HHS Public Access

Author manuscript

Nat Genet. Author manuscript; available in PMC 2016 February 01.

Published in final edited form as:

Nat Genet. 2015 August ; 47(8): 888–897. doi:10.1038/ng.3336.

Genome-wide significant risk associations for mucinous ovarian carcinoma

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

Abstract

Genome-wide association studies have identified several risk associations for ovarian carcinomas (OC) but not for mucinous ovarian carcinomas (MOC). Genotypes from OC cases and controls were imputed into the 1000 Genomes Project reference panel. Analysis of 1,644 MOC cases and 21,693 controls identified three novel risk associations: rs752590 at 2q13 ($P = 3.3 \times 10^{-8}$), rs711830 at 2q31.1 ($P = 7.5 \times 10^{-12}$) and rs688187 at 19q13.2 ($P = 6.8 \times 10^{-13}$). Expression Quantitative Trait Locus (eQTL) analysis in ovarian and colorectal tumors (which are histologically similar to MOC) identified significant eQTL associations for *HOXD9* at 2q31.1 in ovarian ($P = 4.95 \times 10^{-4}$, FDR = 0.003) and colorectal (P = 0.01, FDR = 0.09) tumors, and for *PAX8* at 2q13 in colorectal tumors (P = 0.03, FDR = 0.09). Chromosome conformation capture analysis identified interactions between the *HOXD9* promoter and risk SNPs at 2q31.1. Overexpressing *HOXD9* in MOC cells augmented the neoplastic phenotype. These findings provide the first evidence for MOC susceptibility variants and insights into the underlying biology of the disease.

Ovarian carcinomas (OC) caused approximately 140,000 cancer deaths globally in 2008[1]. Germline mutations in genes conferring high (*BRCA1* and *BRCA2*)[2] and more moderate risk (e.g., *TP53, BRIP1, RAD51D, RAD51C, MLH1, MSH2, PMS1, PMS2* and *MSH6*)[3-56] of OCs are among the best-defined genetic risk factors but explain only 10-15 percent of all OCs[6-8]. More recently, genome-wide association studies (GWAS) have identified multiple regions of the genome harboring common variants (minor allele frequency [MAF] > 0.05) conferring low risk (odds ratios [OR] < 1.5) of invasive OC[9-17]. However, it is increasingly

Corresponding author: Simon A Gayther, PhD, Department of Preventive Medicine, Keck School of Medicine, University of Southern California Norris Comprehensive Cancer Center, Los Angeles, CA, USA. simon.gayther@med.usc.edu.

†Authors contributed equally

Author contributions

LEK, KL, PDPP, SAG and AB wrote the manuscript. CMP and GC-T coordinated and prepared samples for genotyping. LEK, KL, JT and PDPP carried out data analysis including imputation, genotype quality control, association analysis, *in silico* analyses and *in vitro* analyses. QL and MLF performed the eQTL analysis. Remaining authors contributed samples or technical expertise. All authors read and approved the final version.

URLs

1000 Genomes Project, http://www.1000genomes.org/page.php; Australian Ovarian Cancer Study, http://www.aocstudy.org/; Cancer Genome Atlas Project data bioportal, http://www.cbioportal.org/; Linkage Disequilibrium Calculator, https://caprica.genetics.kcl.ac.uk/~ilori/ld_calculator.php; Polymorphism Phenotyping (PolyPhen) v2, http://genetics.bwh.harvard.edu/pph/; Sorting Intolerant From Tolerant (SIFT), http://sift.jcvi.org; The Cancer Genome Atlas Project, http://cancergenome.nih.gov/; UCSC Genome Browser, genome.ucsc.edu/ENCODE/; Wellcome Trust Case Control Consortium, http://www.wtccc.org.uk/.

Competing financial interests

The authors declare no competing financial interests.

recognized that OCs encompass multiple distinct disease histotypes[18] that vary in epidemiologic[19-21] and genetic[22] risk factors, somatic alterations[23,24] and clinical response to platinum-taxane based therapy[18]. Most of the known common risk alleles for OC confer susceptibility to the most common histotype, high grade serous (HGSOC), with one region also associated with the clear cell histotype at genome-wide significance[12]. There are as yet no reports of confirmed genome-wide significant susceptibility loci for the other main histotypes, mucinous and endometrioid OCs.

Mucinous ovarian carcinomas (MOCs) are characterized by multicystic tumors with conspicious amounts of intracellular mucin (usually ≥ 50 percent of the cytoplasm) in more than 90 percent of the tumor cells[25]. Historically, MOCs have been estimated to account for about 12 percent of all OCs but recent refinements in the morphologic assessment indicate that primary invasive MOCs comprise approximately 3 percent of all OCs[26]. This lower prevalence is due to several reasons including consensus by pathologists to separate benign mucinous tumors from invasive MOCs[27] and from pathology guidelines[26,28-30] that aim to distinguish primary invasive MOCs from metastatic carcinomas involving the ovary, in which the majority derive from organs of the gastrointestinal system[26,31]. These criteria, along with the frequent inability to find a non-ovarian primary cancer, suggest that true MOCs develop *de novo* at the ovary and cannot be explained by metastatic lesions. This low incidence has made it challenging to study the etiology and pathogenesis of these tumors.

At the genetic level, MOCs are not associated with germline *BRCA1/BRCA2* mutations. Unlike other OC histotypes, invasive MOCs usually harbor foci of benign or atypical (low malignant potential [LMP]) epithelium, with identical *KRAS* mutations frequently present[32-34], suggesting that this is an early somatic event in a multistep progression model. Normal mucin-secreting cells are not present in the ovary raising uncertainty regarding the cell at risk of transformation. It has been hypothesized that some MOCs originate from foci of benign endocervical-subtype Müllerian metaplasia of the surface epithelium or cortical inclusion cysts[35]. This subtype, however, may be less frequently associated with fully invasive MOCs, which comprise mostly the intestinal subtype[35]. To complicate further the etiology of MOCs, expression analysis of small numbers of MOCs (N = 3–9) associated these tumors more closely to colonic epithelium or colorectal carcinomas (CRC) than to ovarian surface epithelium[36,37], suggesting the pathogenesis of MOCs may be similar to colorectal carcinomas[38]. The current study reports the identification of genetic susceptibility alleles for MOCs, which may help to elucidate genes and biological pathways that are disregulated during MOC development.

Results

Genetic association analyses

We used genotypes from 16,038 ovarian cancer cases and 30,816 controls from various genotyping arrays providing genome-wide coverage (**Table 1**). Participating studies are listed in **Supplementary Table 1**[10,12,39]. We imputed these genotypes into a reference panel from the 1000 Genomes Project to provide observed or imputed genotypes at 15,504,273 variants (**Online Methods, Supplementary Table 2**). Genotype re-imputation without pre-phasing was carried out for regions of interest to improve accuracy (see **Supplementary**

Note). The primary association analyses reported in this paper were based on OCAC-COGS participants of European ancestry and those with invasive or LMP MOC, comprising 1,644 cases (1,003 invasive, 641 LMP) and 21,693 controls (**Table 1**). We identified SNPs in three different regions that were associated with MOC at genome-wide significance (**Table 2**, **Fig. 1 a–c**). Two regions (2q13 and 19q13.2) have not been previously associated with risk for other OC histotypes; the third region (2q31.1) has been reported to be associated with HGSOC[10].

At 2q13, the most strongly associated SNP, rs752590, was imputed (imputation $r^2 = 0.66$, effect allele frequency, EAF = 0.21). It is located 347 bases upstream of *PAX8* (paired box 8) and the effect allele was associated with increased risk for all MOC (OR = 1.34, 95% CI = 1.21–1.49, $P = 3.3 \times 10^{-8}$) (**Table 2**). The risk was similar for invasive and LMP cases (data not shown). At 19q13.2, the most strongly associated SNP, rs688187, was also imputed (imputation $r^2 = 0.55$, EAF = 0.32). It lies approximately 489kb downstream of *IFNL3* (interferon, lambda 3) and the effect allele was associated with decreased risk for all MOC (OR = 0.67, 95% CI = 0. 0.60–0.75, $P = 6.8 \times 10^{-13}$). Again there was little difference in risk between invasive and LMP cases (data not shown).

At 2q31.1, the most significantly associated SNP, rs711830 (EAF = 0.32), is located downstream of the 3' region of *HOXD3* (homeobox D3). The effect allele was associated with increased risk for all MOC (OR = 1.30, 95% CI = 1.20–1.40, $P = 7.5 \times 10^{-12}$) (**Table 2**) with similar risk for invasive and LMP MOC (data not shown). This SNP was also associated with invasive HGSOC (OR = 1.14, $P = 1.9 \times 10^{-13}$) (**Supplementary Table 3**). It is highly correlated ($r^2 = 0.99$) with rs2072590, the variant previously reported for HGSOC[10].

MOCs of extra-ovarian origin are more likely to be stage 3 tumors[26]. In our dataset, only 146/1,644 (8.9 %) MOC cases were stage 3 suggesting that the majority of diagnoses in this study are likely to be true primary ovarian MOCs. Risk estimates were also similar or larger (farther from the null) in women diagnosed with early-stage mucinous tumors compared to stage 3 tumors: OR = 1.39 (95% CI = 1.22–1.58, $P = 5.4 \times 10^{-7}$) for rs752590 at 2q13; OR = 1.28 (95% CI = 1.17–1.41, $P = 6.7 \times 10^{-8}$) for rs711830 at 2q31.1 and OR = 0.65 (95% CI = 0.56–0.75, $P = 5.9 \times 10^{-9}$) for rs688187 at 19q13.2, supporting our initial findings.

Assessment of imputation quality

We assessed the imputation accuracy between the SNPs achieving genome-wide significance by comparing the correlation between observed genotypes with estimated genotype doses from the imputation without pre-phasing for 2,739 OCAC cases of European ancestry (Online Methods). SNPs were selected for genotyping based on pre-phased imputation results. Where primer design failed and the original risk SNP could not be genotyped, a highly correlated ($r^2 = 1$) alternate SNP was selected (**Supplementary Table 4**). The correlation between observed and imputed genotypes was 0.59 for rs6542125 and rs6758928 (proxy SNPs for rs72831838, the top SNP at 2q13 for pre-phased imputed data) and 0.51 for rs35963157 (top SNP at 19q13.2 for pre-phased imputed data). These were close to the expected correlations based on the imputation r^2 (0.64 and 0.55, respectively),

and the similarity of the empirical to the imputation r^2 indicates the results are unlikely to be false positives.

