Title

Identification and molecular characterization of a new ovarian cancer susceptibility locus at 17q21.31.

Journal

Nature communications, 4(1)

ISSN

2041-1723

Authors

Permuth-Wey, Jennifer
Lawrenson, Kate
Shen, Howard C
et al.

Publication Date

2013

DOI

10.1038/ncomms2613

Peer reviewed

HHS Public Access

Author manuscript

Nat Commun. Author manuscript; available in PMC 2013 July 12.

Published in final edited form as:

Nat Commun. 2013 ; 4: 1627. doi:10.1038/ncomms2613.

Identification and molecular characterization of a new ovarian cancer susceptibility locus at 17q21.31

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

Abstract

Epithelial ovarian cancer (EOC) has a heritable component that remains to be fully characterized. Most identified common susceptibility variants lie in non-protein-coding sequences. We hypothesized that variants in the 3′ untranslated region at putative microRNA (miRNA) binding sites represent functional targets that influence EOC susceptibility. Here, we evaluate the association between 767 miRNA binding site single nucleotide polymorphisms (miRSNPs) and EOC risk in 18,174 EOC cases and 26,134 controls from 43 studies genotyped through the Collaborative Oncological Gene-environment Study. We identify several miRSNPs associated with invasive serous EOC risk (OR=1.12, $P=10^{-8}$) mapping to an inversion polymorphism at 17q21.31. Additional genotyping of non-miRSNPs at 17q21.31 reveals stronger signals outside the inversion ($P=10^{-10}$). Variation at 17q21.31 associates with neurological diseases, and our collaboration is the first to report an association with EOC susceptibility. An integrated molecular analysis in this region provides evidence for *ARHGAP27* and *PLEKHM1* as candidate EOC susceptibility genes.

Genome wide association studies (GWAS) have identified hundreds of genetic variants conferring low penetrance susceptibility to cancer[1]. More than 90% of these variants lie in non protein-encoding sequences including non-coding RNAs and regions containing

Correspondence: Jennifer Permuth-Wey, PhD, Moffitt Cancer Center, 12902 Magnolia Dr., MRC-CANT, Tampa, Florida 33612, Jenny.Wey@moffitt.org Telephone: 813-745-5744, Fax: 813-745-6525; Thomas A. Sellers, PhD, Moffitt Cancer Center, 12902 Magnolia Dr, Tampa, Florida 33612, Thomas.Sellers@moffitt.org, Telephone: 813-745-1315, Fax: 813-449-8126; Simon Gayther, PhD, University of Southern California/Keck School of Medicine, Harlyne Norris Research Tower 1450 Biggy Street Los Angeles, CA 90033, gayther@usc.edu, Telephone: 323 442-8112, Fax: 323 442-7787.

*contributed equally to this manuscript

Author Contributions: These authors contributed equally to this work: JPW, KL, and HCS, and ANAM, TAS, and SAG. Writing group: JPW, KL, HCS, AV, ANAM, SAG, TAS, ELG, BLF, SJR, and PDPP. All authors read and approved the final version of the manuscript.

Provision of data and/or samples from contributing studies and institutions: JPW, KL, HCS, AV, JT, ZC, H-YL, YAC, Y- YT, XQ, SJR, RK, JL, NL, MCL, KA, HA-C, NA, AA, SMA, FB, LB, EB, JBS, MWB, MJB, GB, NB, LAB, ABW, RB, RB, QC, IC, JCC, SC, GCT, JQC, MSC, GAC, LSC, FJC, DWC, JMC, ADM, ED, JAD, TD, AdB, MD, DFE, DE, RE, ABE, PAF, DAF, JMF, MGC, AGM, GGG, RMG, JGB, MTG, MG, BG, JG, PH, MH, PH, FH, PH, MH, CH, EH, SH, AJ, AJ, HJ, KK, BYK, SBK, LEK, LAK, FK, GK, CK, SKK, JK, DL, SL, JML, NDL, AL, DAL, DL, JL, BKL, JL, KHL, JL, GL, LFAGM, KM, VM, JRM, UM, FM, KBM, TN, SAN, LN, RBN, HN, SN, HN, KO, SHO, IO, JP, CLP, TP, LMP, MCP, EMP, PR, SPR, HAR, LRR, MAR, AR, IR, IKR, HBS, IS, GS, VS, X-OS, YBS, WS, HS, MCS, BS, DS, RS, S-HT, KLT, DCT, PJT, SST, AMvA, IV, RAV, DV, AV, SW-G, RPW, NW, ASW, EW, LRW, BW, YLW, AHW, Y-BX, HPY, WZ, AZ, FZ, CMP, EI, JMS, AB, BLF, ELG, PDDP, ANAM, TAS, and SAG.

Collated and organized samples for genotyping: SJR and CMP.

Genotyping: JMC, DCT, FB, and DV.

Data analysis: JPW, JT, H-YL, YAC, BLF, MLL, and Y-Y T.

Functional analyses: SAG, ANAM, KL, HC S, AV, JL, RK, and SJR.

Bioinformatics support: ZC, XQ.

regulatory elements (i.e. enhancers, promoters, untranslated regions (UTRs))[1]. The emerging hypothesis is that common variants within non-coding regulatory regions influence expression of target genes, thereby conferring disease susceptibility[1].

MicroRNAs (miRNAs) are short non-coding RNAs that regulate gene expression post-transcriptionally by binding primarily to the 3′ UTR of target messenger RNA (mRNA), causing translational inhibition and/or mRNA degradation[2-4]. MiRNAs have been shown to play a key role in the development of epithelial ovarian cancer (EOC) [2]. We [5,6] and others [7] have found evidence that various miRNA-related single nucleotide polymorphisms (miRSNPs) are associated with EOC risk, suggesting they may be key disruptors of gene function and contributors to disease susceptibility [8,9]. However, studies of miRSNPs that affect miRNA-mRNA binding have been restricted by small sample sizes and therefore have limited statistical power to identify associations at genome wide levels of significance[7-9]. Larger-scale studies and more systematic approaches are warranted to fully evaluate the role of miRSNPs and their contribution to disease susceptibility.

Here, we use the *in silico* algorithms, TargetScan [10,11] and Pictar [12,13] to predict miRNA:mRNA binding regions involving genes and miRNAs relevant to EOC, and align identified regions with SNPs in the dbSNP database (Methods). We then genotype 1,003 miRSNPs (or tagging SNPs with r^2>0.80) in 18,174 EOC cases and 26,134 controls from 43 studies from the Ovarian Cancer Association Consortium (OCAC) (Supplementary Table S1). Genotyping was performed on a custom Illumina Infinium iSelect array designed as part of the Collaborative Oncological Gene-environment Study (COGS), an international effort that evaluated 211,155 SNPs and their association with ovarian, breast, and prostate cancer risk. Our investigation uncovers 17q21.31 as a new susceptibility locus for EOC, and we provide insights into candidate genes and possible functional mechanisms underlying disease development at this locus.

Results

Association analyses

Seven hundred and sixty-seven of the 1,003 miRSNPs passed genotype quality control (QC) and were evaluated for association with invasive EOC risk; most of the miRSNPs that failed QC were monomorphic (see Methods). Primary analysis of 14,533 invasive EOC cases and 23,491 controls of European ancestry revealed four strongly correlated SNPs (r^2=0.99; rs1052587, rs17574361, rs4640231, and rs916793) that mapped to 17q21.31 and were associated with increased risk (per allele odds ratio (OR) = 1.10, 95% CI 1.06-1.13) at a genome-wide level of significance (10^{-7}); no other miRSNPs had associations stronger than $P<10^{-4}$ (Supplementary Fig. S1). The most significant association was for rs1052587 ($P=1.9\times10^{-7}$), and effects varied by histological subtype, with the strongest effect observed for invasive serous EOC cases (OR=1.12, $P=4.6\times10^{-8}$) (Table 1). No heterogeneity in ORs was observed across study sites (Supplementary Fig. S2).

Rs1052587, rs17574361, and rs4640231 reside in the 3′UTR of microtubule-associated protein tau (*MAPT*), KAT8 regulatory NSL complex subunit 1 (*KANSL1/KIAA1267*), and corticotrophin releasing hormone receptor 1 (*CRHR1*) genes, at putative binding sites for

miR-34a, miR-130a, and miR-34c, respectively. The fourth SNP, rs916793, is perfectly correlated with rs4640231 and lies in a non-coding RNA, *MAPT*-antisense 1. 17q21.31 contains a ∼900kb inversion polymorphism[14] (ch 17: 43,624,578-44,525,051 MB, human genome build 37), and all three miRSNPs and the tagSNP are located within the inversion (Fig. 1).

Chromosomes with the non-inverted or inverted segments of 17q21.31, respectively known as haplotype 1 (H1) and haplotype 2 (H2), represent two distinct lineages that diverged ∼3 million years ago and have not undergone any recombination event [14]. The four susceptibility alleles identified here reside on the H2 haplotype that is reported to be rare in Africans and East Asians, but is common (frequency >20%) and exhibits strong linkage disequilibrium (LD) among Europeans [14], consistent with our findings. The H2 haplotype has a frequency of 22% among European women in our primary analysis (Table 1) but only 3.2% and 0.3% among Africans (151 invasive cases, 200 controls) and Asians (716 invasive cases, 1573 controls), respectively.

To increase genomic coverage at this locus, we evaluated an additional 142 non-miRSNPs at 17q21.31 that were also genotyped as part of COGS in the same series of OCAC cases and controls. We also imputed genotypes using data from the 1000 Genomes Project[15]. These approaches identified a second cluster of strongly correlated SNPs (r^2>0.90) in a distinct region proximal to the inversion (centered at chromosome 17: 43.5 MB, human genome build 37) that was more significantly associated with the risk of all invasive EOCs ($P= 10^{-9}$) and invasive serous EOC specifically ($P= 10^{-10}$) than the cluster of identified miRSNPs (Fig. 1). Association results and annotation for SNPs in this second cluster are shown in Supplementary Table S2; this cluster includes three directly genotyped SNPs (rs2077606, rs17631303, and rs12942666), with the strongest association observed for rs2077606 among all invasive cases (OR=1.12, 95% CI: 1.08-1.16), $P=7.8\times10^{-9}$) and invasive serous cases (OR=1.15, 95% CI: 1.12-1.19, $P=3.9\times10^{-10}$). These SNPs were chosen for genotyping in COGS because they had shown evidence of association as modifiers of EOC risk in BRCA1 gene mutation carriers by the Consortium of Investigators of Modifiers of BRCA1/2 (CIMBA)[16]. Several imputed SNPs in strong LD (r^2>0.90) were more strongly associated with risk than their highly correlated genotyped SNPs (Supplementary Table S2). This risk-associated region at 17q21.31 is distinct from a previously reported ovarian cancer susceptibility locus at 17q21[17]; neither the genotyped or imputed SNPs we report here are strongly correlated (maximum r^2= 0.01) with SNPs from the 17q21 locus (spanning 46.2-46.5 MB, build 37).

