Epithelial-Mesenchymal Transition (EMT) Gene Variants and Epithelial Ovarian Cancer (EOC) Risk

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

Citation

Published Version
10.1002/gepi.21921

Citable link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:33840776

Terms of Use
This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Open Access Policy Articles, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#OAP
Epithelial-Mesenchymal Transition (EMT) gene variants and Epithelial Ovarian Cancer (EOC) risk

A full list of authors and affiliations appears at the end of the article.

Abstract

Introduction—Epithelial-mesenchymal transition (EMT) is a process whereby epithelial cells assume mesenchymal characteristics to facilitate cancer metastasis. However, EMT also contributes to the initiation and development of primary tumors. Prior studies that explored the hypothesis that EMT gene variants contribute to EOC risk have been based on small sample sizes and none have sought replication in an independent population.

Methods—We screened 1254 SNPs in 296 genes in a discovery phase using data from a genome-wide association study of EOC among women of European ancestry (1,947 cases and 2,009 controls) and identified 793 variants in 278 EMT-related genes that were nominally \( (p<0.05) \) associated with invasive EOC. These SNPs were then genotyped in a larger study of 14,525 invasive-cancer patients and 23,447 controls. A \( p \)-value \(<0.05\) and a false discovery rate (FDR) \(<0.2\) was considered statistically significant.

Results—In the larger dataset, \( GPC6/GPC5 \) rs17702471 was associated with the endometrioid subtype among Caucasians (OR=1.16, 95%CI=1.07–1.25, \( p=0.0003\), FDR=0.19), while \( F8 \) rs7053448 (OR=1.69, 95%CI=1.27–2.24, \( p=0.0003\), FDR=0.12), \( F8 \) rs7058826 (OR=1.69, 95%CI=1.27–2.24, \( p=0.0003\), FDR=0.12), and \( CAPN13 \) rs1983383 (OR=0.79, 95%CI=0.69–0.90, \( p=0.0005\), FDR=0.12) were associated with combined invasive EOC among Asians. In silico functional analyses revealed that \( GPC6/GPC5 \) rs17702471 coincided with DNA regulatory elements.

Conclusion—These results suggest that EMT gene variants do not appear to play a significant role in the susceptibility to EOC.

Keywords

ovarian cancer; epithelial-mesenchymal transition; single nucleotide polymorphisms

INTRODUCTION

Epithelial ovarian carcinoma (EOC) is the leading cause of gynecological cancer-related mortality in the western world. In the United States, 22,000 cases and 14,300 deaths are estimated in 2014 (American Cancer Society 2014). A family history of the disease is associated with a 2–3 fold increased risk of developing the disease (Auranen et al. 1996),
implicating inherited genetic factors (Lichtenstein et al. 2000, Stratton et al. 1998). Germline mutations in highly penetrant genes such as BRCA1 and BRCA2 are estimated to account for about 40% of the excess familial risk (Narod and Foulkes 2004). Given little evidence for additional highly penetrant genes, the prevailing consensus is that the remaining familial risk reflects common susceptibility alleles that confer low risk. Indeed, genome-wide association studies (GWAS) and large-scale replication studies have identified some common susceptibility alleles that collectively account for another 4% of the excess familial risk (Pharoah et al. 2013). Thus, more variants remain to be identified.

Epithelial-mesenchymal transition (EMT) is a fundamental embryonic process whereby epithelial cells assume mesenchymal characteristics to facilitate migration through the extracellular matrix (ECM) into the interior of the embryo to contribute to the development of internal organs (Davidson, Trope and Reich 2012, Nieto 2013, Thiery et al. 2009). EMT is a complex process, accompanied by changes in the expression of multiple and diverse factors, including cell adhesion molecules, growth factors, metalloproteinases, transcription factors and epithelial markers (Thiery et al. 2009). A role for EMT in cancer has mostly been limited to cancer cell invasion and metastatic spread, where malignant cells lose their epithelial characteristics and assume mesenchymal properties that promote ECM invasion and metastasis (Thiery et al. 2009). However, several lines of evidence support the hypothesis that EMT contributes to the initiation and development of primary tumors: (i) the expression of inducers of EMT in non-invasive neoplastic lesions, (ii) drivers of EMT attenuate key tumor suppressive mechanisms in epithelial cells by affecting the functions of modulators of p53 and retinoblastoma dependent pathway, and (iii) the involvement of EMT in cell differentiation and tissue homeostasis (Nieto 2013, Puisieux, Brabletz and Caramel 2014). It is plausible that variants in EMT genes may alter gene expression and therefore contribute to inter-individual variation in the risk of ovarian cancer.