Risk associations based on observed genotypes for 2,739 cases could not be evaluated, as we did not genotype any control samples. However, SNPs at MOC risk regions at 2q13 and 19q13.2 were not associated with other histotypes of OC (**Supplementary Table 3**). We, therefore, compared the results of a case-only analysis for 151 MOC cases and 2,588 cases with other invasive OC histotypes using imputed genotype doses and observed genotypes (**Supplementary Table 5**). We confirmed associations for the two alternate genotyped SNPs at 2q13, rs6542125 (case-only OR = 1.42, 95% CI = 1.07–1.90, P = 0.01) and rs6758928 (case-only OR = 1.41, 95% CI = 1.06–1.88, P = 0.02). While the ORs from genotyped samples were somewhat attenuated compared to the results for the imputed data, the P-values were smaller suggesting that the imputed association was robust. However, for rs35963157, there was no case-only association using imputed data in this subset of genotyped cases (case-only OR = 0.98, 95% CI = 0.69–1.39, P = 0.93). For the same samples, the association using observed genotypes was in the direction expected based on the full dataset, although not significant (case-only OR = 0.88, 95% CI = 0.67–1.14, P = 0.33). The association was somewhat stronger when an additional 1,274 cases (59 MOC cases and 1,215 other OC histotypes) without imputed data were included (case-only OR = 0.80, 95% CI = 0.64–1.00, P = 0.05) suggesting the observed association for the imputed data is not due to artifacts of imputation.

Functional annotation of variants in risk regions

At each of the three risk regions we identified all SNPs with a 1:100 or greater statistical odds of being the disease-causing variant (**Supplementary Tables 6, 7 and 8**). We annotated these SNPs with respect to exons, introns and untranslated regions (UTR) and epigenetic marks from two ovarian epithelial cell lines and two OC cell lines. Given the biological similarities between some MOC and CRC[40], we also annotated these SNPs for epigenetic marks profiled in a CRC cell line (HCT-116) and normal colonic tissues. The vast majority of the variants lie within non-coding DNA, suggesting they influence the function of non-coding regulatory elements.

At 2q13, there were 55 candidate SNPs spanning a 78.6 kb region encompassing most of *PAX8* and part of *PSD4* (pleckstrin and Sec7 domain containing 4) (**Fig. 2**). Most risk-associated variants were in *PAX8* introns, but the most statistically significant SNP (rs752590) and a moderately correlated variant (rs4849174, r^2 = 0.73) in an RNA polymerase II binding site lie within the *PAX8* proximal promoter. Three SNPs (rs874898, rs1478 and rs1479) lie within the 3' UTR of *PAX8* and so could influence RNA stability. Two SNPs (rs6734610 and rs7585510) lie within the sequence of the *PAX8-AS1* (*PAX8* antisense RNA 1) long non-coding (lnc) transcript and so may impact its stability or function (**Supplementary Table 8**). Eleven (20 percent) and 13 (24 percent) of SNPs lie in enhancer elements detected in ovarian and colonic cells, respectively. Of these, rs2305132, lies within a CTCF binding site detected in multiple cell types, suggesting this variant may be involved in the repression of *PAX8* and/or *PSD4* expression during MOC development.

At 19q13.2, there are 14 candidate SNPs located in and around *IFNL3* and *IFNL4* (**Supplementary Table 8**). Rs11882871 lies within the 3' UTR in *IFNL3* and rs4803222 lies within the 5' UTR in *IFNL4* suggesting they may influence RNA stability[41]. Rs11322783 is a coding SNP and predicted to cause a frameshift change in *IFNL4*, while rs8103142 encodes a missense change (lysine to arginine) in *IFNL3*. Rs8103142 is predicted to be non-deleterious by PolyPhen v2 and SIFT. There were no overlaps between these 14 risk SNPs and regulatory DNA elements.

At 2q31.1, there were 19 candidate causal variants spanning ~27 kb encompassing *HOXD3* and the lnc RNA *HAGLR* (*HOXD* antisense growth-associated long non-coding RNA) (**Fig. 2 and Supplementary Table 8**). There was extensive overlap between SNPs and regulatory elements in this region. Eleven and eight SNPs, respectively, coincided with putative enhancers in ovarian and colonic cells. Rs1051929 encodes a synonymous change in *HOXD3* and five SNPs lie within transcribed regions of the *HAGLR* and *HAGLR-OS* lnc RNA genes.

eQTL analysis

We evaluated associations between risk SNPs and transcript expression for all genes within a 100 kb window centered on the most risk-associated SNP in each region using publicly available data for 339 HGSOCs and 121 CRCs from The Cancer Genome Atlas (TCGA)[42,43]. No data were available for primary MOCs in TCGA. Where genotyping data were not available for a risk-associated SNP, correlated proxies ($r^2 > 0.7$) were evaluated. At 2q13, we detected a significant eQTL association between rs6542127, a variant highly correlated ($r^2 > 0.9$) with 6 risk-associated SNPs, and *PAX8* expression in CRCs ($P = 0.03$, FDR = 0.09) (**Fig. 3**). Rs6542127 was associated with MOC risk (OR = 1.20, $P = 8.81 \times 10^{-6}$). The most significant eQTL association with PAX8 expression at this locus was for rs2863243 ($P = 2.2 \times 10^{-6}$) but this SNP was not associated with MOC risk. However, the third most significant eQTL SNP (rs3748916, $P = 3.1 \times 10^{-4}$) was associated with MOC risk (OR = 0.84, $P = 9.37 \times 10^{-6}$). There were no statistically significant eQTL associations with *PAX8* expression in HGSOCs.

At 2q31.1, the most significant risk SNP (rs711830) in *HOXD3* was significantly associated with *HOXD3* expression in CRCs ($P = 0.01$, FDR = 0.09) but there was no eQTL association for *HOXD3* in HGSOCs. *HOXD9*, approximately 49 kb upstream of rs711830, is a candidate susceptibility gene for HGSOCs (K. Lawrenson et al, unpublished data). Rs711830 genotype was significantly associated with *HOXD9* expression in both HGSOCs ($P = 4.95 \times 10^{-4}$, FDR = 0.003) and CRCs ($P = 0.01$, FDR = 0.09) (**Fig. 3**). Another SNP in the region, rs10188827 ($r^2 = 0.59$ with rs711830), showed a slightly stronger eQTL association in HGSOC ($P = 2.05 \times 10^{-4}$) but a slightly weaker association with MOC risk (OR = 1.29, $P = 3.41 \times 10^{-10}$ for all mucinous cases). There was also a stronger association between rs10188827 and *HOXD9* expression in CRCs ($P = 0.003$), although the strongest eQTL association in CRCs was for another SNP, rs973456 ($P = 5.30 \times 10^{-5}$), which had not been imputed. There were no eQTL associations for genes in the 19q13.2 region in either the HGSOC or CRC tumor datasets.

Functional characterization using *in vitro* models of MOC

The eQTL analyses in HGSOCs and CRCs suggested *HOXD9* is a candidate susceptibility gene and target of MOC risk SNPs. We therefore evaluated the role of *HOXD9* in MOC development. We used chromosome conformation capture (3C) to establish if any of the risk SNPs at 2q31 interacted physically with the *HOXD9* promoter in MOC EFO-27 cell lines . We found interactions for DNA fragments containing rs2072590, rs2857532 and rs4972504. These interactions spanned 31–55kb of genomic DNA and were confirmed by sequencing (**Fig. 4**). These genotyped variants were highly correlated with rs711830 (r^2 = 1, 0.98 and 0.89, respectively), the most significant risk SNP in the region. Of the three SNPs, rs2072590 showed the greatest overlap with epigenetic marks, coinciding with enhancer marks in OC cell lines, colon cancer cells and colonic crypts (**Supplementary Table 8**). Taken together, the results of eQTL and 3C analyses indicated that DNA regions at 2q31.1 harboring MOC risk SNPs are involved in the regulation of *HOXD9* expression. Future studies using genome-editing approaches to manipulate the different alleles of rs2072590 will be needed to evaluate the effects on both regulatory activity and *HOXD9* expression and confirm the role of this SNP and *HOXD9* in MOC development.

We also evaluated the effects of overexpressing *HOXD9* in two MOC cell lines (EFO-27 and GTFR230) using lentiviral transduction of a full length HOXD9 GFP fusion construct. Overexpression of *HOXD9* was confirmed by qPCR and immunofluorescence microscopy (**Fig. 5a**). HOXD9 expression was only detected in the nucleus, whereas in control cells expressing GFP only, GFP signal was detected throughout the cell. *HOXD9* overexpression induced a significant increase in anchorage-independent growth in both MOC cell lines (P = 0.02 in EFO-27 and P = 0.04 in GTFR230, **Fig. 5b**) indicating a role in neoplastic transformation. We observed no effect on cellular invasion and migration (data not shown).

Discussion

GWAS have identified common low-penetrance genetic susceptibility alleles for a multitude of common traits and diseases. As the size and scope of GWAS increase, so the power to identify risk alleles for rare disease subtypes has also increased. For OCs, the vast majority of confirmed risk associations from GWAS were for HGSOC, which accounts for almost two-thirds of all invasive OCs[24]. We report, for the first time, genome-wide significant risk associations for the rarer MOC histotype, identified as part of the largest genetic association study yet performed for OC.

Two of the three susceptibility regions identified for MOCs (2q13 and 19q13.2) are specific to this histotype, which may not be surprising given that MOCs are clinically and biologically distinct from the other OC histotypes. The third region associated with MOC risk (2q31.1) was previously reported as a susceptibility locus for HGSOC[10]. Similarly, the 17q12 risk region encompassing *HNF1B* was reported to be associated with HGSOC and the clear cell OC histotype[12] suggesting that the different OC histotypes have some degree of shared germline genetic etiology despite differences in somatic genetic alterations[22-24], epidemiologic risk factors[19-21] and response to standard chemotherapy[18]. This may reflect the influence of the site of tumor development (i.e., the ovary) and the possible functional

role of risk alleles interacting with common processes involved in malignancy, such as the ovarian microenvironment.

