UCLA
UCLA Previously Published Works

Title

Integrative multi-omics analyses to identify the genetic and functional mechanisms underlying ovarian cancer risk regions.

Permalink

https://escholarship.org/uc/item/59f7n5vk

Journal

American Journal of Human Genetics, 111(6)

Authors

Dareng, Eileen
Coetzee, Simon
Tyrer, Jonathan
et al.

Publication Date

2024-06-06

DOI

10.1016/j.ajhg.2024.04.011

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Integrative multi-omics analyses to identify the genetic and functional mechanisms underlying ovarian cancer risk regions

Authors

Eileen O. Dareng, Simon G. Coetzee, Jonathan P. Tyrer, ..., Michelle R. Jones, Paul D.P. Pharoah, Simon A. Gayther

Correspondence

pp10001@medschl.cam.ac.uk (P.D.P.P.), gayther@uthscsa.edu (S.A.G.)

We discovered five risk regions specific to ovarian cancer subtypes by performing a genome-wide association study involving 26,000 ovarian cancer cases and 105,000 controls. We located 4,008 credible causal variants in ovarian cancer active promoters and enhancers. We also identified susceptibility genes and explored gene-variant interactions.

Dareng et al., 2024, The American Journal of Human Genetics *111*, 1061–1083
June 6, 2024

© 2024 American Society of Human Genetics.

https://doi.org/10.1016/j.ajhg.2024.04.011

CellPress

Integrative multi-omics analyses to identify the genetic and functional mechanisms underlying ovarian cancer risk regions

Eileen O. Dareng,[2,141] Simon G. Coetzee,[1,141] Jonathan P. Tyrer,[4,141] Pei-Chen Peng,[1,141] Will Rosenow,[3,141] Stephanie Chen,[1,5] Brian D. Davis,[1,5] Felipe Segato Dezem,[1] Ji-Heui Seo,[6,7] Robbin Nameki,[8,9] Alberto L. Reyes,[1] Katja K.H. Aben,[10,11] Hoda Anton-Culver,[12] Natalia N. Antonenkova,[13] Gerasimos Aravantinos,[14] Elisa V. Bandera,[15] Laura E. Beane Freeman,[16] Matthias W. Beckmann,[17] Alicia Beeghly-Fadiel,[18] Javier Benitez,[19,20] Marcus Q. Bernardini,[21] Line Bjorge,[22,23] Amanda Black,[24] Natalia V. Bogdanova,[13,25,26] Kelly L. Bolton,[27] James D. Brenton,[28] Agnieszka Budzilowska,[29] Ralf Butzow,[30] Hui Cai,[18] Ian Campbell,[31,32] Rikki Cannioto,[33] Jenny Chang-Claude,[34,35] Stephen J. Chanock,[16] Kexin Chen,[36] Georgia Chenevix-Trench,[37] AOCS Group,[31,38] Yoke-Eng Chiew,[38,39] Linda S. Cook,[40,41] Anna DeFazio,[38,39,42] Joe Dennis,[2] Jennifer A. Doherty,[43] Thilo Dörk,[26] Andreas du Bois,[44,45] Matthias Dürst,[46] Diana M. Eccles,[47] Gabrielle Ene,[21] Peter A. Fasching,[17] James M. Flanagan,[48] Renée T. Fortner,[34] Florentia Fostira,[49] Aleksandra Gentry-Maharaj,[50] Graham G. Giles,[51,52,53] Marc T. Goodman,[54] Jacek Gronwald,[55] Christopher A. Haiman,[56] Niclas Håkansson,[57] Florian Heitz,[44,45,58] Michelle A.T. Hildebrandt,[59] Estrid Høgdall,[60] Claus K. Høgdall,[61] Ruea-Yea Huang,[62] Allan Jensen,[63] Michael E. Jones,[64] Daehee Kang,[65,66] Beth Y. Karlan,[67] Anthony N. Karnezis,[68] Linda E. Kelemen,[69]

(Author list continued on next page)

Summary

To identify credible causal risk variants (CCVs) associated with different histotypes of epithelial ovarian cancer (EOC), we performed genome-wide association analysis for 470,825 genotyped and 10,163,797 imputed SNPs in 25,981 EOC cases and 105,724 controls of European origin. We identified five histotype-specific EOC risk regions (p value $<5 \times 10^{-8}$) and confirmed previously reported associations for 27 risk regions. Conditional analyses identified an additional 11 signals independent of the primary signal at six risk regions (p value $<10^{-5}$). Fine mapping identified 4,008 CCVs in these regions, of which 1,452 CCVs were located in ovarian cancer-related chromatin marks with significant enrichment in active enhancers, active promoters, and active regions for CCVs from each EOC histotype. Transcriptome-wide association and colocalization analyses across histotypes using tissue-specific and cross-tissue datasets identified 86 candidate susceptibility genes in known EOC risk regions and 32 genes in 23 additional genomic regions that may represent novel EOC risk loci (false discovery rate <0.05). Finally, by integrating genome-wide HiChIP interactome analysis with transcriptome-wide association study (TWAS), variant effect predictor, transcription factor ChIP-seq, and motifbreakR data, we identified candidate gene-CCV interactions at each locus. This included risk loci where TWAS identified one or more candidate susceptibility genes (e.g., *HOXD-AS2*, *HOXD8*, and *HOXD3* at 2q31) and other loci where no candidate gene was identified (e.g., *MYC* and *PVT1* at 8q24) by TWAS. In summary, this study describes a functional framework and provides a greater understanding of the biological significance of risk alleles and candidate gene targets at EOC susceptibility loci identified by a genome-wide association study.

Introduction

Epithelial ovarian cancer (EOC) comprises five main histotypes of invasive disease: high-grade serous ovarian cancer (HGSOC), low-grade serous ovarian cancer (LGSOC), mucinous ovarian cancer (MOC), endometrioid (ENOC), and clear-cell carcinoma (CCOC). Previous studies have shown that these histotypes differ in their underlying

[1]Center for Bioinformatics and Functional Genomics and the Cedars Sinai Genomics Core, Cedars-Sinai Medical Center, Los Angeles, CA, USA; [2]Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK; [3]3Department of Public Health Sciences, University of Virginia, Charlottesville, VA, USA; [4]Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK; [5]Applied Genomics, Computation and Translational Core, Cedars-Sinai Medical Center, Los Angeles, CA, USA; [6]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA, USA; [7]The Center for Functional Cancer Epigenetics, Dana-Farber Cancer Institute, Boston, MA, USA; [8]Women's Cancer Program, Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, CA, USA; [9]Division of Gynecologic Oncology, Department of Obstetrics and Gynecology, Cedars-Sinai Medical Center, Los Angeles, CA, USA; [10]Radboud Institute for Health Sciences, Radboud University Medical Center, Nijmegen, the Netherlands; [11]Netherlands Comprehensive Cancer Organisation, Utrecht, the Netherlands; [12]Department of Medicine, Genetic Epidemiology Research Institute, University of California, Irvine, Irvine, CA, USA; [13]N.N. Alexandrov Research Institute of Oncology and Medical Radiology, Minsk, Belarus; [14]Agii Anargiri' Cancer Hospital, Athens, Greece; [15]Cancer Prevention and Control Program, Rutgers Cancer Institute of New Jersey, New Brunswick, NJ, USA; [16]Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes

(Affiliations continued on next page)

© 2024 American Society of Human Genetics.

Catherine J. Kennedy,[38,39,70] Elza K. Khusnutdinova,[71,72] Lambertus A. Kiemeney,[10] Susanne K. Kjaer,[61,63] Jolanta Kupryjanczyk,[29] Marilyne Labrie,[73] Diether Lambrechts,[74,75] Melissa C. Larson,[76] Nhu D. Le,[77] Jenny Lester,[67] Lian Li,[36] Jan Lubiński,[55] Michael Lush,[2] Jeffrey R. Marks,[78] Keitaro Matsuo,[79,80] Taymaa May,[21] John R. McLaughlin,[81] Iain A. McNeish,[82,83] Usha Menon,[50] Stacey Missmer,[84,85,86] Francesmary Modugno,[87,88] Melissa Moffitt,[89] Alvaro N. Monteiro,[90] Kirsten B. Moysich,[91] Steven A. Narod,[92] Tu Nguyen-Dumont,[53,93] Kunle Odunsi,[94,95] Håkan Olsson,[96] N. Charlotte Onland-Moret,[97] Sue K. Park,[65,66,98] Tanja Pejovic,[99,100] Jennifer B. Permuth,[90] Anna Piskorz,[28] Darya Prokofyeva,[72] Marjorie J. Riggan,[101] Harvey A. Risch,[102] Cristina Rodríguez-Antona,[20,103] Mary Anne Rossing,[104,105] Dale P. Sandler,[106] V. Wendy Setiawan,[56] Kang Shan,[107] Honglin Song,[108] Melissa C. Southey,[51,53,93] Helen Steed,[109,110] Rebecca Sutphen,[111] Anthony J. Swerdlow,[64,112] Soo Hwang Teo,[113,114] Kathryn L. Terry,[86,115] Pamela J. Thompson,[116] Liv Cecilie Vestrheim Thomsen,[22,23] Linda Titus,[117] Britton Trabert,[24] Ruth Travis,[118] Shelley S. Tworoger,[90] Ellen Valen,[22,23] Els Van Nieuwenhuysen,[119] Digna Velez Edwards,[120] Robert A. Vierkant,[76] Penelope M. Webb,[121] OPAL Study Group,[121] Clarice R. Weinberg,[122] Rayna Matsuno Weise,[123] Nicolas Wentzensen,[24] Emily White,[105,124] Stacey J. Winham,[125] Alicja Wolk,[57,126] Yin-Ling Woo,[127] Anna H. Wu,[128] Li Yan,[129] Drakoulis Yannoukakos,[49] Nur Zeinomar,[15] Wei Zheng,[18] Argyrios Ziogas,[12] Andrew Berchuck,[101] Ellen L. Goode,[130] David G. Huntsman,[131,132] Celeste L. Pearce,[128,133] Susan J. Ramus,[134,135]

(Author list continued on next page)

of Health, Department of Health and Human Services, Bethesda, MD, USA; [17]Department of Gynecology and Obstetrics, Comprehensive Cancer Center Erlangen-EMN, Friedrich-Alexander University Erlangen-Nuremberg, University Hospital Erlangen, Erlangen, Germany; [18]Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, TN, USA; [19]Human Genetics Group, Spanish National Cancer Research Centre (CNIO), Madrid, Spain; [20]Centre for Biomedical Network Research on Rare Diseases (CIBERER), Instituto de Salud Carlos III, Madrid, Spain; [21]Division of Gynecologic Oncology, University Health Network, Princess Margaret Hospital, Toronto, ON, Canada; [22]Department of Obstetrics and Gynecology, Haukeland University Hospital, Bergen, Norway; [23]Centre for Cancer Biomarkers CCBIO, Department of Clinical Science, University of Bergen, Bergen, Norway; [24]Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA; [25]Department of Radiation Oncology, Hannover Medical School, Hannover, Germany; [26]Gynaecology Research Unit, Hannover Medical School, Hannover, Germany; [27]Division of Biology and Biomedical Sciences, Washington University, St. Louis, MO, USA; [28]Cancer Research UK Cambridge Institute, University of Cambridge, Cambridge, UK; [29]Department of Pathology and Laboratory Diagnostics, Maria Sklodowska-Curie National Research Institute of Oncology, Warsaw, Poland; [30]Department of Pathology, Helsinki University Hospital, University of Helsinki, Helsinki, Finland; [31]Cancer Genetics Laboratory, Research Division, Peter MacCallum Cancer Center, Melbourne, VIC, Australia; [32]Sir Peter MacCallum Department of Oncology, The University of Melbourne, Melbourne, VIC, Australia; [33]Cancer Pathology & Prevention, Division of Cancer Prevention and Population Sciences, Roswell Park Comprehensive Cancer Center, Buffalo, NY, USA; [34]Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany; [35]Cancer Epidemiology Group, University Cancer Center Hamburg (UCCH), University Medical Center Hamburg-Eppendorf, Hamburg, Germany; [36]Department of Epidemiology, Tianjin Medical University Cancer Institute and Hospital, Tianjin, China; [37]Department of Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia; [38]Centre for Cancer Research, The Westmead Institute for Medical Research, Sydney, NSW, Australia; [39]Department of Gynaecological Oncology, Westmead Hospital, Sydney, NSW, Australia; [40]Epidemiology, School of Public Health, University of Colorado, Aurora, CO, USA; [41]Community Health Sciences, University of Calgary, Calgary, AB, Canada; [42]The Daffodil Centre, a joint venture with Cancer Council NSW, The University of Sydney, Sydney, NSW, Australia; [43]Huntsman Cancer Institute, Department of Population Health Sciences, University of Utah, Salt Lake City, UT, USA; [44]Department of Gynecology and Gynecological Oncology; HSK, Dr. Horst-Schmidt Klinik, Wiesbaden, Wiesbaden, Germany; [45]Department of Gynecology and Gynecologic Oncology, Evangelische Kliniken Essen-Mitte (KEM), Essen, Germany; [46]Department of Gynaecology, Jena University Hospital - Friedrich Schiller University, Jena, Germany; [47]Faculty of Medicine, University of Southampton, Southampton, UK; [48]Division of Cancer and Ovarian Cancer Action Research Centre, Department of Surgery and Cancer, Imperial College London, London, UK; [49]Molecular Diagnostics Laboratory, INRASTES, National Centre for Scientific Research 'Demokritos', Athens, Greece; [50]MRC Clinical Trials Unit, Institute of Clinical Trials & Methodology, University College London, London, UK; [51]Cancer Epidemiology Division, Cancer Council Victoria, Melbourne, VIC, Australia; [52]Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, VIC, Australia; [53]Precision Medicine, School of Clinical Sciences at Monash Health, Monash University, Clayton, VIC, Australia; [54]Cancer Prevention and Control Program, Cedars-Sinai Cancer, Cedars-Sinai Medical Center, Los Angeles, CA, USA; [55]Department of Genetics and Pathology, International Hereditary Cancer Center, Pomeranian Medical University, Szczecin, Poland; [56]Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA; [57]Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden; [58]Center for Pathology, Evangelische Kliniken Essen-Mitte, Essen, Germany; [59]Department of Epidemiology, University of Texas MD Anderson Cancer Center, Houston, TX, USA; [60]Department of Pathology, Herlev Hospital, University of Copenhagen, Copenhagen, Denmark; [61]Department of Gynaecology, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark; [62]Center For Immunotherapy, Roswell Park Comprehensive Cancer Center, Buffalo, NY, USA; [63]Department of Virus, Lifestyle and Genes, Danish Cancer Society Research Center, Copenhagen, Denmark; [64]Division of Genetics and Epidemiology, The Institute of Cancer Research, London, UK; [65]Department of Preventive Medicine, Seoul National University College of Medicine, Seoul, Korea; [66]Cancer Research Institute, Seoul National University, Seoul, Korea; [67]David Geffen School of Medicine, Department of Obstetrics and Gynecology, University of California at Los Angeles, Los Angeles, CA, USA; [68]Department of Pathology and Laboratory Medicine, UC Davis Medical Center, Sacramento, CA, USA; [69]Hollings Cancer Center, Medical University of South Carolina, Charleston, SC, USA; [70]The University of Sydney, Sydney, NSW, Australia; [71]Institute of Biochemistry and Genetics of the Ufa Federal Research Centre of the Russian Academy of Sciences, Ufa, Russia; [72]Department of Genetics and Fundamental Medicine, Bashkir State University, Ufa, Russia; [73]Department of Immunology and Cell Biology, FMSS - Université de Sherbrooke, Sherbrooke, QC, Canada; [74]Laboratory for Translational Genetics, Department of Human Genetics, KU Leuven, Leuven, Belgium; [75]VIB Center for Cancer Biology, VIB, Leuven, Belgium; [76]Department of Quantitative Health Sciences, Division of Clinical Trials and Biostatistics, Mayo Clinic, Rochester, MN, USA; [77]Cancer Control Research, BC Cancer, Vancouver, BC, Canada; [78]Department of Surgery, Duke University Hospital, Durham, NC, USA; [79]Division of Cancer Epidemiology and Prevention, Aichi Cancer Center Research Institute, Nagoya, Japan; [80]Division of Cancer Epidemiology, Nagoya University Graduate School of Medicine, Nagoya, Japan; [81]Public Health Ontario, Samuel Lunenfeld Research Institute, Toronto, ON, Canada; [82]Division of Cancer and Ovarian Cancer Action

(Affiliations continued on next page)

