic Unit of Pathology, Department of Neuroscience, University of Sheffield, Sheffield, United Kingdom. ³⁵Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. ³⁶Department of Pathology, Leiden University Medical Center, Leiden, The Netherlands. ³⁷Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands. ³⁸Centre Hospitalier Universitaire de Québec Research Center, Laval University, Québec, Canada. ³⁹Gynaecology Research Unit, Hannover Medical School, Hannover, Germany. ⁴⁰David Geffen School of Medicine, Department of Medicine, Division of Hematology and Oncology, University of California at Los Angeles, Los Angeles, California. ⁴¹Division of Cancer Studies, Breakthrough Breast Cancer Research Centre, Institute of Cancer Research, London, United Kingdom. ⁴²Department of Breast Surgery, Herlev Hospital, Copenhagen University Hospital, Herlev, Denmark. ⁴³Molecular Diagnostics Laboratory, IRRP, National Centre for Scientific Research "Demokritos", Athens, Greece. ⁴⁴International Agency for Research on Cancer, Lyon, France. ⁴⁵Division of Genetics and Epidemiology, Institute of Cancer Research, London, United Kingdom. ⁴⁶Department of Surgery, Oulu University Hospital and University of Oulu, Oulu, Finland. ⁴⁷Environmental Epidemiology of Cancer, Center for Research in Epidemiology and Population Health, INSERM, Villejuif, France. ⁴⁸University Paris-Sud, Villejuif, France. ⁴⁹Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California. ⁵⁰Molecular Genetics of Breast Cancer, German Cancer Research Center, Heidelberg, Germany. ⁵¹Saw Swee Hock School of Public Health, National University of Singapore, Singapore, Singapore. ⁵²Department of Surgery, National University Health System, Singapore. ⁵³Department of Medical Oncology, Erasmus University Medical Center, Rotterdam, The Netherlands. ⁵⁴Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan. ⁵⁵Peter MacCallum Cancer Centre, East Melbourne, Victoria, Australia. ⁵⁶Division of Epidemiology and Prevention, Aichi Cancer Center Research Institute, Aichi, Japan. ⁵⁷Department of Genetics and Pathology, Pomeranian Medical University, Szczecin, Poland. ⁵⁸Department of Preventive Medicine, Seoul National University College of Medicine, Seoul, Korea. ⁵⁹Department of Obstetrics and Gynecology, Helsinki University Central Hospital, University of Helsinki, Helsinki, Finland. ⁶⁰Prosserman Centre for Health Research, Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital, Toronto, Ontario, Canada. ⁶¹Division of Epidemiology, Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada. ⁶²School of Medicine, Institute of Clinical Medicine, Pathology and Forensic Medicine and Cancer Center of Eastern 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 $r^2 > 0.1$ 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; 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 $r^2 > 0.1$ 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). 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 $P < 1 \times 10^{-4}$ 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. Because 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.

University of Eastern Finland, Kuopio, Finland; Imaging Center, Department of Clinical Pathology, Kuopio University Hospital, Kuopio, Finland. ⁶³Vesalius Research Center, Leuven, Belgium. ⁶⁴Laboratory for Translational Genetics, Department of Oncology, University of Leuven, Leuven, Belgium. ⁶⁵University of Hawaii Cancer Center, Honolulu, Hawaii. ⁶⁶Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden. ⁶⁷Division of Health Sciences, Warwick Medical School, Warwick University, Coventry, United Kingdom. ⁶⁸Unit of Medical Genetics, Department of Preventive and Predictive Medicine, Fondazione IRCCS Istituto Nazionale dei Tumori (INT), Milan, Italy. ⁶⁹Department of Oncology - Pathology, Karolinska Institutet, Stockholm, Sweden. ⁷⁰National Center for Tumor Diseases, University of Heidelberg, Heidelberg, Germany. ⁷¹Department of Obstetrics and Gynecology, University of Heidelberg, Heidelberg, Germany. ⁷²Department of Preventive Medicine, Kyushu University Faculty of Medical Sciences, Fukuoka, Japan. ⁷³Anatomical Pathology, The Alfred Hospital, Melbourne, Victoria, Australia. ⁷⁴Division of Gynaecology and Obstetrics, Technische Universität München, Munich, Germany. ⁷⁵Institute of Population Health, University of Manchester, Manchester, United Kingdom. ⁷⁶Beckman Research Institute of City of Hope, Duarte, California. ⁷⁷Department of Genetics, Institute for Cancer Research, Oslo University Hospital, Radiumhospitalet, Ullernchausseen, Oslo, Norway. ⁷⁸K.G. Jebsen Center for Breast Cancer Research, Institute for Clinical Medicine, Faculty of Medicine, University of Oslo, Kirkeveien, Oslo, Norway. ⁷⁹Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota. ⁸⁰Division of Breast Cancer Research, Institute of Cancer Research, London, United Kingdom; Cancer Research, Institute of Cancer Research, London, United Kingdom. ⁸¹IFOM, the FIRC Institute of Molecular Oncology, Milan, Italy. ⁸²Department of Pathology, National University Health System, Singapore. ⁸³National Cancer Institute, Bangkok, Thailand. ⁸⁴Research Oncology, Guy's Hospital, King's College London, London, United Kingdom. ⁸⁵Division of Molecular Gyneco-Oncology, Department of Gynaecology and Obstetrics, University Hospital of Cologne, Cologne, Germany. ⁸⁶Center for Integrated Oncology, University Hospital of Cologne, Cologne, Germany. ⁸⁷Center for Molecular Medicine, University Hospital of Cologne, Cologne,