We identified *PAX8* at 2q13 and *HOXD9* at 2q31.1 as candidate MOC susceptibility genes using eQTL analysis of primary HGSOCs and CRCs. CRCs share some molecular and histologic characteristics with MOC[40]; however, gene expression and the functional mechanisms of risk-associated SNPs may be tissue specific[44]. We were unable to perform eQTL analyses in normal tissues or primary MOCs due to the lack of publicly available datasets for this tumor histotype and uncertainty of the likely cell(s) of origin of MOCs. While the eQTL associations we identified were statistically significant, it is possible that eQTLs exist between MOC risk SNPs and other genes, either within these regions or regulated more distally. For example at 2q13 we also observed regional associations of similar statistical significance to *PAX8* for *PSD4* and a *PAX8*-antisense transcript *LOC654433*, although neither gene has been previously implicated in MOC development. We also found evidence of stronger eQTL associations for SNPs with weaker risk associations at both loci. In addition to disease heterogeneity, eQTL analyses are complicated by intra-tumor heterogeneity due to variation in copy number, methylation and gene expression. Thus, caution needs to be applied when interpreting eQTL data. Additional analyses in larger sample sizes and in tissues more relevant to MOC etiology will be needed to confirm the significance of *HOXD9* and *PAX8* as likely susceptibility genes for MOC at these loci.

Functional studies suggest that *HOXD9* is the target MOC susceptibility gene at 2q31.1, through its interaction with three different regions harboring MOC risk SNPs and its ability to enhance neoplastic phenotypes when overexpressed in MOC cells. *HOXD9* is also a candidate susceptibility gene for HGSOC (K. Lawrenson et al, unpublished data). The results from 3C analysis showed that one of the three *HOXD9* interacting regions in MOC cells (containing rs4972404) also interacts with *HOXD9* in HGSOC cells. This suggests that similar functional mechanisms regulating *HOXD9* expression are acting in both MOC and HGSOC, but that the other two interacting regions were tissue-specific to MOCs, indicating regions that control regulation of *HOXD9* in MOCs but not in HGSOCs. *HOXD9* is a member of the HOX family of transcription factors that are only expressed during embryonic development to control patterning and differentiation. *HOXD9* has not been well characterized in the context of cancer development, although the gene was aberrantly expressed in cervical cancer[45] and has been implicated as a marker of cancer stem cells in glioma[46].

The 2q13 MOC risk region has not previously been associated with risk of other diseases or traits. *PAX8* is a plausible candidate susceptibility gene target at this locus. It encodes a transcription factor important in the development of the Müllerian duct[47] and may be a cell-type lineage marker that distinguishes carcinomas of gynecologic origin (e.g., ovary, uterus, peritoneum and fallopian) from other sites such as the gastrointestinal tract[23,47,48]. PAX8 is overexpressed in the majority of HGSOCs compared to normal ovarian epithelial cells[23], due partly from gene amplification[49], but is expressed in 10 percent[23] to 25 percent (L.E. Kelemen et al, unpublished data) of LMP and invasive MOCs and is not expressed in CRC cell lines[49]. Although the precise role of *PAX8* in cancer development is unclear, *PAX8*

expression may be important for acquiring characteristics that maintain a malignant state, including repression of differentiation programs for specific tissue lineages[49-51].

The 19q13.2 risk region has been associated with impaired clearance of hepatitis C virus[52] and variation in response to hepatitis C therapy in Asians involving *IFNL3* (also known as *IL-28B*)[53]. The most likely functioning risk SNP in this region is upstream of *IFNL3* in the coding region of *IFNL4*: rs11322783 (also annotated as rs368234815 in dbSNP 141) is a predicted truncating variant, suggesting it has a loss-of-function role. The insertion allele turns *IFNL4* into a polymorphic pseudogene and abolishes its activity[54]. The variant rs8103142 causes a nonsynonymous coding change (Arg70Lys) in *IFNL3* but this change was predicted to be non-deleterious *in vitro*[55]. There are no reports implicating *IFNL4* or *IFNL3* specifically in the development of OCs or CRCs, although multiple reports have indicated a role for interleukins more broadly in OC. The 19q13 region has also been associated with structural rearrangements in OCs[56].

The novel risk associations we found at 2q13 and 19q13.2 were identified using imputed genotypes and were based on an estimated imputation r^2 that was moderate for both SNPs. The imputation r^2 is an estimation of the expected correlation between imputed genotypes and actual genotypes. Confirmation genotyping of imputed SNPs in a subset of the samples showed that the estimated imputation correlation was similar to the correlation between imputed and observed genotypes. Furthermore, case-only associations for these SNPs based on observed genotypes were associated with smaller *P*-values, providing support for the imputed genotype associations. While it is possible that imputation may be sensitive to small genotyping errors and differential with respect to case-control status, we would expect this bias to apply equally to all cases and not specifically to MOC cases. Since we did not observe significant associations with the other histotypes, the collective findings suggest that the associations with MOC were not due to biases of the imputation process.

The relatively large number of invasive and LMP MOCs in this study represents a significant strength. We combined genotyping data from patients diagnosed with invasive and LMP MOCs because these tumors are thought to evolve along a morphologic continuum[32-34]. In the three risk regions, the statistical significance of the risk associations was stronger in the combined analysis compared to the LMP dataset alone. However, molecular epidemiologic studies are limited by access to details on clinical presentation and by the difficulty to perform centralized histologic review. Although it is reasonable to assume that most LMP tumors arose primarily in the ovary, review of the histology and clinical records of the cancers might have led to exclusion of some cases that were metastases to the ovaries from non-ovarian primary cancers. To address this, we evaluated risk associations for early-stage MOCs separately and observed similar or larger effect estimates. However, even when all the relevant clinical information is available and immunohistochemistry is performed, it is sometimes impossible to be certain whether an MOC arose in the ovary or elsewhere[40]. Notably, we found no overlap between the risk associations we reported for MOC and those discovered in GWAS of gastrointestinal cancers[57] suggesting that the invasive MOCs in this study were mostly primary OCs rather than metastases.

In summary, we reported the first genome-wide significant alleles to be identified for MOC. The power to detect risk associations for MOCs has so far been limited by the small numbers of MOC cases collected through OCAC. The experiences of GWAS for more common cancers (e.g., breast[58] and prostate[59]) indicate that with larger number of MOC cases, we would expect to identify additional susceptibility alleles for MOC. The functional evaluation that we performed for the MOC risk-associated regions also suggests that future studies are likely to provide new insights into the understanding of the biology of MOC. Finally, because MOC and HGSOC are distinct OC histotypes and can be considered separate diseases[23,24], the identification of *HOXD9* as a potential gene target showing oncogenic characteristics in both MOC and HGSOC can be considered independent evidence for the general role of this gene in oncogenesis.

Online Methods

Genetic association studies

We used genotypes from samples of European ancestry available from several Ovarian Cancer Association Consortium (OCAC) genotyping projects. Data were available for five population-based GWAS of ovarian carcinoma. These included 1,785 cases (206 MOC) and 6,118 controls from a UK-based GWAS ("UK GWAS")[9], 2,166 cases (116 MOC) and 2,564 controls from a GWAS from North America ("US GWAS")[61] including two smaller GWAS from New England-Brigham and Women's Hospital ("NEC/BWH") and the National Cancer Institute-Polish study ("NCI-POL"), and 467 cases (36 MOC) and 441 controls from another GWAS from North America ("Mayo GWAS")[17]. An additional 11,620 cases (1,286 MOC) and 21,693 controls from 41 OCAC studies were genotyped using the COGS array ("OCAC-COGS")[12]. The UK and US GWAS comprised several independent case-control studies, and samples from some of these studies were also subsequently genotyped using the COGS array. After removing duplicate samples, remaining samples represented 43 studies from 11 countries including 16,038 women diagnosed with invasive ovarian carcinoma, 1,644 of whom were diagnosed with MOC, and 30,816 controls from the general population. Informed consent was obtained in each of the individual studies and local human research investigations committees approved each study. Details of the genotyping design for each dataset are shown in **Table 1**. Further details of the component studies are found in **Supplementary Table 1**.

Genotyping and quality control

The final sample sizes for each of the five GWAS and OCAC-COGS are shown in **Table 1**. Details for genotyping and quality control for each dataset are found in the **Supplementary Note**.

Imputation

To impute unobserved genotypes of common variation across the entire genome, we performed imputation separately for OCAC-COGS samples and each of the OCAC GWAS samples that passed QC (**Table 1** and **Supplementary Table 1**). We imputed variants from the 1000 Genomes Project reference panel (Integrated Phase 1, version 3, March 2012 release). To improve computation efficiency we initially used a two-step procedure, which

involved pre-phasing in the first step and imputation of the phased data in the second. We carried out pre-phasing using the SHAPEIT software[62]. Unobserved genotypes were inferred probabilistically with IMPUTE2 version 2[63]. To perform the imputation we divided the data into nonoverlapping segments of approximately 5Mb each. We excluded SNPs from the association analysis if their imputation accuracy was $r^2 < 0.25$ or their MAF was < 0.005. In total, of 15,504,273 SNPs were imputed successfully (**Supplementary Table 2**).

Re-imputation without pre-phasing

Following detection of a genome-wide significant association, we re-imputed regions of 1 Mb surrounding the SNP using IMPUTE2 software[63] without pre-phasing and the most recent 1000 Genomes Project reference panel (June 2014 data release) to improve accuracy. To increase the imputation accuracy even further, we changed some of the default parameters in the imputation procedure. These included an increase of the Markov Chain Monte Carlo iterations to 90 (out of which the first 15 were used as burn-in), an increase of the buffer region to 500 Mb and an increase of the number of haplotypes used as templates when phasing observed genotypes to 100. These changes were applied consistently for all datasets. This two-stage process resulted in different SNPs achieving the most significant P value between pre-phased and re-imputed data. The main findings focus on the associations obtained from re-imputation.