Previous studies of variants in EMT-related genes, such as vascular endothelial growth factor (VEGF) (Polterauer et al. 2007, Schultheis et al. 2008, Goode et al. 2010, Steffensen et al. 2010), matrix-metalloprotease genes (MMPs) (Schildkraut et al. 2009b), TP53 (Schildkraut et al. 2009a, Galic et al. 2007), E-cadherin (Li et al. 2008), nitric oxide synthase (NOS) (Heffler et al. 2002), heparanase (HPN) (Ralph et al. 2007) and PIK3CA (Quaye et al. 2009, Quaye et al. 2008) provide support for the hypothesis. However, the majority of the studies were based on small sample sizes with no replications. Furthermore, variation in association by histological subtype and race was not established. In this study, we selected 793 SNPs from 278 EMT-related genes based on nominal associations with EOC in a genome-wide association study among women of European ancestry and performed a replication study involving 14,525 case subjects with invasive disease and 23,447 controls from 43 sites in the Ovarian Cancer Association Consortium (OCAC), as part of the Collaborative Oncological Gene-environment Study (COGS). This large sample size of the replication study also provided an opportunity to explore associations by histological subtypes of EOC and by race.
MATERIALS AND METHODS

Discovery set

We initially selected SNPs based on screening of the results from the North American genome-wide association study (GWAS). Details of this study have been published elsewhere (Permuth-Wey et al. 2011). Briefly, five case-control studies contributed a total of 1,947 EOC cases and 2,009 controls (of European ancestry) to the North American GWAS. Genotyping for four of the studies was performed using the Illumina 610-quad Beadchip Array, while the fifth study was genotyped using the Illumina 317K and 370K arrays. Additional markers were imputed to account for non-overlapping markers in the two genotyping efforts and to improve genome coverage. A total of 2,508,744 (including the 1254 EMT-related SNPs) out of 2,543,887 SNPs (98.6%) passed quality control.

Collaborative Oncological Gene-environment Study

SNPs that were nominally (P<0.05) associated with ovarian cancer in the discovery set were genotyped in the Collaborative Oncological Gene-environment Study (COGS) that included 43 individual studies from OCAC. Details of this study have been published elsewhere (Pharoah et al. 2013). Briefly, 34 case-control studies contributed 14,525 EOC cases and 23,447 controls of white-European ancestry, 150 cases and 200 controls of African ancestry, and 714 cases and 1574 controls of Asian ancestry. Ancestry was assigned using Local Ancestry in adMixed Populations (Sankararaman, 2008) based on genotype frequency for intercontinental ancestry. Participants with >90% white-European ancestry were defined as European and those with >80% African or Asian ancestry were defined as African or Asian, respectively. Population substructure within each ancestry group was determined by principal components analyses using a set of 37,000 uncorrelated markers. The first five principal components were employed for analyses of Europeans and Asians and the first one for Africans. The COGS dataset was used to perform SNP-specific analyses for each of the four main histological subtypes (serous, endometroid, clear cell and mucinous), and each race.

Independent Replication set

The COGS dataset included 62% (1,207 cases and 1,246 controls) of the samples from women of white-European ancestry that were genotyped in the discovery set. Therefore we excluded all these samples from the COGS dataset to create an independent replication set. This independent replication set was used to replicate SNPs identified for all invasive EOC among women of European ancestry in the discovery dataset.

SNP selection, genotyping and quality control

EMT-related genes were identified through review of the published literature (www.pubmed.gov) and pathways described in the Cancer Genome Anatomy Project. Genotype data from HapMap, Perlegen and the NIEHS resequencing projects were then used to select tagSNPs (r^2 > 0.8), non-synonymous coding SNPs and putative functional SNPs from unrelated Caucasian samples. A total of 15,816 SNPs in 296 EMT genes had results available in the discovery dataset. The final selection of EMT-related SNPs for
genotyping in the large COGS dataset was informed by data from the discovery set and design scores from Illumina as previously described (6). SNPs were ranked based on P-value in the discovery set and the 793 SNPs associated with risk of combined invasive EOC or serous ovarian cancer at a nominal P < 0.05 were selected (Supplementary Table 1). All 793 SNPs were successfully designed for genotyping and included as part of the “candidate SNPs” on the custom Illumina Infinium iSelect array that had a total of 211,155 SNPs designed for COGS (Pharoah et al. 2013).

Genotyping was performed using an Illumina Infinium iSelect BeadChip at McGill University-Génom Québec Innovation Centre and the Mayo Clinic Medical Genome Facility following established quality control criteria (Pharoah et al. 2013). The concordance for 1,251 duplicate samples was 99.6%. Out of the 211,155 SNPs included on the array, 94.5% passed quality control (QC). All 793 SNPs in the present study were genotyped successfully and passed QC.

**In silico analysis using publicly available datasets**

We used publicly available data to test for evidence of functional effects of the SNPs using Haploreg (http://www.broadinstitute.org/mammals/haploreg/haploreg.php), ENCODE (http://genome.ucsc.edu/ENCODE/) and ovary super-enhancer datasets (Ward & Kellis 2012, Hnisz et al. 2013). These datasets were used to identify epigenetic marks coinciding with risk associated SNPs.

**Functional Analyses**

An in vitro model of early-stage ovarian cancer has been previously described (Lawrenson et al. 2011). Briefly, Illumina HT12 gene expression microarrays were used to profile the transcriptome of 3D models of normal ovarian cells immortalized with TERT and overexpressing cMYC and a mutant KRAS or BRAF allele.

**Statistical methods**

Participants’ demographic and clinical characteristics by cancer status were compared using t-tests for continuous variables and chi-square tests for categorical variables. Statistical analysis methods for the discovery dataset have been described previously (Permuth-Wey et al. 2011). Analyses of the independent replication set and the COGS dataset were performed as follows. Associations between each SNP and case-control status were estimated as log-linear per allele odds ratios (ORs) with associated 95% confidence intervals (95% CIs) using unconditional logistic regression that treated the number of alleles of interest as ordinal variables. The ORs specific for each histological subtype was estimated by comparing cases of that subtype to all available controls as reference.

