Fine-mapping of a novel premenopausal breast cancer susceptibility locus at Chr4q31.22 in Caucasian women and validation in African and Chinese women

Mahalakshmi Kumaran[1], Sunita Ghosh[2], Anil A. Joy[2], John R. Mackey[2], Carol E. Cass[2], Wei Zheng[3], Yutaka Yasui[4] and Sambasivarao Damaraju [ID][1,5]

[1]Department of Laboratory Medicine & Pathology, University of Alberta, Edmonton, AB, Canada
[2]Department of Oncology, University of Alberta, Edmonton, AB, Canada
[3]Department of Medicine, Vanderbilt University, Nashville, TN
[4]Department of Epidemiology & Cancer Control, St. Jude Children's Research Hospital, Memphis, TN
[5]Cross Cancer Institute, Alberta Health Services, Edmonton, AB, Canada

We previously identified a novel breast cancer susceptibility variant on chromosome 4q31.22 locus (rs1429142) conferring risk among women of European ancestry. Here, we report replication of findings, validation of the variant in diverse populations and fine-mapping of the associated locus in Caucasian population. The SNP rs1429142 (C/T, minor allele frequency 18%) showed association for the overall breast cancer risk in Stages 1–4 ($n = 4,331$ cases/4271 controls; $p = 4.35 \times 10^{-8}$; odds ratio, $OR_{C\text{-allele}}$,1.25), and an elevated risk among premenopausal women ($n = 1,503$ cases/4271 controls; $p = 5.81 \times 10^{-10}$; $OR_{C\text{-allele}}$ 1.40) in European populations. SNP rs1429142 was associated with premenopausal breast cancer risk in women of African (T/C; p-value 1.45×10^{-02}; $OR_{C\text{-allele}}$ 1.2) but not from Chinese ancestry. Fine-mapping of the locus revealed several potential causal variants which are present within a single association signal, revealed from the conditional regression analysis. Functional annotation of the potential causal variants revealed three putative SNPs rs1366691, rs1429139 and rs7667633 with active enhancer functions inferred based on histone marks, DNase hypersensitive sites in breast cell line data. These putative variants were bound by transcription factors (C-FOS, STAT1/3 and POL2/3) with known roles in inflammatory pathways. Furthermore, Hi-C data revealed several short-range interactions in the fine-mapped locus harboring the putative variants. The fine mapped locus was predicted to be within a single topologically associated domain, potentially facilitating enhancer–promoter interactions possibly leading to the regulation of nearby genes.

Introduction

Breast cancer is the most commonly diagnosed cancer among women worldwide.[1,2] Genome-wide association studies (GWAS) in diverse populations have identified to date approximately 170 common low penetrance variants associated with breast cancer risk.[3] GWAS identified trait-associated SNPs are often in linkage disequilibrium (LD) with putative causal variant(s) contributing to the phenotype.[4] Therefore, it is necessary to comprehensively investigate GWAS identified loci by fine-scale mapping to identify putative causal variants and characterize their functional significance.[5] While fine-mapping approaches are well described in the literature, it is challenging to elucidate

Cancer Epidemiology

What's new?

To date, genome-wide association studies have addressed familial or postmenopausal breast cancer susceptibility variants. However, genetic predisposition for sporadic premenopausal breast cancer risk is unknown. This study reports a novel variant (at 4q31.22) associated with elevated risk for premenopausal breast cancer among women from European and African ancestry. The fine-mapped locus was predicted to be within a single topologically associated domain, potentially facilitating enhancer-promoter interactions leading to the regulation of nearby genes.

the functional relevance of GWAS SNPs, which are predominantly from noncoding regions conferring potential gene regulatory roles. Thus far, 15 breast cancer associated GWAS variants have been fine-mapped and characterized for putative biological roles.[6–20]

We previously reported six putative risk variants[21] for breast cancer from a GWAS in European populations (Alberta, Canada), hereafter described also as Caucasian populations, of which four SNPs were from different chromosomes showing association with sporadic (age of disease onset, >40 Years of age and no family history) breast cancer risk. One SNP rs1429142 on Chr4q31.22, showed consistent associations in three independent cohorts for overall risk (Stages 1–3, $p = 1.5 \times 10^{-7}$ adjusted for body mass index [BMI]; OR 1.28). Analysis based on menopausal status (Stages 1–3) revealed that SNP rs1429142 had an elevated risk for breast cancer among premenopausal women.[22] (BMI adjusted p-value of 6.22×10^{-10} and $OR_{\text{per-allele}}$ of 1.49) compared to postmenopausal women (BMI adjusted p-value of 7.79×10^{-03} and $OR_{\text{per-allele}}$ of 1.17) with a p-value of heterogeneity (p-het) $<10^{-03}$. Of the remaining five SNPs, three are from chromosome 19 and are in LD (*ZNF577* locus), and one each from chromosomes 5 (*ROPN1L* locus) and 16 (*C16orf61* locus).[21,22] SNPs from the loci, *ZNF577* and *ROPN1L* were replicated in three independent stages and hence were also considered further in our study using independent cases from Alberta, Canada ($n = 1,502$; Stage 4, see below) for assessing the overall risk in combined Stages 1–4.

Based on the significant trends of associations in previous findings, we further (*i*) examined the SNPs in a stratified analysis based on menopausal status or family history in the combined Stages 1–4; (*ii*) SNP rs1429142 which showed association in premenopausal women was selected for validation in women of African and Chinese ancestries; and (*iii*) conducted a fine-scale mapping for rs1429142; the goal was to identify the potential causal variants and their putative functions.

Methods
Study population
Written informed consent was obtained from all study participants, and the study protocol was approved by the Health Research Ethics Board of Alberta (HREBA)-Cancer Committee.

Samples from Alberta, Canada (internal dataset, Stages 1–4)
The study includes age-matched breast cancer cases (Stages 1–3, $n = 2,750$) and apparently healthy controls ($n = 4,271$) recruited from the province of Alberta, Canada. The cases utilized in Stages 1–3 were described elsewhere,[21,22] and for the current study, we have accessed additional breast cancer cases (Stage 4, $n = 1,722$) diagnosed between 2002 and 2015. The study inclusion criteria were the same as in the previously adopted. Detailed description of the sample inclusion criteria were described in the supporting document with pertinent demographical and patient clinical characteristics (Supporting Information Table S1).

Patient demographics
Total sample size ($n = 9,028$) for the current study included 4755 (cases) and 4,271 (controls). Among the cases, 35% and 62% were premenopausal and postmenopausal cases (self-declared at the time of diagnosis), respectively. Luminal cancers were predominant (77%) and this frequency was maintained when cases were stratified by menopausal status. Up to 94% of the total breast cancer cases in our study were >40 years of age. Predominant number of familial cases are diagnosed <40 years of age. Further cases were excluded based on the study inclusion criteria and genotyping call rate cutoff, resulting in 4331 cases amenable for association analysis. The cases and controls showed similar frequencies for age and BMI distribution (Supporting Information Table S1 and Fig. S1).

