# Transcriptome and interactome analyses identify the *TP53* interacting gene *RCCD1* as a candidate susceptibility gene at the 15p26.1 breast and ovarian cancer risk locus

Jasmine Plummer*[1,2], Felipe Segato Dezem*[1], Stephanie S. Chen*[1,2], Subash Dhungana[1], Deepika Wali[1], Brian Davis[1,2], Justyna Kanska[1], Niko Safi[1], Ji-Heui Seo[3, 4], Rosario I Corona[5,6], Joellen M Schildkraut[6], Paul DP Pharoah[8,9], Kate Lawrenson[5,6], Simon RV Knott[1,2], Matthew L Freedman[3,4,10], Siddhartha P Kar#[11,12], Simon A Gayther#¶[1,2]

[1] Center for Bioinformatics and Functional Genomics, Department of Biomedical Sciences, Cedars Sinai Medical Center, Los Angeles, CA, USA
[2] Applied Genomics, Computation and Translational Core, Cedars Sinai Cancer, Cedars Sinai Medical Center, Los Angeles, CA, USA
[3] Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA, USA
[4] Center for Functional Cancer Epigenetics, Dana-Farber Cancer Institute, Boston, MA, USA
[5] Women's Cancer Program, Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, CA, USA.
[6] Division of Gynecologic Oncology, Department of Obstetrics and Gynecology, Cedars-Sinai Medical Center, Los Angeles, CA, USA
[7] Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA, USA
[8] Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK
[9] Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK
[10] The Eli and Edythe L. Broad Institute, Cambridge, MA, USA
[11] Medical Research Council Integrative Epidemiology Unit, University of Bristol, Bristol, UK
[12] Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK

*These authors contributed equally to the study
# These authors co-directed the study
¶ Corresponding author: simon.gayther@cshs.org

**KEYWORDS:** Breast Cancer; Ovarian Cancer; GWAS; Genetic fine mapping; Chromosome Conformation Capture; Cell biology modeling; Differential gene expression

## ABSTRACT

Common genetic variation in a region on chromosome 15q26 confers susceptibility to breast and ovarian cancer. The P53 interacting gene RCCD1 in this region is a candidate susceptibility gene for both cancers. In this study, a colocalization analysis of breast and ovarian cancer case-control genetic association studies in over 145,000 and 146,000 controls fine mapped the shared association in this region to 17 pleiotropic credible causal risk variants ($P_{breast}$ < 1.16 x $10^{-14}$ and $P_{ovary}$ < 7.50 x $10^{-7}$).  These variants were strongly associated with the expression of RCCD1 in normal breast and ovarian tissues. Circular chromosome conformation capture (4C) analysis of *RCCD1* in breast and ovarian cancer cells identified similar patterns of *cis*-interaction and significant binding site enrichment for the *BRCA2* interacting gene *EMSY* ($P_{adjusted}$ = 9.24 x $10^{-6}$). The 4C analysis pinpointed a single 2kB RCCD1 cis-interaction that contained two of the 17 shared risk variants. RCCD1 *trans*-interacting regions mapped to previously identified genome wide significant (P < 5 x $10^{-8}$) breast cancer risk loci (1p34.2 and 3p14.1) and to the pleiotropic breast-ovarian cancer risk locus at chromosome 9q34.2. Stable overexpression of RCCD1 in breast and ovarian cancer precursor cells identified 13 and 11 differentially expressed genes (DEGs) respectively associated with breast and ovarian cancer risk at genome-wide significance ($P_{MAGMA}$ < 2.6 x $10^{-6}$ after Bonferroni correction). Eighty-two DEGs shared between breast and ovarian cancer were strongly enriched in TP53 (P = 9.9 x $10^{-4}$), Hippo (P = 2.51 x $10^{-3}$) and TNF signaling (P = 4.7 x $10^{-3}$) pathways.

## INTRODUCTION

Breast and ovarian cancers share common etiologies. Epidemiological studies have long established a role for sex steroid hormones (e.g. oral contraceptive use, hormone replacement therapy) influencing the risks of both cancer types[1,2] . Breast and ovarian cancers also share similar genetic liabilities suggesting a significant contribution of pleiotropy to risk of these cancers. In particular, the *BRCA1* and *BRCA2* genes are responsible for most families in which breast and ovarian cancers cluster together[3–7]. Pleiotropy is also an emerging feature of non-Mendelian forms of these and other cancers. Hundreds of common low penetrance susceptibility alleles have been identified for multiple cancer types using genome wide association studies (GWAS) and there is strong evidence that there is a shared genetic risk component for common alleles across cancers[8–11]. Common variants in several regions are associated with risks of both breast and ovarian cancer, suggesting that similar underlying biology may drive neoplastic development in both cancer types. For example, GWAS performed in breast and ovarian cancer populations have independently identified risk associations for each cancer at the 8q24.21 *CMYC* locus[12], the 5p15.33 *TERT* locus[13] and the 19p13.11 *ANKLE1/ABHD8* locus[14]. A recent meta-analysis combining GWAS data for both cancers identified shared (pleiotropic) risk regions that were not identified when each cancer was evaluated separately [15]. This study performed a GWAS meta-analysis that included 62,533 women with breast cancer and 60,976 controls, and 15,437 women with invasive epithelial ovarian cancer and 30,845 controls and identified two new risk loci at 9q31.1 and 15q26.1 associated with both breast and ovarian cancer were at genome wide levels of significance in the combined breast and ovarian cancer data[15].

The vast majority of disease associated variants lie in non-coding DNA regions and likely target tissue specific epigenomic elements that regulate the expression of a target susceptibility gene and/or critical gene expression networks involved in disease pathogenesis. Where the same genetic variants in a genomic region confer risk of two or more closely-related phenotypes, we

hypothesize that there is shared underlying biology driving disease pathogenesis for the different phenotypes.

The goal of the current study was to interrogate the functional mechanisms underlying 15q26.1 breast and ovarian cancer risk locus and establish if the shared genetic associations identified for these cancers reflect a shared biology. Using the integrated approach illustrated in **Figure 1**, we applied a combination of genetic fine-scale mapping and epigenomic annotation, cell biology modeling of candidate genes identified by quantitative risk variant-gene expression assocation analysis, and *cis-* and *trans*-interaction studies in breast and ovarian cancer precursor and cancer cells to identify shared networks of differential gene expression and transcription factor binding linked to genetic risk.

## MATERIALS AND METHODS

***Genetic fine mapping analysis:*** SNP genotype and sample quality control, ancestry inference, imputation, genome-wide association and meta-analysis procedures for the breast and ovarian cancer GWAS meta-analysis data sets have been described previously[16,17]. All analyses were based on individuals of European ancestry and used 1000 Genomes Phase 3 (Version 5)-imputed or genotyped SNPs[18]. We harmonized effect and non-effect alleles and effect size estimate (beta coefficient) signs across the GWAS meta-analysis summary statistics data sets included in this study. We focused all analyses on 3,609 variants with minor allele frequency > 1% and imputation quality > 0.4 in each data set in the chr15:91009215-92009215 genomic region. Colocalization analysis was performed using the hypothesis prioritization in multi-trait colocalization (HyPrColoc) R package[19]. Expression Quantitative Trait Loci (eQTL) annotation of the credible set SNPs was conducted using the Qtlizer web application[20] for the Genotype Tissue Expression (GTEx) project version 8 database [21] and using an online eQTL browser [22] for The Cancer Genome Atlas (TCGA)[23–25].

***Cell culture:*** The FT282 and FT246 cell lines were established from fresh normal human fallopian tube (FT) tissue gifted by Dr. R. Drapkin[26]. These FT cell lines were maintained in DMEM-F12 medium (Gibco, Invitrogen, Carlsbad) supplemented with 10% fetal bovine serum (Gibco). Normal mammary epithelial cells (MCF10A, MCF12A) were cultured in MEBM medium supplemented with 100 ng/ml cholera toxin. Ovarian cancer cell line UWB 1.289 was purchased from ATCC and was cultured in a mixture (1:1) of 50% RPMI and 50% MEGM media with 3% FBS. The breast cancer cell line MCF7 was maintained in DMEM medium with 10% FBS. BT549 and Kuramochi were cultured in RPMI-1640 media supplemented with 10% FBS. All cell lines were thawed and allowed to recover for one day before expansion. All cell lines were passaged and plated onto 150mm plates. Cells were harvested at 80% confluency to be fixed, pelleted and snap frozen. Two biological replicates were included for each cell line.

***Preparation of 4C-Seq library:*** 4C-Seq libraries were prepared using a protocol that has been modified from our previous method[27]. In brief, cells were single-cell suspended and their chromatin was crosslinked with 1% formaldehyde for 10 min at room temperature (21°C). Crosslinked cells were resuspended in a pre-chilled lysis buffer (10 mM Tris HCl, 10 mM NaCl, 0.2% NP-40) with fresh protease inhibitors (Millipore Sigma). The lysate was then digested with the primary restriction enzyme, DpnII (10,000 units/ml, New England Biolabs), and incubated overnight at 37°C. Next, DpnII-digested DNA was subjected to proximity ligation using 1X T4 DNA ligase (NEB), 1% TritonX-100, 0.1% BSA (NEB) and 4000U of T4 DNA ligase. Ligation was carried out overnight at 16°C followed by overnight cross-link removal with 20mg/ml Proteinase K (Qiagen). The samples were purified using phenol-chloroform and ethanol precipitated

resulting in 3C libraries. The 3C libraries were subjected to a second restriction enzyme digestion using Csp6i (Life Technologies), followed by another proximity ligation using T4 DNA ligase. For each viewpoint, a total of 3.2 μg of the resulting 4C templates was used to carry out a scale-up inverse PCR, of which 16 reactions (200 ng in each) were amplified for 29 cycles and purified using Macherey Nagel Nucleospin Gel and PCR Purification kit (Takara Bio). Sequencing libraries were made of the 4C PCR products using Thruplex DNA-seq kit (Takara Bio). 4C libraries were subjected to Agencourt AMPure XP Bead cleanup (Beckman Coulter) using a bead-to-DNA ratio of 1:1 before high-throughput sequencing on the MiniSeq system (Illumina). Libraries were sequenced 1 X 75bp for a minimum coverage of 2 million read depth. Two technical replicates of each Dpn-Csp6I 3C library were made for each 4 cell lines.

***Primer design for 4C-Seq:*** The inverse primers were designed based on a viewpoint region. UCSC Genome Browser (build GRCh37/hg19) was used to locate the region of interest. Upon loading DpnII and Csp6I tracks, DpnII restriction sites flanking the region of interest were identified and the sequence between the nearest DpnII and Csp6I restriction sites were selected as the viewpoint region. Based on this region, 4C primers were designed with the following settings: optimal primer melting temperature of 57 °C (minimum 55 °C and maximum 60 °C); GC content between 40 and 60%.  For DpnII-Csp6I libraries, 2 baits were designed as follows: 1) forward- TAGCTAGCACTGACAAGATC, reverse- GGATGTCCTCTGTTCTGAGT; and 2) forward- GAGGGCATAGGGTCAAGATC, reverse- CTTCAACCAGGCTCAGCT.

***4C cis-analysis:*** For 4C raw sequencing data, cutadapt (version 2.0) was used to remove PCR primers (forward and reverse) and Illumina adapter from the reads, Burrows-Wheeler Aligner (bwa version 0.7.15-r1140) was employed to align raw sequencing files to the human genome (hg19). Samtools (version 1.9) was used to convert sequence alignment files (sam) to their binary version (bam). To determine statistically significant *cis*-interactions we used Bioconductor's R package r3Cseq to count the reads per restriction fragment.  Using technical replicates from each 4C library, the reads were normalized to RPM and interactions were obtained around 100kb from the view-point (4C bait), interactions closer than 3kb from the bait (promoter of RCCD1, coordinate 91496106 hg19) were removed. *Cis*-Interacting regions (CIRs) that were identified from biological replicates for each cell line, with a q-value < 0.01. Ovarian cancer CIRs were then merged based on the overlap of UWB1.289 and Kuramochi interactions. Breast CIRs were computed as the merged set of interactions in MCF7 and BT49 cell lines. Shared CIRs were identified by overlapping the breast and ovarian CIRs to obtain a unified *CIR* set with 100% of the interactions being shared across all four cell lines.

***4C trans-analysis:*** For *trans*-interactions analysis raw coverage signal from sequencing files was inputted for FDR calculation, where a threshold of 0.01 was determined based on 100 permutations for each chromosome. From these permutation calculations, any window sizes that exceed 500 unique fragment ends were identified as *trans*- interactions[28].  All Interaction regions (*cis* and *trans*) were processed into bed files for plotting and visualization on UCSC Genome Browser as custom tracks.