Thomas A. Sellers,[136] The Ovarian Cancer Association Consortium (OCAC), Matthew L. Freedman,[6,7] Kate Lawrenson,[8,9] Joellen M. Schildkraut,[137] Dennis Hazelett,[138,142] Jasmine T. Plummer,[1,5,142] Siddhartha Kar,[139,140,142] Michelle R. Jones,[1,142] Paul D.P. Pharoah,[2,4,142,*] and Simon A. Gayther[1,142,*]

genetic risk factors, their precursor lesions, patterns of spread, and molecular events during oncogenesis. For example, the risks associated with pathogenic germline mutations in DNA double-strand break repair genes (e.g., *BRCA1*, *BRCA2*, *BRIP1*, *RAD51C*, and *RAD51D*) are higher for HGSOC than for other subtypes[1–3] with HGSOCs also likely deriving from fallopian tube secretory epithelial cells.[4] In contrast, CCOC and ENOC are more likely to develop from endometriosis[5] and are driven by genes that function in chromatin remodeling (e.g., *ARID1A*), DNA mismatch repair (e.g., *MSH2* and *MSH6*),[1] and PI3-kinase signaling pathways.[6,7]

Women with a single first-degree relative affected with ovarian cancer have a 3-fold increased risk of developing EOC compared to the general population.[8] About 40% of the excess familial risk of disease is due to rare variants in high to moderately penetrant genes such as *BRCA1* and *BRCA2*.[9] Multiple common, low penetrance susceptibility alleles for ovarian cancer have also been identified using genome-wide association studies (GWASs) (reviewed in Kar et al.[10]). All but one of these common variants are associated with per allele relative risks of less than 1.6, and taken together, they explain ~6.4% of the inherited component of EOC risk in the population. Thus, the known risk variants explain less than 50% of the excess familial risks, suggesting that additional risk alleles likely exist.[11]

The functional mechanisms of common risk variants for complex traits remains largely unknown. Most common low penetrance risk variants for ovarian cancer lie in non-protein-coding DNA and are more likely to lie in regulatory elements (REs), including enhancers, promoters, and insulators,[12–14] suggesting that many common risk

Research Centre, Department Surgery & Cancer, Imperial College London, London, UK; [83]Institute of Cancer Sciences, University of Glasgow, Glasgow, UK; [84]Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA; [85]Obstetrics and Gynecology Epidemiology Center, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA; [86]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA; [87]Women's Cancer Research Center, Magee-Womens Research Institute and Hillman Cancer Center, Pittsburgh, PA, USA; [88]Department of Obstetrics, Gynecology and Reproductive Sciences, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA; [89]Department of Gynecologic Oncology, Roswell Park Comprehensive Cancer Center, Buffalo, NY, USA; [90]Department of Cancer Epidemiology, Moffitt Cancer Center, Tampa, FL, USA; [91]Department of Cancer Prevention and Control, Roswell Park Cancer Institute, Buffalo, NY, USA; [92]Women's College Research Institute, University of Toronto, Toronto, ON, Canada; [93]Department of Clinical Pathology, The University of Melbourne, Melbourne, VIC, Australia; [94]University of Chicago Medicine Comprehensive Cancer Center, Chicago, IL, USA; [95]Department of Obstetrics and Gynecology, University of Chicago, Chicago, IL, USA; [96]Oncology, Department of Clinical Sciences, Lund University, Lund, Sweden; [97]Julius Center for Health Sciences and Primary Care, University Utrecht, UMC Utrecht, Utrecht, the Netherlands; [98]Integrated Major in Innovative Medical Science, Seoul National University College of Medicine, Seoul, South Korea; [99]Department of Obstetrics and Gynecology, Oregon Health & Science University, Portland, OR, USA; [100]Knight Cancer Institute, Oregon Health & Science University, Portland, OR, USA; [101]Department of Gynecologic Oncology, Duke University Hospital, Durham, NC, USA; [102]Chronic Disease Epidemiology, Yale School of Public Health, New Haven, CT, USA; [103]Hereditary Endocrine Cancer Group, Spanish National Cancer Research Center (CNIO), Madrid, Spain; [104]Program in Epidemiology, Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA; [105]Department of Epidemiology, University of Washington, Seattle, WA, USA; [106]Epidemiology Branch, National Institute of Environmental Health Sciences, NIH, Research Triangle Park, NC, USA; [107]Department of Obstetrics and Gynaecology, Hebei Medical University, Fourth Hospital, Shijiazhuang, China; [108]Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK; [109]Division of Gynecologic Oncology, Department of Obstetrics and Gynecology, University of Alberta, Edmonton, AB, Canada; [110]Section of Gynecologic Oncology Surgery, Alberta Health Services, North Zone, Edmonton, AB, Canada; [111]Epidemiology Center, College of Medicine, University of South Florida, Tampa, FL, USA; [112]Division of Breast Cancer Research, The Institute of Cancer Research, London, UK; [113]Breast Cancer Research Programme, Cancer Research Malaysia, Subang Jaya, Selangor, Malaysia; [114]Department of Surgery, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia; [115]Obstetrics and Gynecology Epidemiology Center, Department of Obstetrics and Gyneoclogy, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA; [116]Samuel Oschin Comprehensive Cancer Institute, Cancer Prevention and Genetics Program, Cedars-Sinai Medical Center, Los Angeles, CA, USA; [117]Muskie School of Public Service, University of Southern Maine, Portland, ME, USA; [118]Cancer Epidemiology Unit, University of Oxford, Oxford, UK; [119]Division of Gynecologic Oncology, Department of Gynecology and Obstetrics, Leuven Cancer Institute, Leuven, Belgium; [120]Division of Quantitative Sciences, Department of Obstetrics and Gynecology, Department of Biomedical Sciences, Women's Health Research, Vanderbilt University Medical Center, Nashville, TN, USA; [121]Population Health Department, QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia; [122]Biostatistics and Computational Biology Branch, National Institute of Environmental Health Sciences, NIH, Research Triangle Park, NC, USA; [123]Cancer Epidemiology Program, University of Hawaii Cancer Center, Honolulu, HI, USA; [124]Fred Hutchinson Cancer Research Center, Seattle, WA, USA; [125]Department of Quantitative Health Sciences, Division of Computational Biology, Mayo Clinic, Rochester, MN, USA; [126]Department of Surgical Sciences, Uppsala University, Uppsala, Sweden; [127]Department of Obstetrics and Gynaecology, University of Malaya Medical Centre, University of Malaya, Kuala Lumpur, Malaysia; [128]Department of Preventive Medicine, Keck School of Medicine, University of Southern California Norris Comprehensive Cancer Center, Los Angeles, CA, USA; [129]Department of Molecular Biology, Hebei Medical University, Fourth Hospital, Shijiazhuang, China; [130]Department of Quantitative Health Sciences, Division of Epidemiology, Mayo Clinic, Rochester, MN, USA; [131]Department of Obstetrics and Gynecology, University of British Columbia, Vancouver, BC, Canada; [132]Department of Molecular Oncology, BC Cancer Research Centre, Vancouver, BC, Canada; [133]Department of Epidemiology, University of Michigan School of Public Health, Ann Arbor, MI, USA; [134]School of Women's and Children's Health, Faculty of Medicine and Health, University of NSW Sydney, Sydney, NSW, Australia; [135]Adult Cancer Program, Lowy Cancer Research Centre, University of NSW Sydney, Sydney, NSW, Australia; [136]TAS Consulting, LLC., Portland, OR, USA; [137]Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA, USA; [138]Samuel Oschin Comprehensive Cancer Institute, The Center for Bioinformatics and Functional Biology, Cedars-Sinai Medical Center, Los Angeles, CA, USA; [139]MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK; [140]Section of Translational Epidemiology, Division of Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK

[141]These authors contributed equally

[142]These authors contributed equally

*Correspondence: pp10001@medschl.cam.ac.uk (P.D.P.P.), gayther@uthscsa.edu (S.A.G.)

https://doi.org/10.1016/j.ajhg.2024.04.011.

variants function by regulating the expression of genes involved in disease pathogenesis. Approaches to understand the functional consequences of risk variants usually focus on (1) identifying the causal variants and the non-coding REs they lie in, (2) identifying the susceptibility gene targets of Res, and (3) identifying how the causal variants, Res, and susceptibility gene targets interact with each other to drive disease biology. Thus, understanding the underlying mechanisms associated with common non-coding risk alleles requires the generation and integration of genetic, transcriptomic, and epigenomic datasets.

It is estimated that there are about one million enhancers in the human genome[15–17] any of which may be cell-type and/or disease specific. It is essential that risk variants are evaluated in the context of regulatory landscapes in tissues relevant to the disease under study.[18,19] Statistical approaches have also been developed to identify putative susceptibility genes in GWAS risk regions by investigating the associations between risk variants and the expression of nearby genes (*cis*-expression quantitative trait locus [*cis*-eQTL] analysis); these studies have identified several candidate susceptibility genes at ovarian cancer risk loci EOC.[20,21] Finally, chromosome conformation capture assays have identified evidence of looping that links candidate genes with candidate REs and/or causal risk alleles at several EOC risk loci.[22–25]

Here, we report on a genetic association analysis in 25,981 EOC cases and 105,724 controls genotyped for >470,000 variants, including a genome-wide backbone.[26] The goals were (1) to identify novel variants associated with the different histotypes of EOC, (2) to fine map EOC risk regions to identify credible causal variants (CCVs) that mediate the function of risk alleles, (3) to estimate the contribution of common genetic variation to the heritability of EOC histotypes in European populations, (4) to identify the regulatory and transcriptional targets of EOC risk alleles through the integration of genetic data with epigenomic and whole-transcriptome datasets generated in ovarian- and non-ovarian-cancer-associated tissues, and (5) to validate credible causal SNP-regulatory-transcriptome interactions using whole-genome interactome analyses.

Methods

Study samples

Genotype data from six Ovarian Cancer Association Consortium (OCAC) and two BCAC genotyping projects were used for these analyses. All participating studies were approved by the relevant local institutional research ethics committee, and all participants provided written informed consent. Details of participating studies are provided in Table S1. Some studies (for example, SEARCH) contributed samples to more than one genotyping project, and some case control sets are a combination of multiple individual studies. A total of 25,981 EOC cases, 40,138 OCAC controls, and 65,586 BCAC controls were of European ancestry and passed quality control. Of the 25,981 EOC cases, histotype distribution was 13,609 (HGSOC), 2,749 (LGSOC), 2,877 (ENOC), 1,427 (CCOC), 2,587 (MOC), and 753 (other EOC, excluding those of low malignant potential). For 1,979 EOC cases, histology and/or grade were unascertained. For these cases, we inferred histotypes by performing polygenic risk score (PRS) modeling. We created three separate PRS models: one for HGSOC, one for LGSOC, and one for MOC. Each PRS model was the sum of risk alleles specifically associated at nominal genome-wide significance with the index histotype but not associated with other histotypes ($p > 0.05$), weighted by the effect size estimates from the current GWAS analysis. There were 16, 5, and 5 variants in the HGSOC, LGSOC, and MOC PRS models, respectively. Association results (Table S2) from the PRS models showed that the unassigned cases were more similar to HGSOC than any of the other histotypes. Therefore, the unassigned cases were subsequently classified as part of the HGSOC histotype. All histotypes, excluding the mucinous histotype, were associated to some degree with either the HGSOC PRS or the LGSOC PRS; therefore, we considered non-mucinous as a separate histotype of interest during association analysis.

Histotype definitions

EOC is a highly heterogeneous phenotype with five main histotypes for invasive disease: HGSOC, LGSOC, MOC, ENOC, and CCOC and two histotypes for borderline disease: serous and mucinous. We performed association analysis on the five main histotypes for invasive disease and a non-mucinous histotype, which is a combination of the main invasive histotypes, excluding MOC.

Genotype data and quality control

SNP genotyping

Genotyping data from the BCAC and OCAC projects have been previously published.[11,27–29] Genotyping was performed at five centers: University of Cambridge, Center for Inherited Disease Research, National Cancer Institute, Genome Quebec, and Mayo Clinic. Raw intensity data files were sent to the data coordination center at the University of Cambridge for centralized genotype calling and quality control. Details of genotype calling are as described in Pharoah et al.[27]

Sample quality control

Samples were excluded if they did not harbor two X chromosomes as determined by lack of heterozygosity on the X chromosome or presence of a Y chromosome. Duplicates and close relatives were identified using an in-house software that calculates a concordance matrix for all individuals. Samples with a concordance >0.86 were flagged as duplicates, and samples with a concordance between 0.74 and 0.86 were flagged as relatives. One sample per duplicate or related pair was retained for this analysis. For case-control pairs, we excluded the control, while for case-case and control-control pairs, we excluded the sample with the lower call rate. Concordance statistics were used to identify cryptic duplicates and expected duplicates whose genotypes did not match. We attempted to resolve these discrepancies with the study investigators. If the discrepancy could not be resolved, both samples were excluded. Samples with genotyping call rate <95% and excessively low or high heterozygosity were excluded.

SNP quality control

OncoArray quality control (QC) guidelines[11] were used to filter SNPs. Only SNPs that passed quality control for all consortia

were used for imputation. We excluded SNPs with a call rate <95%, SNPs deviating from Hardy-Weinberg equilibrium ($p < 1 \times 10^{-7}$ in controls and $p << 1 \times 10^{-12}$ in cases), and SNPs with a concordance <98% among 5,280 duplicate pairs. For imputation, we additionally excluded SNPs with minor allele frequency (MAF) <1% and call rate <98%. Of the 533,631 SNPs manufactured on the array, 470,825 passed quality control and were used for imputation.

Imputation

Imputation was performed using the Michigan Imputation Server,[30] which provides a service for large scale population studies. We used SHAPEIT[31] for pre-phasing and Minimac3[30] for imputation with the first release of the Haplotype Reference Consortium panel.[32] We imputed in batches of 10,000 individuals and analyzed imputation quality statistics together. SNPs with imputation r^2 accuracy <0.3 were excluded. The final dataset after imputation included 10,163,797 variants that passed imputation for each genotyping project sample set.

Principal component analysis

Intercontinental ancestry was calculated using the software package FastPop (http://sourceforge.net/projects/fastpop/) developed for the OncoArray. Only samples with >80% European ancestry were used in these analyses. Principal component analyses for the OncoArray data were carried out using data from 33,661 uncorrelated SNPs (pairwise $r^2 < 0.1$) with MAF >0.05 using an in-house program (available at http://ccge.medschl.cam.ac.uk/software/pccalc). Principal component analysis for the other genotype datasets was carried out as previously described.[27]

Association analysis

We evaluated the association between genotype and disease using logistic regression models by estimating the associations with each additional copy of the minor allele (log-additive models). We carried out initial genome-wide analyses separately for OncoArray, COGS, and the five GWAS datasets and pooled the results using a fixed effects meta-analysis. The analyses were adjusted for country and population substructure by including the eigenvectors of project-specific principal components as covariates in the model (nine for OncoArray; five for COGS; two for UK GWAS; two for the US, BWH, and POL GWAS; and a single principal component for the MAY GWAS). The number of eigenvectors chosen was based on the point of inflection of a scree plot. After imputation of the genotypes, we used genotype dosages in a single logistic regression model with adjustment for each genotyping project/study combination and nineteen principal components. Principal components were set to zero for samples not included in a given project.

Fine mapping

We performed fine mapping in new regions (those not previously reported as detailed in Table S3) with variants associated at $p < 5 \times 10^{-8}$ as well as 26 other variants identified from previous analysis, which replicated in this study (Table S4). Each region was explored to identify additional independent association signals and investigate the genomic localization of causal variants. SNPs within a given genomic region were jointly analyzed to evaluate the simultaneous effects of multiple SNPs, using a 1Mb window centered on the most significant variant, in stepwise conditional analysis. In these models, all SNPs within a region, including SNPs in high LD with the lead SNP, were included. Given the presence of a genome-wide significant variant in the region, the prior probability of an additional variant in the same region is higher than in a region without a genome-wide significant lead SNP, therefore we used a threshold of $p < 10^{-5}$ to identify additional independent signals.

Credible causal variant sets

Putative causal SNPs were selected to create a CCV set, the minimum set of SNPs that contain all causal SNPs in the region, by triaging variants based on p values from the association analysis (for primary association signals) and the conditional analysis (for secondary signals). All SNPs in each region were ranked by the likelihood of association with EOC, including histotype specific analysis. The likelihood of each SNP was then compared with the likelihood of the lead SNP in the region. SNPs with causality odds >1:100 compared with the lead SNP were selected. CCV sets were created for all published hits that replicated in this analysis, primary association signals, and secondary association signals (Table S5) from this analysis. As secondary signals were by definition identified in regions with genome-wide significant SNPs, there was some overlap in the CCV sets identified for the secondary signals with the CCV set identified for the primary signals.

