Pleiotropy-guided transcriptome imputation from normal and tumor tissues identifies new candidate susceptibility genes for breast and ovarian cancer

Siddhartha P. Kar[1,2*†], Daniel P. C. Considine[3*], Jonathan P. Tyrer[3], Jasmine T. Plummer[4], Stephanie Chen[4], Felipe S. Dezem[4], Alvaro N. Barbeira[5], Padma S. Rajagopal[5,6], Will Rosenow[7], Fernando M. Antón[8], Clara Bodelon[9], Jenny Chang-Claude[10,11], Georgia Chenevix-Trench[12], Anna deFazio[13,14], Thilo Dörk[15], Arif B. Ekici[16,17], Ailith Ewing[18], George Fountzilas[19], Ellen L. Goode[20], Mikael Hartman[21,22], Florian Heitz[23,24], Peter Hillemanns[25], Estrid Høgdall[26,27], Claus K. Høgdall[28], Tomasz Huzarski[29,30], Allan Jensen[26], Beth Y. Karlan[31], Elza Khusnutdinova[32,33], Lambertus A. Kiemeney[34], Susanne K. Kjaer[26,35], Rüdiger Klapdor[25], Martin Köbel[36], Jingmei Li[21,37], Clemens Liebrich[38], Taymaa May[39], Håkan Olsson[40], Jennifer B. Permuth[41], Paolo Peterlongo[42], Paolo Radice[43], Susan J. Ramus[44,45], Marjorie J. Riggan[46], Harvey A. Risch[47], Emmanouil Saloustros[48], Jacques Simard[49], Lukasz M. Szafron[50], Cheryl L. Thompson[51], Robert A. Vierkant[52], Stacey J. Winham[52], Wei Zheng[53], Jennifer A. Doherty[54], Andrew Berchuck[55], Kate L. Lawrenson[4,56], Hae K. Im[5], Ani W. Manichaikul[7,57], Paul D. P. Pharoah[3,58], Simon A. Gayther[4], Joellen M. Schildkraut[59]

1 Medical Research Council Integrative Epidemiology Unit, University of Bristol, Bristol, UK
2 Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK
3 Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK
4 Center for Bioinformatics and Functional Genomics, Department of Biomedical Science, Cedars-Sinai Medical Center, Los Angeles, CA
5 Section of Genetic Medicine, Department of Medicine, University of Chicago, Chicago, IL
6 Section of Hematology/Oncology, Department of Medicine, University of Chicago, Chicago, IL
7 Center for Public Health Genomics, University of Virginia, Charlottesville, VA
8 Hospital Clínico San Carlos, Madrid, Spain
9 Divison of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD
10 Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany
11 Cancer Epidemiology Group, University Cancer Center Hamburg (UCCH), University Medical Center Hamburg-Eppendorf, Hamburg, Germany
12 Department of Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Brisbane, Australia
13 Centre for Cancer Research, The Westmead Institute for Medical Research, The University of Sydney, Sydney, Australia
14 Department of Gynaecological Oncology, Westmead Hospital, Sydney, Australia
15 Gynaecology Research Unit, Hannover Medical School, Hannover, Germany
16 Institute of Human Genetics, University Hospital Erlangen, Erlangen, Germany
17 Friedrich-Alexander University Erlangen-Nuremberg, Comprehensive Cancer Center Erlangen, Erlangen, Germany
18 MRC Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine, The University of Edinburgh, Western General Hospital, Edinburgh, UK
19 Laboratory of Molecular Oncology, Hellenic Foundation for Cancer Research, Aristotle University of Thessaloniki School of Medicine, Thessaloniki, Greece
20 Department of Health Science Research, Division of Epidemiology, Mayo Clinic, Rochester, MN
21 Department of Surgery, Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, Singapore
22 Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore
23 Department of Gynecology and Gynecologic Oncology, Kliniken Essen-Mitte/ Evang., Essen, Germany
24 Department of Gynecology, Center for Oncologic Surgery, Charité Campus Virchow-Klinikum, Berlin, Germany
25 Department of Gynecology and Obstetrics, Hannover Medical School, Hannover, Germany
26 Department of Virus, Lifestyle and Genes, Danish Cancer Society Research Center, Copenhagen, Denmark
27 Molecular Unit, Department of Pathology, Herlev Hospital, University of Copenhagen, Copenhagen, Denmark
28 The Juliane Marie Centre, Department of Gynecology, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark

29 Department of Genetics and Pathology, International Hereditary Cancer Center, Pomeranian Medical University, Szczecin, Poland
30 Department of Genetics and Pathology, University of Zielona Góra, Zielona Góra, Poland
31 David Geffen School of Medicine, Department of Obstetrics and Gynecology, University of California at Los Angeles, Los Angeles, CA
32 Institute of Biochemistry and Genetics, Ufa Federal Research Centre of the Russian Academy of Sciences, Ufa, Russia
33 Department of Genetics and Fundamental Medicine, Bashkir State University, Ufa, Russia
34 Radboud Institute for Health Sciences, Radboud University Medical Center, Nijmegen, The Netherlands
35 Department of Gynaecology, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark
36 Department of Pathology and Laboratory Medicine, University of Calgary, Foothills Medical Center, Calgary, Canada
37 Genome Institute of Singapore, Human Genetics, Singapore
38 Department of Obstetrics and Gynecology, Klinikum Wolfsburg, Wolfsburg, Germany
39 Division of Gynecologic Oncology, University Health Network, Princess Margaret Hospital, Toronto, Canada
40 Division of Oncology, Department of Clinical Sciences, Lund University, Lund, Sweden
41 Departments of Cancer Epidemiology and Gastrointestinal Oncology, Moffitt Cancer Center and Research Institute, Tampa, FL
42 Genome Diagnostics Program, IFOM-The FIRC (Italian Foundation for Cancer Research) Institute of Molecular Oncology, Milan, Italy
43 Unit of Molecular Bases of Genetic Risk and Genetic Testing, Department of Research, Fondazione IRCCS Istituto Nazionale dei Tumori (INT), Milan, Italy
44 Adult Cancer Program, Lowy Cancer Research Centre, University of New South Wales, Sydney, Australia
45 The Kinghorn Cancer Centre, Garvan Institute of Medical Research, University of New South Wales, Sydney, Australia
46 Department of Obstetrics and Gynecology, Duke University Medical Center, Durham, NC
47 Department of Chronic Disease Epidemiology, Yale School of Public Health, New Haven, CT
48 Department of Oncology, University Hospital of Larissa, Larissa, Greece
49 Genomics Center, Centre Hospitalier Universitaire de Québec - Université Laval, Research Center, Québec City, Canada
50 Maria Sklodowska-Curie Memorial Cancer Center and Institute of Oncology, Warsaw, Poland
51 Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH
52 Department of Health Sciences Research, Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN
53 Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, TN
54 Department of Population Health Sciences, Huntsman Cancer Institute, Salt Lake City, UT
55 Division of Gynecologic Oncology, Department of Obstetrics and Gynecology, Duke University Medical Center, Durham, NC
56 Women's Cancer Program at the Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, CA
57 Department of Public Health Sciences, University of Virginia, Charlottesville, VA
58 Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK
59 Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA

*** Indicates equal contribution**

† Corresponding author:

Siddhartha Kar, MBBS, PhD, MPH
MRC Integrative Epidemiology Unit
University of Bristol, Oakfield House
Oakfield Grove, Clifton BS8 2BN, UK
Email: siddhartha.kar@bristol.ac.uk

Conflict of Interest statement: The authors declare no potential conflicts of interest.