***ChIP-Sequencing:*** Chromatin Immuno-Precipitation sequencing (ChIP-Seq) was performed as previously reported [29]. Cells were fixed in 1% formaldehyde for 10 minutes and quenched with 2M glycine. Cells were harvested, lysed in buffer, and sonicated using the Covaris E220 evolution Focused-Ultrasonicator (Covaris). Antibodies used included H3K27ac (Diagenode), CTCF (Active Motif), H3K4me1 (abcam) and H3K4me3 (EMD Millipore). The H3k27ac antibody was incubated with 5 μg of chromatin while CTCF, H3k4me1 and H3k4me3 were incubated with

25ug of chromatin at 4°C overnight. Blocked magnetic Dynabeads (Life Technologies) were then added to the antibody-lysate conjugates and incubated at 4°C for 4 hours with rotation. Beads were then washed with RIPA buffer and treated with RNaseA and proteinase K (Qiagen). DNA was eluted from the beads in Tris-EDTA buffer and cleaned up using the Nucleospin Gel and PCR Purification kit (Macherey-Nagel). For every cell line, two independent immunoprecipitations and one input sample were submitted for next-generation sequencing on a HiSeq 5000 (Fulgent Genetics; Temple City, CA). All ChIP-Seq data was processed based on the ENCODE histone ChIP-Seq pipeline[30] with modifications on the peak calling algorithm, to retain peaks with P< 1 x 10^{-9}.

***Functional Annotation of Genome:*** The Statepaintr R package translates epigenomic files into chromatin segmentations using a rule-based-decision matrix combined with an abstraction layer with functional categories[31]. StatePaintR uses the deposited ChIP-Seq files of all uniquely mapping segment boundaries from the start and end coordinates of every peak then evaluates the presence or absence of each functional category. Next the package assesses overlaps of each segment to produce a boolean value. StatePaintR enables rank scoring of all states, allowing prioritization for non-coding variant annotation. ChIP-Seq data of the histone modifications (H3K27Ac-active enhancer/promoter), H3K4me3 (active promoter), H3K4me1(poised promoter/enhancer) and architectural complex protein (CTCF) was integrated for combinatorial patterning of overall chromatin states using Statepaintr in all precursor normal (MCF10A, FT246) and cancer cell lines (MCF7, BT549, UWB1.289 and Kuramochi). Bed files from these Statepaintr chromatin annotations were visualized on UCSC genome browser and IGV viewer.

***Transcription Factor Binding Site (TFBS) and protein interaction analysis:*** We utilized PERFECTOS-APE (PrEdict Regulatory Functional Effect of SNPs by Approximate P-value Estimation), a software built to identify TFs whose *in silico* binding sites are predicted to be significantly affected by a given nucleotide substitution[32]. This tool uses basic and dinucleotide Position Weight Matrices (PWMs) as TFBS and estimates the statistical significance (P value) of a predicted TFBS overlapping a SNP. Next, the method determines if the TFBS binding P values calculated for different homologous alleles differ significantly, such that extremely small or large ratios of P values are indicative of a TFBS existing for only one allele. TFBS motif variant prediction was conducted on the SNPs which intersected PCI/CCIs.

TFBS enrichment analysis was performed using ReMap 2020[33], a curated database of 1135 transcription factors from 5798 ChIP-Seq experiments. To date this is the largest catalogue of experimentally derived regulatory regions by integrating all available DNA-binding assays generating 165M binding regions across the human genome. Genomic positions of the 4C interaction regions were crossed against the entire catalog of *in vitro* TF binding peaks. The collective chromosomal positions for all CCIs were compared to the ReMAP catalog of TF binding peaks to calculate statistical enrichment for TFs. First, *trans*-interacting regions were subdivided into 1kb peaks, in which these chromosomal positions were used as a query for the TFBS enrichment analysis. In this analysis a TF peak was considered for enrichment when the TF ChIP-Seq peak in the catalogue had 100% overlap within the 1kb query peak. The default setting for overlap is 10% but we selected 100% as the most stringent filter. For each queried peak, six genomic regions were shuffled to create a theoretical distribution. The p-value represents the probability to observe an effect at least as extreme as the result, under null hypothesis, which can be interpreted as an estimation of the false positive rate (FPR) and it is computed using the Poisson distribution (and validated empirically with randomized query

regions). The q-value is the correction of the p-value to take into account the multiple testing due to the fact that the query is compared to each TF of the remap catalogue and it is an estimation of the false discovery rate (FDR), utilizing the Benjamini-Yekutieli method. Using ggplot2, the TFs enriched in CCIs were plotted by their Q Significance (-log10 of the Q-value) and frequency in which they occurred in the CCIs. This analysis was also applied to all PCIs to look for TF specifically enriched in these regions. TFs enriched in breast *trans*-interactions were evaluated using the union set of *trans*-interacting regions found in the breast cancer cell lines (MCF7 and BT549). Similarly the union set of *trans*-interactions in the ovarian cancer cell lines (UWB 1.289 and Kuramochi) were used as the regions queried in the ReMap dataset.

***Integrative modelling of chromatin landscapes and cistromes:*** We utilized Lisa (http://lisa.cistrome.org/) to predict the transcriptional regulators (TRs) of differentially expressed gene (DEG) sets. Briefly, Lisa uses publicly deposited histone mark ChIP-seq and chromatin accessibility profiles to construct a chromatin model related to the regulation of the DEG sets from breast and ovarian precursor models of RCCD1 overexpression. Using TR ChIP-seq peaks, this software analyses chromatin models to find the relevant TRs. to prioritize the candidate TRs, Lisa compares the predicted effects on the query and background gene sets using the one-sided Wilcoxon rank-sum test. We used two DEG sets as input for the model: set 1 was the DEG list from FT282$^{RCCD1+}$(n= top 500; P<0,05; FDR<0.01) and set 2 was the DEG list in MCF12A$^{RCCD1+}$ (n = 291; P<0,05; FDR<0.01). For the background gene set, we used all genes annotated on Gencode v.35. For the methods selection, ALL criterion was chosen which corresponds to both ISD-RP for motif and ChIP-Seq and TF ChIP-Seq Peak-RP.

***Luciferase reporter construct and activity assay :*** Reporter vectors were constructed in the pGL3-Promoter luciferase vector (Promega; GenBank Accession # U47298). The desired enhancer regions spanning the 4C interaction (chr15:91,504,893-91,507,207) were amplified from human gDNA by PCR using primer set F1/R3. pGL3-Promoter was linearized with XhoI restriction enzyme. Resulting two linear fragments was assembled using NEBuilder HiFi DNA Assembly kit (New England Biolab Inc.) as recommended by manufacturer to create pGL3-promoter-RCCD1 enhancer (pSD1) with an enhancer element containing risk alleles A for rs8028409 and T for rs8037137. pSD2, pSD3 and pSD4 containing one or both risk allele mutations (A to C and T to A for rs8028409 and rs8037137 respectively) were created using pSD1 as template to generate fragments for HiFi DNA assembly reaction introducing mutation at desired site/s . Primers used to generate fragments and introduce mutation/s are listed in **Supplementary Table S1**. Primer sets F1/R1, F2/R3, PGL3-F/PGL3-R generated 3 fragments for pSD2 plasmid assembly, primer sets F1/R2, F3/R3, PGL3-F/PGL3-R generated 3 fragments for pSD3 plasmid assembly and primer sets F1/R1, F2/R2, F3/R3, PGL3-F/PGL3-R generated 4 fragments for pSD4 plasmid assembly. All plasmid constructs were confirmed by Sanger sequencing. MCF7 and UWB1.289 cells were seeded in 6-well plates at a density of 0.3 million cells with respective media as mentioned above. Transfection was carried out when cells reached 60-70% confluency. 1.9μg firefly luciferase reporter plasmids (Promega; Accession # U47298) and 100ng of Renilla luciferase control plasmid (Promega; Accession # AF025848) was mixed with 3μL of BioT transfection reagent (Bioland Scientific LLC) and transfected the cell lines. Cells were grown in their corresponding media for 48 h. Transfected cells were assayed after 48 h (as indicated for firefly and Renilla luciferase activity using the Dual-Luciferase Reporter Assay System (Promega) in a GloMax® Explorer System (GM3500).

***RCCD1 overexpression plasmid construction and preparation***: Overexpression of RCCD1 was accomplished using a sequence validated RCCD1 cDNA clone (HsCD00936904) This

plasmid along with a control plasmid containing CAGG sequence were purchased from the DNASU repository (Biodesign Institute, Arizona State University, Tempe, AZ, USA). HEK 293T cells were plated onto 6cm plates and used to propagate the virus (RCCD1 overexpression plasmid and empty control plasmid). Twenty-four hours post plating, 30µl of BioT (Bioland Scientific, CA, USA) was added to 100 µL serum free media containing a total of 2µg per well of the desired RCCD1 expression construct or control empty vector plasmid. The BioT transfection and DNA solutions were then combined following manufacturer recommendations. Following 48 h of culture, virus was collected and prepared according to the manufacturer's recommendations. Recipient cells (FT282 and MCF10A) were transduced with viral vectors. After 72h, transduction efficiency was evaluated and estimated by presence of fluorescence reporter from the plasmid.

***RNA isolation, library preparation and sequencing:*** Total RNA was isolated using RNA isolation kit (Machery Nagel) as per manufacturer's instructions. RNA concentration and quality were determined by NanoDrop, Qubit and Agilent Bioanalyzer 2100. Total RNA were used as templates for cDNA libraries and were prepared as poly-A selected libraries using the Truseq stranded mRNA protocol (Illumina, San Diego, CA) and then sequenced on the Illumina Novaseq platform at AGCT Core (Los Angeles, CA) at 2 × 150 bp paired reads for a minimum of 35 million reads.

***RNA-sequencing analysis:*** Raw reads obtained from RNA-Seq were aligned to the transcriptome using STAR (version 2.5.0)[34] / RSEM (version 1.2.25)[35] with default parameters, using a custom human GRCh38 transcriptome reference downloaded from http://www.gencodegenes.org, containing all protein coding and long non-coding RNA genes based on human GENCODE version 23 annotation. All data analysis was performed using R and Bioconductor packages therein. For normalization purposes read count was transformed to TPM. Bioconductor's package DESeq2 was used to compare the RCCD1 overexpression lines to control plasmid groups, performing differential expression analysis based on a model using the negative binomial distribution[36].

***Cell proliferation assay:*** The human cell lines FT282^p53R175H (sgOR10A (Control ox), RCCD1 ox, sgRCCD1 (RCCD1 KO), CAGG ox (Control ox) and normal human breast epithelial cell MCF12A^p53R175H with CAGG ox, RCCD1 ox, RCCD1 KO and Control KO were grown in Dulbecco's modified Eagle's medium/Ham's F12 medium supplemented with 10% FBS and Dulbecco's modified Eagle's medium and Ham's F12 medium, 20 ng/ml Human epidermal growth factor, 100 ng/ml cholera toxin, 0.01 mg/ml bovine insulin and 500 ng/ml hydrocortisone, 95%; horse serum, 5% respectively. Cells were seeded in triplicate at the cell density of 10,000 cells/well in 96-well plate overnight in 200 ul respective media. After overnight incubation, equal volume of the 2X RealTime-Glo™ MT Cell Viability Reagent (Catalog # G9711)(Promega)) were added according to the manufacturer's protocol. We measure the luminescence using the GloMax Microplate reader (Promega). To determine cell growth curve, we continue measuring luminescence at various times over the desired time course (e.g., 1hr, 2hr, 3hr, 4hr, 5hr, 8hr). We plot the luminescence versus different time points using linear regression curve fit. A one way ANOVA was used for statistical analysis of proliferation assays.

**RESULTS**

***Genetic associations with breast and ovarian cancer risk in the 15q26.1 region colocalize to 17 candidate causal risk variants:*** We performed colocalization analysis of the previously

identified breast and ovarian cancer susceptibility signal in the 15q26.1 region[15] using summary genetic association data for 3,609 SNPs in the 1 Mb interval chr15:91009215-92009215 from the Breast Cancer Association Consortium (122,977 overall breast cancer cases and 105,974 controls)[16] and the Ovarian Cancer Association Consortium (22,406 invasive epithelial ovarian cancer cases and 40,941 controls)[17]. Colocalization analysis supported two conclusions (**Figure 2A**): First, that there is a confirmed association with breast and ovarian cancer risk for common genetic variants in the chr15:91009215-92009215 interval (regional posterior probability of association (PPA) > 99%);  second, that the alignment of cross-cancer associations in the region was strongly compatible (posterior probability of colocalization = 80%) with the presence of a single underlying causal signal that is associated with both breast and ovarian cancer risk.

Colocalization analysis fine-mapped the pleiotropic breast and ovarian cancer signal in the region to a credible set of 17 SNPs that together explained > 99% of the shared association in the region (i.e., > 99% of the posterior probability of colocalization; **Figure 2B; Supplementary Table S2**). This joint cross-cancer fine mapping represents a 5 SNP (or 23%) improvement over a recently published large-scale, breast cancer-specific fine mapping effort that refined the breast cancer signal in the region down to 22 SNPs[37].  In a QTL based analysis of 396 breast and 167 ovarian tissues in the genotype-tissue expression (GTEx) project all 17 risk associated SNPs in the credible causal set were associated with the expression of two genes: *RCCD1* ($P_{breast}$ < 1.7 x 10$^{-21}$ and $P_{ovary}$ < 8.3 x 10$^{-6}$) and the long non-coding RNA *PRC1-AS1* ($P_{breast}$ < 7.5 x 10$^{-9}$ and $P_{ovary}$ < 4.2 x 10$^{-9}$) (**Supplementary Table S3**)[21]. The alleles associated with increased risk of breast and ovarian cancer were associated with decreased *RCCD1* expression and increased *PRC1-AS1* expression in both breast and ovarian tissues **(Figure 2C)**.  The original lead SNP at this locus, rs8037137, was among the 17 SNPs in the credible causal set; gene expression associations for *RCCD1* and *PRC1-AS1* with rs8037137 genotypes are shown in **Figure 2C**. We also found that  rs8028409 was the top eQTL for *RCCD1* expression in TCGA breast tumors (P = 3.2 x 10$^{-7}$). As with our findings in GTEX, in TCGA data the A (risk) allele is associated with decreased RCCD1 expression. No expression associations were identified for these SNPs in ovarian tumors from TCGA[22].