Functional annotation and enrichment analyses

Epigenomic datasets for ovarian cancer and their precursor tissues

A collection of epigenomic datasets in 18 ovarian cancer cell lines and precursor cell types were compiled (Table S6). These datasets include previously published and newly generated chromatin immunoprecipitation sequencing (ChIP-seq) for four histone marks (H3K27ac, H3K4me1, H3K4me3, and CTCF), and RNA sequencing (RNA-seq) for gene expression. There are 12 established EOC cell lines that model undifferentiated EOC (HEYA8), HGSOC (CaOV3, UWB1.289, and Kuramochi), LGSOC (VOA1056 and OAW42), CCOC (JHOC5, ES2, and RMG-II), and MOC (GFTR230, MCAS, and EFO27) and six cell lines for three ovarian cancer precursor cell types, comprising of fallopian tube secretory epithelial cells (FTSECs; FT246, FT33, and FT282), ovarian surface epithelial cells (OSECs; IOSE4 and IOSE11), and endometriosis epithelial cells (EECs; EEC16).[33]

RNA-seq data were generated as previously described,[34] and are publicly available at GEO: GSE114332. ChIP-seq for H3K27ac, H3K4me3, and H3K4me1 was performed as previously described.[19] Peak calling was performed using the AQUAS pipeline.[35] Reads were aligned against the reference human genome (GRCh38). Quality control metrics were computed for each individual replicate, including number of reads, percentage of duplicated reads, normalized strand coefficient, relative strand correlation, and fraction of reads in called peaks. Peak calling was performed with macs2 with pooled replicate peaks that overlap 50% or more of each individual replicate selected for the final peak set. When replicates were not available (HEYA8, ES2, and RMG-II) pseudo replicates were formed and pooled peaks selected in the same manner from these pseudo replicates.

Generating chromatin state annotations for ovarian cancer cell lines

We used ChromHMM v1.19[36] to identify chromatin states in 18 EOC-associated cell lines profiled for the histone marks H3K27ac, H3Kme1, H3Kme, CTCF, and RNA-seq (Table S6). The input for each cell line was RNA-seq and ChIP-seq peaks for four histone marks—H3K27ac, H3K4me1, H3K4me3, and CTCF.

Signals were binarized at 200 bp resolution. For chromatin mark ChIP-seq files, we used the BinarizedBed function to convert coordinates with default settings. For RNA-seq files, gene expression was calculated at transcript level and extended to include untranslated region (UTR) and exon regions as intervals for ChromHMM input. Transcripts with expression counts greater than 15 were labeled as 1 (expressed), otherwise 0, as recommended by the edgeR package[37] default filterByExpression value. RNA-seq for FT282 cell line was excluded and treated as "missing" in ChromHMM due to unmatched read length and read depth to other RNA-seq files. Gene coordinates from GENCODE version 34lift37[38] in GRCh37 Human Genome reference were used.

We trained six ChromHMM models, ranging from five to ten total chromatin states, based on our prior experience in chromatin states model training.[19] We selected an eight-state model, as it best captured the known classes of genomic elements such as promoters and transcribed regions,[39] and the larger number of states did show more distinctive interactions. The eight states were labeled as active promoter, active enhancer, active region, weak promoter, weak enhancer, insulator, transcribed region, and low signal based on the emission probability and transmission probability matrix learned from ChromHMM. Active regions represent intervals with H3K27ac mark but absence of H3K4me1 and H3K4me3 marks. We note that low-signal regions were not analyzed in this study because there were no repressive histone marks in our inputs, thereby there was no way to differentiate between low-signal and poised regions.

To create consensus peak sets across a single histotype for enrichment analyses, we retained peaks with at least 50% overlap with at least one other peak in two or more samples from the same histotype group and then stretched boundaries to the edge of each peak in the overlap. Chromatin state peak files were then concatenated, and peak coordinates were merged such that if records within the concatenated file were overlapping, they were further combined into a single consensus peak. In total, we compiled seven sets of consensus peaks: three sets for ovarian cancer precursor cell types (FTSECs, OSECs, and endometrioid epithelial cells), and four sets for EOC histotypes (HGSOC, LGSOC, CCOC, and MOC).

Estimating SNP heritability and partitioned SNP heritability

We estimated the variance explained by known SNP effects, or SNP heritability. We first calculated the total SNP heritability of each EOC histotype by using linkage disequilibrium score regression (LDSC),[40,41] version 1.0.0. Briefly, this method uses GWAS summary statistics to estimate SNP heritability and explicitly models the effect of all other SNPs in the same LD region. It applies a regression model to examine the relationship between the GWAS test statistics from a given SNP and "LD scores." The LD score is measured by the amount of genetic variation of all other SNPs tagged by the specific variant under consideration. In this study, the LD scores were pre-calculated from phased European-ancestry individuals from the 1000 Genomes Project reference panel v3 (downloaded alongside the software from https://data.broadinstitute.org/alkesgroup/LDSCORE/).

We applied stratified LD score regression[41] to examine the polygenic contribution of each functional annotation to SNP heritability. The goal was to estimate the proportion of genome-wide SNP heritability attributable to each functional annotation by combining SNPs overlapping with regions of interest and pruning out the effect of LD. The category-specific enrichment was defined as the proportion of SNP heritability in the category divided by the proportion of the number of SNPs in the same category. Instinctively, a functional category that has a smaller number of SNP overlaps but a larger contribution to SNP heritability will show stronger enrichment. According to the total heritability estimates, only non-MOC (NMOC) and HGSOC GWAS summary statistics have sufficient power to quantify heritability for downstream enrichment analyses. We therefore proceeded with the partitioned-heritability enrichment analyses with these two histotypes.

The partitioned-heritability analyses were performed with two different sets of functional annotations. The first is a full baseline model with 24 general broad functional annotations curated by Finucane et al.,[41] which were inclusive of all publicly available cell types and were post-processed.[42] The 24 annotations used in this analysis include coding, 3′UTR, 5′UTR, promoter, and intron regions from UCSC Genome Browser[42,43]; regions conserved in mammals[44,45]; combined chromHMM and Segway predictions comprising CTCF-bound regions, promoter-flanking, transcribed, transcription start site, strong enhancer, weak enhancer, and repressed annotations[46]; digital genomic footprint and transcription factor binding sites (TFBSs) from ENCODE[42]; open chromatin regions as reflected by DNase I hypersensitivity sites from a union of all cell types and a union of only fetal cell types on ENCODE and Roadmap Epigenomics[47]; FANTOM5 enhancer[15]; H3K27ac,[48] H3K4me1, and H3K4me3 histone marks from a union across cell types on Roadmap Epigenomics[16]; and super-enhancers.[48] The second set of functional annotations include 49 ovarian cancer-related chromatin states annotated by a trained ChromHMM model, as described above.

Prioritization of the likely causal variant using PAINTOR

PAINTOR is a Bayesian approach that combines information from GWAS summary statistics, LD structure, and functional annotations to compute the posterior probability of causality for all SNPs at risk loci.[49] Two PAINTOR models were trained: one on NMOC and one on HGSOC risk loci. Five significant annotations each were chosen for NMOC and HGSOC. The annotations for NMOC included 3′ UTR, TFBS from ENCODE, consensus active promoters in HGSOC, consensus active enhancers in FT, and consensus active regions in EEC. For HGSOC, we chose 3′ UTR, TFBS from ENCODE, consensus active promoters in CCOC, consensus active enhancers in FT, and consensus transcribed regions in FT. As recommended by PAINTOR, chosen annotations were significantly enriched with CCVs and approximately uncorrelated to each other in order to maximize the information content.

Enrichment of credible causal variants in functional annotations

Previously reported and newly identified EOC CCVs were combined to create the full candidate set ($n = 4{,}008$) and then stratified by EOC histotype (NMOC = 246, HGSOC = 3,152, LGSOC = 386, ENOC = 50, and MOC = 174). HGSOC CCVs at 17q21.31 were excluded for this analysis due to its disproportionately large number of CCVs ($n = 2{,}256$). Each set of CCVs underwent enrichment analysis against each of the 49 ovarian cancer-related chromatin states. Enrichment analysis was performed with the FunciVar package (https://github.com/Simon-Coetzee/funcivar[19]) a tool for annotation and functional enrichment of variant sets. In brief, FunciVar first takes two lists of variants as inputs: (1) a list of CCVs as the foreground and (2) a list of control variants as the background. The background list of variants was generated by aggregating SNPs within 2 Mb (1 Mb +/−) of the credible causal set in order to maintain similar genetic architecture (e.g., linkage disequilibrium, regulatory activity, and transcriptional program) as CCVs. FunciVar then intersects each variant with functional annotations. Enrichment probability is modeled under a beta-binomial distribution and reported as the median of enrichment

probability differences between foreground and background variants (labeled as "difference"), within the range of −1 to 1, where 1 means strong enrichment and −1 means strong depletion. The significance of results is reported as probability that foreground variants have more overlaps with the functional annotation than background SNPs, within the range of 0–1, the higher being the more confident.

Candidate susceptibility gene identification

Tissue-specific transcriptome-wide association analyses

We conducted a tissue-specific transcriptome-wide association study (TWAS) using methods and resources that have been described previously.[21] Briefly, we used three EOC-relevant, tissue-specific RNA-seq resources: 284 TCGA HGSOCs[50] and 105 ovarian and 60 fallopian tube epithelial precursor tissues[21] and applied the FUSION TWAS pipeline[51] to genome-wide SNP-level association summary statistics for each of the six EOC histotypes. Transcriptome-wide significant associations were determined using a false discovery rate (FDR) threshold of 5%. The FUSION pipeline integrates colocalization analysis as implemented in the "coloc" tool.[52] We used coloc to further prioritize the TWAS results and only focused on transcriptome-wide significant genes with coloc posterior probability ("PP4") >0.6, which indicates that there is a >60% probability of a shared SNP signal underlying the association in the GWAS for EOC risk and in the gene-expression datasets.

Cross-tissue transcriptome-wide association analyses

We performed transcriptome-wide association analyses by integrating genome-wide SNP-level association summary statistics for each of the six EOC histotypes with GTEx version 8[53] gene-expression quantitative trait locus data from 49 tissue types. Summary association statistics for each histotype were used as inputs to S-MultiXcan[54] in order to identify associations with ovarian cancer risk mediated via gene expression. This method taps the joint effects of gene expression correlation across tissues to boost the power to detect candidate disease susceptibility genes. First, S-PrediXcan[55] was run for each histotype and for each of the 49 tissues in GTEx version 8. The results were then fed into S-MultiXcan, which outputs a list of p values for the association between each gene and each ovarian cancer histotype. Transcriptome-wide significant associations were determined using an FDR threshold of 5%.

We used colocalization implemented in fastENLOC[56,57] in order to refine further the results of S-MultiXcan for each histotype. Summary statistics were annotated with posterior inclusion probabilities and were used as inputs to fastENLOC for each of the 49 GTEx tissues. This provided regional colocalization probabilities (RCPs) for each gene across tissues. These RCPs were merged with the S-MultiXcan results to identify genes with an FDR <0.05 and RCP ≥ 0.1 in any tissue. This fastENLOC RCP threshold indicates that the same SNP signal is likely to underlie the GWAS and gene expression associations.

Interactome analysis of EOC risk regions

H3K27ac HiChIP sequencing

HiChIP utilized four different cell lines, two precursor fallopian tube lines (FT33 and FT246), and two HGSOC cell lines (Kuramochi and UWB1.289). All FT lines were cultured in DMEM:F12 (Corning, Cellgro) with 10% fetal bovine serum (FBS). Kuramochi cells were grown in RPMI 1640 (Gibco) with 10% FBS added. UWB1.289 cells were grown in a 1:1 mixture of MEGM (DMEM:F12, 5% horse serum, 20 ng/mL epidermal growth factor, 0.5 mg/mL hydrocortisone, 100 ng/mL cholera toxin, 10 μg/mL insulin) and RPMI 1640. Cells were grown to 80% confluency, detached with trypsin, and strained through a 40-micron filter. The cell pellet was crosslinked with 1% formaldehyde (Sigma) that was freshly made for 10 min at room temperature with rotation and quenched with 125 mM of ice-cold glycine. Cells were pelleted, washed in PBS, flash frozen in liquid nitrogen, and stored at −80°C.

HiChIP libraries were generated utilizing the Mumbach et al.[58] protocol with the following modifications. All samples were processed as technical replicates. Cells were digested with MBoI for 4 h and mixed intermittently. After digestion, overhangs were filled in with DNA polymerase I, large Klenow fragment for 1 h at 37°C. Samples were ligated using a T4 DNA ligase for 4 h at room temperature, nuclei were lysed using Lysis buffer (PBS pH 7.4, 0.1% SDS, 0.137 M NaCl, 0.5% Na-deoxycholic acid, 1% NP-40[59]) with protease inhibitor, and chromatin was sonicated for 5 min on a Covaris E220. Sonicated chromatin was precleared in a mixture of protein A and G Dynabeads (Thermo Fisher A: 10002D, G: 10004D), and the remaining chromatin incubated in 5.6 μug of H3K27ac antibody (Diagenode C15410196) overnight. Chromatin was then purified using a NucleoSpin PCR clean-up kit (Machery-Nagel 740609) and tagmented with Tn5 enzyme (Illumina 15027865) when captured with streptavidin beads and eluted. Libraries were generated by PCR with a Phusion HF 2x Mastermix (BioLabs MO536S) using Nextera custom oligos. A two-sided size selection with Ampure XP beads (Beckman Coulter, A63881) captured fragments between 300 and 700 bp. Libraries were pooled and assessed by a shallow depth 2 × 150 bp paired-end sequencing run on an Illumina Miniseq, generating 1–5 million reads per replicate. Full depth sequencing of the HiChIP pool was performed on the Illumina Novaseq at 2 ×150 bp, generating 300–400 million paired-end reads per sample.

HiChIP data analysis

Sequencing reads from HiChIP libraries were filtered, and the adapters were removed using cutadapt[60] and mapped to the GRCh38 genome via HiC-Pro v2.11.0.[61] Replicates were merged together, and the HiCPro output files were used for QC metrics and subsequent input for FitHiChIP.[62] FitHiChIP was used to call peaks from Hi-C Pro with default settings. This tool uses all reads from dangling ends, re-ligation and self-circle pairs, as well as all reads with Washington Epigenome browser.[63]

We integrated 6 different datasets/annotations—GWAS p value, distance to promoter, GTEx/FTE/OSE eQTLs, HiChIP looping assays, weighted Remap transcription factor (TF) database, and weighted variant effect predictor (VEP) score. For the latter three, we weighted scores according to biological disruptiveness. TF interactions were weighted higher for variants that break the corresponding motif (as predicted by "motifbreakR") and VEP by disruptiveness (e.g., stop mutations vs. missense). In addition, loop scores are weighted by epigenetic context at loop ends (e.g., "enhancer-promoter" interaction vs. "other enhancer"). We used a three-step procedure to accomplish this. In the first step, genes are selected for each hit region as follows: all genes were selected within the larger of the most significant HiChIP interaction in 1 Mb flanking the top variant (2 Mb total) or 500 kb flanking sequence. Second, we assigned a scale-normalized score for each of the aforementioned assays and annotations for each variant-gene combination, which summed together we call the "gene score" (see Table S7). We then sum the gene scores across variants for each region and multiply by the negative log of the p value, producing a "total score" for each gene in the region, which by

now has been weighted by the preponderance of genetic and epigenomic evidence in its favor (summarized in Table S8). These scores can be affected by various factors, including the length of CCV list, so we applied a softmax function so that genes in each region sum to 1 (the "escore") thus producing a relative confidence level that can be compared between and across different loci and histotypes. Finally, in step 3, we organized the top CCV-gene pairs into a table (Table S8) that reports the variance of the escores by region ("region score"), which can be interpreted as related to the confidence in our predictions, i.e., low variance regions probably do not distinguish well between genes vs. high-variance regions in which one or two candidates stand out above the rest.

Functional analysis of PAX8 in ovarian cancer cell lines

We used a small interfering RNA (siRNA) pool to knockdown the expression of PAX8 in the HGSOC line OVCAR4 compared to a scrambled control pool. Count matrices of siRNA scramble control, and PAX8 were downloaded from GEO: GSE150443. Counts were normalized as transcripts per million and analyzed for differential analysis with DESEQ2.[64]

Results

Genetic association and fine mapping analyses of EOC risk regions

Genotype data were available for 25,981 EOC cases and 105,724 controls of European origin (Table S1), including 13,609 HGSOC cases, 2,749 LGSOC cases, 2,877 ENOC cases, 1,427 CCOC cases, and 2,587 MOC cases, as well as 753 cases of other EOC histotypes (excluding those of low malignant potential) confirmed by histology. Histology and/or grade information were not available for 1,979 cases, and so we inferred histotype for these cases using PRS modeling (Table S2). This indicated that the unassigned EOC cases are more similar to HGSOC than any of the other histotypes, and so we classified them as such, increasing the number of HGSOC cases to 15,588, the total number of N. PRS modeling also suggested that the genetic architecture of MOC was substantially different from the other histotypes, and so EOC cases were also classified either as MOC or as NMOC (n = 23,394) (Figure 1A). After quality control analysis, 470,825 SNPs were available for imputation using the Haplotype Reference Consortium reference panel.[65] Thirty-two million SNPs were imputed of which 10,163,797 passed quality control (MAF > 0.01, imputation r^2 > 0.3). These SNPs were evaluated for association with risk of the different histotype groups.