Germany. ⁸⁸Center of Familial Breast and Ovarian Cancer, University Hospital of Cologne, Cologne, Germany. ⁸⁹School of Public Health, China Medical University, Taichung, Taiwan. ⁹⁰Taiwan Biobank, Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan. ⁹¹Department of Pathology, The University of Melbourne, Melbourne, Victoria, Australia. ⁹²Division of Genetics and Epidemiology and Division of Breast Cancer Research, Institute of Cancer Research, London, United Kingdom. ⁹³Cancer Research Initiatives Foundation, Subang Jaya, Selangor, Malaysia. ⁹⁴Breast Cancer Research Unit, Cancer Research Institute, University Malaya Medical Centre, Kuala Lumpur, Malaysia. ⁹⁵Department of Molecular Virology, Immunology, and Medical Genetics, Comprehensive Cancer Center, The Ohio State University, Columbus, Ohio. ⁹⁶Department of Surgical Oncology, Leiden University Medical Center, Leiden, The Netherlands. ⁹⁷Wellcome Trust Centre for Human Genetics and Oxford Biomedical Research Centre, University of Oxford, Oxford, United Kingdom. ⁹⁸Department of Clinical Genetics, Erasmus University Medical Center, Rotterdam, The Netherlands. ⁹⁹Laboratory of Cancer Genetics and Tumor Biology, Department of Clinical Chemistry, University of Oulu, Oulu, Finland. ¹⁰⁰Laboratory of Cancer Genetics and Tumor Biology, Northern Finland Laboratory Centre NordLab, Oulu, Finland. ¹⁰¹Servicio de Oncología Médica, Hospital Universitario La Paz, Madrid, Spain. ¹⁰²Shanghai Municipal Center for Disease Control and Prevention, Shanghai, PR China.

Note: Supplementary data for this article are available at Cancer Epidemiology, Biomarkers & Prevention Online (<http://cebp.aacrjournals.org/>).

Corresponding Author: Wei Zheng, Vanderbilt Epidemiology Center, Vanderbilt University School of Medicine, 2525 West End Avenue, 8th Floor, Nashville, TN 37203. Phone: 615-936-0682; Fax: 615-936-8241; E-mail: wei.zheng@vanderbilt.edu

doi: 10.1158/1055-9965.EPI-15-0363

©2015 American Association for Cancer Research.

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 Anderson 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

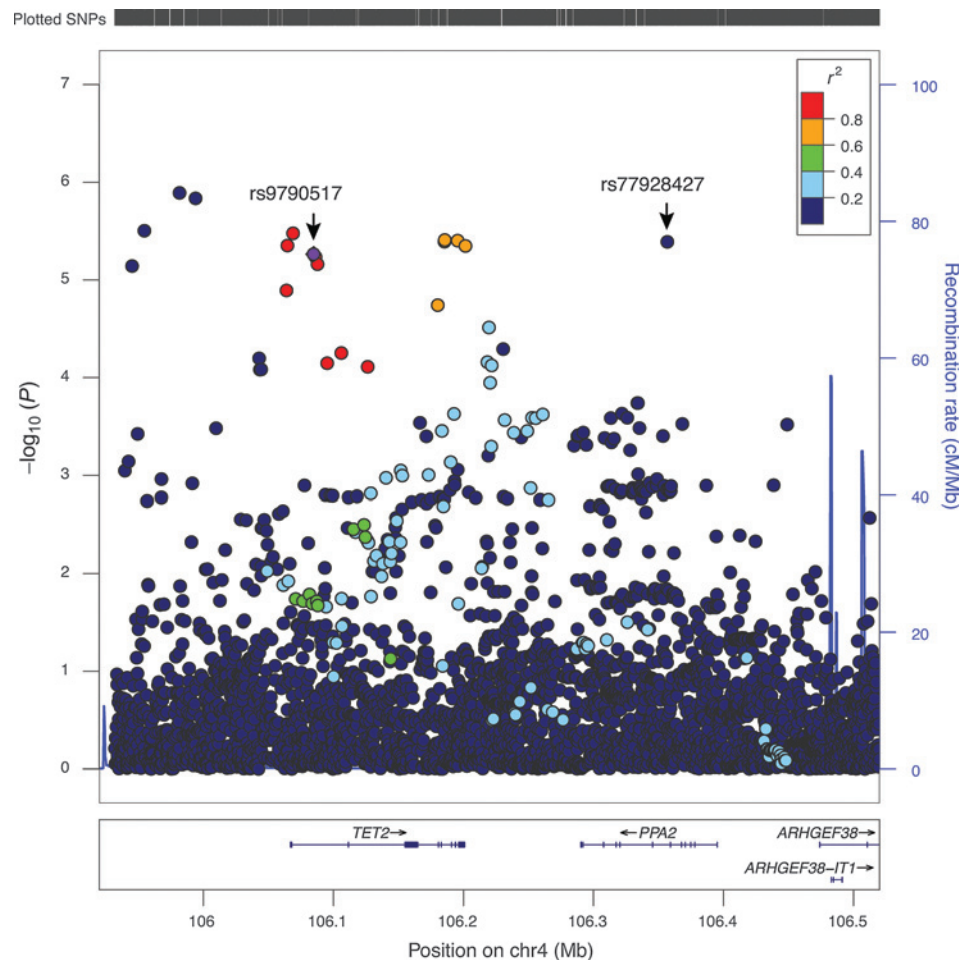
We downloaded RNA-Seq V2 data (level 3) of 1,006 breast cancer tumor tissues from The Cancer Genome Atlas (TCGA) data portal (34). DNA methylation data measured by the Illumina HumanMethylation450 BeadChip were also retrieved from TCGA level 3 data. We also downloaded level 3 SNP data genotyped using 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 CbioPortal (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 \log_2 -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

Figure 1. Regional plot of genetic variants associated with breast cancer risk at 4q24. The index SNP rs9790517 is plotted in diamond purple. The LD (r^2) for the index SNP with each SNP was computed on the basis of 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).



Guo et al.

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 (r^2)	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	106,084,778	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	106,356,761	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	106,084,778	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	106,356,761	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	106,084,778	T/C	0.22	—	1.04 (0.99-1.08)	0.16	1.02 (0.98-1.07)	0.3396
2	rs77928427	106,356,761	A/C	0.24	0.04	1.05 (1.01-1.09)	0.03	1.04 (1.00-1.09)	0.0508

Abbreviation: RAF, risk allele frequency.

^aIndex SNP.^bRisk/reference allele; risk alleles are shown in bold.^c r^2 for LD with the index SNP rs9790517.^dAdjusted for age, study, and the first six and an additional PC for LMBC study.^eIncluded both top SNPs and adjusted for other top SNPs, age, study sites, and the first six and an additional PC for LMBC study.^fA total of 23 SNPs cannot be excluded using LR < 1/100 as candidate causal variants (see Supplementary Table S1).^gA total of 4 SNPs cannot be excluded using LR < 1/100 as candidate causal variants (see Supplementary Table S1).

