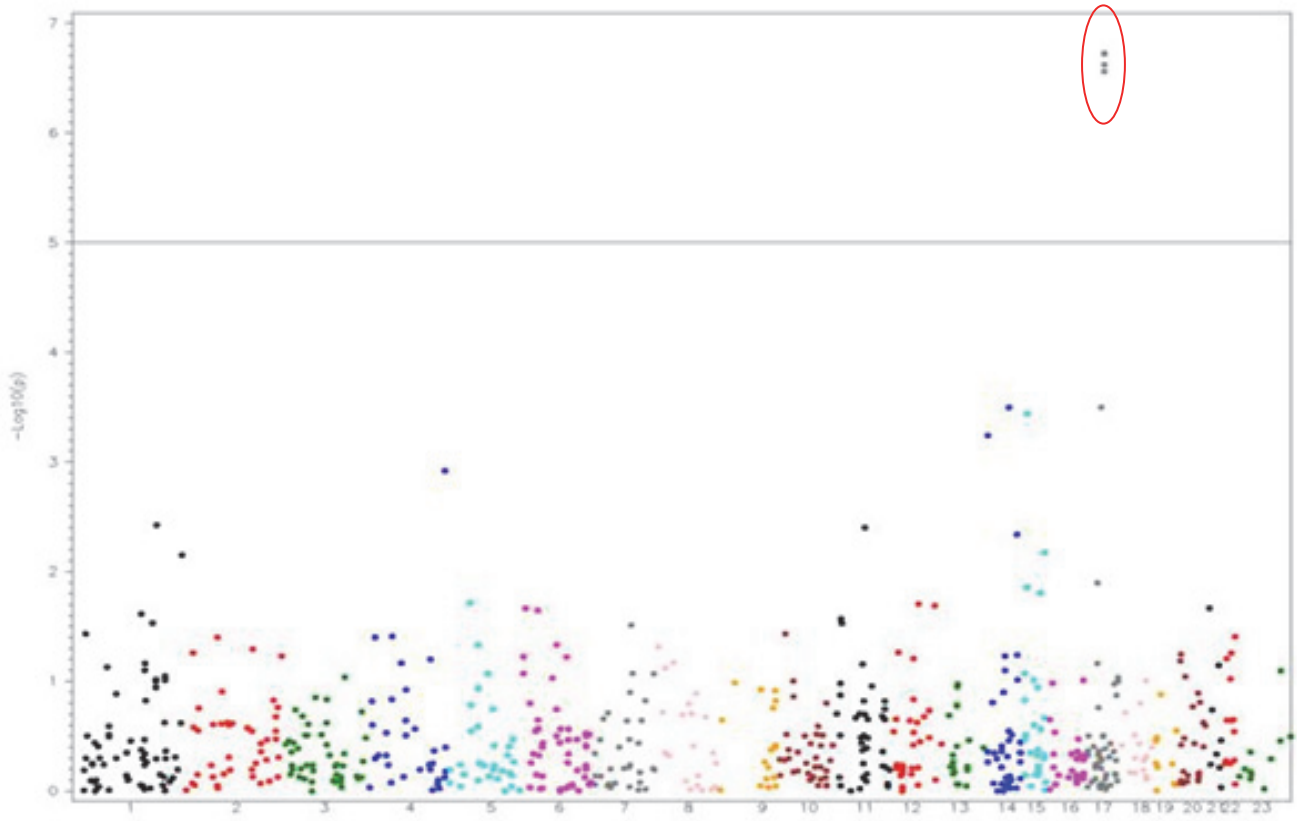
Supplementary Information

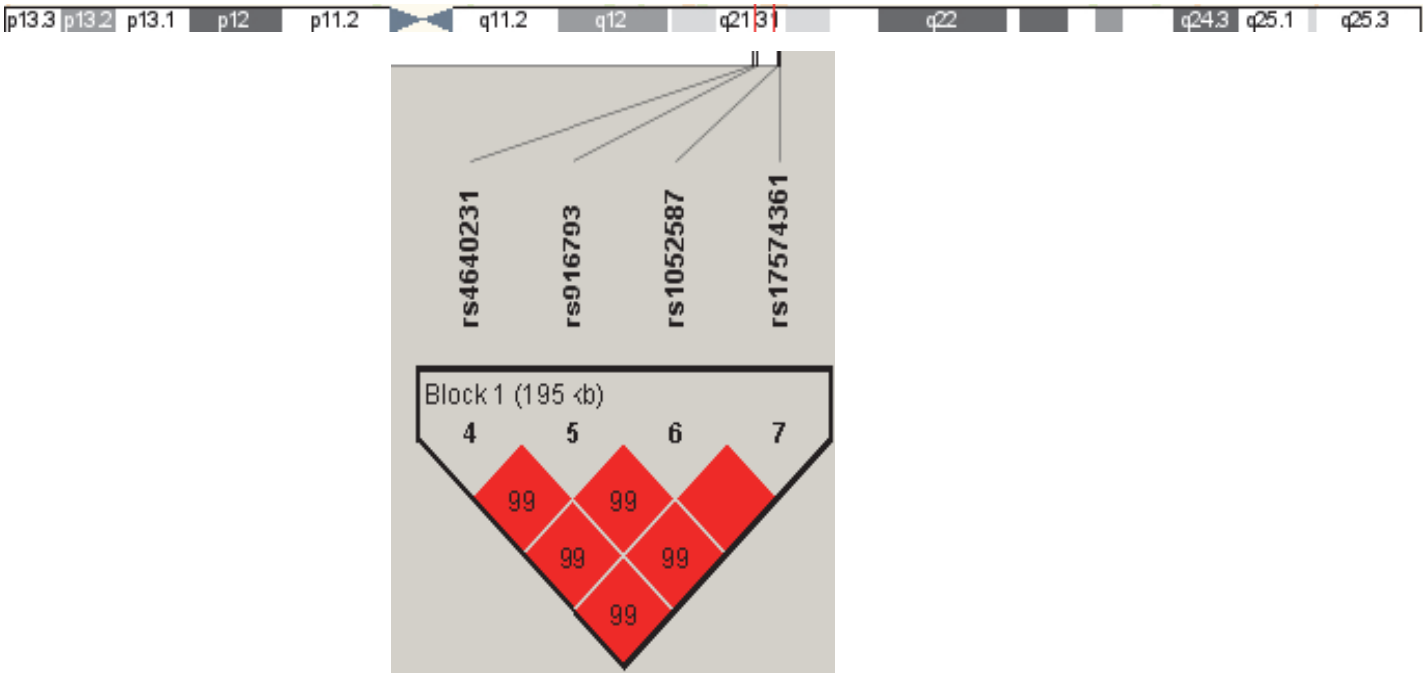
Identification and molecular characterization of a new ovarian cancer susceptibility locus at 17q21.31

Permuth Wey et al.

Supplementary Figure S1. Association results among invasive epithelial ovarian cancer cases.



Chr 17



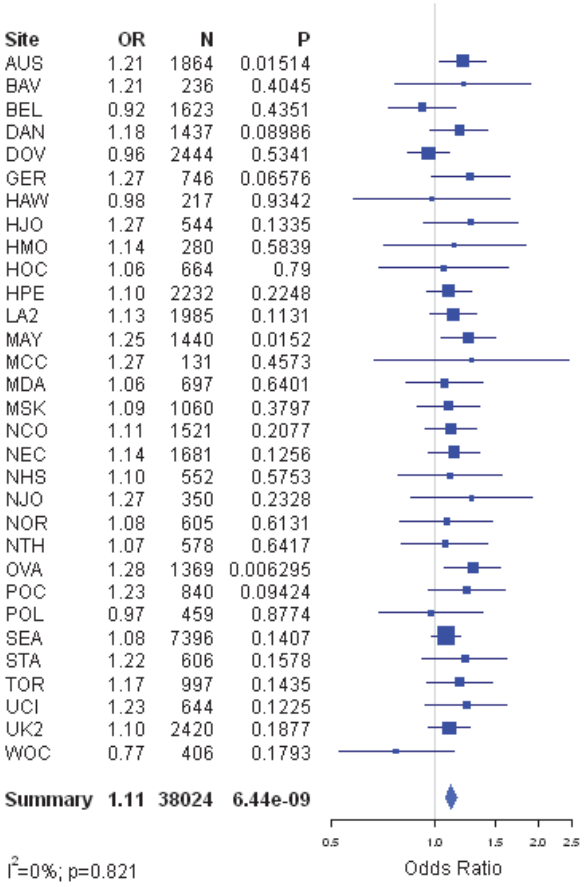
Supplementary Figure S1: Association results among invasive epithelial ovarian cancer cases.

Manhattan plot showing the –log10 p value for the 767 miRSNPs (or tagSNPs) passing genotyping quality control criteria against their chromosomal positions revealed four overlapping signals of genome-wide significance that mapped to chromosome 17, and to band 17q21.31 in particular. A linkage disequilibrium (LD) plot was generated in Haploview 4.1 using the solid spline LD method. Numbers represent r2 * 100, and are based on genome build 36.3; r2 = 0.99 or 1= red=completely correlated.