External datasets
We have accessed the external GWAS datasets from published studies for replication and validation. For the independent replication stage, we have accessed postmenopausal women of European ancestry from the Cancer Genetic Markers of Susceptibility (CGEMS) study ($n = 2,287$). Similarly, for validation stage, we have accessed breast cancer cases and controls from African diaspora study ($n = 3,766$) and Shanghai Breast Cancer Consortium ($n = 4,870$). Detailed description of the study cohorts and genotyping platforms utilized in these studies are described in Supporting Information.

DNA extraction and genotyping
Genomic DNA was extracted from buffy coat samples using a commercially available Qiagen Tm kit (Mississauga, ON, Canada). Genotyping was performed using Sequenom iPLEX

Gold platform (San Diego, CA) and utilized the services provided by McGill University and Genome Quebec Innovation Center, Montreal, QC, Canada.

SNP selection and genotyping

Stage 1 of our study and whole genome genotype data was generated using Human Affymetrix SNP 6.0 array (906,600 SNPs) for 348 cases and 348 controls and was reported earlier. Principal component analysis was used to identify outliers ($n = 72$) and the remaining 624 samples clustered with HapMap population of Caucasian ancestry.[21] We applied a call rate filter (>99%) and assessed for deviations from Hardy–Weinberg equilibrium (cut-off of $p < 0.001$ on controls). We also performed identity by descent analysis[23] based on the genotypes to identify cryptic relatedness (with pairwise correlation $r^2 > 0.25$). Human Affymetrix SNP 6.0 array has 40,146 SNPs on chromosome (Chr) 4 and 209 SNPs in 1 MB region was used for imputation. We used GTOOL for flipping the strand for the SNPs genotyped from the minus strand in Affymetrix to the same strand convention as the reference panel. Followed by strand flipping Chr4 was phased using SHAPEIT algorithm[24] prior to imputation. For imputation, we used the best guess method, implemented within IMPUTE2 algorithm[25] and the 1000 Genomes panel based on diverse populations was used as the reference for imputation.

We imputed 952,002 SNPs on Chr 4 with imputation info score >0.7. SNPs imputed were filtered for genotype call rate >95% and minor allele frequency (MAF) >1%. We selected 2019 SNPs in the 1 MB region flanking the index SNP rs1429142 and tagSNP were selected from the locus. Of the 2019 SNPs, 209 are genotypes from the Affymetrix platform and the rest are imputed SNPs. Instead of genotyping all the 2019 SNPs across all samples as cost-effective strategy, we selected SNPs that will give coverage across the 1 MB region and that enabled second round of imputation in all samples from Stages 1–4. We used Tagger, a SNP selection tool implemented within Haploview ver4.2 and selected 63 tagSNPs. Multiplex assay system on Sequenom iPLEX Gold platform was validated for 56 SNPs (including SNP rs1429142). We genotyped all cases and controls from Stages 1–4, and 4,331 cases and 4,271 controls passed genotyping (Supporting Information Table S3). The 56 SNPs (spanning Chr4:147,802,550–148,781,409, hg19 build) were in LD ($r^2 > 0.2$) with rs1429142. SNP call rates for 56 SNPs were >92%. We also estimated the imputation and genotyping concordance for these 56 tagSNPs in the Stage 1 samples; all the SNPs had a correlation (r^2) of >0.80, of which 44 SNPs had r^2 of >0.90. We included several technical replicates for each SNP within the genotyping batch, and genotype concordance was 100%. We estimated the concordance between genotyping batches (previous genotype calls for Stage 1–3 samples) which also showed 100% concordance.

We reimputed data based on 56 SNPs in the premenopausal cases ($n = 1,503$) and controls ($n = 4,271$), as the focus of this investigation was on assessing breast cancer risk and replicating previous findings. We imputed 1,715 SNPs using one-phase imputation approach with imputation info score value >0.7. After applying the genotyping quality filter, 587 SNPs were retained with 85% genotype call rate and minor allele frequency ≥5% for fine-mapping association analysis.

Statistical analysis

We used correlation/trend test with one degree of freedom (df) for unadjusted analysis in the association study between cases and controls. Unconditional logistic regression was used to estimate the odds ratio (OR) with 95% confidence interval (adjusted for BMI). Here, we report association statistics for both unadjusted and BMI adjusted analysis. Even though BMI is an independent risk factor for breast cancer, we examined if the risk associated with identified variants are in any way modified by BMI. Subgroup analysis was carried out based on menopausal status, disease stage (I, II *vs.* III), grade (high *vs.* low) and molecular subtype (luminal A *vs.* non-luminal A). *p*-Heterogeneity was estimated between the subgroups. Association of rs1429142 with BMI was also carried out independently as a quantitative trait. All association analyses were performed using Golden Helix SNP & Variation suite and Plink v1.07.[26] Conditional logistic regression analysis was conducted with adjustments for the highly associated variants (rs13134510, rs1366691, rs1429139 and rs12501429) using binary logistic regression analysis in PLINK. Likelihood ratio analyses were carried out using IBM SPSS Statistics (IBM Corp. Released 2013. IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY: IBM Corp) to identify the potential causal variants. The top associated SNP rs13134510 was used as a reference, to test fine-mapped SNPs with 4 degrees of freedom. We excluded SNPs with *p*-value >0.01.

In silico predictions for functional relevance of the fine-mapped SNPs

To elucidate the functional relevance, the associated fine-mapped variants were annotated (breast cancer risk variants at *p*-value <0.05). The annotations were from different data sources: ENCODE (Encyclopedia of DNA Elements),[27] Roadmap Epigenomics consortium[28] available through Regulome DB ver1.1,[29] HaploReg v4.1[30] and Washington University Epigenome Browser (https://epigenomegateway.wustl.edu/). The variants were annotated using RegulomeDB and those with scores of 1–4 were further annotated for histone marks such as H3K4me1, H3K4me3, indicative of enhancer and promoter activity, respectively. We used the histone marks data generated in normal breast epithelial cell lines such as mammary epithelial primary cells (HMEC), breast variant human mammary epithelial cells (vHMEC) and breast myoepithelial primary cells. We also annotated for DNase Hypersensitivity sites, which are informative about the open chromatin state in the breast epithelial cell lines. For transcription factor (TF) binding, we used the ChIP-seq datasets generated for the breast cancer cell lines MCF10A-ER-Src, HMEC and MCF7 (ENCODE and Roadmap databases). Polymorphisms potentially affecting the TF binding

motifs were predicted using position weighted matrix (PWM) for each variant, when applicable. We accessed the ENCODE Hi-C datasets for HMEC and ChIA-PET data for POL2A and CTCF in the MCF-7 cell line. TAD domain predictions based on the Hi-C data was predicted using the 3D genome browser[31] (http://promoter.bx.psu.edu/hi-c/view.php). Interaction arcs based on the ChIA-PET data was generated based on the Washington University Epigenome Browser. We also captured the expression of nearby genes (~2 MB spanning the SNP rs1429142) based on the RNA-Seq for the HMEC cell line.

Expression quantitative trait loci analysis
Expression quantitative trait loci (eQTL) data for normal breast tissues and heart left ventricle were used for the interpretation of the results based on GTEx database (GTEx portal was accessed on 07/04/2018, GTEx analysis V7 [dbGaP Accession phs000424.v7.p2]). eQTL based on lymphoblastoid cell lines were inferred from ENCODE project.