Genotype clustering was poor for rs2077606, but clustering was good for its correlated SNP, rs12942666 (r^2=0.99), and so results for this SNP are presented instead (Supplementary Fig. S2; Table 1). Subgroup analysis revealed marginal evidence of association for rs12942666 with endometrioid (P=0.04), but not mucinous or clear cell EOC subtypes (Table 1), and results were consistent across studies (Supplementary Fig. S4). Rs12942666 is correlated with the top-ranked miRSNP, rs1052587 (r^2=0.76) (Fig. 1). To evaluate whether associations observed for rs12942666 and rs1052587 represented independent signals, stepwise logistic regression was used; only rs12942666 was retained in the model. This

suggests that the cluster which includes rs12942666 is driving the association with EOC risk that was initially identified through the candidate miRSNPs.

Functional and molecular analyses

To evaluate functional evidence for candidate genes, risk-associated SNPs, and regulatory regions at 17q21.31, we examined a one megabase region centered on rs12942666 using a combination of locus specific and genome-wide assays and *in silico* analyses of publicly available datasets, including The Cancer Genome Atlas (TCGA) Project[18] (see Methods). Rs12942666 and many of its correlated SNPs lie within introns of Rho GTPase activating protein 27 (*ARHGAP27*) or its neighboring gene, pleckstrin homology domain containing, family M (with RUN domain) member 1 (*PLEKHM1*) (Supplementary Table S2). There are another 15 known protein-coding genes within the region: *KIF18B, C1QL1, DCAKD, NMT1, PLCD3, ABCB4, HEXIM1, HEXIM2, FMNL1, C17orf46, MAP3K14, C17orf69, CRHR1, IMP5*, and *MAPT* (Fig. 2a).

To evaluate the likelihood that one or more genes within this region represent target susceptibility gene(s), we first analyzed expression, copy number variation, and methylation involving these genes in EOC tissues and cell lines (Fig. 2b-g; Supplementary Tables S3 and S4). Most genes showed significantly higher expression ($P<10^{-4}$) in EOC cell lines versus normal ovarian cancer-precursor tissues (OCPTs); *ARHGAP27* showed the most pronounced difference in gene expression between cancer and normal cells ($P=10^{-16}$) (Fig. 2b and Supplementary Table S3). For nine genes, we also found overexpression in primary high-grade serous (HGS) EOC tumors versus normal ovarian tissue in at least one of two publicly available datasets, The Cancer Genome Atlas (TCGA) of 568 tumors [18]and/or the Gene Expression Omnibus (GEO) series GSE18520 dataset consisting of 53 tumors[19] (Fig. 2c and Supplementary Table S3). Analysis of DNA copy number variation in TCGA revealed frequent loss of heterozygosity in this region rather than gains (Supplementary Fig. 5a-b; Supplementary Methods). We observed significant hypomethylation ($P<0.01$) in ovarian tumors compared to normal tissue for *DCAKD, PLCD3, ACBD4, FMNL1*, and *PLEKHM1* (Fig. 2d and Supplementary Table S4), which is consistent with the overexpression observed for *DCAKD, PLCD3*, and *FMNL1*. Taken together, these data suggest that the mechanism underlying overexpression may be epigenetic rather than based on copy number alterations.

We evaluated associations between genotypes for the top risk SNP rs12942666 (or a tagSNP) and expression of all genes in the region (expression quantitative trait locus (eQTL) analysis) in normal OCPTs, lymphoblastoid cell lines (LCLs), and primary tumors from TCGA. We observed significant eQTL associations ($P<0.05$) in normal OCPTs only for *ARHGAP27* ($P=0.04$) (Fig. 2e; Supplementary Table S3). Because rs12942666 was not genotyped in tissues analyzed in TCGA, we used data for its correlated SNP rs2077606 ($r^2=0.99$) to evaluate eQTLs in tumor tissues. Rs2077606 genotypes were strongly associated with *PLEKHM1* expression in primary HGS-EOCs ($P=1\times10^{-4}$) (Fig. 2f; Supplementary Table S3). We also detected associations between rs12942666 (and rs2077606) genotypes and methylation for *PLEKHM1* and *CRHR1* in primary tumors ($P=0.020$ and 0.001, respectively) using methylation quantitative trait locus (mQTL) analyses (Fig. 2g; Supplementary Table S4). Finally, the Catalogue of Somatic Mutations in

Cancer (COSMIC) database [20] showed that nine genes in the region, including *PLEKHM1*, have functionally significant mutations in cancer, although for most genes mutations were not reported in ovarian carcinomas (Supplementary Table S3).

Taken together, these data suggest that several genes at the 17q21.31 locus may play a role in EOC development. The risk-associated SNPs we identified fall within non-coding DNA, suggesting the functional SNP(s) may be located within an enhancer, insulator, or other regulatory element that regulates expression of one of the candidate genes we evaluated. One hypothesis emerging from these molecular analyses is that rs12942666 (or a correlated SNP) mediates regulation of *PLEKHM1*, a gene implicated in osteopetrosis and endocytosis [21] and/or *ARHGAP27*, a gene that may promote carcinogenesis through dysregulation of Rho/Rac/Cdc42-like GTPases [22]. To identify the most likely candidate for being the causal variant at 17q21.31, we compared the difference between log-likelihoods generated from un-nested logistic regression models for rs12942666 and each of 198 SNPs in a 1 MB region featured in Supplementary Table 2. As expected, the log likelihoods were very similar due to the strong LD; no SNPs emerged as having a likelihood ratio greater than 20 for being the causal variant.

To explore the possible functional significance of rs12942666 and strongly correlated variants (r^2>0.80), we then generated a map of regulatory elements around rs12942666 using ENCODE data and FAIRE-seq analysis of OCPTs (Supplementary Methods). We observed no evidence of putative regulatory elements coinciding with rs12942666 or correlated SNPs (Fig. 3a). A map of regulatory elements in the entire 1 MB region can be seen in Supplementary Fig. 5c-f. We subsequently used *in silico* tools (ANNOVAR[23], SNPinfo[24], and SNPnexus[25]) to evaluate the putative function of possible causal SNPs (Supplementary Methods). Of 50 SNPs with possible functional roles, more than 30 reside in putative transcription factor binding sites (TFBS) within or near *PLEKHM1* or *ARHGAP27*; 12 SNPs may affect methylation or miRNA binding, and two are non-synonymous coding variants predicted to be of no functional significance (Supplementary Table S2).

Since most of the top-ranked 17q21.31 SNPs with putative functions (including two of the top directly genotyped SNPs, rs2077606 and rs17631303), are predicted to lie in TFBS (Supplementary Table S2), we used the *in silico* tool, JASPAR [26] to further examine TFBS coinciding with these SNPs. Two SNPs scored highly in this analysis (Supplementary Table S5); the first, rs12946900, lies in a GAGGAA motif and canonical binding site for *SPIB*, an Ets family member[27]. Ets factors have been implicated in the development of ovarian cancer and other malignancies[28], but little evidence supports a specific role for *SPIB* in EOC etiology. The second hit was for rs2077606, which lies in an E-box motif CACCTG at the canonical binding site for *ZEB1* (chr. 10p11.2), a zinc-finger E-box binding transcription factor that represses E-cadherin[29,30] and contributes to epithelial-mesenchymal transition in EOCs [31].

We analyzed expression of *SPIB* and *ZEB1* in primary ovarian cancers using TCGA data; we found no significant difference in *SPIB* expression in tumors compared to normal tissues (Fig. 3bi). In contrast, *ZEB1* expression was significantly lower in primary HGS-EOCs

compared to normal tissues (P=0.005) (Fig. 3bii). We validated this finding using qPCR analysis in 123 EOC and OCPT cell lines (P=8.8 ×10^{-4}) (Fig. 3biii). Since rs2077606 lies within an intron of *PLEKHM1*, this gene is a candidate target for *ZEB1* binding at this site. Our eQTL analysis also suggests *ARHGAP27* is a strong candidate *ZEB1* target at this locus; *ARHGAP27* expression is highest in OCPT cell lines carrying the minor allele of rs2077606 (P=0.034) (Figure 3ci). Although we observed no eQTL associations between rs2077606 and *ZEB1* expression in LCLs (Figure 3cii), we found evidence of eQTL between rs2077606 and *ZEB1* expression in HGS-EOCs (P=0.045) (Figure 3ciii). *ZEB1* binding at the site of the common allele is predicted to repress gene expression while loss of *ZEB1* binding conferred by the minor allele may enable expression of *ARHGAP27*, consistent with the eQTL association in OCPTs (Fig. 3ci). Although this data supports a repressor role for *ZEB1* in EOC development and suggests ARHGAP27 may be a functional target of rs2077606 (or a correlated SNP) in OCPTs through trans-regulatory interactions with *ZEB1*, it is important to investigate additional hypotheses as we continue to narrow down the list of target susceptibility genes, SNPs, and regulatory mechanisms that contribute to EOC susceptibility at this locus.

Discussion

The present study represents the largest, most comprehensive investigation of the association between putative miRSNPs in the 3′ untranslated region and cancer risk. This and the systematic follow-up to evaluate associations with EOC risk for non-miRSNPs in the region identified 17q21.31 as a new susceptibility locus for EOC. Although the miRSNPs identified here may have some biological significance, our findings suggest that other types of variants in non-coding DNA, especially non-miRSNPs at the 17q21.31 locus, are stronger contributors to EOC risk. It is possible, however, that highly significant miRSNPs exist that were not identified in our study because a) they were not pre-selected for evaluation (i.e. they do not reside in a binding site involving miRNAs or genes with known relevance to EOC, or they reside in regions other than the 3′UTR[3,4]) and/or b) they were very rare and could not be designed or detected with our genotyping platform and sample size, respectively. Despite these limitations, the homogeneity between studies of varying designs and populations in the OCAC and the genome-wide levels of statistical significance imply that all detected associations are robust. Furthermore, molecular correlative analyses of genes within the region suggest that cis-acting genetic variants influencing non-coding DNA regulatory elements, miRNAs, and/or methylation underlie disease susceptibility at the 17q21.31 locus. Finally, these studies point to a subset of candidate genes (i.e. *PLEKHM1*, *ARHGAP27*) and transcription factors (i.e. ZEB1) that may influence EOC initiation and development.

This novel locus is one of eleven loci now identified that contains common genetic variants conferring low penetrance susceptibility to EOC in the general population [17,32,33,34]. Genetic variants at several of these loci influence risks of more than one cancer type, suggesting that several cancers may share common mechanisms. For example, alleles at 5p15.33 and 19p13.1 are associated with estrogen-receptor-negative breast cancer and serous EOC susceptibility [32,35], and variants at 8q24 are associated with risk of EOC and other cancers [17,36]. Genetic variation at 17q21.31 is also associated with frontotemporal

dementia-spectrum disorders, Parkinson's disease, developmental delay, and alopecia [37-42]. Through COGS, the CIMBA also recently identified 17q21.31 variants as modifying EOC risk in *BRCA1* and *BRCA2* carriers ($P<10^{-8}$ in *BRCA1/2* combined)[16]. In particular, rs17631303, which is perfectly correlated with rs2077606 and rs12942666, was among the top-ranking SNPs detected by CIMBA[16]. Consistent with our findings, CIMBA also provide data that suggests EOC risk is associated with altered expression of one or more genes in the 17q21.31 region[16]. Thus, results from this large-scale collaboration support a role for this locus in both *BRCA1/2* and non-*BRCA1/2* mediated EOC development. Before these findings can be integrated with variants from other confirmed loci and non-genetic factors to predict women at greatest risk of developing EOC and provide options for medical management of these risks, continued efforts will be needed to fine map the 17q21.31 region and to fully characterize the functional and mechanistic effects of potential causal SNPs in disease etiology and development.