***RCCD1 interacts with shared epigenomic targets and two colocalized risk variants in breast and ovarian cancer cells:*** We hypothesized that *RCCD1* is the target susceptibility gene of genetic risk allele(s) at 15q26.1 regulated by the same genetic and epigenomic mechanisms in breast and ovarian cancers. To evaluate this, we first integrated the 17 risk variants with epigenomic features from profiling breast and ovarian precursor and cancer cell lines to identify putative regulatory targets of these variants **(Table 1)**. This identified intersections between 9/17 risk variants and different epigenomic features including active enhancers, poised enhancers, active regions and CTCF-bound regions, and with numerous transcription factors binding site (TFBS) motifs for 14/17 risk variants (**Table1; Supplementary Tables S4 and S5).** We then used circular chromosome conformation capture (4C) analysis to identify which, if any, of these regions physically interacts with *RCCD1*. The promoter of *RCCD1* was used to bait interacting regions in the breast cancer cell lines MCF7 and BT549; and in the high grade serous ovarian cancer (HGSOC) cell lines UWB1.289 and Kuramochi.  We identified a total of 65 *cis*-interacting regions (CIRs) with *RCCD1* including the same 9 CIRs in all four cell lines (**Figure 3A; Supplementary Table S6**). One of these nine CIRs in a region of approximately 2kb (chr15:91504892-91507207) harbored 2/17 credible causal risk variants: rs8028409 ($P_{breast\ risk}$ = 4.73 x 10$^{-15}$ and $P_{ovaran\ risk}$ = 3.26 x 10$^{-7}$) and rs8037137 ($P_{breast\ risk}$ = 6.52 x 10$^{-15}$ and $P_{ovarian\ risk}$ = 4.61 x 10$^{-7}$). Neither of these SNPs coincided with putative enhancers, promoters or CTCF-bound regions, but were predicted *in silico* to disrupt TFBS motifs for 16

factors (**Figure 3A; Table1; Supplementary Table S4**). We also used ReMap data to evaluate all local regions that interact with *RCCD1* for TFBS motif enrichment based on experimentally observed binding sites identified with ChIP-seq. CIRs were most significantly enriched for TFBSs with known roles in cancer including the BRCA2 interacting TF EMSY ($P_{adjusted}$ = 9.24 x 10$^{-6}$), the STAT1 and NF-κB regulator PIAS1 ($P_{adjusted}$ = 6.74 x 10$^{-3}$) and ZNF316 ($P_{adjusted}$ = 6.74 x 10$^{-3}$) **(Figure 3B)**.

We tested the functional effects of different allelic combinations of rs8028409 and rs8037137 on RCCD1 expression using luciferase reporter assays in breast (MCF7) and ovarian (UWB1.289) cancer cells. Four constructs each containing different allelic combinations of rs8028409 (A-risk/T-protective) and rs8037137 (T-risk/C-protective) were transfected into MCF7 and UWB1.289 cells. Constructs containing only the risk alleles of rs8028409 and rs8037137 were associated with the lowest RCCD1 expression in both breast and ovarian cancer cells. Constructs expressing only protective alleles were associated with the highest RCCD1 expression ($P_{breast}$ <0.0001 and $P_{ovarian}$ = 0.02 respectively) while constructs containing different combinations of protective/risk allele for each SNP were associated with intermediate RCCD1 expression **(Figure 3C)**.

***RCCD1 4C trans interactions in breast and ovarian cancer risk are enriched for disease associated transcription factor binding sites:*** We used 4C data in the breast and ovarian cancer cells to identify long-range *trans* interacting regions (TIRs) between RCCD1 and regulatory elements throughout the genome. In total, we found 165 TIRs across all four cancer cell lines with 30 and 46 TIs respectively identified in MCF7 and BT549 breast cancer cells, 66 and 23 TIRs respectively identified in UWB1.289 and Kuramochi ovarian cancer cells and three TIRs shared across all four cell lines (**Supplementary Table S7**). Five other TIRs were shared between both breast cancer cell lines, and nine other TIRs were shared between both ovarian cancer cell lines **(Figure 4A)**. We performed enrichment analyses of these TIRs with respect to TFBS motifs from ReMap data. There were significant enrichments for many TFBSs shown to play a role in cancer development generally and breast and ovarian cancer biology specifically. For breast cancer TIRs, the most significantly enriched TFBS motifs included the transcriptional regulator chromobox protein (CBX1), the P53 degradation E3 ligase protein (RNF2), the DNA double-strand break repair protein RAD21 and the retinoblastoma protein (RB1)**(Figure 4C; Supplementary Table S8)**. For ovarian cancer, the most significantly enriched TFBS motifs included the MYC Associated Factor X (MAX) protein, the stromal antigen 1 protein(STAG1) and bromodomain transcriptional regulatory (BRD4) **(Figure 4D; Supplementary Table S9)**.

***Phenotypic and molecular analysis in breast and ovarian cancer precursor cells support a role for RCCD1 in early stage neoplastic development:*** Expression QTL analyses suggest that the risk alleles of variants at 15q26.1 are associated with a decrease in *RCCD1* expression (**Figure 2C**); but in breast and ovarian tumors and in cancer cell lines the evidence suggests that *RCCD1* is overexpressed during tumor development (**Supplementary Figure 1A**)[23,24]. Therefore, we tested the effects of modulating *RCCD1* expression by both overexpressing and knocking down the gene separately in models of breast (MCF12A) and ovarian (FT282) cancer precursor cells. We first generated single cell clonal models of *RCCD1* overexpression (MCF12A$^{RCCD1+}$ and FT282$^{RCCD1+}$) and confirmed overexpression of both *RCCD1* transcript and protein (**Supplementary Figure 1B**). MCF12A$^{RCCD1+}$ and FT282$^{RCCD1+}$ cells were significantly more proliferative than MCF12A and FT282 cells (P = 1.7 X 10$^{-3}$ and 5.1 X 10$^{-3}$ respectively) (**Figure 5A & B**). Next, we used CRISPR/Cas9 to knockdown *RCCD1* expression (MCF12A$^{RCCD1-}$ and FT282$^{RCCD1-}$), derived single cell clones and confirmed loss of *RCCD1* transcript and protein expression (**Supplementary Figure 1C**). *RCCD1* knockdown induced a

strong growth suppression in MCF12A$^{RCCD1-}$ cells (P = 4.5 X 10$^{-3}$ ; **Figure 5A**) but an increase in proliferation in FT282$^{RCCD1-}$ cells (P = 1.1 X 10$^{-3}$; **Figure 5B**).

Because the phenotypic effects associated with RCCD1 overexpression were similar in MCF12A and FT282 cells, we chose to perform whole transcriptome profiling of MCF12A$^{RCCD1+}$ and FT282$^{RCCD1+}$ clones by RNA-seq. We identified 291 differentially expressed genes (DEGs) (|log2FC | >0.5, P$_{adjusted}$< 0.05) between MCF12A and MCF12A$^{RCCD1+}$ cells of which 138 genes were up-regulated and 153 genes were down-regulated **(Figure 5C, Supplementary Table S10)**. Comparing FT282 with FT282$^{RCCD1+}$ cells, we identified 2,469 DEGs (| log2FC |> 0.5, p$_{adjusted}$< 0.05), of which 1,084 genes were up-regulated and 1,385 genes were down-regulated in **(Figure 5D, Supplementary Table S10).** Eighty-two DEGs were shared between breast and ovarian cancer precursor cell line models. All DEGs were annotated using the KEGG database to identify biologically relevant pathways (**Supplementary Table S11-13**). In MCF12A$^{RCCD1+}$ models we identified significant enrichments for pathways involved in PPAR (peroxisome proliferator-activated receptor) signaling (P = 0.01), the pro-inflammatory cytokine IL-17 (interleukin 17A) signaling (P = 0.02) and steroid hormone biosynthesis (P = 0.03) **(Supplementary Table S11)**. In FT282$^{RCCD1+}$ cells we identified significant enrichments for pathways involved in Ras (P = 9.4 X 10$^{-5}$), Rap1 (P = 9.83 x 10$^{-5}$) and p53 signaling pathways (P = 9.9 x 10$^{-5}$) **(Supplementary Table S12).** Pathway analyses of the 82 DEGs that were shared between breast and ovarian *RCCD1* overexpression models identified significant enrichment of genes in the TP53 (P = 9.9 x 10$^{-4}$), Hippo (P = 2.51 x 10$^{-3}$) and TNF signaling (P = 4.65 x 10$^{-3}$) pathways **(Supplementary Table S13)**.

***Evidence that RCCD1 interacts with and operates in the TP53 pathway in breast and ovarian cancer:*** We further explored the putative role of RCCD1 in the TP53 pathway. RCCD1 is a known partner of the histone H3K36 demethylase KDM8 (JMJD5)[38] which interacts with the DNA-binding domain of TP53 to negatively regulate its activity (**Figure 6A**). This finding was mainly driven by enrichment for TP53-associated DEGs identified in FT282$^{RCCD1+}$ models. Several DEGs identified in these models are closely related partners of TP53 signaling including: MDM2, a critical negative regulator of P53; the cyclin-dependent kinase inhibitor CDKN2A; and CCND1 and CCND2, two regulators of cyclin-dependent kinases (**Figure 6B**). In two independent clones of RCCD1 knockdown in FT282 cells stably expressing a functional mutation of TP53 (R175H), mutant TP53 was significantly depleted (**Figure 6C**). We analyzed RCCD1 expression with respect to somatic, functional TP53 mutations in primary breast and ovarian cancers using TCGA data[39]. Both breast and ovarian cancers with increased RCCD1 expression were significantly more likely to harbor functional TP53 mutations than tumors with low RCCD1 expression ($P_{breast}$ = 8 x 10$^{-12}$ and $P_{ovary}$ = 4 x 10$^{-3}$) (**Figure 6D**).

***Differentially expressed genes identified in RCCD1 overexpression models are associated with breast and ovarian risk alleles:*** We used GWAS data generated by the Breast and Ovarian Cancer Association Consortia to map common variants across the genome to genes and performed gene-level association analysis using MAGMA[40,41]. This identified 281 genes associated with breast cancer risk and 42 genes associated with ovarian cancer risk at genome-wide significance ($P_{MAGMA}$ < 2.62 x 10$^{-6}$; threshold adjusted for testing 19,100 genes). Thirteen of the 281 breast cancer associated genes and 11/42 ovarian cancer associated genes from MAGMA showed significant gene expression perturbations (P$_{adjusted}$ < 0.05) in MCF12A$^{RCCD1+}$ cells and FT282$^{RCCD1+}$ cells, respectively (counts exclude *RCCD1*; **Supplementary Table S14**). There was a |log2FC| > 0.5 fold change in expression of *TRIM31* in the MHC region and *CDYL2* and *NOTCH2NL* in MCF12A$^{RCCD1+}$ cells and *ANKLE1, BNC2, HOXB3*, and *LRRC37A* in FT282$^{RCCD1+}$ cells. Several differentially expressed genes in

FT282[RCCD1+] cells were located at pleiotropic breast and ovarian cancer risk loci, such as *ANKLE1* and *ANO8* at the 19p13.11 locus[14]; *LRRC37A, KANSL1, ARL17A,* and *PLEKHM1* at the 17q21.31 locus[42,43], and *TERT* at the 5p15.33 locus [13]. The *HOXB* cluster and *BNC2* have previously been identified as having specific functional roles in mediating ovarian cancer risk[44,45]. *MLLT10* was differentially expressed in both MCF12A[RCCD1+] and FT282[RCCD1+] cells ($P_{adjusted}$ < 0.05) and was significantly associated with both breast and ovarian cancer risk in the MAGMA analyses ($P_{MAGMA}$ < 2.62 x 10^-6).

***Integrating transcriptomic and RCCD1 interactome data validates 4C trans-interactions identified in breast and ovarian cancers:*** To validate the functional significance of RCCD1-TIRs in breast and ovarian cancers, we integrated these regions with the differential expressed genes identified after overexpression of *RCCD1* in MCF12A and FT282 cells. We identified six DEGs(PCBP3, BSCL2, COL18A1, NUP210 ,POLR2G and PARP10) in MCF12A[RCCD1+] cells that overlap TIRs in breast cancer cell lines and eighteen DEGs (EEF1G, BCDIN3D, TUBD1,PPM1D, RNF43, BCAS4, TSHZ2, DOK5 DOCK3, MKRN2OS, HRH1, RPL26L1, VDAC1, SKP1, ZDHHC11, RPS10, PRKRIP1, RASA4) that overlap TIRs in ovarian cancer cell lines**.** This enrichment of DEGs in RCCD1-TIRs in breast and ovarian cancer cells was significantly more than expected by chance based on the simulation of more than a thousand random permutations (Wilcoxon P = 0.049; **Figure 4B)**.