All models included variables for study site and for the first five (for Europeans and Asians) or one (for Africans) ancestry-specific principal components as covariates. False discovery rates (FDR) were calculated to control for multiple testing. Statistical tests were two-sided with an alpha level < 0.05 and an FDR < 0.2 considered statistically significant. All statistical analyses were implemented with SAS/Genetics version 9.2 (SAS Institute, NC).
RESULTS

A detailed description of the OCAC studies and study participants in the discovery (Permuth-Wey et al. 2011) and COGS (Pharoah et al. 2013) datasets are provided elsewhere. In the large COGS dataset, the majority of women (93.5%, n=37,972) were of European ancestry; only 0.9% (n=350) and 5.6% (n=2288) were of African or Asian ancestry, respectively (Table 1). As expected, the proportion of serous histological subtype was higher than the proportions for other subtypes for all race groups.

In the first step, we screened 15,816 SNPs in 296 EMT-related candidate genes in the GWAS data set, which consisted of women of European ancestry only. A total of 793 SNPs in 278 genes were nominally (p<0.05) associated with the risk of all invasive EOC. Seven of the 793 SNPs showed a nominal consistent (in direction) association with risk of invasive EOC among women of European ancestry in an independent replication dataset (Table 2). However, none of the associations for these SNPs or additional SNPs in the corresponding genes remained statistically significant after controlling for multiple testing (Table 2).

Previous studies have revealed that genetic associations with EOC can differ, sometimes in the opposite direction, by histology. Since this effect can mask associations, we performed analyses for the four main histological subtypes, separately in the COGs dataset only because the discovery set did not have information on all the histological subtypes (Supplementary Table 1). This analysis revealed a statistically significant association, after controlling for multiple testing, at glypican 6/glypican 5 (GPC6/GPC5) rs17702471 (OR=1.16, 95%CI=1.07–1.25, p=0.0003, FDR=0.19) for the endometroid subtype among Caucasians (Table 3).

Analyses for Asians and African Americans separately in the COGs dataset because the discovery set did not have information on these two races. We observed associations at F8 rs7053448 (OR=1.69, 95%CI=1.27–2.24, p=0.0003, FDR=0.12), F8 rs7058826 (OR=1.69, 95%CI=1.27–2.24, p=0.0003, FDR=0.12), and CAPN13 rs1983383 (OR=0.79, 95%CI=0.69–0.90, p=0.0005, FDR=0.12) for combined invasive EOC among Asians (Table 4). However, F8 rs7053448 and rs7058826 were highly correlated (r²=0.7). No statistical significant associations were observed for women of African ancestry.

In silico functional analysis of variants and candidate genes

We used publicly available data, including Haploreg, ENCODE and ovary super-enhancer datasets, to test for evidence of functionality of the four SNPs (GPC6/GPC5 rs17702471, F8 rs7053448, F8 rs7058826 and CAPN13 rs1983383) that were significant after controlling for multiple testing. GPC6/GPC5 rs17702471 coincided with functional elements in the datasets (Table 5).

Although none of the seven SNPs (NRP2 rs3771044, HEG1/MUC13 rs10460829, GHSP/TNFSF10 rs2100143, SEMA3C/LOC100128317 rs10954593, SHH rs172310, SEMA4B rs8030039 and IGF1R rs10794486) that showed consistent association between the discovery and independent replication datasets remained statistically significant after multiple testing, we evaluated their functionality because of their consistent associations.
Functional analyses indicated a strong evidence of regulatory activity at the NRP2 rs3771044 polymorphism, which coincided with DNase I hypersensitivity peaks and transcription factor binding sites in 62 different cell lines (Hnisz et al., 2013), plus the footprint of a super-enhancer (Figure 1), active in ovarian tissue and known to target NRP2. NRP2 (neuropilin 2) is a transmembrane glycoprotein that interacts with TGF-β1 signaling to promote cancer progression (Grandclement et al. 2011).

**DISCUSSION**

The current study sought to test the hypothesis that genetic variants in the EMT process influence risk of EOC. Seven variants showed nominal consistent associations with EOC in the discovery and replication sets among women of European ancestry, but none of them reached the threshold for statistical significance after controlling for multiple testing in the combined COGS dataset. However, four SNPs showed significant associations, after controlling for multiple testing, in the endometrioid subtype among Caucasians (GPC6/GPC5 rs17702471), and for combined invasive EOC among Asians (F8 rs7053448, F8 rs7058826 and CAPN13 rs1983383). Rs17702471, an intergenic SNP located between GPC5 and GPC6 was associated with the endometrioid subtype among women of European-ancestry. Glypicans (GPC) are a family of heparan sulphate proteoglycans (HSPGs) that are attached to the plasma membrane and regulate cell proliferation and division (De Cat and David 2001, Filmus and Selleck 2001) and have been previously shown to be involved in the development and metastasis of various types of human cancer ((Yang et al. 2013, Zhang et al. 2011)). Variants in GPC5/GPC6 have been implicated in lung (Li et al. 2010, Liu et al. 2014), but not ovarian cancer. Rs17702471 coincides with potential regulatory elements and further studies evaluating its role in the endometroid histological subtype of ovarian cancer are warranted.