Methods

Study population

Forty-three individual OCAC studies contributed samples and data to the COGS initiative. Nine of the 43 participating studies were case-only (GRR, HSK, LAX, ORE, PVD, RMH, SOC, SRO, UKR); cases from these studies were pooled with case-control studies from the same geographic region. The two national Australian case-control studies were combined into a single study to create 34 case-control sets. Details regarding the 43 participating OCAC studies are summarized in Supplementary Table S1. Briefly, cases were women diagnosed with histologically confirmed primary EOC (invasive or low malignant potential), fallopian tube cancer, or primary peritoneal cancer ascertained from population- and hospital-based studies and cancer registries. The majority of OCAC cases (>90%) do not have a family history of ovarian or breast cancer in a first-degree relative, and most have not been tested for *BRCA1/2* mutations as part of their parent study. Controls were women without a current or prior history of ovarian cancer with at least one ovary intact at the reference date. All studies had data on disease status, age at diagnosis/interview, self-reported racial group, and histologic subtype. Most studies frequency-matched cases and controls on age-group and race.

Selection of Candidate Genes and SNPs

To increase the likelihood of identifying miRSNPs with biological relevance to EOC, we reviewed published literature and consulted public databases to generate two lists of candidate genes: 1) 55 miRNAs reported to be deregulated in EOC tumors compared to normal tissue in at least one study [43-46], and 2) 665 genes implicated in the pathogenesis of EOC through gene expression analyses [47,48], somatic mutations [49], or genetic association studies [50,51]. Many genes were identified through the Gene Prospector database[51], a web-based application that selects and prioritizes potential disease-related genes using a highly curated, up-to-date database of genetic association studies.

Using each candidate gene list as input, we identified putative sites of miRNA:mRNA binding with the computational prediction algorithms TargetScan version 5.1 [10,11] and

PicTar [12,13] and Supplementary Methods). Each algorithm generated start and end coordinates for regions of miRNA binding, and database SNP (dbSNP)[52] version 129 was mined to identify SNPs falling within the designated binding regions. Of 3,246 unique miRSNPs that were identified, 1102 obtained adequate design scores using Illumina's Assay Design Tool. The majority (n=1085, 98.5%) of the 1102 SNPs resided in predicted sites of miRNA binding (and therefore represent miRSNPs), while the remainder (n=17) are tagSNPs ($r^2 > 0.80$) for miRSNPs that were not designable or had poor to moderate design scores. Ninety nine of the 1102 SNPs failed during custom assay development, leaving a total of 1,003 SNPs that were designed and genotyped.

Genotyping and QC

The candidate miRSNPs selected for the current investigation were genotyped using a custom Illumina Infinium iSelect Array as part of the international Collaborative Oncological Gene-environment Study (COGS), an effort to evaluate 211,155 genetic variants for association with the risk of ovarian, breast, and prostate cancer. Samples and data were included from several consortia, including OCAC, the Breast Cancer Association Consortium (BCAC), the Consortium of Investigators of Modifiers of *BRCA1/2* (CIMBA), and the Prostate Cancer Association Group to Investigate Cancer- Associated Alterations in the Genome (PRACTICAL). Although one of the primary goals of COGS was to replicate and fine-map findings from pooled genome-wide association studies (GWAS) from each consortia, this effort also aimed to genotype candidate SNPs of interest (such as the miRSNPs). The genotyping and QC process has been described recently in our report of OCAC's pooled GWAS findings[34]. Briefly, COGS genotyping was conducted at six centers, two of which were used for OCAC samples: McGill University and Génome Québec Innovation Centre (Montréal, Canada) (n=19,806) and Mayo Clinic Medical Genomics Facility (n=27,824). Each 96-well plate contained 250ng genomic DNA (or 500 ng whole genome-amplified DNA). Raw intensity data files were sent to the COGS data coordination center at the University of Cambridge for genotype calling and QC using the GenCall algorithm.

Sample QC—One thousand two hundred and seventy three OCAC samples were genotyped in duplicate. Genotypes were discordant for greater than 40 percent of SNPs for 22 pairs. For the remaining 1,251 pairs, concordance was greater than 99.6 percent. In addition we identified 245 pairs of samples that were unexpected genotypic duplicates. Of these, 137 were phenotypic duplicates and judged to be from the same individual. We used identity-by-state to identify 618 pairs of first-degree relatives. Samples were excluded according to the following criteria: 1) 1,133 samples with a conversion rate (the proportion of SNPs successfully called per sample) of less than 95 percent; 2) 169 samples with heterozygosity >5 standard deviations from the intercontinental ancestry specific mean heterozygosity; 3) 65 samples with ambiguous sex; 4) 269 samples with the lowest call rate from a first-degree relative pair 5) 1,686 samples that were either duplicate samples that were non-concordant for genotype or genotypic duplicates that were not concordant for phenotype. A total of 44,308 eligible subjects including 18,174 cases and 26,134 controls were available for analysis.

SNP QC—The process of SNP selection by the participating consortia has been summarized previously[34]. In total, 211,155 SNP assays were successfully designed, including 23,239 SNPs nominated by OCAC. Overall, 94.5% of OCAC-nominated SNPs passed QC. SNPs were excluded if: (1) the call rate was less than 95% with MAF > 5% or less than 99% with MAF < 5% (n=5,201); (2) they were monomorphic upon clustering (n=2,587); (3) p values of HWE in controls were less than 10^{-7} (n=2,914); (4) there was greater than 2% discordance in duplicate pairs (n=22); (5) no genotypes were called (n=1,311). Of 1,003 candidate miRSNPs genotyped, 767 passed QC criteria and were available for analysis; the majority of miRSNPs that were excluded were monomorphic (n=158, 67%). Genotype intensity cluster plots were visually inspected for the most strongly associated SNPs.

Population stratification

HapMap DNA samples for European (CEU, n=60), African (YRI, n=53) and Asian (JPT +CHB, n=88) populations were also genotyped using the COGS iSelect. We used the program LAMP [53] to estimate intercontinental ancestry based on the HapMap (release no. 23) genotype frequency data for these three populations. Eligible subjects with greater than 90 percent European ancestry were defined as European (n=39,773) and those with greater than 80 percent Asian or African ancestry were defined as Asian (n=2,382) or African respectively (n=387). All other subjects were defined as being of mixed ancestry (n=1,766). We then used a set of 37,000 unlinked markers to perform principal components analysis within each major population subgroup. To enable this analysis on very large sample sizes we used an in-house program written in C++ using the Intel MKL libraries for eigenvectors (available at http://ccge.medschl.cam.ac.uk/software/).

Tests of association

We used unconditional logistic regression treating the number of minor alleles carried as an ordinal variable (log-additive model) to evaluate the association between each SNP and EOC risk. Separate analyses were carried out for each ancestry group. The model for European subjects was adjusted for population substructure by including the first 5 eigenvalues from the principal components analysis. African- and Asian- ancestry-specific estimates were obtained after adjustment for the first two components representing each respective ancestry. Due to the heterogeneous nature of EOC, subgroup analysis was conducted to estimate genotype-specific odds ratios for serous carcinomas (the most predominant histologic subtype) and the three other main histological subtypes of EOC: endometrioid, mucinous, and clear cell. Separate analyses were also carried out for each study site, and site-specific ORs were combined using a fixed-effect meta-analysis. The I^2 test of heterogeneity was estimated to quantify the proportion of total variation due to heterogeneity across studies, and the heterogeneity of odds ratios between studies was tested with Cochran's Q statistic. The R statistical package ‘r-meta’ was used to generate forest plots. Statistical analysis was conducted in PLINK[54].

Imputation of genotypes at 17q21.31

To increase genomic coverage, we imputed genotype data for the 17q21.31 region (chr17: 40,099,001-44,900,000, human genome build 37) with IMPUTE2.2 [55] using phase 1 haplotype data from the January 2012 release of the 1000 genome project data [15]. For each imputed genotype the expected number of minor alleles carried was estimated (as weights). IMPUTE provides estimated allele dosage for SNPs that were not genotyped and for samples with missing data for directly genotyped SNPs. Imputation accuracy was estimated using an r^2quality metric. We excluded imputed SNPs from analysis where the estimated accuracy of imputation was low ($r^2<0.3$).

Functional studies and *in silico* analysis of publicly available datasets

We performed the following assays for each gene in the one megabase region centered on the most significant SNP at the 17q21.31 locus (see Supplementary Methods): gene expression analysis in EOC cell lines (n=51) compared to normal cell lines from ovarian cancer precursor tissues (OCPTs)[56], including ovarian surface epithelial cells (OSECs) and fallopian tube secretory epithelial cells (FTSECs) (n=73), and CpG island methylation analysis in high grade serous ovarian cancer (HGS-EOC) tissues (n=106) and normal tissues (n=7). Genes in the region were also evaluated *in silico* by mining publicly available molecular data generated for primary EOCs and other cancer types, including The Cancer Genome Atlas (TCGA) analysis of 568 HGS EOCs[18], the Gene Expression Omnibus series GSE18520 dataset of 53 HGS EOCs [19], and the Catalogue Of Somatic Mutations In Cancer (COSMIC) database[20].