Finally, we used the online tool LISA (Cistrome) to evaluate the TFs and chromatin regulators that may be responsible for the perturbation of DEGs found in MCF12A[RCCD1+] and FT282[RCCD1+] cells . In MCF12A[RCCD1+] cells 13 out of the top 50 TFs overlapped ReMAP and Cistrome predictions (including YY1, HDAC2, EGR1, STAG1, HEXIM1, MBD3, ELF1 and CHD2) **(Supplementary Table S15)** and in FT282[RCCD1+] cells 11 of the top 50 TFs overlapped ReMAP and Cistrome predictions from the DEGs identified in FT282[RCCD1+] cells (including NR2F1, SP1, ERG, ESR1, RUNX1 and HIF1A) **(Supplementary Table S16).** We also identified several TFs that were shared between breast and ovarian cancer cells in both our trans TF enrichment and Cistrome analyses including MYC, RAD21, CTCF, MAX and AR.


**DISCUSSION**

We have shown that the genetic architecture of common variant risk associations in the 15q26.1 region is the same in both breast and ovarian cancer suggesting that similar functional mechanisms at this locus drive disease development for both cancer types. We focused on RCCD1 (regulator of chromosome condensation domain-containing protein 1) in this region as the most likely breast and ovarian cancer susceptibility gene and target of risk variation following expression quantitative trait locus (eQTL) analysis which found significant associations between risk genotypes and *RCCD1* expression in normal breast and ovarian tissues. However, we cannot rule out that other genes in the region may be the target susceptibility gene operating in collaboration with or independently from *RCCD1* and our eQTL analyses also identified significant associations with the lncRNA *PRC1-AS1* (Protein Regulator of cytokinesis 1-antisense 1) in normal breast and ovarian tissues. *PRC1-AS1* overlaps the *PRC1* gene, and as such may play a role in regulating *PRC1* transcript expression. While we did not identify a significant eQTL for *PRC1* in this analysis, a previous transcriptome wide association study (TWAS) has identified a significant splice-QTL for *PRC1* in breast tumors which fully explained the GWAS signal in ovarian cancer at the 15q26.1[46]. In addition, recent studies have suggested that *PRC1* overexpression may be associated with the development of multiple tumor types[47–49]

and in one study *PRC1* mRNA and protein expression were upregulated in high‑grade serous ovarian cancers, particularly in tumors from patients without *BRCA1/BRCA2* mutations[50].

*RCCD1* is a highly plausible breast and ovarian cancer susceptibility gene. By partnering with the histone H3K36 demethylase KDM8 (JMJD5)[38],RCCD1 interacts with the DNA-binding domain of TP53 to negatively regulate its activity, and is an important complex for spindle organization, chromosome segregation and accurate mitotic division [38]. Germline pathogenic mutations in TP53 are responsible for the autosomal dominant Li-Fraumeni syndrome (LFS), which is most commonly associated with early onset sarcoma, and malignancies of the brain, breast and adrenal glands[51]. The risk of breast cancer in women with LFS is estimated to be about 50% by age 60[52]. Somatic P53 mutations are also essential for HGSOC development, are common aberrations in triple negative breast cancer[25,24], and are associated with *BRCA1* and *BRCA2* mutant cancers[53,54]. In this study, we provide functional evidence that RCCD1 interacts with TP53 in early stage neoplastic development. Knockout of wild-type *RCCD1* expression in fallopian tube cells expressing the functional *TP53* hotspot mutation (R175H) resulted in depletion of TP53 protein expression. Consistent with this, primary breast and ovarian tumors analyzed by TCGA [39] with low *RCCD1* expression are significantly less likely to also harbor pathogenic TP53 mutations than tumors with higher *RCCD1* expression.

The interaction with TP53 suggests that *RCCD1* plays an important role in cell cycle regulation and DNA damage repair[55]. Other data from this study support this hypothesis. The same *RCCD1 cis*-interacting regions identified in both breast and ovarian cancer cell lines were significantly enriched for *EMSY* binding site motifs. *EMSY* is a transcriptional repressor that directly interacts with the DNA double strand break repair protein BRCA2, a well characterized breast and ovarian cancer susceptibility gene. Thus, EMSY likely dysregulates DNA damage repair to induce genomic instability. Immunofluorescence analysis in the normal mammary epithelial cell line MCF-10A and breast cancer cell line MCF-7 shows that EMSY localizes to the nucleus after exposure to ionizing radiation; there is also direct evidence that EMSY locates to sites of DNA lesions after DNA damage[56,57]. In another study, overexpression of *EMSY* disrupted the BRCA2/RAD51 pathway in response to DNA-damage, leading to the hypothesis that amplification of EMSY mimics BRCA2 deficiency by overriding RPA and PALB2, crippling the BRCA2/RAD51 complex at site of DNA damage[58].

*RCCD1 trans*-4C interaction analyses identified genomic regions that were highly significantly enriched in motifs for TFs that are known to be involved in breast and/or ovarian cancer biology including in the three shared regions identified in four breast and ovarian cancer cell lines. This included *RAD21* an essential gene that functions in the DNA double strand break repair pathway, further supporting the putative functional role of *RCCD1* in DNA repair. A recent study has shown that mutant p53 stimulates cell invasion through its interaction with RAD21 in ovarian cancer cells, while *RAD21* expression is associated with poor prognosis in *BRCA2* associated breast cancers [59]. ESR1 (estrogen receptor 1) was also enriched in *RCCD1 trans*-interacting regions in breast and ovarian cancer cells. *ESR1* is a nuclear receptor that is activated by the sex hormone estrogen. *ESR1* aberrations play a significant role in the development of breast cancers and their progression to metastasis and resistance[60,61]. Germline variants at the *ESR1* locus on chromosome 6q25 are associated with breast cancer risk[62], including in women carrying *BRCA1* and *BRCA2* mutations[63] and variations in mammographic density[64], probably through their regulation of *ESR1* gene expression[65]. In breast cancer cells we found enrichment for several notable genes and TFs involved in cancer development including *RB1* and *FOXA1*. The critical role of *RB1* (retinoblastoma protein 1) as a driver of neoplastic development is unequivocal. Alterations of *RB1* or components regulating the CDK-RB-E2F pathway have been

identified in nearly every human malignancy and lead to excessive cell growth through loss of cell cycle control. FOXA1 (Forkhead box protein A1) is a member of the forkhead class of DNA-binding proteins and hepatocyte nuclear factor transcriptional activators. It has a well characterised role in breast cancer development acting as a pioneer factor in estrogen receptor (ER) positive breast cancer, where its expression may indicate ER+ cancers that undergo rapid reprogramming of ER signaling are associated with poor outcomes and treatment resistance [66]. In ovarian cancer cells TFs that were enriched included *MYC, NR2F1 and KDM5B*. Genomic amplification of *MYC* is known to be a major driver of ovarian cancer development[67,68] while the *CMYC* locus on chromosome 8q24.1 is the most significant pleiotropic risk locus so far identified in breast, ovarian and several other cancers[12,69–73]. *NR2F1* (nuclear receptor subfamily 2 group F, member 1) binds directly with ESR1 and is a member of the nuclear hormone receptor family of steroid hormone receptors, consistent with the underlying etiology and known role of steroid hormones in the risk of ovarian cancer[1,2]. Finally, *KDM5B* (*JARID1B*) regulates *EMSY* via its interaction with the micro-RNA *miR-31*[74] and induces re-expression of *HEXIM1*, which was also enriched in in RCCD1 *trans*-interacting regions in ovarian cancer cells[75].

Differential gene expression analysis after *RCCD1* overexpression identified 82 genes in both breast and ovarian cancer precursor cell line models that may be associated with the early stages of neoplastic development of breast and ovarian cancers (**Supplementary Table S17**). As expected, *RCCD1* was one of these genes. For others, there is substantial evidence they are altered during breast/ovarian cancer initiation and development including: CCND2 (Cyclin D2), a regulator of cyclin-dependent kinases that appears to be frequently methylated in breast cancers affecting drug resistance[76,77] and in which common germline variants may affect both risk and survival for ovarian cancer [78,79]; WNT5A, which is central to many processes in cancer development including cellular senescence, proliferation, invasion, migration, inflammation, metastasis and chemo-resistance[80], increases cell migration and invasion in breast cancers[81-82] and is upregulated in ovarian cancer promoting epithelial-mesenchymal transition[83,84]; and the cadherin CDH11 is overexpressed in about 15% of breast cancers and is associated with poor prognosis [85]. Some genes may represent novel, possibly overlooked candidate biomarkers for screening and/or targets for therapeutic intervention. These include the Ras Protein Activator Like 1 gene, RASAL1 (Ras pathway activation is a critical component of ovarian, breast and many other cancers[86]); the Wilms Tumor interacting protein, WTIP (the Wilms Tumor 1 gene is a highly specific diagnostic and prognostic marker for ovarian cancers)[87]; the kalikrein -related peptidase KLK5 (one of a family of secreted proteins evaluated as screening biomarkers, including KLK3, or prostate specific antigen, which is routinely used to screen for prostate cancer)[88]; and the death associated protein kinase gene DAPK1, a candidate tumor suppressor gene and drug target that promotes the growth of P53 mutant cancers[89].

The 82 differentially expressed genes shared between breast and ovarian cancer models were also significantly enriched in three of the most well established pathways in cancer: TP53, Hippo and TNF signaling. Aside from TP53, which has been discussed at length already, the Hippo pathway is involved in regulating cell proliferation and promoting apoptosis in organ development and has a critical role in stem cell and tissue specific progenitor cell self-renewal and expansion[90]. Many of the pathway components in the Hippo signaling network are recognized as tumor suppressor genes and oncogenes that are perturbed in human cancers, including Fat4 which is mutated in breast cancer[91] and YAP in ovarian cancers[92,93] . TNF (tumor necrosis factor) signaling has a primary role in the regulation of immune cells and is involved in a wide range of physiological processes including inducing fever, apoptotic cell death, cachexia, inflammation and tumor inhibition. Dysregulation of TNF production has been implicated in cancer development[94] with breast and ovarian cancers often accompanied by inflammatory

processes characterized by the presence of proinflammatory cytokines including TNF-α[95,96]. Immunotherapy has become a major focus of targeted treatments against a wide range of cancers and within this field members of the TNF receptor superfamily have also been proposed as novel therapeutic targets[97]. Some of the pathways were the same as those identified as enriched networks in *RCCD1* 4C *trans*-interaction studies, in particular steroid hormone biosynthesis in breast cancer and P53 associated pathways in ovarian cancer, even though these networks were derived using different methods and experimental models at different stages of disease pathogenesis.

Strikingly, our data provide evidence that connects 4C *trans*-interaction networks and gene level associations from modelling *RCCD1* overexpression to credible causal risk variants at already identified breast, ovarian and pleiotropic breast-ovarian risk regions. Our results suggest cross-talk between regulatory elements in risk regions with gene promoters at other risk regions on different chromosomes. Recently, Pritchard and colleagues [98] proposed a model where genetic contributions to complex traits are partitioned into direct effects from core genes and indirect effects from peripheral genes acting in *trans*. The data we present here indicates that there are likely to be direct functional *trans*-relationships between different GWAS risk loci for any given phenotype. Interestingly, past studies have shown these types of contacts are known to be within interchromosomal enhancer–promoter networks in developmental cell differentiation[99,100] and have been proposed to contribute to transcriptional heterogeneity within cell populations[101,102]. In one study, aberrant *TAL1* expression in human T-cell acute lymphoblastic leukemia (ALL) was mediated by interchromosomal interaction between the *TAL1* promoter on chromosome 1 with a regulatory element called *TIL16* on chromosome 1[103] indicating that *trans*-interactions can capture a target gene's ability to co-regulate the expression of other genes. This supports the hypothesis that weak *trans*-eQTL SNPs contribute to their modest heritability effects through their regulation of peripheral genes to alter core gene expression[98]. If these core genes tend to be co-regulated, then the model also predicts that peripheral variation effects can be increased with a large proportion of heritability driven by these weak *trans* effects.

In summary, we describe a functional framework for the analysis of pleiotropic GWAS risk loci, based on the identification of *trans*-interaction networks combined with differential gene expression modeling of candidate susceptibility genes, to connect different susceptibility loci through their shared biology. The data we present supports the hypothesis that RCCD1 is the susceptibility gene target for both breast and ovarian cancer at the 15q26.1 pleiotropic risk locus, and suggests that by looking at RCCD1 interaction networks throughout the genome we can identify additional risk loci and susceptibility genes for these phenotypes. This may be particularly relevant for studying rare phenotypes where the ability to identify new genome wide risk associations based on case-control genotype analyses alone is restricted by sample size.