Results
Association of GWAS-identified SNP rs1429142 at Chr4q31.22 with overall and premenopausal breast cancer risk in women of European ancestry
In our previous study, we reported a novel SNP rs1429142 associated with overall breast cancer risk and the SNP conferred elevated risk among premenopausal women of Caucasian ancestry. The SNP is located at Chr4:148289398 (GRCh37/hg19), with minor allele "C" (MAF ~18%) among the Caucasian population. In the combined analysis (Table 1) for overall breast cancer risk (Stages 1–4; total n = 4,331 cases/4271 controls), SNP rs1429142 showed a genome level significance with adjusted p-value 4.35×10^{-08} and OR of 1.25 (1.15–1.35). The genome-wide significance threshold was calculated based on testing 782,838 SNPs for association in Stage 1 study ($0.05/782,838 = 6.4 \times 10^{-8}$).

In a subgroup analysis (samples from Stages 1–4) based on menopausal status, the association of rs1429142 with premenopausal breast cancer risk in women of Caucasian ancestry reached genome level significance with an adjusted p-value of 5.81×10^{-10} and OR of 1.40 (1.26–1.56), as was also demonstrated in our previous study. However, the association among postmenopausal women in our population was moderate (OR of 1.17 [1.07–1.28], p-value of 7.81×10^{-04}) (Table 1). The p-value for the test of heterogeneity comparing the ORs between premenopausal and postmenopausal women was statistically significant at 1.84×10^{-02} (Supporting Information Table S2a), consistent with our earlier findings.[22]

The SNP rs1429142 was initially identified to be associated with sporadic breast cancer (Stages 1 and 2). In subsequent replication studies (Stages 3 and 4), we recruited cases irrespective of family history. Stratified association analysis of SNP rs1429142 was conducted based on family history. The p-het (0.37) between these strata were not significant, however the SNP showed a trend of elevated risk among cases

without family history (n = 1,886 cases/4,271 controls, p-value 5.09×10^{-8} OR 1.31) compared to cases with family history (n = 1,640 cases/4,271 controls, p-value 1.86×10^{-4} OR 1.21; Supporting Information Table S2a), validating the original study premise. Subgroup analysis based on clinicopathological features such as molecular subtype (luminal $vs.$ non-luminal), tumor grade (high $vs.$ low), and stage (<III $vs.$ ≥III did not show any trends of elevated risk between the strata (Supporting Information Table S2a).

Based on the insights gained for the stratified analysis (family history or menopausal status) for rs1429142, we extended the analysis for additional SNPs reported from our earlier GWAS (rs1092913 on Chr5, rs3848562 on Chr19). In the analysis based on cases with no family history (sporadic) $vs.$ controls, the SNPs rs1092913 and rs3848562 showed genome-wide significance. However, there were no statistically significant differences (p-het) in the risk between the cases with or without family history (Supporting Information Table S2b). Additionally, the SNP rs1092913 showed higher association with premenopausal breast cancer compared to postmenopausal, although there was no statistically significant difference in the risk between the strata (Supporting Information Table S2c). Based on these analyses, SNP rs1429142 on Chr4 is thus a novel variant conferring statistically significant higher risk for premenopausal breast cancer. Therefore, rs1429142 was considered for further validation and fine-mapping.

We independently tested for the association of rs1429142 in Cancer Genetic Markers of Susceptibility dataset (CGEMs; 1,144 cases/1,143 controls) comprising postmenopausal women. The SNP rs1429142 did not show statistical significance (OR 1.05; p-value = 6.8×10^{-01}, Table 1).

Association of SNP rs1429142 with premenopausal breast cancer risk in women of African and Chinese ancestry
The association of SNP rs1429142 was tested using datasets from the African Diaspora study. SNP rs1429142 has a T/C polymorphism in the African population with a minor allele (T) frequency of 25%. Since C allele is a risk allele in Caucasian population, we present all our association study findings with reference to C allele. We initially tested rs1429142 in 1607 cases/2041 controls for overall risk of breast cancer and the SNP did not show statistically significant association (p-value 6.08×10^{-01}). Interestingly, in the stratified analysis, SNP rs1429142 was associated with breast cancer risk among premenopausal women (p-value 1.45×10^{-02}; OR of 1.2 [1.03–1.40]). Risk for postmenopausal women was not statistically significant (8.56×10^{-01}).

We examined the association of SNP rs1429142 (C/T polymorphism, C allele is the minor allele) in Chinese ancestries using datasets from the Shanghai Breast Cancer Genetic Study. We analyzed 2,731 cases and 2,139 controls and the overall association was not statistically significant (p-value = 2.50×10^{-01}). The SNP was also not significant in the stratified analysis based on menopausal status, that is,

Table 1. Replication and validation of SNP rs1429142 at Chr4q31.22 and association with premenopausal breast cancer risk

	Sample size, *n*	Status	Risk allele/allele frequency	*p*-value	Allelic OR [95% CI]
Replication (Caucasian populations)					
Caucasian, Stages 1–4[1] (Canada)	**4,331 cases/4,271 controls**	**Overall**	**C/0.18**	**4.35E−08**	**1.25 [1.15–1.35]**
	1,503 cases/4,271 controls	**Premenopausal**	**C/0.17**	**5.81E−10**	**1.40 [1.26–1.56]**
	2,700 cases/4,271 controls	Postmenopausal	C/0.18	7.81E−04	1.17 [1.07–1.28]
Caucasian (CGEMs study)	1,144 cases/1,143 controls	Postmenopausal	C/0.17	6.80E−01	1.05[0.89–1.22]
Validation (Diverse populations)					
African Diaspora	1,607 cases/2,041 controls	Overall	C/0.75	6.08E−01	1.03 [0.92–1.14]
	645 cases/2,041 controls	**Premenopausal**	**C/0.75**	**1.45E−02**	**1.21 [1.04–1.40]**
	663 cases/2,041 controls	Postmenopausal	T/0.75	8.56E−01	1.01 [0.88–1.17]
Chinese (Shanghai Breast Cancer Study)	2,731 cases/2,139 controls	Overall	C/0.36	2.50E−01	1.05 [0.96–1.13]
	1,577 cases/2,139 controls	Premenopausal	C/0.36	6.00E−01	1.03 [0.93–1.14]
	1,154 cases/2,139 controls	Postmenopausal	C/0.36	2.20E−01	1.08 [0.96–1.22]

The text and numbers indicated in bold highlight the novel findings.
[1]Indicates the association analysis adjusted for body mass index (BMI) available for cases and controls in Canadian populations. BMI information was not available for external cohorts. Table summarizes the overall association in Caucasian populations (Stages 1–4 from Alberta, Canada) and the results stratified for menopausal status are also indicated. Association in postmenopausal women from CGEMS study is shown. SNP rs1429142 is validated in diverse ethnic populations. For SNP rs1429142, the minor allele is C in the Caucasian and Chinese populations (C/T), whereas it is T in the African population (T/C). Note that the frequencies of the minor alleles across the populations are different. The results are presented with respect to the risk allele "C".

premenopausal (*p*-value = 6×10^{-01}) and postmenopausal (*p*-value = 2.2×10^{-01}).