Imputation quality

Lymphocyte DNA from 4,013 OCAC cases of European ancestry, of which 2,739 were represented in COGS, was genotyped using the iPLEX Gold assay (Sequenom, Inc) in order to compare the accuracy of imputed and actual genotype doses. SNPs showing genome-wide significance were selected for genotyping from the pre-phased imputed results. Where primer design failed for the imputed pre-phased GWAS SNP, an alternate SNP was selected for genotyping within a region of high linkage disequilibrium (LD). Using the r^2 coefficient, we compared the imputation accuracy between genotype doses derived from imputation using both the pre-phased and re-imputed results and those derived from actual genotyping.

Statistical analysis

Association testing was restricted to OCAC-COGS participants of European ancestry and those with invasive or LMP MOC, resulting in 1,644 cases (1,003 invasive-only and 641 LMP) and 21,693 controls. Genotypes (both typed and imputed) for each SNP were used to estimate allele frequencies and pair-wise LD between two variants was estimated with r^2 values based on the 1000 Genomes Project reference panel using an online program (see URLs).

We estimated per-allele log odds ratios (OR) and 95% confidence intervals (CI) between each SNP and MOC risk using unconditional logistic regression, where number of variant alleles carried was treated as an ordinal variable (log-additive, co-dominant model). The likelihood ratio statistic was used to examine association; this statistic has been shown to have greater power for rare variants than alternatives such as the Wald test or the score test [64]. The logistic regression model was adjusted for study stratum and population substructure by including the first five eigenvalues from principal components analyses (see

ref. [12]). Analyses were performed separately for combined invasive and LMP MOC and invasive-only MOC. Statistical P-values were two-sided and, unless stated otherwise, were implemented with STATA version 13.0 (StataCorp LP) and SAS version 9 (SAS Institute). P-values $< 5 \times 10^{-8}$ were considered to be genome-wide significant.

Molecular and functional analyses

Analysis of epigenetic biofeatures—We used in-house FAIREseq and H3K27ac plus H3K3me1 ChIPseq data for "normal" ovarian epithelial cell lines and serous ovarian cancer cell lines (G. Coetzee et al, unpublished data). HCT116 and primary tissue data (colon and ovarian) were downloaded as BED files from Hnisz et al[60]. CTCF data were obtained from ENCODE using the UCSC Genome Browser (see URLs).

Expression quantitative trait locus (eQTL) analyses—For each MOC SNP, we first identified correlated variants ($r^2 > 0.7$) in the 1000 Genomes European ancestry population. Germline genotypes of 339 high grade serous ovarian cystadenocarcinoma and 121 colorectal cancer samples were downloaded from The Cancer Genome Atlas (TCGA) data portal and samples were selected for inclusion in the eQTL analyses using EIGENSTRAT[65]. For each sample, tumor gene expression profiles, somatic copy number and CpG methylation data were downloaded from TCGA portal (see URLs). Expression profiles were adjusted for somatic copy number and CpG methylation variation as previously described[66,67]. For each SNP we evaluated the correlation between the genotype and adjusted expression levels of candidate genes. A false-discovery-rate (FDR) below 0.1 was considered to be a significant *cis*-association.

Cell culture—The EFO-27 cell line is commercially available and was originally derived from a solid metastasis of a mucinous papillary ovarian carcinoma[68]. The GTFR230 mucinous ovarian cancer cell line was created in house and derived from a stage IC low grade primary mucinous ovarian cancer collected as part of the Gynecological Tissue and Fluid Repository (GTFR) at the University of Southern California (USC). We routinely profile cell lines using STR profiling (using the Powerplex 16 panel) at the University of Arizona Genetics Core facility and screen for mycoplasma using a mycoplasma-specific PCR. A piece of the tissue removed during surgery was transferred to the cell culture laboratory, minced into small pieces (1–2 mm diameter) and placed into NOSE-CM[69] which consists of Medium 199 mixed in a 1:1 ratio with MCDB105 (Sigma), 15% fetal bovine serum (FBS, Hyclone), 10 ng/ml epidermal growth factor (EGF), 34 μg/ml bovine pituitary extract (Life Technologies), 500ng/ml hydrocortisone and 5 μg/ml insulin (Sigma). Cells were confirmed to be epithelial in origin by staining for cytokeratin.

Chromosome conformation capture (3C)—3C was performed as previously described[70]. A sample of 10 million EFO-27 cells was fixed with 1% formaldehyde, then lysed in 10 mM Tris-HCl pH 8, 10 mM NaCl, 0.2% Nonidet P-40. Nuclei were pelleted and then resuspended in 1X restriction enzyme buffer with 0.1% SDS and 1.6% Triton-X. Fifteen hundred units of Csp6i (Fermentas) enzyme were added. Samples were incubated at 37 °C overnight to digest after which 1.5% SDS was added before de-crosslinking samples by incubating at 65 °C for 30 minutes. Digested genomic DNA was then ligated by

incubating samples in ligation buffer plus 4000U T4 DNA ligase (NEB) for 24 hr at 16°C. Control samples received no ligase. Following ligation, samples were de-crosslinked by overnight incubation at 65 °C with proteinase K. Chromosome conformation capture libraries were extracted using phenol/chloroform, the DNA precipitated using ethanol, and desalted using Microcon Ultra Cell YM –100 columns. Two primers were designed at the *HOXD9* promoter and one primer for each restriction fragment of interest (Supplementary Table 9). PCR was performed using HotStar Taq polymerase (Qiagen), using the following conditions: 5 min at 94 °C, 42 cycles of (20s at 94 °C, 20s at 57–59 °C, and 30s at 72 °C), then 10 min at 72 °C. PCR products were run on a 1.2% agarose gel. PCR products were either cleaned up using the QIAgen QIAquick PCR Purification kit or were gel purified using the QIAgen QIAquick Gel Extraction kit. Cleaned products were sequenced from both ends by Genewiz.

In vitro modeling of HOXD9 overexpression—*HOXD9*-GFP and GFP lentiviral vectors were purchased from Genecopeia. Lentiviral supernatants were produced by co-transfecting HEK293T cells with vectors of interest, plus lentiviral packaging vectors. Supernatants were snap frozen and stored at –80°C until use. EFO-27 and GTFR230 cells were transduced with *HOXD9*-GFP and GFP viral supernatants in the presence of 4 g/ml polybrene (Sigma). Cells expressing GFP alone were controls. RNA was extracted from selected cells using the Zymo Quick-RNA kit, reverse transcribed using the MMLV enzyme (Promega) and gene expression analysis performed using TaqMan probes (Life Technologies). Gene expression for *HOXD9* was normalized to expression of *GAPDH* and beta-actin using the delta-delta Ct method. For immunofluorescence, cells were fixed on glass coverslips using 4% paraformaldehdye and nuclei stained using Hoechst. Cells were imaged using a Zeiss Axio Imager Z1 fluorescent microscope. Positive stained cells were selected with puromycin at concentrations of 1 mg/ml (EFO-27) and 400 ng/ml (GTFR230). Anchorage-independent growth assays were performed in 6-well plates by re-suspending 20,000 cells per well in media containing 0.33% Noble agar and 1 mg/ml bacto-peptone (both Sigma) and plating atop 3 ml of media containing 0.6% Noble agar and 1 mg/ml bacto-peptone. Assays were incubated at 37 °C for 28 days before fixing and staining with 1 mg/ml p-iodo tetrazolium violet (Sigma) dissolved in 100% methanol (VWR). For invasion and migration assays, Cultrex 96-well invasion and migration kits were used according to the manufacturer's instructions.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Authors

Linda E. Kelemen[1,†], Kate Lawrenson[2,†], Jonathan Tyrer[3], Qiyuan Li[4,5], Janet M. Lee[2], Ji-Heui Seo[5], Catherine M. Phelan[6], Jonathan Beesley[7], Xiaoqin Chen[7], Tassja J. Spindler[2], Katja K.H. Aben[8,9], Hoda Anton-Culver[10], Natalia Antonenkova[11], Australian Cancer Study[12], Australian Ovarian Cancer Study Group[12], Helen Baker[13], Elisa V. Bandera[14], Yukie Bean[15,16], Matthias W. Beckmann[17], Maria Bisogna[18], Line Bjorge[19,20], Natalia Bogdanova[21], Louise A.

