Title

GWAS meta-analysis and replication identifies three new susceptibility loci for ovarian cancer.

Permalink

https://escholarship.org/uc/item/200214g1

Journal

Nature Genetics, 45(4)

Authors

Antonenkova, Natalia
Armasu, Sebastian
Baglietto, Laura
et al.

Publication Date

2013-04-01

DOI

10.1038/ng.2564

Peer reviewed

GWAS meta-analysis and replication identifies three new susceptibility loci for ovarian cancer

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

Abstract

Genome wide association studies (GWAS) have identified four susceptibility loci for epithelial ovarian cancer (EOC) with another two loci being close to genome-wide significance. We pooled data from a GWAS conducted in North America with another GWAS from the United Kingdom. We selected the top 24,551 SNPs for inclusion on the iCOGS custom genotyping array. Follow-up genotyping was carried out in 18,174 cases and 26,134 controls from 43 studies from the Ovarian Cancer Association Consortium. We validated the two loci at 3q25 and 17q21 previously near genome-wide significance and identified three novel loci associated with risk; two loci associated with all EOC subtypes, at 8q21 (rs11782652, $P=5.5\times10^{-9}$) and 10p12 (rs1243180; $P=1.8\times10^{-8}$), and another locus specific to the serous subtype at 17q12 (rs757210; $P=8.1\times10^{-10}$). An integrated molecular analysis of genes and regulatory regions at these loci provided evidence for functional mechanisms underlying susceptibility that implicates *CHMP4C* in the pathogenesis of ovarian cancer.

Evidence from twin and family studies suggests an inherited genetic component to EOC risk [1,2]. Rare, high-penetrance allele of genes such as *BRCA1* and *BRCA2* account for about 40 percent of the excess familial risk [3] and GWAS have recently identified common risk alleles at 9p22, 8q24, 2q31, and 19p13 [4-6] with two additional loci at 3q25 and 17q21 that approached genome-wide significance [6]. However these only explain 4 percent of the excess familial risk, and more loci probably exist.

We therefore pooled the data from two GWAS to inform the selection of SNPs for a large-scale replication. The North American study comprised four independent case-control studies that included 1,952 EOC cases and 2,052 controls. The second study was a two-phase multi-center GWAS that included 1,817 EOC cases and 2,354 controls in the first phase and 4,162 EOC cases and 4,810 controls in the second phase. We carried out a fixed effects meta-analysis from the two GWAS for ~2.5 million genotyped or imputed SNPs. We selected 24,551 SNPs associated with the risk of either all-histology (11,647 SNPs) or serous ovarian cancer (12,904 SNPs) based on ranked *P*-values. Assays were designed for 23,239 SNPs and included on a custom Illumina Infinium iSelect array ("iCOGS") comprising 211,155 SNPs designed by the Collaborative Oncological Gene-environment Study (COGS) to evaluate genetic variants for association with the risk of breast, ovarian and prostate cancer. These SNPs were then genotyped in cases and controls from 43 individual studies from the Ovarian Cancer Association Consortium (OCAC) that were grouped into 34 case-control strata (**Table 1** and **Supplementary tables 1 and 2**). These included most of the samples genotyped in the initial GWAS.

Results

Association analyses

After applying quality control filters (see **Online methods**), we tested 22,252 SNPs for association with risk of all invasive EOC and for serous invasive EOC in 18,174 EOC cases (10,316 serous cases) and 26,134 controls. The primary analyses were based on the subjects of European ancestry (16,283 cases and 23,491 controls).

The associations of the four SNPs at 2q31, 8q24, 9p22 and 19p13 previously reported at genome-wide significance were all confirmed (**Supplementary table 3**). SNPs at the two other loci previously reported near to genome-wide significant (at 3q25 and 17q21)[6] were also confirmed. The associated SNP at 3q25 reported by Goode and colleagues (rs2665390) failed design, but a correlated SNP, rs7651446 (r^2=0.61) was highly significantly associated with invasive EOC (effect allele frequency (EAF)=0.050, per-allele OR = 1.44, 95% CI: 1.35-1.53, P=1.5×10^{-28}) as was rs9303542 at 17q21 (EAF=0.27, OR = 1.12, 95% CI: 1.08-1.16, P=6.0×10^{-11}). The Manhattan plots for all invasive EOC and serous invasive EOC, after excluding 176 SNPs from the six known loci, are shown in **Figure 1**. We identified three new loci at genome-wide significance (P<5×10^{-8}), two of which were significant for all invasive EOC (8q21 and 10p12) and another for invasive serous EOC only (17q12) (**Table 2**). The genotype clustering for the top hits at 8q21 (rs11782652) and 17q12 (rs757210) was good, but clustering of the top hit at 10p12 (rs7084454) was poor (**Supplementary Fig. 1**). Clustering for a second, highly correlated SNP (r^2= 0.86) at this locus (rs1243180) was good and so the results for this SNP are presented instead.

The most significant association for all invasive EOC was rs11782652 at 8q21 (odds ratio (OR) = 1.19, 95% CI: 1.12-1.26, P=5.5×10^{-9}). This SNP had been selected for replication because it was associated with all invasive EOC in the combined GWAS data (OR =1.20, 95% CI: 1.07-1.36, P=0.0025). This SNP is not correlated with any other SNP in HapMap (**Supplementary fig. 2**) and was the only SNP in the selected for genotyping in COGS. Effects varied by histological subtype (P=0.0002), with the strongest effect in the serous

subtype (**Table 2**). There was little evidence for heterogeneity of the association by ancestry (P=0.55) or between the 31 European studies included (P=0.13, **Supplementary fig. 3**). Rs1243180 at 10p12 was selected for replication because it was associated with all invasive EOC in the combined GWAS data (OR =1.11, 95% CI: 1.04-1.19, P=0.0027) and was also associated with risk of all invasive EOC in the replication data (OR = 1.10, 95% CI: 1.06-1.13, P=1.8×10^{-8}). There is strong LD in this region and 32 other SNPs in the region were selected for replication (**Supplementary fig. 4**). T here was some heterogeneity of effects by tumor subtype (P=0.0007) but not by study (P=0.65, **Supplementary fig. 5**) or by population (P=0.12). At 17q12, rs757210 was selected for replication because it was associated with serous EOC in the combined GWAS data (OR =1.13, 95% CI: 1.04-1.23, P=0.0026) and was most strongly associated with the serous subtype in the replication data (OR = 1.12, 95% CI 1.08-1.17, P=9.6×10^{-10}). Eight other SNPs in the region were selected for replication in COGS (**Supplementary fig. 6**). The association with all invasive EOC was much weaker (OR = 1.05, 95% CI 1.02-1.09, P=9×10^{-4}); there was substantial heterogeneity by tumor subtype (P=<0.0001) with the risk allele for serous EOC being associated with a reduced risk of both clear cell and mucinous EOC (**Table 2**). Data from a set of fine mapping SNPs genotyped in this region suggest that this apparent paradox is caused by the presence of two independent loci for serous and clear cell cancer and that the top hit at each of these loci is correlated with rs757210 (CL Pearce, personal communication). There was also heterogeneity by ancestry for the serous subtype (P=0.034) with the risk allele associated with a reduced disease risk in the subjects of mixed ancestry origin, and some between-study heterogeneity (P=0.038, **Supplementary fig. 7**).

Functional and molecular analyses

The most significant risk-associated SNPs for the three novel EOC susceptibility loci are located in non-coding DNA sequences, but these may only be markers for the true causal variant(s), which could be functional coding variants, or variants in non-coding DNA elements or non-coding RNAs, influencing the expression of nearby target genes (*cis-*regulatory effects). They may also act on genes through more distal regulation (*trans-*regulatory effects) [7-11]. In order to identify the most likely functional SNP and target gene we evaluated the putative functional role in EOC for all genes in a one-megabase region centred on the most significant risk-SNP at each locus. We used a combination of locus specific and genome-wide assays to characterize the transcribed genes (see **Online methods** and **Supplementary fig. 8)** and regulatory elements (**Supplementary fig. 9**) within susceptibility regions to evaluate putative functional mechanisms and identify candidate EOC susceptibility gene(s) at each locus.