Therefore, rs1429142 is a novel premenopausal risk variant with a high effect size for breast cancer in the Caucasian population (OR 1.40) relative to the GWAS variants reported thus far. This variant was also validated in premenopausal African women (Table 1). These findings warrant further fine-scale mapping of the locus to identify potential causal variant(s) and their putative roles in conferring breast cancer susceptibility.

Identification of potential causal variants by fine-scale mapping of Chr4q31.22

We performed a fine-scale mapping of SNP rs1429142 to identify putative causal variants. We fine-mapped a ~1 MB region, 147,802,550–148,781,409 (GRCh37/hg19) flanking the SNP, rs1429142 located at Chr4:148289389. The 1 Mb region had 209 SNPs from the Affymetrix array, we adopted imputation and genotyping approaches to increase the SNP density from 209 SNPs in 1 MB region to 1,715 SNPs at the imputation info score cutoff of >0.7. Furthermore, filtering based on 587 SNPs were retained based on >85% genotype call rate and MAF ≥5%.

Association testing of 587 fine-mapped SNPs in the premenopausal cases and controls identified 135 SNPs with *p*-value of <0.05 and 49 SNPs at $<10^{-8}$ (Fig. 1 and Supporting Information Table S4, *p* values unadjusted and adjusted for BMI). Four SNPs (rs13134510, rs1366691, rs1429139 and rs12501429) had *p*-values of $<10^{-11}$. All these four fine-mapped SNPs were in LD with the originally identified SNP rs1429142. SNP rs13134510 showed the highest statistical significance (unadjusted *p*-value 1.11×10^{-12}). Conditional

regression analysis based on these four SNPs did not reveal any additional independent signals (Supporting Information Fig. S2a–d and Table S5).

We used multiple methods, tools and annotation algorithms described below to assess the functional relevance of the associated and fine-mapped SNPs.

1. *Log-likelihood ratio analysis*: This was carried out as an independent pruning method which revealed five SNPs with a *p*-value of >0.05. These five SNPs were excluded and the remaining 130 SNPs (including the top four SNPs showing the highest association) were identified as potentially causal variants showing a statistical significance at *p* < 0.01 (Supporting Information Table S6).

2. *LD mapping*: Given the expected small LD block patterns in African populations and the statistical significance observed among premenopausal women, we refined the fine-mapped region (130 SNPs) using the HapMap dataset. We noted that the Caucasian population had fewer but larger LD blocks consisting of the fine-mapped SNPs and the GWAS SNP rs1429142 (Supporting Information Fig. S3a). As expected, we observed multiple smaller LD blocks in the African populations in the fine-mapped region in contrast to the Caucasian populations. The fine-mapped variants (130 SNPs) were scattered across multiple LD blocks in the African population. In the African population, 10 of the highly significant fine-mapped SNPs (*p*-value <10^{-10}; rs1366691, rs1429139, rs12501429, rs1583003, rs2163012, rs2163011, rs12498595, rs13120678, rs1366679 and rs1313 4510) were clustered in a single LD block and the remaining SNPs including the GWAS index SNP rs1429142 were scattered over multiple LD blocks (Supporting Information Fig. S3b). This contrasts with the Caucasian population
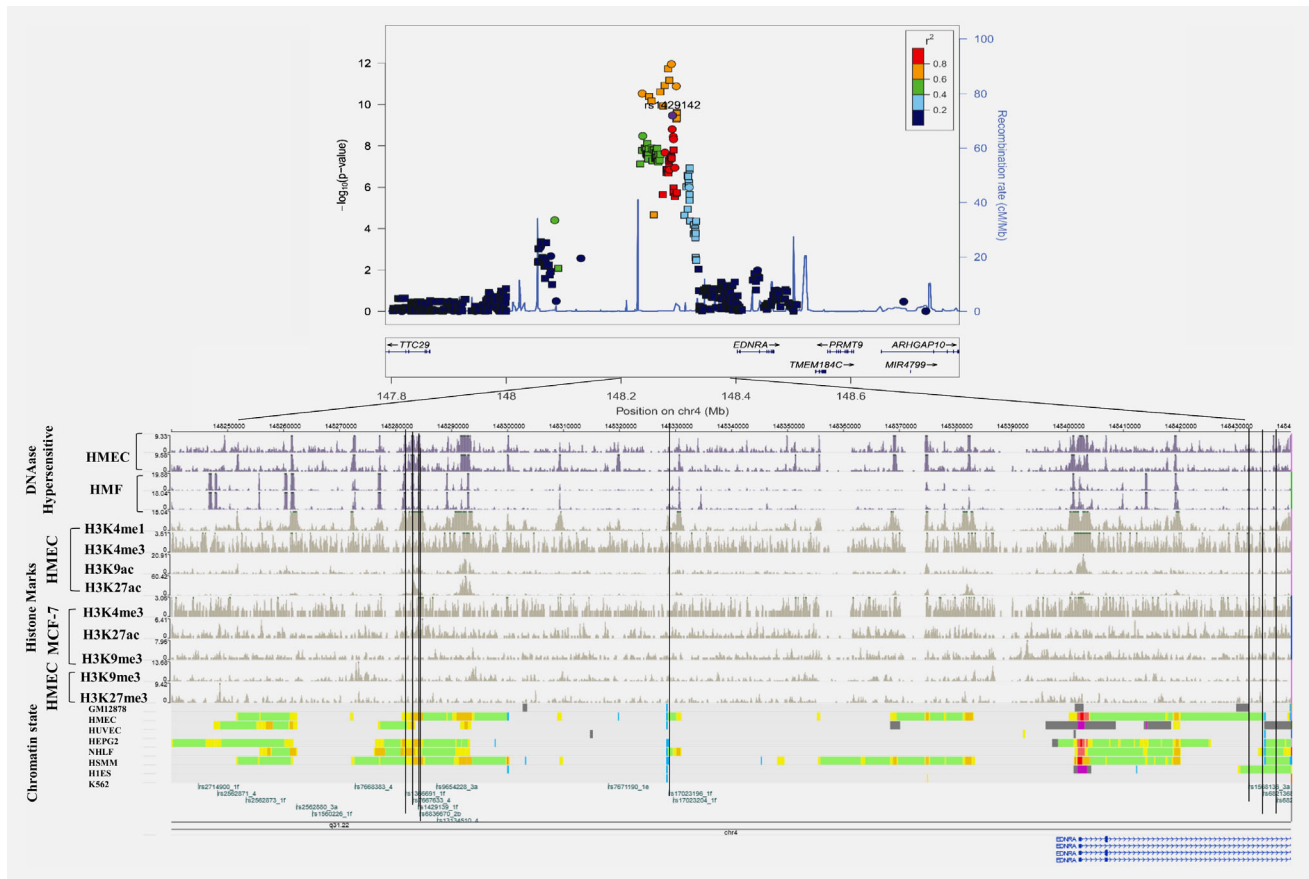
Figure 1. Association of the fine-mapped SNPs with premenopausal breast cancer risk and their functional annotation. This figure represents the association of the fine-mapped SNPs with premenopausal breast cancer risk and the functional relevance of the SNP is indicated in cell line data. The top panel indicates the locus zoom plot with an association *p*-value (log scale) on the *y*-axis and genomic location on the *x*-axis. The 587 fine-mapped SNPs are represented as squares (imputed) and circles (genotyped) and the LD (r2) between the SNPs were indicated according to the color scale. The GWAS SNP rs1429142 is indicated. The bottom panel indicates the functional relevance of the fine-mapped SNPs inferred using human breast cell lines (HMEC, HMF and MCF-7). The DNase hypersensitive sites (HMEC, HMF), histone marks (HMEC and MCF-7) and chromatin states (Encode cell lines) were inferred from corresponding cell lines. The SNPs with RegulomeDb score (1–4) are indicated.

wherein the index SNP along with all the 10 highly associated SNPs were found in a single LD block.