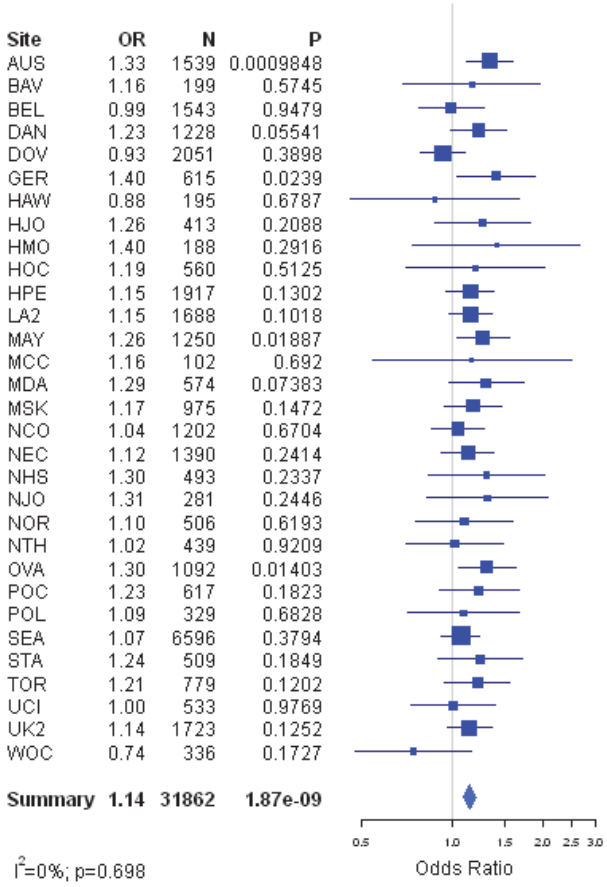
(Cases: n=14,533; controls: n=23,491)

Supplementary Figure S2. Study specific odds ratios for the association between rs1052587 genotype and epithelial ovarian cancer risk.

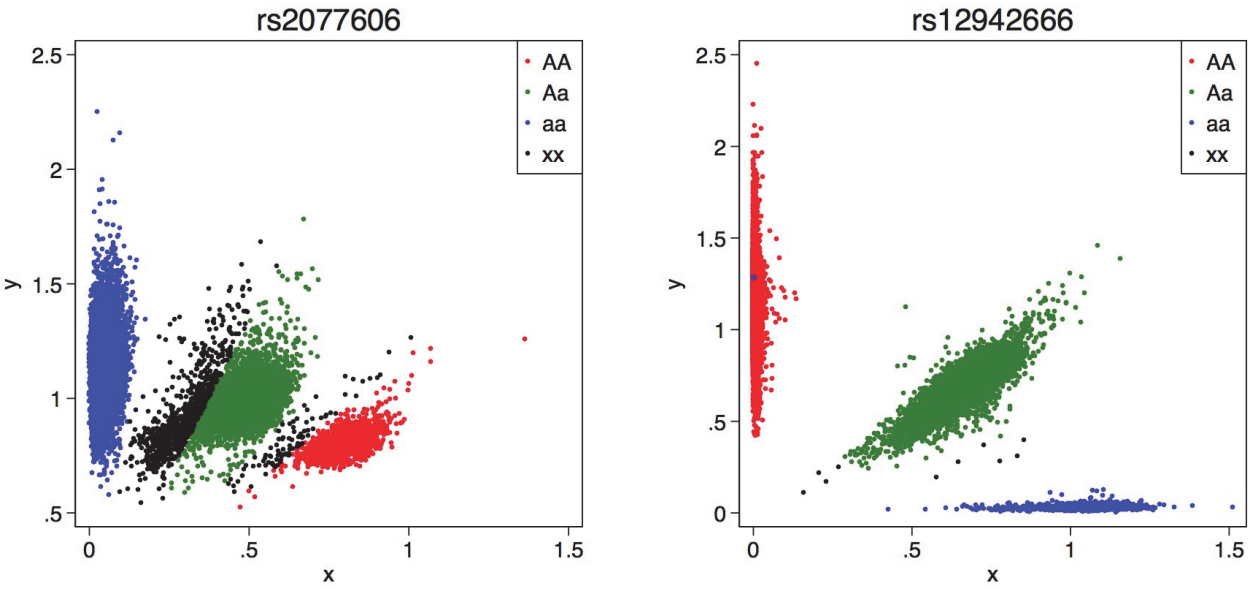
a. Susceptibility to invasive epithelial ovarian cancer.



b. Susceptibility to invasive serous carcinoma.



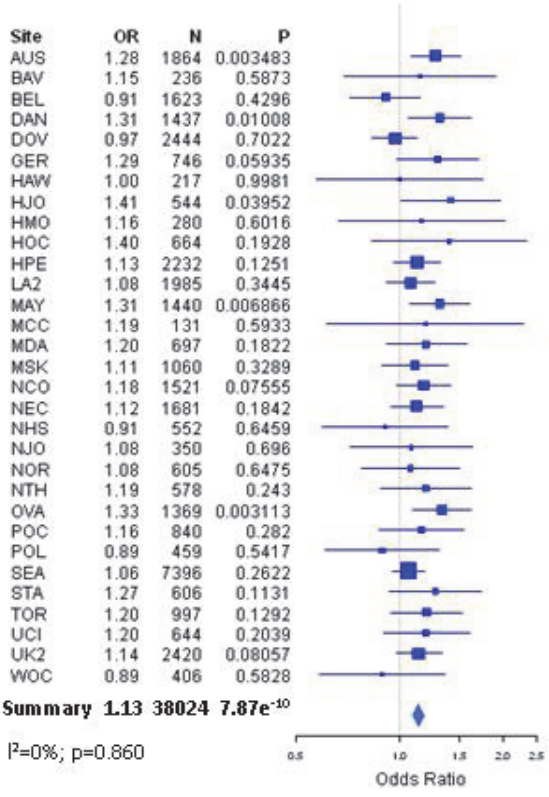
Supplementary Figure S3. Supplementary Figure S3: Signal intensity cluster plots for the most strongly associated non-miRSNPs at 17q21.31.



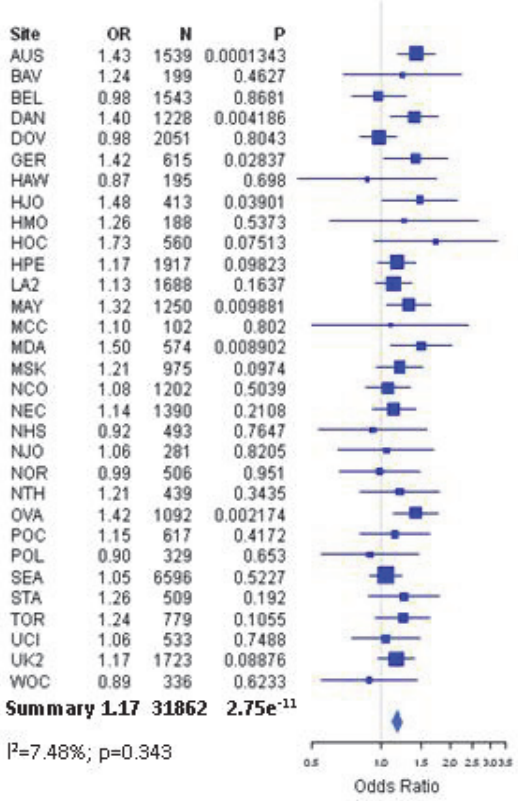
Genotype clustering was poor for rs2077606, but was good for its most strongly correlated SNP (r2=0.99), rs12942666. Results of association and molecular analyses are similar for both SNPs (Supplementary Tables 2, 3, and 4). Both SNPs passed quality control metrics.

Supplementary Figure S4. Study specific odds ratios for the association between rs12942666 and epithelial ovarian cancer susceptibility.

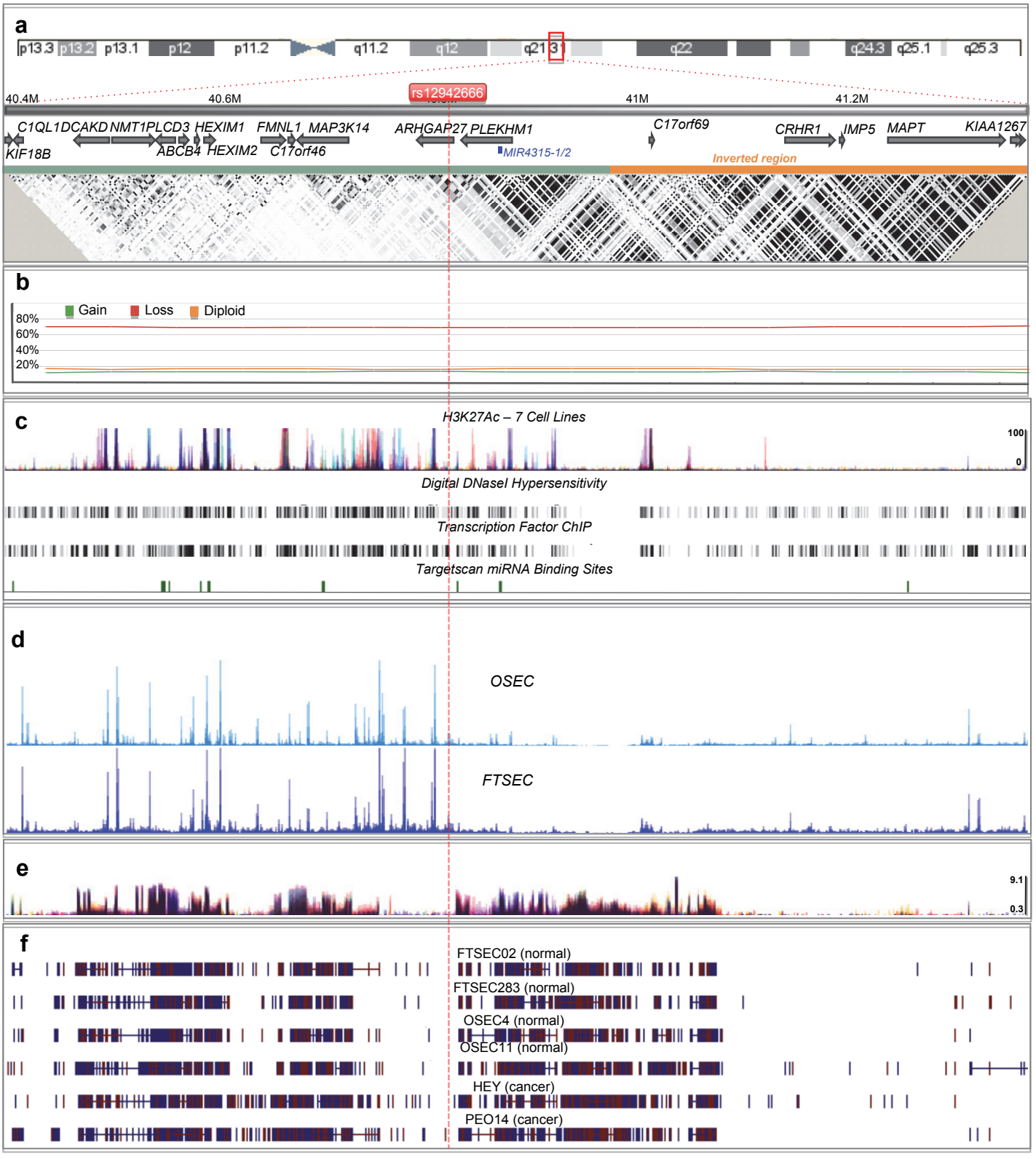
a **Susceptibility to invasive epithelial ovarian cancer.**



b **Susceptibility to invasive serous carcinoma.**



Supplementary Figure S5. Functional map of the 17q21.31 EOC susceptibility locus.