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 $P < 1 \times 10^{-4}$ (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; $P = 5.44 \times 10^{-6}$], 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 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 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 on the basis of 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.

Table 2. Haplotype analyses of the lead SNPs in two independent signals in relation to breast cancer risk among women of European ancestry

	Signal				
	rs9790517 ^a	rs77928427	% ^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}

^aIndex SNP.^bHaplotype frequency.^cAdjusted 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 (r^2) ^c	OR (95% CI) ^d	P_{trend} ^d	RAF	LD (r^2) ^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}

Abbreviation: RAF, risk allele frequency.

^aIndex SNP^bRisk/reference allele; risk alleles are shown in bold.^c r^2 for LD with the index SNP rs9790517 in Asians and Africans, respectively.^dAdjusted for age, study, and the first six PC and an additional PC for LMBC study.

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 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 ($r^2 = 0.98$ 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 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 Anal-

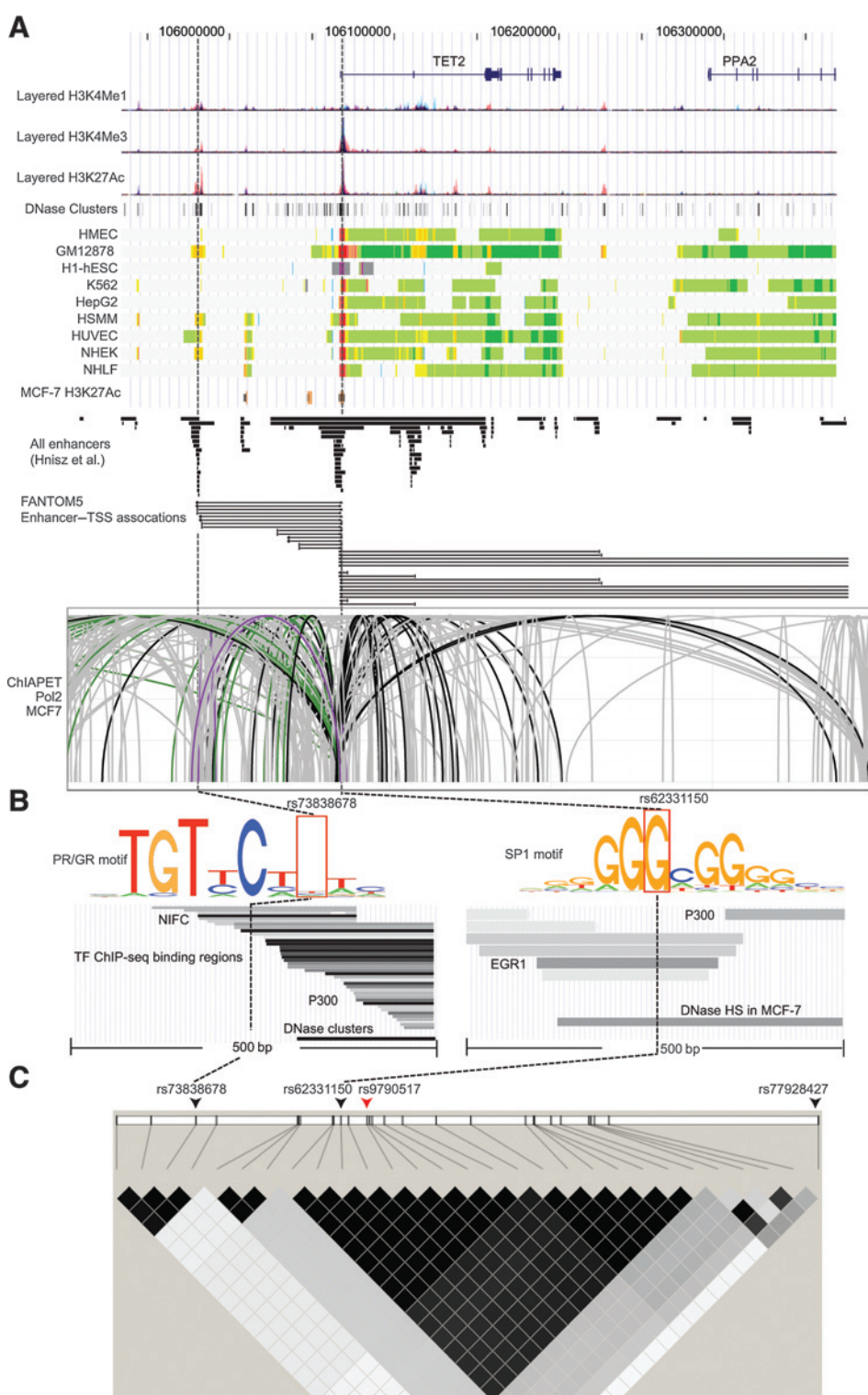
ysis of Gene Expression (CAGE) from the FANTOM5 project (23). We found that rs62331150 and rs73838678 reside in regulatory elements of enhancers associated with TSS and TSS of *TET2* in multiple cells (Fig. 2A). We also examined potential functional chromatin interactions between distal and proximal 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 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 (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.

Gene expression analyses

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 *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 (Fig. 3).

Guo et al.

**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 and colleagues 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 (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; ref. 27). For ChIA-PET track, black lines represented interactions with the promoter region ($-1,500/+500$) of *TET2*, and gray lines represent chromatin interactions that do not involve the *TET2* promoter region. Purple and green lines represent interactions within ± 500 bp 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 DHS. 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 on the basis of the genotype data from the BCAC.

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).