## ACKNOWLEDGEMENTS

**FIGURES AND TABLES**

**Table 1: Details of the 17 credible causal risk variants (in order (5' to 3') of nucleotide location) identified by colocalization analyses of breast and ovarian cancer data at the 15q26.1 locus**

| SNP ID | Nucleotide position (build 37) | P-Value breast cancer | P-Value ovarian cancer | SNP Score [a] | # TF motifs intersections | Tissue specific epigenomic intersections [b] |
|---|---|---|---|---|---|---|
| *rs8028409* [c] | *91506422* | *4.73 x 10^{-15}* | *3.26 x 10^{-7}* | *0.046* | *13* | *-* |
| *rs8037137* [c] | *91506637* | *6.52 x 10^{-15}* | *4.61 x 10^{-7}* | *0.023* | *3* | *-* |
| rs112770790 | 91507495 | 3.44 x 10^{-15} | 4.97 x 10^{-7} | 0.039 | 7 | EPR |
| rs77554484 | 91509215 | 1.28 x 10^{-15} | 5.73 x 10^{-7} | 0.090 | - | AR |
| rs2290202 | 91512267 | 1.87 x 10^{-15} | 4.38 x 10^{-7} | 0.073 | 11 | - |
| rs2890156 | 91513157 | 5.65 x 10^{-15} | 4.81 x 10^{-7} | 0.022 | 14 | - |

| | | | | | | |
|---|---|---|---|---|---|---|
| rs12595025 | 91520131 | $4.03 \times 10^{-15}$ | $4.20 \times 10^{-7}$ | 0.039 | 15 | - |
| rs12594925 | 91520287 | $2.64 \times 10^{-15}$ | $3.54 \times 10^{-7}$ | 0.070 | 16 | CTCF |
| rs113343095 | 91522092 | $3.37 \times 10^{-15}$ | $4.54 \times 10^{-7}$ | 0.056 | 16 | AR; EAR |
| rs8026714 | 91522253 | $6.05 \times 10^{-15}$ | $5.20 \times 10^{-7}$ | 0.021 | - | EPR; AR; EAR |
| *rs2301825*[d] | *91528070* | $2.24 \times 10^{-15}$ | $4.37 \times 10^{-7}$ | *0.070* | *4* | *EAR* |
| rs150333865 | 91528875 | $3.83 \times 10^{-15}$ | $6.46 \times 10^{-7}$ | 0.026 | 24 | EAR |
| rs60290423 | 91529452 | $3.21 \times 10^{-15}$ | $7.50 \times 10^{-7}$ | 0.027 | 4 | EPR |
| rs12594752 | 91531995 | $1.86 \times 10^{-15}$ | $2.13 \times 10^{-7}$ | 0.159 | 1 | - |
| rs76119208 | 91535329 | $2.40 \times 10^{-15}$ | $2.20 \times 10^{-7}$ | 0.129 | 29 | EPR |
| rs78029804 | 91538920 | $3.17 \times 10^{-15}$ | $2.46 \times 10^{-7}$ | 0.098 | - | - |
| rs3826033 | 91548601 | $1.16 \times 10^{-14}$ | $7.14 \times 10^{-7}$ | 0.008 | 15 | - |

[a] Ranking score for the top 17 risk associated SNPs in breast and ovarian cancer from co-localization analysis (i.e., posterior probability of colocalization at each SNP). These 17 SNPs explain > 99% of the shared association in the region
[b] EAR: Active enhancer region; EPR: Poised enhancer region; CTCF: Architectural complex region; AR: Active region
[c] Intersects a 4C interacting region in breast and ovarian cancer cells (Figure 3A)
[d] Intersects a 4C interacting region in breast and ovarian precursor cells (Figure 3A)

**Figure 1: The experimental pipeline used for the genetic and functional analysis of the 15q26.1 (*RCCD1*) breast and ovarian cancer susceptibility locus**.

**Figure 2: Fine mapping and expression quantitative trait locus analysis of the chromosome 15p26.1 breast and ovarian cancer risk locus: (A)** Regional association plot for the chr15:91009215-92009215 interval derived from independent genetic association analysis performed breast and ovarian cancer case-control studies and from a meta-analysis that combines breast and ovarian cancer cases and controls to identify 17 credible causal risk variants that are shared between both cancers; **(B)** Regional localization plot showing the location of the 17 credible causal risk variants with respect to three nearest by genes, *RCCD1*, *PRC1* and the lncRNA *PRC1-AS1*; **(C)** Expression QTL analysis in normal breast and ovarian tissues from GTEx. Violin plots show allele-expression associations for rs8037137 index SNP. The T (risk associated) allele is significantly associated with decreased expression of *RCCD1* and increased expression of *PRC1-AS1* in both breast and ovarian tissues.

**Figure 3: Circular chromosome conformation capture (4C) analysis anchored to *RCCD1* in breast and ovarian precursor and cancer cell lines. (A)** Genomic locus plot of chr 15: 91470-91450 showing. RefSeq genes annotated from hg19 build of UCSC genome browser.

Shared 4C interactions with RCCD1 in breast and ovarian cancer cell lines (magenta loops) and breast and ovarian precursors (blue loops). Combined tracks of enhancer-associated H3K27ac histone marks catalogued in breast (pink peaks) and ovarian (dark teal peaks) cancer and precursor cell lines show the 4C-RCCD1 landscape extending upstream and downstream of RCCD1, with respect to the colocalized credible causal risk variants (vertical dark blue lines). Interacting peaks that overlap 2/17 risk variants (rs8037137 and rs8028409) in cancer cell lines and 1/17 risk variants (rs2301825) in precursor cell lines are highlighted by the pink and purple vertical bars respectively. Chromatin states annotated in breast (MCF7) and ovarian (UWB1.289) cancer cell lines: Active region (AR): active promoter region (PAR); poised promoter region (PPR); active enhancer region (EAR); poised promoter region (PPR); architectural complex region (CTCF). **(B)** Enrichment of TFs ChIP-Seq peaks found in *cis*-interactions of breast cancer (MCF7, BT549) cell lines. Color designates the statistical significance (Q value) of the TF enrichment with highest (red) to lowest E value (blue). Circle sizes depict the number of overlaps the TFs intersect with *cis*-interactions; **(C)** Enrichment of TFs ChIP-Seq peaks found in *cis*-interactions of (B) ovarian cancer (UWB1.289, Kuramochi) cell lines.

**Figure 4: *Trans*-4C interaction and network analysis for RCCD1 in breast and ovarian cancer cells. (A)** Circos plot of genome wide RCCD1 *trans*-interactions (FDR<0.01) in breast cancer cell lines (MCF7, BT549) and ovarian cancer cell lines (UWB.1289, Kuramochi) are depicted as red and green lines, respectively, looping in the center. Interactions are aligned with respect to chromosomes (outer track), and breast and ovarian cancer GWAS risk loci (P≥5x10$^{-8}$) which intersect *trans*-interactions (inner circle). **(B)** Wilcoxon text of DEGs enriched in RCCD1-TIRs in breast and ovarian cancer cells based on the simulation of more than a thousand random permutations (P = 0.049)**(C)** Enrichment of TF ChIP-Seq peaks found in *trans*-interactions of breast cancer cell lines MCF7 and BT549. Color of the circles represent the statistical significance of the TF enrichment with red representing the highest and blue the lowest. Circle sizes represent the total numbers of overlaps between TFs and *trans*-interactions in breast cancer; **(D)** Enrichment of TF ChIP-Seq peaks found in *trans*-interactions of ovarian cancer cell lines UWB1.289 and Kuramochi. Color codes are the same as in (B).

**Figure 5:  Phenotypic analysis of RCCD1 overexpression models in precursor breast and ovarian cancer cells**. **(A)** Cell proliferation is significantly increased in precursor **(A)** breast (MCF12A RCCD1 OE) and **(B)** ovarian (FT282-RCCD1 OE) cell models of RCCD1 overexpression compared to control (CONTROL OE). Cell proliferation decreased in RCCD1 knockout cell models in breast (MCF12A RCCD1 KO)  but increased in ovarian cancer models (FT282 RCCD1 KO) compared to control cells (OR104A KO). Heat map of color coded expression of DEGs (| log2FC |> 0.5, p$_{adjusted}$<0.05) from overexpression of RCCD1 in **(C)** breast cancer precursor cells (MCF12A$^{RCCD1+}$) and **(D)** ovarian cancer precursor cells (FT282$^{RCCD1+}$). Upregulated genes are depicted in yellow and down regulated genes are shown in green.

**Figure 6: RCCD1 is a putative regulator of the P53 pathway in breast and ovarian cancer cells (A)** RCCD1 operates in the P53 pathways through its interaction with KDM8. As part of this complex RCCD1 also regulates HIF1A signaling, transcriptional activation via H3K36me2, epithelial mesenchymal transition via TGF beta signaling, cell proliferation via CCNA1, DNA damage response and Androgen receptor signaling. **(B)** Pathway analysis of DEGs from RCCD1 overexpression models are enriched in key proteins of the TP53 signaling pathway.
**(C)** Western blot analysis after CRISPR/Cas9 knockout of RCCD1 in fallopian tube epithelial cells expressing the R175H P53 mutation (FT282$^{P53:R175H}$). Two different RCCD1 knockout

clones are compared to knockdown of a control gene, the olfactory receptor OR10A4. After RCCD1 knockdown and clonal selection, mutant TP53 is depleted in FT282$^{P53:R175H}$ cells indicating downregulated expression of mutant TP53 is selected for in the presence of RCCD1 knockout.

**(C)** RCCD1 expression in TCGA breast and ovarian tumors based on TP53 mutation status. These plots were obtained using the UALCAN web tool (http://ualcan.path.uab.edu/cgi-bin/TCGAExResultNew2.pl?genenam=RCCD1&ctype=BRCA. Briefly, in UALCAN, TP53 mutation status was obtained from TCGA whole exome sequencing data. Mutation Annotation Format files (derived from VarScan2) from the Genomic Data Commons portal were matched with TCGA RNA-seq data and the analysis performed stratifying the samples with/without TP53 mutation.

**Supplementary Table S1:** Primers used for luciferase enhancer assay to create risk and non risk SNPs (rs802840, rs8037137)

**Supplementary Table S2:** Colocalization fine-mapping of the breast and ovarian cancer datasets at the 15q26 breast-ovarian cancer risk locus identifies 17 credible SNPs (rows 2-18) explaining > 99% of the shared genetic association in the region.

**Supplementary Table S3:** Quantitative trait locus (QTL) analysis for all 17 credible causal SNPs in the 15q26 risk region and their association with expression of the RCCD1 gene and the PRC-AS1 lncRNA within the region.

**Supplementary Table S4:** Variant effect prediction of altered TFBSs by broken TF motifs of risk SNPs which intersect with 4C interactions using PERFECTOS-APE.

**Supplementary Table S5:** Epigenomic annotations of all 17 credible causal SNPs in the 15q26 breast and ovarian cancer risk region by tissue type. Chromatin state annotations were compiled from H3K27Ac (active promoter/enhancer), H3K4me3(active promoter) , H3k4me1 (poised enhancer/promoter), CTCF (architectural protein)

**Supplementary Table S6:** *cis*-4C interaction enrichments in breast, ovarian, and combined breast-ovarian normal and cancer cell lines.

**Supplementary Table S7:** *trans*-4C interaction enrichments in breast (MCF7, BT549) and ovarian (UWB1.289, Kuramochi) cancer cell lines (FDR<0.01).

**Supplementary Table S8:** TFBS enrichment of *trans*-interactions from breast cancer cell lines (MCF7, BT549) in ReMAP 2020 dataset (P<0.01).

**Supplementary Table S9:** TFBS enrichment of *trans*-interactions from ovarian cancer cell lines (UWB1.289, Kuramochi) in ReMAP 2020 dataset (P<0.01).

**Supplementary Table S10:** Differentially expressed genes, ranked by P-value, identified by RNA-Seq analysis, after overexpression of RCCD1 in breast and ovarian precursor cell lines(MCF12A$^{RCCD1+}$, FT282$^{RCCD1+}$ respectively).

**Supplementary Table S11**: KEGG pathway enrichment in breast precursor cells (MCF12A$^{RCCD1+}$)

**Supplementary Table S12:** KEGG pathway enrichment in ovarian precursor cells (FT282$^{RCCD1+}$)

**Supplementary Table S13**: KEGG pathway enrichment in overexpression of RCCD1 in breast (MCF12A$^{RCCD1}$) and ovarian (FT282$^{RCCD1}$) precursor cell lines

**Supplementary Table S14:** Genes with P$_{adjusted}$ < 0.05 and PMAGMA < 2.6 x 10-6

**Supplementary Table S15:** TF Enrichment from LISA Cistrome Analysis in MCF12A$^{RCCD1+}$

**Supplementary Table S16:** TF Enrichment from LISA Cistrome Analysis in FT282$^{RCCD1+}$

**Supplementary Table S17:** Directional effects of DEGs shared between RCCD1 overexpression in breast and ovarian precursor cells (MCF12A$^{RCCD1+}$ and FT282$^{RCCD1+}$); (adjusted P value<0.05)

**Supplementary Figure 1: (A)** RT-PCR analysis of breast and ovarian normal and cancer cell lines. RCCD1 shows higher expression in triple negative breast cancer cell lines compared to normal mammary epithelial cells, in low grade serous ovarian cancer cell lines compared to the likely precursors -fallopian tube secretory epithelial cells and ovarian surface epithelial cells;**(B)** Western blot analysis after CRISPR/Cas9 knockout of RCCD1 in breast (MCF12A$^{RCCD1-}$) and ovarian (FT282$^{RCCD1-}$) cancer precursor cells compared to knockout of the olfactory receptor ORF104A. After RCCD1 knockout and clonal selection, RCCD1 is depleted in MCF12A and FT282 cells, but not in cells after knockout of ORF104A; **(C)(i)** RT-PCR analysis confirming increased RCCD1 expression after stable overexpression of full length RCCD1 and clonal derivation in breast (MCF12A) and ovarian (FT282) cancer precursor cells. (ii) Western blotting analysis confirms over expression of RCCD1 protein in breast (MCF12A$^{RCCD1+}$) and ovarian (FT282$^{RCCD1+}$) cancer precursor cells

## REFERENCES

Bibliography

1. Beral, V., Million Women Study Collaborators, Bull, D., Green, J., and Reeves, G. (2007). Ovarian cancer and hormone replacement therapy in the Million Women Study. Lancet *369*, 1703–1710.