Supplementary Figure S5. Functional map of the 17q21.31 EOC susceptibility locus. **(a)** Genomic map of a one-megabase region centered around the most statistically significant SNP, rs12942666. Linkage disequilibrium structure of the 1Mb region from Haploview showing R2 associations between SNPs (in grey scale). The darker the color, the higher the R2 score. The location and approximate size of all known protein coding genes (grey) and non-coding RNA sequences (blue) are shown relative to the most significant SNP (red dashed line). **(b)** Analysis of somatic DNA copy number variation across the region generated by Affymetrix 6.0 SNP array analysis of 568 high-grade serous ovarian tumors analyzed by the cancer genome atlas (TCGA) project. Red = proportion of tumors showing loss; green = proportion of tumors showing gain/amplification; Yellow = proportion of tumors that are diploid (no loss or gain). **(c)** Analysis of ENCODE to evaluate putative regulatory DNA or non-coding RNA elements at susceptibility loci. Data were generated for non-ovarian cancer associated cell lines and show H3K27Ac regulatory marks, DNaseI Hypersensitivity elements, transcription factor binding site information from chromatin immunoprecipitation (ChIP) analysis, and TargetScan analysis of miRNA binding sites. **(d)** Formaldehyde assisted isolation of regulatory elements sequencing (FAIRE-seq) performed in normal ovarian surface epithelial cells (OSECs) and fallopian tube secretory epithelial cells (FTSECs). From genome wide FAIRE-seq, the profile of peaks of open chromatin (and putative regulatory sites) are illustrated for the region. **(e)** ENCODE RNA sequencing analysis (RNA-seq) of non-ovarian cancer associated cell lines shows the spectrum of coding transcripts across the region. **(f)** RNA-seq analysis of normal OSECs, FTSECs and epithelial ovarian cancer cell lines shows the spectrum of coding transcripts across the region in ovarian cancer associated tissues.

Supplementary Table S1: Description of individual OCAC studies and case-control sets[1].

Study Name	Study abbreviation	Study Location	Study Type	Total Number of Subjects[2]		% of subjects of European Ancestry	Number of invasive cases of European ancestry
				Cases	Controls		All histologies (serous)
Australian Ovarian Cancer Study/Australian Cancer Study (Ovarian Cancer)	AUS	Australia	Population based/ case-control	949[c]	1011	95.1	886 (561)
Bavarian Ovarian Cancer Cases and Controls	BAV	Southeast Germany	Population based/ case-control	98 [b]	143	100	93 (56)
Belgian Ovarian Cancer Study	BEL	Belgium, University Hospital Leuven	Hospital based/ Case-control	277[c]	1352	99.6	274 (194)
Diseases of the Ovary and their Evaluation	DOV	USA: 13 counties in western Washington state	Population based/ case-control	1350 [c]	1606	92.0	903 (525)
Oregon Ovarian Cancer Registry	ORE	Portland, Oregon	Case only	70 [b]	0	92.9	54 (39)
German Ovarian Cancer Study	GER	Germany: Baden-Württemberg and Rhineland-Palatinate	Population based/ case-control	216 [c]	413	99.5	189 (95)
Dr. Horst Schmidt Kliniken	HSK	Germany	Case only	155 [c]	0	98.7	144 (107)
Hawaii Ovarian Cancer Case-Control Study	HAW	USA: Hawaii	Population based/ case-control	338 [b]	601	25.4[c]	60 (38)
Hannover-Jena Ovarian Cancer Study	HJO	Germany	Hospital based/ Case-control	286 [c]	274	99.5	271 (140)
Hannover-Minsk Ovarian Cancer Study	HMO	Belarus	Hospital-based/Case-control	144 [c]	140	98.6	142 (50)
Helsinki Ovarian Cancer Study	HOC	Helsinki, Finland	Case-control	226 [c]	447	99.9	217 (113)
Hormones and Ovarian Cancer PrEdiction (HOPE) Study	HOP	USA: West Pennsylvania, Northeast Ohio, West New York	Population based/ Case-control	759 [c]	1501	97.0	654 (377)

Supplementary Table S1

Study Name	Study abbreviation	Study Location	Study Type	Total Number of Subjects[2]		% of subjects of European Ancestry	Number of invasive cases of European ancestry
				Cases	Controls		All histologies (serous)
Gilda Radner Familial Ovarian Cancer Registry	GRR	USA	Familial cancer/ Case only	115 [b]	0	97.4	112 (74)
Hospital-based Research Program at Aichi Cancer Center	JPN	Japan: Nagoya City	Case-control	76 [c]	81	0[c]	-
Women's Cancer Program at the Samuel Oschin Comprehensive Cancer Institute	LAX	USA: Southern California	Case only	330 [c]	0	84.2	278 (217)
Los Angeles County Case-control studies of Ovarian Cancer-1	USC	Los Angeles County	Population based/ Case-control	1260 [c]	1370	71.8	660 (424)
MALignant OVArian Cancer	MAL	Denmark	Population based/ Case-control	571 [b]	829	99.9	440 (272)
Danish Pelvic Mass Study	PVD	Denmark	Case only	172 [b]	0	98.2	169 (128)
Malaysia Ovarian Cancer Study	MAS	Malaysia	Hospital-based / Case-control	106 [c]	106	0[c]	-
Mayo Clinic Ovarian Cancer Case-Control Study	MAY	USA: North Central (MN, SD, ND, IL, IA, WI)	Clinic based/ Case-control	791 [c]	753	98.5	697 (507)
Melbourne Collaborative Cohort Study	MCC	Melbourne, Australia	Cohort/ Nested case-control	64 [c]	68	99.2	63 (34)
MD Anderson Ovarian Cancer Study	MDA	USA: Texas	Hospital based/ Case-control	323 [c]	385	98.5	313 (190)
Memorial Sloan-Kettering Cancer Center	MSK	USA: New York City	Hospital-based/Case-control	556 [b]	697	84.6	467 (382)
North Carolina Ovarian Cancer	NCO	USA: Central and eastern North	Population based/	1063 [b]	984	82.8	729 (410)

Supplementary Table S1

Study Name	Study abbreviation	Study Location	Study Type	Total Number of Subjects[2]		% of subjects of European Ancestry	Number of invasive cases of European ancestry
				Cases	Controls		All histologies (serous)
Study		Carolina (48 counties)	Case-control				
New England Case Control Study	NEC	USA: New Hampshire and Eastern Massachusetts	Population based/ Case-control	949 [b]	1049	95.8	672 (381)
Nurses' Health Study	NHS	USA	Cohort/ Nested case-control	147 [b]	429	99.3	127 (68)
New Jersey Ovarian Cancer Study	NJO	USA: New Jersey (six counties)	Population-based/Case-control	190 [c]	194	91.2	169 (100)
University of Bergen, Haukeland University Hospital, Norway	NOR	Norway	Case-control	248 [b]	371	99.5	234 (135)
Nijmegen Ovarian Cancer Study	NTH	Eastern part of the Netherlands	Case-control	265 [c]	323	98.6	255 (116)
Ovarian Cancer in Alberta and British Columbia	OVA	Alberta and British Columbia, Canada	Population-based/Case-control	855 [c]	810	91.5	621 (344)
Polish Ovarian Cancer Study	POC	Poland: Szczecin, Poznan, Opole, Rzeszów	Hospital-based/Case-control	423 [c]	417	100	423 (200)
Polish Ovarian Cancer Case Control Study	POL	Poland, Warszaw and Lodz	Population based/ Case-control	236 [b]	223	100	236 (106)
Study of Epidemiology and Risk Factors in Cancer Heredity	SEA	UK: East Anglia and West Midlands	Population based/ Case-control	1496 [c]	6067	99.1	1372 (572)
Family Registry for Ovarian Cancer and Genetic Epidemiology of Ovarian Cancer	STA	USA: Six counties in the San Francisco Bay area	Population based/ Case-control	293 [b]	404	88.4	257 (160)
Shanghai Women's Health	SWH	Shanghai, China	Cohort/Nested	135 [c]	891	0[c]	-

Supplementary Table S1

Study Name	Study abbreviation	Study Location	Study Type	Total Number of Subjects[2]		% of subjects of European Ancestry	Number of invasive cases of European ancestry
				Cases	Controls		All histologies (serous)
Study			case-control				
Toronto Ovarian Cancer Study	TOR	Canada: Province of Ontario	Population based	559 [c]	443	99.5	557 (339)
University of California Irvine Ovarian Study	UCI	USA: Southern California (Orange and San-Diego, Imperial Counties)	Population based/ Case-control	507 [b]	425	84.2	277 (166)
United Kingdom Ovarian Cancer Population Study	UKO	United Kingdom (England, Wales and Northern Ireland)	Population based/ Case-control	718 [c]	1123	98.1	702 (353)
Royal Marsden Hospital Ovarian Cancer Study	RMH	UK: London	Hospital based/ Case only	152 [b]	0	95.4	144 (49)
UK Familial Ovarian Cancer Registry	UKR	UK: National	Case only/ Familial Register	48 [b]	0	97.9	47 (23)
Southampton Ovarian Cancer Study	SOC	United Kingdom, Wessex region	Case only/ Hospital based	295 [c]	0	97.2	266 (102)
Scottish Randomised Trial in Ovarian Cancer	SRO	Coordinated through clinical trials unit, Glasgow UK from patients recruited world-wide	Case only from clinical trial	159 [b]	0	98.7	157 (92)
Warsaw Ovarian Cancer Study	WOC	Poland: Warsaw and central Poland	Hospital-based/Case-control	204 [b]	204	100	202 (132)
TOTAL				**18,174**	**26,134**	**89.8**	**14,533 (8,371)**