3. *Putative regulatory functions for the causal variants*: We have annotated all 130 variants for functional relevance. We used RegulomeDB-ver1.1 (Supporting Information Tables S7 and S8) and HaploReg-v4.1 (Supporting Information Table S9) for functional annotations. We identified 19 SNPs (Supporting Information Table S8) with Regulome scores between 1 and 4 (1 being the most informative); these are derived from composite scores from the inferred regulatory functional states such as DNase hypersensitivity sites, transcription factor binding, chromatin state, histone marks and changes in the binding motifs of the bound proteins. Among the 19 SNPs with putative regulatory functions, five SNPs (with *p* values): rs1366691 (1.91×10^{-12}), rs1429139 (6.64×10^{-12}), rs7667633 (5.05×10^{-08}), rs6836670 (1.41×10^{-07}) and rs17023196 (1.01×10^{-04}) were predicted to have enhancer roles inferred

from chromatin marks (or posttranslational modification of histone proteins). The combination of the chromatin marks was used to predict the enhancer functions using the method chromHMM (multivariate hidden Markov model). The chromatin state at the locus of interest harbored the histone marks: H3K4me1, H3K27ac and H3K9ac, captured by ChIP-seq assay in normal breast cell lines: mammary epithelial primary cells (HMEC) and breast variant human mammary epithelial cells (vHMEC; Supporting Information Table S9). There was evidence of DNase hypersensitivity peaks near these SNPs captured in HMEC, vHMEC and breast myoepithelial primary cells (Supporting Information Table S9).

Among the 19 SNPs that were annotated for putative regulatory functions, we noted SNPs rs1568136, rs6821368 and rs6822565 were present within the intron of the *EDNRA* gene.

The histone marks at these loci indicated weak transcriptional activity in HMEC, vHMEC and breast myoepithelial primary cells. Additionally, we noted that SNP rs1568136 affected the binding of transcription factors such as EN1 and SNP rs6821368 affected binding of NF-AT, SOX, HDAC2, HOXA4, PAX-4, POU2F2, POU3F2 and SIN3AK-20 (Supporting Information Table S9) judged from the position weighted matrix (PWM) scores.

4. *Binding of transcription factors at the SNP sites*: The dataset from the ENCODE project offered further insights into binding of transcription factors (TFs) at three SNPs, rs1366691, rs7667633 and rs7668383. Evidence for binding of three TFs (FOS, STAT3 and POL2A) at these sites was obtained from the MCF10A-Er-Src cell line (derived from parental MCF-10A cells which are negative for estrogen receptor expression). However, MCF10A-Er-Src contains a variant of the Src kinase oncoprotein that is fused to the ligand binding domain of the estrogen receptor and is induced by adding tamoxifen (TAM; Supporting Information Table S8). Src expression leads to transformation of cells as evidenced by visible morphological changes between 24 and 36 hr. ENCODE project has also captured binding of TFs to target sites in TAM-treated and untreated cells at 4-,12- and 36-hr time intervals. Based on the ChIP-sequencing, FOS binding was noted to be high at rs1366691, rs7667633 and rs7668383 loci in the TAM-treated group relative to the untreated group when analyzed at different time intervals in the MCF10A-Er-Src cell line (Fig. 2).

In summary, the evidence presented from the various methods described above indicated that a select number of SNPs (1 and 2) among the fine-mapped region appeared to be active enhancer domains judged from the collective experimental evidence (3 and 4) from various cell lines (epigenetic marks and transcription factor binding). We identified three SNPs; rs1366691 and rs1429139 (at *p*-value $<10^{-10}$) and rs7667633 (at *p*-value 10^{-08}) which are likely the causal SNPs. Our conclusions are based on the strengths of association and functionality as enhancers (inferred from chromatin state and binding of transcription factors). These loci may exhibit complex long or short-range DNA interactions, and such interactions between the enhancer(s) and promoters may contribute to the overall regulatory effects.

Gene regulation by short-range DNA interactions
The fine-mapped region was interrogated for possible short-range interactions based on the Hi-C data available for HMEC cell line. The fine-mapped regions harbored multiple interactions with the neighboring region and were predicted to be present within the topologically associated domain (TAD; Supporting Information Fig. S4a). TAD consists of the regions of the DNA that preferentially interact with each other. The interactions are predominantly seen within the TAD boundaries and are less likely to interact outside of the TAD.[32] Since TADs are derived by complex DNA looping and interactions, they play a role in gene regulation, wherein the promoters interact with the local enhancer elements. CCCTC-binding factor (CTCF) and Cohesin (a multisubunit protein complex) are the common DNA binding proteins often known to be enriched in the TAD regions. DNA looping is mediated by the binding of CTCF proteins mediating the physical contact of the domains. We analyzed the data from the chromatin interaction analysis by paired-end tag (ChIA-PET) data generated from MCF-7 enriched for CTCF and POL2 (Supporting Information Fig. S4b). We observed multiple interactions between fine-mapped SNPs and upstream promoter elements of nearby genes including *EDNRA*, *PRMT10*, *ARHGAP10* and *TMEM18C* (potential eQTLs, Supporting Information Table S10). Further experiments are needed to gain mechanistic insights on the regulation of the target genes and interactions with the identified potential causal variants.

Discussion
We report three potential causal variants (rs1366691, rs1429139 and rs7667633) from fine-mapping and annotation analysis which are strongly associated with premenopausal breast cancer risk. The effect size for the three novel variants are in line with the originally described index SNP rs1429142 (OR 1.4, Table 1 and Supporting Information Table S4). GWAS literature identified fewer variants with the effect sizes in the range 1.25–1.4. Limited GWASs addressed sporadic breast cancer without emphasis to menopausal status,[33–35] or focused predominantly on postmenopausal women with the familial component.

Despite several GWAS findings reported in breast cancer literature, rs1429142 was never reported as a risk variant. We ascribe this to our stratified analysis approach with an emphasis on premenopausal risk. A recent breast cancer association study reported by Michailidou *et al.*,[36] the team utilized iCOGS and OncoArrays with a sample size of 108,067 cases and 88,386 controls. The data was accessible through the consortia (Breast Cancer Association Consortium [BCAC]). When interrogated for the summary statistics, the SNP rs1429142 did not show association with breast cancer risk (*p* = 0.19). However, upon closer examination, we identified that a larger proportion of cases were from postmenopausal women (46%) and 19% were from premenopausal women, with the remaining cases were unknown for menopausal status (35%, but the age distribution suggested that majority of these cases are likely postmenopausal). In our study, 35% were premenopausal cases, 62% postmenopausal and only 3% of cases as unknown for menopausal status. Risk allele frequency in Caucasian population was at 16% for controls, and 21% and 19% for premenopausal and postmenopausal cases, respectively. Risk allele frequency differences and the disproportionate number of postmenopausal cases relative to premenopausal cases may affect the observed overall association statistics.