Abstract

Familial, genome-wide association (GWAS), and sequencing studies and genetic correlation analyses have progressively unraveled the shared or pleiotropic germline genetics of breast and ovarian cancer. In this study, we aimed to leverage this shared germline genetics to improve the power of transcriptome-wide association studies (TWAS) to identify candidate breast cancer and ovarian cancer susceptibility genes. We built gene expression prediction models using the PrediXcan method in 681 breast and 295 ovarian tumors from The Cancer Genome Atlas and 211 breast and 99 ovarian normal tissue samples from the Genotype-Tissue Expression project and integrated these with GWAS meta-analysis data from the Breast Cancer Association Consortium (122,977 cases/105,974 controls) and the Ovarian Cancer Association Consortium (22,406 cases/40,941 controls). The integration was achieved through novel application of a pleiotropy-guided conditional/conjunction false discovery rate approach for the first time in the setting of a TWAS. This identified 14 new candidate breast cancer susceptibility genes spanning 11 genomic regions and 8 new candidate ovarian cancer susceptibility genes spanning 5 genomic regions at conjunction FDR < 0.05 that were > 1 Mb away from known breast and/or ovarian cancer susceptibility loci. We also identified 38 candidate breast cancer susceptibility genes and 17 candidate ovarian cancer susceptibility genes at conjunction FDR < 0.05 at known breast and/or ovarian susceptibility loci. Overlaying candidate causal risk variants identified by GWAS fine mapping onto expression prediction models for genes at known loci suggested that the association for 55% of these genes was driven by the underlying GWAS signal.

Significance

The 22 new genes identified by our cross-cancer analysis represent promising candidates that further elucidate the role of the transcriptome in mediating germline breast and ovarian cancer risk.

Introduction

The last three decades have witnessed major advances in our understanding of the shared inherited genetic basis of breast and ovarian cancer. The identification of rare inherited mutations in *BRCA1* (1) and *BRCA2* (2) that confer high risks of developing both breast and ovarian cancer has directly opened up the identification of novel oncogenic mechanisms leading to the development of poly ADP ribose polymerase inhibitor therapy (3). The findings from genome-wide association studies (GWAS) have demonstrated that there is a strong genetic correlation between breast and ovarian cancer (4) and have identified several genomic regions containing common (minor allele frequency > 1%) variants that confer risk of developing both breast and ovarian cancer (5,6).

Transcriptome-wide association studies (TWAS) represent the latest study design for the identification of disease-associated susceptibility genes. TWAS involve establishing robust multi-variant models for the component of somatic (normal or tumor) gene expression that is regulated by germline genetic variation in a smaller data set where both germline genotype and somatic transcriptomic data are available. These models are then used to impute the germline genetically regulated component of gene expression into a larger GWAS data set where measured gene expression is unavailable but that offers significantly improved power to identify genes associated with disease risk where such risk may be mediated by expression. Moving from single variants (GWAS) to genes (TWAS) as the unit of association reduces the multiple testing burden. The use of gene expression provides a readily accessible read-out of the functional basis of the identified association in contrast to GWAS-identified risk variants that predominantly reside in non-coding regions of the genome (7).

PrediXcan is a method developed recently for conducting TWAS (8). TWAS methods have been applied to single cancer types before, including breast cancer (9,10) and ovarian cancer (11,12). Here we present the first application of PrediXcan, and indeed broadly of TWAS, in the pleiotropic cross-cancer setting. We used the normal and tumor breast- and ovary-specific gene expression and matched germline genotype data sets to generate tissue-specific PrediXcan models and first imputed these models into GWAS data for the corresponding cancers (i.e., from breast tissue-derived models into breast cancer GWAS and likewise for the ovarian models). We then imputed models across cancer types (i.e., from breast tissue-derived models into ovarian cancer GWAS and *vice versa*). Finally, we implemented a powerful conjunction false discovery rate approach (13,14) that has been applied previously to GWAS (15–18), but not to TWAS, to leverage the combined GWAS sample of over 145,000 breast and ovarian cancer cases. We identify new candidate breast and ovarian cancer susceptibility genes in regions not previously implicated by GWAS or TWAS analyses of these cancers.

Methods

Matched germline genotype – normal/tumor gene expression data sets

We used data for 211 normal breast tissue samples and 99 normal ovarian tissue samples from the Genotype-Tissue Expression (GTEx) project (version 7 release; (19)). Germline genotypes in the GTEx data had been called from whole-genome sequencing (Illumina HiSeq X) and gene expression was profiled using RNA-Sequencing (Illumina TruSeq). We also used data from 681 breast cancer (20) and 295 high-grade serous ovarian cancer [HGSOC; (21)] cases from The Cancer Genome Atlas (TCGA) network. Germline genotypes in the TCGA data had been called from genotyping arrays (Affymetrix SNP 6.0) and gene expression was profiled using RNA-Sequencing (Illumina HiSeq 2000). Imputation of TCGA germline genotypes using the 1000 Genomes version 5 reference panel was performed as described previously (22,23). The TCGA sample sizes reported here refer to only those samples that had > 95% European ancestry. Ancestry was estimated using the Local Ancestry in adMixed Populations tool (LAMP version 2.5; (24)). Downstream PrediXcan modelling (described below) used variants imputed with quality > 0.8 that had a minor allele frequency > 5% in TCGA data sets.

Genome-wide association data sets

Summary statistics from genome-wide association meta-analyses were obtained from the Breast Cancer Association Consortium (BCAC; (22)) and the Ovarian Cancer Association Consortium (OCAC; (23)). The breast cancer susceptibility data were based on 122,977 cases and 105,974 controls, including 21,468 estrogen receptor (ER)-negative cases. The ovarian cancer susceptibility data were based on 22,406 epithelial ovarian cancer cases and 40,941 controls, including 13,037 HGSOC cases. We harmonised the signs of the effect size estimates and aligned them to the same effect allele in the breast and ovarian cancer GWAS data sets. We retained 9,530,997 variants with minor allele frequency > 1% and imputation quality > 0.4 in both data sets for S-PrediXcan analyses. All individuals in these studies were of genetically inferred European ancestry.