At the 8q21 locus, the strongest associated SNP, rs11782652, is located in the first intron of *CHMP4C*. We imputed genotypes in the region to the 1000GP and tested all variants with MAF>0.02 for association. The log-likelihoods of the regression models were compared and eight SNPs with a log-likelihood within 6.91 of the most strongly associated SNP - equivalent to an odds of 1000:1 - were considered the most likely candidates for being the causal variant. Six of them lie in introns of *CHMP4C* but *in silico* analysis provided functional evidence for only one (rs74544416), which contains a putative SOX9 binding site. One is an indel (4 nucleotides) at the exon-intron border (rs137960856; alleles -/

GTGA) but it is unlikely to have a functional impact because the next four nucleotides are also GTGA. Thus, even in the deleted allele the corrected exonic sequences are retained and it is not expected to affect splicing. The eighth SNP, rs35094336, is predicted to result in a coding change from Ala to Thr that may be functionally relevant (Polyphen2 score: 0.997). This residue is located in a C-terminal amphipatic alpha helix conserved in all CHMP4 proteins and is important for binding to ALIX, a protein involved in the ESCRT[12]. Further studies will be necessary to determine whether this change is of functional significance and has an impact on ovarian cancer biology. ENCODE data from non-ovarian cancer associated tissues, FAIRE-seq data and mapping of enhancer elements generated in normal serous ovarian cancer precursor cells suggests there are two regulatory regions that may be influenced by risk associated SNPs; one at the *CHMP4C* promotor and the other in intron one of *CHMP4C* (**Figure 2**).

We found no evidence of a correlation between rs11782652 genotype and gene expression in normal ovarian/fallopian tube epithelial cells for any of the nine genes in the region (FABP5, PMP2, FABP4, FABP12, IMPA1, SLC10A5, ZFAND1, CHMP4C and *SNX16*); but there was a highly statistically significant association between rs11782652 and *CHMP4C* expression in primary EOCs ($P=3.9\times10^{-14}$) and in transformed lymphocytes (P=0.012). We also found evidence of association for rs11782652 with methylation status (mQTL) for three genes in tumour tissue; *ZFAND1* (P=0.003), *CHMP4C* (P=0.001) and *SNX16* (P=0.001). However, only *CHMP4C* methylation was negative correlated with expression (P=0.036).

Three genes in the region, *FABP5 CHMP4C* and *SNX16*, were significantly overexpressed in both EOC cell lines compared to normal tissues (P=0.002, 4.8×10^{-9} and 5.9×10^{-4} respectively) and, where data were available, in primary EOC tissues. Finally the Catalogue Of Somatic Mutations In Cancer (COSMIC) database showed that four genes in the region *IMPA1, ZFAND1, CHMP4C* and *SNX16* have functionally significant mutations in cancer, with the last three genes mutated in ovarian carcinoma (**Supplementary fig. 10**). These four genes form a highly connected co-expression network across different experimental conditions (**Supplementary fig. 11**).

Taken together these data suggest that several genes at the 8q21 locus may play a role in the somatic development of EOC; but the cumulative evidence indicate that *CHMP4C* (chromatin-modifying protein 4C), is the most likely candidate susceptibility gene. This is supported by previously published data on the function of *CHMP4C*. *CHMP4C* is involved in the final steps of cell division, co-ordinating midbody resolution with the abscission checkpoint [13] and its transcription is regulated by TP53 to enhance exosome production[14]. A more recently study has shown that *CHMP4C* is frequently overexpressed in ovarian tumor tissues, with the suggestion that it may represent a diagnostic tumor marker and therapeutic -target for patients with the disease[15].

At the 10p12 locus, six known genes (*NEBL, C10Orf113, C10Orf114, C10Orf140, MLLT10* and *DNAJC1*) span the one megabase region around rs1243180, which lies in an intron of *MLLT10* (**Figure 3**). Based on data imputed from the 1000 Genomes Project 57 SNPs are candidates for being functionally significant variants. This includes variants in the 3′UTR of

c10orf114 and the 5′UTR of c10orf140, and a synonymous variant in *MLLT10*. Forty-six SNPs lie in introns of the *MLLT10* gene, and the remaining eight are intergenic. *In silico* analyses found little or no functional evidence that any of these SNPs, including the most risk associated SNP rs1243180. However, after FAIREseq analysis of normal serous ovarian cancer precursor cells, one of these SNPs, rs10828252 (r^2=0.87 with rs1243180), was found to coincide with a region of open chromatin, which probably corresponds to the promotor of *MLLT10* (**Figure 3**). Although rs10828252 is not positioned directly at the apex of the signal, but rather within the upstream portion, it is well established that open chromatin at transcription start sites of genes result from the coordinated influence of numerous transcription factors (TF) binding within the vicinity, and therefore it is highly plausible that rs10828247 is modulating one of these TF binding sites. The resulting shape and position of the FAIRE-seq signal may therefore represent the resulting effects of putative TF binding at rs10828247 working in concert with others in close proximity.

This finding suggests a possible mechanism for susceptibility to EOC at this locus through subtle variations in promotor regulation of *MLLT10*. However, eQTL analysis found no significant associations between genotypes of either rs1243180 or rs10828252 and *MLLT10* expression in normal tissues. We did observe eQTL associations for two other coding genes in the region, *NEBL* and *C10orf114* (P=0.04 and 0.03 respectively). *C10orf114* expression was also associated with rs1243180 genotypes in primary EOCs (P=0.02) as was *C10orf140* (P=0.02). Methylation at both *C10orf140* and *MLLT10* was associated with rs1243180 genotype in primary EOC tissues (P=0.03 and 0.05 respectively), and both genes show significant negative correlation between methylation and expression (P=0.0016 and 0.002 respectively). *C10orf140* also shows a significant difference in methylation in tumors compared to normal tissue (P=1.9×10^{-5}).

Four genes (*NEBL, C10orf114, C10orf140* and *MLLT10*) were significantly overexpressed in EOC cell lines compared to normal tissues (P 0.01) of which two, *C10orf114*, and *MLLT10*, also showed elevated overexpression in primary EOCs (**Figure 3**). Correlations between gene expression and DNA copy number variation at this locus in primary EOCs suggest that overexpression of the *C10orf114* and *MLLT10* genes are driven by copy number variation. *NEBL* is the only gene with reported mutations in ovarian cancer (**Supplementary fig. 10**). Together, these data suggest that *NEBL, C10orf114, C10orf140* and *MLLT10* gene may all play a role in ovarian cancer development, and any of these four genes could be the target susceptibility gene at this locus. However, there is no other evidence to implicate *C10orf114* or *C10orf140* in EOC or to suggest a rationale that may underlie disease susceptibility. More is known about the function of *NEBL* and *MLLT10* although neither gene has previously been implicated in ovarian cancer. *MLLT10* (mixed-lineage leukemia (trithorax homolog, Drosophila) translocated to 10) encodes a transcription factor and has been identified as a partner gene involved in several chromosomal rearrangements resulting in leukaemia[16]. More than 60 MLL fusion partner genes have been described at the molecular level, including a recently reported fusion *NEBL-MLL*[17]. *NEBL* (Nebulette) encodes a nebulin like protein that is abundantly expressed in cardiac muscle, and has been implicated in the genetics of sudden cardiac death syndrome and cardiac remodeling[18]. This evidence does not support directly a role for these genes in ovarian cancer, but the first

common gene-fusion in serous ovarian cancers (ESRRA-C11orf20) has been reported recently and provides an underlying hypothesis for the involvement of genes at this locus in EOC development[19].

At chromosome 17q12, the most significant SNP, rs757210, lies in an intron of *HNF1B* and is associated with serous subtype ovarian cancer. Based on data imputed from the 1000 Genomes Project nine SNPs are candidates for being the causal variant. SNPs in this region have been associated with diabetes [20], endometrial cancer[21] and prostate cancer[22]. There are thirteen genes in the 1MB region on either side of this SNP (ACACA, C17orf78, TADA2L, DUSP14, AP1GBP1, DOX52, HNF1B, TBC1D3F, TBC1D3 MRPL45, GRP179, SOCS7, and ARHGAP23). Seven of these (DUSP14, HNF1B, TBC1D3, TBC1D3F, MRPL45, SOCS7 and ARHGAP23) were overexpressed in EOC cell lines and primary tumors compared to normal tissues (**Figure 4** and **Supplementary fig. 8.iii**) indicating they may play a role in EOC. *HNF1B* is a strong candidate gene target at this locus; it has been extensively studied in EOC and is used as a biomarker for subtype stratification of EOC tumors[23], in particular to distinguish clear cell from other EOC subtypes. Consistent with this, overexpression of *HNF1B* in EOC cell lines was largely driven by higher expression in clear cell EOC cell lines (Figure 4)[24]. However, *HNF1B* shows lower expression in primary serous EOCs compared to normal tissues, which may suggest a different role for this gene in clear cell compared to serous tumours [25]. The phenotypic consequences of *HNF1B* knockdown in clear cell EOC cell lines also suggest it may behave as an oncogene in the development of this subtype.[26]. We found no correlation between *HNF1B* expression and DNA copy number variation at this locus in primary EOCs; but a highly statistically significant inverse correlation between *HNF1B* expression and methylation ($P=2.1\times10^{-6}$) which implies the mechanism for overexpression of this gene is epigenetic.