The SNP rs1429142 was shown to be associated with overall breast cancer risk, as well as an enhanced risk among
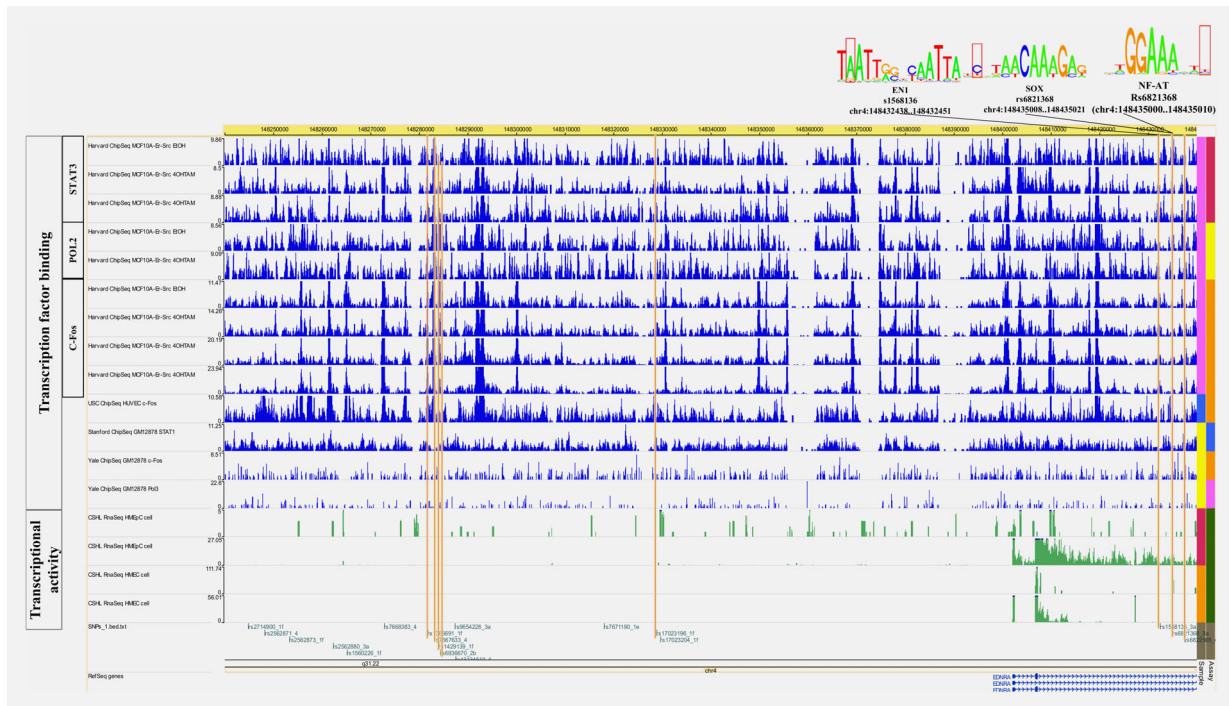
Figure 2. Transcriptional activity at the fine-mapped locus. The figure represents transcriptional activity at the fine-mapped locus. The binding of the transcription factors (left top corner) was determined using ChIP-Seq data capturing the binding of FOS, STAT1/3 and Pol2/3 were described in breast cell lines (MCF10A-Er-Src, HMEC) and Encode cell lines. Similarly, transcriptional activity (left bottom panel) estimated from the RNA-seq data generated in HMEC cell line. The binding of the transcription factors (right-side top) such as EN1, SOX and NF-AT may potentially be affected by polymorphism in the intron of the *EDNRA* gene estimated from position weighted matrix. The source of the data is shown in the column (ChIP-seq for c-FOS, POL2, STAT3) based on MCF10A-Er-Src were generated from Harvard, for the encode cell lines: c-FOS captured in HUVEC from University of Southern California; STAT1 captured in GM12878 from Stanford University; C-FOS and Pol3 captured in GM12878 from Yale University. Figure was generated based on the output from the browser http://epigenomegateway.wustl.edu/browser/

premenopausal women (Stages 1–4, Table 1). The overall breast cancer risk conferred by SNP rs1429142 was not affected by luminal status, tumor grade or stage (Supporting Information Table S2a). In an independent analysis, we showed that the SNP rs1429142 was not associated with estrogen receptor (ER) status (*p*-het between ER-positive *vs.* ER-negative cases, Supporting Information Table S2a). The majority of the GWAS identified SNPs in earlier studies were shown to confer risk in women with ER-positive disease[36,37] and in postmenopausal cases.[33] Also consistent with earlier findings, the SNP rs1092913 on Chr5 and rs3848562 on Chr 19 showed higher association with sporadic breast cancer, even though *p*-heterogeneity was not significant. These SNPs warrants further investigations.

The minor allele and/or the MAF of SNP rs1429142 showed variations across the populations. Among the Chinese populations, the minor allele was C with a frequency of 30%. The overall association as well as cases stratified by menopausal status did not show statistically significant associations in women of Chinese ancestry.

Among the African populations, an allele reversal was noted wherein C is the major allele and T is the minor allele with 75% and 25% frequencies, respectively. In the overall association, SNP rs1429142 was not associated with breast

cancer, however, in the subgroup analysis, its association was significant among premenopausal breast cancer risk (*p*-value <0.05). The C allele remained the risk allele across different populations (Table 1), an observation that aligns with the higher prevalence of premenopausal breast cancer among women of African ancestry.[38,39]

In the fine-scale mapping of the associated region at the Chr4q31.22 locus, we identified 587 SNPs within the 1 Mb region flanking SNP rs1429142. Of the 587 SNPs, 135 were associated with premenopausal breast cancer risk. Conditional regression analysis did not reveal any independently associated signals. Likelihood analysis retained 130 as putatively causal SNPs with *p* values <0.01. The fine-mapped region and the SNPs showing association with premenopausal breast cancer risk were present within fewer but large LD blocks in the Caucasian population, whereas there were multiple but smaller LD blocks for the same region in the African population. These findings agree with the higher level of recombination events and resultant decay of LD in African populations (Supporting Information Fig. S3), consistent with current knowledge of LD in diverse populations.