PrediXcan model development and S-PrediXcan analyses

We built genetically regulated gene expression prediction models using the elastic net regularization approach implemented in PrediXcan and validated these models using tenfold cross-validation (8). Essentially, this generates a list of variants for each gene where model construction is successful and each variant in the list is assigned a weight reflecting its influence on its target gene expression. Genes with models where the nested tenfold cross-validated correlation between predicted and

actual levels of expression was > 10% (predictive performance $r^2 > 0.01$) and *P*-value of the correlation test was < 0.05 were retained. These models were adjusted for the latent determinants of gene expression variation (referred to hereafter as "PEER factors"), which were identified using the Probabilistic Estimation of Expression Residuals (PEER; version 1.3) method (25). We adjusted for 60 and 45 PEER factors for TCGA breast and ovarian cancer data, respectively. The choice of these numbers is a function of sample size and consistent with recommendations (8,25). *ESR1* expression was also included as a covariate in the construction of breast cancer models to account for estrogen receptor status and its influence on the expression of individual genes. For the GTEx version 7 data sets, we downloaded pre-computed PrediXcan models from predictdb.org. Our pipeline for processing the TCGA data sets, including the application of PEER factors, was designed to be consistent with the pipeline used to generate the pre-computed GTEx PrediXcan models. S-PrediXcan refers to the application of the PrediXcan gene expression models, specifically the variant weights from elastic net combined into multi-variant gene-level instruments, to summary statistics GWAS data sets and has been described in detail before (8).

Conditional and conjunction false discovery rate analyses

We obtained *P*-values for association of predicted expression of each gene with breast cancer risk and with ovarian cancer risk. We then computed the false discovery rate (FDR) for gene-breast cancer risk association conditional on gene-ovarian cancer risk association (as conditional $FDR_{\text{Breast Cancer}|\text{Ovarian Cancer}}$). This is the probability that a gene is not associated with breast cancer risk given the P-values for association with both breast cancer risk and ovarian cancer risk. The analogous conditional FDR for gene-ovarian cancer risk association was also calculated ($FDR_{\text{Ovarian Cancer}|\text{Breast Cancer}}$). Finally, the conjunctional FDR estimate, which is conservatively defined as the maximum of the two conditional FDR values, was computed. This process minimizes the effect of a single phenotype (in this case, breast or ovarian cancer) driving the shared association signal. It allows the power of pleiotropic associations to be tapped for genetic discovery unlike a traditional FDR approach that is informed solely by the distribution of *P*-values for a single phenotype. We used the R implementation of the conditional FDR method available from github.com/KehaoWu/GWAScFDR. The conditional and conjunctional FDR method has been described extensively elsewhere (13–18), but not applied before to the TWAS setting. The overall study design is summarized in Figure 1.

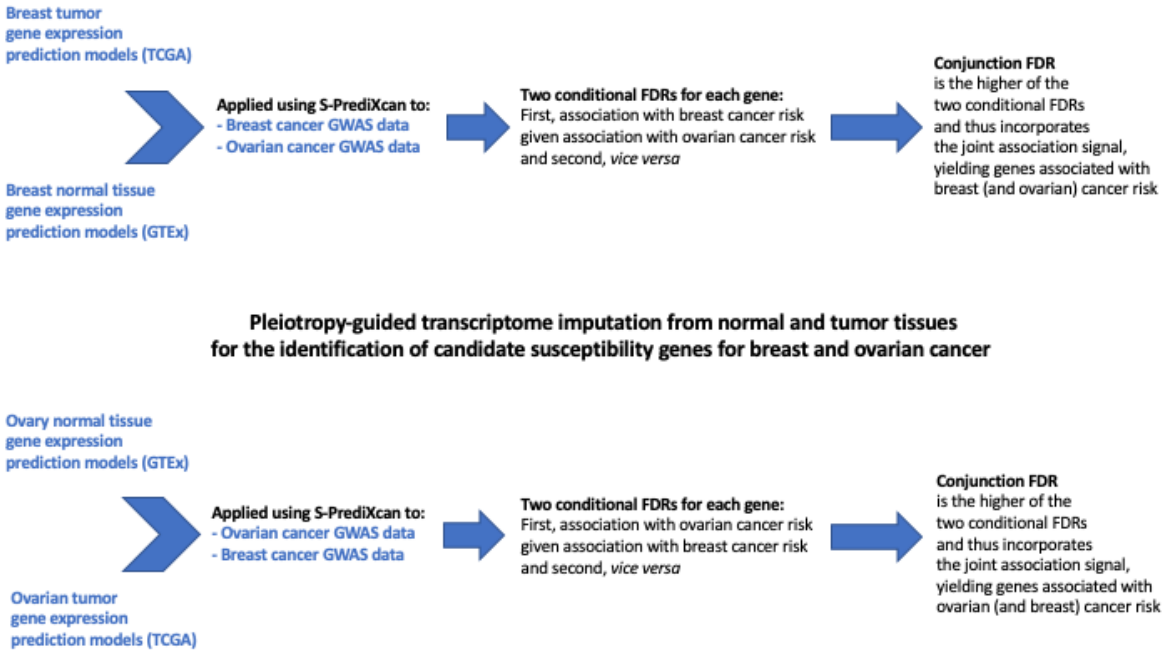

Figure 1: Flowchart providing an overview of the data sets used and the various steps in the analysis. GTEx: Genotype-Tissue Expression project; TCGA: The Cancer Genome Atlas; GWAS: genome-wide association study; FDR: false discovery rate.

Fine-mapped candidate causal risk variant data sets

We examined the overlap between variants in the breast gene expression prediction models and a published list of fine-mapped candidate causal risk variants for breast cancer (26). This was done to follow-up genes that we identified in genomic regions that are known to be associated with breast cancer risk under the intuition that gene-level association signals identified by S-PrediXcan that demonstrate such overlap with fine-mapped variants are likely being driven by the GWAS association signal in the same region.

Fine-mapped candidate causal risk variants lists for breast cancer were obtained from Fachal et al (26). Briefly, Fachal et al. fine-mapped 150 known breast cancer susceptibility regions using dense genotype data on women participating in the BCAC and in the Consortium of Investigators of Modifiers of BRCA1/2 (CIMBA). Stepwise multinomial logistic regression was used to identify independent association signals in each region. Credible causal variants within each signal were defined as being within a 100-fold likelihood of the top conditional variant to delineate the variants driving the GWAS associations in each region.

Results

Development of tissue/tumor-specific gene expression prediction models

We built genetically regulated gene expression predictor models using matched germline genotype and tumor gene expression data from TCGA by applying elastic net regularization as implemented in the PrediXcan software. Genes with models where the nested tenfold cross-validated correlation between predicted and actual levels of expression was > 10% (predictive performance $r^2 > 0.01$) and *P*-value of the correlation test was < 0.05 were retained in line with best practice quality control recommendations by the developers of PrediXcan (8). We constructed and evaluated predictor models that met these criteria for 4,457 genes based on 681 TCGA breast tumor samples and for 2,705 genes based on 295 TCGA ovarian tumor samples. We obtained pre-computed genetically

regulated gene expression predictor models that met the same criteria (predictive performance r^2 > 0.01; correlation test P < 0.05) in matched germline genotype and normal tissue gene expression data from the GTEx Project. Specifically, the pre-computed data included 5,274 genes modelled based on 211 GTEx breast tissue samples and 3,034 genes modelled based on 99 GTEx ovarian tissue samples.