RNAseq analysis of normal ovarian cancer precursor tissues indicates that *HNF1B* is expressed at extremely low levels (**Figure 4**), which restricts the extent to which the function of this gene in normal ovarian cancer precursor tissues can be studied. We found no evidence for eQTL associations between rs757210 and the expression of any gene in normal tissues throughout the region, but we observed a strong mQTL association between rs757210 and *HNF1B* methylation (P=0.009) (**Figure 4; Supplementary table 4**) in primary serous EOCs. The minor (risk) allele of rs757210 was associated with lower methylation and therefore is predicted to be associated with increased *HNF1B* expression. In the absence of additional functional data it is difficult to interpret these findings; but given the possible role of *HNF1B* as an oncogene in the development of clear cell ovarian cancer, it may be that increased *HNF1B* expression at an early stage in the development of ovarian cancer precursor tissues, driven by the risk variant(s) underlying susceptibility, has increased oncogenic activity in the proportion of serous ovarian that carry this allele.

Overall, the functional data we have generated do not point strongly to any one gene at 17q12 as the functionally relevant gene mediating the genetic association with disease. However, when combined with a large body of previous work data implicating *HNF1B* in ovarian cancer development, it suggests that this is the strongest candidate, and the mQTL and methylation-expression associations suggest a role for genetic variants influencing *HNF1B* expression and disease susceptibility through epigenetic regulation.

Discussion

This study has demonstrated the strength of large-scale collaboration in genetic association studies. We have identified three novel common alleles that confer susceptibility to EOC and confirmed two loci that had been previously reported at near genome-wide significance. Molecular analyses of genes at these loci, combining publicly available datasets and systematic, large-scale experiments point to a small number of candidate gene targets that may play a role in EOC initiation and development. However, the effects of the novel loci were modest, and together they explain less than 1 per cent of the excess familial risk of EOC, with about 4 per cent explained by all known loci with common susceptibility alleles. The lack of heterogeneity between studies of varying designs, carried out in different populations and the high levels of statistical significance indicate that these are robust associations. Fewer common susceptibility loci have now been found for EOC than for several other common cancers including breast, colorectal and prostate cancer[27]. It seems unlikely that the underlying genetic architecture for EOC susceptibility is substantially different from other cancers. This suggests that a key factor limiting our ability to detect susceptibility loci is sample size – the power of this study to detect risk alleles across a range of likely effect sizes is modest (**Supplementary fig. 12**). However, EOC is less common than these cancers and has a higher mortality and recruiting extremely large numbers of cases will be difficult. Disease heterogeneity will also reduce power if a substantial proportion of EOC susceptibility alleles are subtype specific. All EOC susceptibility loci so far identified are strongly associated with serous EOC, which is also the most common subtype. Both the discovery and replication phases of this study were weighted towards identifying serous associated risk alleles. It seems likely that additional common risk loci for clear cell, endometrioid and mucinous EOC subtypes also exist and await identification.

Several EOC susceptibility alleles have now been identified that increased risk of multiple cancers. For example, an increased risk of estrogen-receptor-negative breast cancer is associated with the EOC susceptibility allele at 19p13[5,28] and the EOC susceptibility allele at the 17q12 locus reported in this manuscript is also associated with risk of endometrial cancer[21] and prostate cancer[22]. Several of the loci containing EOC susceptibility alleles have been found to harbour different susceptibility alleles for other cancers. For example, Michailidou and colleagues [29] have found an association between rs7072776 at 10p12 and breast cancer. This SNP is ~120kb centromeric to and partially correlated with rs1243180 (r^2=0.51). Michailidou and colleagues also report an association of rs11780156 at 8q24 with breast cancer. The new locus lies ~300kb telomeric of the known locus for ovarian cancer (rs10088218)[6], but is uncorrelated with it (r^2=0.02). Both loci lie ~400kb 3′ of *MYC*. Previous GWAS have identified multiple loci 5′ of *MYC* and associated with different cancer types, including a locus for breast cancer. These associations may reflect tissue-specific regulation of key genes and understanding the functional mechanisms underlying genetic associations at the same locus for different phenotypes may provide insights into more general mechanisms of disease etiology and cancer development.

Assuming a log-additive model of interaction between loci, the currently know loci (**Table 2 and Supplementary Table 3**) define a genetic risk profile with a combined variance for the log relative risk distribution of 0.057. Such a distribution has limited discriminatory ability;

the estimated relative risks at the 5th and 95th centiles are 0.63 and 1.48 respectively. However, based on what is known about the architecture of genetic susceptibility for other cancers it is likely that many more common susceptibility alleles exist. The discovery of genetic association with ovarian cancer may be enhanced by taking advantage functional annotation data and the analysis of gene-gene and gene-environment interactions using a genome-wide approach. Continued international efforts are needed to establish new case-control studies, to expand existing case-control studies and to improve the pathological characterisation of the cases in these studies in order to unravel the inherited genetic basis of the disease. In combination with rarer risk alleles and other risk factors, genetic profiling may provide sufficient discrimination to justify targeted ovarian cancer prevention.

URLs

1000 Genomes Project: http://www.1000genomes.org/page.php

COSMIC: http://www.sanger.ac.uk/genetics/CGP/cosmic/

GeneMANIA: http://genemania.org/

MACH: http://www.sph.umich.edu/csg/abecasis/MACH/

NCBI Unigene: http://www.ncbi.nlm.nih.gov/unigene

The Cancer Genome Atlas Project: http://cancergenome.nih.gov/

The Cancer Genome Atlas Project data bioprtal: http://www.cbioportal.org/

The Wellcome Trust Case Control Consortium: http://www.wtccc.org.uk/

Online Methods

SNP selection

We combined the results from two ovarian cancer GWAS from North America and the UK. Details of these studies have been published previously [4,31] and are described in the **Supplementary note**. In order to account for different marker sets and to improve genome coverage, imputation to HapMap2 was performed using 60 CEU founders as reference. Data on 2,508,744 genotyped SNPs or SNPs imputed with $r^2 > 0.3$ were available for analysis.

The North American and UK studies were analysed separately and the results combined using fixed effects meta-analysis. The 2.5 million SNPs were ranked according to the *P*-values for each of four analyses performed: North America study only (all invasive and serous histology) and combined GWAS meta-analysis (all invasive and serous histology). The minimal ranking for each SNP was obtained across the four sets of results. SNPs with minor allele frequency less than 3 percent or SNPs that were already genotyped or in perfect LD with UK GWAS phase 2 SNPs were excluded. We acquired the design score for each SNP using the Illumina Assay Design Tool and removed SNPs that were redundant or predicted to perform poorly. In total, 24,552 SNPs were included on the iCOGS custom genotyping array (see Supplementary materials).

Study populations

A total of 47,630 samples from 43 studies in OCAC were genotyped of which 44,308 passed QC, including 18,174 (10,315 serous) cases and 26,134 controls (**Supplementary table 1**). The HapMap samples for European (CEU, n=60), African (YRI, n=53) and Asian (JPT +CHB, n=88) populations were also genotyped using the iCOGS array.

SNP genotyping

Genotyping was conducted using an Illumina Infinium iSelect BeadChip in six centres, of which two were used for OCAC - McGill University and Génome Québec Innovation Centre (n=19,806) and the Mayo Clinic Medical Genome Facility (n=27,824). Each 96-well plate contained 250 ng genomic DNA (or 500 ng whole-genome amplified DNA). Raw intensity data files for all consortia were sent to the COGS data co-ordination centre at the University of Cambridge for centralized genotype calling and QC. Genotypes were called using GenCall [32]. Initial calling used a cluster file generated using 270 samples from Hapmap2. These calls were used for ongoing QC checks during the genotyping. To generate the final calls used for the data analysis, we first selected a subset of 3,018 individuals, including samples from each of the genotyping centres, each of the participating consortia, and each major ethnicity. Only plates with a consistently high call rate in the initial calling were used. The HapMap samples and ~160 samples that were known positive controls for rare variants on the array were used to generate a cluster file that was then applied to call the genotypes for the remaining samples. We also investigated two other calling algorithms: Illumnus [33] and GenoSNP [34], but manual inspection of a sample of SNPs with discrepant calls indicated that GenCall was invariably superior.

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 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. Thus, a total of 44,308 subjects including 18,174 cases and 26,134 controls were available for analysis. Of these, 2,482 had been in the North American GWAS, 1,641 were in the phase 1 of the UK GWAS and 8,463 in Phase 2 of the UK GWAS.