1 Studies combined into single case controls sets are indicated by contiguous shading (AOCS+ACS; DOV+ORE; HOP+GRR; GER+HSK; LAX+USC; MAL+PVD; UKO+RMH+UKR+SOC+SRO)

2 Totals represent the number of subjects passing genotyping quality control criteria.

a All or most of the cases were confirmed to have a histological subtype of epithelial ovarian cancer by a pathologist.

b Information about the diagnosis was primarily acquired from a cancer registry, pathology report, or the medical record.

c All or the majority of subjects from this site are of Asian ancestry: HAW (59.2%), JPN (100%), MAS (80.3%), SWH (100%)

Supplementary Table S2. Association results and functional annotation for SNPs highly correlated (R^2>0.90) with rs12942666 and associated with invasive serous epithelial ovarian cancer risk (P<10-9)

Chr 17 a coordinate	dbSNP name (if applicable)b	Imp.	R^2	Maj. allele	Min. allele	MAF	Imp. Accur.	OR	L95	U95	P-value	Putative Function(s) c	Location_1 d	Gene_1 e	Location_2 d	Gene_2 e	Type f **
43529293	rs2077606	No	0.99	G	A	0.19	1.00	1.15	1.12	1.19	3.91E-10	TFBS	intron||5'U/S	PLEKHM1			
43516402	rs17631303	No	0.99	A	G	0.19	1.00	1.15	1.12	1.19	4.67E-10	TFBS	intron	PLEKHM1			
43499839	rs12942666	No	1.00	A	G	0.19	1.00	1.15	1.11	1.19	1.04E-09	N/A	intron	ARHGAP27			
43504525	[b]	Yes		A	AC	0.16	0.84	1.19	1.15	1.23	8.05E-11						
43504524		Yes		C	CA	0.16	0.83	1.19	1.15	1.23	1.09E-10						
43551151	rs12950965 [b]	Yes	0.95	C	G	0.17	0.89	1.18	1.14	1.21	1.14E-10	N/A	intron	PLEKHM1	3'D/S	MIR4315 1/2	
43565599	rs62065444 [b]	Yes	0.96	T	C	0.18	0.94	1.17	1.13	1.21	1.16E-10	N/A	intron||5'U/S	PLEKHM1			
43550107	rs2090847 [b]	Yes	0.98	A	G	0.18	0.95	1.16	1.13	1.20	1.35E-10	N/A	intron	PLEKHM1			
43567337	rs1879586 [b]	Yes	0.96	C	G	0.17	0.92	1.17	1.13	1.21	1.54E-10	TFBS	intron	PLEKHM1			
43563894	rs62065442 [b]	Yes	0.98	T	C	0.19	0.97	1.16	1.12	1.20	1.58E-10	N/A	intron	PLEKHM1			
43551523	rs62065403	Yes	0.98	T	C	0.19	0.98	1.16	1.12	1.20	1.77E-10	N/A	intron	PLEKHM1	3'D/S	MIR4315 1/2	
43484876		Yes		C	CT	0.17	0.95	1.17	1.13	1.20	1.79E-10						
43555253	rs56192752	Yes	0.98	A	G	0.19	0.97	1.16	1.12	1.20	1.84E-10	N/A	intron	PLEKHM1	3'D/S	U7	
43556982	rs55643511	Yes	0.98	G	A	0.19	0.97	1.16	1.12	1.20	1.86E-10	N/A	intron	PLEKHM1	5'U/S	U7	
43557612	rs62065437	Yes	0.98	G	A	0.19	0.97	1.16	1.12	1.20	1.86E-10	N/A	intron	PLEKHM1	5'U/S	U7	
43552717	rs12452076	Yes	0.98	G	C	0.19	0.97	1.16	1.12	1.20	1.97E-10	splice||cons.||TFBS	NONCOD||COD	PLEKHM1	3'D/S	MIR4315 1/2	syn
43568927		Yes		T	TA	0.18	0.95	1.16	1.13	1.20	1.97E-10						
43565840	rs62065445	Yes	0.98	G	A	0.19	0.97	1.16	1.12	1.20	1.99E-10	N/A	intron||5'U/S	PLEKHM1			
43563093	rs62065441	Yes	0.97	A	G	0.18	0.95	1.16	1.13	1.20	2.04E-10	N/A	intron	PLEKHM1			
43551321	rs56015792	Yes	0.98	G	C	0.19	0.98	1.16	1.12	1.20	2.06E-10	N/A	intron	PLEKHM1	3'D/S	MIR4315 1/2	
43549526	rs17631676	Yes	0.98	A	G	0.19	0.98	1.16	1.12	1.20	2.07E-10	N/A	intron	PLEKHM1			
43551613	rs62065404	Yes	0.98	C	T	0.19	0.98	1.16	1.12	1.20	2.07E-10	N/A	intron	PLEKHM1	3'D/S	MIR4315 1/2	
43552537	rs71238846	Yes	0.98	G	A	0.19	0.98	1.16	1.12	1.20	2.08E-10	N/A	NONCOD||COD	PLEKHM1	3'D/S	MIR4315 1/2	syn
43549608		Yes	0.98	G	A	0.19	0.97	1.16	1.12	1.20	2.18E-10	N/A	intron	PLEKHM1			
43548481	rs55652155	Yes	0.98	A	G	0.19	0.98	1.16	1.12	1.20	2.19E-10	N/A	intron	PLEKHM1			
43558092	rs62065438	Yes	0.98	C	T	0.18	0.97	1.16	1.12	1.20	2.2E-10	N/A	intron	PLEKHM1	5'U/S	U7	
43556652	rs62065436	Yes	0.98	G	A	0.19	0.97	1.16	1.12	1.20	2.22E-10	N/A	intron	PLEKHM1	5'U/S	U7	
43521193	rs117793085	Yes	0.99	A	G	0.18	0.97	1.16	1.12	1.20	2.25E-10	N/A	intron	PLEKHM1			
43570893	rs1879583	Yes	0.96	C	T	0.18	0.93	1.16	1.13	1.20	2.28E-10	TFBS	intergen	PLEKHM1	N/A	LOC644354	
43551546	rs77132763	Yes	0.98	C	T	0.19	0.97	1.16	1.12	1.20	2.31E-10	N/A	intron	PLEKHM1	3'D/S	MIR4315 1/2	
43567175	rs62065446	Yes	0.98	T	G	0.18	0.97	1.16	1.12	1.20	2.33E-10	TFBS	intron||5'U/S	PLEKHM1	intergen||5'U/S	PLEKHM1	
43548424	rs55671319	Yes	0.98	A	G	0.19	0.97	1.16	1.12	1.20	2.43E-10	N/A	intron	PLEKHM1			
43568928	rs111392251	Yes	0.97	T	A	0.18	0.95	1.16	1.12	1.20	2.53E-10	N/A	5'U/S	PLEKHM1			
43569770	rs62065448	Yes	0.97	C	T	0.19	0.96	1.16	1.12	1.20	2.63E-10	TFBS	intergen||5'U/S	PLEKHM1	N/A	LOC100132027	
43509316	rs62064653	Yes	0.99	C	T	0.19	1.00	1.16	1.12	1.19	2.66E-10	CPG	N/A	LOC201175	intron||5'U/S	ARHGAP27	
43509310	rs62064652	Yes	0.98	A	G	0.19	0.97	1.16	1.12	1.20	2.82E-10	CPG	N/A	LOC201175	intron||5'U/S	ARHGAP27	