Imputation of gene expression into GWAS and pleiotropy-guided FDR control

We used the GTEx normal breast tissue-derived prediction models to impute genetically regulated gene expression in a genome-wide association meta-analysis involving 122,977 breast cancer cases and 105,974 controls using S-PrediXcan. We tested for association between imputed gene expression and breast cancer risk. We also used the same GTEx breast tissue-based models to impute gene expression in a genome-wide association meta-analysis of 22,406 ovarian cancer cases and 40,941 controls and test for association between imputed expression and ovarian cancer risk. For these two steps, we applied the conditional FDR method to the S-PrediXcan gene-level association *P*-values to correct for testing 5,274 genes in each analysis. This yielded two conditional FDR values: one for association with breast cancer risk given association with ovarian cancer risk and the other for association with ovarian cancer risk given association with breast cancer risk. Finally, we took the larger of the two values for each gene as a conservative estimate of its conjunction FDR to identify candidate breast cancer susceptibility genes at conjunction FDR < 0.05. We refer to these genes as candidate breast cancer susceptibility genes because they were identified on the basis of gene expression predictor models derived from breast tissue. However, the conditional-conjunction FDR analysis effectively borrowed information from pleiotropic associations with inherited susceptibility to a second cancer type (in this case ovarian cancer) in addition to the primary cancer type (breast cancer) and these genes may be considered as risk genes for the second cancer as well. These steps were repeated for three other ordered combinations of data sets: TCGA breast tumor tissue-breast cancer GWAS-ovarian cancer GWAS to identify candidate breast cancer susceptibility genes; GTEx normal ovarian tissue-ovarian cancer GWAS-breast cancer GWAS and TCGA ovarian tumor tissue-ovarian cancer GWAS-breast cancer GWAS to identify candidate ovarian cancer susceptibility genes. We also replaced the overall breast cancer GWAS and all invasive ovarian cancer GWAS used in the four data set combinations described above with ER-negative breast cancer GWAS (21,468 cases/105,974 controls) and HGSOC GWAS (13,037 cases/22,406 controls), respectively. This helped identify additional candidate breast and ovarian cancer susceptibility genes driven by subtype-specific associations at conjunction FDR < 0.05.

For each gene, coverage was defined as the percentage of the number of variants included in its expression prediction model that were also captured in the genome-wide association meta-analysis. The coverage was >= 80% for at least 93% of the genes in each of the four matched germline genotype and normal or tumor gene expression data sets used to build the predictor models indicating that for most genes, most of the corresponding model variants available were used. In each ordered analytic combination of data sets (e.g., GTEx normal breast tissue-breast cancer GWAS-ovarian cancer GWAS) we observed that, in general, for progressively smaller S-PrediXcan *P*-values of the second cancer type, the true discovery rate for association with the primary cancer type approached 100% at progressively larger S-PrediXcan *P*-values for the primary cancer type (Figure 2 and Supplementary Figure 1). This was consistent with substantial shared gene-level associations for breast and ovarian cancer risk and these shared signals being tapped by the conditional-conjunction FDR method to power candidate susceptibility gene discovery.

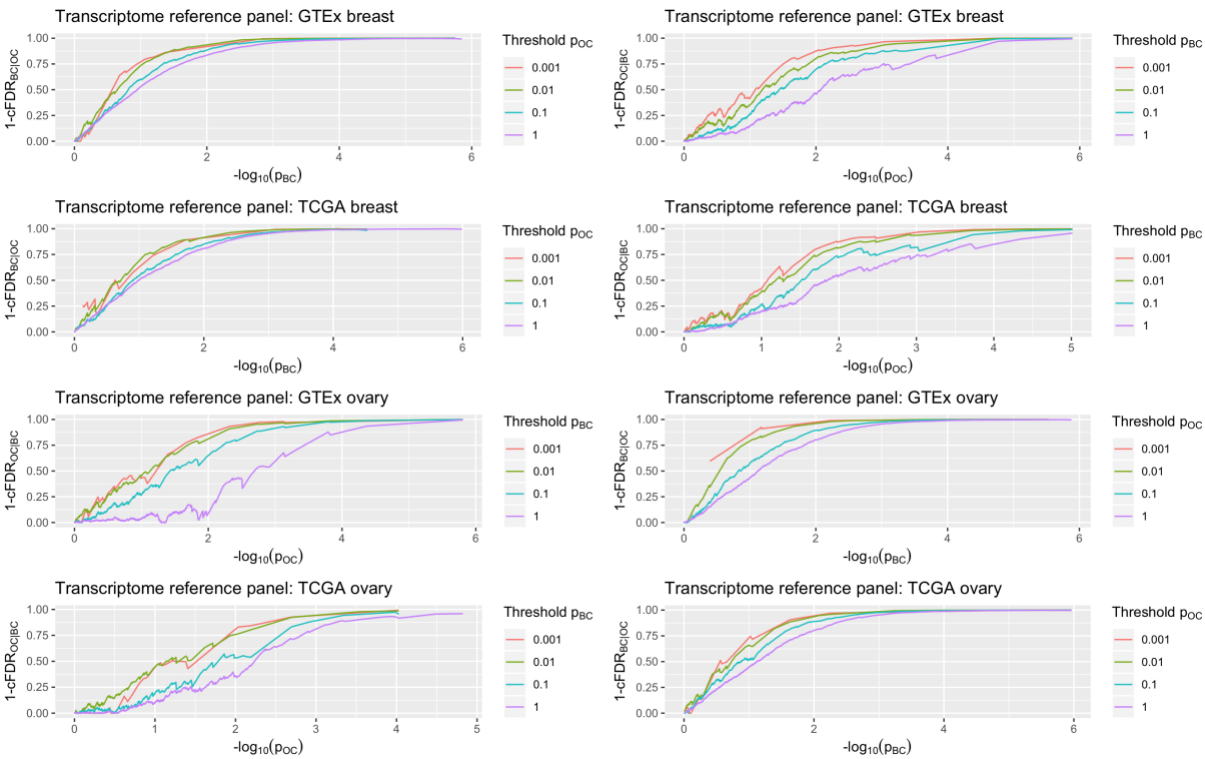

Figure 2: True discovery rate against the negative logarithm (base 10) of the *P*-value for each cancer for subsets of genes based on strength of association with the other cancer. The Y-axis of each plot is the true discovery rate which is defined as 1 – conditional false discovery rate (cFDR). For a given ordered analytic combination of data sets (e.g., GTEx normal breast tissue as transcriptome reference panel-breast cancer GWAS-ovarian cancer GWAS, plotted in the upper left hand corner) we observed that, in general, for progressively smaller S-PrediXcan *P*-values of the second cancer type (indicated by the key "Threshold p" next to each plot), the true discovery rate (Y-axis) for association with the primary cancer type approached 100% at progressively larger S-PrediXcan *P*-values for the primary cancer type (X-axis; negative logarithm (base 10) of the *P*-values). BC: overall breast cancer risk; OC: all invasive ovarian cancer risk. Only *P*-values > 10^{-6} are plotted on the X-axis.

Identification of new candidate breast cancer and ovarian cancer susceptibility genes

We identified 14 new candidate breast cancer susceptibility genes at the conjunction FDR < 0.05 threshold (Table 1 and Supplementary Table 1). The 14 genes were distributed between 11 genomic regions > 1 Mb apart from each other (Table 1). These genes have not been reported as susceptibility genes in any prior TWAS of breast cancer risk and are > 1 Mb away from published genome-wide significant lead variants for breast cancer susceptibility (27). For ovarian cancer, we identified 8 novel candidate susceptibility genes at conjunction FDR < 0.05 (Table 2 and Supplementary Table 2). The 8 genes were located across 5 genomic regions > 1 Mb apart from each other (Table 2). These genes have not been reported as candidate risk genes in any previously reported TWAS of ovarian cancer risk and are > 1 Mb away from published genome-wide significant lead variants for ovarian cancer susceptibility (23).