SNP QC

Of 211,155 SNP assays successfully designed and included on the array, SNPs were excluded according to the following criteria: (1) 1,311 SNPs without a genotype call; (2)

2,857 monomorphic SNPS; (3) 5,201 SNPs with a call rate less than 95 percent and MAF > 0.05 or call rate less than 99 percent with MAF < 0.05; (4) 2,194 SNPs showing evidence of deviation of genotype frequencies from Hardy-Weinberg equilibrium (p < 10^{-7}); (5) 22 SNPS with greater than 2 percent discordance in duplicate pairs. Overall, 94.5 percent passed QC. Genotype intensity cluster plots were visually inspected for the most strongly associated SNPs at each newly identified locus.

Statistical methods

We used the program LAMP[35] to assign intercontinental ancestry based on the genotype frequencies for the European, Asian and African populations. Subjects with greater than 90 percent European ancestry were defined as European (n=39,944) and those with greater than 80 percent Asian or African ancestry were defined as being Asian (n=2,388) and African respectively (n=387). All other subjects were defined as being of mixed ancestry (n=1,770). We then used a set of 37,000 unlinked markers to perform principal components analysis within each major population subgroup [36]. To enable this analysis on very large samples we used an in-house programme written in C++ using the Intel MKL libraries for eigenvectors (available at http://ccge.medschl.cam.ac.uk/software/).

Unconditional logistic regression treating the number of alternate alleles carried as an ordinal variable (log-additive, co-dominant model) was used to evaluate the association between each SNP and ovarian cancer risk. A likelihood ratio test was used to test for association, and per-allele log odds ratios and 95 percent confidence limits were estimated. The likelihood ratio test has been shown to have greater power than alternatives such as the Wald test and score test for rare variants[37]. Separate analyses were carried out for each ancestry group. The model for European subjects was adjusted for study and population substructure by including study-specific indicators and the first five eigenvalues from the principal components analysis in the model. For analysis of the Asian and other ancestry groups, the first five ancestry-specific principal components were included in the model, and one principal component was included in the model for analysis of subjects of African ancestry. The number of principal components was chosen based on the position of the inflexion of the principal components scree plot (**Supplementary fig. 13**). We tested for sub-type specific heterogeneity by comparing genotype frequencies in the four case subtypes using the Kruskal-Wallis test. We tested for heterogeneity of ORs by study and ancestry using the method of Breslow and Day [38].

In order to assess the magnitude of confounding due to cryptic population substructure, we tested the 147,722 SNPs that not been selected as candidates for ovarian cancer susceptibility. Inflation in the test statistics (λ) was estimated by dividing the median of the test statistic by 0.455 (the median for the χ^2 distribution on 1df). The inflation was converted to an equivalent inflation for a study with 1000 cases and 1000 controls (λ_{1000}) by adjusting by effective study size, namely

$$\lambda_{1000} = 1 + 500\left(\lambda - 1\right)/\sum_k \left(\frac{1}{n_k} + \frac{1}{m_k}\right)^{-1}$$

where n is the number of cases and m the number of controls in each study stratum, k. In analyses restricted to European subjects and adjusted only for study there was a small inflation of the test statistics ($\lambda = 1.13$, $\lambda_{1000} = 1.007$). This was reduced to 1.078 ($\lambda_{1000} = 1.004$) after adjusting for five principal components. Heterogeneity of odds ratios between studies was tested with Cochran's Q statistic.

Functional studies

We performed the following assays for each gene in the one megabase region centred on the most significant SNP at each locus (see **Supplementary note** for details): (1) Gene expression in EOC cell lines (n=50) and normal precursor cells and tissues for ovarian cancers - ovarian surface epithelial cells (OSE) and fallopian tube secretory epithelial (FTSE) cells (n=73); (2) CpG island methylation analysis in high grade serous EOC (n=106) and normal tissues (n=7).

We also evaluated these genes *in silico*, using bioinformatics tools to mine publicly available somatic genetic data generated for primary EOCs and other cancer types. These were The Cancer Genome Atlas (TCGA) data for ~500 high grade serous EOCs (gene expression, somatic mutation, DNA copy number variation, eQTL, and methylation data) [25] and the Catalogue Of Somatic Mutations In Cancer (COSMIC) [39] analysis of mutations in genes curated from the published literature and data from whole genome resequencing of cancer samples, undertaken by the Cancer Genome Project. We generated co-expression networks for genes in each locus using GeneMANIA, a large dataset of gene expression studies (n=154) [40].

All these data enabled us to: (1) compare gene expression in tumor and normal epithelium (EOC cell lines/normal cell lines and TCGA tumors/normal tissue), (2) test for association between copy number alteration and gene expression at each locus (3) compare gene methylation status in tumor and normal tissue (4) carry out a gene expression quantitative trait locus (eQTL) analysis to evaluate associations between germline genotype and gene expression in lymphoblastoid cell lines, normal serous EOC precursor tissues and tumors, and (5) carry out a methylation quantitative trait locus (mQTL) analysis to evaluate associations between germline genotype and gene methylation in tumors (**Supplementary figs. 8i-iii**).

We used data from the ENCyclopedia Of DNA Elements (ENCODE)[7] to evaluate the overlap between regulatory elements in non-coding regions and risk associated SNPs at the three loci. ENCODE describes regulatory DNA elements (e.g. enhancers, insulators and promotors) and non-coding RNAs (e.g. micro-RNAs, long non-coding and piwi-interacting RNAs) that may be targets for susceptibility alleles (**Supplementary figs. 9i-iii**). However, ENCODE does not include data for EOC associated tissues, and activity of such regulatory elements often varies in a tissue specific manner [7,41]. Therefore, we profiled the spectrum of non-coding regulatory elements in OSE and FTSE cells using a combination of formaldehyde assisted isolation of regulatory elements sequencing (FAIRE-seq) and RNA sequencing (RNA-seq) (**Supplementary figs. 9i-iii**). We also analysed regulatory regions in early-stage transformed OSE cells. For all regulatory biofeatures spanning the one megabase

region at each locus, we evaluated their overlap with the most strongly associated SNP and all SNPs correlated with an r^2 0.8.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Authors

Paul D. P. Pharoah[1,2,*], Ya-Yu Tsai[3,*], Susan J. Ramus[4,*], Catherine M. Phelan[3,*], Ellen L. Goode[5], Kate Lawrenson[4], Melissa Price[3], Brooke L. Fridley[5], Jonathan P. Tyrer[1], Howard Shen[4], Rachel Weber[6], Rod Karevan[4], Melissa C. Larson[7], Honglin Song[1], Daniel C. Tessier[8,9], François Bacot[8,9], Daniel Vincent[8,9], Julie M. Cunningham[10], Joe Dennis[2], Ed Dicks[1], Australian Cancer Study[11], Australian Ovarian Cancer Study Group[11], Katja K. Aben[12,13], Hoda Anton-Culver[14], Natalia Antonenkova[15], Sebastian M. Armasu[7], Laura Baglietto[16,17], Elisa V. Bandera[18], Matthias W. Beckmann[19], Michael J. Birrer[20,21], Greg Bloom[3], Natalia Bogdanova[22], James D. Brenton[23], Louise A. Brinton[24], Angela Brooks-Wilson[25], Robert Brown[26], Ralf Butzow[27,28], Ian Campbell[29,30], Michael E Carney[31], Renato S. Carvalho[3], Jenny Chang-Claude[32], Y. Anne Chen[3], Zhihua Chen[3], Wong-Ho Chow[33], Mine S. Cicek[5], Gerhard Coetzee[34], Linda S. Cook[35], Daniel W. Cramer[36,37], Cezary Cybulski[38], Agnieszka Dansonka-Mieszkowska[39], Evelyn Despierre[40], Jennifer A Doherty[41], Thilo Dörk[22], Andreas du Bois[42,43], Matthias Dürst[44], Diana Eccles[45], Robert Edwards[46,47], Arif B. Ekici[48], Peter A. Fasching[19,49], David Fenstermacher[3], James Flanagan[26], Yu-Tang Gao[50], Montserrat Garcia-Closas[51,52], Aleksandra Gentry-Maharaj[53], Graham Giles[16,17,54], Anxhela Gjyshi[3], Martin Gore[55], Jacek Gronwald[38], Qi Guo[1], Mari K Halle[56,57], Philipp Harter[42,43], Alexander Hein[19], Florian Heitz[42,43], Peter Hillemanns[58], Maureen Hoatlin[59], Estrid Høgdall[60,61], Claus K. Høgdall[62], Satoyo Hosono[63], Anna Jakubowska[38], Allan Jensen[60], Kimberly R. Kalli[64], Beth Y. Karlan[65], Linda E. Kelemen[66,67], Lambertus A. Kiemeney[12,13,68], Susanne Krüger Kjaer[60,62], Gottfried E. Konecny[49], Camilla Krakstad[56,57], Jolanta Kupryjanczyk[39], Diether Lambrechts[69], Sandrina Lambrechts[40], Nhu D. Le[70], Nathan Lee[4], Janet Lee[4], Arto Leminen[28], Boon Kiong Lim[71], Jolanta Lissowska[72], Jan Lubiński[38], Lene Lundvall[62], Galina Lurie[31], Leon F.A.G. Massuger[73], Keitaro Matsuo[63], Valerie McGuire[74], John R McLaughlin[75,76], Usha Menon[53], Francesmary Modugno[46,77], Kirsten B. Moysich[78], Toru Nakanishi[79], Steven A. Narod[80], Roberta B. Ness[81], Heli Nevanlinna[28], Stefan Nickels[32], Houtan Noushmehr[34,82], Kunle Odunsi[78], Sara Olson[83], Irene Orlow[83], James Paul[84], Tanja Pejovic[85,86], Liisa M Pelttari[28], Jenny Permuth-Wey[3], Malcolm C Pike[4,83], Elizabeth M Poole[37,87], Xiaotao Qu[3], Harvey A. Risch[88], Lorna Rodriguez-Rodriguez[18], Mary Anne Rossing[89,90], Anja Rudolph[32], Ingo Runnebaum[44], Iwona K Rzepecka[39], Helga B. Salvesen[56,57], Ira Schwaab[91], Gianluca Severi[16,17], Hui Shen[82], Vijayalakshmi Shridhar[10], Xiao-Ou Shu[92,93], Weiva Sieh[74], Melissa C. Southey[94], Paul Spellman[95], Kazuo Tajima[63], Soo-Hwang Teo[96,97], Kathryn L. Terry[36,37], Pamela J Thompson[31], Agnieszka Timorek[98,99], Shelley S. Tworoger[37,87], Anne M. van Altena[73], David Van Den