Functional scoring revealed five SNPs (rs1366691, rs1429139, rs7667633, rs6836670 and rs17023196) at highest predicted levels of functionality (i.e., as enhancers). The DNase hypersensitivity

peaks revealed an open chromatin state at these loci. In addition, the histone methylation pattern, H3K4me1 and acetylation of H3K9ac and H3K27ac suggested potential enhancer roles based on HMEC, vHMEC and breast myoepithelial primary cell lines. To decipher transcription factors binding at these loci, we utilized the ChIP-Seq data from ENCODE for the MCF10-src cell line. The characteristic feature of MCF10-Src cells is that upon transformation by Tamoxifen induction, the cells exhibit increased motility, invasion, formation of foci, formation of single cell colonies, mammospheres and formation of tumor in mouse xenografts.[40,41] Based on the ENCODE data, transcription factors including FOS, STAT3 and POL2RA were bound to SNPs rs136691, r7667633 and rs7668383 from among the fine-mapped loci. These results suggested active enhancer regions at the putative causal loci which potentially regulate the expression of downstream target genes flanking the index SNP. For instance, the nearest target gene identified was *EDNRA*, located 2 kb downstream of putative causal SNP rs1366691.

STAT3 protein is a well-characterized transcription factor implicated in many cancer types.[42–44] STAT3 expression alone was sufficient to initiate tumorigenesis, and its overexpression brings about transformation of both human fibroblast[45] and MCF10 derived (MCF10-ER-Src)[46] cell lines. Induction of Src expression transforms the cells, conferring the phenotypic changes characteristics of cancers.[40,41] The process of transformation involves epigenetic switch and inflammatory pathway gene expressions. STAT3 exclusively binds to open chromatin regions and regulates expression of NFKB1 which in turn regulates expression of IL6, a cascade of events that is part of the well-characterized feedback loop involving these transcription factors and inflammatory mediators,[47] a hallmark in tumorigenesis cascade of events. Often STAT3 and FOS proteins coregulate the transcription of genes. In our study, STAT3 and FOS bound to the sequences at SNP sites, rs1366691 and rs7667633 in the MCF10-ER-Src cell line during the process of transformation.

Since the fine-mapped variants were predicted to have an enhancer function, they are likely to influence promoters of the nearby genes by DNA looping. Based on the DNA interaction profiles generated in HMEC cells, we confirmed that the fine-mapped loci have multiple local interactions and were present within TAD domains. TAD domains, which were recently described,[32] consist of regions of DNA that are likely to interact with each other within the TAD boundaries. These are complex mechanisms of gene regulation and TAD domains are conserved across tissues and species.[32,48]

Several SNPs from the fine-mapped region appeared to be eQTLs (in different tissues other than breast) regulating nearby genes *ENDRA*, *ARHGAP10* present within ~800 kb distance (Supporting Information Table S10). ENDRA is well known for its role in vasoconstriction and in arterial diseases. However, these genes are also often noted to be dysregulated in cancer; EDNRA bound by endothelin-1 triggers a cascade of

signaling pathways leading to proliferation,[49] angiogenesis,[50] invasion/tumor progression[51,52] and inhibition of cell death,[53] when activated by Hypoxia induced factor 1-Alpha. Overexpression of EDNRA has been observed in several cancer types[49,52,53] and is an independent predictor of prognosis.[54] Similarly, ARHGAP10 belongs to the family of Rho GTPase-activating proteins that are known to play a role in cell cytoskeleton organization, cellular migration and adhesion, regulation of transcription.[55] ARHGAP10 was associated with invasive breast cancer prognosis,[56] ovarian[57] and lung cancers.[58] ARHGAP10 is often downregulated in tumors and may play a role as a tumor suppressor.[57,58] The eQTL role for the fine-mapped variants in breast tissues warrants further work and is recognized as a potential limitation for generalizability of the findings.

The fine-mapped variants in our study are common polymorphisms (MAF 18%). A higher sample size might have enabled the identification of low frequency putative causal variants within the susceptibility locus to gain additional biological insights.[5,18] Due to the challenges in the functional characterization of the fine-mapped loci, only a limited number of breast cancer studies successfully identified the target genes (*FGFR2*,[11] *CCND1*,[10] *MAP3K1*,[13] *TERT*,[9] *IGFBP5*,[12] *TET2*,[14] *STXBP4*[16]) with role in breast cancer etiology.

In summary, we have identified three potential causal variants (rs1366691, rs1429139 and rs7667633) strongly associated with premenopausal breast cancer risk and the variants appear to have enhancer functions, likely regulating the nearby target genes. It is not clear on the biological mechanisms underlying the observed higher risk for premenopausal women, and further experimental evidence is warranted. The novel locus associated with premenopausal breast cancer in our study and a fine-mapping analysis of the locus revealed binding of transcription factors known to play a role in inflammatory pathways, also a common etiological basis of many cancers.

Author contributions

Study concept: SD. Experiments and analysis: MK. Article preparation: SD and MK. Data collection and analysis on Chinese subjects from the Shanghai Breast Cancer Consortium: WZ. Statistical input and consultations: YY and SG. Input for breast cancer clinical perspectives: JRM and AAJ. Investigations and extensive editing: CEC. Feedback, analysis and conclusions: all authors.

Cancer Epidemiology

Cancer Epidemiology

References

1. Canadian Cancer Society's Advisory Committee on Cancer Statistics. *Breast Cancer Statistics.* Toronto, ON: Canadian Cancer Society, 2017.

2. Siegel R, Ma J, Zou Z, et al. Cancer statistics, 2014. *CA Cancer J Clin* 2014;64:9–29.

3. Welter D, MacArthur J, Morales J, et al. The NHGRI GWAS catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res* 2014;42: D1001–6.

4. McCarthy MI, Abecasis GR, Cardon LR, et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet* 2008;9:356–69.

5. Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. *Nature* 2009;461:747–53.

6. Udler MS, Meyer KB, Pooley KA, et al. FGFR2 variants and breast cancer risk: fine-scale mapping using African American studies and analysis of chromatin conformation. *Hum Mol Genet* 2009; 18:1692–703.

7. Udler MS, Ahmed S, Healey CS, et al. Fine scale mapping of the breast cancer 16q12 locus. *Hum Mol Genet* 2010;19:2507–15.

8. Chen F, Chen GK, Millikan RC, et al. Fine-mapping of breast cancer susceptibility loci characterizes genetic risk in African Americans. *Hum Mol Genet* 2011;20:4491–503.

9. Michailidou K, Hall P, Gonzalez-Neira A, et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat Genet* 2013;45: 353–61.

10. French JD, Ghoussaini M, Edwards SL, 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.

11. Meyer KB, Reilly M, Michailidou K, 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.

12. Ghoussaini M, Edwards SL, Michailidou K, et al. Evidence that breast cancer risk at the 2q35 locus is mediated through IGFBP5 regulation. *Nat Commun* 2014;9:16193.

13. Glubb DM, Maranian MJ, Michailidou K, 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.

14. Guo X, Long J, Zeng C, et al. Fine-scale mapping of the 4q24 locus identifies two independent loci associated with breast cancer risk. *Cancer Epidemiol Biomarkers Prev* 2015;24:1680–91.

15. Orr N, Dudbridge F, Dryden N, et al. Fine-mapping identifies two additional breast cancer susceptibility loci at 9q31.2. *Hum Mol Genet* 2015; 24:2966–84.

16. Darabi H, Beesley J, Droit A, et al. Fine scale mapping of the 17q22 breast cancer locus using dense SNPs, genotyped within the collaborative oncological gene-environment study (COGs). *Sci Rep* 2016;6:32512.