Table 1: New candidate breast cancer susceptibility genes identified by pleiotropy-guided S-PrediXcan analysis.

Gene	Genomic Region	*P*-value BC	*P*-value OC	Conditional FDR BC|OC	Conditional FDR OC|BC	Conjunction FDR
Transcriptome reference panel: GTEx breast (normal) | primary GWAS: overall BC risk (second GWAS: all invasive OC risk)						
ZSCAN29	15q15.3	1.8E-04	9.1E-04	4.1E-04	6.0E-03	6.0E-03
STRCP1	15q15.3	1.6E-03	1.8E-03	4.1E-03	1.9E-02	1.9E-02
AC011330.5	15q15.3	5.4E-04	2.5E-03	1.8E-03	2.2E-02	2.2E-02
STRC	15q15.3	1.4E-04	3.8E-03	6.1E-04	2.2E-02	2.2E-02
ZNF276	16q24.3	3.7E-06	4.7E-03	2.4E-05	2.2E-02	2.2E-02
RGS19	20q13.33	1.1E-03	5.4E-03	4.3E-03	4.3E-02	4.3E-02
RNFT1	17q23.1	2.4E-04	8.7E-03	1.3E-03	4.7E-02	4.7E-02
C15orf65	15q21.3	2.2E-03	5.9E-03	8.2E-03	4.7E-02	4.7E-02
Transcriptome reference panel: TCGA breast (tumor) | primary GWAS: overall BC risk (second GWAS: all invasive OC risk)						
GMNC	3q28	2.6E-03	1.2E-03	6.1E-03	2.0E-02	2.0E-02
ESRP2	16q22.1	1.9E-02	9.6E-04	4.3E-02	3.3E-02	4.3E-02
BHLHA15	7q21.3	8.5E-05	7.0E-03	5.5E-04	4.9E-02	4.9E-02
SCGB1D2	11q12.3	3.5E-04	5.5E-03	2.0E-03	4.9E-02	4.9E-02
Transcriptome reference panel: TCGA breast (tumor) | primary GWAS: ER-negative BC risk (second GWAS: HGSOC risk)						
ETAA1	2p14	3.0E-03	1.5E-03	2.0E-02	2.0E-02	2.0E-02
ATP8B4	15q21.2	1.6E-03	2.2E-03	1.5E-02	2.4E-02	2.4E-02

Abbreviations: BC, breast cancer; OC, ovarian cancer; FDR, false discovery rate; ER, estrogen receptor; HGSOC, high-grade serous ovarian cancer.

Table 2: New candidate ovarian cancer susceptibility genes identified by pleiotropy-guided S-PrediXcan analysis.

Gene	Genomic Region	*P*-value OC	*P*-value BC	Conditional FDR OC|BC	Conditional FDR BC|OC	Conjunction FDR
Transcriptome reference panel: GTEx ovary (normal) | primary GWAS: all invasive OC risk (second GWAS: overall BC risk)						
STRCP1	15q15.3	7.2E-04	6.4E-05	3.1E-03	8.5E-05	3.1E-03
CPNE1	20q11.22	1.2E-03	7.2E-05	5.0E-03	9.9E-05	5.0E-03
AC011330.5	15q15.3	1.7E-03	2.6E-05	5.8E-03	4.5E-05	5.8E-03
CCNE1	19q12	1.9E-03	3.2E-03	1.4E-02	4.4E-03	1.4E-02
CATSPER2P1	15q15.3	4.8E-03	1.9E-04	1.8E-02	4.1E-04	1.8E-02
UQCC1	20q11.22	3.8E-03	2.5E-03	2.8E-02	4.7E-03	2.8E-02
Transcriptome reference panel: TCGA ovary (tumor) | primary GWAS: all invasive OC risk (second GWAS: overall BC risk)						
CPNE1	20q11.22	2.0E-03	9.0E-05	2.0E-02	4.9E-04	2.0E-02
Transcriptome reference panel: GTEx ovary (normal) | primary GWAS: HGSOC risk (second GWAS: ER-negative BC risk)						
CCNE1	19q12	1.7E-03	2.0E-04	5.9E-03	1.5E-03	5.9E-03
STRCP1	15q15.3	9.2E-03	3.2E-04	3.1E-02	3.9E-03	3.1E-02
HEATR3	16q12.1	4.3E-03	3.1E-02	4.6E-02	4.4E-02	4.6E-02
Transcriptome reference panel: TCGA ovary (tumor) | primary GWAS: HGSOC risk (second GWAS: ER-negative BC risk)						
THSD7A	7p21.3	1.5E-03	1.2E-02	2.8E-02	4.3E-02	4.3E-02

Abbreviations: BC, breast cancer; OC, ovarian cancer; FDR, false discovery rate; ER, estrogen receptor.

Candidate breast cancer and ovarian cancer susceptibility genes at known GWAS loci

We identified 38 candidate breast cancer susceptibility genes that were located within 1 Mb of a published lead variant associated at genome-wide significance with breast cancer risk (Supplementary Table 3; (27)). Four of the 38 genes have also been reported in previously published TWAS (Supplementary Table 3; (9,10)). The 38 genes were spread across 12 genomic regions > 1 Mb apart from each other. Overlaying fine-mapped candidate causal breast cancer risk variants on breast gene expression predictor model variants showed that for 21/38 (55%) genes the prediction model variants included at least one fine-mapped candidate causal variant (Supplementary Tables 3 and 4). This suggested that for these genes the GWAS association signal was driving the S-PrediXcan signal. We also identified three additional genes that were > 1 Mb away from known GWAS loci that have previously been reported as TWAS loci for breast cancer risk (Supplementary Table 3; (9,10)).

For ovarian cancer, we identified 17 candidate susceptibility genes that were located within 1 Mb of a published lead variant associated at genome-wide significance with ovarian cancer risk (Supplementary Table 5; (23)). Six of these genes have also been reported in a previously published TWAS for ovarian cancer (Supplementary Table 5; (11,12)). The 17 genes span 5 different genomic regions > 1 Mb apart.

Discussion

In this study, we used the conditional and conjunctional FDR as a novel tool to systematically improve the power of breast cancer and ovarian cancer candidate susceptibility gene discovery in a PrediXcan-based TWAS. While gene expression prediction models based on multiple tissue types has been the more common approach to improving TWAS power (11,28), the conditional/conjunction FDR approach gains power through the incorporation of multiple related GWAS data sets into a TWAS analysis. We investigated the shared inherited genetic basis of these two cancer types by integrating normal and tumor tissue-specific transcriptomic data sets with large-scale genome-wide association meta-analysis findings for susceptibility to breast cancer and ovarian cancer. We identified 11 new genomic regions associated with breast cancer risk and five new regions linked to ovarian cancer risk.