Berg[4], Ignace Vergote[40], Robert A. Vierkant[5], Allison F. Vitonis[36], Shan Wang-Gohrke[100], Nicolas Wentzensen[24], Alice S. Whittemore[74], Elisabeth Wik[56,57], Boris Winterhoff[101], Yin Ling Woo[71], Anna H Wu[4], Hannah P. Yang[24], Wei Zheng[92,93], Argyrios Ziogas[14], Famida Zulkifli[96,97], Marc T. Goodman[31], Per Hall[102], Douglas F Easton[1,2], Celeste L Pearce[4], Andrew Berchuck[103], Georgia Chenevix-Trench[104], Edwin Iversen[105], Alvaro N.A. Monteiro[3], Simon A. Gayther[4], Joellen M. Schildkraut[6,106,§], and Thomas A. Sellers[3,§]

Affiliations

[1] The Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK. [2] The Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. [3] Department of Cancer Epidemiology, Division of Population Sciences, Moffitt Cancer Center, Tampa, FL, USA. [4] Department of Preventive Medicine, Keck School of Medicine, University of Southern California Norris Comprehensive Cancer Center, Los Angeles, California, USA. [5] Department of Health Science Research, Division of Epidemiology, Mayo Clinic, Rochester, Minnesota, USA. [6] Department of Community and Family Medicine, Duke University Medical Center, Durham, NC, USA. [7] Department of Health Science Research, Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, Minnesota, USA. [8] McGill University, Montreal, Canada. [9] Génome Québec Innovation Centre, Montreal, Canada. [10] Department of Laboratory Medicine and Pathology, Division of Anatomic Pathology, Mayo Clinic, Rochester, Minnesota, USA. [12] Department of Epidemiology, Biostatistics and HTA, Radboud University Medical Centre, Nijmegen, Netherlands. [13] Comprehensive Cancer Center The Netherlands, Utrecht, The Netherlands. [14] Department of Epidemiology, Center for Cancer Genetics Research and Prevention, School of Medicine, University of California Irvine, Irvine, California, USA. [15] Byelorussian Institute for Oncology and Medical Radiology Aleksandrov N.N., Minsk, Belarus. [16] Cancer Epidemiology Centre, The Cancer Council Victoria, Melbourne, Australia. [17] Centre for Molecular, Environmental, Genetic and Analytical Epidemiology, University of Melbourne, Australia. [18] The Cancer Institute of New Jersey, Robert Wood Johnson Medical School, New Brunswick, NJ, USA. [19] University Hospital Erlangen, Department of Gynecology and Obstetrics, Friedrich-Alexander-University Erlangen-Nuremberg, Comprehensive Cancer Center, Erlangen, Germany. [20] Harvard Medical School, Boston, Massachusetts, United States of America [21] Massachusetts General Hospital Cancer Center, Massachusetts General Hospital, Boston, MA, United States of America. [22] Gynaecology Research Unit, Hannover Medical School, Hannover, Germany. [23] Cambridge Research Institute Li Ka Shing Centre, Cambridge, UK. [24] Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda MD, USA. [25] Genome Sciences Centre, BC Cancer Agency, Vancouver, BC, Canada. [26] Department of Surgery and Cancer, Imperial College London, London, UK. [27] Department of Pathology, Helsinki University Central Hospital, Helsinki, Finland. [28] Department of Obstetrics and Gynecology, Helsinki University Central Hospital, Helsinki, Finland. [29] Cancer Genetics Laboratory,

Research Division, Peter MacCallum Cancer Centre, Melbourne, Australia. [30] Department of Pathology, University of Melbourne, Parkville, Victoria, Australia. [31] Cancer Epidemiology Program, University of Hawaii Cancer Center, Hawaii, USA. [32] German Cancer Research Center, Division of Cancer Epidemiology, Heidelberg, Germany. [33] Division of Cancer Etiology and Genetics, National Cancer Institute, Bethesda, MD, USA. [34] Department of Microbiology and Preventive Medicine, University of Southern California, Norris Comprehensive Cancer Center, Los Angeles, CA. [35] Department of Internal Medicine, University of New Mexico, Albuquerque, NM, USA. [36] Obstetrics and Gynecology Epidemiology Center, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA. [37] Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, USA. [38] International Hereditary Cancer Center, Department of Genetics and Pathology. Pomeranian Medical University, Szczecin, Poland. [39] Department of Molecular Pathology, The Maria Sklodowska-Curie Memorial Cancer Center and Institute of Oncology, Warsaw, Poland. [40] Division of Gynecologic Oncology, Department of Obstetrics and Gynaecology, Leuven Cancer Institute, Leuven, Belgium. [41] Section of Biostatistics and Epidemiology, The Geisel School of Medicine at Dartmouth, Lebanon, NH, USA. [42] Department of Gynecology and Gynecologic Oncology, Dr. Horst Schmidt Kliniken Wiesbaden, Wiesbaden, Germany. [43] Department of Gynecology and Gynecologic Oncology, Kliniken Essen-Mitte/ Evang. Huyssens-Stiftung/ Knappschaft GmbH, Essen, Germany. [44] Department of Gynecology, Jena University Hospital, Jena, Germany. [45] Faculty of Medicine, University of Southampton, University Hospital Southampton, UK. [46] Department of Obstetrics, Gynecology and Reproductive Sciences, University of Pittsburgh, Pittsburgh, PA, USA. [47] Womens Cancer Program, Magee-Womens Research Institute, Pittsburg, PA, USA. [48] Institute of Human Genetics, Friedrich-Alexander-University Erlangen-Nuremberg, Erlangen, Germany. [49] Department of Medicine, Division of Hematology and Oncology, David Geffen School of Medicine, University of California at Los Angeles, Los Angeles, USA. [50] Department of Epidemiology, Shanghai Cancer Institute, Shanghai, China. [51] Division of Genetics and Epidemiology, Institute of Cancer Research, Sutton, UK. [52] Breakthrough Breast Cancer Research Centre, London, UK. [53] Gynaecological Cancer Research Centre, UCL EGA Institute for Women's Health, London, United Kingdom. [54] Department of Epidemiology and Preventive Medicine, Monash University, Melbourne, Australia. [55] Gynecological Oncology Unit, The Royal Marsden Hospital, United Kingdom. [56] Department of Gynecology and Obstetrics, Haukeland University Hospital, Bergen, Norway. [57] Department of Clinical Medicine, University of Bergen, Bergen, Norway. [58] Clinics of Obstetrics and Gynaecology, Hannover Medical School, Hannover, Germany. [59] Department of Biochemistry and Molecular Biology, Oregon Health and Science University, Portland, USA. [60] Virus, Lifestyle and Genes, Danish Cancer Society Research Center, Copenhagen, Denmark. [61] Molecular Unit, Department of Pathology, Herlev Hospital, University of Copenhagen, Copenhagen, Denmark. [62] The Juliane Marie Centre, Department of Obstetrics and Gynecology, Rigshospitalet, Copenhagen, Denmark. [63] Division of