We used these data to 1) compare gene expression between a) EOC cell lines and normal cell lines and b) tumor tissue and normal tissue from TCGA, 2) compare gene methylation status in HGS-EOCs and normal tissue, 3) conduct gene expression quantitative trait locus (eQTL) analyses to evaluate genotype-gene expression associations in normal OCPTs, lymphoblastoid cells, and HGS-EOCs, and 4) conduct methylation quantitative trait locus (mQTL) analyses in HGS-EOCs to evaluate genotype-gene methylation associations. Data from ENCyclopedia Of DNA Elements (ENCODE) [57] were used to evaluate the overlap between regulatory elements in non-coding regions and risk-associated SNPs. ENCODE describes regulatory DNA elements (e.g. enhancers, insulators and promotors) and non-coding RNAs (e.g. miRNAs, long non-coding and piwi-interacting RNAs) that may be targets for susceptibility alleles. However, ENCODE does not include data for EOC associated tissues, and activity of such regulatory elements often varies in a tissue specific manner [57,58]. Therefore, we profiled the spectrum of non-coding regulatory elements in OSECs and FTSECs using a combination of formaldehyde assisted isolation of regulatory elements sequencing (FAIRE-seq) and RNA sequencing (RNA-seq) (Supplementary Methods).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Authors

Jennifer Permuth-Wey[1,*], Kate Lawrenson[2,*], Howard C. Shen[2,*], Aneliya Velkova[1], Jonathan P. Tyrer[3], Zhihua Chen[4], Hui-Yi Lin[5], Y. Ann Chen[5], Ya-Yu Tsai[1], Xiaotao Qu[4], Susan J. Ramus[2], Rod Karevan[2], Janet Lee[2], Nathan Lee[2], Melissa C. Larson[6], Katja K. Aben[7,8], Hoda Anton-Culver[9], Natalia Antonenkova[10], Antonis Antoniou[11], Sebastian M. Armasu[6], Australian Cancer Study[12], Australian Ovarian Cancer Study[12,13], François Bacot[14], Laura Baglietto[15,16], Elisa V. Bandera[17], Jill Barnholtz-Sloan[18], Matthias W. Beckmann[19], Michael J. Birrer[20], Greg Bloom[4], Natalia Bogdanova[21], Louise A. Brinton[22], Angela Brooks-Wilson[23], Robert Brown[24], Ralf Butzow[25,26], Qiuyin Cai[27], Ian Campbell[13,28], Jenny Chang-Claude[29], Stephen Chanock[22], Georgia Chenevix-Trench[12], Jin Q. Cheng[30], Mine S. Cicek[31], Gerhard A. Coetzee[32], Consortium of Investigators of Modifiers of BRCA1/2[33,34], Linda S. Cook[35], Fergus J. Couch[34], Daniel W. Cramer[36], Julie M. Cunningham[31], Agnieszka Dansonka-Mieszkowska[37], Evelyn Despierre[38], Jennifer A Doherty[39], Thilo Dörk[21], Andreas du Bois[40,41], Matthias Dürst[42], Douglas F Easton[11,43], Diana Eccles[44], Robert Edwards[45], Arif B. Ekici[46], Peter A. Fasching[19,47], David A. Fenstermacher[4], James M. Flanagan[24], Montserrat Garcia-Closas[48], Aleksandra Gentry-Maharaj[49], Graham G. Giles[15,16,50], Rosalind M. Glasspool[51], Jesus Gonzalez-Bosquet[52], Marc T. Goodman[53], Martin Gore[54], Bohdan Górski[55], Jacek Gronwald[55], Per Hall[56], Mari K. Halle[57,58], Philipp Harter[40,41], Florian Heitz[40,41], Peter Hillemanns[59], Maureen Hoatlin[60], Claus K. Høgdall[61], Estrid Høgdall[62,63], Satoyo Hosono[64], Anna Jakubowska[55], Allan Jensen[63], Heather Jim[65], Kimberly R. Kalli[66], Beth Y. Karlan[67], Stanley B. Kaye[68], Linda E. Kelemen[69], Lambertus A. Kiemeney[7,8,70], Fumitaka Kikkawa[71], Gottfried E. Konecny[47], Camilla Krakstad[57,58], Susanne Krüger Kjaer[61,63], Jolanta Kupryjanczyk[37], Diether Lambrechts[72,73], Sandrina Lambrechts[38], Johnathan M. Lancaster[52], Nhu D. Le[74], Arto Leminen[26], Douglas A. Levine[75], Dong Liang[76], Boon Kiong Lim[77], Jie Lin[78], Jolanta Lissowska[79], Karen H. Lu[80], Jan Lubiński[55], Galina Lurie[81], Leon F.A.G. Massuger[82], Keitaro Matsuo[64], Valerie McGuire[83], John R McLaughlin[84,85], Usha Menon[49], Francesmary Modugno[78,80,86], Kirsten B. Moysich[87], Toru Nakanishi[88], Steven A. Narod[89], Lotte Nedergaard[90], Roberta B. Ness[91], Heli Nevanlinna[26], Stefan Nickels[29], Houtan Noushmehr[32,92], Kunle Odunsi[93], Sara H. Olson[94], Irene Orlow[94], James Paul[51], Celeste L Pearce[2], Tanja Pejovic[95,96], Liisa M. Pelttari[26], Malcolm C. Pike[2,94], Elizabeth M. Poole[97,98], Paola Raska[18], Stefan P. Renner[19], Harvey A. Risch[99], Lorna Rodriguez-Rodriguez[17], Mary Anne Rossing[100,101], Anja Rudolph[29], Ingo B. Runnebaum[42], Iwona K. Rzepecka[37], Helga B. Salvesen[57,58], Ira Schwaab[102], Gianluca Severi[15,16], Vijayalakshmi Shridhar[103], Xiao-Ou Shu[27], Yurii B. Shvetsov[81], Weiva Sieh[83], Honglin Song[3], Melissa C. Southey[104], Beata Spiewankiewicz[105], Daniel Stram[2], Rebecca Sutphen[106], Soo-Hwang Teo[77], Kathryn L. Terry[36], Daniel C. Tessier[14], Pamela J. Thompson[53], Shelley S. Tworoger[97,98], Anne M. van Altena[82], Ignace Vergote[38], Robert A. Vierkant[31], Daniel Vincent[14], Allison F. Vitonis[36], Shan Wang-Gohrke[107], Rachel Palmieri Weber[108], Nicolas Wentzensen[22], Alice S. Whittemore[83], Elisabeth Wik[57,58], Lynne R. Wilkens[81], Boris Winterhoff[109], Yin Ling

Woo[77], Anna H. Wu[2], Yong-Bing Xiang[110], Hannah P. Yang[22], Wei Zheng[27], Argyrios Ziogas[111], Famida Zulkifli[77], Catherine M. Phelan[1], Edwin Iversen[112], Joellen M. Schildkraut[108,113], Andrew Berchuck[114], Brooke L. Fridley[115], Ellen L. Goode[31], Paul D. P. Pharoah[11,43], Alvaro N.A. Monteiro[1], Thomas A. Sellers[1], and Simon A. Gayther[2]

Affiliations

[1]Department of Cancer Epidemiology, Division of Population Sciences, Moffitt Cancer Center, Tampa, FL, USA, 33612 [2]Department of Preventive Medicine, Keck School of Medicine, University of Southern California Norris Comprehensive Cancer Center, Los Angeles, CA, USA, 90033 [3]Department of Oncology, University of Cambridge, Cambridge, CB1 8RN, UK [4]Department of Biomedical Informatics, Moffitt Cancer Center, Tampa, FL, USA, 33612 [5]Department of Biostatistics, Moffitt Cancer Center, Tampa, FL, USA, 33612 [6]Department of Health Science Research, Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN, USA, 55905 [7]Department of Epidemiology, Biostatistics and HTA, Radboud University Medical Centre, Nijmegen, HB 6500, Netherlands [8]Comprehensive Cancer Center, the Netherlands, Utrecht, Amsterdam, 1066CX, The Netherlands [9]Department of Epidemiology, Director of Genetic Epidemiology Research Institute, UCI Center of Medicine, University of California Irvine, Irvine, CA, USA, 92697 [10]Byelorussian Institute for Oncology and Medical Radiology Aleksandrov N.N., 223040, Minsk, Belarus [11]Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, CB1 8RN, UK [12]Queensland Institute of Medical Research, Brisbane QLD 4006, Australia [13]Cancer Genetics Laboratory, Research Division, Peter MacCallum Cancer Centre, Melbourne, VIC 3002, Australia [14]McGill University and Génome Québec Innovation Centre, Montréal (Québec) Canada, H3A 0G1 [15]Cancer Epidemiology Centre, The Cancer Council Victoria, Melbourne, Carlton VIC 3053, Australia [16]Centre for Molecular, Environmental, Genetic and Analytical Epidemiology, University of Melbourne, Melbourne, VIC 3010, Australia [17]The Cancer Institute of New Jersey, Robert Wood Johnson Medical School, New Brunswick, NJ, USA, 08901 [18]Case Comprehensive Cancer Center, Case Western Reserve University School of Medicine, Cleveland, OH, USA, 44195 [19]University Hospital Erlangen, Department of Gynecology and Obstetrics, Friedrich-Alexander-University Erlangen-Nuremberg, Comprehensive Cancer Center, Erlangen, 91054, Germany [20]Massachusetts General Hospital, Boston, MA, USA, 02114 [21]Gynaecology Research Unit, Hannover Medical School, Hannover, 30625, Germany [22]Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda MD, USA, 20892 [23]Genome Sciences Centre, BC Cancer Agency, Vancouver, BC, Canada, V5Z 1L3 [24]Department of Surgery and Cancer, Imperial College London, London, SW7 2AZ, UK [25]Department of Pathology, Helsinki University Central Hospital, Helsinki, Finland, 00530 [26]Department of Obstetrics and Gynecology, University of Helsinki and Helsinki University Central Hospital, Helsinki, Finland, 00530 [27]Vanderbilt Epidemiology Center and Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, TN, USA, 37232 [28]Department of Pathology,

University of Melbourne, Parkville, VIC 3053, Australia [29]German Cancer Research Center, Division of Cancer Epidemiology, 69120, Heidelberg, Germany [30]Department of Interdisciplinary Oncology, Moffitt Cancer Center, Tampa, FL, USA, 33612 [31]Division of Epidemiology, Department of Health Sciences Research, Mayo Clinic College of Medicine, Rochester, MN, USA, 55905 [32]Department of Urology, Microbiology and Preventive Medicine, University of Southern California, Norris Comprehensive Cancer Center, Los Angeles, CA, USA, 90089 [33]Cancer Research UK, Genetic Epidemiology Unit, Dept of Public Health & Primary Care, University of Cambridge, Strangeways Research Lab, Cambridge, CB1 8RN, UK [34]Department of Laboratory of Medicine and Pathology, Mayo Clinic, Rochester, MN, USA, 55905 [35]Division Epidemiology and Biostatistics, University of New Mexico, Albuquerque, NM, USA, 87131 [36]Obstetrics and Gynecology Epidemiology Center, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA, 02115 [37]Department of Molecular Pathology, The Maria Sklodowska-Curie Memorial Cancer Center and Institute of Oncology, Warsaw, Poland, 02-781 [38]Division of Gynecologic Oncology, Department of Obstetrics and Gynaecology and Leuven Cancer Institute, University Hospitals Leuven, Leuven, Belgium, 3000 [39]Section of Biostatistics and Epidemiology, The Geisel School of Medicine at Dartmouth, Lebanon, NH, USA, 03755 [40]Department of Gynecology and Gynecologic Oncology, Dr. Horst Schmidt Klinik Wiesbaden, 65199, Wiesbaden, Germany [41]Department of Gynecology and Gynecologic Oncology, Kliniken Essen-Mitte, 45136, Essen, Germany [42]Department of Gynecology and Obstetrics, Jena University Hospital, 07743, Jena, Germany [43]Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, CB1 8RN, UK [44]Faculty of Medicine, University of Southampton, University Hospital Southampton, SO17 1BJ, UK [45]Maggee Women's Hospital, Pittsburg, PA, USA, 15213 [46]Institute of Human Genetics, Friedrich-Alexander-University Erlangen-Nuremberg, 91054, Erlangen, Germany [47]Department of Medicine, Division of Hematology and Oncology, David Geffen School of Medicine, University of California at Los Angeles, Los Angeles, CA, USA, 90095 [48]Sections of Epidemiology and Genetics at the Institute of Cancer Research and Breakthrough Breast Cancer Research Centre, London, UK, SW7 3RP [49]Gynaecological Cancer Research Centre, UCL EGA Institute for Women's Health, London, NW1 2BU, United Kingdom [50]Department of Epidemiology and Preventive Medicine, Monash University, Melbourne, VIC 3806, Australia [51]The Beatson West of Scotland Cancer Centre, Glasgow, G12 0YN, UK [52]Department of Women's Oncology, Moffitt Cancer Center, Tampa, FL, USA, 33612 [53]Samuel Oschin Comprehensive Cancer Center Institute, Cedars Sinai Medical Center, Los Angeles, CA, USA, 90048 [54]Gynecological Oncology Unit, The Royal Marsden Hospital, London, SW3 6JJ, United Kingdom [55]International Hereditary Cancer Center, Department of Genetics and Pathology, Pomeranian Medical University, Szczecin, Poland, 70-115 [56]Department of Epidemiology and Biostatistics, Karolinska Istitutet, Stockholm, Sweden, 171-77 [57]Department of Gynecology and Obstetrics, Haukeland University Hospital, Bergen, HB 5006, Norway [58]Department of Clinical Medicine, University of