43569083	rs62065447	Yes	0.97	C	T	0.19	0.96	1.16	1.12	1.20	2.87E-10	TFBS	U/S||5'U/S	PLEKHM1	N/A	LOC644354	
43552921	rs12452273	Yes	0.97	C	T	0.19	0.94	1.16	1.12	1.20	2.97E-10	cons.	NONCOD||COD	PLEKHM1	5'U/S	MIR4315 1/2	syn
43564222	rs9899111	Yes	0.97	T	G	0.19	0.96	1.16	1.12	1.20	2.97E-10	N/A	intron	PLEKHM1	N/A	LOC440456	
43569001		Yes		CA	C	0.19	0.96	1.16	1.12	1.20	3.14E-10						
43534353	rs2960000	Yes	0.98	T	C	0.18	0.96	1.16	1.12	1.20	3.26E-10	N/A	intron	PLEKHM1	intron	AC091132.1	
43493504	rs71373560	Yes	0.97	C	T	0.18	0.96	1.16	1.12	1.20	3.39E-10	N/A	intron	ARHGAP27			
43516739	rs35591873	Yes	0.99	G	A	0.18	1.00	1.16	1.12	1.19	3.57E-10	TFBS	intron	PLEKHM1			
43527323	rs112538459	Yes	1.00	C	T	0.19	1.00	1.15	1.12	1.19	3.58E-10	N/A	intron	PLEKHM1			
43551083	rs55746869	Yes	0.97	C	T	0.18	0.96	1.16	1.12	1.20	3.58E-10	N/A	intron	PLEKHM1	3'D/S	MIR4315 1/2	
43522361	rs62065378	Yes	1.00	T	C	0.19	1.00	1.15	1.12	1.19	3.76E-10	N/A	intron	PLEKHM1			
43566931	rs3972613 [b]	Yes	0.93	A	G	0.16	0.86	1.18	1.14	1.22	3.78E-10	TFBS	intron||5'U/S	PLEKHM1	N/A	LOC100132027	
43525346		Yes		TAG	T	0.19	1.00	1.15	1.12	1.19	3.8E-10						
43525022	rs71373572	Yes	1.00	G	A	0.19	1.00	1.15	1.12	1.19	3.82E-10	TFBS	intron	PLEKHM1			
43570680	rs1879582	Yes	0.95	C	T	0.17	0.91	1.17	1.13	1.21	3.83E-10	TFBS	intergen	PLEKHM1	N/A	LOC644354	
43514954	rs62064655	Yes	1.00	G	A	0.19	1.00	1.15	1.12	1.19	3.85E-10	TFBS||miRNA	NONCOD||intron||3'u	PLEKHM1			
43523385		Yes		T	TC	0.19	1.00	1.15	1.12	1.19	3.89E-10						
43515927	rs35354512	Yes	1.00	C	T	0.19	1.00	1.15	1.12	1.19	3.98E-10	TFBS	intron	PLEKHM1			
43536408	rs55703888	Yes	1.00	C	T	0.19	1.00	1.15	1.12	1.19	4.05E-10	N/A	intron	PLEKHM1	intron	AC091132.1	
43536743	rs56005713	Yes	1.00	C	T	0.19	1.00	1.15	1.12	1.19	4.05E-10	N/A	intron	PLEKHM1	intron	AC091132.1	
43519564	rs34887474	Yes	1.00	C	A	0.19	1.00	1.15	1.12	1.19	4.08E-10	N/A	intron	PLEKHM1			
43513441	rs11012	Yes	1.00	C	T	0.19	1.00	1.15	1.12	1.19	4.21E-10	GWAS||TFBS||miRNA	NONCOD||COD||3'utr	PLEKHM1	5'U/S	ARHGAP27	nonsyn
43515846	rs34363898	Yes	1.00	C	T	0.19	1.00	1.15	1.12	1.19	4.24E-10	TFBS	intron	PLEKHM1			
43541627	rs2139890	Yes	1.00	C	A	0.19	1.00	1.15	1.12	1.19	4.24E-10	N/A	intron	PLEKHM1	3'D/S	AC091132.1	
43512206	rs12946900	Yes	1.00	A	G	0.19	1.00	1.15	1.12	1.19	4.25E-10	TFBS	3'D/S	PLEKHM1	intergen||5'U/S	ARHGAP27	
43512439	rs55790407	Yes	1.00	G	C	0.19	1.00	1.15	1.12	1.19	4.25E-10	TFBS	3'D/S	PLEKHM1	5'U/S	ARHGAP27	
43513551	rs9730	Yes	1.00	C	G	0.19	1.00	1.15	1.12	1.19	4.25E-10	TFBS||miRNA	NONCOD||intron||3'u	PLEKHM1	5'U/S	ARHGAP27	
43508223	rs12939187	Yes	1.00	G	T	0.19	1.00	1.15	1.12	1.19	4.28E-10	Regulate||CPG	N/A	LOC201175	intron||5'U/S	ARHGAP27	
43508616	rs34104358	Yes	1.00	G	T	0.19	1.00	1.15	1.12	1.19	4.29E-10	CPG	N/A	LOC201175	intron||5'U/S	ARHGAP27	
43539035	rs62065392	Yes	0.99	G	A	0.19	1.00	1.15	1.12	1.19	4.32E-10	N/A	intron	PLEKHM1	intron	AC091132.1	
43524526	rs113575082	Yes	0.99	C	T	0.19	1.00	1.15	1.12	1.19	4.33E-10	N/A	intron	PLEKHM1			
43507403	rs7220206	Yes	0.99	G	A	0.19	1.00	1.15	1.12	1.19	4.37E-10	Regulate||CPG	N/A	LOC201175	intron||NONCOD|	ARHGAP27	
43511435	rs34465449	Yes	1.00	C	A	0.19	1.00	1.15	1.12	1.19	4.37E-10	N/A	3'D/S	PLEKHM1	intergen||intron||	ARHGAP27	
43538991	rs62065390	Yes	0.99	G	A	0.19	1.00	1.15	1.12	1.19	4.37E-10	N/A	intron	PLEKHM1	intron	AC091132.1	
43487574	rs35327136	Yes	0.98	C	A	0.18	0.99	1.16	1.12	1.19	4.39E-10	N/A	intron	ARHGAP27			
43538993	rs62065391	Yes	0.99	G	C	0.19	0.99	1.16	1.12	1.19	4.48E-10	N/A	intron	PLEKHM1	N/A	LOC440456	
43544379	rs55663797	Yes	1.00	G	A	0.19	1.00	1.15	1.12	1.19	4.5E-10	N/A	intron	PLEKHM1			
43503000	rs62064651	Yes	0.99	G	A	0.19	1.00	1.15	1.12	1.19	4.51E-10	TFBS	N/A	LOC201175	5'utr||intron||5'U,	ARHGAP27	
43517252	rs35489312	Yes	1.00	T	C	0.19	1.00	1.15	1.12	1.19	4.56E-10	N/A	intron	PLEKHM1			
43515885	rs36114997	Yes	0.99	A	G	0.19	1.00	1.15	1.12	1.19	4.62E-10	TFBS	intron	PLEKHM1			
43507649	rs7222444	Yes	0.99	A	G	0.18	0.99	1.16	1.12	1.19	4.65E-10	CPG||splice||cons.||T	N/A	LOC201175	5'utr||NONCOD||i	ARHGAP27	
43503294	rs7209501	Yes	0.99	A	C	0.19	1.00	1.15	1.12	1.19	4.67E-10	TFBS	N/A	LOC201175	U/S||intron||5'U/	ARHGAP27	
43539968	rs4325608	Yes	0.99	G	A	0.18	0.99	1.16	1.12	1.19	4.68E-10	N/A	intron	PLEKHM1	N/A	LOC440456	

43508303	rs34018943	Yes	1.00	A	G	0.19	1.00	1.15	1.12	1.19	4.7E-10	cons.||CPG	N/A	LOC201175	intron||5'U/S	ARHGAP27	
43541656	rs41353445	Yes	0.99	C	T	0.19	0.99	1.15	1.12	1.19	4.83E-10	N/A	intron	PLEKHM1			
43541656	rs41353445	Yes	0.99	C	T	0.19	0.99	1.15	1.12	1.19	4.83E-10	N/A	N/A	LOC440456	3'D/S	AC091132.1	
43521161	rs62065376	Yes	0.99	C	G	0.19	1.00	1.15	1.12	1.19	4.89E-10	N/A	intron	PLEKHM1			
43545893	rs1879581	Yes	0.99	T	C	0.19	1.00	1.15	1.12	1.19	4.91E-10	splice	intron||NONCOD||CO	PLEKHM1			syn
43538807	rs62065389	Yes	0.99	T	G	0.19	0.99	1.15	1.12	1.19	5.02E-10	N/A	intron	PLEKHM1	intron	AC091132.1	
43527025	rs62065380	Yes	0.99	C	T	0.19	1.00	1.15	1.12	1.19	5.04E-10	N/A	intron	PLEKHM1			
43556807	rs55925547	Yes	0.97	T	C	0.19	0.97	1.16	1.12	1.19	5.1E-10	N/A	intron	PLEKHM1	5'U/S	U7	
43525365	rs62065379	Yes	0.99	C	T	0.19	1.00	1.15	1.12	1.19	5.21E-10	N/A	intron	PLEKHM1			
43538523	rs111423688	Yes	0.99	G	A	0.19	0.99	1.15	1.12	1.19	5.24E-10	N/A	intron	PLEKHM1	intron	AC091132.1	
43500477	rs62064645	Yes	0.99	G	A	0.19	1.00	1.15	1.12	1.19	5.31E-10	N/A	intron	ARHGAP27			
43513896	rs62064654	Yes	0.99	C	T	0.19	1.00	1.15	1.12	1.19	5.34E-10	TFBS||miRNA	NONCOD||intron||3'u	PLEKHM1			
43568280		Yes		G	GGGA	0.19	0.94	1.16	1.12	1.19	5.39E-10						
43534694	rs62065385	Yes	0.99	C	T	0.19	1.00	1.15	1.12	1.19	5.46E-10	N/A	intron	PLEKHM1	intron	AC091132.1	
43520272	rs62065374	Yes	0.99	G	A	0.19	1.00	1.15	1.12	1.19	5.5E-10	N/A	intron	PLEKHM1	N/A	LOC440456	
43569245	rs113322852	Yes	0.97	T	A	0.19	0.95	1.16	1.12	1.19	5.54E-10	N/A	intergen||5'U/S	PLEKHM1			
43484903	rs1808189	Yes	0.90	T	C	0.17	0.92	1.16	1.13	1.20	5.62E-10	N/A	intron||5'U/S	ARHGAP27			
43510187	rs55642947	Yes	0.99	G	C	0.19	1.00	1.15	1.12	1.19	5.67E-10	splice||CPG	N/A	LOC201175	5'utr||NONCOD||!	ARHGAP27	
43517054	rs62064657	Yes	0.99	G	A	0.19	1.00	1.15	1.12	1.19	5.72E-10	N/A	intron	PLEKHM1			
43574935	rs62066460	Yes	0.82	C	T	0.21	0.89	1.16	1.12	1.19	5.93E-10	TFBS	intergen	PLEKHM1	N/A	LOC644354	
43490853	rs62064643	Yes	0.99	G	A	0.18	0.99	1.15	1.12	1.19	5.95E-10	N/A	intron	ARHGAP27			
43503284	rs76344126	Yes	0.99	G	A	0.19	1.00	1.15	1.12	1.19	5.99E-10	N/A	U/S||intron||5'U/S	ARHGAP27			
43546057	rs62065399	Yes	0.99	G	T	0.18	1.00	1.15	1.12	1.19	6.01E-10	N/A	intron	PLEKHM1			
43512318	rs56168933	Yes	0.99	G	A	0.18	0.99	1.15	1.12	1.19	6.14E-10	TFBS	3'D/S	PLEKHM1	5'U/S	ARHGAP27	
43539437	rs62065393	Yes	0.97	C	G	0.17	0.93	1.16	1.13	1.20	6.45E-10	N/A	intron	PLEKHM1	intron	AC091132.1	
43488382	rs8071011	Yes	0.98	C	T	0.19	0.99	1.15	1.12	1.19	6.86E-10	N/A	intron||5'U/S	ARHGAP27			
43509778	rs56020833	Yes	0.97	C	T	0.19	0.99	1.15	1.11	1.19	6.88E-10	CPG	N/A	LOC201175	intron||5'U/S	ARHGAP27	
43552812	rs2521859	Yes	0.93	A	G	0.16	0.88	1.17	1.13	1.21	7.32E-10	cons.||Regulate||TFBS	NONCOD||COD	PLEKHM1	5'U/S	MIR4315 1/2	syn
43556862	rs56159231	Yes	0.97	C	T	0.19	0.95	1.16	1.12	1.19	7.4E-10	N/A	intron	PLEKHM1	5'U/S	U7	
43540449	rs111435106	Yes	0.96	G	C	0.16	0.90	1.17	1.13	1.21	7.54E-10	N/A	intron	PLEKHM1	intron	AC091132.1	
43563305	rs112995489	Yes	0.93	T	G	0.16	0.87	1.17	1.14	1.21	7.67E-10	N/A	intron	PLEKHM1			
43495852	rs34286926	Yes	0.99	A	G	0.19	1.00	1.15	1.11	1.19	7.92E-10	N/A	intron	ARHGAP27			
43574882	rs62065459	Yes	0.83	T	G	0.20	0.90	1.16	1.12	1.19	8.03E-10	TFBS	intergen	PLEKHM1	N/A	LOC644354	
43492357	rs12940792	Yes	0.98	C	T	0.19	1.00	1.15	1.11	1.19	8.14E-10	N/A	intron	ARHGAP27			
43493101	rs12947718	Yes	0.99	G	A	0.19	1.00	1.15	1.11	1.19	8.2E-10	N/A	intron	ARHGAP27			
43544206	rs2959992	Yes	0.94	G	C	0.16	0.89	1.17	1.13	1.21	8.2E-10	N/A	intron	PLEKHM1	N/A	LOC440456	
43563304	rs112275793	Yes	0.93	A	C	0.16	0.88	1.17	1.13	1.21	8.38E-10	N/A	intron	PLEKHM1			
43493560	rs34915103	Yes	0.99	C	G	0.19	1.00	1.15	1.11	1.19	8.4E-10	N/A	intron	ARHGAP27			
43493835	rs12942951	Yes	0.98	T	C	0.18	0.98	1.15	1.12	1.19	8.57E-10	N/A	intron	ARHGAP27			
43491003	rs34063617	Yes	0.98	G	A	0.19	1.00	1.15	1.11	1.19	8.59E-10	N/A	intron	ARHGAP27			
43488792	rs62064641	Yes	0.98	T	C	0.19	0.99	1.15	1.11	1.19	8.78E-10	N/A	intron||5'U/S	ARHGAP27			
43495216	rs35884427	Yes	0.99	G	A	0.19	1.00	1.15	1.11	1.19	8.93E-10	N/A	intron	ARHGAP27			