17. Horne HN, Chung CC, Zhang H, et al. Fine-mapping of the 1p11.2 breast cancer susceptibility locus. *PLoS One* 2016;11:e0160316.

18. Shi J, Zhang Y, Zheng W, et al. Fine-scale mapping of 8q24 locus identifies multiple independent risk variants for breast cancer. *Int J Cancer* 2016; 139:1303–17.

19. Zeng C, Guo X, Long J, et al. Identification of independent association signals and putative functional variants for breast cancer risk through fine-scale mapping of the 12p11 locus. *Breast Cancer Res* 2016;18:64.

20. Betts JA, Moradi Marjaneh M, Al-Ejeh F, et al. Long noncoding RNAs CUPID1 and CUPID2 mediate breast cancer risk at 11q13 by modulating the response to DNA damage. *Am J Hum Genet* 2017;101:255–66.

21. Sehrawat B, Sridharan M, Ghosh S, et al. Potential novel candidate polymorphisms identified in genome-wide association study for breast cancer susceptibility. *Hum Genet* 2011;130:529–37.

22. Sapkota Y, Yasui Y, Lai R, et al. Identification of a breast cancer susceptibility locus at 4q31.22 using a genome-wide association study paradigm. *PLoS One* 2013;8:e62550.

23. Pang AW, MacDonald JR, Pinto D, et al. Towards a comprehensive structural variation map of an individual human genome. *Genome Biol* 2010; 11:R52.

24. Delaneau O, Marchini J, Zagury JF. A linear complexity phasing method for thousands of genomes. *Nat Methods* 2011;9:179–81.

25. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* 2009;5:e1000529.

26. Purcell S, Neale B, Todd-Brown K, et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;81:559–75.

27. Consortium EP. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012;489:57–74.

28. Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, et al. Integrative analysis of 111 reference human epigenomes. *Nature* 2015;518: 317–30.

29. Boyle AP, Hong EL, Hariharan M, et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res* 2012; 22:1790–7.

30. 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–D34.

31. Wang Y, Zhang B, Zhang L, et al. The 3D genome browser: a web-based browser for visualizing 3D genome organization and long-range chromatin interactions. *Genome Biol* 2018;19:151.

32. Dixon JR, Selvaraj S, Yue F, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 2012; 485:376–80.

33. Hunter DJ, Kraft P, Jacobs KB, et al. A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nat Genet* 2007;39:870–4.

34. Stacey SN, Manolescu A, Sulem P, et al. Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat Genet* 2007;39:865–9.

35. Zheng W, Long J, Gao YT, et al. Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. *Nat Genet* 2009;41: 324–8.

36. Michailidou K, Lindstrom S, Dennis J, et al. Association analysis identifies 65 new breast cancer risk loci. *Nature* 2017;551:92–4.

37. Lilyquist J, Ruddy KJ, Vachon CM, et al. Common genetic variation and breast cancer risk-past, present, and future. *Cancer Epidemiol Biomarkers Prev* 2018;27:380–94.

38. Sighoko D, Bah E, Haukka J, et al. Population-based breast (female) and cervix cancer rates in The Gambia: evidence of ethnicity-related variations. *Int J Cancer* 2010;127:2248–56.

39. Sighoko D, Kamate B, Traore C, et al. Breast cancer in pre-menopausal women in West Africa: analysis of temporal trends and evaluation of risk factors associated with reproductive life. *Breast* 2013;22:828–35.

40. Aziz N, Cherwinski H, McMahon M. Complementation of defective colony-stimulating factor 1 receptor signaling and mitogenesis by Raf and v-Src. *Mol Cell Biol* 1999;19:1101–15.

41. Soule HD, Maloney TM, Wolman SR, et al. Isolation and characterization of a spontaneously immortalized human breast epithelial cell line, MCF-10. *Cancer Res* 1990;50:6075–86.

42. Bowman T, Garcia R, Turkson J, et al. STATs in oncogenesis. *Oncogene* 2000;19:2474–88.

43. Frank DA. STAT3 as a central mediator of neoplastic cellular transformation. *Cancer Lett* 2007; 251:199–210.

44. Yu H, Pardoll D, Jove R. STATs in cancer inflammation and immunity: a leading role for STAT3. *Nat Rev Cancer* 2009;9:798–809.

45. Bromberg JF, Wrzeszczynska MH, Devgan G, et al. Stat3 as an oncogene. *Cell* 1999;98: 295–303.

46. Dechow TN, Pedranzini L, Leitch A, et al. Requirement of matrix metalloproteinase-9 for the transformation of human mammary epithelial cells by Stat3-C. *Proc Natl Acad Sci U S A* 2004; 101:10602–7.

47. Fleming JD, Giresi PG, Lindahl-Allen M, et al. STAT3 acts through pre-existing nucleosome-depleted regions bound by FOS during an epigenetic switch linking inflammation to cancer. *Epigenetics Chromatin* 2015;8:7.

48. Vietri Rudan M, Barrington C, Henderson S, et al. Comparative hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. *Cell Rep* 2015;10:1297–309.

49. Zhang WM, Zhou J, Ye QJ. Endothelin-1 enhances proliferation of lung cancer cells by increasing intracellular free Ca^{2+}. *Life Sci* 2008;82: 764–71.

50. Wulfing P, Kersting C, Tio J, et al. Endothelin-1-, endothelin-A-, and endothelin-B-receptor expression is correlated with vascular endothelial growth factor expression and angiogenesis in breast cancer. *Clin Cancer Res* 2004;10:2393–400.

51. Rosano L, Cianfrocca R, Masi S, et al. Beta-arrestin links endothelin A receptor to beta-catenin signaling to induce ovarian cancer cell invasion and metastasis. *Proc Natl Acad Sci U S A* 2009;106:2806–11.

52. Wilson JL, Burchell J, Grimshaw MJ. Endothelins induce CCR7 expression by breast tumor cells via endothelin receptor A and hypoxia-inducible factor-1. *Cancer Res* 2006;66:11802–7.

53. Del Bufalo D, Di Castro V, Biroccio A, et al. Endothelin-1 protects ovarian carcinoma cells against paclitaxel-induced apoptosis: requirement for Akt activation. *Mol Pharmacol* 2002;61: 524–32.

54. Wulfing P, Diallo R, Kersting C, et al. Expression of endothelin-1, endothelin-A, and endothelin-B receptor in human breast cancer and correlation

with long-term follow-up. *Clin Cancer Res* 2003;9: 4125–31.

55. Jaffe AB, Hall A. Rho GTPases: biochemistry and biology. *Annu Rev Cell Dev Biol* 2005;21:247–69.

56. Azzato EM, Pharoah PD, Harrington P, et al. A genome-wide association study of prognosis in breast cancer. *Cancer Epidemiol Biomarkers Prev* 2010;19:1140–3.

57. Luo N, Guo J, Chen L, et al. ARHGAP10, down-regulated in ovarian cancer, suppresses tumorigenicity of ovarian cancer cells. *Cell Death Dis* 2016; 7:e2157.

58. Teng JP, Yang ZY, Zhu YM, et al. The roles of ARHGAP10 in the proliferation, migration and invasion of lung cancer cells. *Oncol Lett* 2017;14: 4613–8.

Cancer Epidemiology