Epidemiology and Prevention, Aichi Cancer Center Research Institute, Nagoya, Aichi, Japan. [64] Department of Medical Oncology, Mayo Clinic, Rochester, Minnesota, USA. [65] Women's Cancer Program at the Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, California. [66] Department of Population Health Research, Alberta Health Services-Cancer Care, Calgary, Alberta, Canada. [67] Department of Medical Genetics, University of Calgary, Calgary, AB, Canada. [68] Department of Urology, Radboud University Medical Centre, Nijmegen, Netherlands. [69] Vesalius Research Center, University of Leuven, Belgium. [70] Cancer Control Research, BC Cancer Agency, Vancouver, BC, Canada. [71] Department of Obstetrics and Gynaecology, University Malaya Medical Centre, University Malaya, Kuala Lumpur, Malaysia. [72] The Maria Sklodowska-Curie Memorial Cancer Center, Warsaw, Poland. [73] Department of Gynaecology, Radboud University Medical Centre, Nijmegen, Netherlands. [74] Department of Health Research and Policy - Epidemiology, Stanford University School of Medicine, Stanford CA, USA. [75] Dalla Lana School of Public Health, Faculty of Medicine, University of Toronto, Canada. [76] Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Ontario, Canada. [77] Department of Epidemiology, University of Pittsburgh, Pittsburgh, PA, USA. [78] Department of Cancer Prevention and Control, Roswell Park Cancer Institute, Buffalo, NY. [79] Department of Gynecologic Oncology, Aichi Cancer Center Central Hospital, Nagoya, Aichi, Japan. [80] Women's College Research Institute, University of Toronto, Toronto, Ontario, Canada. [81] The University of Texas School of Public Health, Houston, TX, USA. [82] USC Epigenome Center, Keck School of Medicine, University of Southern California, Norris Comprehensive Cancer Center, Los Angeles, CA. [83] Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, New York, New York, USA. [84] The Beatson West of Scotland Cancer Centre, Glasgow, UK. [85] Department of Obstetrics and Gynecology, Oregon Health and Science University, Portland, USA. [86] Knight Cancer Institute, Oregon Health and Science University, Portland, Oregon, USA. [87] Brigham and Women's Hospital, Boston, MA, USA. [88] Department of Epidemiology and Public Health, Yale University School of Public Health and School of Medicine. [89] Department of Epidemiology, University of Washington, Seattle, WA, USA. [90] Program in Epidemiology, Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA. [91] Institut für Humangenetik Wiesbaden, Wiesbaden, Germany. [92] Vanderbilt Epidemiology Center, Vanderbilt University School of Medicine, Nashville, TN, USA. [93] Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, TN, USA. [94] Genetic Epidemiology Laboratory, Department of Pathology, The University of Melbourne, Melbourne, Victoria, Australia. [95] Molecular and Medical Genetics, Oregon Health and Science University, Portland, Oregon, USA. [96] Cancer Research Initiatives Foundation, Sime Darby Medical Centre, Subang Jaya, Malaysia. [97] University Malaya Medical Centre, University Malaya, Kuala Lumpur, Malaysia. [98] Department of Obstetrics, Gynecology and Oncology, IInd Faculty of Medicine, Warsaw Medical University, Warsaw, Poland. [99] Brodnowski Hospital, Warsaw, Poland. [100]

Department of Obstetrics and Gynecology, University of Ulm, Ulm, Germany. [101] Department of Obstetrics and Gynecology, Mayo Clinic, Rochester, Minnesota, USA. [102] Department of Epidemiology and Biostatistics, Karolinska Istitutet, Stockholm, Sweden. [103] Department of Obstetrics and Gynecology, Duke University Medical Center, Durham, North Carolina, USA. [104] Cancer Division, Queensland Institute of Medical Research, Herston, QLD, Australia. [105] Department of Statistics, Duke University, Durham, NC, USA. [106] Cancer Prevention, Detection and Control Research Program, Duke Cancer Institute, Durham, North Carolina, USA.

Acknowledgements

References

1. Lichtenstein P, et al. Environmental and heritable factors in the causation of cancer-- analyses of cohorts of twins from Sweden, Denmark and Finland. N.Engl.J.Med. 2000; 343:78–85. [PubMed: 10891514]

2. Stratton JF, Pharoah PDP, Smith SK, Easton DF, Ponder BAJ. A systematic review and meta-analysis of family history and risk of ovarian cancer. British Journal of Obstetrics and Gynaecology. 1998; 105:493–499. [PubMed: 9637117]

3. Antoniou AC, Easton DF. Risk prediction models for familial breast cancer. Future Oncol. 2006; 2:257–74. [PubMed: 16563094]

4. 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]

5. Bolton KL, et al. Common variants at 19p13 are associated with susceptibility to ovarian cancer. Nat.Genet. 2010; 42:880–4. [PubMed: 20852633]

6. Goode EL, et al. A genome-wide association study identifies susceptibility loci for ovarian cancer at 2q31 and 8q24. Nat.Genet. 2010; 42:874–879. [PubMed: 20852632]

7. Ernst J, et al. Mapping and analysis of chromatin state dynamics in nine human cell types. Nature. 2011; 473:43–9. [PubMed: 21441907]

8. Jia L, et al. Functional enhancers at the gene-poor 8q24 cancer-linked locus. PLoS Genet. 2009; 5:e1000597. [PubMed: 19680443]

9. Kim MJ, et al. Functional characterization of liver enhancers that regulate drug-associated transporters. Clin Pharmacol Ther. 2011; 89:571–8. [PubMed: 21368754]

10. Wasserman NF, Aneas I, Nobrega MA. An 8q24 gene desert variant associated with prostate cancer risk confers differential in vivo activity to a MYC enhancer. Genome Res. 2010; 20:1191–7. [PubMed: 20627891]

11. Wright JB, Brown SJ, Cole MD. Upregulation of c-MYC in cis through a large chromatin loop linked to a cancer risk-associated single-nucleotide polymorphism in colorectal cancer cells. Mol Cell Biol. 2010; 30:1411–20. [PubMed: 20065031]

12. McCullough J, Fisher RD, Whitby FG, Sundquist WI, Hill CP. ALIX-CHMP4 interactions in the human ESCRT pathway. Proc.Natl.Acad.Sci.U.S.A. 2008; 105:7687–91. [PubMed: 18511562]

13. Carlton JG, Caballe A, Agromayor M, Kloc M, Martin-Serrano J. ESCRT-III governs the Aurora B-mediated abscission checkpoint through CHMP4C. Science. 2012; 336:220–5. [PubMed: 22422861]

14. Yu X, Riley T, Levine AJ. The regulation of the endosomal compartment by p53 the tumor suppressor gene. The FEBS journal. 2009; 276:2201–12. [PubMed: 19302216]

15. Nikolova DN, et al. Genome-wide gene expression profiles of ovarian carcinoma: Identification of molecular targets for the treatment of ovarian carcinoma. Molecular medicine reports. 2009; 2:365–84. [PubMed: 21475838]

16. Caudell D, Aplan PD. The role of CALM-AF10 gene fusion in acute leukemia. Leukemia. 2008; 22:678–85. [PubMed: 18094714]

17. Coser VM, et al. Nebulette is the second member of the nebulin family fused to the MLL gene in infant leukemia. Cancer Genet.Cytogenet. 2010; 198:151–4. [PubMed: 20362230]

18. Ram R, Blaxall BC. Nebulette mutations in cardiac remodeling: big effects from a small mechanosensor. J Am Coll Cardiol. 2010; 56:1503–5. [PubMed: 20951327]

19. Salzman J, et al. ESRRA-C11orf20 is a recurrent gene fusion in serous ovarian carcinoma. PLoS Biol. 2011; 9:e1001156. [PubMed: 21949640]

20. Voight BF, et al. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. Nat.Genet. 2010; 42:579–89. [PubMed: 20581827]

21. Spurdle AB, et al. Genome-wide association study identifies a common variant associated with risk of endometrial cancer. Nat.Genet. 2011; 43:451–454. [PubMed: 21499250]

22. Elliott KS, et al. Evaluation of association of HNF1B variants with diverse cancers: collaborative analysis of data from 19 genome-wide association studies. PLoS One. 2010; 5:e10858. [PubMed: 20526366]

23. Kato N, Sasou S, Motoyama T. Expression of hepatocyte nuclear factor-1beta (HNF-1beta) in clear cell tumors and endometriosis of the ovary. Mod Pathol. 2006; 19:83–9. [PubMed: 16258507]

24. Tsuchiya A, et al. Expression profiling in ovarian clear cell carcinoma: identification of hepatocyte nuclear factor-1 beta as a molecular marker and a possible molecular target for therapy of ovarian clear cell carcinoma. Am.J.Pathol. 2003; 163:2503–12. [PubMed: 14633622]