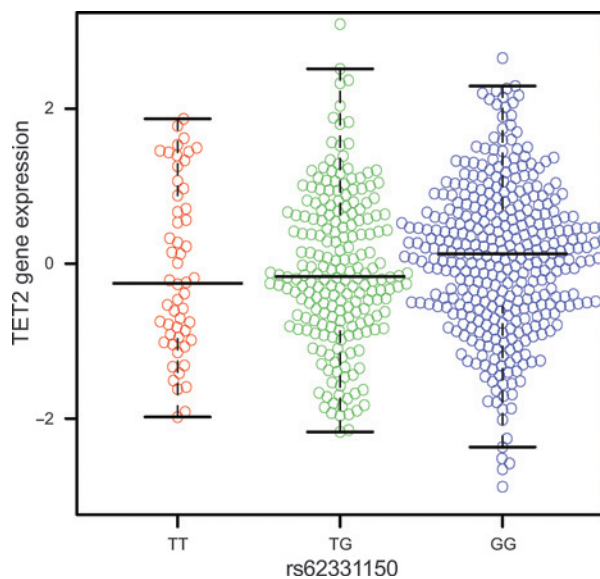


Figure 3. The association between SNP rs62331150 and *TET2* expression in breast cancer tissues from TCGA. The reference allele G of rs62331150 is significantly associated with the increased gene expression relative to the risk allele T.

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 *TET2*, 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 TF sites. In particular, it lies in the ER 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.

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

Conception and design: C. Zeng, M. Ghossaini, R.L. Milne, H. Anton-Culver, H. Brauch, T. Brüning, G.G. Giles, J.L. Hopper, J. Li, R.K. Schmutzler, C.-Y. Shen, M.J. Shrubsole, M.C. Southey, J. Simard, G. Chenevix-Trench, A.M. Dunning, D.F. Easton, W. Zheng

Development of methodology: X. Guo, C. Zeng, M. Ghossaini, J.L. Hopper, D. Lambrechts, N. Orr, R.K. Schmutzler, M.C. Southey, R.A.E.M. Tollenaar, D.F. Easton

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): C. Zeng, R.L. Milne, X.-O. Shu, I.L. Andrulis, H. Anton-Culver, V. Arndt, M.W. Beckmann, J. Benitez, W. Blot, N. Bogdanova, S.E. Bojesen, H. Brauch, H. Brenner, A. Broeks, B. Burwinkel, H. Cai, J. Chang-Claude, J.-Y. Choi, F.J. Couch, A. Cox, S.S. Cross, K. Czene, P. Devilee, T. Dörk, P.A. Fasching, O. Fletcher, H. Flyger, V. Gaborieau, M. Garcia-Closas, G.G. Giles, P. Guénel, C.A. Haiman, J.L. Hopper, H. Ito, A. Jakubowska, N. Johnson, D. Kang, S. Khan, J.A. Knight, L. Le Marchand, J. Li, A. Lindblom, A. Lophatananon, J. Lubinski, A. Mannermaa, S. Manoukian, S. Margolin, F. Marme, K. Matsuo, A. Meindl, K. Muir, S.L. Neuhausen, H. Nevanlinna, S. Nord, J.E. Olson, P. Peterlongo, T.C. Putti, A. Rudolph, S. Sangrajrang, E.J. Sawyer,

Guo et al.

M.K. Schmidt, R.K. Schmutzler, C.-Y. Shen, M.J. Shrubsole, M.C. Southey, A. Swerdlow, S. Hwang Teo, B. Thienpont, A.E. Toland, I.P.M. Tomlinson, T. Truong, A. van den Ouweland, R. Winqvist, A. Wu, C. Har Yip, M.P. Zamora, Y. Zheng, P. Hall, P.D.P. Pharoah, J. Simard, G. Chenevix-Trench, A.M. Dunning, D.F. Easton

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): X. Guo, J. Long, C. Zeng, K. Michailidou, X.-O. Shu, J. Beesley, S.P. Kar, A. Beeghly-Fadiel, S.E. Bojesen, H. Cai, S. Canisius, H. Darabi, A. Droit, U. Hamann, J.L. Hopper, C.-N. Hsiung, D. Lambrechts, C.A. McLean, S. Nord, R.K. Schmutzler, C.-Y. Shen, J. Shi, M.C. Southey, R.A.E.M. Tollenaar, C.-c. Tseng, W. Wen, P. Hall, J. Simard, D.F. Easton, W. Zheng

Writing, review, and/or revision of the manuscript: X. Guo, C. Zeng, K. Michailidou, R.L. Milne, Q. Cai, J. Beesley, I.L. Andrulis, H. Anton-Culver, V. Arndt, M.W. Beckmann, A. Beeghly-Fadiel, J. Benitez, W. Blot, S.E. Bojesen, H. Brauch, H. Brenner, L. Brinton, T. Brüning, B. Burwinkel, H. Cai, J. Chang-Claude, J.-Y. Choi, F.J. Couch, A. Cox, S.S. Cross, K. Czene, H. Darabi, A. Droit, T. Dörk, P.A. Fasching, H. Flyger, F. Fostira, M. García-Closas, G.G. Giles, U. Hamann, M. Hartman, A. Hollestelle, J.L. Hopper, A. Jakubowska, N. Johnson, M. Kabisch, S. Khan, J.A. Knight, V.-M. Kosma, D. Lambrechts, L. Le Marchand, J. Li, A. Lindblom, A. Lophatananon, A. Mannermaa, K. Muir, S.L. Neuhausen, H. Nevanlinna, J.E. Olson, P. Peterlongo, A. Rudolph, E.J. Sawyer, M.K. Schmidt, R.K. Schmutzler, M.J. Shrubsole, M.C. Southey, A. Swerdlow, B. Thienpont, A.E. Toland, A. van den Ouweland, R. Winqvist, A. Wu, M.P. Zamora, P.D.P. Pharoah, J. Simard, G. Chenevix-Trench, D.F. Easton, W. Zheng

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): C. Zeng, M.K. Bolla, Q. Wang, H. Anton-Culver, M.W. Beckmann, A. Beeghly-Fadiel, S.E. Bojesen, H. Brauch, B. Burwinkel, M. Grip, U. Hamann, A. Hollestelle, J.L. Hopper, S. Khan, J. Lubinski, A. Mannermaa, S. Manoukian, F. Marme, A. Meindl, A. Rudolph, M.K. Schmidt, M.C. Southey, A.E. Toland, R.A.E.M. Tollenaar, T. Truong, C.-c. Tseng, Y. Zheng, P. Hall, W. Zheng