We identified 14 novel candidate breast cancer susceptibility genes (Table 1). Many of these genes have a strong biological rationale for involvement in breast carcinogenesis and are in or near genomic regions associated with other cancer types or potential cancer risk factors. For example, the *ZNF276* intronic variant rs12925026 is associated at genome-wide significance with non-melanoma skin cancer (29). *ZNF276* overlaps *FANCA* in a tail-to-tail manner (30). The genetically regulated predictor model for *ZNF276* expression was fit using gene expression measured in GTEx breast tissues but neither this data set nor any of the other data sets could capture a predictor model for *FANCA* expression. *FANCA* encodes one of eight subunits that together form the core Fanconi Anemia (FA) complex that repairs blockages in DNA replication due to cross-linking (31). Several members of the FA family of proteins have been implicated in breast and ovarian cancer predisposition including, *BRCA1* (*FANCS*), *BRCA2* (*FANCD1*), *BRIP1* (*FANCJ*), *PALB2* (*FANCN*), *RAD51C* (*FANCO*) and *FANCM* and it is possible that *FANCA* may represent another or possibly the true target breast cancer susceptibility gene in this region given this biological function and its overlap with *ZNF276* (31,32). *ZNF276* in its own right has also been implicated as a candidate tumor suppressor gene in breast cancer (30).

Other candidate breast cancer susceptibility genes we identified include *ESRP2*, which encodes an epithelial cell-specific regulator of splicing of the breast cancer susceptibility gene *FGFR2* (33,34) and *SCGB1D2*, which encodes lipophilin B that is known to be expressed in both breast and

ovarian tumors (35). Lipophilin B is tightly co-expressed with and forms a covalent complex with Mammaglobin A encoded by *SCGB2A2*, the gene next to *SCGB1D2* (35). Mammaglobin A may be used to detect disseminated or circulating tumor cells and is under investigation as a potential immunotherapeutic target in breast cancer (36). However, we were unable to develop gene expression prediction models for *SCGB2A2* in breast normal or tumor tissues. *BHLHA15* encodes an estrogen-regulated transcription factor that is required to maintain mammary gland differentiation in mice (37). *ETAA1* harbors lead variants associated at genome-wide significance with pancreatic cancer (38), and the hormone-related traits of age at menopause (39) and male-pattern baldness (40). It encodes an activator of ATR kinase that accumulates at DNA damage sites and promotes replication fork progression and integrity (41). While our pleiotropy-guided transcriptome imputation study was ongoing, a genome-wide association meta-analysis for breast cancer risk that was performed in parallel identified lead variants rs79518236 (184 kb from *BHLHA15*) and rs9712235 (244 kb from *ETAA1*) at genome-wide significance only on addition of 10,407 breast cancer cases and 7,815 controls to the Michailidou et al. data set used here ((42), unpublished pre-print). There were no known GWAS associations for breast cancer risk in these regions until the larger GWAS meta-analysis and our concomitant identification of the same regions using gene expression imputation into a smaller GWAS underscores the power of leveraging expression data to bolster genetic discovery.

We identified 11 new candidate ovarian cancer susceptibility genes (Table 2). As with breast cancer, there is strong support for a role of several genes in ovarian cancer pathogenesis and many of these genes are in regions of the genome that harbor pleiotropic associations with other cancer types. Variants immediately upstream of *CCNE1* are associated at genome-wide significance with bladder cancer risk (43). *CCNE1* amplification is believed to be an early event in the development of ovarian cancer (44) and is a frequent somatic event in HGSOCs that do not carry homologous recombination DNA repair pathway defects (45). *CCNE1* amplification is also associated with poor prognosis in triple negative breast tumors (46) and it is worth noting that we observed the stronger conjunction FDR association signal for *CCNE1* in the pleiotropy-informed analysis that was based on the HGSOC and ER-negative breast cancer susceptibility GWAS data sets (Table 2). This study is the first to suggest a role for *CCNE1* in conferring ovarian cancer risk. Intronic variants in *HEATR3* are associated at genome-wide significance with glioma in European ancestry individuals (47) and with squamous cell esophageal carcinoma in East Asian ancestry individuals (48). *HEATR3* was also identified by a TWAS of glioma susceptibility (49). Intronic variants in *THSD7A* are associated with epithelial ovarian cancer risk in East Asians (50), albeit not at genome-wide significance (lead variant rs10260419 $P = 1 \times 10^{-7}$). Gene expression prediction models derived from breast and ovarian tissues both implicated the 15q15.3 region as a new breast and ovarian cancer susceptibility region on imputation with these models into the breast and ovarian cancer GWAS data. Our analysis suggested several genes in this region (Tables 1 and 2), with the pseudogene *STRCP1* as the only common gene across breast and ovarian tissues. *STRCP1* overlaps the protein coding gene *STRC*, also identified in the breast tissue-based analysis (Table 1), and variants in *STRC* have previously been associated with lung cancer risk (lung cancer lead variant rs35028925 $P = 2 \times 10^{-6}$) (51).

In this analysis, we chose to label the identified genes as candidate breast cancer susceptibility genes if they were identified on integrating the GTEx or TCGA breast expression prediction models with the breast cancer GWAS and incorporating pleiotropic information from the ovarian cancer GWAS and *vice versa* for candidate ovarian cancer susceptibility genes. However, application of the conjunction FDR over and above the conditional FDR in principle identified genes associated with both cancer types by tapping into GWAS data from both cancers. Therefore, in a sense all these genes may well be regarded as candidate breast *and* ovarian cancer susceptibility genes.

We identified 38 candidate breast cancer susceptibility genes and 17 candidate ovarian cancer susceptibility genes in regions previously implicated by GWAS for breast cancer and ovarian cancer, respectively (Supplementary Tables 3 and 5). The identification of a large number of genes in these regions is unsurprising given that GWAS associations are the key determinant of the S-PrediXcan signal. However, we were able to take full advantage of extensive fine-scale mapping data generated by the Breast Cancer Association Consortium to separately pinpoint those genes where a fine-mapped candidate causal GWAS risk variant for breast cancer was incorporated in the breast PrediXcan model suggesting it drives the gene-based association. Overall, we found this to be the case for 55% of the candidate susceptibility genes identified by PrediXcan in the breast cancer susceptibility regions identified by GWAS. Comprehensive functional follow up of the 19p13.11 breast and ovarian cancer GWAS region suggests that *ABHD8* and *ANKLE1* are the most likely targets in this region (5). While there was no overlap between S-PrediXcan model variants for *ABHD8* and *ANKLE1* and fine-mapped risk variants in this region, S-PrediXcan did detect both genes as candidate causal susceptibility genes, with *ANKLE1* being the only gene that made the cut in both breast and ovarian tissues, suggesting that S-PrediXcan applied to pleiotropic gene dense regions such as 19p13.11 does help short-list the key targets even in the absence of overlap with fine-mapped variants. A total of 21/38 breast and 13/17 ovarian cancer candidate susceptibility genes in the published GWAS regions were clustered at 17q21.31, reflecting the unique long-distance linkage disequilibrium structure of this region (52). This phenomenon has also led to clustering of associations at 17q21.31 in previous TWAS of breast or ovarian cancer risk (9,11).