Brinton[22], Angela Brooks-Wilson[23,24], Fiona Bruinsma[25], Ralf Butzow[26,27], Ian G. Campbell[28,30], Karen Carty[31], Jenny Chang-Claude[32], Y. Ann Chen[33], Zhihua Chen[33], Linda S. Cook[34], Daniel W. Cramer[35,36], Julie M. Cunningham[37], Cezary Cybulski[38], Agnieszka Dansonka-Mieszkowska[39], Joe Dennis[13], Ed Dicks[13], Jennifer A. Doherty[40], Thilo Dörk[21], Andreas du Bois[41,42], Matthias Dürst[43], Diana Eccles[44], Douglas T. Easton[13], Robert P. Edwards[45], Ursula Eilber[32], Arif B. Ekici[46], Svend Aage Engelholm[47], Peter A. Fasching[17,48], Brooke L. Fridley[49], Yu-Tang Gao[50], Aleksandra Gentry-Maharaj[51], Graham G. Giles[25,52,53], Rosalind Glasspool[31], Ellen L. Goode[54], Marc T. Goodman[55,56], Jacek Grownwald[38], Patricia Harrington[3], Philipp Harter[41,42], Hanis Nazihah Hasmad[57], Alexander Hein[17], Florian Heitz[41,42], Michelle A.T. Hildebrandt[58], Peter Hillemanns[59], Estrid Hogdall[60,61], Claus Hogdall[62], Satoyo Hosono[63], Edwin S. Iversen[64], Anna Jakubowska[38], Allan Jensen[61], Bu-Tian Ji[22], Beth Y Karlan[65], Melissa Kellar[15,16], Joseph L. Kelley[45], Lambertus A. Kiemeney[8], Camilla Krakstad[19,20], Susanne K. Kjaer[61,62], Jolanta Kupryjanczyk[39], Diether Lambrechts[66,67], Sandrina Lambrechts[68], Nhu D. Le[69], Alice W. Lee[2], Shashi Lele[70], Arto Leminen[27], Jenny Lester[65], Douglas A. Levine[18], Dong Liang[71], Jolanta Lissowska[72], Karen Lu[73], Jan Lubinski[38], Lene Lundvall[62], Leon F.A.G. Massuger[74], Keitaro Matsuo[75], Valerie McGuire[76], John R. McLaughlin[77], Iain McNeish[78], Usha Menon[51], Francesmary Modugno[45,79,80], Joanna Moes-Sosnowska[39], Kirsten B. Moysich[70], Steven A. Narod[81], Lotte Nedergaard[62], Roberta B. Ness[82], Heli Nevanlinna[27], Mat Adenan Noor Azmi[83], Kunle Odunsi[84], Sara H. Olson[85], Irene Orlow[85], Sandra Orsulic[65], Rachel Palmieri Weber[86], James Paul[31], Celeste Leigh Pearce[2,87], Tanja Pejovic[15,16], Liisa M. Pelttari[27], Jennifer Permuth-Wey[6], Malcolm C. Pike[2,85], Elizabeth M. Poole[35,88], Susan J. Ramus[2], Harvey A. Risch[89], Barry Rosen[90], Mary Anne Rossing[91,92], Joseph H. Rothstein[76], Anja Rudolph[32], Ingo B. Runnebaum[43], Iwona K. Rzepecka[39], Helga B. Salvesen[19,20], Joellen M. Schildkraut[86,93], Ira Schwaab[94], Xiao-Ou Shu[95], Yurii B Shvetsov[96], Nadeem Siddiqui[97], Weiva Sieh[76], Honglin Song[3,13], Melissa C. Southey[29], Lara Sucheston[70], Ingvild L. Tangen[19,20], Soo-Hwang Teo[57,98], Kathryn L. Terry[35,36], Pamela J Thompson[55,56], Shelley S. Tworoger[35,88], Anne M. van Altena[74], Els Van Nieuwenhuysen[68], Ignace Vergote[68], Robert A. Vierkant[99], Shan Wang-Gohrke[100], Christine Walsh[65], Nicolas Wentzensen[22], Alice S. Whittemore[76], Kristine G. Wicklund[91], Lynne R. Wilkens[96], Sawicki Wlodzimierz[101], Yin-Ling Woo[83,98], Xifeng Wu[58], Anna H. Wu[2], Hannah Yang[22], Wei Zheng[95], Argyrios Ziogas[10], Thomas A. Sellers[6], Matthew L. Freedman[5], Georgia Chenevix-Trench[7], Paul D. Pharoah[3,13], Simon A. Gayther[2], and Andrew Berchuck[102] **on behalf of the Ovarian Cancer Association Consortium**

Affiliations

[1] Department of Public Health Sciences, College of Medicine, Medical University of South Carolina, Charleston, SC, USA [2] Department of Preventive Medicine, Keck School of Medicine, University of Southern California Norris Comprehensive Cancer Center, Los Angeles, CA, USA [3] The Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK [4] Medical

College, Xiamen University, Xiamen, China [5] Department of Medical Oncology, The Center for Functional Cancer Epigenetics, Dana-Farber Cancer Institute, Boston, MA, USA [6] Department of Cancer Epidemiology, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL, USA [7] Department of Cancer Genetics, QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia [8] Radboud University Medical Centre, Radboud Institute for Health Sciences, Nijmegen, The Netherlands [9] Comprehensive Cancer Centre, The Netherlands, Utrecht, The Netherlands [10] Department of Epidemiology, Genetic Epidemiology Research Institute, School of Medicine, University of California Irvine, Irvine, California, USA [11] Byelorussian Institute for Oncology and Medical Radiology Aleksandrov N.N., Minsk, Belarus [12] A list of members is provided in the Supplementary Note [13] The Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK [14] Cancer Prevention and Control Program, Rutgers Cancer Institute of New Jersey, The State University of New Jersey, New Brunswick, NJ, USA [15] Department of Obstetrics & Gynecology, Oregon Health & Science University, Portland, OR, USA [16] Knight Cancer Institute, Oregon Health & Science University, Portland, OR, USA [17] Department of Gynecology and Obstetrics, University Hospital Erlangen, Friedrich-Alexander University Erlangen-Nuremberg, Comprehensive Cancer Center, Erlangen, Germany [18] Gynecology Service, Department of Surgery, Memorial Sloan Kettering Cancer Center, New York, NY, USA [19] Department of Gynecology and Obstetrics, Haukeland University Horpital, Bergen, Norway [20] Centre for Cancer Biomarkers, Department of Clinical Science, University of Bergen, Bergen, Norway [21] Gynaecology Research Unit, Hannover Medical School, Hannover, Germany [22] Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA [23] Canada's Michael Smith Genome Sciences Centre, British Columbia Cancer Agency, Vancouver, BC, Canada [24] Department of Biomedical Physiology and Kinesiology, Simon Fraser University, Burnaby, BC, Canada [25] Cancer Epidemiology Centre, Cancer Council Victoria, Melbourne, VIC, Australia [26] Department of Pathology, University of Helsinki and Helsinki University Hospital, Helsinki, Finland [27] Department of Obstetrics and Gynecology, University of Helsinki and Helsinki University Hospital, Helsinki, Finland [28] Cancer Genetics Laboratory, Research Division, Peter MacCallum Cancer Centre, St Andrews Place, East Melbourne, VIC, Australia [29] Department of Pathology, University of Melbourne, Melbourne, VIC, Australia [30] Sir Peter MacCallum Department of Oncology, University of Melbourne, Melbourne, VIC, Australia [31] Cancer Research UK Clinical Trials Unit, The Beatson West of Scotland Cancer Centre, Glasgow, UK [32] Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany [33] Department of Biostatistics, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL, USA [34] Division of Epidemiology and Biostatistics, Department of Internal Medicine, University of New Mexico, Albuquerque, NM, USA [35] Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA [36] Obstetrics and Gynecology Epidemology Center, Brigham and Women's Hospital, Boston, MA, USA [37] Department of Laboratory Medicine and Pathology, Division of

Experimental Pathology, Mayo Clinic, Rochester, MN, USA [38] International Hereditary Cancer Center, Department of Genetics and Pathology, Pomeranian Medical University, Szczecin, Poland [39] Department of Pathology and Laboratory Diagnostics, The Maria Sklodowska-Curie Memorial Cancer Center and Institute of Oncology, Warsaw, Poland [40] Department of Epidemiology, The Geisel School of Medicine at Dartmouth, Lebanon, NH, USA [41] Department of Gynecology and Gynecologic Oncology, Kliniken Essen-Mitte/ Evang. Huyssens-Stiftung/ Knappschaft GmbH, Essen, Germany [42] Department of Gynecology and Gynecologic Oncology, Dr. Horst Schmidt Kliniken Wiesbaden, Wiesbaden, Germany [43] Department of Gynecology, Jena University Hospital-Friedrich Schiller University, Jena, Germany [44] Faculty of Medicine, University of Southampton, Southampton, UK [45] Department of Obstetrics, Gynecology and Reproductive Sciences, Division of Gynecologic Oncology and Ovarian Cancer Center of Excellence, University of Pittsburgh, Pittsburgh, PA, USA [46] Institute of Human Genetics, Friedrich-Alexander-University Erlangen-Nuremberg, Comprehensive Cancer Center Erlangen Nuremberg, Erlangen, Germany [47] Department of Radiation Oncology, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark [48] Department of Medicine, Division of Hematology and Oncology, David Geffen School of Medicine, University of California at Los Angeles, Los Angeles, CA, USA [49] Department of Biostatistics, University of Kansas Medical Center, Kansas City, KS, USA [50] Shanghai Cancer Institute, Shanghai, China [51] Gynaecological Cancer Research Centre, Department of Women's Cancer, Institute for Women's Health, University College London, London, UK [52] Centre for Epidemiology and Biostatistics, University of Melbourne, VIC, Australia [53] Department of Epidemiology and Preventive Medicine, Monash University, Melbourne, VIC, Australia [54] Department of Health Sciences Research, Division of Epidemiology, Mayo Clinic, Rochester, MN, USA [55] Cancer Prevention and Control, Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, California, USA [56] Community and Population Health Research Institute, Department of Biomedical Sciences, Cedars-Sinai Medical Center, Los Angeles, California, USA [57] Cancer Research Initiatives Foundation, Sime Darby Medical Centre, Subang Jaya, Malaysia [58] Department of Epidemiology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA [59] Clinics of Obstetrics and Gynaecology, Hannover Medical School, Hannover, Germany [60] Department of Pathology, Herlev Hospital, University of Copenhagen, Copenhagen, Denmark [61] Department of Virus, Lifestyle and Genes, Danish Cancer Society Research Centre, Copenhagen, Denmark [62] Department of Gynaecology, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark [63] Division of Epidemiology and Prevention, Aichi Cancer Center Research Institute, Nagoya, Aichi, Japan [64] Department of Statistical Science, Duke University, Durham, NC, USA [65] Women's Cancer Program at the Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, CA, USA [66] Vesalius Research Center, University of Leuven, Leuven, Belgium [67] Laboratory for Translational Genetics, Department of Oncology, University of Leuven, Belgium [68] Division of Gynecologic Oncology,