Study supervision: M. Ghousaini, X.-O. Shu, H. Anton-Culver, P. Guénel, M. Hartman, J.L. Hopper, M.J. Shrubsole, M.C. Southey, D.F. Easton, W. Zheng
Other (original member and principal investigator within the German Human Genome Project): T. Brüning

Other (responsible for SGBCC, a study within BCAC): M. Hartman

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); Rosalind A. Eeles, Ali Amin Al Olama, Zsófia Kote-Jarai, Sara Benloch (PRACTICAL); Antonis Antoniou, Lesley McGuffog, Ken Offit (CIMBA); Andrew Lee, and Ed Dicks, Craig Luccarini and the staff of the Centre for Genetic Epidemiology Laboratory, Daniel C. Tessier, Francois Bacot, Daniel Vincent, Sylvie LaBoissière, Frederic Robidoux and the staff of the McGill University and Génome Québec Innovation Centre, Sune F. Nielsen and the staff of the Copenhagen DNA laboratory, Julie M. Cunningham, Sharon A. Windebank, Christopher A. Hilker, Jeffrey Meyer and the staff of Mayo Clinic Genotyping Core Facility, Maggie Angelakos, Judi Maskiell, Ellen van der Schoot (Sanquin Research); Emiel Rutgers, Senno Verhoef, Frans Hogervorst, the Thai Ministry of Public Health (MOPH); Dr. Prat Boonyawongviroj (former Permanent Secretary of MOPH), Dr. Pornthep Siriwanarungsan (Department Director-General of Disease Control), Michael Schrauder, Matthias Rübner, Sonja Oeser, Silke Landrith, Eileen Williams, Elaine Ryder-Mills, Kara Sargus, Niall McNerney, Gabrielle Colleran, Andrew Rowan, Angela Jones, Christof Sohn, Andeas Schneeweiß, Peter Bugert, the Danish Breast Cancer Group, Núria Álvarez, the CTS Steering Committee (including Leslie Bernstein, James Lacey, Sophia Wang, Huiyan Ma, Yani Lu and Jessica Clague DeHart at the Beckman Research Institute of the City of Hope; Dennis Deapen, Rich Pinder, Eunjung Lee and Fred Schumacher at the University of Southern California; Pam Horn-Ross, Peggy Reynolds and David Nelson at the Cancer Prevention Institute of California; and Hannah Park at the University of California Irvine), Hartwig Ziegler, Sonja Wolf, Volker Hermann, The GENICA network [Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Stuttgart, and University of Tübingen, Germany; (HB, Wing-Yee Lo, Christina Justenhoven), German Cancer Consortium (DKTK) and German Cancer Research Center (DKFZ) [HB] Department of Internal Medicine, Evangelische Kliniken Bonn gGmbH, Johanniter Krankenhaus, Bonn, Germany

(Yon-Dschun Ko, Christian Baisch), Institute of Pathology, University of Bonn, Germany (Hans-Peter Fischer), Molecular Genetics of Breast Cancer, Deutsches Krebsforschungszentrum (DKFZ) Heidelberg, Germany (Ute Hamann), Institute for Prevention and Occupational Medicine of the German Social Accident Insurance, Institute of the Ruhr University Bochum (IPA), Germany (TB, Beate Pesch, Sylvia Rabstein, Anne Lotz), Institute of Occupational Medicine and Maritime Medicine, University Medical Center Hamburg-Eppendorf, Germany (Volker Harth)], Tuomas Heikkinen, Irja Erkkilä, Kirsimari Aaltonen, Karl von Smitten, Natalia Antonenkova, Peter Hillemanns, Hans Christiansen, Eija Myöhänen, Helena Kemiläinen, Heather Thorne, Eveline Niedermayr, the AOCs Management Group (D Bowtell, G Chenevix-Trench, A deFazio, D Gertig, A Green, P Webb), the ACS Management Group (A. Green, P. Parsons, N. Hayward, P. Webb, D. Whiteman), the LAABC data collection team, especially Annie Fung and June Yashiki, Gilian Peuteman, Dominiek Smeets, Thomas Van Brussel, Kathleen Corthouts, Nadia Obi, Judith Heinz, Sabine Behrens, Ursula Eilber, Muhabbet Celik, Til Olchers, Siranoush Manoukian, Bernard Peissel, Giulietta Scuvera, Daniela Zaffaroni, Bernardo Bonanni, Irene Feroce, Angela Maniscalco, Alessandra Rossi, Loris Bernard, the personnel of the Cogentech Cancer Genetic Test Laboratory, The Mayo Clinic Breast Cancer Patient Registry, Martine Tranchant, Marie-France Valois, Annie Turgeon, Lea Heguy, Phuah Sze Yee, Peter Kang, Kang In Nee, Shivaani Mariapun, Yoon Sook-Yee, Daphne Lee, Teh Yew Ching, Nur Aishah Mohd Taib, Meeri Otsukka, Kari Mononen, Teresa Selander, Nayana Weerasooriya, OFBCR staff, E. Krol-Warmerdam, J. Molenaar, J. Blom, Louise Brinton, Neonila Szeszenia-Dabrowska, Beata Peplonska, Witold Zatonski, Pei Chao, Michael Stagner, Petra Bos, Jannet Blom, Ellen Crepin, Anja Nieuwlaat, Annette Heemskerck, the Erasmus MC Family Cancer Clinic, Sue Higham, Simon Cross, Helen Cramp, Dan Connley, Sabapathy Balasubramanian, Ian Brock, The Eastern Cancer Registration and Information Centre, the SEARCH and EPIC teams, Michael Kerin, Nicola Miller, Niall McNerney, Gabrielle Colleran (BIGGS), Pierre Kerbrat; Patrick Arveux; Romuald Le Scodan; Yves Raoul; Pierre Laurent-Puig; Claire Mulot (CECILE), Christa Stegmaier and Katja Butterbach (ESTHER), Natalia Antonenkova, Peter Hillemanns, Hans Christiansen and Johann H. Karstens (HMBCS), Gilian Peuteman, Dominiek Smeets, Thomas Van Brussel and Kathleen Corthouts (LMBC), Dieter Flesch-Janys, Petra Seibold, Judith Heinz, Nadia Obi, Alina Vrieling, Sabine Behrens, Ursula Eilber, Muhabbet Celik, Til Olchers and Stefan Nickels (MARIE). We wish to thank Paolo Radice, Bernard Peissel, and Daniela Zaffaroni of the Fondazione IRCCS Istituto Nazionale dei Tumori (INT); Bernardo Bonanni, Monica Barile and Irene Feroce of the Istituto Europeo di Oncologia (IEO) and Loris Bernard and the personnel of the Cogentech Cancer Genetic Test Laboratory. Cancer Council Victoria acknowledges the Traditional Owners of the land and waters throughout Victoria and pays respect to them, their culture and their Elders past, present, and future. The authors thank Martine Tranchant (Cancer Genomics Laboratory, CHU de Québec Research Center), Marie-France Valois, Annie Turgeon and Lea Heguy (McGill University Health Center, Royal Victoria Hospital; McGill University) for DNA extraction, sample management, and skillful technical assistance. J.S. is Chairholder of the Canada Research Chair in Oncogenetics. OBCC thanks Katri Pylkäs, Arja Jukkola-Vuorinen, Saila Kauppila, Kari Mononen, and Meeri Otsukka for data collection and sample preparation. Craig Luccarini, Don Conroy, Caroline Baynes, Kimberley Chua, the Ohio State University Human Genetics Sample Bank and Robert Pilarski. Data on SCCS cancer cases used in this publication were provided by the: Alabama Statewide Cancer Registry; Kentucky Cancer Registry, Lexington, KY; Tennessee Department of Health, Office of Cancer Surveillance; Florida Cancer Data System; North Carolina Central Cancer Registry, North Carolina Division of Public Health; Georgia Comprehensive Cancer Registry; Louisiana Tumor Registry; Mississippi Cancer Registry; South Carolina Central Cancer Registry; Virginia Department of Health, Virginia Cancer Registry; Arkansas Department of Health, Cancer Registry.