43500621	rs62064647	Yes	0.99	G	A	0.19	1.00	1.15	1.11	1.19	8.97E-10	N/A	intron	ARHGAP27		
43487424	rs35389313	Yes	0.98	T	C	0.19	0.99	1.15	1.11	1.19	9.07E-10	N/A	intron	ARHGAP27		
43487217	rs36078910	Yes	0.98	G	A	0.19	0.99	1.15	1.11	1.19	9.28E-10	N/A	intron	ARHGAP27		
43548321	rs55930887	Yes	0.85	T	C	0.16	0.79	1.18	1.14	1.22	9.46E-10	N/A	intron	PLEKHM1		
43496465	rs56378631	Yes	0.98	T	C	0.19	1.00	1.15	1.11	1.19	9.82E-10	N/A	intron	ARHGAP27		
43497210	rs12952764	Yes	0.98	G	A	0.19	1.00	1.15	1.11	1.19	1.01E-09	N/A	intron	ARHGAP27		
43476807	rs56220387	Yes	0.97	A	G	0.18	0.98	1.15	1.12	1.19	1.03E-09	N/A	intron	ARHGAP27		
43484496	rs62064603	Yes	0.97	C	T	0.19	0.98	1.15	1.11	1.19	1.04E-09	N/A	intron||5'U/S	ARHGAP27		
43502012	rs55914643	Yes	0.98	T	C	0.18	0.99	1.15	1.11	1.19	1.05E-09	TFBS	intron	ARHGAP27		
44236321		Yes		C	G	0.12	0.58	1.25	1.21	1.28	1.05E-09					
43499328	rs35519908	Yes	0.98	T	C	0.19	1.00	1.15	1.11	1.19	1.1E-09	N/A	intron	ARHGAP27		
43501442	rs7222389	Yes	0.98	T	C	0.19	1.00	1.15	1.11	1.19	1.12E-09	N/A	intron	ARHGAP27		
43501940	rs56212100	Yes	0.98	G	A	0.19	1.00	1.15	1.11	1.19	1.12E-09	N/A	intron	ARHGAP27		
43500587	rs62064646	Yes	0.97	G	A	0.19	1.00	1.15	1.11	1.19	1.15E-09	N/A	intron	ARHGAP27		
43501591	rs55648326	Yes	0.98	T	C	0.19	1.00	1.15	1.11	1.19	1.15E-09	N/A	intron	ARHGAP27		
43572896	rs5026246	Yes	0.84	T	G	0.19	0.86	1.16	1.12	1.20	1.19E-09	TFBS	intergen	PLEKHM1	N/A	LOC644354
43502111	rs62064649	Yes	0.98	A	G	0.19	1.00	1.15	1.11	1.19	1.2E-09	TFBS	intron	ARHGAP27		
43502241	rs35626715	Yes	0.98	T	C	0.19	1.00	1.15	1.11	1.19	1.2E-09	cons.||TFBS	intron	ARHGAP27		
43473307	rs62064597	Yes	0.96	C	G	0.18	0.96	1.16	1.12	1.19	1.22E-09	CPG	3'D/S||intron	ARHGAP27		
43485551	rs62064637	Yes	0.97	T	G	0.19	0.98	1.15	1.11	1.19	1.27E-09	N/A	intron	ARHGAP27		
43557848		Yes			T	0.18	0.92	1.16	1.12	1.20	1.28E-09					
43495235	rs12952504	Yes	0.98	C	G	0.19	0.99	1.15	1.11	1.19	1.32E-09	N/A	intron	ARHGAP27		
43563349	rs111960572	Yes	0.94	G	A	0.17	0.87	1.17	1.13	1.21	1.34E-09	N/A	intron	PLEKHM1		
43553496		Yes		AAT	A	0.19	0.91	1.15	1.12	1.19	1.4E-09					
43572419	rs56328224	Yes	0.83	C	T	0.21	0.91	1.15	1.11	1.19	1.45E-09	TFBS	intergen	PLEKHM1	N/A	LOC644354
43483551	rs56236914	Yes	0.97	C	T	0.18	0.98	1.15	1.11	1.19	1.48E-09	N/A	intron||5'U/S	ARHGAP27		
43570407	rs62065450	Yes	0.96	C	T	0.18	0.89	1.16	1.12	1.20	1.54E-09	TFBS	intergen	PLEKHM1	N/A	LOC644354
43540472	rs111876069	Yes	0.96	A	T	0.17	0.89	1.17	1.13	1.20	1.62E-09	N/A	intron	PLEKHM1	intron	AC091132.1
43475929	rs2028078	Yes	0.96	G	A	0.18	0.98	1.15	1.11	1.19	1.66E-09	N/A	intron	ARHGAP27		
43551428		Yes		AT	A	0.19	0.90	1.16	1.12	1.19	1.66E-09					
43574972	rs2959948	Yes	0.86	T	G	0.16	0.78	1.18	1.14	1.22	1.68E-09	TFBS	intergen	PLEKHM1	N/A	LOC644354
43480701	rs62064600	Yes	0.97	G	A	0.18	0.98	1.15	1.11	1.19	1.72E-09	N/A	3'D/S||intron	ARHGAP27		
43484598	rs73984391	Yes	0.96	G	T	0.19	0.98	1.15	1.11	1.19	1.74E-09	N/A	intron||5'U/S	ARHGAP27		
43479748	rs55793500	Yes	0.96	T	C	0.18	0.98	1.15	1.11	1.19	1.76E-09	N/A	3'D/S||intron	ARHGAP27		
43522775		Yes		TTGTC	T	0.18	0.95	1.16	1.12	1.19	1.79E-09					
43685269		Yes		G	A	0.09	0.51	1.29	1.25	1.33	1.86E-09					
43915497		Yes		C	T	0.29	0.98	1.13	1.09	1.17	1.86E-09					
44313269		Yes		T	C	0.13	0.63	1.22	1.18	1.26	1.95E-09					
43474668	rs62064598	Yes	0.96	C	G	0.18	0.98	1.15	1.11	1.19	2.01E-09	N/A	3'D/S||intron	ARHGAP27		
43573231	rs2903705	Yes	0.82	C	G	0.20	0.89	1.15	1.12	1.19	2.01E-09	N/A	intergen	PLEKHM1	N/A	LOC644354
43573419	rs62065453	Yes	0.83	C	T	0.21	0.91	1.15	1.11	1.19	2.04E-09	TFBS	intergen	PLEKHM1	N/A	LOC644354
43463493	rs79724577	Yes	0.96	A	C	0.18	0.98	1.15	1.11	1.19	2.11E-09	N/A	intergen	MAP3K14		