Bergen, 5006, Bergen, Norway [59]Clinics of Obstetrics and Gynaecology, Hannover Medical School, 30625, Hannover, Germany [60]Department of Biochemistry and Molecular Biology, Oregon Health and Science University, Portland, OR, USA, 97239 [61]The Juliane Marie Centre, Department of Obstetrics and Gynecology, Rigshospitalet, Copenhagen, 2100, Denmark [62]Department of Pathology, Molecular Unit, Herlev Hospital, University of Copenhagen, Denmark, 2730 [63]Virus, Lifestyle and Genes, Danish Cancer Society Research Center, DK-2100, Copenhagen, Denmark [64]Division of Epidemiology and Prevention, Aichi Cancer Center Research Institute, Nagoya, Aichi, 464-8681, Japan [65]Department of Health Outcomes and Behavior, Moffitt Cancer Center, Tampa, FL, USA, 33612 [66]Department of Medical Oncology, Mayo Clinic, Rochester, MN, USA, 55905 [67]Women's Cancer Program at the Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, CA, USA, 90048 [68]Section of Medicine, Institute of Cancer Research, Sutton, SM2 5NG, UK [69]Department of Popluation Health Research, Alberta Health Services-Cancer Care, Calgary, Alberta, Canada and Departments of Medical Genetics and Oncology, University of Calgary, Calgary, AB, Canada, T2N 2T9 [70]Department of Urology, Radboud University Medical Centre, Nijmegen, HB 6500, Netherlands [71]Department of Obsterics and Gynecology, Nagoya University Graduate School of Medicine, Nagoya, Aichi, 466-8550, Japan [72]Vesalius Research Center, VIB, 3000, Leuven, Belgium [73]Laboratory for Translational Genetics, Department of Oncology, University of Leuven, 3000, Leuven, Belgium [74]Cancer Control Research, BC Cancer Agency, Vancouver, BC, Canada, G12 0YN [75]Gynecology Service, Department of Surgery, Memorial Sloan-Kettering Cancer Center, New York, NY, USA, 10021 [76]College of Pharmacy and Health Sciences, Texas Southern University, Houston, Texas, USA, 77044 [77]Department of Obstetrics and Gynaecology, University Malaya Medical Centre, University Malaya, 59100 Kuala Lumpur, Federal Territory of Kuala Lumpur, Malaysia [78]Department of Epidemiology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA, 77030 [79]Department of Cancer Epidemiology and Prevention, The Maria Sklodowska-Curie Memorial Cancer Center, 02-781, Warsaw, Poland [80]Department of Gynecologic Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA, 77030 [81]Cancer Epidemiology Program, University of Hawaii Cancer Center, Hawaii, USA, 96813 [82]Department of Gynaecology, Radboud University Medical Centre, Nijmegen, HB 6500, Netherlands [83]Department of Health Research and Policy - Epidemiology, Stanford University School of Medicine, Stanford CA, USA, 94305 [84]Dalla Lana School of Public Health, Faculty of Medicine, University of Toronto, ON, M5T 3M7, Canada [85]Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Ontario, Canada, M5G 1X5 [86]Women's Cancer Research Program, Magee-Womens Research Institute and University of Pittsburgh Cancer Institute, Pittsburgh, PA, USA, 15213 [87]Department of Cancer Prevention and Control, Roswell Park Cancer Institute, Buffalo, NY, USA, 14263 [88]Department of Gynecologic Oncology, Aichi Cancer Center Central Hospital, Nagoya, Aichi, Nagoya, 464-8681, Japan [89]Women's College Research Institute, University of Toronto, Toronto, Ontario, Canada, M5G 1N8 [90]Department

of Pathology, Rigshospitalet, University of Copenhagen, 2100, Denmark [91]The University of Texas School of Public Health, Houston, TX, USA, 77030 [92]USC Epigenome Center, Keck School of Medicine, University of Southern California, Norris Comprehensive Cancer Center, Los Angeles, CA, 90089 [93]Department of Gynecologic Oncology, Roswell Park Cancer Institute, Buffalo, NY, USA, 14263 [94]Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, New York, NY, USA, 10065 [95]Department of Obstetrics and Gynecology, Oregon Health and Science University, Portland, OR, USA, 97239 [96]Knight Cancer Institute, Oregon Health and Science University, Portland, OR, USA, 97239 [97]Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, USA, 02115 [98]Channing Laboratory, Harvard Medical School and Brigham and Women's Hospital, Boston, MA, USA, 02115 [99]Department of Epidemiology and Public Health, Yale University School of Public Health and School of Medicine, New Haven, CT, USA, 06520 [100]Program in Epidemiology, Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA, 98109 [101]Department of Epidemiology, University of Washington, Seattle, WA, USA, 98109 [102]Institut für Humangenetik Wiesbaden, 65187, Wiesbaden, Germany [103]Department of Laboratory Medicine and Pathology, Division of Anatomic Pathology, Mayo Clinic, Rochester, MN, USA, 55905 [104]Genetic Epidemiology Laboratory, Department of Pathology, The University of Melbourne, Melbourne, VIC 3053, Australia [105]Department of Gynecologic Oncology, The Maria Sklodowska-Curie Memorial Cancer Center and Institute of Oncology, Warsaw, Poland, 02-781 [106]Pediatrics Epidemiology Center, College of Medicine, University of South Florida, Tampa, FL, USA, 33612 [107]Department of Obstetrics and Gynecology, University of Ulm, Ulm, 89081, Germany [108]Department of Community and Family Medicine, Duke University Medical Center, Durham, NC, USA, 27708 [109]Department of Obstetrics and Gynecology, Mayo Clinic, Rochester, MN, USA, 55905 [110]Shanghai Cancer Institute, Shanghai, China, 2200-25 [111]Department of Epidemiology, Center for Cancer Genetics Research and Prevention, School of Medicine, University of California Irvine, Irvine, California, USA, 92697 [112]Department of Statistical Science, Duke University, Durham, NC, USA, 27708 [113]Cancer Prevention, Detection and Control Research Program, Duke Cancer Institute, Durham, North Carolina, USA, 27708-0251 [114]Department of Obstetrics and Gynecology, Duke Comprehensive Cancer Center, Durham, NC, USA, 27708 [115]Department of Biostatistics, University of Kansas Medical Center, Kansas City, KS, USA, 66160

Acknowledgments

We thank all the individuals who took part in this study and all the researchers, clinicians and administrative staff who have made possible the many studies contributing to this work. In particular, we thank: D. Bowtell, P Webb, A. deFazio, D. Gertig, A. Green, P. Parsons, N. Hayward, and D. Whiteman (AUS); D. L. Wachter, S. Oeser, S. Landrith (BAV); G. Peuteman, T. Van Brussel and D. Smeets (BEL); the staff of the genotyping unit, S LaBoissière and F Robidoux (McGill University and Génome Québec Innovation Centre); U. Eilber and T. Koehler (GER); L. Gacucova (HMO); P. Schu¨rmann, F. Kramer, T.-W. Park-Simon, K. Beer-Grondke and D. Schmidt (HJO); G.L. Keeney, 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, U. Chandran, A. Samoila, and Y. Bensman (NJO); M. Insua and R. Evey (Moffitt); M. Sherman, A. Hutchinson, N. Szeszenia- Dabrowska, B. Peplonska, W. Zatonski, A. Soni, P. Chao and M. Stagner (POL); C. Luccarini, P. Harrington the SEARCH team and ECRIC (SEA); the Scottish Gynaecological Clinical Trails group and SCOTROC1 investigators (SRO); W-H. Chow, Y-T. Gao, G. Yang, B-T. Ji (SWH); I. Jacobs, M. Widschwendter, E. Wozniak, N. Balogun, A. Ryan and J. Ford (UKO); M. Notaridou (USC); C. Pye (UKR); and V. Slusher (U19).

The COGS project is funded through a European Commission's Seventh Framework Programme grant (agreement number 223175 - HEALTH-F2-2009-223175). 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). The scientific development and funding for this project were in part supported by the US National Cancer Institute (R01-CA-114343 and R01-CA114343-S1) and the Genetic Associations and Mechanisms in Oncology (GAME-ON): a NCI Cancer Post-GWAS Initiative (U19-CA148112).

This study made use of data generated by the Wellcome Trust Case Control consortium. A full list of the investigators who contributed to the generation of the data is available from http://www.wtccc.org.uk/. Funding for the project was provided by the Wellcome Trust under award 076113. The results published here are in part based upon data generated by The Cancer Genome Atlas Pilot Project established by the National Cancer Institute and National Human Genome Research Institute. Information about TCGA and the investigators and institutions who constitute the TCGA research network can be found at http://cancergenome.nih.gov/.

D.F.E. is a Principal Research Fellow of Cancer Research UK G.C.-T. and P.M.W. are supported by the National Health and Medical Research Council. BK holds an American Cancer Society Early Detection Professorship (SIOP-06-258-01-COUN). LEK is supported by a Canadian Institutes of Health Research Investigator award (MSH-87734). A.C.A. is Cancer Research-UK Senior Cancer Research Fellow.