F8 rs7053448, F8 rs7058826 and CAPN13 rs1983383 were associated with all invasive EOC in Asians. Coagulation factor VIII (F8) participates in the intrinsic pathway of blood coagulation. A direct role for F8 has not been established for cancer. Calpain (CAPN) 13 is part of a family of cytosolic calcium-activated proteases involved in apoptosis, cell division and modulation of integrin-cytoskeletal interaction. Although the associations for these SNPs met our threshold for statistical significance, the analyses were exploratory and also based on a small sample size. Thus future studies are warranted to confirm these initial findings.

Although statistical evidence did not support SNPs that showed consistent nominal associations in the discovery and the independent replication datasets, functional data for NRP2 rs3771044 may suggest a role in ovarian cancer. This SNP coincides with DNase I hypersensitivity peaks and transcription factor binding sites and a super-enhancer active in ovarian tissues (Hnisz et al. 2013). Analysis of the expression of NRP2 using TCGA data revealed a significantly higher expression in high-grade serous tumors compared to normal control tissues (data not shown). Neurophilins are transmembrane glycoproteins that are involved in several signaling pathways leading to cytoskeletal organization, angiogenesis and cancer progression (Prahst et al. 2008, Soker et al. 2002, Sulpice et al. 2008). A role for
NRP2 in ovarian cancer has not been reported previously, thus warranting further evaluation of its role in ovarian cancer risk.

Our study is the largest one to date to evaluate EMT-related gene variants in ovarian cancer. The study also benefits from other strengths including discovery and replication phases, control for population substructure within each ethnicity, biological plausibility of selected genes, evaluation of a comprehensive list of SNPs in EMT-related genes, evaluation of histological subtypes and race (albeit exploratory) and the tight quality control on the genotype data. A limitation of our study is that the analysis in the other ethnic/racial groups was based on small numbers and the associations may not be reliable. Future studies with larger numbers of African and Asians are therefore warranted.

The results suggest that EMT gene variants do not appear to play a significant role in the susceptibility to EOC.

Supplementary Material
Refer to Web version on PubMed Central for supplementary material.

Authors

Genet Epidemiol. Author manuscript; available in PMC 2016 January 21.
Affiliations

1Department of Cancer Epidemiology, Moffitt Cancer Center, Tampa, FL, USA
2Clinical and Translational Research Organization, All Children’s Hospital Johns Hopkins Medicine, St Petersburg, FL
3Department of Biostatistics and Bioinformatics, Moffitt Cancer Center, Tampa, FL, USA
4Department of Public Health and Primary Care, The Centre for Cancer Epidemiology, University of Cambridge, Strangeways Research Laboratory, Cambridge, UK
5Department of Preventive Medicine, Keck School of Medicine, University of Southern California Norris Comprehensive Cancer Center, Los Angeles, CA, USA
6Department for Health Evidence, Radboud University Medical Center, Radboud Institute for Health Sciences, Nijmegen, The Netherlands
7Comprehensive Cancer Center The Netherlands, Nijmegen, The Netherlands
8Genetic Epidemiology Research Institute, UCI Center for Cancer Genetics Research and Prevention, School of Medicine, Department of Epidemiology, University of California Irvine, Irvine, CA, USA
9Byelorussian Institute for Oncology and Medical Radiology Aleksandrov N.N., Minsk, Belarus
10Cancer Epidemiology Centre, Cancer Council Victoria, Melbourne, Australia
11Cancer Prevention and Control, Rutgers Cancer Institute of New Jersey, New Brunswick, NJ, USA
12Department of Obstetrics & Gynecology, Oregon Health & Science University, Portland, OR, USA
13Knight Cancer Institute, Oregon Health & Science University, Portland, OR, USA
14Department of Gynecology and Obstetrics, University Hospital Erlangen, Friedrich-Alexander-University, Erlangen-Nuremberg Comprehensive Cancer Center, Erlangen EMN, Germany
15Gynecology Service, Department of Surgery, Memorial Sloan-Kettering Cancer Center, New York, NY, USA
16Department of Gynecology and Obstetrics, Haukeland University Hospital, Bergen, Norway
17Centre for Cancer Biomarkers, Department of Clinical Medicine, University of Bergen, Bergen, Norway
18Radiation Oncology Research Unit, Hannover Medical School, Hannover, Germany
19Division of Cancer
and Oncology, Los Angeles, CA, USA 49Biostatistics and Informatics Shared Resource, University of Kansas Medical Center, Kansas City, KS, USA 50Department of Epidemiology, Shanghai Cancer Institute, Shanghai, China 51Women's Cancer, UCL EGA Institute for Women's Health, London, UK 52Centre for Epidemiology and Biostatistics, School of Population and Global Health, The University of Melbourne, Melbourne, Australia 53Cancer Prevention and Control, Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, CA, USA 54Community and Population Health Research Institute, Department of Biomedical Sciences, Cedars-Sinai Medical Center, Los Angeles, CA, USA 55Cancer Research Initiatives Foundation, Sime Darby Medical Center, Subang Jaya, Malaysia 56Department of Epidemiology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA 57Department of Gynaecology, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark 58Department of Virus, Lifestyle and Genes, Danish Cancer Society Research Center, Copenhagen, Denmark 59Molecular Unit, Department of Pathology, Herlev Hospital, University of Copenhagen, Copenhagen, Denmark 60Division of Epidemiology and Prevention, Aichi Cancer Center Research Institute, Nagoya, Aichi, Japan 61Department of Statistics, Duke University, Durham, NC, USA 62Women's Cancer Program at the Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, CA, USA 63Department of Health Outcomes and Behavior, Moffitt Cancer Center, Tampa, FL, USA 64Vesalius Research Center, VIB, University of Leuven, Leuven, Belgium 65Laboratory for Translational Genetics, Department of Oncology, University of Leuven, Belgium 66Cancer Control Research, BC Cancer Agency, Vancouver, BC, Canada 67Department of Cancer Prevention and Control, Roswell Park Cancer Institute, Buffalo, NY, USA 68College of Pharmacy and Health Sciences, Texas Southern University, Houston, TX, USA 69Department of Obstetrics and Gynaecology, University Malaya Medical Centre, University Malaya, Kuala Lumpur, Malaysia 70Department of Cancer Epidemiology and Prevention, M. Sklodowska-Curie Memorial Cancer Center and Institute of Oncology, Warsaw, Poland 71Department of Gynecologic Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA 72Radboud University Medical Center, Radboud Institute for Molecular Life Sciences, Nijmegen, The Netherlands 73Department of Health Research and Policy - Epidemiology, Stanford University School of Medicine, Stanford, CA, USA 74Public Health Ontario, Toronto, ON, Canada 75Women's Cancer Research Program, Magee-Women's Research Institute and University of Pittsburgh Cancer Institute, Pittsburgh, PA, USA 76Department of Obstetrics, Gynecology and Reproductive Sciences, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA 77Department of Pathology, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark 78The University of Texas School of Public Health, Houston, TX, USA 79Department of Gynecologic Oncology, Roswell Park Cancer Institute, Buffalo, NY 80Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, New York, NY, USA 81Department of Community and Family Medicine, Duke University Medical Center, Durham, NC, USA 82Department of Epidemiology, University of Michigan, 1415
Acknowledgments