Grant Support

The work conducted for this project at Vanderbilt Epidemiology Center is supported in part by NIH grant R37CA070867 and endowment funds for 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 Union COST programme (BM0606). Genotyping of the iCOGS array was funded by the European Union (HEALTH-F2-2009-223175), Cancer Research UK (C8197/A16565 and C1287/A10710), the

Canadian Institutes of Health Research for the "CIHR Team in Familial Risks of Breast Cancer" program and the Ministry of Economic Development, Innovation and Export Trade of Quebec (PSR-SIIRI-701). Additional support for the iCOGS infrastructure was provided by the NIH (CA128978) and Post-Cancer GWAS initiative (1U19 CA148537, 1U19 CA148065 and 1U19 CA148112—the GAME-ON initiative), the Department of Defence (W81XWH-10-1-0341), Komen Foundation for the Cure, the Breast Cancer Research Foundation, and the Ovarian Cancer Research Fund. This work was supported by grant UM1 CA164920 from the National Cancer Institute. The ABCFS was also supported by the National Health and Medical Research Council of Australia, the New South Wales Cancer Council, the Victorian Health Promotion Foundation (Australia), and the Victorian Breast Cancer Research Consortium. J.L.H. is a National Health and Medical Research Council (NHMRC) Senior Principal Research Fellow and M.C.S. is a NHMRC Senior Research Fellow. The OFBCR work was also supported by the Canadian Institutes of Health Research "CIHR Team in Familial Risks of Breast Cancer" program. The ABCS was funded by the Dutch Cancer Society Grant no. NKI2007-3839 and NKI2009-4363. The ACP study is funded by the Breast Cancer Research Trust, UK. The work of the BBCC was partly funded by ELAN-Programme of the University Hospital of Erlangen. The BBCC is funded by Cancer Research UK and Breakthrough Breast Cancer and acknowledges NHS funding to the NIHR Biomedical Research Centre, and the National Cancer Research Network (NCRN). E.J. Sawyer is supported by NIHR Comprehensive Biomedical Research Centre, Guy's & St. Thomas' NHS Foundation Trust in partnership with King's College London, UK. Core funding to the Wellcome Trust Centre for Human Genetics was provided by the Wellcome Trust (090532/Z/09/Z). I.P.M. Tomlison is supported by the Oxford Biomedical Research Centre. The BSUCH study was supported by the Dietmar-Hopp Foundation, the Helmholtz Society and the German Cancer Research Center (DKFZ). The CECILE study was funded by Fondation de France, Institut National du Cancer (INCa), Ligue Nationale contre le Cancer, Agence Nationale de Sécurité Sanitaire (ANSES), Agence Nationale de la Recherche (ANR). The CGPS was supported by the Chief Physician Johan Boserup and Lise Boserup Fund, the Danish Medical Research Council, and Herlev Hospital. The CNIO-BCS was supported by the Genome Spain Foundation, the Red Temática de Investigación Cooperativa en Cáncer and grants from the Asociación Española Contra el Cáncer and the Fondo de Investigación Sanitario (PI11/00923 and PI081120). The Human Genotyping-CEGEN Unit, CNIO is supported by the Instituto de Salud Carlos III. Divyansh Agarwal was supported by a Fellowship from the Michael Manzella Foundation (MMF) and was a participant in the CNIO Summer Training Program. The CTS was initially supported by the California Breast Cancer Act of 1993 and the California Breast Cancer Research Fund (contract 97-10500) and is currently funded through the National Institutes of Health (R01 CA77398). Collection of cancer incidence data was supported by the California Department of Public Health as part of the statewide cancer reporting program mandated by California Health and Safety Code Section 103885. H. Anton-Culver receives support from the Lon V Smith Foundation (LVS39420). The ESTHER study was supported by a grant from the Baden Württemberg Ministry of Science, Research and Arts. Additional cases were recruited in the context of the VERDI study, which was supported by a grant from the German Cancer Aid (Deutsche Krebshilfe). The GENICA was funded by the Federal Ministry of Education and Research (BMBF) Germany grants 01KW9975/5, 01KW9976/8, 01KW9977/0 and 01KW0114, the Robert Bosch Foundation, Stuttgart, Deutsches Krebsforschungszentrum (DKFZ), Heidelberg, Institute for Prevention and Occupational Medicine of the German Social Accident Insurance, Institute of the Ruhr University Bochum (IPA), as well as the Department of Internal Medicine, Evangelische Kliniken Bonn gGmbH, Johanner Krankenhaus Bonn, Germany. The HEBCS was supported by the Helsinki University Central Hospital Research Fund, Academy of Finland (266528), the Finnish Cancer Society, The Nordic Cancer Union and the Sigrid Juselius Foundation. The HERPACC was supported by a Grant-in-Aid for Scientific Research on Priority Areas from the Ministry of Education, Science, Sports, Culture and Technology of Japan, by a Grant-in-Aid for the Third Term Comprehensive 10-Year Strategy for Cancer Control from Ministry Health, Labour and Welfare of Japan, by a research grant from Takeda Science Foundation, by Health and Labour Sciences Research Grants for Research on Applying Health Technology from Ministry Health, Labour and Welfare of Japan and by National Cancer Center Research and Development Fund. The HMBCS was supported by the Rudolf Bartling Foundation. Financial support for KARBAC was provided through the regional agreement on medical training and clinical research (ALF) between Stockholm County Council and Karolinska Institutet, the Stockholm Cancer Foundation and the Swedish Cancer Society.