25. Integrated genomic analyses of ovarian carcinoma. Nature. 2011; 474:609–15. [PubMed: 21720365]

26. Kato N, Motoyama T. Hepatocyte nuclear factor-1beta(HNF-1beta) in human urogenital organs: its expression and role in embryogenesis and tumorigenesis. Histol Histopathol. 2009; 24:1479–86. [PubMed: 19760597]

27. Hindorff, LA.; Junkins, HA.; Hall, PA.; Mehta, JP.; Manolio, TA. A catalogue of published genome-wide association studies. http://www.genome.gov/gwastudies (accessed

28. Antoniou AC, et al. A locus on 19p13 modifies risk of breast cancer in BRCA1 mutation carriers and is associated with hormone receptor-negative breast cancer in the general population. Nat.Genet. 2010; 42:885–92. [PubMed: 20852631]

29. Michailidou K. Large-scale genotyping identifies 41 new breast cancer susceptibility loci. Nat.Genet. ((submitted)).

30. 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; 12:317–25. [PubMed: 20360942]

31. Permuth-Wey J, et al. LIN28B Polymorphisms Influence Susceptibility to Epithelial Ovarian Cancer. Cancer Res. 2011; 71:3896–3903. [PubMed: 21482675]

32. Kermani BG. Artificial intelligence and global normalization methods for genotyping. 2008

33. Teo YY, et al. A genotype calling algorithm for the Illumina BeadArray platform. Bioinformatics. 2007; 23:2741–6. [PubMed: 17846035]

34. Giannoulatou E, Yau C, Colella S, Ragoussis J, Holmes CC. GenoSNP: a variational Bayes within-sample SNP genotyping algorithm that does not require a reference population. Bioinformatics. 2008; 24:2209–14. [PubMed: 18653518]

35. Sankararaman S, Sridhar S, Kimmel G, Halperin E. Estimating local ancestry in admixed populations. Am.J.Hum.Genet. 2008; 82:290–303. [PubMed: 18252211]

36. Price AL, et al. Principal components analysis corrects for stratification in genome-wide association studies. Nat.Genet. 2006; 38:904–9. [PubMed: 16862161]

37. Xing G, Lin CY, Wooding SP, Xing C. Blindly Using Wald's Test Can Miss Rare Disease-Causal Variants in Case-Control Association Studies. Ann.Hum.Genet. 2012

38. Breslow, NE.; Day, NE. Statistical methods in cancer research Volume 1 - the analysis of case-control studies. International Agency for Research on Cancer; Lyon: 1980.

39. Forbes SA, et al. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. Nucleic Acids Res. 2011; 39:D945–50. [PubMed: 20952405]

40. Mostafavi S, Ray D, Warde-Farley D, Grouios C, Morris Q. GeneMANIA: a real-time multiple association network integration algorithm for predicting gene function. Genome Biol. 2008; 9(Suppl 1):S4. [PubMed: 18613948]

41. Heintzman ND, et al. Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature. 2009; 459:108–12. [PubMed: 19295514]
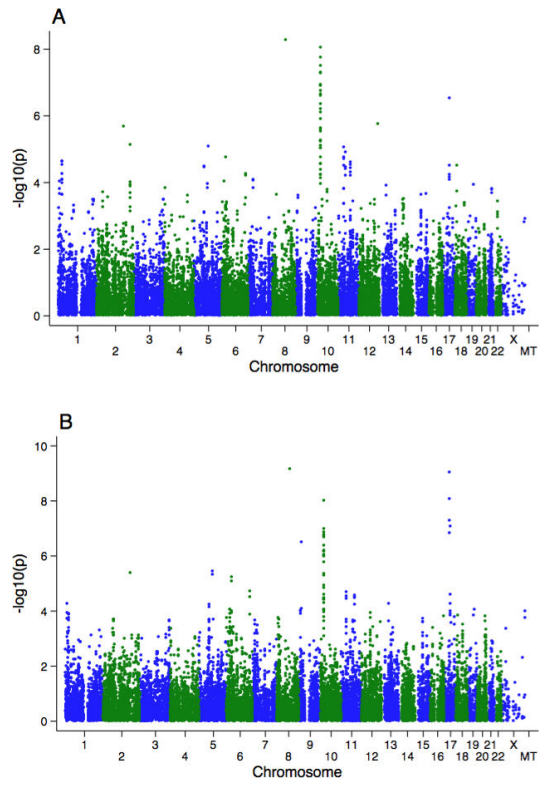
Figure 1.
Manhattan plot showing association between genotype of 22,076 SNPs and risk of A) all invasive epithelial ovarian cancer and B) serous invasive epithelial ovarian cancer
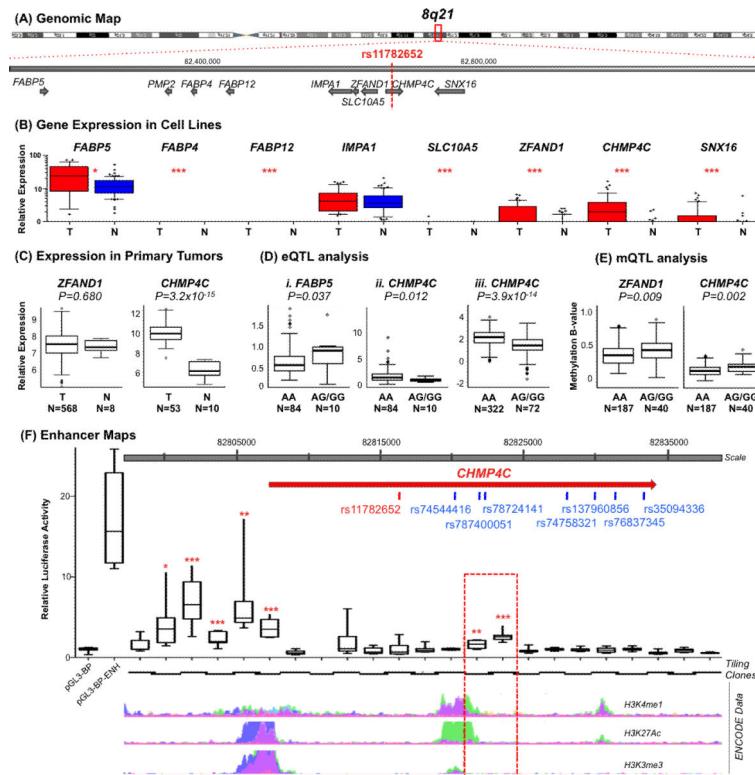
Figure 2. Summary of the functional analyses of the 8q21 locus
(A) Genomic map of a one-megabase region at 8q21 centered on the most statistically significant SNP, rs11782652. The location and size of all nine known protein-coding genes (grey) in the region are shown relative to the location of rs11782652 (red dashed line). **(B)** Expression analysis for all genes at the 8q21 locus performed in epithelial ovarian cancer (EOC) cell lines (n=50) and normal ovarian surface epithelial cell (OSEC) plus fallopian tube secretory epithelial cell (FTSEC) lines (n=73) illustrating the relative levels expression for each gene in tumor (T) compared to normal (N) cell lines (*p<0.05, ***p<0.001). **(C)** The *ZFAND1* result from cell line studies does not replicate in the TCGA nor the MD Anderson primary tumor expression dataset. However, increased expression of *CHMP4C* gene in primary, high-grade serous ovarian tumors (T) compared to normal (N) tissues was confirmed in expression data for primary tissues (MD Anderson data). **(D)** Expression quantitative trait locus (eQTL) analysis: Gene expression is shown relative to the germline genotypes for individuals carrying minor/heterozygous allele (AG/GG) and common alleles (AA) for rs11782652. (i) *FABP5* and (ii) *CHMP4C* show positive eQTL associations in lymphoblastoid cell lines. (iii) A highly significant eQTL association with rs11782652 genotype and *CHMP4C* expression is also seen in primary tumors (TCGA data). **(E)** Methylation quantitative trait locus (mQTL) analysis showing methylation status in 277 high grade serous ovarian cancers relative to genotypes for rs11782652. **(F)** Functional enhancer mapping: A 40kb region was tested for the presence of enhancer regions by transfection of 2kb tiling clones into immortalized OSECs[30]. Activity of the luciferase reporter is shown as fold change in luciferase activity relative to pGL3-BP control. A novel enhancer region in OSECs is indicated by the red box. See Supplementary figure 8i and 9i for additional molecular analysis of all genes at this locus.
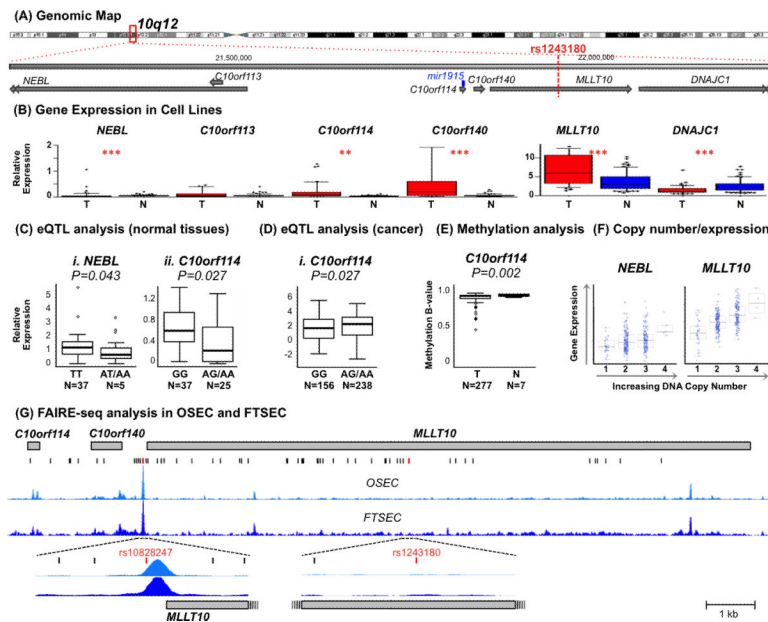
Figure 3. Summary of the functional analysis of the 10p12 locus
(**A**) Genomic map of a one-megabase region at 10p12 centred on the most statistically significant SNP, rs1243180. The location and approximate size of all six known protein-coding genes (grey) and one microRNA (blue) in the region are shown relative to the location of rs1243180 (red dashed line). (**B**) Expression analysis for all genes at this locus performed in EOC cell lines (T) and normal (N) OSEC and FTSEC lines illustrating the relative expression levels for each gene (** p<0.01, ***p<0.001). (**C**) eQTL analysis demonstrates significant associations between genotype at rs1243180 and expression of *(i) NEBL* and *(ii) C10orf114* in early passage primary OSEC/FTSEC cultures. (**D**) *(i)* A positive eQTL association between *C10orf114* genes expression and genotype at rs7098100 (r^2=0.86with rs1243180) was also observed in primary high-grade serous ovarian tumors (TCGA). (**E**) Methylation analysis of 277 high-grade serous ovarian tumors (T) compared to normal ovarian tissues (N) (n=7); methylation at CpG sites near to *C10orf114* show significantly less methylation than in normal samples, which may suggest that the increased expression of this gene in tumors is partly due to loss of methylation. (**F**) Expression versus copy number in TCGA tumors. *MLLT10* and *NEBL* genes show a trend for higher levels of gene expression found in tumors with increased DNA copy number at 10p12; 1 = heterozygous loss; 2 = Diploid; 3 = Copy number gain; 4 = Amplification (**G**) Formaldehyde assisted isolation of regulatory elements sequencing (FAIRE-seq) performed in normal OSECs and FTSECs, relative to all SNPs correlated with r^2 0.8 to rs1243180. The SNP rs10828247 coincides with a region of open chromatin at the 5′ end of the MLLT10 gene. See Supplementary figure 8ii and 9ii for additional information for all genes at this locus.
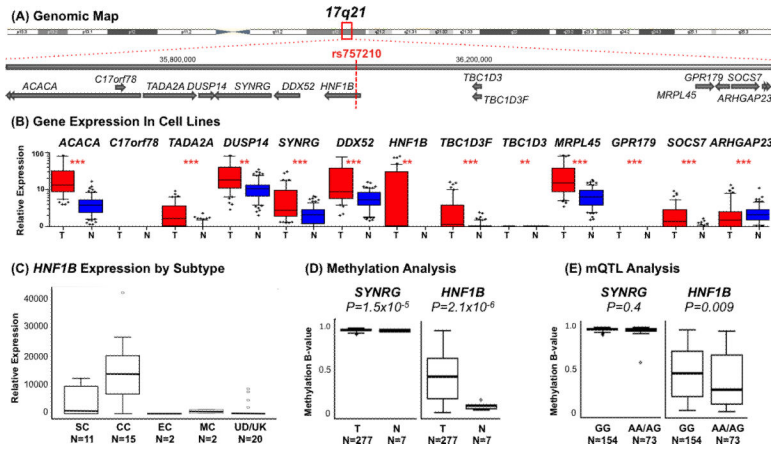
Figure 4. Summary of the functional analysis of the 17q locus