We identified five novel EOC risk associations (genomic regions where no EOC risk association had previously been reported at genome-wide significance (p value $<5 \times 10^{-8}$) (Tables S3 and S4) (Table 1; Figure 1B). Risk-associated SNPs in these regions were common (MAF >0.05), well imputed ($r^2 > 0.90$), and had low probabilities of being false-positive associations (Bayes false discovery probability [BFDP] < 5%), assuming a prior probability of association of 0.0001 and upper-likely odds ratio (OR) under the alternative hypothesis of 1.2. Two regions were associated with HGSOC risk, rs4149419 at 4q13.3 (OR 0.93 [95% confidence interval (CI) 0.90–0.95], p value = 2.7×10^{-8}) and rs7851336 at 9p22.1 (OR 1.10 [95% CI 1.07–1.13], p value = 2.9×10^{-10}); two with risk of NMOC, rs336126 at 5q11.2 (OR 0.92 [95% CI 0.90–0.94], p value = 6.4×10^{-11}) and rs2070368 at 21q22.12 (OR 0.94 [95% CI 0.92–0.96] p value = 1.1×10^{-8}); and one with risk of MOC, rs72827480 at 2q14.2 (OR 0.85 [95% CI 0.80–0.90], p value = 2.7×10^{-8}). We confirmed the associations at genome-wide significance for 19 of 28 previously reported EOC risk regions,[11,23,27,66–71] six (of nine) risk loci identified in meta-analyses of data from OCAC and the Consortium of Investigators of Modifiers of BRCA1/2 (CIMBA),[11,72] and two (of five) risk associations conferring risk of both breast and ovarian cancer[73] (Table 2; Figure 1C; Tables S3 and S4).

We also performed conditional analysis to fine map risk variants at all new and previously reported risk regions, adjusting for the most statistically significant SNP in each risk region. The index SNP in each region was imputed with r^2 >0.90 and estimated allele frequency >0.05, identifying 11 signals at six risk regions with p values $<10^{-5}$ (of which four had a p value $<5.0 \times 10^{-8}$) that were independent of the primary signal (Table 1). Independent risk signals were identified for two or more ovarian cancer histotypes in five regions, notably at 8q24.21 where we replicated the previously reported risk associations for HGSOC (chr8:129541931 [genome build 37])[21] and NMOC (chr8:128817883 [genome build 37])[11] and identified three additional independent risk associations: rs6470494 (chr8:128087904 [genome build 37]) associated with MOC (OR 0.87 [95% CI 0.82–0.92], p value = 5.7×10^{-6}), rs7833298 (chr8:129080657 [genome build 37]) associated with HGSOC (OR 0.92 [95% CI 0.90–0.95], p value = 1.2×10^{-9}), and rs77235147 (chr8:129982731 [genome build 37]) associated with NMOC (OR 0.91 [95% CI 0.87–0.95], p value = 7.2×10^{-6}) (positions shown are in hg19; Figure 2A). Another example was at 2q31.1 where we replicated previously reported associations with MOC (chr2:177037311 [genome build 37])[23] and NMOC (chr2:177039578 [genome build 37]).[68] Subsequent conditional analysis revealed a previously unidentified independent risk association with MOC for rs2594950 (chr2:177508622 [genome build 37]; OR 0.87 [95% CI 0.81–0.92], p value = 6.5×10^{-6}) (Figure 2B). The other regions associated with risks of two or more EOC histotypes were 2q13a (HGSOC and MOC), 5p15.33 (HGSOC, LGSOC, and NMOC), and 17q12 (HGSOC and CCOC) (Tables 1 and 2).

Fine mapping identified 4,008 CCVs at the 32 new or replicated risk regions and the 11 secondary risk regions (Table S5). Of the 4,008 CCVs, 169 (4.2%) were in known gene transcripts and included 60 CCVs in 3′ UTRs, 50 CCVs in 5′ UTRs, 32 coding missense variants, and 39 synonymous variants (Table S9). Of the CCVs in 3′ UTRs, 13 were in *MAPT* at the 17q21.31 risk locus, 5 were in *PAX8*

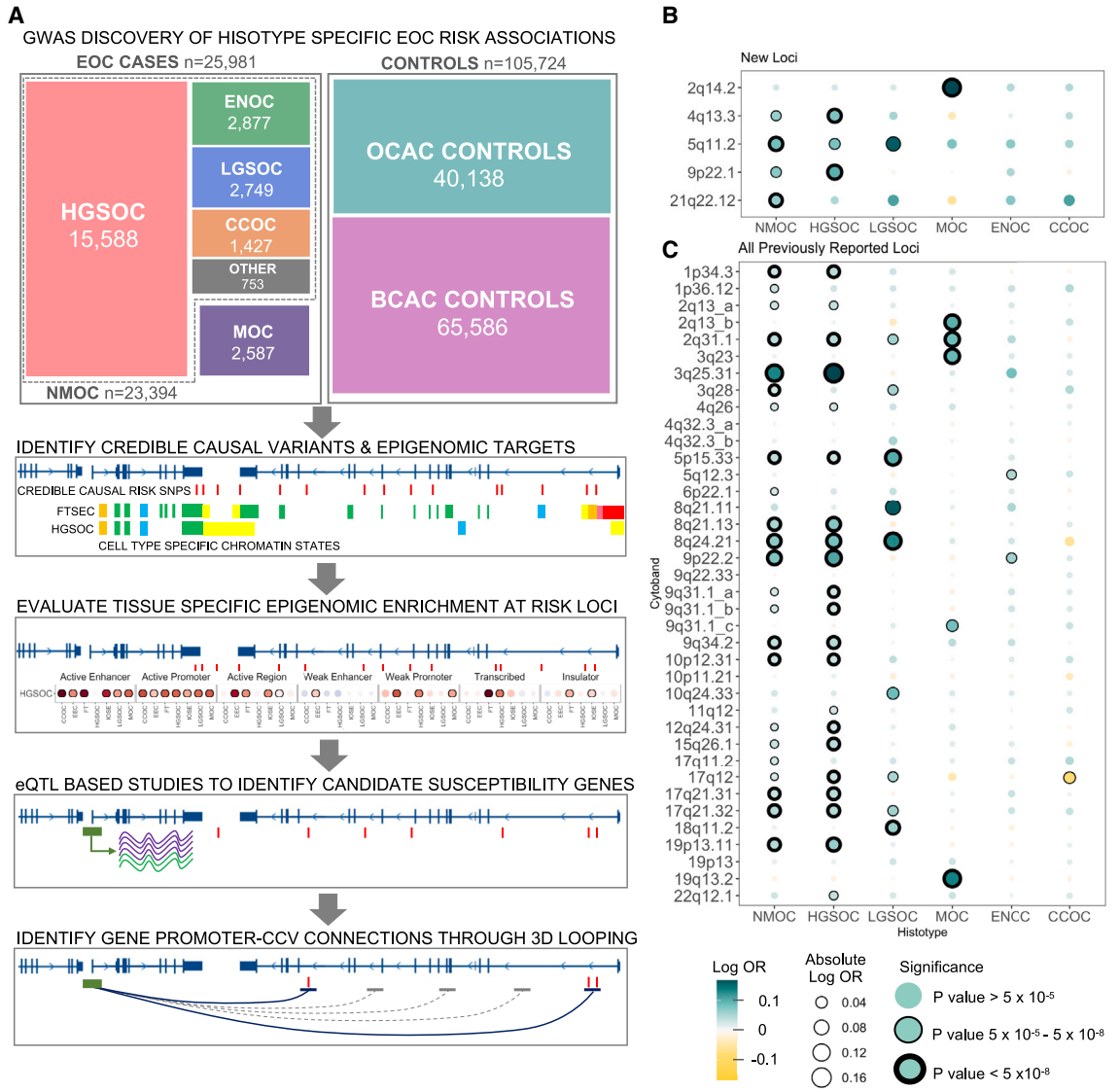

Figure 1. Study design for the identification and functional analysis of common low penetrance risk variants for epithelial ovarian cancer (EOC)

(A) Study design: HGSOC, high-grade serous ovarian cancer; ENOC, endometrioid ovarian cancer; LGSOC, low-grade serous ovarian cancer; CCOC, clear-cell ovarian cancer; MOC, mucinous ovarian cancer; NMOC, all non-mucinous ovarian cancer; fine mapping of risk regions and epigenomic annotation to identify credible causal risk variants (CCVs) at each risk locus; cell-type-specific epigenomic enrichment and partitioning heritability analysis; expression QTL-based approaches to identify candidate susceptibility genes in EOC risk regions; and 3D looping analysis to identify gene-CCV interactions at EOC risk loci.

(B) Genome-wide association analyses identified five novel EOC risk loci at $p < 5 \times 10^{-8}$ for different EOC histotypes; bubble plot illustrates both the effect size and the statistical significance value for each region by histotype.

(C) Risk associations by histotype for previously reported EOC risk regions including 22 regions for which the association signal replicates in one or more histotypes at $p < 5 \times 10^{-8}$ and 11 risk regions that do not replicate at genome-wide significance.

at the 2q13 locus, and 5 were in *ANKLE 1* at the 19p13.11 locus. Of the CCVs in 5′ UTRs, 25 were in *MAPT*, 10 were in *KANSL1* at the 17q21.31 risk locus, and 7 were in *SMC2* at the 9q31.3 locus. Finally, of the 32 missense variants, 7 were predicted to be deleterious by consensus from the Sorting Intolerant From Tolerant (SIFT),[74] Polymorphism Phenotyping (Polyphen),[75] and Combined Annotation Dependent Depletion[76] algorithms, including 3 CCVs in *ANKLE1* and 1 CCV in *CHECK2*, *CHMP4C*, *KANSL1*, and *MAPT*, respectively.

Estimating SNP heritability and partitioning heritability for EOC risk alleles

We estimated the genome-wide contribution of risk variants to narrow-sense heritability (i.e., the proportion of phenotypic variance attributable to the additive effects of common GWAS SNPs) by LDSC using genome-wide GWAS summary statistics for each EOC histotype group. The genome-wide SNP-heritability estimates per histotype were 4.6% for NMOC cases, 4% for HGSOC, 1.6% for MOC, 1.3% for LGSOC, 1% for ENOC, and 0.8% for

Table 1. Histotype specific EOC risk associations identified in this study

Analysis	Locus	Histotype	Index SNP rsid	Index SNP position[a]	Effect allele	Effect allele freq.	r^{2}[b]	OR (95% CI)[c]	*p* value	BFDP[d] (%)	# of CCVs[e]	Nearest gene to lead SNP
Primary	2q14.2	MOC	rs72827480	121146501	C	0.36	0.94	0.85 (0.80–0.90)	2.7×10^{-8}	3	4	*INHBB*
	4q13.3	HGSOC	rs4149419	70592242	T	0.34	0.93	0.93 (0.90–0.95)	2.7×10^{-8}	2	85	*SULT1B1*
	5q11.2	NMOC	rs336126	54476556	A	0.72	0.93	0.92 (0.90–0.94)	6.4×10^{-11}	<1	10	*CDC20B*
	9p22.1	HGSOC	rs7851336	19030883	C	0.25	0.95	1.10 (1.07–1.13)	2.9×10^{-10}	<1	47	*SAXO1*
	21q22.12	NMOC	rs2070368	36080398	C	0.39	0.97	0.94 (0.92–0.96)	1.1×10^{-8}	<1	6	*CLIC6*
Conditional	2q13a	LGSOC	rs1470053	111915946	T	0.16	1.00	0.84 (0.78–0.90)	2.6×10^{-6}	22	6	*BCL2L11*
	2q13b	HGSOC	rs895412	113973964	C	0.49	0.95	1.08 (1.04–1.11)	1.3×10^{-7}	<1	87	*PAX8*
	2q31.1	MOC	rs2594950	177508622	C	0.31	0.91	0.87 (0.81–0.92)	6.5×10^{-6}	28	63	*MTX2*
	5p15.33	HGSOC	rs10069690	1279790	T	0.28	0.99	1.09 (1.05–1.13)	2.1×10^{-6}	9	2	*TERT*
	5p15.33	LGSOC	rs2853677	1287194	A	0.51	0.99	0.80 (0.74–0.86)	7.8×10^{-10}	<1	1	
	5p15.33	NMOC	rs2853669	1295349	G	0.30	0.99	0.92 (0.90–0.95)	4.9×10^{-8}	<1	6	
	8q24.21	MOC	rs6470494	128087904	C	0.69	1.00	0.87 (0.82–0.92)	5.7×10^{-6}	70	4	*PCAT1/CASC1*
	8q24.21	HGSOC	rs7833298	129080657	C	0.38	1.00	0.92 (0.90–0.95)	1.2×10^{-9}	<1	10	*PVT1*
	8q24.21	NMOC	rs77235147	129982731	T	0.07	0.98	0.91 (0.87–0.95)	7.2×10^{-6}	27	47	*CCDC26*
	9p22.1	HGSOC	rs10810671	16914835	C	0.27	1.00	1.08 (1.05–1.11)	6.5×10^{-10}	<1	20	*BNC2*
	19p13.11	HGSOC	rs12982058	17409380	T	0.48	0.99	0.92 (0.89–0.96)	4.1×10^{-6}	15	44	*ABHD8*

Primary analysis: EOC risk loci ($p < 5 \times 10^{-8}$) at genomic regions not previously reported; conditional analysis: independent EOC risk associations ($p < 1 \times 10^{-5}$) identified at previously genomic regions previously reported to be associated with EOC.
[a]Build 37 position.
[b]Average imputation accuracy estimate.
[c]Odds ratio (95% confidence interval).
[d]Bayesian false discovery probability based on a prior probability on the alternative hypothesis of 1:10,000 and an upper likely effect size (odds ratio) of 1.2.
[e]CCVs, credible causal variants.

Table 2. Re-analysis of previously reported genome-wide significant risk loci

Cytoband	Histotype	Lead SNP previous analysis[a]	Lead SNP reanalysis	Effect allele	Effect allele freq.	R^{2}[b]	OR (95% CI)[c]	*p* value	BFDP[d] (%)	# of CCVs[e]	Nearest gene to lead SNP
1p34.3	HGSOC	38,096,421	38,082,122	A	0.254	0.99	1.12 (1.09–1.15)	1.9E-14	<1	13	*RSPO1*
2q13a	MOC	113,972,945	113,979,364	A	0.150	0.92	1.31 (1.21–1.41)	4.5E-12	<1	18	*PAX8-AS1*
2q13b	HGSOC	111,818,658	111,782,834	A	0.159	0.99	1.10 (1.06–1.13)	7.6E-08	4	38	*ACOXL; BCL2L11*
2q31.1	MOC	177,037,311	177,508,622	C	0.332	0.91	0.87 (0.82–0.93)	1.0E-5	86	63	*HOXD3*
	NMOC	177,042,633	177,039,578	T	0.680	0.99	0.90 (0.88–0.92)	4.9e-20	<1	16	*HOXD-AS1*
3q23	MOC	138,849,543	138,849,543	C	0.288	0.86	1.29 (1.21–1.37)	8.1E-15	<1	10	*MRPS22*
3q25.31	HGSOC	156,435,640	156,402,487	T	0.049	1.00	1.59 (1.52–1.68)	8.0E-71	<1	61	–
3q28	NMOC	190,525,516	190,531,882	A	0.303	0.97	0.93 (0.91–0.96)	6.9E-9	1	40	–
5p15.33	LGSOC	1,279,790	1,279,790	T	0.259	0.99	1.22 (1.15–1.29)	2.4E-10	<1	2	*TERT*
	HGSOC		1,279,790	T	0.259	0.99	1.14 (1.11–1.17)	1.0E-19	<1	2	
	LGSOC	1,285,974	1,285,974	A	0.330	0.98	1.32 (1.24–1.39)	1.8E-21	<1	1	
	HGSOC		1,285,974	A	0.330	0.98	1.09 (1.06–1.12)	2.4E-10	<1	1	–
8q21.13	NMOC	82,668,818	82,653,644	G	0.068	1.00	1.15 (1.10–1.19)	2.2E-11	<1	8	*CHMP4C*
8q24.21	NMOC	129,541,931	129,541,931	A	0.130	1.00	0.83 (0.81–0.86)	2.1E-28	<1	49	*LINC00826*
	LGSOC		129,541,931	A	0.130	1.00	0.70 (0.64–0.77)	2.0E-15	<1	49	
	HGSOC		129,541,931	A	0.130	1.00	0.80 (0.77–0.83)	8.2E-30	<1	49	
	NMOC	128,817,883	128,817,883	G	0.457	0.99	1.08 (1.05–1.10)	9,8E-12	<1	18	*PVT1; MYC*
	HGSOC		128,817,883	G	0.457	0.99	1.09 (1.06–1.12)	1.7E-11	<1	18	–
9p22.2	NMOC	16,914,716	16,914,716	A	0.205	1.00	0.81 (0.79–0.83)	7.8E-51	<1	8	*BNC2*
	HGSOC		16,914,716	A	0.205	1.00	0.75 (0.72–0.77)	2.7E-68	<1	8	
9q31.1	NMOC	106,856,793	106,866,703	C	0.552	0.99	1.06 (1.04–1.09)	1.3E-08	1	135	*SMC2*
	HGSOC		106,866,703	C	0.552	0.99	1.08 (1.05–1.11)	4.0E-09	<1	135	
9q34.2	NMOC	136,155,000	136,155,000	T	0.197	0.96	1.10 (1.07–1.13)	7.6E-13	<1	9	*ABO*
	HGSOC		136,155,000	T	0.197	0.96	1.13 (1.10–1.17)	1.6E-14	<1	9	
10p12.31	NMOC	21,821,274	21,821,274	A	0.331	0.99	1.08 (1.06–1.11)	2.0E-12	<1	13	*MLLT10*
	HGSOC	–	21,821,274	A	0.331	0.99	1.08 (1.06–1.11)	1.5E-09	<1	13	–
12q24.31	NMOC	121,403,724	121,403,724	A	0.576	0.99	0.95 (0.93–0.97)	3.0E-07	17	31	*HNF1A-AS1*
12q24.31	HGSOC	121,403,724	12,140,3724	A	0.576	0.99	0.93 (0.91–0.95)	7.0E-09	1	31	*HNF1A-AS1*