Individual acknowledgements by study:

We thank all the individuals who took part in this study and all the researchers, clinicians and technical and administrative staff who have made possible the many studies contributing to this work. In particular, we thank: D. Bowtell, A. deFazio, D. Gertig, A. Green, P. Parsons, N. Hayward, P. Webb and D. Whiteman (AUS); G. Peuteman, T. Van Brussel and D. Smeets (BEL); the staff of the genotyping unit, S LaBoissiere and F Robidoux (Genome Quebec); U. Eilber (GER); L. Gacucova (HMO); P. Schurmann, F. Kramer, W. Zheng, T. W. Park, Simon, K. Beer-Grondke and D. Schmidt (HJO); S. Windebank, C. Hilker and J. Vollenweider (MAY); the state cancer registries of AL, AZ, AR, CA, CO, CT, DE, FL, GA, HI, ID, IL, IN, IA, KY, LA, ME, MD, MA, MI, NE, NH, NJ, NY, NC, ND, OH, OK, OR, PA, RI, SC, TN, TX, VA, WA, and WYL (NHS); L. Paddock, M. King, L. Rodriguez-Rodriguez, A. Samoila, and Y. Bensman (NJO); M. Sherman, A. Hutchinson, N. Szczesniak—Dabrowska, B. Peplonska, W. Zatonski, A. Soni, P. Chao and M. Stagner (POL); C. Luccarini, P. Harrington the SEARCH team and ECRIC (SEA); R. Royer, S. Zhang (TOR), I. Jacobs, M. Widschwendter, E. Wozniak, N. Balogun, A. Ryan and J. Ford (UKO); Carole Pye (UKR); A. Amin Al Olama, K. Michlайдou, R. Kuchenbaker (COGS).

Main funding:

The scientific development and funding for this project were funded by the following: NIH R01 CA-1491429 (Phelan PI); the US National Cancer Institute (R01-CA076016); the COGS project is funded through a European Commission’s Seventh Framework Programme grant (agreement number 223175 HEALTH F2 2009-223175); the Genetic Associations and Mechanisms in Oncology (GAME-ON): a NCI Cancer Post-GWAS Initiative (U19-
CA148112); the Ovarian Cancer Association Consortium is supported by a grant from the Ovarian Cancer Research Fund thanks to donations by the family and friends of Kathryn Sladek Smith (PPD/RPCI.07).

Investigator-specific funding:
E.K.A was supported by a Cancer prevention fellowship from the National Cancer Institute (R25T CA147832) during the conduct of this study. K.L. is supported by a K99/R00 grant from the National Cancer Institute (Grant number 1K99CA184415-01). C.-T. is supported by the National Health and Medical Research Council; B.K. holds an ACS Early Detection Professorship (SIOP—06—258—01—COUN); L.E.K. is supported by a Canadian Institute of Health Research New Investigator Award (MSH—87734). AWL is supported by NIEHS T32 training grant (T32ES013678).