(A) Genomic map of a one-megabase region at 17q12 centered on the most statistically significant SNP, rs757210. The location and approximate size of all thirteen known protein-coding genes (grey) in the region are shown relative to the location of rs757210 (red dashed line). **(B)** Expression analysis for all genes at this locus performed in ovarian tumor (T) cell lines and normal (N) OSEC and FTSEC primary cultures illustrates the relative levels of expression for each gene (** p<0.01, ***p<0.001). **(C)** Overexpression of *HNF1B* detected in EOC cells compared to OSECs/FTSECs was largely driven by the high expression of this gene in clear cell EOC cell lines. *HNF1B* is an established clear cell EOC biomarker [23]**(D)** Methylation analysis of 277 high-grade serous ovarian tumors (T) compared to normal ovarian tissues (N) (n=7) shows significant hypermethylation of CpG sites upstream of *SYNRG* and *HNF1B* in tumors (T) compared to normal (N) tissues. **(E)** mQTL analysis showing methylation status of *SYNRG* and *HNF1B* genes in primary high grade serous ovarian cancers relative to germline genotypes for individuals carrying minor/heterozygous allele (AG/GG) and common alleles (AA) for rs757210. Only methylation at the *HNF1B* gene shows a significant association with genotype at this locus. See Supplementary figure 8iii and 9iii for additional information for all genes at this locus.

Table 1

Summary of samples and SNPs genotyped

Phase	Study population	No of studies	No. of cases	No. of controls	Illumina genotyping platform	SNPs		
						Genotyped	Passed QC	Imputed
Pooled GWAS	North American	5	1,952	2,042	610K/317K	620,901	599,179	1,909,565
	UK phase 1	6	1,817	2,354	610K/550K	620,901	507,094	2,001,650
	UK phase 2	10	4,162	4,810	Custom iSelect	23,590	21,955	-
COGS*	OCAC	43	18,549	26,134	Custom iSelect	23.239	22,252	-

*
Includes 2,482 samples in the North American GWAS, 1,641 samples in the phase 1 of the UK GWAS and 8,463 samples in Phase 2 of the UK GWAS

Table 2

Odds ratio and tests of association by histological subtype and population for most strongly associated SNPs at 8q21, 10p12 and 17q12

SNP	Locus	Ref/alt allele	Risk allele frequency	Subtype	Ancestry	Odds ratio	95% CI	P-value
rs11782652	8q21	a>g	0.07	All invasive	European	1.19	1.12 – 1.26	5.5×10^{-9}
			0.07	Serous	European	1.24	1.16 – 1.33	7.0×10^{-10}
			0.07	Endometrioid	European	1.04	0.92 – 1.19	0.50
			0.07	Clear cell	European	1.12	0.95 – 1.33	0.18
			0.07	Mucinous	European	0.95	0.78 – 1.15	0.58
			0.00	Serous	Asian	-	-	-
			0.07	Serous	African	1.73	0.92 – 3.28	0.91
			0.06	Serous	Mixed	1.15	0.81 – 1.64	0.43
rs1243180	10p12	t>a	0.31	All invasive	European	1.10	1.06 – 1.13	1.8×10^{-8}
			0.31	Serous	European	1.11	1.07 – 1.15	1.4×10^{-7}
			0.31	Endometrioid	European	1.08	1.00 – 1.15	0.038
			0.31	Clear cell	European	1.09	0.99 – 1.19	0.091
			0.31	Mucinous	European	0.97	0.87 – 1.07	0.50
			0.03	Serous	Asian	1.41	0.82 – 2.43	0.21
			0.06	Serous	African	1.11	0.50 – 2.48	0.80
			0.25	Serous	Mixed	0.87	0.72 – 1.07	0.18
rs757210	17q12	g>a	0.37	All invasive	European	1.05	1.02 – 1.09	0.00090
			0.37	Serous	European	1.12	1.08 – 1.17	8.1×10^{-10}
			0.37	Endometrioid	European	0.98	0.91 – 1.04	0.47
			0.37	Clear cell	European	0.80	0.72 – 0.88	3.9×10^{-6}
			0.37	Mucinous	European	0.89	0.81 – 0.99	0.027
			0.29	Serous	Asian	1.10	0.87 – 1.37	0.43
			0.53	Serous	African	1.08	0.75 – 1.54	0.69
			0.37	Serous	Mixed	0.86	0.73 – 1.02	0.093