(*Continued on next page*)

Table 2. *Continued*

Cytoband	Histotype	Lead SNP previous analysis[a]	Lead SNP reanalysis	Effect allele	Effect allele freq.	R^{2b}	OR (95% CI)[c]	*p* value	BFDP[d] (%)	# of CCVs[e]	Nearest gene to lead SNP
15q26.1	HGSOC	91,506,637	91,535,329	T	0.133	0.99	0.89 (0.86–0.93)	3.2E-9	<1	36	–
17q12	HGSOC	36,099,840	36,100,767	G	0.602	1.00	0.91 (0.89–0.94)	2.9E-12	<1	13	–
17q21.31	NMOC	44,187,257	44,187,257	G	0.216	0.99	1.09 (1.07–1.12)	1.6E-12	<1	2256	*KANSL1*
	HGSOC		44,187,257	G	0.216	0.99	1.13 (1.09–1.16)	2.0E-15	<1	2256	–
17q21.32	NMOC	46,500,673	46,472,432	C	0.273	1.00	1.13 (1.10–1.16)	1.2E-24	<1	90	*SKAP1*
	HGSOC		46,472,432	C	0.273	1.00	1.13 (1.10–1.16)	7.0E-19	<1	90	–
18q11.2	LGSOC	21,405,553	21,425,852	C	0.620	0.92	0.84 (0.80–0.89)	6.0E-9	1	145	*LAMA3*
19p13.11	HGSOC	17,390,291	17,390,291	C	0.297	1.00	1.19 (1.16–1.22)	8.8E-38	<1	13	*BABAM1*
19q13.2	MOC	39,738,787	39,738,787	T	0.307	0.88	0.69 (0.64–0.74)	8.5E-28	<1	4	*IFNL4*

[a]Build 37 position.
[b]Average imputation accuracy estimate.
[c]Odds ratio (95% confidence interval).
[d]Bayesian false discovery probability based on a prior probability on the alternative hypothesis of 1:10,000 and an upper likely effect size (odds ratio) of 1.2.
[e]CCVs, credible causal variants.

Figure 2. Regional plots show primary and secondary risk association signals for different histotypes at the 8q24.1 and 3q31.1 risk loci

SNPs are colored by histotype with light points representing *p* values from primary (unadjusted) analyses and dark points representing *p* values from CCVs.

(A) At the 8q24.21 risk locus we identified two independent risk signals for HGSOC (dark red points), two independent risk signals for NMOC (dark blue points), and one signal for MOC (dark orange points at 128,087,904). *p* values from unadjusted association analyses are plotted beneath CCVs in light red, blue, and orange for HGSOC, NMOC, and MOC, respectively.

(B) At the 2q31.1 risk locus, primary signals for the HGSOC, MOC, and NMOC histotypes were co-localized to the same region with evidence for a secondary signal for MOC located approximately 500 kb distal to the primary signal for the same histotype.

CCOC (Figure 3A). We found strong genetic correlations (correlation >0.5 and *p* value <0.05) between HGSOC and all other histotypes except MOC (Table S10). Heritability was partitioned across functional annotations in NMOC and HGSOC histotype groups only due to the low SNP-heritability estimates for other histotypes. We first partitioned genome-wide SNP heritability across functional annotations agnostic to cell type. SNPs in REs such as promoters, enhancers, and known TF binding sites contributed significantly to SNP-heritability estimates (Figure 3B). Promoters showed enrichment of up to 5% of risk SNPs, accounting for 28% of the SNP heritability

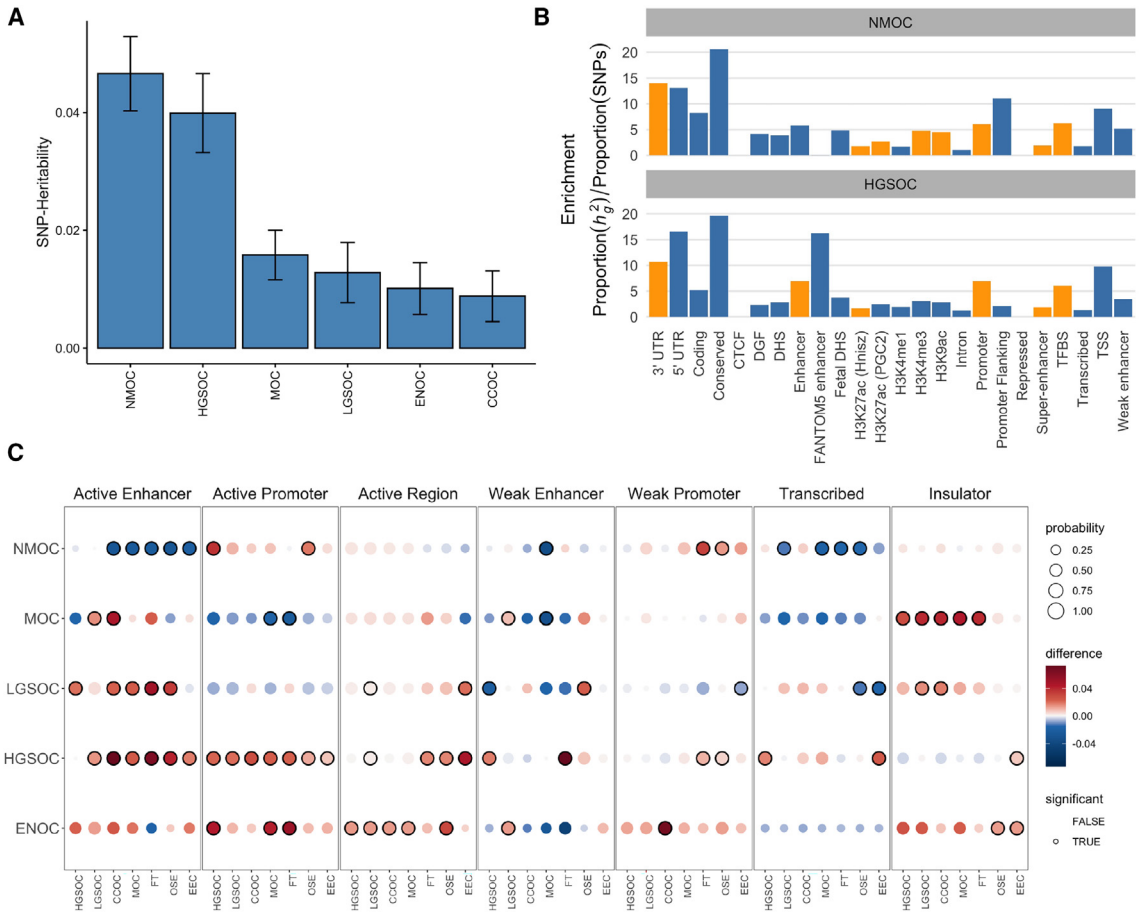

Figure 3. Estimates of SNP-heritability (h_g^2) and enrichment in regulatory features

(A) Overall SNP-heritability estimates for each EOC histotype. The error bars represent 95% confidence intervals.

(B) Enrichment of 24 functional annotations for NMOC and HGSOC risk loci calculated as the proportion of estimated SNP-heritability explained by the proportion of SNPs in each of several functional categories. Statistically significant annotations ($p < 0.05$) are shown in orange.

(C) Enrichment analyses for CCVs by EOC histotype (HGSOC, high-grade serous ovarian cancer; LGSOC, low-grade serous ovarian cancer; CCOC, clear-cell ovarian cancer; MOC, mucinous ovarian cancer; FT, fallopian tube epithelial cells; OSE, ovarian surface epithelial cells; and EEC, endometriosis epithelial cells). Enriched histotype-specific chromatin states are shown in red; depleted chromatin states are shown in blue; density of the color indicates strength of enrichment/depletion; size of the circle indicates the probability of enrichment, circles outlined met significance (probability > 0.98).

for NMOC (6.1-fold enrichment, p value = 0.01) and 32% of SNP heritability for HGSOC (7.0-fold enrichment, p value = 7.5×10^{-3}). Next, genome-wide SNP heritability was partitioned across ChromHMM chromatin states available for 18 EOC precursor or cell lines related to EOC histotypes (Table S11; Figure S1). The highest proportion of heritability accounted for by NMOC and HGSOC summary statistics was in active promoters in EOC and fallopian tube cell lines (Table S11; Figure S2).

PAINTOR[49] was used to compute the posterior probability of causality for all SNPs at NMOC and HGSOC risk loci by combining information from summary statistics, linkage disequilibrium structure, and functional annotations. This identified rs1562315 at 2q31 near *HAGLR*, rs1491585 and rs13429413 at 2q13 near *PAX8-AS1*, rs3745185 and rs4609972 at 19p13 near *BABAM1*, and rs6005807 at 22q12.2 near *TTC28* as putative causal risk alleles (Table S12).

Enrichment of CCVs in chromatin states defined in ovarian precursor and cancer cell types

Using FunciVar, we integrated the 4,008 CCVs with epigenomic features in ovarian precursor and cancer cell types. These analyses excluded the 17q21.31 risk region for which fine mapping identified a disproportionately large number of CCVs (n = 2,256 of a total of 4,008 CCVs genome wide). Thus, we evaluated 1,752 CCVs across 32 risk regions, including 246 CCVs associated with NMOC, 896 CCVs associated with HGSOC, 386 CCVs associated with LGSOC, 174 CCVs associated with MOC, and 50 CCVs associated with ENOC (Table S5). We identified histotype-specific enrichment for CCVs at HGSOC and LGSOC risk loci in both enhancers and promoters active across multiple EOC cell types. HGSOC CCVs were significantly enriched in active promoters in all EOC cell lines with significant enrichment for promoters characterized in fallopian tube cell lines at nine of 11 HGSOC risk loci

and in HGSOC cell lines at seven HGSOC risk loci. At eight of these loci, CCVs intersected a single promoter. There was also enrichment for CCVs at both LGSOC and HGSOC risk loci in active REs across all ovarian cancer cell types (Figure 3C; Table S13). We note that the NMOC CCVs are qualitatively different from the HGSOC CCVs and so the different enrichment patterns may simply reflect a different underlying biology. At several loci, CCVs for different histotypes were enriched in the same chromatin states, for example the 2q13a HGSOC and LGSOC risk locus (Table S14) where CCVs for both histotypes were enriched in active enhancers and the active promoter of *ACOXL*. At other loci, enrichment analyses identified different putative target genes, for example at 2q31.1 where CCVs for NMOC intersect active enhancers and promoters in the *HOXD3*, *HOXD1*, *HAGLR*, and *HAGLROS* gene clusters; CCVs for MOC intersected active enhancers and promoters for several miRNA and other non-coding RNA transcripts (Table S14). Functional annotation tracks and CCV information is available in a custom UCSC Genome Browser Session available at https://genome.ucsc.edu/s/pengp/ovariancancerregulatoryatlas.

Transcriptome-wide association analyses to identify candidate EOC susceptibility genes

We performed TWAS[51,54] to identify candidate susceptibility genes associated with risk alleles for the six histotype groups and to identify gene associations in regions not associated with EOC risk at genome-wide significance. First, we performed EOC-specific TWAS that was based on genotyping and RNA-seq data for 284 HGSOCs performed by TCGA,[50] 105 ovarian surface epithelial, and 60 fallopian tube epithelial precursor tissues.[21] Second, we performed a cross-tissue TWAS based on genotyping and RNA-seq data from 49 non-cancerous tissue types profiled by GTEx.[53] All TWAS analyses were coupled with colocalization analyses,[52,56,57] and we highlight only transcriptome-wide significant genes (FDR <0.05) for which there is concomitant evidence of colocalization between gene expression and genetic-association signals.

Tissue-specific TWAS identified 18 genes at transcriptome-wide significance (p value $\leq 4.2 \times 10^{-5}$ and FDR ≤ 0.05). This included five genes that were over 1 Mb away from any EOC index SNP, suggesting they represent new risk regions (Tables 3 and S15). The five genes included *THSD7A*, which was associated with HGSOC risk, in a region previously reported as a sub-genome-wide significant EOC risk region in Asian women (rs10260419—nearest coding gene *THSD7A*, p value = 10^{-7})[77] and *IRF5* associated with LGSOC risk, a gene that was reported to be associated with endometriosis risk in a previous TWAS.[78]

Cross-tissue TWAS identified 83 genes at transcriptome-wide significance (p value $\leq 2.2 \times 10^{-4}$ and FDR ≤ 0.045). This included 27 genes in 19 distinct genomic regions that were over 1 Mb away from any EOC index SNP (Tables 3 and S16). Some of these candidate genes have been identified as candidate eQTL associations in other studies: for example, *HOXD9* at 2q31.1,[25] *CHMP4C* at 8q21.13,[27] and *HAUS6* at 9p22.1,[21] while other genes represent candidate susceptibility genes at risk loci not previously reported, notably *SMC2* (structural maintenance of chromosomes 2) at 9p31.1. Sixty-four of the genes within 1 Mb of a genome-wide significant index risk SNP were identified as putative functional targets in previous studies, including *PAX8* at 2q14.1,[23] *TERT* at 5p15.33,[69] *BNC2* at 9p22.3,[22] *RCCD1* at 15q26.1,[79] *ABHD8* at 19p13.11,[24] and *HNF1B* at 17q12.[71] Other candidate EOC susceptibility genes identified in this study include *TIPARP* at 3q25.31 (p value = 2.6×10^{-65} in HGSOC) and *SKAP1* at 17q21.1 (p value = 5×10^{-17} in HGSOC).

Gene-regulatory interaction analyses in EOC risk regions

To identify *cis*-interactions between CCVs, candidate genes, and REs in EOC risk regions, we performed HiChIP-sequencing[58] for the histone mark H3K27ac in two normal-derived fallopian tube secretory epithelial cell lines (FT33 and FT246) and two HGSOC cell lines (UWB1.289 and Kuramochi). To quantify these interactions, we first summarized the evidence for every CCV with each gene within 500 kb flanking each fine-mapped risk locus, integrating GWAS and TWAS statistics, HiChIP-seq data, weighted ReMap TF data and VEP scores[61] paired with MotifbreakR.[18] This provided a "CCV/gene" score for each region (Table S7). We then summed the scores of all CCVs for each gene to produce a composite score (a "region/gene" score) reflecting the relative weight of evidence of interaction for every gene-CCV pair at each risk locus. These data are given in Table S8. In the combined MotifbreakR and TF ChIP-seq analysis, we identified CCVs that broke or created a TF motif located within a ChIP-seq-verified TF binding site in the ReMap database for any cell type and weighted the results based on looping class. The weighting prioritizes active enhancer-to-promoter looping interactions over inactive etc., see methods for details. The results of this analysis are presented in Table S17, and MotifbreakR results are in Table S18. Python library pygenometracks[80] was used to plot and visualize the most significant gene-regulatory interactions in each risk region.