In conclusion, the powerful combination of pleiotropic breast and ovarian cancer GWAS data sets with transcriptome imputation from normal and tumor breast and ovarian tissues identified a total of 16 novel genomic loci (22 new genes) associated with breast and ovarian cancer risks. Fine-mapping in larger GWAS data sets and deeper laboratory-based functional follow-up studies of these new loci and candidate genes have the potential to provide fresh insights into the common biological underpinnings of breast and ovarian cancer.

Data Access Statement

Genome-wide summary genetic association statistics from BCAC are available at: http://bcac.ccge.medschl.cam.ac.uk/bcacdata/oncoarray/gwas-icogs-and-oncoarray-summary-results/

Genome-wide summary genetic association statistics from OCAC are available at: https://www.ebi.ac.uk/gwas/downloads/summary-statistics

PrediXcan prediction models trained on the GTEx version 7 data (breast and ovarian tissues) are available here: http://hakyimlab.org/post/2017/v7-v6p-analysis/

PrediXcan prediction models trained on the TCGA data (breast and ovarian tumors) will be made publicly available upon publication.

Acknowledgments

The analyses presented in this manuscript were funded by grant number R01CA211574 from the United States National Institutes of Health/National Cancer Institute.

For GTex:

The GTEx Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The data used for the analyses described in this manuscript can be obtained from dbGaP via accession number phs000424.

For TCGA:

The results published here are in part based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga.

For BCAC:

The breast cancer genome-wide association analyses were supported by the Government of Canada through Genome Canada and the Canadian Institutes of Health Research, the 'Ministère de l'Économie, de la Science et de l'Innovation du Québec' through Genome Québec and grant PSR-SIIRI-701, The National Institutes of Health (U19 CA148065, X01HG007492), Cancer Research UK (C1287/A10118, C1287/A16563, C1287/A10710) and The European Union (HEALTH-F2-2009-223175 and H2020 633784 and 634935). All studies and funders are listed in Michailidou et al (2017).

For OCAC:

The ovarian cancer genome-wide association meta-analyses were supported by the US National Institutes of Health (CA1X01HG007491-01 (C.I. Amos), U19-CA148112 (T.A. Sellers), R01-CA149429 (C.M. Phelan) and R01-CA058598 (M.T. Goodman); Canadian Institutes of Health Research (MOP-86727 (L.E. Kelemen)) and the Ovarian Cancer Research Fund (A. Berchuck). The COGS project was funded through a European Commission's Seventh Framework Programme grant (agreement number 223175 - HEALTH-F2–2009-223175). All studies and funders are listed in Phelan et al (2017).

References

1. Miki Y, Swensen J, Shattuck-Eidens D, Futreal PA, Harshman K, Tavtigian S, et al. A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. Science. 1994;266:66–71.

2. Wooster R, Bignell G, Lancaster J, Swift S, Seal S, Mangion J, et al. Identification of the breast cancer susceptibility gene BRCA2. Nature. 1995;378:789–92.

3. Lord CJ, Ashworth A. PARP inhibitors: Synthetic lethality in the clinic. Science. 2017;355:1152–8.

4. Jiang X, Finucane HK, Schumacher FR, Schmit SL, Tyrer JP, Han Y, et al. Shared heritability and functional enrichment across six solid cancers. Nat Commun. 2019;10:431.

5. Lawrenson K, Kar S, McCue K, Kuchenbaeker K, Michailidou K, Tyrer J, et al. Functional mechanisms underlying pleiotropic risk alleles at the 19p13.1 breast-ovarian cancer susceptibility locus. Nat Commun. 2016;7:12675.

6. Kar SP, Beesley J, Amin Al Olama A, Michailidou K, Tyrer J, Kote-Jarai Zs, et al. 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. 2016;6:1052–67.

7. Freedman ML, Monteiro ANA, Gayther SA, Coetzee GA, Risch A, Plass C, et al. Principles for the post-GWAS functional characterization of cancer risk loci. Nat Genet. 2011;43:513–8.

8. Gamazon ER, Wheeler HE, Shah KP, Mozaffari SV, Aquino-Michaels K, Carroll RJ, et al. A gene-based association method for mapping traits using reference transcriptome data. Nat Genet. 2015;47:1091–8.

9. Wu L, Shi W, Long J, Guo X, Michailidou K, Beesley J, et al. A transcriptome-wide association study of 229,000 women identifies new candidate susceptibility genes for breast cancer. Nat Genet. 2018;50:968–78.

10. Ferreira MA, Gamazon ER, Al-Ejeh F, Aittomäki K, Andrulis IL, Anton-Culver H, et al. Genome-wide association and transcriptome studies identify target genes and risk loci for breast cancer. Nat Commun. 2019;10:1741.

11. Gusev A, Lawrenson K, Lin X, Lyra PC, Kar S, Vavra KC, et al. A transcriptome-wide association study of high-grade serous epithelial ovarian cancer identifies new susceptibility genes and splice variants. Nat Genet. 2019;51:815–23.

12. Lu Y, Beeghly-Fadiel A, Wu L, Guo X, Li B, Schildkraut JM, et al. A Transcriptome-Wide Association Study Among 97,898 Women to Identify Candidate Susceptibility Genes for Epithelial Ovarian Cancer Risk. Cancer Res. 2018;78:5419–30.

13. Smeland OB, Frei O, Shadrin A, O'Connell K, Fan C-C, Bahrami S, et al. Discovery of shared genomic loci using the conditional false discovery rate approach. Hum Genet. 2019;

14. Sun L, Craiu RV, Paterson AD, Bull SB. Stratified false discovery control for large-scale hypothesis testing with application to genome-wide association studies. Genet Epidemiol. 2006;30:519–30.

15. Yokoyama JS, Wang Y, Schork AJ, Thompson WK, Karch CM, Cruchaga C, et al. Association Between Genetic Traits for Immune-Mediated Diseases and Alzheimer Disease. JAMA Neurol. 2016;73:691–7.

16. Witoelar A, Jansen IE, Wang Y, Desikan RS, Gibbs JR, Blauwendraat C, et al. Genome-wide Pleiotropy Between Parkinson Disease and Autoimmune Diseases. JAMA Neurol. 2017;74:780–92.

17. Andreassen OA, Djurovic S, Thompson WK, Schork AJ, Kendler KS, O'Donovan MC, et al. Improved detection of common variants associated with schizophrenia by leveraging pleiotropy with cardiovascular-disease risk factors. Am J Hum Genet. 2013;92:197–209.

18. Smeland OB, Frei O, Kauppi K, Hill WD, Li W, Wang Y, et al. Identification of Genetic Loci Jointly Influencing Schizophrenia Risk and the Cognitive Traits of Verbal-Numerical

Reasoning, Reaction Time, and General Cognitive Function. JAMA Psychiatry. 2017;74:1065–75.

19. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. Nat Genet. 2013;45:580–5.

20. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. Nature. 2012;490:61–70.

21. Cancer Genome Atlas Research Network. Integrated genomic analyses of ovarian carcinoma. Nature. 2011;474:609–15.

22. Michailidou K, Lindström S, Dennis J, Beesley J, Hui S, Kar S, et al. Association analysis identifies 65 new breast cancer risk loci. Nature. 2017;551:92–4.