The KBCP was financially supported by the special Government Funding (EVO) of Kuopio University Hospital grants, Cancer Fund of North Savo, the Finnish Cancer Organizations, the Academy of Finland, and by the strategic funding of the University of Eastern Finland. kConFab is supported by grants from the National Breast Cancer Foundation, the NHMRC, the Queensland Cancer Fund, the Cancer Councils of New South Wales, Victoria, Tasmania, and South Australia, and the Cancer Foundation of Western Australia. The kConFab Clinical Follow Up Study was funded by the NHMRC (145684, 288704, 454508). Financial support for the AOCS was provided by the United States Army Medical Research and Materiel Command (DAMD17-01-1-0729), the Cancer Council of Tasmania and Cancer Foundation of Western Australia, and the NHMRC (199600). G.C.T. and P.W. are supported by the NHMRC. LAABC is supported by grants (1RB-0287, 3PB-0102, 5PB-0018 and 10PB-0098) from the California Breast Cancer Research Program. Incident breast cancer cases were collected by the USC Cancer Surveillance Program (CSP), which is supported under subcontract by the California Department of Health. The CSP is also part of the National Cancer Institute's Division of Cancer Prevention and Control Surveillance, Epidemiology, and End Results Program, under contract number N01CN25403. LMBC is supported by the 'Stichting tegen Kanker' (232-2008 and 196-2010). D. Lambrechts is supported by the FWO and the KULPFV/10/016-SymBioSysII and by an ERC consolidator grant. The MARIE study was supported by the Deutsche Krebshilfe e.V. (70-2892-BR I, 106332, 108253, 108419), the Hamburg Cancer Society, the German Cancer Research Center, and the Federal Ministry of Education and Research (BMBF) Germany [01KH0402]. MBCSG is supported by grants from the Italian Association for Cancer Research (AIRC) and by funds from the Italian citizens who allocated a 5/1000 share of their tax payment in support of the Fondazione IRCCS Istituto Nazionale Tumori, according to Italian laws (INT-Institutional strategic projects "5 × 1000"). The MCBCS was supported by the NIH grants (CA122340, CA128978) and a Specialized Program of Research Excellence (SPORE) in Breast Cancer (CA116201), the Breast Cancer Research Foundation, and a generous gift from the David F. and Margaret T. Grohne Family Foundation and the Ting Tsung and Wei Fong Chao Foundation. MCCS cohort recruitment was funded by VicHealth and Cancer Council Victoria. The MCCS was further supported by Australian NHMRC grants 209057, 251553, and 504711 and by infrastructure provided by Cancer Council Victoria. The MEC was supported by NIH grants CA63464, CA54281, CA098758, and CA132839. The work of MTLGBCS was supported by the Quebec Breast Cancer Foundation, the Canadian Institutes of Health Research for the "CIHR Team in Familial Risks of Breast Cancer" program—grant CRN-87521 and the Ministry of Economic Development, Innovation and Export Trade—grant PSR-SIIRI-701. MYBRCA is funded by research grants from the Malaysian Ministry of Science, Technology and Innovation (MOSTI), Malaysian Ministry of Higher Education (UM.C/HIR/MOHE/06), and Cancer Research Initiatives Foundation (CARIF). Additional controls were recruited by the Singapore Eye Research Institute, which was supported by a grant from the Biomedical Research Council (BMRC08/1/35/19<tel:08/1/35/19>550), Singapore and the National Medical Research Council, Singapore (NMRC/CG/SERI/2010). The NBCS was supported by grants from the Norwegian Research council (155218/V40, 175240/S10 to Anne-Lise Borresen-Dale, FUGE-NFR 181600/V11 to V.-M. Kosma and a Swizz Bridge Award to Anne-Lise Borresen-Dale). S. Nord has a carrier grant from the Health Region South East (HSØ, grant nr 2014061). The NBHS was supported by NIH grant R01CA100374. Biological sample preparation was conducted the Survey and Biospecimen Shared Resource, which is supported by P30 CA68485. OBSC was supported by the Academy of Finland (grant number 250083, 122715 and Center of Excellence grant number 251314), the Finnish Cancer Foundation, the Sigrid Juselius Foundation, the University of Oulu, the University of Oulu Support Foundation, and the special Governmental EVO funds for Oulu University Hospital-based research activities. The ORIGO study was supported by the Dutch Cancer Society (RUL 1997-1505) and the Biobanking and Biomolecular Resources Research Infrastructure (BBMRI-NL CP16). The PBCS was funded by Intramural Research Funds of the National Cancer Institute, Department of Health and Human Services, USA. pKARMA is a combination of the KARMA and LIBRO-1 studies. KARMA was supported by Mårit and Hans Rausings Initiative Against Breast Cancer. KARMA and LIBRO-1 were supported the Cancer Risk Prediction Center (CRiSP; www.crispcenter.org), a Linnaeus Centre (Contract ID 70867902) financed by the Swedish Research Council. The RBCS was funded by the Dutch Cancer Society (DDHK 2004-3124, DDHK 2009-4318). SASBAC was supported by funding from the Agency for Science, Technology, and Research of Singapore (A*STAR), the US National Institute of Health (NIH), and the Susan G. Komen Breast Cancer Foundation. KC was