2. Beral, V., and Million Women Study Collaborators (2003). Breast cancer and hormone-replacement therapy in the Million Women Study. Lancet *362*, 419–427.

3. Futreal, P.A., Liu, Q., Shattuck-Eidens, D., Cochran, C., Harshman, K., Tavtigian, S., Bennett, L.M., Haugen-Strano, A., Swensen, J., and Miki, Y. (1994). BRCA1 mutations in primary breast and ovarian carcinomas. Science *266*, 120–122.

4. Gayther, S.A., Warren, W., Mazoyer, S., Russell, P.A., Harrington, P.A., Chiano, M., Seal, S., Hamoudi, R., van Rensburg, E.J., Dunning, A.M., et al. (1995). Germline mutations of the

BRCA1 gene in breast and ovarian cancer families provide evidence for a genotype-phenotype correlation. Nat. Genet. *11*, 428–433.

5. Gayther, S.A., Mangion, J., Russell, P., Seal, S., Barfoot, R., Ponder, B.A., Stratton, M.R., and Easton, D. (1997). Variation of risks of breast and ovarian cancer associated with different germline mutations of the BRCA2 gene. Nat. Genet. *15*, 103–105.

6. Lancaster, J.M., Wooster, R., Mangion, J., Phelan, C.M., Cochran, C., Gumbs, C., Seal, S., Barfoot, R., Collins, N., Bignell, G., et al. (1996). BRCA2 mutations in primary breast and ovarian cancers. Nat. Genet. *13*, 238–240.

7. Ford, D., Easton, D.F., Stratton, M., Narod, S., Goldgar, D., Devilee, P., Bishop, D.T., Weber, B., Lenoir, G., Chang-Claude, J., et al. (1998). Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families. The Breast Cancer Linkage Consortium. Am. J. Hum. Genet. *62*, 676–689.

8. Jiang, X., Finucane, H.K., Schumacher, F.R., Schmit, S.L., Tyrer, J.P., Han, Y., Michailidou, K., Lesseur, C., Kuchenbaecker, K.B., Dennis, J., et al. (2019). Shared heritability and functional enrichment across six solid cancers. Nat. Commun. *10*, 431.

9. Lindström, S., Finucane, H., Bulik-Sullivan, B., Schumacher, F.R., Amos, C.I., Hung, R.J., Rand, K., Gruber, S.B., Conti, D., Permuth, J.B., et al. (2017). Quantifying the Genetic Correlation between Multiple Cancer Types. Cancer Epidemiol. Biomarkers Prev. *26*, 1427–1435.

10. Rashkin, S.R., Graff, R.E., Kachuri, L., Thai, K.K., Alexeeff, S.E., Blatchins, M.A., Cavazos, T., Corley, D.A., Emami, N.C., Hoffman, J.D., et al. (2019). Pan-Cancer Study Detects Novel Genetic Risk Variants and Shared Genetic Basis in Two Large Cohorts. BioRxiv.

11. Fehringer, G., Kraft, P., Pharoah, P.D., Eeles, R.A., Chatterjee, N., Schumacher, F.R., Schildkraut, J.M., Lindström, S., Brennan, P., Bickeböller, H., et al. (2016). Cross-Cancer Genome-Wide Analysis of Lung, Ovary, Breast, Prostate, and Colorectal Cancer Reveals Novel Pleiotropic Associations. Cancer Res. *76*, 5103–5114.

12. Ghoussaini, M., Song, H., Koessler, T., Al Olama, A.A., Kote-Jarai, Z., Driver, K.E., Pooley, K.A., Ramus, S.J., Kjaer, S.K., Hogdall, E., et al. (2008). Multiple loci with different cancer specificities within the 8q24 gene desert. J Natl Cancer Inst *100*, 962–966.

13. Bojesen, S.E., Pooley, K.A., Johnatty, S.E., Beesley, J., Michailidou, K., Tyrer, J.P., Edwards, S.L., Pickett, H.A., Shen, H.C., Smart, C.E., et al. (2013). Multiple independent variants at the TERT locus are associated with telomere length and risks of breast and ovarian cancer. Nat. Genet. *45*, 371–384, 384e1.

14. Lawrenson, K., Kar, S., McCue, K., Kuchenbaeker, K., Michailidou, K., Tyrer, J., Beesley, J., Ramus, S.J., Li, Q., Delgado, M.K., et al. (2016). Functional mechanisms underlying pleiotropic risk alleles at the 19p13.1 breast-ovarian cancer susceptibility locus. Nat. Commun. *7*, 12675.

15. Kar, S.P., Beesley, J., Amin Al Olama, A., Michailidou, K., Tyrer, J., Kote-Jarai, Zs., Lawrenson, K., Lindstrom, S., Ramus, S.J., Thompson, D.J., et al. (2016). Genome-Wide Meta-Analyses of Breast, Ovarian, and Prostate Cancer Association Studies Identify Multiple

New Susceptibility Loci Shared by at Least Two Cancer Types. Cancer Discov. *6*, 1052–1067.

16. Michailidou, K., Lindström, S., Dennis, J., Beesley, J., Hui, S., Kar, S., Lemaçon, A., Soucy, P., Glubb, D., Rostamianfar, A., et al. (2017). Association analysis identifies 65 new breast cancer risk loci. Nature *551*, 92–94.

17. Phelan, C.M., Kuchenbaecker, K.B., Tyrer, J.P., Kar, S.P., Lawrenson, K., Winham, S.J., Dennis, J., Pirie, A., Riggan, M.J., Chornokur, G., et al. (2017). Identification of 12 new susceptibility loci for different histotypes of epithelial ovarian cancer. Nat. Genet. *49*, 680–691.

18. 1000 Genomes Project Consortium, Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., et al. (2015). A global reference for human genetic variation. Nature *526*, 68–74.

19. Foley, C.N., Staley, J.R., Breen, P.G., Sun, B.B., Kirk, P.D., Burgess, S., and Howson, J.M. (2019). A fast and efficient colocalization algorithm for identifying shared genetic risk factors across multiple traits. BioRxiv.

20. Munz, M., Wohlers, I., Simon, E., Busch, H., Schaefer, A.S., and Erdmann, J. (2018). Qtlizer: comprehensive QTL annotation of GWAS results. BioRxiv.

21. GTEx Consortium (2013). The Genotype-Tissue Expression (GTEx) project. Nat. Genet. *45*, 580–585.

22. Lim, Y.W., Chen-Harris, H., Mayba, O., Lianoglou, S., Wuster, A., Bhangale, T., Khan, Z., Mariathasan, S., Daemen, A., Reeder, J., et al. (2018). Germline genetic polymorphisms influence tumor gene expression and immune cell infiltration. Proc Natl Acad Sci USA *115*, E11701–E11710.

23. Cancer Genome Atlas Research Network (2011). Integrated genomic analyses of ovarian carcinoma. Nature *474*, 609–615.

24. Cancer Genome Atlas Network (2012). Comprehensive molecular portraits of human breast tumours. Nature *490*, 61–70.

25. Berger, A.C., Korkut, A., Kanchi, R.S., Hegde, A.M., Lenoir, W., Liu, W., Liu, Y., Fan, H., Shen, H., Ravikumar, V., et al. (2018). A Comprehensive Pan-Cancer Molecular Study of Gynecologic and Breast Cancers. Cancer Cell *33*, 690-705.e9.

26. Karst, A.M., Jones, P.M., Vena, N., Ligon, A.H., Liu, J.F., Hirsch, M.S., Etemadmoghadam, D., Bowtell, D.D.L., and Drapkin, R. (2014). Cyclin E1 deregulation occurs early in secretory cell transformation to promote formation of fallopian tube-derived high-grade serous ovarian cancers. Cancer Res. *74*, 1141–1152.

27. Chen, L., Huang, M., Plummer, J., Pan, J., Jiang, Y.Y., Yang, Q., Silva, T.C., Gull, N., Chen, S., Ding, L.W., et al. (2020). Master transcription factors form interconnected circuitry and orchestrate transcriptional networks in oesophageal adenocarcinoma. Gut *69*, 630–640.

28. Splinter, E., de Wit, E., van de Werken, H.J.G., Klous, P., and de Laat, W. (2012). Determining long-range chromatin interactions for selected genomic sites using 4C-seq technology: from fixation to computation. Methods *58*, 221–230.

29. Adler, E.K., Corona, R.I., Lee, J.M., Rodriguez-Malave, N., Mhawech-Fauceglia, P., Sowter, H., Hazelett, D.J., Lawrenson, K., and Gayther, S.A. (2017). The PAX8 cistrome in epithelial ovarian cancer. Oncotarget *8*, 108316–108332.

30. Davis, C.A., Hitz, B.C., Sloan, C.A., Chan, E.T., Davidson, J.M., Gabdank, I., Hilton, J.A., Jain, K., Baymuradov, U.K., Narayanan, A.K., et al. (2018). The Encyclopedia of DNA elements (ENCODE): data portal update. Nucleic Acids Res. *46*, D794–D801.

31. Coetzee, S.G., Ramjan, Z., Dinh, H.Q., Berman, B.P., and Hazelett, D.J. (2020). StateHub-StatePaintR: rapid and reproducible chromatin state evaluation for custom genome annotation. F1000Res. *7*, 214.

32. E. Vorontsov, I., V. Kulakovskiy, I., Khimulya, G., D. Nikolaeva, D., and J. Makeev, V. (2015). PERFECTOS-APE - Predicting Regulatory Functional Effect of SNPs by Approximate P-value Estimation. In Proceedings of the International Conference on Bioinformatics Models, Methods and Algorithms, (SCITEPRESS - Science and and Technology Publications), pp. 102–108.

33. Chèneby, J., Ménétrier, Z., Mestdagh, M., Rosnet, T., Douida, A., Rhalloussi, W., Bergon, A., Lopez, F., and Ballester, B. (2020). ReMap 2020: a database of regulatory regions from an integrative analysis of Human and Arabidopsis DNA-binding sequencing experiments. Nucleic Acids Res. *48*, D180–D188.

34. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

35. Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics *12*, 323.

36. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

37. Fachal, L., Aschard, H., Beesley, J., Barnes, D.R., Allen, J., Kar, S., Pooley, K.A., Dennis, J., Michailidou, K., Turman, C., et al. (2020). Fine-mapping of 150 breast cancer risk regions identifies 191 likely target genes. Nat. Genet. *52*, 56–73.

38. Marcon, E., Ni, Z., Pu, S., Turinsky, A.L., Trimble, S.S., Olsen, J.B., Silverman-Gavrila, R., Silverman-Gavrila, L., Phanse, S., Guo, H., et al. (2014). Human-chromatin-related protein interactions identify a demethylase complex required for chromosome segregation. Cell Rep. *8*, 297–310.

39. Chandrashekar, D.S., Bashel, B., Balasubramanya, S.A.H., Creighton, C.J., Ponce-Rodriguez, I., Chakravarthi, B.V.S.K., and Varambally, S. (2017). UALCAN: A portal for facilitating tumor subgroup gene expression and survival analyses. Neoplasia *19*, 649–658.

40. de Leeuw, C.A., Mooij, J.M., Heskes, T., and Posthuma, D. (2015). MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput. Biol. *11*, e1004219.

41. de Leeuw, C., Sey, N.Y.., Posthuma, D., and Won, H. (2020). A response to Yurko et al: H-MAGMA, inheriting a shaky statistical foundation, yields excess false positives. BioRxiv.

42. Wu, L., Shi, W., Long, J., Guo, X., Michailidou, K., Beesley, J., Bolla, M.K., Shu, X.-O., Lu,

Y., Cai, Q., et al. (2018). A transcriptome-wide association study of 229,000 women identifies new candidate susceptibility genes for breast cancer. Nat. Genet. *50*, 968–978.

43. Kar, S., Considine, D., Tyrer, J., Plummer, J., Chen, S., Dezem, F., Barbeira, A., Rajagopal, P., Rosenow, W., Anton, F., et al. (2020). Pleiotropy-guided transcriptome imputation from normal and tumor tissues identifies new candidate susceptibility genes for breast and ovarian cancer. BioRxiv.

44. Kar, S.P., Tyrer, J.P., Li, Q., Lawrenson, K., Aben, K.K.H., Anton-Culver, H., Antonenkova, N., Chenevix-Trench, G., Australian Cancer Study, Australian Ovarian Cancer Study Group, et al. (2015). Network-Based Integration of GWAS and Gene Expression Identifies a HOX-Centric Network Associated with Serous Ovarian Cancer Risk. Cancer Epidemiol. Biomarkers Prev. *24*, 1574–1584.

45. Buckley, M.A., Woods, N.T., Tyrer, J.P., Mendoza-Fandiño, G., Lawrenson, K., Hazelett, D.J., Najafabadi, H.S., Gjyshi, A., Carvalho, R.S., Lyra, P.C., et al. (2019). Functional analysis and fine mapping of the 9p22.2 ovarian cancer susceptibility locus. Cancer Res. *79*, 467–481.