43573649	rs62065455	Yes	0.83	C	T	0.20	0.89	1.15	1.11	1.19	2.15E-09	N/A	intergen	PLEKHM1		
43539723	rs2684526	Yes	0.92	C	T	0.15	0.86	1.17	1.13	1.21	2.37E-09	N/A	intron	PLEKHM1	intron	AC091132.1
43543075		Yes		C	CA	0.19	0.84	1.16	1.12	1.20	2.51E-09					
43495420	rs12936645	Yes	0.88	A	T	0.18	0.81	1.17	1.13	1.20	2.61E-09	N/A	intron	ARHGAP27		
43498181	rs34792542	Yes	0.95	A	G	0.17	0.94	1.16	1.12	1.20	2.63E-09	N/A	intron	ARHGAP27		
43493834		Yes		T	TC	0.18	0.93	1.16	1.12	1.19	2.77E-09					
43498180	rs71363531	Yes	0.95	C	T	0.17	0.93	1.16	1.12	1.20	2.83E-09	N/A	intron	ARHGAP27		
43496828	rs12946723	Yes	0.97	C	T	0.19	0.97	1.15	1.11	1.18	3.14E-09	N/A	intron	ARHGAP27		
44672951		Yes		G	A	0.14	0.68	1.20	1.16	1.24	3.25E-09					
44313610		Yes		A	T	0.15	0.68	1.20	1.16	1.24	3.33E-09					
44346721		Yes		A	T	0.16	0.73	1.19	1.15	1.22	3.43E-09					
43471489	rs4763	Yes	0.95	G	A	0.18	0.96	1.15	1.11	1.19	3.53E-09	cons.||miRNA	3'D/S||NONCOD||3'ut	ARHGAP27		
43682377		Yes		A	G	0.18	0.77	1.17	1.13	1.21	3.59E-09					
43682393		Yes		A	G	0.18	0.77	1.17	1.13	1.21	3.7E-09					
44229415		Yes		T	C	0.15	0.70	1.19	1.15	1.23	4.59E-09					
43824382		Yes		T	G	0.21	0.94	1.14	1.11	1.18	5.07E-09					
43460374		Yes		A	G	0.15	0.87	1.17	1.13	1.21	5.09E-09					
43546442		Yes		A	G	0.06	0.48	1.38	1.34	1.42	5.23E-09					
43460181		Yes		G	A	0.15	0.88	1.17	1.13	1.21	5.63E-09					
43665612		Yes			A	0.20	0.81	1.16	1.12	1.19	5.7E-09					
44218138		Yes		C	T	0.15	0.71	1.19	1.15	1.23	5.94E-09					
44348326		Yes		G	A	0.17	0.78	1.17	1.13	1.21	6.2E-09					
44243407		Yes		G	A	0.15	0.68	1.19	1.16	1.23	6.25E-09					
43794209		Yes		G	C	0.21	0.95	1.14	1.10	1.18	6.51E-09					
44313627		Yes		G	A	0.14	0.65	1.20	1.17	1.24	6.76E-09					
44234064		Yes		G	A	0.15	0.71	1.19	1.15	1.22	6.97E-09					
43778315		Yes		CA	C	0.22	0.98	1.13	1.10	1.17	7.25E-09					
44312530		Yes		G	A	0.14	0.68	1.20	1.16	1.23	7.56E-09					
44324572		Yes		A	G	0.23	0.98	1.13	1.09	1.17	7.98E-09					
44231932		Yes		G	A	0.14	0.64	1.21	1.17	1.24	8.27E-09					
43794182		Yes		G	A	0.21	0.96	1.14	1.10	1.18	8.54E-09					
43741138		Yes		C	CA	0.62	0.66	0.88	0.84	0.91	8.72E-09					
43839951		Yes		A	T	0.22	1.00	1.13	1.09	1.17	9.04E-09					
43671471		Yes		T	C	0.23	0.90	1.14	1.10	1.18	9.21E-09					
43773523		Yes		ATT	A	0.21	0.95	1.14	1.10	1.18	9.7E-09					
43671737		Yes		A	C	0.23	0.89	1.14	1.10	1.18	9.9E-09					
43671739		Yes		A	G	0.23	0.89	1.14	1.10	1.18	9.9E-09					

Note:SNPs are sorted by imputation status (directly genotyped or imputed), and then by P-value. Blank cells signify data is unavailable

Abbreviations and Nomenclature: Imp.=Imputed; Imp. Accur.=Imputation accuracy was estimated using an r^2 quality metric and derived from IMPUTE2.2; OR=Odds ratio; CI=Confidence interval; MAF=minor allele frequency; TFBS=SNP resides in a putative transcription factor binding site; Cons.=Conserved=SNP is located in a regior that is evolutionarily conserved across species; NONCOD=non-coding; COD=coding; CpG=SNP is located in a CpG island; Splice=The SNP has the potential to cause splice site variants; U/S=upstream; D/S=downstream; miRNA=SNP is located in a putative miRNA binding site; Regulate=SNP is located in a potential region where a regulatory

element can bind; GWAS=the SNP has been associated with a GWAS of another phenotype;LoseFunc=SNP is non-synonymous and has the potential to affect protein functior via an amino acid substitution; utr=untranslated region; intergen=intergenic

a Human genome build 37

b This SNP was considered to be a likely candidate for being the causal variant when comparing the difference in log-likelihoods of logistic regression model between this SNP and rs12942666 (10<likelihood ratio<20).

c Three in silico SNP functional annotation tools (ANNOVAR, SNPInfo, and SNPnexus) were used to generate these predictions, and supplemented with information from JASPAR **(Supplementary Table S5).**

d. Location shows the relative position of the SNP to its adjacent gene. This information is directly retrieved from each in silico prediction tool/server. Since different transcripts from different resources (RefSeq, Ensemble, UCSC Genome Browser) are used to annotate a SNP's relative position, prediction results are kept intact and combined to provide as much information as possible.

e. The name of the gene containing the SNP or the nearest gene.

f. If the SNP is synonymous (syn)or non-synonymous (nonsyn), rs11012 is likely tolerated/benign per SIFT prediction. See ** below about another nonsyn SNP predicted to be benign. For additional details about how predictions are made, please refer to each individual tool:

ANNOVAR, http://www.openbioinformatics.org

SNPinfo, http://www.niehs.nih.gov/snpinfo

SNPnexus, http://www.snp-nexus.org

** There are two SNPs (rs62071393 and rs12949256) with putative functions that were not added to this table because they are less correlated with rs12942666 ($0.80<R^2<0.90$). rs62071393 resides in a putative TFBS near PLEKHM1, and rs12949256 is a non-synonymous SNP in or near ARHGAP27. Polyphen predicts that rs12949256 is benign

Supplementary Table S3: Summary of gene expression and somatic analyses of EOC tumor and normal tissue and cell lines for genes at the 17q21.31 locus

Gene	RNAseq a		Relative expression b			Gene expression in tumor v normal d			Expression QTL			Somatic mutation	
	OCPTs	Cancer	OCPTs	All EOC v OCPTs	All EOC v OCPTs	Serous EOC v OCPTs c	Cell lines	TCGA e	GEO f	rs12924666 g		rs2077606 h		
	Average normalized reads			fold change	P-value	P-value	direction	direction	direction	OCPT	LCL	TCGA tumors	OvCa	Any cancer
KIF18B	1.0	5.34	1.25	5.1	**$<2.2x10^{-16}$**	**$3.0x10^{-12}$**	**Up**	**Up**	**Up**	0.75	0.85		No	**Yes**
C1QL1	0.1	8.09	0.66	6.8	**$2.4x10^{-07}$**	**$2.1x10^{-04}$**	**Up**	Same	Up	0.20	0.25	**0.04**	No	No
DCAKD	2.5	19.07	8.73	1.9	**$3.1x10^{-06}$**	**0.004**	**Up**	Same	Inconsistent	0.93	0.29	0.35	No	No
NMT1	26.1	31.49	11.89	2.2	**$2.4x10^{-06}$**	**0.002**	**Up**	Inconsistent	Inconsistent	0.99	0.47	0.10	No	**Yes**
PLCD3	1.5	19.96	3.03	4.6	**$1.7x10^{-13}$**	**$1.9x10^{-09}$**	**Up**	N/A	Inconsistent	0.79	0.44	0.37	No	No
ACBD4	0.8	0.21	0.05	6.3	**$1.3x10^{-14}$**	**$1.7x10^{-10}$**	**Up**	Same	N/A	0.99	0.68	0.83	**Yes**	**Yes**
HEXIM1	4.9	27.81	5.50	3.9	**$5.6x10^{-12}$**	**$1.4x10^{-08}$**	**Up**	Inconsistent	Same	0.42	0.25	0.36	**Yes**	**Yes**
HEXIM2	0.6	1.13	0.05	30.7	**$<2.2x10^{-16}$**	**$1.7x10^{-14}$**	**Up**	N/A	**Up**	0.36	0.31	0.33	No	No
FMNL1	0.5	2.72	0.83	2.0	**$2.5x10^{-05}$**	**0.005**	**Up**	Up	**Down**	0.68	0.59	0.10	No	No
C17orf46	0.0	0.00	0.00	6.9	**0.048**	0.565	Same	N/A	**Up**	0.59	0.67	0.35	No	**Yes**
MAP3K14	0.5	1.06	0.08	13.3	**$<2.2x10^{-16}$**	**$7.9x10^{-13}$**	**Up**	Same	Same	0.30	0.26	0.63	No	**Yes**
ARHGAP27	0.0	0.88	0.00	851.0	**$<2.2x10^{-16}$**	**$9.2x10^{-16}$**	**Up**	N/A	**Up**	**0.04**	0.10	0.90	No	No
PLEKHM1	1.5	4.23	1.80	1.8	**0.004**	0.098	**Up**	Inconsistent	Inconsistent	0.77	0.89	**$1x10^{-4}$**	No	**Yes**
CRHR1	0.0	0.14	0.00	37.6	**$2.0x10^{-04}$**	**0.008**	**Up**	**Up**	Inconsistent	0.27	0.07	0.53	No	No
IMP5	0.0	0.01	0.00	45.5	**$7.9x10^{-05}$**	**0.009**	**Up**	N/A	**Up**	0.13	0.16	0.91	No	**Yes**
MAPT	0.1	1.18	0.05	8.1	**$4.4x10^{-9}$**	**$1.1X10^{-05}$**	**Up**	Same	**Up**	0.92	0.61	0.06	No	**Yes**