Funding of the constituent studies was provided by the American Cancer Society (CRTG-00-196-01-CCE); the California Cancer Research Program (00-01389V-20170, N01-CN25403, 2II0200); the Canadian Institutes for Health Research (MOP-86727); 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, C536/A13086, C536/A6689); the Celma Mastry Ovarian Cancer Foundation; the Danish Cancer Society (94-222-52); the Norwegian Cancer Society, Helse Vest, the Norwegian Research Council; ELAN Funds of the University of Erlangen-Nuremberg; the Eve Appeal; the Helsinki University Central Hospital Research Fund; Imperial Experimental Cancer Research Centre (C1312/A15589); the Ovarian Cancer Research Fund; Nationaal 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 Radboud University Nijmegen Medical Centre; the Polish Ministry of Science and Higher Education (4 PO5C 028 14, 2 PO5A 068 27); the Roswell Park Cancer Institute Alliance Foundation; the US National Cancer Institute (K07-CA095666, K07-CA143047, K22-CA138563, N01-CN55424, N01-PC067001, N01-PC035137, P01-CA017054, P01-CA087696, P50-CA105009, P50-CA136393, R01-CA014089, R01-CA016056, R01-CA017054, R01-CA049449, R01-CA050385, R01-CA054419, R01-CA058598, R01-CA058860, R01-CA061107, R01-CA061132, R01-CA063682, R01-CA064277, R01-CA067262, R01-CA071766, R01-CA074850, R01-CA076016, R01-CA080742, R01-CA080978, R01-CA087538, R01-CA092044, R01-095023, R01-CA106414, R01-CA122443, R01-CA136924, R01-CA112523, R01-CA114343, R01-CA126841, R01-CA149429, R03-CA113148, R03-CA115195, R37-CA070867, R37-CA70867, R01-CA83918, U01-CA069417, U01-CA071966, P30-CA15083, PSA 042205, and Intramural research funds); the US Army Medical Research and Materiel Command (DAMD17-98-1-8659, DAMD17-01-1-0729, DAMD17-02-1-0666, DAMD17-02-1-0669, W81XWH-10-1-02802); the Department of Defense Ovarian Cancer Research Program (W81XWH-07-1-0449); the National Health and Medical Research Council of Australia (199600 and 400281); the German Federal Ministry of Education and Research of Germany Programme of Clinical Biomedical Research (01 GB 9401); the state of Baden-Württemberg through Medical Faculty of the University of Ulm (P.685); the German Cancer Research Center; Pomeranian Medical University; the Minnesota Ovarian Cancer Alliance; the Mayo Foundation; the Fred C. and Katherine B. Andersen Foundation; the Malaysian Ministry of Higher Education (UM.C/HlR/MOHE/06) and Cancer Research Initiatives 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.

Access to genotype data for SNPs that were not nominated by OCAC was provided by FJC and ACA on behalf of CIMBA BRCA1 GWAS investigators and was funded by the US National Cancer Institute (R01 CA128978) and U.S. Department of Defense Ovarian Cancer Idea award (W81XWH-10-1-0341). CIMBA BRCA1 GWAS investigators include: Irene L. Andrulis, Ontario Cancer Genetics Network, Cancer Care Ontario and Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Canada; Trinidad Caldes, Molecular Oncology Laboratory, Hospital Clínico San Carlos, Madrid, Spain; Maria Adelaide Caligo, Section of Genetic Oncology, University Hospital of Pisa, Pisa, Italy; Olga Sinilnikova for GEMO Study Collaborators, Cancer Genetics Network

"Groupe Genetique et Cancer", Federation Nationale des Centres de Lutte Contre le Cancer, Lyon, France; Thomas V. O. Hansen, Genomic Medicine, Department of Clinical Biochemistry, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark; Matti Rookus and Frans Hogervorst for HEBON Investigators, Department of Epidemiology, The Netherlands Cancer Institute, Amsterdam, The Netherlands; Anna Jakubowska and Jan Lubinski, International Hereditary Cancer Center, Department of Genetics and Pathology, Pomeranian Medical University, Szczecin, Poland; Susan Peock for EMBRACE Investigators, Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Strangeways Research Laboratory, Worts Causeway, Cambridge, United Kingdom; Georgia Chenevix-Trench for kConFab Investigators, Peter MacCallum Cancer Centre, East Melbourne, Victoria, Australia; BCFR Investigators, Breast Cancer Family Registry, Epidemiology and Genetics Research Program, DCCPS, National Cancer Institute, Rockville, Maryland, USA; Katherine L. Nathanson and Susan Domchek, Departments of Medicine and Medical Genetics and Abramson Cancer Center, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, USA; Heli Nevanlinna, Department of Obstetrics and Gynecology, Helsinki University Central Hospital, Helsinki, Finland; Kenneth Offit, Memorial Sloan Kettering Cancer Center, New York, New York, USA; Ana Osorio and Javier Benitez, Human Genetics Group, Human Cancer Genetics Programme, Spanish National Cancer Centre, Madrid, Spain; Paolo Radice for CONSIT investigators, Unit of Molecular Bases of Genetic Risk and Genetic Testing, Department of Preventive and Predictive Medicine, Fondazione IRCCS Istituto Nazionale Tumori (INT); IFOM, Fondazione Istituto FIRC di Oncologia Molecolare, Milan, Italy; Christian F. Singer, Division of Special Gynecology, Medical University of Vienna, Vienna, Austria; SWE-BRCA Investigators, Karolinska University Hospital, Stockholm, Sweden; Rita Schmutzler for GC-HBOC investigators, Center of Familial Breast and Ovarian Cancer, Department of Obstetrics and Gynaecology and Center for Integrated Oncology (CIO), University of Cologne, Cologne, Germany; Andrew Godwin, Department of Pathology and Laboratory Medicine, University of Kansas Medical Center, Kansas City, Kansas; Ignacio Blanco and Conxi Lazaro, Hereditary Cancer Program, Instituto Catalan de Oncología, Barcelona, Spain; Marco Montagna, Immunology and Molecular Oncology Unit, Istituto Oncologico Veneto IOV–IRCCS, Padua, Italy; Mary S. Beattie, Cancer Risk Program, Departments of Medicine, Epidemiology, and Biostatistics, University of California at San Francisco, San Francisco, California, USA; Antonis C. Antoniou and Douglas F. Easton, Centre for Cancer Genetic Epidemiology, University of Cambridge, Cambridge, United Kingdom; Fergus J. Couch, Mayo Clinic, Rochester, Minnesota, USA.

References

1. Freedman ML, et al. Principles for the post-GWAS functional characterization of cancer risk loci. Nat Genet. 2011; 43:513–518. [PubMed: 21614091]
2. Dahiya N, Morin PJ. MicroRNAs in ovarian carcinomas. Endocr Relat Cancer. 2010; 17:F77–89. [PubMed: 19903743]
3. Lytle JR, Yario TA, Steitz JA. Target mRNAs are repressed as efficiently by microRNA-binding sites in the 5' UTR as in the 3' UTR. Proc Natl Acad Sci U S A. 2007; 104:9667–9672. [PubMed: 17535905]
4. Lee I, et al. New class of microRNA targets containing simultaneous 5'-UTR and 3'-UTR interaction sites. Genome Res. 2009; 19:1175–1183. [PubMed: 19336450]
5. Permuth-Wey J, et al. LIN28B polymorphisms influence susceptibility to epithelial ovarian cancer. Cancer Res. 2011
6. Permuth-Wey J, et al. MicroRNA processing and binding site polymorphisms are not replicated in the Ovarian Cancer Association Consortium. Cancer Epidemiol Biomarkers Prev. 2011; 20:1793–1797. [PubMed: 21636674]
7. Liang D, et al. Genetic variants in MicroRNA biosynthesis pathways and binding sites modify ovarian cancer risk, survival, and treatment response. Cancer Res. 2010; 70:9765–9776. [PubMed: 21118967]
8. Ryan BM, Robles AI, Harris CC. Genetic variation in microRNA networks: the implications for cancer research. Nat Rev Cancer. 2010; 10:389–402. [PubMed: 20495573]
9. Sethupathy P, Collins FS. MicroRNA target site polymorphisms and human disease. Trends Genet. 2008; 24:489–497. [PubMed: 18778868]
10. Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell. 2005; 120:15–20. [PubMed: 15652477]
11. TargetScanHuman. 2009. http://genes.mit.edu/targetscan
12. Krek A, et al. Combinatorial microRNA target predictions. Nat Genet. 2005; 37:495–500. [PubMed: 15806104]
13. PicTar. 2009. pictar.mdc-berlin.de/

14. Stefansson H, et al. A common inversion under selection in Europeans. Nat Genet. 2005; 37:129–137. [PubMed: 15654335]
15. 1,000 Genomes. 2012. http://www.1000genomes.org/page.php
16. Couch FJ, Wang X, McGuffog L, Lee A, Olswold C, Kuchenbacecker KB, et al. Genome-wide association study in BRCA1 mutation carriers identifies novel loci associated with breast and ovarian cancer risk. PLoS Genetics. XYZ. in press.
17. Goode EL, et al. A genome-wide association study identifies susceptibility loci for ovarian cancer at 2q31 and 8q24. Nat Genet. 2010
18. Integrated genomic analyses of ovarian carcinoma. Nature. 2011; 474:609–615. [PubMed: 21720365]
19. Mok SC, et al. A gene signature predictive for outcome in advanced ovarian cancer identifies a survival factor: microfibril-associated glycoprotein 2. Cancer Cell. 2009; 16:521–532. [PubMed: 19962670]
20. Forbes SA, et al. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. Nucleic Acids Res. 2011; 39:D945–950. [PubMed: 20952405]
21. Tabata K, et al. Rubicon and PLEKHM1 negatively regulate the endocytic/autophagic pathway via a novel Rab7-binding domain. Mol Biol Cell. 2010; 21:4162–4172. [PubMed: 20943950]
22. Katoh Y, Katoh M. Identification and characterization of ARHGAP27 gene in silico. Int J Mol Med. 2004; 14:943–947. [PubMed: 15492870]
23. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010; 38:e164. [PubMed: 20601685]
24. Xu Z, Taylor JA. SNPinfo: integrating GWAS and candidate gene information into functional SNP selection for genetic association studies. Nucleic Acids Res. 2009; 37:W600–605. [PubMed: 19417063]
25. Dayem Ullah AZ, Lemoine NR, Chelala C. SNPnexus: a web server for functional annotation of novel and publicly known genetic variants (2012 update). Nucleic Acids Res. 2012
26. JASPAR. 2012. http://jaspar.cgb.ki.se/
27. Ray D, et al. Characterization of Spi-B, a transcription factor related to the putative oncoprotein Spi-1/PU.1. Molecular and cellular biology. 1992; 12:4297–4304. [PubMed: 1406622]
28. Fujimoto J, et al. Clinical implications of expression of ETS-1 related to angiogenesis in metastatic lesions of ovarian cancers. Oncology. 2004; 66:420–428. [PubMed: 15331930]
29. Spaderna S, et al. The transcriptional repressor ZEB1 promotes metastasis and loss of cell polarity in cancer. Cancer research. 2008; 68:537–544. [PubMed: 18199550]
30. Peinado H, Olmeda D, Cano A. Snail, Zeb and bHLH factors in tumour progression: an alliance against the epithelial phenotype? Nature reviews Cancer. 2007; 7:415–428. [PubMed: 17508028]
31. Bendoraite A, et al. Regulation of miR-200 family microRNAs and ZEB transcription factors in ovarian cancer: evidence supporting a mesothelial-to-epithelial transition. Gynecol Oncol. 2010; 116:117–125. [PubMed: 19854497]
32. Bolton KL, et al. Common variants at 19p13 are associated with susceptibility to ovarian cancer. Nat Genet. 2010
33. Song H, et al. A genome-wide association study identifies a new ovarian cancer susceptibility locus on 9p22.2. Nat Genet. 2009; 41:996–1000. [PubMed: 19648919]
34. Pharoah, et al. GWAS meta-analysis and replication identifies three novel common susceptibility loci for ovarian cancer. Nat Genet. XYZ. in press.
35. Couch FJ, et al. Common variants at the 19p13.1 and ZNF365 loci are associated with ER subtypes of breast cancer and ovarian cancer risk in BRCA1 and BRCA2 mutation carriers. Cancer Epidemiol Biomarkers Prev. 2012; 21:645–657. [PubMed: 22351618]
36. Ghoussaini M, et al. Multiple loci with different cancer specificities within the 8q24 gene desert. J Natl Cancer Inst. 2008; 100:962–966. [PubMed: 18577746]
37. Coppola G, et al. Evidence for a Role of The Rare p.A152T Variant in MAPT in increasing the Risk for FTD-Spectrum and Alzheimer's Diseases. Hum Mol Genet. 2012
38. Ghidoni R, et al. The H2 MAPT haplotype is associated with familial frontotemporal dementia. Neurobiol Dis. 2006; 22:357–362. [PubMed: 16410051]