46. Gusev, A., Lawrenson, K., Lin, X., Lyra, P.C., Kar, S., Vavra, K.C., Segato, F., Fonseca, M.A.S., Lee, J.M., Pejovic, T., et al. (2019). A transcriptome-wide association study of high-grade serous epithelial ovarian cancer identifies new susceptibility genes and splice variants. Nat. Genet. *51*, 815–823.

47. Chen, J., Rajasekaran, M., Xia, H., Zhang, X., Kong, S.N., Sekar, K., Seshachalam, V.P., Deivasigamani, A., Goh, B.K.P., Ooi, L.L., et al. (2016). The microtubule-associated protein PRC1 promotes early recurrence of hepatocellular carcinoma in association with the Wnt/β-catenin signalling pathway. Gut *65*, 1522–1534.

48. Zhang, B., Shi, X., Xu, G., Kang, W., Zhang, W., Zhang, S., Cao, Y., Qian, L., Zhan, P., Yan, H., et al. (2017). Elevated PRC1 in gastric carcinoma exerts oncogenic function and is targeted by piperlongumine in a p53-dependent manner. J. Cell. Mol. Med. *21*, 1329–1341.

49. Zhan, P., Xi, G.-M., Liu, H.-B., Liu, Y.-F., Xu, W.-J., Zhu, Q., Zhou, Z.-J., Miao, Y.-Y., Wang, X.-X., Jin, J.-J., et al. (2017). Protein regulator of cytokinesis-1 expression: prognostic value in lung squamous cell carcinoma patients. J. Thorac. Dis. *9*, 2054–2060.

50. Bu, H., Li, Y., Jin, C., Yu, H., Wang, X., Chen, J., Wang, Y., Ma, Y., Zhang, Y., and Kong, B. (2020). Overexpression of PRC1 indicates a poor prognosis in ovarian cancer. Int. J. Oncol. *56*, 685–696.

51. Varley, J.M. (2003). Germline TP53 mutations and Li-Fraumeni syndrome. Hum. Mutat. *21*, 313–320.

52. Masciari, S., Dillon, D.A., Rath, M., Robson, M., Weitzel, J.N., Balmana, J., Gruber, S.B., Ford, J.M., Euhus, D., Lebensohn, A., et al. (2012). Breast cancer phenotype in women with TP53 germline mutations: a Li-Fraumeni syndrome consortium effort. Breast Cancer Res. Treat. *133*, 1125–1130.

53. Smith, P.D., Crossland, S., Parker, G., Osin, P., Brooks, L., Waller, J., Philp, E., Crompton, M.R., Gusterson, B.A., Allday, M.J., et al. (1999). Novel p53 mutants selected in BRCA-associated tumours which dissociate transformation suppression from other wild-type p53

functions. Oncogene *18*, 2451–2459.

54. Ramus, S.J., Bobrow, L.G., Pharoah, P.D., Finnigan, D.S., Fishman, A., Altaras, M., Harrington, P.A., Gayther, S.A., Ponder, B.A., and Friedman, L.S. (1999). Increased frequency of TP53 mutations in BRCA1 and BRCA2 ovarian tumours. Genes Chromosomes Cancer *25*, 91–96.

55. Hsia, D.A., Tepper, C.G., Pochampalli, M.R., Hsia, E.Y.C., Izumiya, C., Huerta, S.B., Wright, M.E., Chen, H.-W., Kung, H.-J., and Izumiya, Y. (2010). KDM8, a H3K36me2 histone demethylase that acts in the cyclin A1 coding region to regulate cancer cell proliferation. Proc Natl Acad Sci USA *107*, 9671–9676.

56. Konstantinopoulos, P.A., Ceccaldi, R., Shapiro, G.I., and D'Andrea, A.D. (2015). Homologous recombination deficiency: exploiting the fundamental vulnerability of ovarian cancer. Cancer Discov. *5*, 1137–1154.

57. Fradet-Turcotte, A., Sitz, J., Grapton, D., and Orthwein, A. (2016). BRCA2 functions: from DNA repair to replication fork stabilization. Endocr. Relat. Cancer *23*, T1–T17.

58. Cousineau, I., and Belmaaza, A. (2011). EMSY overexpression disrupts the BRCA2/RAD51 pathway in the DNA-damage response: implications for chromosomal instability/recombination syndromes as checkpoint diseases. Mol. Genet. Genomics *285*, 325–340.

59. Ahn, J.-H., Kim, T.J., Lee, J.H., and Choi, J.-H. (2017). Mutant p53 stimulates cell invasion through an interaction with Rad21 in human ovarian cancer cells. Sci. Rep. *7*, 9076.

60. Lei, J.T., Shao, J., Zhang, J., Iglesia, M., Chan, D.W., Cao, J., Anurag, M., Singh, P., He, X., Kosaka, Y., et al. (2018). Functional Annotation of ESR1 Gene Fusions in Estrogen Receptor-Positive Breast Cancer. Cell Rep. *24*, 1434-1444.e7.

61. Robinson, D.R., Wu, Y.-M., Vats, P., Su, F., Lonigro, R.J., Cao, X., Kalyana-Sundaram, S., Wang, R., Ning, Y., Hodges, L., et al. (2013). Activating ESR1 mutations in hormone-resistant metastatic breast cancer. Nat. Genet. *45*, 1446–1451.

62. Purrington, K.S., Slager, S., Eccles, D., Yannoukakos, D., Fasching, P.A., Miron, P., Carpenter, J., Chang-Claude, J., Martin, N.G., Montgomery, G.W., et al. (2014). Genome-wide association study identifies 25 known breast cancer susceptibility loci as risk factors for triple-negative breast cancer. Carcinogenesis *35*, 1012–1019.

63. Antoniou, A.C., Kartsonaki, C., Sinilnikova, O.M., Soucy, P., McGuffog, L., Healey, S., Lee, A., Peterlongo, P., Manoukian, S., Peissel, B., et al. (2011). Common alleles at 6q25.1 and 1p11.2 are associated with breast cancer risk for BRCA1 and BRCA2 mutation carriers. Hum. Mol. Genet. *20*, 3304–3321.

64. Lindström, S., Thompson, D.J., Paterson, A.D., Li, J., Gierach, G.L., Scott, C., Stone, J., Douglas, J.A., dos-Santos-Silva, I., Fernandez-Navarro, P., et al. (2014). Genome-wide association study identifies multiple loci associated with both mammographic density and breast cancer risk. Nat. Commun. *5*, 5303.

65. Dunning, A.M., Michailidou, K., Kuchenbaecker, K.B., Thompson, D., French, J.D., Beesley, J., Healey, C.S., Kar, S., Pooley, K.A., Lopez-Knowles, E., et al. (2016). Breast cancer risk

variants at 6q25 display different phenotype associations and regulate ESR1, RMND1 and CCDC170. Nat. Genet. *48*, 374–386.

66. Ross-Innes, C.S., Stark, R., Teschendorff, A.E., Holmes, K.A., Ali, H.R., Dunning, M.J., Brown, G.D., Gojis, O., Ellis, I.O., Green, A.R., et al. (2012). Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. Nature *481*, 389–393.

67. Jung, M., Russell, A.J., Kennedy, C., Gifford, A.J., Australian Ovarian Cancer Study Group, Mallitt, K.-A., Sivarajasingam, S., Bowtell, D.D., DeFazio, A., Haber, M., et al. (2018). Clinical importance of myc family oncogene aberrations in epithelial ovarian cancer. JNCI Cancer Spectr *2*, pky047.

68. Fallah, Y., Brundage, J., Allegakoen, P., and Shajahan-Haq, A.N. (2017). MYC-Driven Pathways in Breast Cancer Subtypes. Biomolecules *7*,.

69. Easton, D.F., Pooley, K.A., Dunning, A.M., Pharoah, P.D.P., Thompson, D., Ballinger, D.G., Struewing, J.P., Morrison, J., Field, H., Luben, R., et al. (2007). Genome-wide association study identifies novel breast cancer susceptibility loci. Nature *447*, 1087–1093.

70. Thomas, G., Jacobs, K.B., Yeager, M., Kraft, P., Wacholder, S., Orr, N., Yu, K., Chatterjee, N., Welch, R., Hutchinson, A., et al. (2008). Multiple loci identified in a genome-wide association study of prostate cancer. Nat. Genet. *40*, 310–315.

71. Zanke, B.W., Greenwood, C.M.T., Rangrej, J., Kustra, R., Tenesa, A., Farrington, S.M., Prendergast, J., Olschwang, S., Chiang, T., Crowdy, E., et al. (2007). Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. Nat. Genet. *39*, 989–994.

72. Cole, M.D. (2014). MYC association with cancer risk and a new model of MYC-mediated repression. Cold Spring Harb. Perspect. Med. *4*, a014316.

73. Pharoah, P.D.P., Tsai, Y.-Y., Ramus, S.J., Phelan, C.M., Goode, E.L., Lawrenson, K., Buckley, M., Fridley, B.L., Tyrer, J.P., Shen, H., et al. (2013). GWAS meta-analysis and replication identifies three new susceptibility loci for ovarian cancer. Nat. Genet. *45*, 362–370, 370e1.

74. Viré, E., Curtis, C., Davalos, V., Git, A., Robson, S., Villanueva, A., Vidal, A., Barbieri, I., Aparicio, S., Esteller, M., et al. (2014). The breast cancer oncogene EMSY represses transcription of antimetastatic microRNA miR-31. Mol. Cell *53*, 806–818.

75. Montano, M.M., Yeh, I.-J., Chen, Y., Hernandez, C., Kiselar, J.G., de la Fuente, M., Lawes, A.M., Nieman, M.T., Kiser, P.D., Jacobberger, J., et al. (2019). Inhibition of the histone demethylase, KDM5B, directly induces re-expression of tumor suppressor protein HEXIM1 in cancer cells. Breast Cancer Res. *21*, 138.

76. Fischer, H., Chen, J., Skoog, L., and Lindblom, A. (2002). Cyclin D2 expression in familial and sporadic breast cancer. Oncol. Rep. *9*, 1157–1161.

77. Hung, C.-S., Wang, S.-C., Yen, Y.-T., Lee, T.-H., Wen, W.-C., and Lin, R.-K. (2018). Hypermethylation of CCND2 in lung and breast cancer is a potential biomarker and drug target. Int. J. Mol. Sci. *19*,.

78. Gayther, S.A., Song, H., Ramus, S.J., Kjaer, S.K., Whittemore, A.S., Quaye, L., Tyrer, J., Shadforth, D., Hogdall, E., Hogdall, C., et al. (2007). Tagging single nucleotide polymorphisms in cell cycle control genes and susceptibility to invasive epithelial ovarian cancer. Cancer Res. *67*, 3027–3035.

79. Song, H., Hogdall, E., Ramus, S.J., Dicioccio, R.A., Hogdall, C., Quaye, L., McGuire, V., Whittemore, A.S., Shah, M., Greenberg, D., et al. (2008). Effects of common germ-line genetic variation in cell cycle genes on ovarian cancer survival. Clin. Cancer Res. *14*, 1090–1095.

80. Asem, M.S., Buechler, S., Wates, R.B., Miller, D.L., and Stack, M.S. (2016). Wnt5a signaling in cancer. Cancers (Basel) *8*,.

81. Kobayashi, Y., Kadoya, T., Amioka, A., Hanaki, H., Sasada, S., Masumoto, N., Yamamoto, H., Arihiro, K., Kikuchi, A., and Okada, M. (2018). Wnt5a-induced cell migration is associated with the aggressiveness of estrogen receptor-positive breast cancer. Oncotarget *9*, 20979–20992.

82. Han, B., Zhou, B., Qu, Y., Gao, B., Xu, Y., Chung, S., Tanaka, H., Yang, W., Giuliano, A.E., and Cui, X. (2018). FOXC1-induced non-canonical WNT5A-MMP7 signaling regulates invasiveness in triple-negative breast cancer. Oncogene *37*, 1399–1408.

83. Ford, C.E., Punnia-Moorthy, G., Henry, C.E., Llamosas, E., Nixdorf, S., Olivier, J., Caduff, R., Ward, R.L., and Heinzelmann-Schwarz, V. (2014). The non-canonical Wnt ligand, Wnt5a, is upregulated and associated with epithelial to mesenchymal transition in epithelial ovarian cancer. Gynecol. Oncol. *134*, 338–345.

84. Qi, H., Sun, B., Zhao, X., Du, J., Gu, Q., Liu, Y., Cheng, R., and Dong, X. (2014). Wnt5a promotes vasculogenic mimicry and epithelial-mesenchymal transition via protein kinase Cα in epithelial ovarian cancer. Oncol. Rep. *32*, 771–779.

85. Assefnia, S., Dakshanamurthy, S., Guidry Auvil, J.M., Hampel, C., Anastasiadis, P.Z., Kallakury, B., Uren, A., Foley, D.W., Brown, M.L., Shapiro, L., et al. (2014). Cadherin-11 in poor prognosis malignancies and rheumatoid arthritis: common target, common therapies. Oncotarget *5*, 1458–1474.

86. Simanshu, D.K., Nissley, D.V., and McCormick, F. (2017). RAS proteins and their regulators in human disease. Cell *170*, 17–33.

87. Carter, J.H., Deddens, J.A., Mueller, G., Lewis, T.G., Dooley, M.K., Robillard, M.C., Frydl, M., Duvall, L., Pemberton, J.O., and Douglass, L.E. (2018). Transcription factors WT1 and p53 combined: a prognostic biomarker in ovarian cancer. Br. J. Cancer *119*, 462–470.