Funding of included studies:
Funding of the constituent studies was provided by the California Cancer Research Program (00—01389V—20170, N01—CN25403, 2II020); the Canadian Institutes of Health Research (MOP—86727); Cancer Australia; Cancer Council Victoria; Cancer Council Queensland; Cancer Council New South Wales; Cancer Council South Australia; Cancer Council Tasmania; Cancer Foundation of Western Australia; the Cancer Institute of New Jersey; Cancer Research UK (C490/A6187, C490/A10119, C490/A10124); the Danish Cancer Society (94—222—52); the ELAN Program of the University of Erlangen—Nuremberg; the Eve Appeal; the Helsinki University Central Hospital Research Fund; Helse Vest; the Norwegian Cancer Society; the Norwegian Research Council; the Ovarian Cancer Research Fund; National Kankerplan of Belgium; Grant—in—Aid for the Third Term Comprehensive 10—Year Strategy For Cancer Control from the Ministry of Health Labour and Welfare of Japan; the L & S Milken Foundation; the Polish Ministry of Science and Higher Education (4 POSC 02 14, 2 PO SA 068 27); the Roswell Park Cancer Institute Alliance Foundation; the US National Cancer Institute (K07—CA095666, K07—CA143047, K22—CA138563, N01—CN55424, N01—PC67001, N01—PC035137, R01—CA017054, R01—CA087696, P00—CA072720, P50—CA105099, P50—CA136393, R01—CA014089, R01—CA016565); the US Public Health Service (PSA—042205); The National Health and Medical Research Council of Australia (199600 and 400281); the German Federal Ministry of Education and Research of Germany Programme of Biomedical Research (01GB 9401); the State of Baden—Wurttemberg through Medical Faculty of the University of Ulm (P685); the Minnesota Ovarian Cancer Alliance; the Mayo Foundation; the Fred C. and Katherine B. Andersen Foundation; the Lon V. Smith Foundation (LVS—39420); the Oak Foundation; the OHSU Foundation; the Mermaid I project; the Rudolf—Bartling Foundation; the UK National Institute for Health Research Biomedical Research Centres at the University of Cambridge, Imperial College London, University College Hospital “Womens Health Theme” and the Royal Marsden Hospital; WorkSafeBC 14.

References


Figure 1. Annotation of regulatory elements at rs3771044
This variant lies within a super-enhancer detected in ovarian tissues (data from Hsinz et al). Using data from ENCODE Regulation tracks, we identified evidence of cell type specific H3K4me1 marks. This SNP also coincides with ChIPseq marks for multiple transactions factors (TFs) including EP300, NFKB and RUNX3. This variant lies with a DNasel hypersensitive region detected in 109 samples.
Table 1

Demographic and clinical characteristics of study participants in the COGS replication study

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>European descent</th>
<th>Asian descent</th>
<th>African descent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Controls (n=23,447) N (%)</td>
<td>Invasive Cases (n=14,525) N (%)</td>
<td>p-value²</td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean ± SD</td>
<td>55.6± 11.9</td>
<td>58.1± 11.3</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>&lt;40</td>
<td>2027 (8.7)</td>
<td>748 (5.2)</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>40–49</td>
<td>4771 (20.6)</td>
<td>2544 (17.6)</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>50–59</td>
<td>7403 (31.9)</td>
<td>4537 (31.3)</td>
<td></td>
</tr>
<tr>
<td>60–69</td>
<td>6098 (26.3)</td>
<td>4324 (29.8)</td>
<td></td>
</tr>
<tr>
<td>≥70</td>
<td>2892 (12.5)</td>
<td>2343 (16.2)</td>
<td></td>
</tr>
<tr>
<td>Family history of ovarian cancer</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>15425 (92.0)</td>
<td>8634 (82.4)</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Yes</td>
<td>1351 (8.0)</td>
<td>1849 (17.6)</td>
<td></td>
</tr>
<tr>
<td>Histological subtypes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Serous</td>
<td>N/A</td>
<td>8368 (57.6)</td>
<td>N/A</td>
</tr>
<tr>
<td>Endometroid</td>
<td>2067 (14.2)</td>
<td>112 (15.7)</td>
<td>16 (10.7)</td>
</tr>
<tr>
<td>Clear Cell</td>
<td>1024 (7.1)</td>
<td>103 (14.4)</td>
<td>6 (4.0)</td>
</tr>
<tr>
<td>Mucinous</td>
<td>944 (6.5)</td>
<td>60 (8.4)</td>
<td>11 (7.3)</td>
</tr>
<tr>
<td>Others²</td>
<td>2122 (14.6)</td>
<td>190 (26.6)</td>
<td>28 (18.7)</td>
</tr>
</tbody>
</table>

¹ for the first degree relatives

² t-test for a continuous variable and chi-square test for a categorical variable

³ include mixed cell, other specified epithelial, undifferentiated, unknown (but known to be epithelial)
## Table 2

EMT SNPs significantly associated with invasive ovarian cancer risk among Caucasians