Department of Obstetrics and Gynaecology and Leuven Cancer Institute, University Hospitals Leuven, Leuven, Belgium [69] Cancer Control Research, British Columbia Cancer Agency, Vancouver, BC, Canada [70] Department of Cancer Prevention and Control, Roswell Park Cancer Institute, Buffalo, NY, USA [71] College of Pharmacy and Health Sciences, Texas Southern University, Houston, TX, USA [72] Department of Cancer Epidemiology and Prevention, M. Sklodowska-Curie Memorial Cancer Center and Institute of Oncology, Warsaw, Poland [73] Department of Gynecologic Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA [74] Radboud University Medical Centre, Department of Obstetrics and Gynecology, Nijmegen, The Netherlands [75] Department of Preventive Medicine, Kyushu University Faculty of Medical Science, Nagoya, Aichi, Japan [76] Department of Health Research and Policy - Epidemiology, Stanford University School of Medicine, Stanford, CA, USA [77] Public Health Ontario, Toronto, ON, Canada [78] Institute of Cancer Sciences, University of Glasgow, Wolfson Wohl Cancer Research Centre, Beatson Institute for Cancer Research, Glasgow, UK [79] Department of Epidemiology, University of Pittsburgh Graduate School of Public Health, Pittsburgh, PA, USA [80] Womens Cancer Research Program, Magee-Womens Research Institute and University of Pittsburgh Cancer Institute, Pittsburgh, PA, USA [81] Women's College Research Institute, University of Toronto, Toronto, ON, Canada [82] The University of Texas School of Public Health, Houston, TX, USA [83] Department of Obstetrics and Gynaecology, University Malaya Medical Centre, University Malaya, Kuala Lumpur, Malaysia [84] Department of Gynecological Oncology, Roswell Park Cancer Institute, Buffalo, NY, USA [85] Memorial Sloan Kettering Cancer Center, Department of Epidemiology and Biostatistics, New York, NY, USA [86] Department of Community and Family Medicine, Duke University Medical Center, Durham, NC, USA [87] Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI, USA [88] Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA [89] Department of Chronic Disease Epidemiology, Yale School of Public Health, New Haven, CT, USA [90] Department of Obstetrics and Gynecology, University of Toronto, ON, Canada [91] Program in Epidemiology, Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA [92] Department of Epidemiology, University of Washington, Seattle, WA, USA [93] Duke Cancer Institute, Durham, NC, USA [94] Praxis für Humangenetik, Wiesbaden, Germany [95] Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center and Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, TN, USA [96] Cancer Epidemiology Program, University of Hawaii Cancer Center, Hawaii, USA [97] Department of Gynaecological Oncology, Glasgow Royal Infirmary, Glasgow, UK [98] University Malaya Cancer Research Institute, Faculty of Medicine, University Malaya Medical Centre, University Malaya, Kuala Lumpur, Malaysia [99] Department of Health Sciences Research, Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN, USA [100] Department of Obstetrics and Gynecology, University of Ulm, Ulm, Germany [101] Department of Obstetrics, Gynecology and Oncology, IInd Faculty of

Medicine, Warsaw Medical University and Brodnowski Hospital, Warsaw, Poland [102]
Department of Obstetrics and Gynecology, Duke University Medical Center, Durham, NC, USA

Acknowledgements

We thank all the individuals who took part in this study and all the researchers, clinicians and technical and administrative staff who made possible the many studies contributing to this work (a full list is provided in the **Supplementary Note**). 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 supported in part by the Genetic Associations and Mechanisms in Oncology (GAME-ON) and a National Cancer Institute Cancer Post-GWAS Initiative (U19-CA148112). Details of the funding of individual investigators and studies are provided in the **Supplementary Note**. This study made use of data generated by the Wellcome Trust Case Control consortium; funding for the project was provided by the Wellcome Trust under award 076113. A full list of the investigators who contributed to the generation of the data is available from the website (see URLs). The results published here are based in part on data generated by The Cancer Genome Atlas Pilot Project established by the National Cancer Institute and National Human Genome Research Institute; information about The Cancer Genome Atlas (TCGA) and the investigators and institutions who constitute the TCGA research network can be found on the website (see URLs).

References

1. Ferlay J, et al. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. Int. J. Cancer. 2010; 127:2893–917. [PubMed: 21351269]
2. Hall JM, et al. Linkage of early-onset familial breast cancer to chromosome 17q21. Science. 1990; 250:1684–9. [PubMed: 2270482]
3. Walsh T, et al. Mutations in 12 genes for inherited ovarian, fallopian tube, and peritoneal carcinoma identified by massively parallel sequencing. Proc. Natl. Acad. Sci. U. S. A. 2011; 108:18032–7. [PubMed: 22006311]
4. Lynch HT, et al. Hereditary nonpolyposis colorectal cancer (Lynch syndromes I and II). II. Biomarker studies. Cancer. 1985; 56:939–51. [PubMed: 4016686]
5. Lynch HT, Conway T, Lynch J. Hereditary ovarian cancer. Pedigree studies, Part II. Cancer Genet. Cytogenet. 1991; 53:161–83. [PubMed: 2065292]
6. Boyd J, Rubin SC. Hereditary ovarian cancer: molecular genetics and clinical implications. Gynecol. Oncol. 1997; 64:196–206. [PubMed: 9038264]
7. Narod SA, et al. Hereditary and familial ovarian cancer in southern Ontario. Cancer. 1994; 74:2341–6. [PubMed: 7922985]
8. Risch HA, et al. Population BRCA1 and BRCA2 mutation frequencies and cancer penetrances: a kin-cohort study in Ontario, Canada. J. Natl. Cancer Inst. 2006; 98:1694–706. [PubMed: 17148771]
9. 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]
10. Goode EL, et al. A genome-wide association study identifies susceptibility loci for ovarian cancer at 2q31 and 8q24. Nat. Genet. 2010; 42:874–9. [PubMed: 20852632]
11. Bolton KL, et al. Common variants at 19p13 are associated with susceptibility to ovarian cancer. Nat. Genet. 2010; 42:880–4. [PubMed: 20852633]
12. Pharoah PD, et al. GWAS meta-analysis and replication identifies three new susceptibility loci for ovarian cancer. Nat. Genet. 2013; 45:362–70. 370e1–2. [PubMed: 23535730]
13. Bojesen SE, et al. Multiple independent variants at the TERT locus are associated with telomere length and risks of breast and ovarian cancer. Nat. Genet. 2013; 45:371–84. 384e1–2. [PubMed: 23535731]
14. Permuth-Wey J, et al. Identification and molecular characterization of a new ovarian cancer susceptibility locus at 17q21.31. Nat. Commun. 2013; 4:1627. [PubMed: 23535648]

15. Couch FJ, et al. Genome-wide association study in BRCA1 mutation carriers identifies novel loci associated with breast and ovarian cancer risk. PLoS Genet. 2013; 9:e1003212. [PubMed: 23544013]
16. Chen K, et al. Genome-wide association study identifies new susceptibility loci for epithelial ovarian cancer in Han Chinese women. Nat. Commun. 2014; 5:4682. [PubMed: 25134534]
17. Kuchenbaecker KB, et al. Identification of six new susceptibility loci for invasive epithelial ovarian cancer. Nat. Genet. 2015; 47:164–71. [PubMed: 25581431]
18. Vaughan S, et al. Rethinking ovarian cancer: recommendations for improving outcomes. Nat. Rev. Cancer. 2011; 11:719–25. [PubMed: 21941283]
19. Risch HA, Marrett LD, Jain M, Howe GR. Differences in risk factors for epithelial ovarian cancer by histologic type. Results of a case-control study. Am. J. Epidemiol. 1996; 144:363–72. [PubMed: 8712193]
20. Pearce CL, et al. Association between endometriosis and risk of histological subtypes of ovarian cancer: a pooled analysis of case-control studies. Lancet Oncol. 2012; 13:385–94. [PubMed: 22361336]
21. Faber MT, et al. Cigarette smoking and risk of ovarian cancer: a pooled analysis of 21 case-control studies. Cancer Causes Control. 2013; 24:989–1004. [PubMed: 23456270]
22. Alsop K, et al. BRCA mutation frequency and patterns of treatment response in BRCA mutation-positive women with ovarian cancer: a report from the Australian Ovarian Cancer Study Group. J. Clin. Oncol. 2012; 30:2654–63. [PubMed: 22711857]
23. Kobel M, et al. Ovarian carcinoma subtypes are different diseases: implications for biomarker studies. PLoS Med. 2008; 5:e232. [PubMed: 19053170]
24. Gilks CB, et al. Tumor cell type can be reproducibly diagnosed and is of independent prognostic significance in patients with maximally debulked ovarian carcinoma. Hum. Pathol. 2008; 39:1239–51. [PubMed: 18602670]
25. Tavassoulu, FA.; Devilee, P. World Health Organization classification of tumors. Pathology and genetics of tumors of the breast and female genital organs. IARC Press; Lyon: 2003.
26. Seidman JD, Kurman RJ, Ronnett BM. Primary and metastatic mucinous adenocarcinomas in the ovaries: incidence in routine practice with a new approach to improve intraoperative diagnosis. Am. J. Surg. Pathol. 2003; 27:985–93. [PubMed: 12826891]
27. Ronnett BM, et al. Mucinous borderline ovarian tumors: points of general agreement and persistent controversies regarding nomenclature, diagnostic criteria, and behavior. Hum. Pathol. 2004; 35:949–60. [PubMed: 15297962]
28. Ronnett BM, et al. Disseminated peritoneal adenomucinosis and peritoneal mucinous carcinomatosis. A clinicopathologic analysis of 109 cases with emphasis on distinguishing pathologic features, site of origin, prognosis, and relationship to "pseudomyxoma peritonei". Am. J. Surg. Pathol. 1995; 19:1390–408. [PubMed: 7503361]
29. Lee KR, Young RH. The distinction between primary and metastatic mucinous carcinomas of the ovary: gross and histologic findings in 50 cases. Am. J. Surg. Pathol. 2003; 27:281–92. [PubMed: 12604884]
30. Yemelyanova AV, Vang R, Judson K, Wu LS, Ronnett BM. Distinction of primary and metastatic mucinous tumors involving the ovary: analysis of size and laterality data by primary site with reevaluation of an algorithm for tumor classification. Am. J. Surg. Pathol. 2008; 32:128–38. [PubMed: 18162780]
31. Hart WR. Mucinous tumors of the ovary: a review. Int. J. Gynecol. Pathol. 2005; 24:4–25. [PubMed: 15626914]
32. Cuatrecasas M, Villanueva A, Matias-Guiu X, Prat J. K-ras mutations in mucinous ovarian tumors: a clinicopathologic and molecular study of 95 cases. Cancer. 1997; 79:1581–6. [PubMed: 9118042]
33. Pieretti M, et al. Heterogeneity of ovarian cancer: relationships among histological group, stage of disease, tumor markers, patient characteristics, and survival. Cancer Invest. 2002; 20:11–23. [PubMed: 11852993]
34. Ichikawa Y, et al. Mutation of K-ras protooncogene is associated with histological subtypes in human mucinous ovarian tumors. Cancer Res. 1994; 54:33–5. [PubMed: 8261457]