39. Koolen DA, et al. A new chromosome 17q21.31 microdeletion syndrome associated with a common inversion polymorphism. Nat Genet. 2006; 38:999–1001. [PubMed: 16906164]
40. Tobin JE, et al. Haplotypes and gene expression implicate the MAPT region for Parkinson disease: the GenePD Study. Neurology. 2008; 71:28–34. [PubMed: 18509094]
41. Li R BF, Kiefer AK, Steffanson H, Nyholt DR, et al. Six Novel Susceptibility Loci for Early-Onset Androgenic Alopecia and Their Unexpected Association with Common Diseases. PLoS Genet. 2012; 8(5):e1002746. [PubMed: 22693459]
42. Edwards TL, et al. Genome-wide association study confirms SNPs in SNCA and the MAPT region as common risk factors for Parkinson disease. Ann Hum Genet. 2010; 74:97–109. [PubMed: 20070850]
43. Dahiya N, et al. MicroRNA expression and identification of putative miRNA targets in ovarian cancer. PLoS ONE. 2008; 3:e2436. [PubMed: 18560586]
44. Iorio MV, et al. MicroRNA signatures in human ovarian cancer. Cancer Res. 2007; 67:8699–8707. [PubMed: 17875710]
45. Nam EJ, et al. MicroRNA expression profiles in serous ovarian carcinoma. Clin Cancer Res. 2008; 14:2690–2695. [PubMed: 18451233]
46. Yang H, et al. MicroRNA expression profiling in human ovarian cancer: miR-214 induces cell survival and cisplatin resistance by targeting PTEN. Cancer Res. 2008; 68:425–433. [PubMed: 18199536]
47. Tothill RW, et al. Novel molecular subtypes of serous and endometrioid ovarian cancer linked to clinical outcome. Clin Cancer Res. 2008; 14:5198–5208. [PubMed: 18698038]
48. Zorn KK, et al. Gene expression profiles of serous, endometrioid, and clear cell subtypes of ovarian and endometrial cancer. Clin Cancer Res. 2005; 11:6422–6430. [PubMed: 16166416]
49. Landen CN Jr, Birrer MJ, Sood AK. Early events in the pathogenesis of epithelial ovarian cancer. J Clin Oncol. 2008; 26:995–1005. [PubMed: 18195328]
50. Fasching PA, et al. Role of genetic polymorphisms and ovarian cancer susceptibility. Mol Oncol. 2009; 3:171–181. [PubMed: 19383379]
51. Yu W, Wulf A, Liu T, Khoury MJ, Gwinn M. Gene Prospector: an evidence gateway for evaluating potential susceptibility genes and interacting risk factors for human diseases. BMC Bioinformatics. 2008; 9:528. [PubMed: 19063745]
52. NCBI dbSNP database. 2009. http://ncbi.nlm.nih.gov/SNP
53. Sankararaman S, Sridhar S, Kimmel G, Halperin E. Estimating local ancestry in admixed populations. Am J Hum Genet. 2008; 82:290–303. [PubMed: 18252211]
54. Purcell S, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007; 81:559–575. [PubMed: 17701901]
55. Howie B, Marchini J, Stephens M. Genotype imputation with thousands of genomes. G3 (Bethesda). 2011; 1:457–470. [PubMed: 22384356]
56. Lawrenson K, et al. Senescent fibroblasts promote neoplastic transformation of partially transformed ovarian epithelial cells in a three-dimensional model of early stage ovarian cancer. Neoplasia. 2010; 12:317–325. [PubMed: 20360942]
57. Ernst J, et al. Mapping and analysis of chromatin state dynamics in nine human cell types. Nature. 2011; 473:43–49. [PubMed: 21441907]
58. Heintzman ND, et al. Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature. 2009; 459:108–112. [PubMed: 19295514]

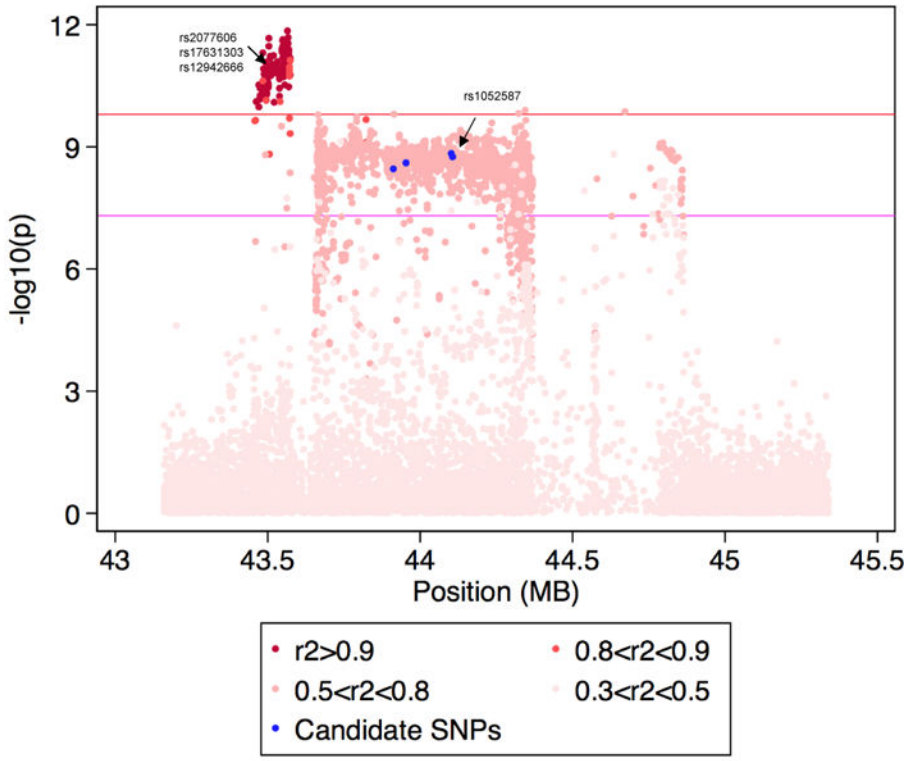

Figure 1. Regional association plot for genotyped and imputed SNPs at 17q21.31

The middle portion of the plot contains the region of the inversion polymorphism (ch 17: 43,624,578-44,525,051, hg build 37), with the four blue dots representing the candidate miRSNPs (rs4640231, rs1052587, and rs17574361) and the tagSNP, rs916793. rs1052587 in the 3′UTR of *MAPT* has the strongest signal ($P=4.6\times10^{-8}$) among the miRSNPs. The cluster on the left side of the plot (around 43.5 MB) contains highly correlated SNPs ($r^2=0.99$), including three directly genotyped intronic SNPs, rs2077606 and rs17631303 in *PLEKHM1* ($P=3.9\times10^{-10}$ and $P=4.7\times10^{-10}$, respectively), and rs12942666 in ARHGAP27 ($P=1.0\times10^{-9}$). The linkage disequilibrium between each plotted SNP and the top-ranked SNP in the region with the best clustering, rs12942666, is depicted by the color scheme; the deeper the color red, the stronger the correlation between the plotted SNP and rs12942666. The top miRSNP, rs1052587, is moderately correlated ($r^2=0.76$) with rs2077606, rs17631303, and rs12942666 in our study population. (n=8,371 invasive serous cases and n= 23,491 controls, of European ancestry).

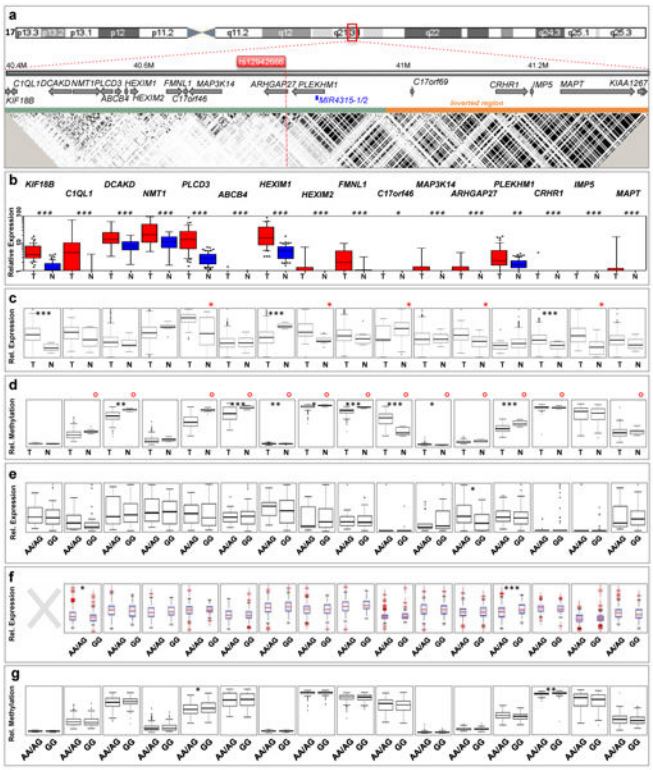

Figure 2. Expression and methylation analyses at the 17q21.31 ovarian cancer susceptibility locus

(a) Genomic map and LD structure. The location and approximate size of 17 known protein coding genes (grey) and one microRNA (blue) in the region are shown relative to the location of rs12942666. Orange indicates the location of the inversion polymorphism, and green indicates the region outside the inversion.

(b) Gene expression (EOC and normal cell lines). Gene expression analysis in Epithelial Ovarian Cancer (EOC) cell lines (T; n=51) compared to normal ovarian surface epithelial cells (OSECs) and fallopian tube secretory epithelial cells (FTSEC) (N; n=73) (* $p<0.05$, ** $p<0.01$, *** $p<0.001$).