These analyses identified the top candidates at each risk region based on evidence from genetic fine mapping (GWAS), genotype-gene expression associations identified by TWAS, chromatin architecture (H3K27Ac HiChIP), and TF binding (Remap/motifbreakR). At some risk loci, we identified strong evidence of interactions between CCVs and candidate genes that we identified by TWAS. For example, Figure 4 illustrates the 2q31 risk region, where both tissue-specific and cross-tissue TWAS identified several significant candidate susceptibility genes (Tables S15 and S16). The strongest interactions in this region were for two CCVs associated with NMOC risk and *HOXD-AS2* gene expression, *HOXD8* associated with RUNX2 TF binding sites, and for *HOXD3* associated with

Table 3. Genes identified by transcriptome-wide association analyses that were located over 1 Mb away from genome-wide significant EOC risk loci

Histotype	TWAS	Gene	Chr	Start[a]	End[a]	*p* value	FDR[b]	RCP/ColocPP4[c]
HGSOC	multi-tissue	*ST3GAL3*	1	44,171,495	44,396,837	8.1E-06	2.7E-03	0.71
HGSOC	multi-tissue	*TRIM46*	1	155,146,263	155,157,447	2.2E-05	6.5E-03	0.14
HGSOC	multi-tissue	*MUC1*	1	155,161,993	155,162,707	4.1E-06	1.5E-03	0.18
HGSOC	multi-tissue	*THBS3*	1	155,165,379	155,178,842	1.6E-04	3.7E-02	0.17
HGSOC	multi-tissue	*TDRD5*	1	179,560,748	179,660,407	8.0E-05	2.0E-02	0.42
HGSOC	multi-tissue	*PIGG*	4	492,999	533,989	1.7E-05	5.3E-03	0.17
HGSOC	multi-tissue	*MTO1*	6	74,171,301	74,218,959	2.0E-04	4.2E-02	0.14
HGSOC	multi-tissue	*AC058791.1*	7	130,561,568	130,598,069	1.8E-04	4.0E-02	0.47
HGSOC	multi-tissue	*PANK1*	10	91,339,254	91,405,329	1.4E-04	3.2E-02	0.11
HGSOC	multi-tissue	*RP11-80H5.7*	10	91,454,052	91,457,685	3.1E-05	8.8E-03	0.13
HGSOC	multi-tissue	*RP11-20L19.1*	12	91,202,984	91,204,466	1.0E-04	2.5E-02	0.15
HGSOC	multi-tissue	*CHMP6*	17	78,965,398	78,983,317	8.1E-06	2.7E-03	0.74
HGSOC	multi-tissue	*CTD-2561B21.11*	17	78,991,769	79,001,879	1.0E-04	2.5E-02	0.24
HGSOC	multi-tissue	*BAIAP2*	17	79,008,922	79,091,232	1.4E-04	3.3E-02	0.54
MOC	multi-tissue	*RP11-161D15.3*	4	174,845,358	174,912,175	5.2E-06	9.0E-03	0.37
NMOC	multi-tissue	*ST3GAL3*	1	44,171,495	44,396,837	1.4E-04	3.0E-02	0.44
NMOC	multi-tissue	*TNN*	1	175,036,969	175,117,211	2.5E-06	9.3E-04	0.20
NMOC	multi-tissue	*TLR6*	4	38,825,336	38,858,438	3.8E-05	1.0E-02	0.14
NMOC	multi-tissue	*FAM114A1*	4	38,869,298	38,947,360	9.0E-05	2.2E-02	0.15
NMOC	multi-tissue	*NAT2*	8	18,248,792	18,258,728	1.6E-04	3.4E-02	0.10
NMOC	multi-tissue	*FRMPD1*	9	37,650,951	37,746,901	1.3E-04	2.9E-02	0.18
NMOC	multi-tissue	*NDUFA12*	12	95,290,831	95,397,524	1.7E-05	4.8E-03	0.13
NMOC	multi-tissue	*FGD6*	12	95,470,525	95,611,258	5.5E-05	1.5E-02	0.29
NMOC	multi-tissue	*SYNE2*	14	64,228,617	64,693,151	1.2E-04	2.8E-02	0.27
NMOC	multi-tissue	*TEFM*	17	29,224,354	29,233,256	3.8E-06	1.3E-03	0.45
NMOC	multi-tissue	*RGS19*	20	62,704,534	62,711,341	4.6E-05	1.2E-02	0.34
NMOC	multi-tissue	*OPRL1*	20	62,711,435	62,731,996	1.5E-04	3.2E-02	0.20
NMOC	multi-tissue	*BIK*	22	43,506,756	43,525,718	5.5E-05	1.5E-02	0.54
CCOC	tissue-specific	*DEPDC5*	22	32,149,944	32,303,012	4.2E-05	2.0E-02	0.90
HGSOC	tissue-specific	*THSD7A*	7	11,414,173	11,871,824	1.1E-04	1.2E-02	0.77
HGSOC	tissue-specific	*FLJ37201*	10	91,451,057	91,457,685	1.7E-04	1.5E-02	0.89
LGSOC	tissue-specific	*GSK3B*	3	119,540,804	119,813,264	9.9E-06	5.2E-03	0.67
LGSOC	tissue-specific	*IRF5*	7	128,577,769	128,590,086	5.9E-07	9.2E-04	0.99
NMOC	tissue-specific	*FLJ37201*	10	91,451,057	91,457,685	1.7E-04	1.7E-02	0.65

[a]Build 37 position.
[b]Only FDR <0.05 results are shown, which was the threshold used to identify transcriptome-wide significant genes.
[c]Colocalization probabilities used were RCP (threshold RCP >0.1) for multi-tissue MultiXcan transcriptome-wide association analysis and ColocPP4 (threshold ColocPP4 >0.6) for tissue-specific FUSION transcriptome-wide association analysis.

enhancers and *LINC01116*. CCVs associated with MOC were located about 500 kb distal to the CCV cluster associated with NMOC. The strongest interaction for these CCVs was also with HOXD-AS2. Other interactions were found with *HAGLROS*, *LINC01116*, and *HOXD1*. Thus, the genes with the strongest interactions were the same for NMOC and MOC CCVs even though the top CCVs for each subtype were separated by hundreds of kilobases, implying cell-type-specific dysregulation of the same gene in these histologically distinct cancers.

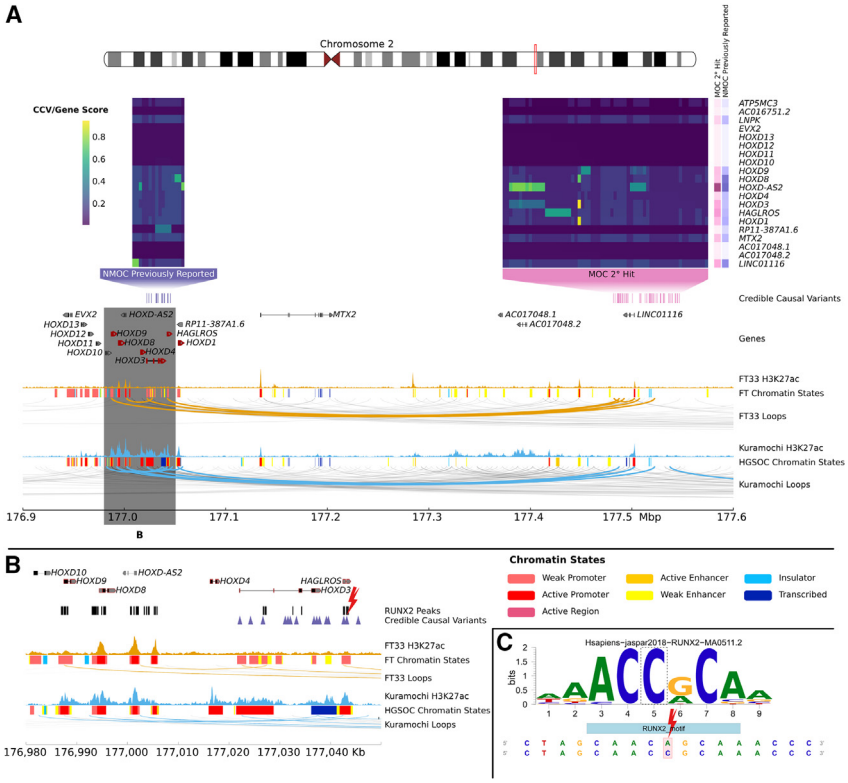

Figure 4. Looping interactions between candidate genes, CCVs, and REs at the 2q31.1 locus

(A) Locus plot of 2q31.1 showing the integration of genetic fine mapping data (NMOC CCVs in purple, MOC CCVs in pink) with chromatin state calls by cell/tissue type. The profiles of *cis*-interactions in the fallopian tube cell line FT33 and HGSOC cell line Kuramochi were very similar. A heatmap shows the CCV/gene scores associated with each gene in the region. The greater the intensity the stronger the evidence of the interaction. This is summarized across CCVs separately for each histotype to the right of the heatmap.

(B) The gray box zooms into a region of strongest interaction between *HOXD-AS2* and two CCVs associated with NMOC risk about 500 kb distal to the gene. A cluster of NMOC CCVs are located in an active enhancer/promoter region of *HAGLROS* in both FT33 and Kuramochi cells. Kuramochi cells exhibit looping between the CCV rs6755766 (chr2:177,043,205) within a Remap ChIP-seq RUNX2 peak that breaks a RUNX2 motif.

(C) Highlights this motif disruption.

At other risk loci we identified strong evidence of interactions between CCVs and candidate genes not identified by TWAS (i.e., eQTL-negative genes). Notably, at the 8q24 risk region, there were several independent, histotype-specific risk signals (Figure 5). At this locus, CCVs for the newly discovered secondary signal for HGSOC are located in untranscribed regions of *PVT1* and show prominent interaction with the *MYC* promoter. This region is enriched for TEAD3/4 and PAX8 TFBSs with one CCV (chr8:129076573:C>T [genome build 37], rs13255292) predicted to disrupt a TEAD3/4 motif that shows evidence for binding of TEAD3/4 and cobinding between TEAD3/4 and PAX8 at the *MYC* promoter. Consistent with this, depletion of *PAX8* expression in the HGSOC line OVCAR4 using an siRNA pool against *PAX8* significantly perturbed the expression of *MYC* (log2FoldChange = −0.55; adjusted p value = 5.3×10^{-7}) (Figure 5B). By contrast, CCVs for NMOC located downstream and proximal to *MYC* showed evidence of interactions with *PVT1*. The strongest region/gene score prediction for both genes jointly was for a single CCV for HGSOC (chr8:129069820:G>A [genome build 37], rs35916594), which disrupts a CREB1 MotifbreakR-TFBS. CREB1 also binds to both *MYC* and *PVT1* promoters. Two CCVs distal to *LINC00824* representing the primary independent signal in the region show strongest looping to the *MYC* promoter.

Discussion

We have increased the statistical power over previous ovarian cancer GWASs by using genotyping data from the OCAC with additional controls combined with improved imputation using the HRC reference panel (about 65,000 haplotypes). We identified five new EOC risk loci at genome-wide significance. Increase in sample size was provided by an increase in the number of controls compared to previous analyses together with improved imputation for some variants. We also identified 11 signals at six risk regions that were independent of the primary signal and confirmed 24 previously reported EOC risk associations. The BFDP was $\leq$3% for association signals at all five new risk loci and <10% for six of the 11 independent signals at previously identified EOC risk loci, indicating a high likelihood that the majority of these risk associations are real.

Fifteen previously reported genome-wide significant risk associations did not replicate at nominal genome-wide significance. Of these, two were reported in a study of Han Chinese women,[81] three were published from a multi-cancer meta-analysis[73] and so may not be expected replicate for ovarian cancer alone without a substantial increase in sample size, one was only reported in ovarian cancer occurring in individuals with *BRCA1* mutations,[82] and nine were reported either for OCAC data alone or from a meta-analysis of OCAC and CIMBA data.[11,72] For all nine of these risk loci, p values of the original associations were close to the nominal threshold for genome-wide significance. An increase in sample size in the current study combined with the winner's curse effect in the original data is a likely explanation for why these p values are now slightly larger and no longer reach genome-wide significance.

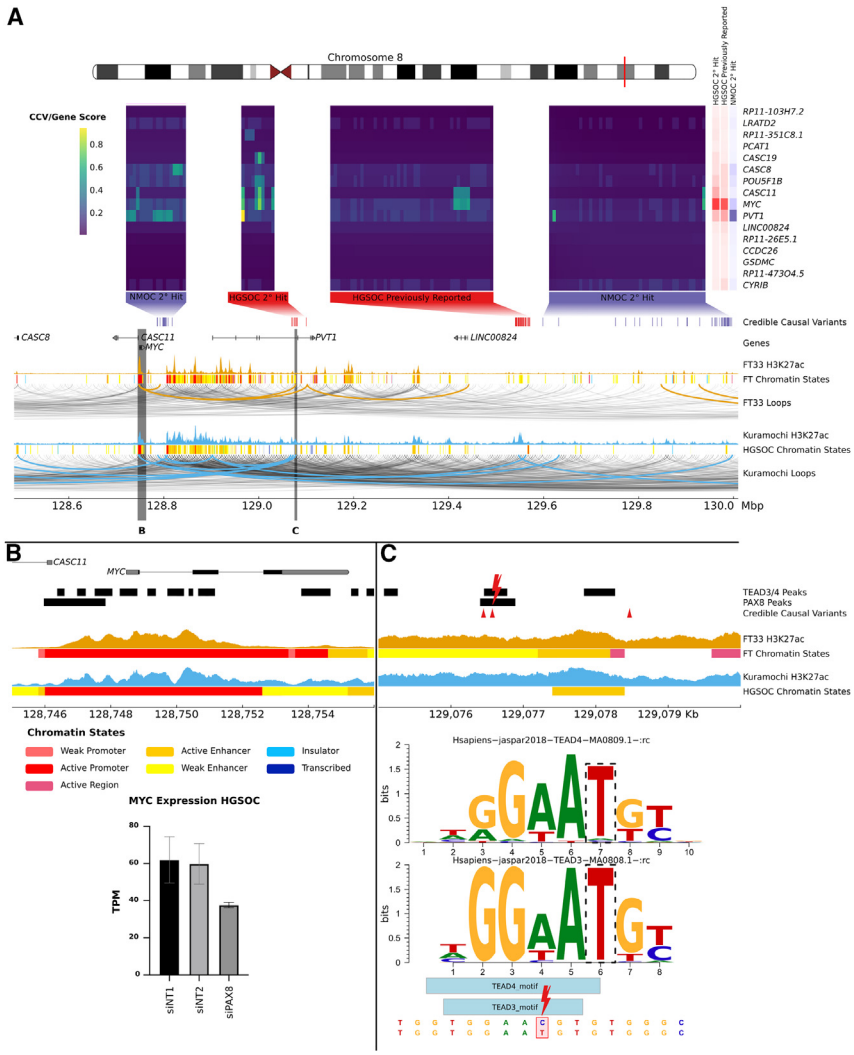

Figure 5. Looping interactions between candidate genes, CCVs, and REs at the 8q24.1 risk locus

(A) Locus plot of 8q24.1 showing the integration of genetic fine mapping data (HGSOC CCVs in red, NMOC CCVs in green) with chromatin state calls by cell/tissue type, and *cis*-interaction analysis in the fallopian tube cell line FT33 and HGSOC cell line Kuramochi. A heatmap shows the CCV/gene scores associated with each gene in the region. The greater the intensity, the stronger the evidence of and interaction. This is summarized across CCVs separately for each histotype to the right of the heatmap.

(B and C) Zoom in on the two regions highlighted: (B) highlights chromatin state calls in around the *MYC* gene, which shows greatest interaction with HGSOC SNPs located in region C. Epigenomic profiling indicate activation of the *MYC* promoter in both FT33 and Kuramochi cell lines. HiChIP identifies a strong interaction with a single HGSOC CCV that disrupts a Tead3/4 TF motif in a remap PAX8-TEAD3/4 ChIPseq co-binding site (see black bars in top tracks) in (C), implicating a biological connection between MYC and *PAX8*/TEAD3/4. Consistent with this, knock down of *PAX8* (siPAX8) in HGSOC cells resulted in a reduction in *MYC* expression compared to two controls (siNT1 and siNT2) (error bars show standard error) (B).

Our findings confirm previous reports that genetic risk factors reflect the underlying disease heterogeneity. Of the 16 new risk associations we identified (five from primary and 11 from conditional analyses), seven were most strongly associated with HGSOC, two with LGSOC, and three with MOC. The remaining four risk associations were most strongly associated with all NMOCs (i.e., HGSOC, LGSOC, CCOC, and ENOC cases combined). While there are clear differences in the histotype-specific risks for different genetic variants, there was a much greater distinction between MOC compared to other histotypes. For some loci, the most strongly associated histotype was different from the histotype with the largest effect estimate. For example, rs336126 and rs2070368 were most strongly associated with NMOCs but had the largest effect sizes for LGSOC. We found some evidence that alleles that increased risk of one histotype, may confer protection for another histotype. For example, at the 17q12 risk locus, the same risk variants that were associated with an increased risk of HGSOC at genome-wide levels showed marginal evidence for protection against CCOC and MOC, respectively.

More than 99% of CCVs in EOC risk-regions lie in non-protein coding DNA. We hypothesized that the CCVs are enriched in REs that show specificity in ovarian cancer-associated tissues. More than 80% of REs shared across tissues are in promoters whereas >80% of cell-type-specific REs are enhancers. We found CCVs were enriched in both promoters and enhancers characterized by epigenomic profiling of ovarian normal precursor and cancer cell lines, suggesting they play a role in both tissue-agnostic and tissue-specific gene regulation; previous observations have suggested that cell-type-specific enhancers likely drive the spatial and temporal diversity of gene expression for ovarian cancer-associated tissue types.[14,19,83] This included several likely causal risk variants and novel associated transcripts, most notably rs1491585 and rs13429413 at 2q13 associated with *PAX8-AS1* expression, rs10069690 at 5p15.33 associated with *TERT* expression, and rs3745185 and rs4609972 at 19p13.11 associated with expression of the *BRCA1*-interacting gene, *BABAM1*. At several other risk loci, cross-tissue TWAS confirmed the results of previous findings, for example, *CHMP4C* at 8q21.13,[21] *ABHD8* at 19p13.11,[24]

and *MLLT10* at 10p12.31.[27] While there may have been a loss of tissue specificity in our cross-tissue TWAS, the substantially greater sample size compared to our tissue specific TWAS has likely provided a significant increase in statistical power to identify candidate gene associations.