35. Feeley KM, Wells M. Precursor lesions of ovarian epithelial malignancy. Histopathology. 2001; 38:87–95. [PubMed: 11207821]
36. Heinzelmann-Schwarz VA, et al. A distinct molecular profile associated with mucinous epithelial ovarian cancer. Br. J. Cancer. 2006; 94:904–13. [PubMed: 16508639]
37. Marquez RT, et al. Patterns of gene expression in different histotypes of epithelial ovarian cancer correlate with those in normal fallopian tube, endometrium, and colon. Clin. Cancer Res. 2005; 11:6116–26. [PubMed: 16144910]
38. Vogelstein B, et al. Genetic alterations during colorectal-tumor development. N. Engl. J. Med. 1988; 319:525–32. [PubMed: 2841597]
39. Shen H, et al. Epigenetic analysis leads to identification of HNF1B as a subtype-specific susceptibility gene for ovarian cancer. Nat. Commun. 2013; 4:1628. [PubMed: 23535649]
40. Kelemen LE, Kobel M. Mucinous carcinomas of the ovary and colorectum: different organ, same dilemma. Lancet Oncol. 2011; 12:1071–80. [PubMed: 21616717]
41. Chen JM, Ferec C, Cooper DN. A systematic analysis of disease-associated variants in the 3' regulatory regions of human protein-coding genes II: the importance of mRNA secondary structure in assessing the functionality of 3' UTR variants. Hum. Genet. 2006; 120:301–33. [PubMed: 16807757]
42. Integrated genomic analyses of ovarian carcinoma. Nature. 2011; 474:609–15. [PubMed: 21720365]
43. Comprehensive molecular characterization of human colon and rectal cancer. Nature. 2012; 487:330–7. [PubMed: 22810696]
44. Brown CD, Mangravite LM, Engelhardt BE. Integrative modeling of eQTLs and cis-regulatory elements suggests mechanisms underlying cell type specificity of eQTLs. PLoS Genet. 2013; 9:e1003649. [PubMed: 23935528]
45. Li H, Huang CJ, Choo KB. Expression of homeobox genes in cervical cancer. Gynecol. Oncol. 2002; 84:216–21. [PubMed: 11812077]
46. Tabuse M, et al. Functional analysis of HOXD9 in human gliomas and glioma cancer stem cells. Mol. Cancer. 2011; 10:60. [PubMed: 21600039]
47. Laury AR, et al. PAX8 reliably distinguishes ovarian serous tumors from malignant mesothelioma. Am. J. Surg. Pathol. 2010; 34:627–35. [PubMed: 20414098]
48. Laury AR, et al. A comprehensive analysis of PAX8 expression in human epithelial tumors. Am. J. Surg. Pathol. 2011; 35:816–26. [PubMed: 21552115]
49. Cheung HW, et al. Systematic investigation of genetic vulnerabilities across cancer cell lines reveals lineage-specific dependencies in ovarian cancer. Proc. Natl. Acad. Sci. U. S. A. 2011; 108:12372–7. [PubMed: 21746896]
50. Muratovska A, Zhou C, He S, Goodyer P, Eccles MR. Paired-Box genes are frequently expressed in cancer and often required for cancer cell survival. Oncogene. 2003; 22:7989–97. [PubMed: 12970747]
51. Di Palma T, Lucci V, de Cristofaro T, Filippone MG, Zannini M. A role for PAX8 in the tumorigenic phenotype of ovarian cancer cells. BMC Cancer. 2014; 14:292. [PubMed: 24766781]
52. Duggal P, et al. Genome-wide association study of spontaneous resolution of hepatitis C virus infection: data from multiple cohorts. Ann. Intern. Med. 2013; 158:235–45. [PubMed: 23420232]
53. Ochi H, et al. IL-28B predicts response to chronic hepatitis C therapy--fine-mapping and replication study in Asian populations. J. Gen. Virol. 2011; 92:1071–81. [PubMed: 21228123]
54. Key FM, et al. Selection on a Variant Associated with Improved Viral Clearance Drives Local, Adaptive Pseudogenization of Interferon Lambda 4 (IFNL4). PLoS Genet. 2014; 10:e1004681. [PubMed: 25329461]
55. Urban TJ, et al. IL28B genotype is associated with differential expression of intrahepatic interferon-stimulated genes in patients with chronic hepatitis C. Hepatology. 2010; 52:1888–96. [PubMed: 20931559]
56. Pejovic T. Genetic changes in ovarian cancer. Ann. Med. 1995; 27:73–8. [PubMed: 7742004]

57. Houlston RS, et al. Meta-analysis of three genome-wide association studies identifies susceptibility loci for colorectal cancer at 1q41, 3q26.2, 12q13.13 and 20q13.33. Nat. Genet. 2010; 42:973–7. [PubMed: 20972440]
58. Michailidou K, et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. Nat. Genet. 2013; 45:353–61. 361e1–2. [PubMed: 23535729]
59. Eeles RA, et al. Identification of 23 new prostate cancer susceptibility loci using the iCOGS custom genotyping array. Nat. Genet. 2013; 45:385–91. 391e1–2. [PubMed: 23535732]
60. Hnisz D, et al. Super-enhancers in the control of cell identity and disease. Cell. 2013; 155:934–47. [PubMed: 24119843]
61. Permuth-Wey J, et al. LIN28B polymorphisms influence susceptibility to epithelial ovarian cancer. Cancer Res. 2011; 71:3896–903. [PubMed: 21482675]
62. Delaneau O, Marchini J, Zagury JF. A linear complexity phasing method for thousands of genomes. Nat. Methods. 2012; 9:179–81. [PubMed: 22138821]
63. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. 2009; 5:e1000529. [PubMed: 19543373]
64. 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; 76:168–77. [PubMed: 22256951]
65. Price AL, et al. Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. 2006; 38:904–9. [PubMed: 16862161]
66. Li Q, et al. Integrative eQTL-based analyses reveal the biology of breast cancer risk loci. Cell. 2013; 152:633–41. [PubMed: 23374354]
67. Li Q, et al. Expression QTL-based analyses reveal candidate causal genes and loci across five tumor types. Hum. Mol. Genet. 2014; 23:5294–302. [PubMed: 24907074]
68. Kunzmann R, Holzel F. Karyotype alterations in human ovarian carcinoma cells during long-term cultivation and nude mouse passage. Cancer Genet. Cytogenet. 1987; 28:201–12. [PubMed: 3476184]
69. Li NF, et al. A modified medium that significantly improves the growth of human normal ovarian surface epithelial (OSE) cells in vitro. Lab. Invest. 2004; 84:923–31. [PubMed: 15077121]
70. Pomerantz MM, et al. The 8q24 cancer risk variant rs6983267 shows long-range interaction with MYC in colorectal cancer. Nat. Genet. 2009; 41:882–4. [PubMed: 19561607]

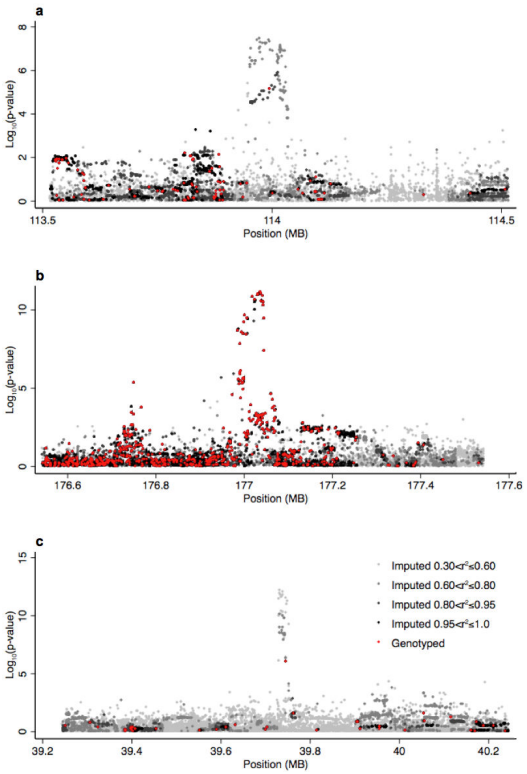

Figure 1.

Manhattan plots showing association between risk of MOC and the genotypes of SNPs in a 1Mb region of re-imputation surrounding the most significantly associated SNP at (a) 2q13 (top SNP: rs752590), (b) 2q31.1 (top SNP: rs711830) and (c) 19q13.2 (top SNP: rs688187). Sample size is 1,644 cases and 21,693 controls. Red dots indicate a genotyped SNP in COGS, progressively darker grey dots indicate SNPs with pre-phased imputation r^2 values between 0.30 and 0.60, 0.60 and 0.80 and 0.80 to 0.95, respectively, and black dots indicate SNPs with pre-phased imputation r^2 values between 0.95 and 1.0.

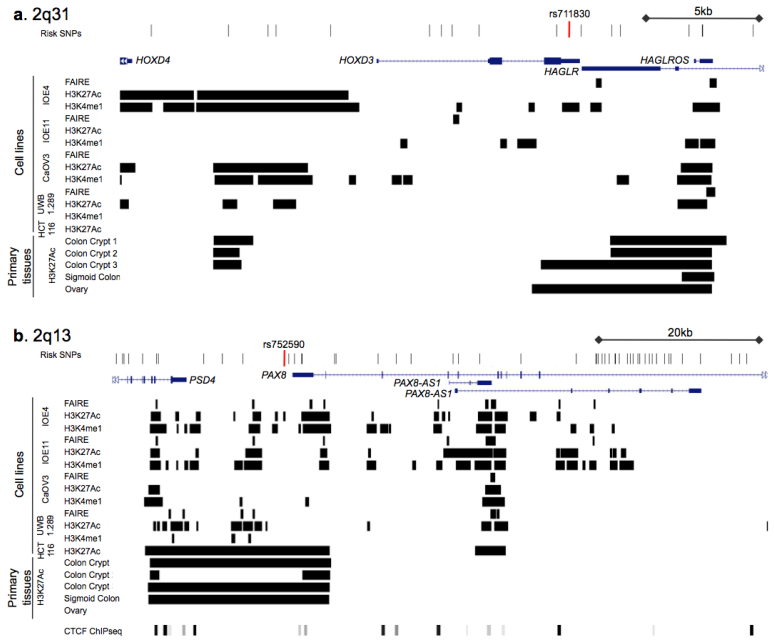

Figure 2.