<table>
<thead>
<tr>
<th>SNP</th>
<th>Nearest gene</th>
<th>Location</th>
<th>Chr</th>
<th>Minor allele</th>
<th>MAF</th>
<th>P</th>
<th>OR(95% CI)</th>
<th>MAF</th>
<th>P</th>
<th>OR(95% CI)</th>
<th>MAF</th>
<th>P</th>
<th>OR(95% CI)</th>
<th>FDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs3771044</td>
<td>NRP2</td>
<td>intron</td>
<td>2</td>
<td>G</td>
<td>0.13</td>
<td>0.022</td>
<td>1.17 (1.02–1.34)</td>
<td>0.11</td>
<td>0.014</td>
<td>1.06 (1.01–1.12)</td>
<td>0.12</td>
<td>0.002</td>
<td>1.08 (1.03–1.13)</td>
<td>0.5</td>
</tr>
<tr>
<td>rs10460829</td>
<td>HEG1/MUC13</td>
<td>Intergenic</td>
<td>3</td>
<td>A</td>
<td>0.21</td>
<td>0.015</td>
<td>1.15 (1.03–1.29)</td>
<td>0.21</td>
<td>0.023</td>
<td>1.05 (1.01–1.09)</td>
<td>0.21</td>
<td>0.005</td>
<td>1.06 (1.02–1.10)</td>
<td>0.55</td>
</tr>
<tr>
<td>rs2100143</td>
<td>GHSR/TNFSF10</td>
<td>Intergenic</td>
<td>3</td>
<td>G</td>
<td>0.48</td>
<td>0.028</td>
<td>0.9 (0.82–0.99)</td>
<td>0.49</td>
<td>0.044</td>
<td>0.97 (0.94–1.00)</td>
<td>0.48</td>
<td>0.013</td>
<td>0.96 (0.93–0.99)</td>
<td>0.55</td>
</tr>
<tr>
<td>rs10954593</td>
<td>SEMA3C/LOC100128317</td>
<td>Intergenic</td>
<td>7</td>
<td>G</td>
<td>0.19</td>
<td>0.033</td>
<td>1.14 (1.01–1.27)</td>
<td>0.19</td>
<td>0.034</td>
<td>1.05 (1.00–1.09)</td>
<td>0.19</td>
<td>0.011</td>
<td>1.05 (1.01–1.09)</td>
<td>0.55</td>
</tr>
<tr>
<td>rs172310</td>
<td>SHH</td>
<td>flanking_5</td>
<td>7</td>
<td>A</td>
<td>0.31</td>
<td>0.039</td>
<td>1.11 (1.01–1.23)</td>
<td>0.30</td>
<td>0.01</td>
<td>1.05 (1.01–1.08)</td>
<td>0.30</td>
<td>0.002</td>
<td>1.05 (1.02–1.09)</td>
<td>0.55</td>
</tr>
<tr>
<td>rs8030039</td>
<td>SEMA4B</td>
<td>intron</td>
<td>15</td>
<td>A</td>
<td>0.35</td>
<td>0.019</td>
<td>1.12 (1.02–1.23)</td>
<td>0.35</td>
<td>0.033</td>
<td>1.04 (1.00–1.07)</td>
<td>0.35</td>
<td>0.021</td>
<td>1.04 (1.01–1.07)</td>
<td>0.55</td>
</tr>
<tr>
<td>rs10794486</td>
<td>IGF1R</td>
<td>intron</td>
<td>15</td>
<td>G</td>
<td>0.3</td>
<td>0.043</td>
<td>1.11 (1.00–1.22)</td>
<td>0.30</td>
<td>0.004</td>
<td>1.05 (1.02–1.09)</td>
<td>0.30</td>
<td>0.002</td>
<td>1.06 (1.02–1.09)</td>
<td>0.5</td>
</tr>
</tbody>
</table>

Abbreviations: Chr=chromosome; MAF=minor allele frequency; OR=odds ratio; CI=confidence interval.

\( ^{a} \)Includes the replication dataset and 62% of the discovery dataset
Table 3

Association between \textit{GPC6/GPC5} rs17702471 and histological subtypes of epithelial ovarian cancer among Caucasians

<table>
<thead>
<tr>
<th>Histological subtype</th>
<th>P</th>
<th>OR(95% CI)</th>
<th>FDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Endometroid</td>
<td>0.00026</td>
<td>1.16 (1.07–1.25)</td>
<td>0.19</td>
</tr>
<tr>
<td>Clear Cell</td>
<td>0.013</td>
<td>1.15 (1.03–1.28)</td>
<td>0.80</td>
</tr>
<tr>
<td>Mucinous</td>
<td>0.13</td>
<td>1.09 (0.98–1.22)</td>
<td>0.94</td>
</tr>
<tr>
<td>Serous</td>
<td>0.21</td>
<td>1.03 (0.98–1.08)</td>
<td>0.77</td>
</tr>
<tr>
<td>All invasive</td>
<td>0.019</td>
<td>1.05 (1.01–1.09)</td>
<td>0.55</td>
</tr>
</tbody>
</table>
### Table 4

Associations between \textit{F8} rs7053448 and rs7058826 and \textit{CAPN13} rs1983383 and all invasive epithelial ovarian cancer by race