It has been estimated in studies of breast cancer that thousands of common germline genetic variants contribute to the burden of heritability for the disease.[28] If the same is true for ovarian cancer, then our studies to date have identified only a small fraction of the excess heritable risk for the disease. As with other GWASs, our ability to identify additional risk regions for ovarian cancer at $p < 5 \times 10^{-8}$ is limited by sample size. In this study, we used genotype data for more than 26,000 ovarian cancer cases and over 100,000 healthy control subjects. Modeling studies of the underlying genetic architecture of cancer suggest that a quadrupling of the sample size would be needed to be able to explain a substantial fraction of the polygenic component of ovarian cancer risk.[84] Our data also highlight the significance of disease heterogeneity in identifying risk variants. The different histotypes appear to be driven by different biology and different underlying genetics. It is not surprising that the vast majority of risk alleles were identified for the two most commonly designated histotype groups, HGSOC and NMOC, for which genotype data were available for the largest sample sizes (15,588 and 23,394 cases respectively). We failed to identify any genome-wide significant risk loci associated specifically with endometrioid (n = 2,877) or clear-cell (n = 1,427) ovarian cancers. In contrast, and despite the small sample sizes, we were able to identify several risk regions associated with low-grade serous (n = 2,749) and mucinous (n = 2,587) ovarian cancers perhaps indicating that a diagnosis of these histotypes is more definitive and/or that their biology and underlying genetics are sufficiently distinct from other histotypes to identify histotype-specific risk signals. Identifying additional common low-risk variants for ovarian cancer will remain a challenge without substantially increasing the sample size of these studies. Applying additional "weighting" to sub genome-wide significant risk regions based on functional evidence from epigenomic and gene expression studies is one approach that may provide evidence for true risk associations. For example, in this study, cross-tissue TWAS identified 19 distinct genomic regions that were over 1 Mb away from any genome-significant index SNP but for which there was significant evidence of at least one genotype-gene expression association signal suggesting that the integration of genetics and gene expression data has identified several novel risk loci.

Chromatin interaction data from genome-wide interactome methods like HiChIP-seq provide context-specific resolution to validate candidate genes identified by TWAS in risk regions and their predicted connections to CCVs and REs. These methods can also identify novel candidate genes in risk regions where no TWAS association has been found (i.e., TWAS-negative risk regions). HiChIP offers a window into latent or "potential" target genes because looping is relatively static and not consistently altered in response to perturbations,[85,86] and looping can inform on the context-specific nature of TWAS as shown for autoimmune disorders.[87] In the current study, we highlight two examples—one a TWAS-positive locus (2q31), the other a TWAS-negative locus (8q24)—where the integration of disease-specific looping data has identified likely susceptibility gene targets. At the 2q31 region, which is associated with risk of NMOC and MOC histotypes, TWAS identified several significant gene associations in the HOXD family of transcription factors, many of which were shared across the different histotypes. HiChIP enabled us to identify *HOXD-AS2* as the most likely target gene for these because its interaction with CCVs for both NMOC and MOC suggests it may be a key tissue-specific regulator of gene expression of one or more *HOXD* genes in the region. Interestingly, the top-scoring CCVs for each gene are complementary and non-overlapping, which suggest the risk in this region could be mediated independently by several genes as a haplotype. In the 8q24 region, HiChIP identified strong evidence of interactions between the HGSOC risk variants and the *MYC* promoter and between the NMOC risk variants and the long non-coding lncRNA *PVT1*. There is substantial evidence that both genes are involved in EOC development,[88,89] and *MYC* plays a critical role in the pathogenesis of several cancer types.[90] CCVs in the region that show strong interactions with MYC are also predicted to affect binding to TEAD3/4 sites that interact with PAX8. This is supported by experiments in which we knocked down the expression of *PAX8* in an HGSOC line and observed a concomitant reduction in *MYC* expression, suggesting a functional link between PAX8 binding and MYC transcriptional control in HGSOCs. PAX8 is possibly the most widely recognized lineage-specific TF in the development of fallopian tube epithelia and its neoplastic progression to HGSOCs. Consistent with this, a recent study identified significant genome-wide enrichment of non-coding somatic mutations in the TF motif for TEAD4 and its binding partner PAX8 and a burden of *cis*-REs associated with PAX8 as the most frequently mutated set of enhancers in ovarian cancers.[83]

Thus, HiChIP-seq may distinguish between likely causal susceptibility genes and false-positive TWAS associations and/or co-regulated genes in the same region, although there remain challenges in using this method; it has a limited resolution (~5 kb), which may reduce its sensitivity for detecting interacting regions containing short REs, and in this study, HiChIP-seq was directed to regulatory features marked only by H3K27ac (i.e., it is not agnostic to all epigenomic features). Consequently, a proportion of *cis* gene-regulatory interactions in ovarian cancer risk regions will likely have been missed.

Assigning variants to genes remains the defining challenge for post-GWAS research. The target gene scoring

approach that we adopted in this study relied on triangulating across six lines with statistical and functional evidence. It is worth noting that the absence of a "gold standard" or "truth set" of target genes makes it particularly difficult to evaluate the ultimate effectiveness of our scoring, and we present our results simply as one measure of confidence in a target. For example, other variant-to-gene approaches have been benchmarked on metabolite GWAS data, taking advantage of the fact that target genes in metabolite GWAS frequently converge on well-annotated biochemical pathways.[91] However, such benchmarking is not straightforward in the context of complex diseases such as cancer. While we only considered gene expression QTLs here, future approaches could be extended to incorporate other emerging cell-type-specific QTLs such as chromatin[92] and DNA methylation.[93] The next functional steps in prioritizing credible causal risk variants and genes emerging from our work will likely involve a combination of genome editing, single-cell profiling, and germline-somatic data integration to further characterize the genetic and epigenomic mechanisms underlying each risk locus. In addition, by editing credible causal risk variants and REs to create isogenic cell lines that differ only by risk genotype, we will be able to measure how specific genotypes impact target gene expression.[94] Given the germline genetic heterogeneity that we find underpins the different histotypes of EOC, single-cell RNA sequencing of ovarian surface, fallopian tube, and other related tissues, coupled with statistical disease relevance scores,[95] has the potential to pinpoint the specific cellular origins of each subtype. Finally, the pipeline of variant-to-function post-GWAS studies in EOC will likely require an evaluation of the downstream effects of GWAS-identified target gene expression on somatic mutations and mutational signatures.[96]

In summary, we describe a comprehensive genetic association study for ovarian cancer. It expands upon previous findings indicating that different genetic alleles confer risks to different ovarian cancer histotypes but also that several risk regions confer risk to two or more histotypes, suggesting there are similarities in the underlying biology driving the disease spectrum of ovarian cancer. Imputation identified credible causal risk alleles in each region, which show a significant enrichment with disease-specific epigenomic features that represent candidate regulatory targets of risk alleles. TWAS identified several plausible ovarian cancer candidate susceptibility genes in genome-wide significant risk regions but also gene associations within multiple regions at sub-genome wide significance, suggesting TWAS is an approach that can be used for the further discovery of disease-associated risk alleles. By combining genetic, TWAS, and epigenomic profiling with HiChIP-seq interactome analysis, we also show the likely functional mechanisms underlying several EOC risk loci, representing hitherto unrecognized aspects of biology in the development of different EOC histotypes. Taken together, our approach represents a functional framework for understanding the biological mechanisms underlying common low penetrance risk regions for ovarian and other cancers in the future.

Data and code availability

The summary statistics generated in the genome-wide association analysis generated in this study are available at the European Genome Phenome Archive (www.ebi.ac.uk/gwas/home) under dataset accession identifier GEO: GCP000499. The functional genomics datasets are available in a custom UCSC Genome Browser Session: https://genome.ucsc.edu/s/pengp/OvarianCancerRegulatoryAtlas, which includes CCVs (for NMOC, HGSOC, ENOC, LGSOC, MOC, and CCOC), ovarian cancer-related chromatin states (for FT, IOSE, NMOC, HGSOC, ENOC, LGSOC, MOC, and CCOC), PAINTOR predicted causal SNPs (for NMOC and HGSOC), and chromatin interactions (for FT33, FT246, Kutamochi, and UWB1.289). Data for H3K27Ac HiChIP interactions are available in the Gene Expression Omnibus (GEO) under project GEO: GSE255142. Data availability of histone ChIP-seq are detailed in Table S6.

Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.ajhg.2024.04.011.

Acknowledgments

The analyses presented in this manuscript have been funded in part by NIH/NCI grants: R01CA207456, R01CA204954, R01CA211707, R01CA211575, R01CA207456, R21CA220078, R00CA256519, and U19C8804/A7058. A portion of funding was provided as part of an Institutional commitment to The Center for Bioinformatics and Functional Genomics at Cedars-Sinai Medical Center.

Declaration of interests

The authors declare no competing interests.

Received: July 13, 2022
Accepted: April 16, 2024
Published: May 8, 2024

Web resources

CADD, https://cadd.gs.washington.edu/
Cutadapt, https://cutadapt.readthedocs.io/en/stable/
DESeq2, https://doi.org/10.1186/s13059-014-0550-8
GitHub, www.github.com/macs3-project/MACS
GTEx, https://www.gtexportal.org/home/
Haplotype Reference Consortium, https://www.sanger.ac.uk/collaboration/haplotype-reference-consortium/#:~:text=The%20Haplotype%20Reference%20Consortium%20
Macs2, https://github.com/macs3-project/MACS
Michigan Imputation Server, https://imputationserver.readthedocs.io/en/latest/

Minimac3, https://genome.sph.umich.edu/wiki/Minimac3
PolyPhen: http://genetics.bwh.harvard.edu/pph2/
Pygenometracks: https://github.com/deeptools/pyGenomeTracks
ReMap: https://remap2022.univ-amu.fr/
SIFT: https://bioinformaticshome.com/tools/descriptions/SIFT.html#gsc.tab=0
SHAPE-IT: http://www.griv.org/shapeit/
S-prediXan: https://github.com/hakyimlab/PrediXcan
Variant Effect Predictor: https://useast.ensembl.org/info/docs/tools/vep/index.html

References

1. Song, H., Cicek, M.S., Dicks, E., Harrington, P., Ramus, S.J., Cunningham, J.M., Fridley, B.L., Tyrer, J.P., Alsop, J., Jimenez-Linan, M., et al. (2014). The contribution of deleterious germline mutations in BRCA1, BRCA2 and the mismatch repair genes to ovarian cancer in the population. Hum. Mol. Genet. *23*, 4703–4709.
2. Ramus, S.J., Song, H., Dicks, E., Tyrer, J.P., Rosenthal, A.N., Intermaggio, M.P., Fraser, L., Gentry-Maharaj, A., Hayward, J., Philpott, S., et al. (2015). Germline mutations in the BRIP1, BARD1, PALB2, and NBN genes in women with ovarian cancer. J. Natl. Cancer Inst. *107*, djv214.
3. Song, H., Dicks, E., Ramus, S.J., Tyrer, J.P., Intermaggio, M.P., Hayward, J., Edlund, C.K., Conti, D., Harrington, P., Fraser, L., et al. (2015). Contribution of germline mutations in the RAD51B, RAD51C, and RAD51D genes to ovarian cancer in the population. J. Clin. Oncol. *33*, 2901–2907.
4. Vang, R., Shih, I.-M., and Kurman, R.J. (2013). Fallopian tube precursors of ovarian low- and high-grade serous neoplasms. Histopathology *62*, 44–58.
5. Gounaris, I., and Brenton, J.D. (2015). Molecular pathogenesis of ovarian clear cell carcinoma. Future Oncol. *11*, 1389–1405.
6. Samartzis, E.P., Noske, A., Dedes, K.J., Fink, D., and Imesch, P. (2013). ARID1A mutations and PI3K/AKT pathway alterations in endometriosis and endometriosis-associated ovarian carcinomas. Int. J. Mol. Sci. *14*, 18824–18849.
7. Yamamoto, S., Tsuda, H., Takano, M., Tamai, S., and Matsubara, O. (2012). Loss of ARID1A protein expression occurs as an early event in ovarian clear-cell carcinoma development and frequently coexists with PIK3CA mutations. Mod. Pathol. *25*, 615–624.
8. Stratton, J.F., Pharoah, P., Smith, S.K., Easton, D., and Ponder, B.A. (1998). A systematic review and meta-analysis of family history and risk of ovarian cancer. Br. J. Obstet. Gynaecol. *105*, 493–499.
9. Jervis, S., Song, H., Lee, A., Dicks, E., Tyrer, J., Harrington, P., Easton, D.F., Jacobs, I.J., Pharoah, P.P.D., and Antoniou, A.C. (2014). Ovarian cancer familial relative risks by tumour subtypes and by known ovarian cancer genetic susceptibility variants. J. Med. Genet. *51*, 108–113.
10. Kar, S.P., Berchuck, A., Gayther, S.A., Goode, E.L., Moysich, K.B., Pearce, C.L., Ramus, S.J., Schildkraut, J.M., Sellers, T.A., and Pharoah, P.D.P. (2018). Common genetic variation and susceptibility to ovarian cancer: current insights and future directions. Cancer Epidemiol. Biomarkers Prev. *27*, 395–404.
11. 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.
12. Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J., et al. (2012). Systematic localization of common disease-associated variation in regulatory DNA. Science *337*, 1190–1195.
13. Hazelett, D.J., Rhie, S.K., Gaddis, M., Yan, C., Lakeland, D.L., Coetzee, S.G., Henderson, B.E., Noushmehr, H., et al.; Ellipse/GAME-ON consortium; and Practical consortium (2014). Comprehensive functional annotation of 77 prostate cancer risk loci. PLoS Genet. *10*, e1004102.
14. Coetzee, S.G., Shen, H.C., Hazelett, D.J., Lawrenson, K., Kuchenbaecker, K., Tyrer, J., Rhie, S.K., Levanon, K., Karst, A., Drapkin, R., et al. (2015). Cell-type-specific enrichment of risk-associated regulatory elements at ovarian cancer susceptibility loci. Hum. Mol. Genet. *24*, 3595–3607.
15. Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T., et al. (2014). An atlas of active enhancers across human cell types and tissues. Nature *507*, 455–461.
16. ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74.
17. Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., Ching, C.W., et al. (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature *459*, 108–112.
18. Coetzee, S.G., Coetzee, G.A., and Hazelett, D.J. (2015). motifbreakR: an R/Bioconductor package for predicting variant effects at transcription factor binding sites. Bioinformatics *31*, 3847–3849.
19. Jones, M.R., Peng, P.-C., Coetzee, S.G., Tyrer, J., Reyes, A.L.P., Corona, R.I., Davis, B., Chen, S., Dezem, F., Seo, J.-H., et al. (2020). Ovarian Cancer Risk Variants Are Enriched in Histotype-Specific Enhancers and Disrupt Transcription Factor Binding Sites. Am. J. Hum. Genet. *107*, 622–635.
20. Lu, Y., Beeghly-Fadiel, A., Wu, L., Guo, X., Li, B., Schildkraut, J.M., Im, H.K., Chen, Y.A., Permuth, J.B., Reid, B.M., et al. (2018). A Transcriptome-Wide Association Study Among 97,898 Women to Identify Candidate Susceptibility Genes for Epithelial Ovarian Cancer Risk. Cancer Res. *78*, 5419–5430.
21. 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.
22. 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.
23. Kelemen, L.E., Lawrenson, K., Tyrer, J., Li, Q., Lee, J.M., Seo, J.-H., Phelan, C.M., Beesley, J., Chen, X., Spindler, T.J., et al. (2015). Genome-wide significant risk associations for mucinous ovarian carcinoma. Nat. Genet. *47*, 888–897.
24. 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.
25. Lawrenson, K., Li, Q., Kar, S., Seo, J.-H., Tyrer, J., Spindler, T.J., Lee, J., Chen, Y., Karst, A., Drapkin, R., et al. (2015). Cis-eQTL