Guo et al.

financed by the Swedish Cancer Society (5128-B07-01PAF). The SBCGS was supported primarily by NIH grants R01CA64277, R01CA148667, and R37CA70867. Biological sample preparation was conducted the Survey and Biospecimen Shared Resource, which is supported by P30 CA68485. The SBCS was supported by Yorkshire Cancer Research S305PA, 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 participates 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 supported 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 (2012-0000347). SGBCC is funded by the NUS 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: 05/1/21/19/425. SKKDKFZS is supported by the DKFZ. The SZBCS was supported by Grant PBZ_KBN_122/P05/2004. 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), the Breast Cancer 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 Chao 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 & Technology: ARISTEIA. The TWBCS is supported by the Institute of Biomedical Sciences, Academia Sinica and the National Science Council, Taiwan. The UKBGS is funded by Breakthrough Breast Cancer and the Institute of Cancer Research (ICR). ICR acknowledges NHS funding to the NIHR Biomedical Research Centre.

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

References

- 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. *Nat Genet* 2013;45:353–61, 61e1–2.
- iCOGS [Internet]. Cambridge (UK): center for cancer genetic epidemiology, department of public health and primary care/department of oncology, University of Cambridge. [cited 2015 Sept. 21]. Available from: <http://ccge.medschl.cam.ac.uk/research/consortia/icogs/>.
- Tefferi A, Lim KH, Levine R. Mutation in TET2 in myeloid cancers. *N Engl J Med* 2009;361:1117; author reply 1118.
- Delhommeau F, Dupont S, Della Valle V, James C, Trannoy S, Masse A, et al. Mutation in TET2 in myeloid cancers. *N Engl J Med* 2009;360:2289–301.
- Cancer Genome Atlas Research N. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med* 2013;368:2059–74.
- Figuerola 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;18:553–67.
- Seshagiri S, Stawiski EW, Durinck S, Modrusan Z, Storm EE, Conboy CB, et al. Recurrent R-spondin fusions in colon cancer. *Nature* 2012;488:660–4.
- 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;486:400–4.
- Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res* 2012;22:1790–7.
- 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. *Nat Genet* 2012;44:1137–41.
- 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. *Nat Genet* 2011;43:1193–201.
- 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. *Nat Genet* 2013;45:371–84, 84.e1–2.
- 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. *Am J Hum Genet* 2013;93:1046–60.
- 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. *Hum Mol Genet* 2013;22:2520–8.
- Gong J, Schumacher F, Lim U, Hindorf LA, Haessler J, Buyske S, et al. Fine Mapping and Identification of BMI Loci in African Americans. *Am J Hum Genet* 2013;93:661–71.
- 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. *Am J Hum Genet* 2013;92:489–503.
- 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. *Nat Genet* 2013;45:319–24.
- 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. *Am J Hum Genet* 2015;96:5–20.
- 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. *Nat Commun* 2014;4:4999.
- Breast Cancer Association Consortium (BCAC) [Internet]. Cambridge (UK): genetic epidemiology unit, department of public health and primary care, University of Cambridge, UK. [cited 2015 Sept. 21]. Available from: <http://www.srl.cam.ac.uk/consortia/bcac/>.
- Eeles RA, Kote-Jarai Z, AlOlama AA, Giles GG, Guy M, Severi G, et al. Identification of seven new prostate cancer susceptibility loci through a genome-wide association study. *Nat Genet* 2009;41:1116–21.
- HapMap project [Internet]. Bethesda (MD): National Institutes of Health, National Library of Medicine, National Center for Biotechnology Information. [cited 2015 Sept. 21]. Available from: <http://hapmap.ncbi.nlm.nih.gov/>.
- 1000 Genomes [Internet]. Bethesda (MD): national institutes of health, national library of medicine, national center for biotechnology information. c2008–2012 [cited 2015 Sept. 21]. Available from: <http://browser.1000genomes.org/>.
- IMPUTE v.2.2 [Internet]. Oxford (UK): Oxford University. [cited 2015 Sept. 21]. Available from: https://mathgen.stats.ox.ac.uk/impute/impute_v2.html.
- R version 2.13.0 [Internet]. Vienna (Austria): The R Foundation. [cited 2015 Sept. 21]. Available from: <http://www.r-project.org/>.
- 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. [cited 2015 Sept. 21]. Available from: <http://genome.ucsc.edu/ENCODE/>.
- Ernst J, Kellis M. ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods* 2012;9:215–6.

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>.
29. RegulomeDB [Internet]. Stanford (CA): Center for Genomics and Personalized Medicine at Stanford University. [cited 2015 Sept. 21]. Available from: <http://regulome.stanford.edu/>.
30. HaploReg v2 [Internet]. Cambridge (MA):The Broad Institute of MIT and Harvard. [cited 2015 Sept. 21]. 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 Res* 2012;40:D930–4.
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;155:934–47.
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;507:455–61.
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;464:768–72.
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;152:633–41.
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;486:346–52.
39. Shabalin AA. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics* 2012;28:1353–8.
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). *Cell Oncol* 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* 2007;26:996–1000.
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 Res* 2004;32:3065–9.
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 Dev* 2006;20:2513–26.
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. *Proc Natl Acad Sci U S A* 2011;108:14566–71.
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.
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;466:1129–33.
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;468:839–43.
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;8:200–13.
49. Schoofs T, Berdel WE, Muller-Tidow C. Origins of aberrant DNA methylation in acute myeloid leukemia. *Leukemia* 2014;28:1–14.
50. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Glied TP, et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* 2010;26:2336–7.

Cancer Epidemiology, Biomarkers & Prevention

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

Xingyi Guo, Jirong Long, Chenjie Zeng, et al.

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](https://doi.org/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, 3 of which you can access for free at:
<http://cebp.aacrjournals.org/content/24/11/1680.full#ref-list-1>

Citing articles This article has been cited by 2 HighWire-hosted articles. Access the articles at:
<http://cebp.aacrjournals.org/content/24/11/1680.full#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, use this link
<http://cebp.aacrjournals.org/content/24/11/1680>.
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.