<table>
<thead>
<tr>
<th>SNP</th>
<th>Nearest gene</th>
<th>Chr</th>
<th>Race</th>
<th>Minor allele</th>
<th>MAF</th>
<th>P</th>
<th>OR (95% CI)</th>
<th>FDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs7053448</td>
<td>\textit{F8}</td>
<td>23</td>
<td>Asian</td>
<td>G</td>
<td>0.06</td>
<td>0.00034</td>
<td>1.69 (1.27–2.24)</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>African American</td>
<td>G</td>
<td>0.13</td>
<td>0.91</td>
<td>0.98 (0.63–1.52)</td>
<td>0.97</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Caucasian</td>
<td>G</td>
<td>0.18</td>
<td>0.45</td>
<td>0.98 (0.95–1.03)</td>
<td>0.79</td>
</tr>
<tr>
<td>rs7058826</td>
<td>\textit{F8}</td>
<td>23</td>
<td>Asian</td>
<td>A</td>
<td>0.06</td>
<td>0.00034</td>
<td>1.69 (1.27–2.24)</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>African American</td>
<td>A</td>
<td>0.10</td>
<td>0.47</td>
<td>0.98 (0.51–1.36)</td>
<td>0.94</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Caucasian</td>
<td>A</td>
<td>0.15</td>
<td>0.74</td>
<td>0.99 (0.95–1.04)</td>
<td>0.84</td>
</tr>
<tr>
<td>rs1983383</td>
<td>\textit{CAPN13}</td>
<td>2</td>
<td>Asian</td>
<td>A</td>
<td>0.45</td>
<td>0.00049</td>
<td>0.79 (0.69–0.90)</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>African American</td>
<td>A</td>
<td>0.49</td>
<td>0.11</td>
<td>0.99 (0.96–1.03)</td>
<td>0.83</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Caucasian</td>
<td>A</td>
<td>0.34</td>
<td>0.67</td>
<td>0.99 (0.96–1.03)</td>
<td>0.83</td>
</tr>
</tbody>
</table>
### Table 5

Identification of regulatory DNA elements coinciding with risk variants

<table>
<thead>
<tr>
<th>rsID</th>
<th>Chr</th>
<th>Position</th>
<th>Nearest Gene</th>
<th>Location</th>
<th>ENCODE</th>
<th>ROADMAP</th>
<th>‡ Super Enhancer</th>
<th>Motifs</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs3771044</td>
<td>2</td>
<td>206577411</td>
<td>NRP2</td>
<td>Intrinsic</td>
<td></td>
<td></td>
<td></td>
<td>BATF; ERα; Maf</td>
</tr>
<tr>
<td>rs10460829</td>
<td>3</td>
<td>124665905</td>
<td>MUC13</td>
<td>Intergenic</td>
<td></td>
<td></td>
<td></td>
<td>Gmeb1; p300</td>
</tr>
<tr>
<td>rs2100143</td>
<td>3</td>
<td>172202876</td>
<td>TNFSF10</td>
<td>Intergenic</td>
<td></td>
<td></td>
<td></td>
<td>ERα; HNF4; LRH1; MIF1; RXR; LXR; RXRA; SF1</td>
</tr>
<tr>
<td>rs10513702</td>
<td>3</td>
<td>172186953</td>
<td>GHSR</td>
<td>Intergenic</td>
<td></td>
<td></td>
<td></td>
<td>GR, HMG-1Y; Hoxa3; Myc; Nkx6-1; Pax-4; SIX5; XB1-</td>
</tr>
<tr>
<td>rs172310</td>
<td>7</td>
<td>155615627</td>
<td>SHH</td>
<td>Intergenic</td>
<td></td>
<td></td>
<td></td>
<td>RXRA; SP1; TR4</td>
</tr>
<tr>
<td>rs10954593</td>
<td>7</td>
<td>81036747</td>
<td>AC008163.4</td>
<td>Intergenic</td>
<td></td>
<td></td>
<td></td>
<td>Ets; NF-kappaB</td>
</tr>
<tr>
<td>rs288761</td>
<td>7</td>
<td>155633481</td>
<td>SHH</td>
<td>Intergenic</td>
<td></td>
<td></td>
<td></td>
<td>RXRA</td>
</tr>
<tr>
<td>rs1233556</td>
<td>7</td>
<td>155600417</td>
<td>SHH</td>
<td>Intrinsic</td>
<td></td>
<td></td>
<td></td>
<td>Ets; NF-kappaB</td>
</tr>
<tr>
<td>rs17702471</td>
<td>13</td>
<td>93877117</td>
<td>GPC6</td>
<td>Intergenic</td>
<td></td>
<td></td>
<td></td>
<td>RXRA</td>
</tr>
<tr>
<td>rs10794486</td>
<td>15</td>
<td>99301535</td>
<td>IGF1R</td>
<td>Intrinsic</td>
<td></td>
<td></td>
<td></td>
<td>Ets1; GR; OTX; Pax7</td>
</tr>
<tr>
<td>rs8030039</td>
<td>15</td>
<td>90767489</td>
<td>SEMA4B</td>
<td>Intrinsic</td>
<td></td>
<td></td>
<td></td>
<td>Ir; SR</td>
</tr>
<tr>
<td>rs7053448</td>
<td>X</td>
<td>154193211</td>
<td>F8</td>
<td>Intrinsic</td>
<td></td>
<td></td>
<td></td>
<td>Fos; Foxo2; Foxo2; Foxo1; Fox1; Pax31; Fox3f1; Fox5f1; Sox15; Sox19; Sox2; TATA; Zfp105</td>
</tr>
<tr>
<td>rs7058826</td>
<td>X</td>
<td>154194989</td>
<td>F8</td>
<td>Intrinsic</td>
<td></td>
<td></td>
<td></td>
<td>DMRT1; ERα; Ets2; Foxa; RREB-1</td>
</tr>
</tbody>
</table>