analysis and functional validation of candidate susceptibility genes for high-grade serous ovarian cancer. Nat. Commun. *6*, 8234.
26. Amos, C.I., Dennis, J., Wang, Z., Byun, J., Schumacher, F.R., Gayther, S.A., Casey, G., Hunter, D.J., Sellers, T.A., Gruber, S.B., et al. (2017). The oncoarray consortium: A network for understanding the genetic architecture of common cancers. Cancer Epidemiol. Biomarkers Prev. *26*, 126–135.
27. 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*, 362–370e3702.
28. 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.
29. Michailidou, K., Hall, P., Gonzalez-Neira, A., Ghoussaini, M., Dennis, J., Milne, R.L., Schmidt, M.K., Chang-Claude, J., Bojesen, S.E., Bolla, M.K., et al. (2013). Large-scale genotyping identifies 41 new loci associated with breast cancer risk. Nat. Genet. *45*, 353–361e3612.
30. Das, S., Forer, L., Schönherr, S., Sidore, C., Locke, A.E., Kwong, A., Vrieze, S.I., Chew, E.Y., Levy, S., McGue, M., et al. (2016). Next-generation genotype imputation service and methods. Nat. Genet. *48*, 1284–1287.
31. Delaneau, O., Coulonges, C., and Zagury, J.-F. (2008). Shape-IT: new rapid and accurate algorithm for haplotype inference. BMC Bioinf. *9*, 540.
32. Loh, P.-R., Danecek, P., Palamara, P.F., Fuchsberger, C., A Reshef, Y., K Finucane, H., Schoenherr, S., Forer, L., McCarthy, S., Abecasis, G.R., et al. (2016). Reference-based phasing using the Haplotype Reference Consortium panel. Nat. Genet. *48*, 1443–1448.
33. Hernandez, L., Kim, M.K., Lyle, L.T., Bunch, K.P., House, C.D., Ning, F., Noonan, A.M., and Annunziata, C.M. (2016). Characterization of ovarian cancer cell lines as in vivo models for preclinical studies. Gynecol. Oncol. *142*, 332–340.
34. Reyes, A.L.P., Silva, T.C., Coetzee, S.G., Plummer, J.T., Davis, B.D., Chen, S., Hazelett, D.J., Lawrenson, K., Berman, B.P., Gayther, S.A., and Jones, M.R. (2019). GENAVi: a shiny web application for gene expression normalization, analysis and visualization. BMC Genom. *20*, 745.
35. Landt, S.G., Marinov, G.K., Kundaje, A., Kheradpour, P., Pauli, F., Batzoglou, S., Bernstein, B.E., Bickel, P., Brown, J.B., Cayting, P., et al. (2012). ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. Genome Res. *22*, 1813–1831.
36. Ernst, J., and Kellis, M. (2012). ChromHMM: automating chromatin-state discovery and characterization. Nat. Methods *9*, 215–216.
37. Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.
38. Frankish, A., Diekhans, M., Ferreira, A.-M., Johnson, R., Jungreis, I., Loveland, J., Mudge, J.M., Sisu, C., Wright, J., Armstrong, J., et al. (2019). GENCODE reference annotation for the human and mouse genomes. Nucleic Acids Res. *47*, D766–D773.
39. Kasowski, M., Kyriazopoulou-Panagiotopoulou, S., Grubert, F., Zaugg, J.B., Kundaje, A., Liu, Y., Boyle, A.P., Zhang, Q.C., Zakharia, F., Spacek, D.V., et al. (2013). Extensive variation in chromatin states across humans. Science *342*, 750–752.
40. Bulik-Sullivan, B.K., Loh, P.-R., Finucane, H.K., Ripke, S., Yang, J., Schizophrenia Working Group of the Psychiatric Genomics Consortium, Patterson, N., Daly, M.J., Price, A.L., and Neale, B.M. (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat. Genet. *47*, 291–295.
41. Finucane, H.K., Bulik-Sullivan, B., Gusev, A., Trynka, G., Reshef, Y., Loh, P.-R., Anttila, V., Xu, H., Zang, C., Farh, K., et al. (2015). Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat. Genet. *47*, 1228–1235.
42. Gusev, A., Lee, S.H., Trynka, G., Finucane, H., Vilhjálmsson, B.J., Xu, H., Zang, C., Ripke, S., Bulik-Sullivan, B., Stahl, E., et al. (2014). Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases. Am. J. Hum. Genet. *95*, 535–552.
43. Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. Genome Res. *12*, 996–1006.
44. Lindblad-Toh, K., Garber, M., Zuk, O., Lin, M.F., Parker, B.J., Washietl, S., Kheradpour, P., Ernst, J., Jordan, G., Mauceli, E., et al. (2011). A high-resolution map of human evolutionary constraint using 29 mammals. Nature *478*, 476–482.
45. Ward, L.D., and Kellis, M. (2012). Evidence of abundant purifying selection in humans for recently acquired regulatory functions. Science *337*, 1675–1678.
46. Hoffman, M.M., Ernst, J., Wilder, S.P., Kundaje, A., Harris, R.S., Libbrecht, M., Giardine, B., Ellenbogen, P.M., Bilmes, J.A., Birney, E., et al. (2013). Integrative annotation of chromatin elements from ENCODE data. Nucleic Acids Res. *41*, 827–841.
47. Trynka, G., Sandor, C., Han, B., Xu, H., Stranger, B.E., Liu, X.S., and Raychaudhuri, S. (2013). Chromatin marks identify critical cell types for fine mapping complex trait variants. Nat. Genet. *45*, 124–130.
48. Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-André, V., Sigova, A.A., Hoke, H.A., and Young, R.A. (2013). Super-enhancers in the control of cell identity and disease. Cell *155*, 934–947.
49. Kichaev, G., Yang, W.-Y., Lindstrom, S., Hormozdiari, F., Eskin, E., Price, A.L., Kraft, P., and Pasaniuc, B. (2014). Integrating functional data to prioritize causal variants in statistical fine-mapping studies. PLoS Genet. *10*, e1004722.
50. Cancer Genome Atlas Research Network (2011). Integrated genomic analyses of ovarian carcinoma. Nature *474*, 609–615.
51. Gusev, A., Ko, A., Shi, H., Bhatia, G., Chung, W., Penninx, B.W.J.H., Jansen, R., de Geus, E.J.C., Boomsma, D.I., Wright, F.A., et al. (2016). Integrative approaches for large-scale transcriptome-wide association studies. Nat. Genet. *48*, 245–252.
52. Giambartolomei, C., Vukcevic, D., Schadt, E.E., Franke, L., Hingorani, A.D., Wallace, C., and Plagnol, V. (2014). Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet. *10*, e1004383.
53. GTEx Consortium (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science *369*, 1318–1330.
54. Barbeira, A.N., Pividori, M., Zheng, J., Wheeler, H.E., Nicolae, D.L., and Im, H.K. (2019). Integrating predicted transcriptome from multiple tissues improves association detection. PLoS Genet. *15*, e1007889.

55. Barbeira, A.N., Dickinson, S.P., Bonazzola, R., Zheng, J., Wheeler, H.E., Torres, J.M., Torstenson, E.S., Shah, K.P., Garcia, T., Edwards, T.L., et al. (2018). Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. Nat. Commun. *9*, 1825.
56. Pividori, M., Rajagopal, P.S., Barbeira, A., Liang, Y., Melia, O., Bastarache, L., Park, Y., Consortium, G., Wen, X., and Im, H.K. (2020). PhenomeXcan: Mapping the genome to the phenome through the transcriptome. Sci. Adv. *6*.
57. Wen, X., Pique-Regi, R., and Luca, F. (2017). Integrating molecular QTL data into genome-wide genetic association analysis: Probabilistic assessment of enrichment and colocalization. PLoS Genet. *13*, e1006646.
58. Mumbach, M.R., Rubin, A.J., Flynn, R.A., Dai, C., Khavari, P.A., Greenleaf, W.J., and Chang, H.Y. (2016). HiChIP: efficient and sensitive analysis of protein-directed genome architecture. Nat. Methods *13*, 919–922.
59. Johnson, D.S., Mortazavi, A., Myers, R.M., and Wold, B. (2007). Genome-wide mapping of in vivo protein-DNA interactions. Science *316*, 1497–1502.
60. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. j. *17*, 10.
61. 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.
62. Bhattacharyya, S., Chandra, V., Vijayanand, P., and Ay, F. (2019). Identification of significant chromatin contacts from HiChIP data by FitHiChIP. Nat. Commun. *10*, 4221.
63. Li, D., Hsu, S., Purushotham, D., Sears, R.L., and Wang, T. (2019). WashU Epigenome Browser update 2019. Nucleic Acids Res. *47*, W158–W165.
64. 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.
65. McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A.R., Teumer, A., Kang, H.M., Fuchsberger, C., Danecek, P., Sharp, K., et al. (2016). A reference panel of 64,976 haplotypes for genotype imputation. Nat. Genet. *48*, 1279–1283.
66. Song, H., Ramus, S.J., Tyrer, J., Bolton, K.L., Gentry-Maharaj, A., Wozniak, E., Anton-Culver, H., Chang-Claude, J., Cramer, D.W., DiCioccio, R., et al. (2009). A genome-wide association study identifies a new ovarian cancer susceptibility locus on 9p22.2. Nat. Genet. *41*, 996–1000.
67. Bolton, K.L., Tyrer, J., Song, H., Ramus, S.J., Notaridou, M., Jones, C., Sher, T., Gentry-Maharaj, A., Wozniak, E., Tsai, Y.-Y., et al. (2010). Common variants at 19p13 are associated with susceptibility to ovarian cancer. Nat. Genet. *42*, 880–884.
68. Goode, E.L., Chenevix-Trench, G., Song, H., Ramus, S.J., Notaridou, M., Lawrenson, K., Widschwendter, M., Vierkant, R.A., Larson, M.C., Kjaer, S.K., et al. (2010). A genome-wide association study identifies susceptibility loci for ovarian cancer at 2q31 and 8q24. Nat. Genet. *42*, 874–879.
69. 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*, 371–384e1-2.
70. Permuth-Wey, J., Lawrenson, K., Shen, H.C., Velkova, A., Tyrer, J.P., Chen, Z., Lin, H.-Y., Chen, Y.A., Tsai, Y.-Y., Qu, X., et al. (2013). Identification and molecular characterization of a new ovarian cancer susceptibility locus at 17q21.31. Nat. Commun. *4*, 1627.
71. Shen, H., Fridley, B.L., Song, H., Lawrenson, K., Cunningham, J.M., Ramus, S.J., Cicek, M.S., Tyrer, J., Stram, D., Larson, M.C., et al. (2013). Epigenetic analysis leads to identification of HNF1B as a subtype-specific susceptibility gene for ovarian cancer. Nat. Commun. *4*, 1628.
72. Kuchenbaecker, K.B., Ramus, S.J., Tyrer, J., Lee, A., Shen, H.C., Beesley, J., Lawrenson, K., McGuffog, L., Healey, S., Lee, J.M., et al. (2015). Identification of six new susceptibility loci for invasive epithelial ovarian cancer. Nat. Genet. *47*, 164–171.
73. Kar, S.P., Beesley, J., Amin Al Olama, A., Michailidou, K., Tyrer, J., Kote-Jarai, Z., 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.
74. Ng, P.C., and Henikoff, S. (2003). SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res. *31*, 3812–3814.
75. Adzhubei, I., Jordan, D.M., and Sunyaev, S.R. (2013). Predicting functional effect of human missense mutations using PolyPhen-2. Curr. Protoc. Hum. Genet. *Chapter 7*, Unit7.20.
76. Rentzsch, P., Witten, D., Cooper, G.M., Shendure, J., and Kircher, M. (2019). CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Res. *47*, D886–D894.
77. Lawrenson, K., Song, F., Hazelett, D.J., Kar, S.P., Tyrer, J., Phelan, C.M., Corona, R.I., Rodríguez-Malavé, N.I., Seo, J.-H., Adler, E., et al. (2019). Genome-wide association studies identify susceptibility loci for epithelial ovarian cancer in east Asian women. Gynecol. Oncol. *153*, 343–355.
78. Mortlock, S., Corona, R.I., Kho, P.F., Pharoah, P., Seo, J.-H., Freedman, M.L., Gayther, S.A., Siedhoff, M.T., Rogers, P.A.W., Leuchter, R., et al. (2022). A multi-level investigation of the genetic relationship between endometriosis and ovarian cancer histotypes. Cell Rep. Med. *3*, 100542.
79. Plummer, J., Dezem, F.S., Chen, S.S., Dhungana, S., Wali, D., Davis, B., Kanska, J., Safi, N., Seo, J.H., Corona, R.I., et al. (2020). 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. Preprint at bioRxiv. https://doi.org/10.1101/2020.09.29.319699.
80. Lopez-Delisle, L., Rabbani, L., Wolff, J., Bhardwaj, V., Backofen, R., Grüning, B., Ramírez, F., and Manke, T. (2021). pyGenomeTracks: reproducible plots for multivariate genomic datasets. Bioinformatics *37*, 422–423.
81. Chen, K., Ma, H., Li, L., Zang, R., Wang, C., Song, F., Shi, T., Yu, D., Yang, M., Xue, W., et al. (2014). Genome-wide association study identifies new susceptibility loci for epithelial ovarian cancer in Han Chinese women. Nat. Commun. *5*, 4682.
82. Couch, F.J., Wang, X., McGuffog, L., Lee, A., Olswold, C., Kuchenbaecker, K.B., Soucy, P., Fredericksen, Z., Barrowdale, D., Dennis, J., et al. (2013). Genome-wide association study in BRCA1 mutation carriers identifies novel loci associated with breast and ovarian cancer risk. PLoS Genet. *9*, e1003212.
83. Corona, R.I., Seo, J.-H., Lin, X., Hazelett, D.J., Reddy, J., Fonseca, M.A.S., Abassi, F., Lin, Y.G., Mhawech-Fauceglia, P.Y., Shah, S.P., et al. (2020). Non-coding somatic mutations converge on the PAX8 pathway in ovarian cancer. Nat. Commun. *11*, 2020.

84. Zhang, Y.D., Hurson, A.N., Zhang, H., Choudhury, P.P., Easton, D.F., Milne, R.L., Simard, J., Hall, P., Michailidou, K., Dennis, J., et al. (2020). Assessment of polygenic architecture and risk prediction based on common variants across fourteen cancers. Nat. Commun. *11*, 3353.
85. Jeng, M.Y., Mumbach, M.R., Granja, J.M., Satpathy, A.T., Chang, H.Y., and Chang, A.L.S. (2019). Enhancer Connectome Nominates Target Genes of Inherited Risk Variants from Inflammatory Skin Disorders. J. Invest. Dermatol. *139*, 605–614.
86. Baca, S.C., Takeda, D.Y., Seo, J.-H., Hwang, J., Ku, S.Y., Arafeh, R., Arnoff, T., Agarwal, S., Bell, C., O'Connor, E., et al. (2021). Reprogramming of the FOXA1 cistrome in treatment-emergent neuroendocrine prostate cancer. Nat. Commun. *12*, 1979.
87. Simeonov, D.R., Gowen, B.G., Boontanrart, M., Roth, T.L., Gagnon, J.D., Mumbach, M.R., Satpathy, A.T., Lee, Y., Bray, N.L., Chan, A.Y., et al. (2017). Discovery of stimulation-responsive immune enhancers with CRISPR activation. Nature *549*, 111–115.
88. Guan, Y., Kuo, W.-L., Stilwell, J.L., Takano, H., Lapuk, A.V., Fridlyand, J., Mao, J.-H., Yu, M., Miller, M.A., Santos, J.L., et al. (2007). Amplification of PVT1 contributes to the pathophysiology of ovarian and breast cancer. Clin. Cancer Res. *13*, 5745–5755.
89. Reyes-González, J.M., and Vivas-Mejía, P.E. (2021). c-MYC and Epithelial Ovarian Cancer. Front. Oncol. *11*, 601512.
90. Duffy, M.J., O'Grady, S., Tang, M., and Crown, J. (2021). MYC as a target for cancer treatment. Cancer Treat Rev. *94*, 102154.
91. Stacey, D., Fauman, E.B., Ziemek, D., Sun, B.B., Harshfield, E.L., Wood, A.M., Butterworth, A.S., Suhre, K., and Paul, D.S. (2019). ProGeM: a framework for the prioritization of candidate causal genes at molecular quantitative trait loci. Nucleic Acids Res. *47*, e3.
92. Baca, S.C., Singler, C., Zacharia, S., Seo, J.-H., Morova, T., Hach, F., Ding, Y., Schwarz, T., Huang, C.-C.F., Anderson, J., et al. (2022). Genetic determinants of chromatin reveal prostate cancer risk mediated by context-dependent gene regulation. Nat. Genet. *54*, 1364–1375.
93. Oliva, M., Demanelis, K., Lu, Y., Chernoff, M., Jasmine, F., Ahsan, H., Kibriya, M.G., Chen, L.S., and Pierce, B.L. (2023). DNA methylation QTL mapping across diverse human tissues provides molecular links between genetic variation and complex traits. Nat. Genet. *55*, 112–122.
94. Spisák, S., Lawrenson, K., Fu, Y., Csabai, I., Cottman, R.T., Seo, J.-H., Haiman, C., Han, Y., Lenci, R., Li, Q., et al. (2015). CAUSEL: an epigenome- and genome-editing pipeline for establishing function of noncoding GWAS variants. Nat. Med. *21*, 1357–1363.
95. Zhang, M.J., Hou, K., Dey, K.K., Sakaue, S., Jagadeesh, K.A., Weinand, K., Taychameekiatchai, A., Rao, P., Pisco, A.O., Zou, J., et al. (2022). Polygenic enrichment distinguishes disease associations of individual cells in single-cell RNA-seq data. Nat. Genet. *54*, 1572–1580.
96. Chen, Z., Wen, W., Beeghly-Fadiel, A., Shu, X.-O., Díez-Obrero, V., Long, J., Bao, J., Wang, J., Liu, Q., Cai, Q., et al. (2019). Identifying Putative Susceptibility Genes and Evaluating Their Associations with Somatic Mutations in Human Cancers. Am. J. Hum. Genet. *105*, 477–492.