(c) Gene expression (Primary EOCs and Normal Tissue). Boxplots of The Cancer Genome Atlas (TCGA) Affymetrix U133A-array based gene expression in primary high-grade serous ovarian tumors (T; n=568) and normal fallopian tube tissues (N; n=8). Where data were not available in TCGA, gene expression data from the Gene Expression Omnibus series GSE18520 dataset containing 53 high-grade serous tumors and 10 normal ovarian tissues are shown (indicated by a red asterisk).

(d) Methylation (Primary Tumors and Normal Tissue). Methylation analysis of 106 high-grade serous ovarian tumors compared to normal ovarian tissues (n=7). Methylation data were generated for CpG site(s) associated with each gene using the Illumina 450 methylation array. Pairwise analysis of methylation for an individual CpG for each gene is based on the CpG with most significant inverse relationship to gene expression (i.e. cis negative), for a subset of 43 tumors having available gene expression data. Statistically significant cis-negative probes are indicated by a red open circle.

(e) Expression quantitative trait locus (eQTL) analysis (OSECs/FTSECs). eQTL analysis comparing expression for each gene to genotype for the most statistically significant SNP at 17q21.31 (rs12942666), for 73 normal OSEC/FTSEC lines. Data are presented as box plots comparing expression levels in cases carrying rare homozygotes/heterozygotes, with cases homozygous for the common allele.

(f) Expression quantitative trait locus (eQTL) analysis (Primary EOCs). eQTL analysis comparing expression for each gene to genotype using level 3 gene expression profiling data from Agilent 244K custom arrays and level 2 genotype data from the Illumina 1M-Duo BeadChip for 568 high-grade serous ovarian cancer patients from TCGA. In all panels * $p<0.05$, ** $p<0.01$, *** $p<0.001$. Grey X's indicate data not available. Here, genotype data for rs2077606 is used (rather than rs12942666) because rs12942666 was not genotyped in the TCGA dataset.

(g) Methylation quantitative trait locus (mQTL) analysis (Primary EOCs). mQTL analysis showing methylation status in 227 high-grade serous EOCs relative to rs12942666 genotype.

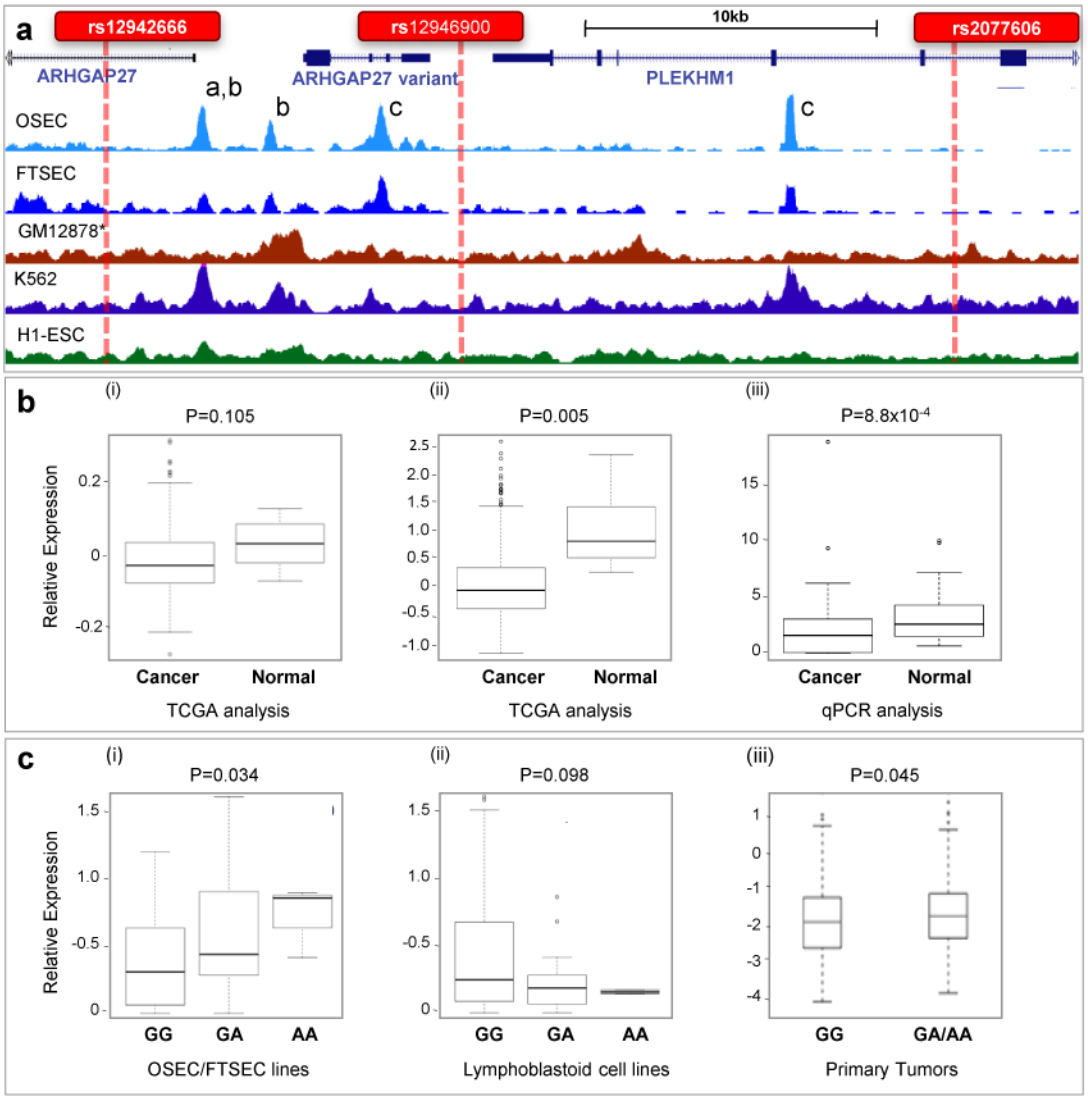

Figure 3. eQTL associations between the rs2077606 susceptibility SNP at 17q21

(a) Analysis of the chromatin landscape at *ARHGAP27* and *PLEKHM1* in normal ovarian surface epithelial and fallopian tube secretory epithelial cells (OSECs/FTSECs) by formaldehyde assisted isolation of regulatory elements sequencing (FAIRE-seq). Alignment with ENCODE FAIRE-seq tracks (shown) and ChIP-seq tracks (not shown) from non-EOC related cell lines reveals open chromatin peaks corresponding to (a) promoters (b) CTCF insulator binding sites and (c) H3K4me3 signals, suggestive of a dynamic regulatory region. An H3K4me3 signal at a coding *ARHGAP27* mRNA variant (c) located between the genes is highly pronounced in OSEC/FTSEC, suggesting tissue-specific expression and function. Several of the top-ranking SNPs fall within transcription factor binding sites (TFFS) (Supplementary Table S2). rs12942666 did not coincide with TFBS, but tightly linked SNPs, rs12946900 and rs2077606 fell within predicted binding sites for SPIB and ZEB1, respectively.

(b) We analyzed the expression of *SPIB* and *ZEB1* in primary high-grade serous tumors from TCGA and found (i) no significant change in SPIB expression but (ii) significant

down-regulation of *ZEB1* in tumors compared to normal tissues. (iii) QPCR analysis of *ZEB1* expression in 73 OCPT and 50 EOC cell lines replicated the finding that *ZEB1* expression is lower in cancer cell lines compared to normal precursor tissues. (c) eQTL analysis in OSECs/FTSECs for different alleles of rs2077606. There was a (i) significant eQTL for *ARHGAP27*, with the minor (A) allele being associated with increased *ARHGAP27* expression (P=0.034), (ii) no evidence of an association between rs2077606 genotypes and *ARHGAP27* expression in lymphoblastoid cell lines suggesting this association may be tissue-specific. (iii) We observed a borderline significant eQTL association between *ZEB1* mRNA and rs2077606 in tumors from TCGA, with the minor risk allele also associated with lower expression.

Nat Commun. Author manuscript; available in PMC 2013 July 12.

Table 1

Tests of association by histological subtype for directly genotyped and imputed SNPs at 17q21.31 most strongly associated with invasive epithelial ovarian cancer risk among Europeans

SNP Major>minor allele allele	Coordinate[a]	MAF	Subtype	Number of cases (versus 23,491 controls)	Per-allele OR (95% CI)[b]	P-value
rs1052587[c] (T>C)	44102604	0.22	All Invasives	14,533	1.10 (1.06-1.13)	1.9×10^{-7}
			Serous	8,371	1.12 (1.08-1.17)	4.6×10^{-8}
			Endometrioid	2,068	1.11 (1.04-1.19)	5.2×10^{-3}
			Clear Cell	1,025	0.98 (0.88-1.09)	0.68
			Mucinous	944	1.07 (0.96-1.20)	0.22
rs12942666[d] (A>G)	43499839	0.22	All Invasives	14,533	1.11 (1.07-1.15)	3.3×10^{-8}
			Serous	8,371	1.15 (1.11-1.20)	1.0×10^{-9}
			Endometrioid	2,068	1.10 (1.02-1.18)	0.04
			Clear Cell	1,025	1.04 (0.92-1.14)	0.61
			Mucinous	944	1.04 (0.92-1.16)	0.55
rs2960000[e] (T>C)	43534353	0.18	All Invasives	14,533	1.12 (1.08-1.16)	4.2×10^{-9}
			Serous	8,371	1.16 (1.12-1.20)	3.3×10^{-10}
			Endometrioid	2,068	1.12 (1.03-1.20)	0.01
			Clear Cell	1,025	1.05 (0.93-1.16)	0.44
			Mucinous	944	1.03 (0.90-1.15)	0.65

Abbreviations: MAF=minor allele frequency in controls; OR=Odds ratio; CI=Confidence intervals

[a] Genome build NCBI B37/human genome build 19 assembly.

[b] OR and 95% CI per copy of the minor allele, with adjustment for the first five eigenvalues from principal components analysis.

[c] rs1052587 is the most statistically significant miRNA binding site SNP among all invasives and serous; it resides in a putative miRNA binding site between microtubule-associated protein tau (*MAPT*) and miR- 34a-5p (chr 1:9134225-9134425).

[d] rs12942666 is a SNP at 17q21.31 that was directly genotyped as part of COGs; it is in strong linkage disequilibrium (r^2=0.99) with two other 17q21.31 SNPs that were directly genotyped but had less optimal clustering: rs2077606 ($P=3.9 \times 10^{-10}$ for the serous subtype) and rs17631303 ($P=4.7 \times 10^{-10}$ for the serous subtype).

[e] rs2960000 represents the most statistically significant SNP at 17q21.31 (among all invasives) that was imputed from the 1000 genome Project reference panel with an R-squared quality metric of 95% or greater (http://www.1000genomes.org/page.php).