The epigenetic landscape of MOC risk regions. All candidate causal SNPs are shown, the most significantly associated risk SNP is indicated in red. Data for normal immortalized ovarian epithelial (IOE) cells, serous ovarian cancer cells (CaOV3, UWB1.289, used in the absence of analogous data for mucinous ovarian cancer cells) colon cancer cells (HCT-116) and normal tissue from Hniez et al[60] (colonic crypts, sigmoid colon and ovary). At both (a) 2q31.1 and (b) 2q13 there is extensive overlap between regulatory biofeatures and risk SNPs. We also included collated ENCODE ChIPseq data for CTCF at 2q13, since PAX8 is rarely expressed in invasive mucinous ovarian cancer, SNPs that coincide with repressor marks could be the most relevant for this disease subtype.

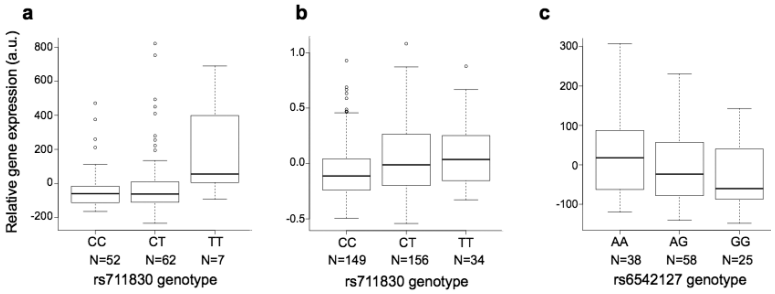

Figure 3.

Expression quantitative trait locus (eQTL) at MOC risk regions. Boxplots show the median (horizontal line), 1st to 3rd quartiles of expression (box) and 1.5 × the interquartile range (whiskers) in arbitrary units (a.u.). Tumor data are from The Cancer Genome Atlas for 339 high grade serous ovarian cancers (HGSOC) and 121 colorectal cancers (CRC). Significant associations were found between *HOXD9* gene expression at the 2q31.1 region and genotypes of the risk SNP rs711830 in (a) CRC ($P = 0.01$) and (b) HGSOC ($P = 4.95 \times 10^{-4}$). (c) A significant association was found between *PAX8* gene expression at the 2q13 region and genotypes of the risk SNP rs6542127 in CRC ($P = 0.03$). P-values are corrected for false-discovery-rate and considered to be a significant below 0.1.

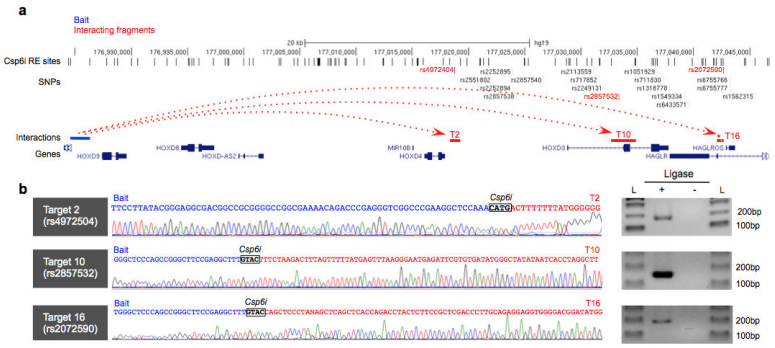

Figure 4.

Chromosome conformation capture (3C) at the 2q31.1 region containing *HOXD9* performed in the EFO27 mucinous ovarian cancer cell line. Csp6i restriction enzyme fragments containing the MOC risk SNPs rs4972504 (T2 fragment), rs2857532 (T10 fragment) and rs2072590 (T16 fragment) show evidence of interaction with the *HOXD9* promoter region, defined as 1.5 kb upstream of the transcription start site. (a) Schematic showing the locus, all SNPs with a 1:100 chance of being the causal variant, the locations of the interacting fragments (horizontal red bars), the risk SNPs located within each interacting region (vertical red bars) and the *HOXD9* promoter bait region (blue bar). (b) PCR confirmation of bait-target interactions (+ lanes) of the predicted size (L=ladder) and absence in controls (– lanes) without the addition of ligase. DNA sequencing chromatograms show interactions between the bait region (blue sequence) ligated to the target region (red sequence) and the intervening Csp6i restriction enzyme site (black box).

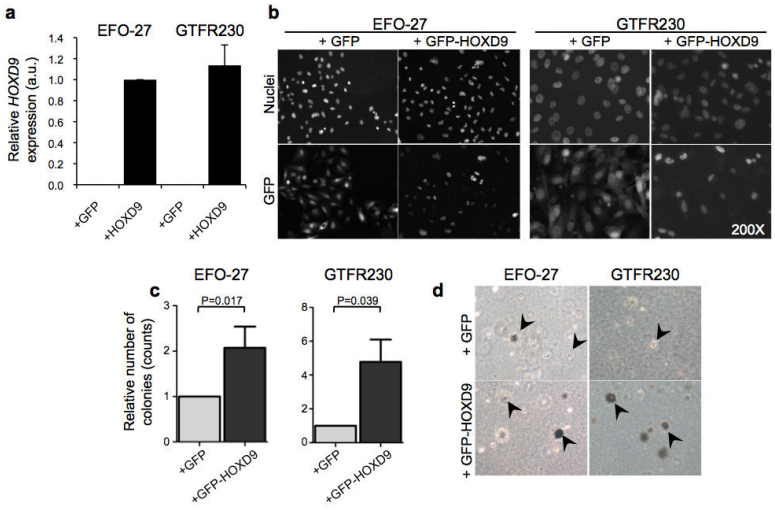

Figure 5.

Modeling the effects of HOXD9 overexpression in two *in vitro* models of MOC. Overexpression of *HOXD9* was confirmed by (a) qPCR in arbitrary units (a.u.), error bars are standard deviation (SD) and represent two independent experiments and (b) immunofluorescence microscopy (200 × magnification). Overexpression of HOXD9 is associated with increased anchorage independent growth. (c) Data shown are mean relative colony numbers from three independent experiments ± SD. Two-tailed paired t-test. (d) Phase-contrast images of colony growth. Examples of colonies are indicated with arrowheads.

Table 1

Summary of genotyping datasets used for imputation*, European samples

Study Set†	Controls, N	All Ovarian Cases, N	Mucinous Cases All (Invasive-only)	Genotyping Platform	Genotyping Center	Number SNPs passing QC
U.K. GWAS	6,118	1,785	206 (184)	Illumina 610K (cases) Illumina 1.2M and 550K (controls)	Illumina Corporation Sanger Centre	507,094 507,094
U.S. GWAS	1,867	1,813	100 (99)	Illumina 610K	Mayo Clinic Medical Genome Facility	556,480
NEC/BWH GWAS	142	132	0	Illumina 317K	National Cancer Institute	305,690
NCI-POL GWAS	555	221	16 (13)	Illumina 550K	National Cancer Institute	527,435
Mayo GWAS	441	467	36 (11)	HumanOmni 2.5 BeadChip	Mayo Clinic Medical Genome Facility	1,587,042
OCAC-COGS	21,693‡	11,620‡	1,286 (696)‡	Illumina custom iSelect ~211K	McGill University and Génome Québec Innovation Centre and Mayo Clinic Medical Genome Facility	199,570
Total	**30,816**	**16,038**	**1,644 (1,003)**			

*All datasets were used for imputation; however, association analysis was based on all mucinous ovarian cancer cases (N=1,644) and controls genotyped in OCAC-COGS (N=21,693)

†See Supplementary Table 1 for details of individual studies

‡Number of unique samples after exclusion of duplicates also genotyped in the other five GWAS

Table 2

Association testing in OCAC samples participating in COGS, 1,644 MOC cases and 21,693 controls

							Invasive+LMP MOC N=1,644			Invasive-only MOC N=1,003		
SNP	**Locus**	**Position**	**Nearest Gene**	**EAF***	**Imputation r^2 †**	**Ref/Alt Allele**	**OR**	**95% CI**	**P**	**OR**	**95% CI**	**P**
rs752590	2q13	113972945	*PAX8*	0.21	0.66	A/G	1.34	1.21–1.49	3.3×10^{-8}	1.31	1.17–1.46	2.9×10^{-6}
rs72831838 ‡	2q13	114016401	*PAX8*	0.14	0.65	C/T	1.38	1.23–1.56	1.2×10^{-7}	1.39	1.22–1.58	7.5×10^{-7}
rs6542125 (alternate) §	2q13	114018826	*PAX8*	0.15	0.64	G/A	1.38	1.23–1.56	9.0×10^{-8}	1.37	1.21–1.56	1.2×10^{-6}
rs6758928 (alternate) §	2q13	114019350	*PAX8*	0.15	0.64	G/A	1.38	1.23–1.55	9.4×10^{-8}	1.37	1.21–1.56	1.3×10^{-6}
rs711830 ¶	2q31.1	177037311	*HOXD3*	0.32	1	C/T	1.30	1.20–1.40	7.5×10^{-12}	1.26	1.15–1.39	7.7×10^{-7}
rs688187	19q13.2	39732752	*IFNL3*	0.32	0.55	G/A	0.67	0.60–0.75	6.8×10^{-13}	0.71	0.63–0.80	1.3×10^{-8}
rs35963157 //	19q13.2	39745695	*IFNL4*	0.34	0.55	G/GC	0.68	0.61–0.76	3.0×10^{-12}	0.72	0.64–0.81	2.9×10^{-8}

*Effect allele frequency

†Imputation r^2 was estimated without pre-phasing using the 1000 Genomes Project reference panel

‡Most significant SNP on 2q13 in pre-phased imputation results; ranked #19 without pre-phasing

§Alternate SNPs on 2q13 were correlated with rs72831838 (r^2=1) and selected for genotyping in lymphocyte DNA due to failed primer design for rs72831838; rs6542125 is ranked #13 and rs6758928 is ranked #14 without pre-phasing

¶Genotyped SNP

//Most significant SNP on 19q13.2 in pre-phased imputation results; ranked #8 without pre-phasing