23. Phelan CM, Kuchenbaecker KB, Tyrer JP, Kar SP, Lawrenson K, Winham SJ, et al. Identification of 12 new susceptibility loci for different histotypes of epithelial ovarian cancer. Nat Genet. 2017;49:680–91.

24. Sankararaman S, Sridhar S, Kimmel G, Halperin E. Estimating local ancestry in admixed populations. Am J Hum Genet. 2008;82:290–303.

25. Stegle O, Parts L, Piipari M, Winn J, Durbin R. Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. Nat Protoc. 2012;7:500–7.

26. Fachal L, Aschard H, Beesley J, Barnes DR, Allen J, Kar S, et al. Fine-mapping of 150 breast cancer risk regions identifies 191 likely target genes. Nat Genet. 2020;52:56–73.

27. Lilyquist J, Ruddy KJ, Vachon CM, Couch FJ. Common Genetic Variation and Breast Cancer Risk-Past, Present, and Future. Cancer Epidemiol Biomarkers Prev. 2018;27:380–94.

28. Barbeira AN, Pividori M, Zheng J, Wheeler HE, Nicolae DL, Im HK. Integrating predicted transcriptome from multiple tissues improves association detection. PLoS Genet. 2019;15:e1007889.

29. Visconti A, Duffy DL, Liu F, Zhu G, Wu W, Chen Y, et al. Genome-wide association study in 176,678 Europeans reveals genetic loci for tanning response to sun exposure. Nat Commun. 2018;9:1684.

30. Wong JCY, Gokgoz N, Alon N, Andrulis IL, Buchwald M. Cloning and mutation analysis of ZFP276 as a candidate tumor suppressor in breast cancer. J Hum Genet. 2003;48:668–71.

31. D'Andrea AD. Susceptibility pathways in Fanconi's anemia and breast cancer. N Engl J Med. 2010;362:1909–19.

32. Figlioli G, Bogliolo M, Catucci I, Caleca L, Lasheras SV, Pujol R, et al. The FANCM:p.Arg658* truncating variant is associated with risk of triple-negative breast cancer. NPJ Breast Cancer. 2019;5:38.

33. Warzecha CC, Sato TK, Nabet B, Hogenesch JB, Carstens RP. ESRP1 and ESRP2 are epithelial cell-type-specific regulators of FGFR2 splicing. Mol Cell. 2009;33:591–601.

34. Meyer KB, Maia A-T, O'Reilly M, Teschendorff AE, Chin S-F, Caldas C, et al. Allele-specific up-regulation of FGFR2 increases susceptibility to breast cancer. PLoS Biol. 2008;6:e108.

35. Zafrakas M, Petschke B, Donner A, Fritzsche F, Kristiansen G, Knüchel R, et al. Expression analysis of mammaglobin A (SCGB2A2) and lipophilin B (SCGB1D2) in more than 300 human tumors and matching normal tissues reveals their co-expression in gynecologic malignancies. BMC Cancer. 2006;6:88.

36. Ghersevich S, Ceballos MP. Mammaglobin A: review and clinical utility. Adv Clin Chem. 2014;64:241–68.

37. Zhao Y, Johansson C, Tran T, Bettencourt R, Itahana Y, Desprez P-Y, et al. Identification of a basic helix-loop-helix transcription factor expressed in mammary gland alveolar cells and required for maintenance of the differentiated state. Mol Endocrinol. 2006;20:2187–98.

38. Childs EJ, Mocci E, Campa D, Bracci PM, Gallinger S, Goggins M, et al. Common variation at 2p13.3, 3q29, 7p13 and 17q25.1 associated with susceptibility to pancreatic cancer. Nat Genet. 2015;47:911–6.

39. Kichaev G, Bhatia G, Loh P-R, Gazal S, Burch K, Freund MK, et al. Leveraging Polygenic Functional Enrichment to Improve GWAS Power. Am J Hum Genet. 2019;104:65–75.

40. Yap CX, Sidorenko J, Wu Y, Kemper KE, Yang J, Wray NR, et al. Dissection of genetic variation and evidence for pleiotropy in male pattern baldness. Nat Commun. 2018;9:5407.

41. Bass TE, Luzwick JW, Kavanaugh G, Carroll C, Dungrawala H, Glick GG, et al. ETAA1 acts at stalled replication forks to maintain genome integrity. Nat Cell Biol. 2016;18:1185–95.

42. Zhang H, Ahearn T, Lecarpentier J, Barnes D, Beesley J, Qi G, et al. Genome-wide association study identifies 32 novel breast cancer susceptibility loci from overall and subtype-specific analyses. 2019;

43. Rothman N, Garcia-Closas M, Chatterjee N, Malats N, Wu X, Figueroa JD, et al. A multi-stage genome-wide association study of bladder cancer identifies multiple susceptibility loci. Nat Genet. 2010;42:978–84.

44. Karst AM, Jones PM, Vena N, Ligon AH, Liu JF, Hirsch MS, et al. Cyclin E1 deregulation occurs early in secretory cell transformation to promote formation of fallopian tube-derived high-grade serous ovarian cancers. Cancer Res. 2014;74:1141–52.

45. Bowtell DDL. The genesis and evolution of high-grade serous ovarian cancer. Nat Rev Cancer. 2010;10:803–8.

46. Zhao Z-M, Yost SE, Hutchinson KE, Li SM, Yuan Y-C, Noorbakhsh J, et al. CCNE1 amplification is associated with poor prognosis in patients with triple negative breast cancer. BMC Cancer. 2019;19:96.

47. Melin BS, Barnholtz-Sloan JS, Wrensch MR, Johansen C, Il'yasova D, Kinnersley B, et al. Genome-wide association study of glioma subtypes identifies specific differences in genetic susceptibility to glioblastoma and non-glioblastoma tumors. Nat Genet. 2017;49:789–94.

48. Wu C, Kraft P, Zhai K, Chang J, Wang Z, Li Y, et al. Genome-wide association analyses of esophageal squamous cell carcinoma in Chinese identify multiple susceptibility loci and gene-environment interactions. Nat Genet. 2012;44:1090–7.

49. Atkins I, Kinnersley B, Ostrom QT, Labreche K, Il'yasova D, Armstrong GN, et al. Transcriptome-Wide Association Study Identifies New Candidate Susceptibility Genes for Glioma. Cancer Res. 2019;79:2065–71.

50. Lawrenson K, Song F, Hazelett DJ, Kar SP, Tyrer J, Phelan CM, et al. Genome-wide association studies identify susceptibility loci for epithelial ovarian cancer in east Asian women. Gynecol Oncol. 2019;153:343–55.

51. McKay JD, Hung RJ, Han Y, Zong X, Carreras-Torres R, Christiani DC, et al. Large-scale association analysis identifies new lung cancer susceptibility loci and heterogeneity in genetic susceptibility across histological subtypes. Nat Genet. 2017;49:1126–32.

52. de Jong S, Chepelev I, Janson E, Strengman E, van den Berg LH, Veldink JH, et al. Common inversion polymorphism at 17q21.31 affects expression of multiple genes in tissue-specific manner. BMC Genomics. 2012;13:458.