88. Catalona, W.J., Richie, J.P., Ahmann, F.R., Hudson, M.A., Scardino, P.T., Flanigan, R.C., DeKernion, J.B., Ratliff, T.L., Kavoussi, L.R., Dalkin, B.L., et al. (1994). Comparison of digital rectal examination and serum prostate specific antigen in the early detection of prostate cancer: results of a multicenter clinical trial of 6,630 men. J. Urol. *151*, 1283–1290.

89. Zhao, J., Zhao, D., Poage, G.M., Mazumdar, A., Zhang, Y., Hill, J.L., Hartman, Z.C., Savage, M.I., Mills, G.B., and Brown, P.H. (2015). Death-associated protein kinase 1 promotes growth of p53-mutant cancers. J. Clin. Invest. *125*, 2707–2720.

90. Zhao, B., Tumaneng, K., and Guan, K.-L. (2011). The Hippo pathway in organ size control, tissue regeneration and stem cell self-renewal. Nat. Cell Biol. *13*, 877–883.

91. Qi, C., Zhu, Y.T., Hu, L., and Zhu, Y.-J. (2009). Identification of Fat4 as a candidate tumor suppressor gene in breast cancers. Int. J. Cancer *124*, 793–798.

92. Lin, X., Spindler, T.J., de Souza Fonseca, M.A., Corona, R.I., Seo, J.-H., Dezem, F.S., Li, L., Lee, J.M., Long, H.W., Sellers, T.A., et al. (2019). Super-Enhancer-Associated LncRNA UCA1 Interacts Directly with AMOT to Activate YAP Target Genes in Epithelial Ovarian Cancer. IScience *17*, 242–255.

93. Hall, C.A., Wang, R., Miao, J., Oliva, E., Shen, X., Wheeler, T., Hilsenbeck, S.G., Orsulic, S., and Goode, S. (2010). Hippo pathway effector Yap is an ovarian cancer oncogene. Cancer Res. *70*, 8517–8525.

94. Locksley, R.M., Killeen, N., and Lenardo, M.J. (2001). The TNF and TNF receptor superfamilies: integrating mammalian biology. Cell *104*, 487–501.

95. Martínez-Reza, I., Díaz, L., and García-Becerra, R. (2017). Preclinical and clinical aspects of TNF-α and its receptors TNFR1 and TNFR2 in breast cancer. J. Biomed. Sci. *24*, 90.

96. Thompson, M.S., and Mok, S.C. (2009). Immunopathogenesis of ovarian cancer. Minerva Med. *100*, 357–370.

97. Vanamee, É.S., and Faustman, D.L. (2017). TNFR2: A novel target for cancer immunotherapy. Trends Mol. Med. *23*, 1037–1046.

98. Liu, X., Li, Y.I., and Pritchard, J.K. (2019). Trans effects on gene expression can drive omnigenic inheritance. Cell *177*, 1022-1034.e6.

99. Monahan, K., Horta, A., and Lomvardas, S. (2019). LHX2- and LDB1-mediated trans interactions regulate olfactory receptor choice. Nature *565*, 448–453.

100. Zhao, Z., Tavoosidana, G., Sjölinder, M., Göndör, A., Mariano, P., Wang, S., Kanduri, C., Lezcano, M., Sandhu, K.S., Singh, U., et al. (2006). Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. Nat. Genet. *38*, 1341–1347.

101. Noordermeer, D., de Wit, E., Klous, P., van de Werken, H., Simonis, M., Lopez-Jones, M., Eussen, B., de Klein, A., Singer, R.H., and de Laat, W. (2011). Variegated gene expression caused by cell-specific long-range DNA interactions. Nat. Cell Biol. *13*, 944–951.

102. Schoenfelder, S., and Fraser, P. (2019). Long-range enhancer-promoter contacts in gene expression control. Nat. Rev. Genet. *20*, 437–455.

103. Patel, B., Kang, Y., Cui, K., Litt, M., Riberio, M.S.J., Deng, C., Salz, T., Casada, S., Fu, X., Qiu, Y., et al. (2014). Aberrant TAL1 activation is mediated by an interchromosomal interaction in human T-cell acute lymphoblastic leukemia. Leukemia *28*, 349–361.
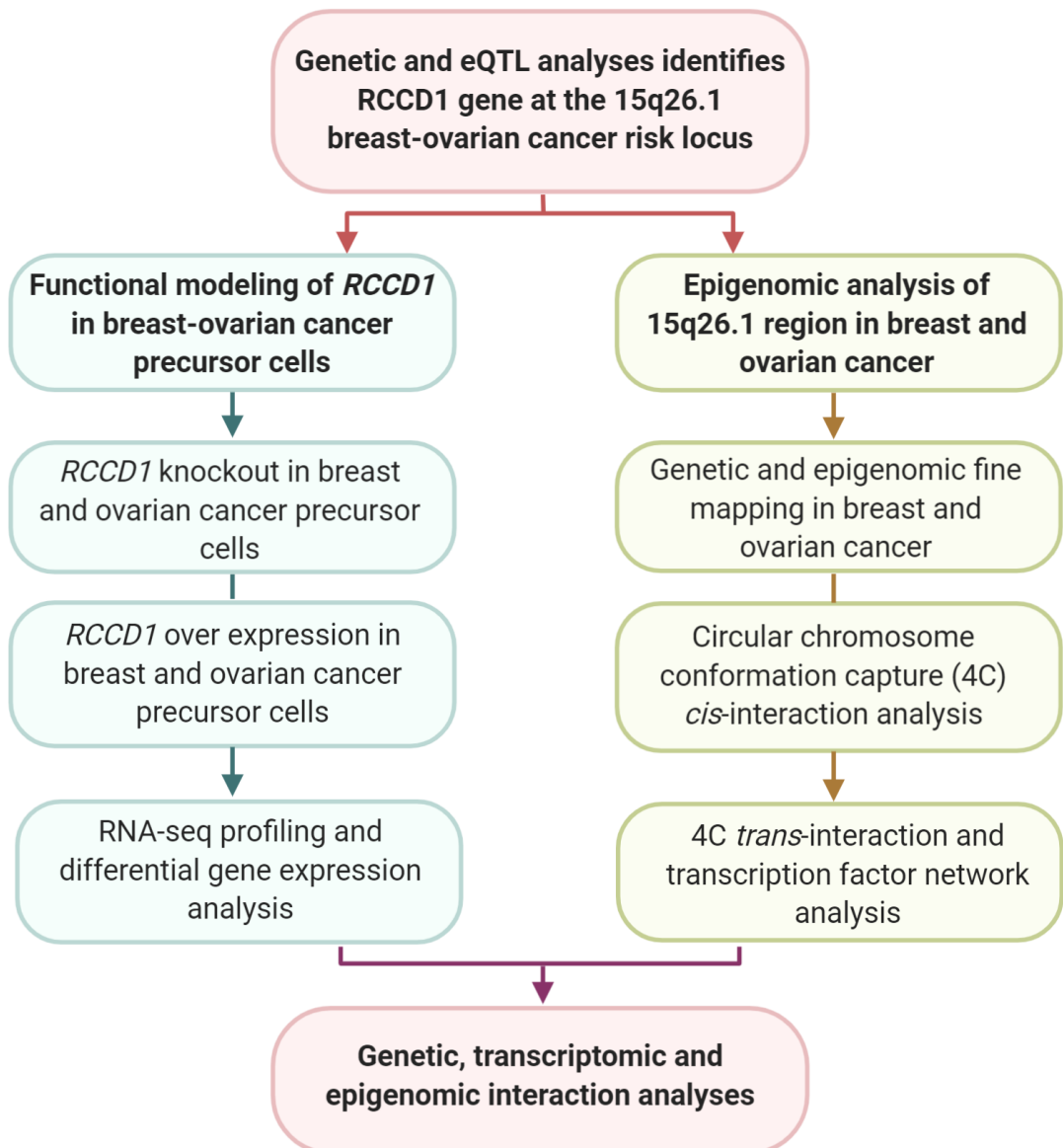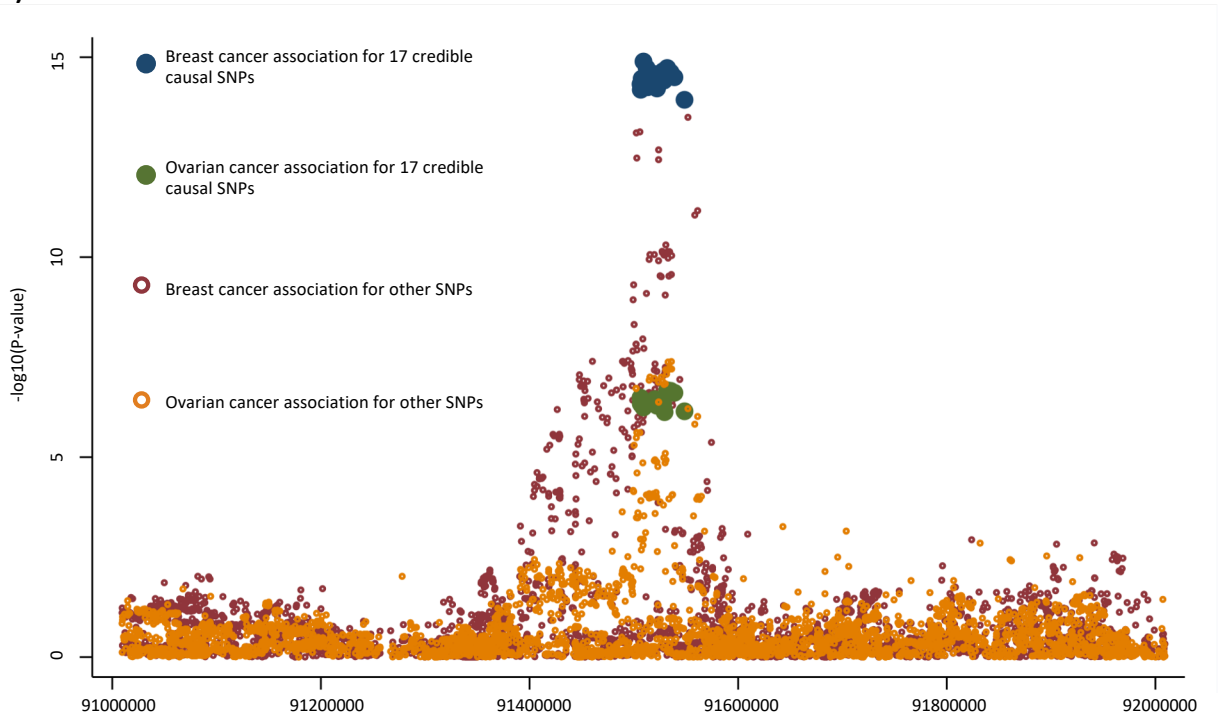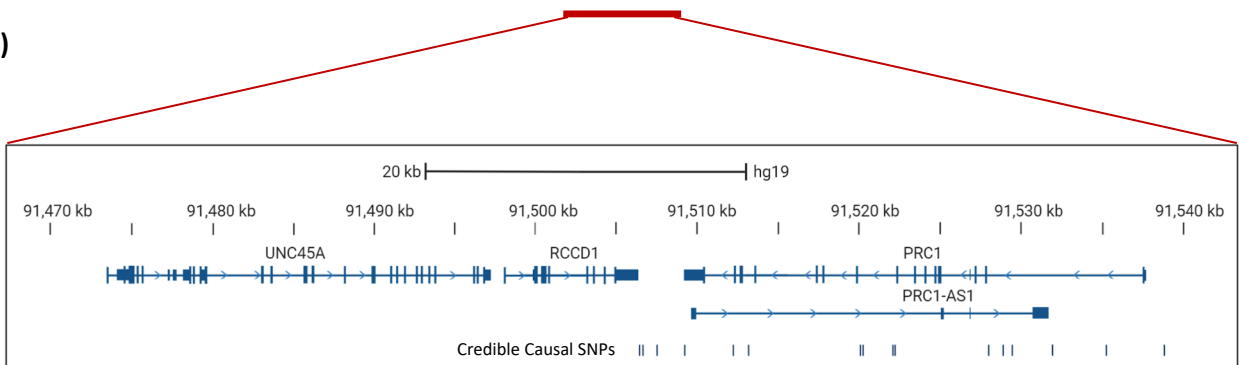
**FIGURE 1**

# FIGURE 2

**(A)**



**(B)**



**(C)**

# FIGURE 3

**(A)**



**(B)**



**(C)**

# FIGURE 4



**(A)**

- Breast trans interaction
- Ovarian trans interaction
- Shared trans interaction

**(B)** DEGs in Trans vs DEGs in Random simulated trans

4C Trans

Random

**(C)** Breast Trans TFBS

Q-significance
370
308
245
182
120

# of overlaps
470
372
275
178
80

Mapped peaks ratio

**(D)** Ovarian Trans TFBS

Q-significance
140
115
90
65
40

# of overlaps
200
160
130
80
40

Mapped peaks ratio

**FIGURE 5**



**(A)** Breast Precursor

**(B)** Ovarian Precursor

Control OE
RCCD1 OE
RCCD1 KO
Control KO

**(C)** Breast DEGs (RCCD1 OE vs Control)

**(D)** Ovarian DEGs (RCCD1 OE vs Control)

# FIGURE 6