Abbreviations: OCPTs=ovarian cancer precursor tissues which include ovarian surface epithelial cells (OSECs) and fallopian tube secretory epithelial cells (FTSECs)
LCL=lymphoblastoid cell lines
Red bold indicates statistically significant
a average of two OSEC lines
b Gene expression relative to β-actin and GAPDH n 50 EOC cell lines and 73 normal OSEC/FTSEC lines
c enriched for serous (removed known clear cell and mucinous cell lines)
d Up or down regulation in tumor samples, red bold text indicates siginificant differences in one or more probes and remaining probes in same direction, black text indicates if there is a non sigificant trend in direction for all probes for gene
e TCGA data: gene expression in 568 serous EOC and 8 fallopian tube (normal) samples
f GEO series GSE18520 data: gene expression in 53 serous EOC and 10 ovary (normal) samples
g rs12924666 OCPTs AA=44, AG/GG=19, LCL AA=62, AG/GG=31, TCGA data is not available for this SNP
h rs2077606 OCPTs GG=43, AG/AA=18, LCL GG=63, AG/AA=31, TCGA GG=244 AG/AA=132

Supplementary Table S4: Summary of methylation data in high-grade serous (HGS) EOCs and normal tissues for genes at the 17q21.31 locus

Gene			High grade serous (n=106)		Normal (n=7)			Methylation (mean beta), HGS Cases (N=227)				Methylation (mean beta), HGS Cases (N=227)			
								rs12924666				rs2077606			
	Methylation Probe ID a	P-value	Mean	Std Dev	Mean	Std Dev	Adjusted p-value	Common allele homozygote	Heterozygote	Rare allele homozygote	P-value	Common allele homozygote	Heterozygote	Rare allele homozygote	P-value
								n=147	n=73	n=7		n=147	n=73	n=6	
KIF18B															
C1QL1	cg05294159	**0.003**	0.29	0.12	0.33	0.05	0.086	0.27	0.27	0.31	0.501	0.27	0.27	0.25	0.377
DCAKD	cg03287877	**0.001**	0.69	0.12	0.82	0.03	**0.004**	0.67	0.68	0.64	0.724	0.67	0.68	0.66	0.581
NMT1	cg00852245	0.056	0.16	0.11	0.15	0.04	0.699	0.14	0.15	0.15	0.310	0.14	0.15	0.14	0.373
PLCD3	cg24806326	**$7.5x10^{-5}$**	0.57	0.13	0.81	0.05	**$4.3x10^{-6}$**	0.53	0.56	0.62	0.086	0.53	0.56	0.62	0.107
ACBD4	cg04970158	**0.009**	0.72	0.15	0.87	0.04	**0.007**	0.72	0.73	0.72	0.864	0.72	0.73	0.73	0.799
HEXIM1	cg18053085	**0.006**	0.05	0.01	0.04	0.01	**0.01**	0.05	0.05	0.05	0.457	0.05	0.05	0.05	0.355
HEXIM2	cg03874026	**0.034**	0.88	0.05	0.91	0.02	0.061	0.87	0.87	0.88	0.513	0.87	0.87	0.87	0.624
FMNL1	cg09264140	**0.015**	0.79	0.07	0.87	0.01	**$4.2x10^{-5}$**	0.77	0.78	0.77	0.888	0.77	0.78	0.76	0.774
C17orf46	cg01357958	**0.036**	0.64	0.14	0.33	0.1	**$4.2x10^{-4}$**	0.62	0.6	0.72	0.391	0.62	0.6	0.72	0.338
MAP3K14	cg19411214	**0.014**	0.04	0.02	0.02	0.01	**0.015**	0.04	0.03	0.05	0.476	0.04	0.03	0.05	0.376
ARHGAP27	cg20615344	**0.012**	0.11	0.03	0.11	0.03	0.798	0.09	0.1	0.1	0.796	0.09	0.1	0.1	0.749
PLEKHM1	cg14154330	**0.041**	0.41	0.08	0.51	0.07	**$7.9x10^{-4}$**	0.4	0.37	0.4	**0.029**	0.4	0.37	0.39	**0.021**
CRHR1	cg15413793	0.317	0.88	0.03	0.87	0.02	0.242	0.86	0.88	0.8	**0.001**	0.86	0.88	0.78	**0.002**
IMP5	cg26656751	0.121	0.73	0.16	0.83	0.05	0.200	0.72	0.69	0.77	0.361	0.72	0.69	0.78	0.372
MAPT	cg26413900	**0.046**	0.34	0.12	0.32	0.07	0.766	0.32	0.3	0.36	0.302	0.32	0.3	0.35	0.254

a CpG with the most significant negative correlations between CpG and RNA expression, build 37; N=43; adjusted for age and histology

Supplementary Table S5: JASPAR transcription factor binding site prediction at SNPs correlated with rs12942666, sorted by relative score a

SNP	Position	TF site	Score	Relative Score a	Strand	Motif	Location	Gene
rs12946900	**chr17:43512206**	**SPIB**	**10.47**	**1.00001472**	1	**AGAGGAA**	**intergenic**	
rs2077606	**chr17:43529293**	**ZEB1**	**9.262**	**1.000002602**	**-1**	**CACCTG**	**intron**	
		Arnt	10.351	0.999997275	-1	CACGTG		
rs3972613	chr17:43566931	NFIC	9.697	0.999985774	-1	TTGGCA	intron	PLEKHM1
		GATA2	6.651	0.999974952	-1	GGATA		
		ETS1	7.633	0.992426668	-1	TTTCCT		
		FOXC1	7.249	0.987548393	1	GGTGAGTA		
rs35354512	chr17:43515927	FOXC1	7.209	0.985310893	-1	AGTAAGTA	intron	PLEKHM1
		Pax2	8.666	0.984299863	-1	AGTCACGT		
		Arnt::Ahr	9.197	0.982781685	-1	CGCGTG		
		USF1	10.923	0.980450844	-1	CACGTGA		
		USF1	10.923	0.980450844	1	CACGTGA		
rs62065448	chr17:43569770	NFE2L1::MafG	8.072	0.967939714	1	TATGAC	intergenic	
rs62066460	chr17:43574935	Pax2	8.232	0.967257225	1	GGTCACGC	intergenic	
rs62065450	chr17:43570407	MAX	11.926	0.964626738	1	TATCACGTGA	intergenic	
		BRCA1	7.554	0.961793168	1	CCAACCC		
rs62071393	chr17:43574229	FOXC1	6.764	0.96041871	-1	ACTGAGTA	intergenic	
	chr17:43512439	Pax2	7.832	0.951549724	1	GGTCACGT		
rs2959948	chr17:43574972	NFIC	8.238	0.951118625	-1	TTGGCG	intergenic	
		MAX	11.413	0.948953058	-1	AGTCACGTGA		
rs2521859	chr17:43552812	NFIC	7.969	0.942108849	1	TTGGCC	coding-synon L (TTG) --> L (CTG)	PLEKHM1
		Hltf	7.32	0.939204052	-1	TTCCATTTCC		
rs62064649	chr17:43502111	Gfi	10.29	0.931170979	1	AAAATCTCAG	intron	ARHGAP27
		NFATC2	9.49	0.931041561	-1	TTTTCCT		
rs56328224	chr17:43572419	YY1	7.027	0.930737645	-1	ACCATG	intergenic	
rs35591873	chr17:43516739	Hltf	7.075	0.930569352	-1	GAACTTATTG	intron	PLEKHM1
		Hltf	7.046	0.929547286	-1	CCCCTTTTCC		
		FOXC1	6.138	0.925401842	-1	ATAAAGTA		
rs55790407	chr17:43512439	MAX	10.61	0.924419012	1	GGTCACGTGC	intergenic	
rs9730	chr17:43513551	Zfx	13.281	0.92417516	1	AGGGCTTGGGCCTG	3' UTR	PLEKHM1
		Pax2	7.117	0.923472567	1	TATCACGT		
rs71373572	chr17:43525022	Myb	8.824	0.918440927	-1	CGCCGTTG	intron	PLEKHM1
		ETS1	5.977	0.916040416	1	GTTCCT		

rs62065447	chr17:43569083	HIF1A::ARNT	8.295	0.913365499	-1	GGGCGTGG	intergenic	
rs7209501	chr17:43503294	GATA3	6.215	0.90883429	-1	GGATAG	intergenic	
rs12452076	chr17:43552717	ETS1	5.777	0.906815023	-1	TTTCCC	coding-synon S (TCC) --> S (TCG)	PLEKHM1
rs35626715	chr17:43502241	SPI1	8.725	0.906109929	1	AGAAAGT	intron	ARHGAP27
rs62064655	chr17:43514954	Zfx	12.225	0.90578414	1	TGAGCCAGGGCCTG	3' UTR	PLEKHM1
rs62065453	chr17:43573419	FOXC1	5.782	0.905488096	-1	CGCCTGTA	intergenic	
		Nkx2-5	6.961	0.905137724	1	ATAAGTT		
		SPI1	8.637	0.903177735	-1	AGGAACT		
		HIF1A::ARNT	7.948	0.903013641	-1	GCGCGTGA		
rs62065459	chr17:43574882	IRF1	12.913	0.902652702	1	CATAGTGAAACC	intergenic	
rs55914643	chr17:43502012	-	-	-	-	-	intron	ARHGAP27
rs62064651	chr17:43503000	-	-	-	-	-	5' UTR	ARHGAP27
rs7222444	chr17:43507649	-	-	-	-	-	5' UTR	ARHGAP27
rs56168933	chr17:43512318	-	-	-	-	-	intergenic	
rs11012	chr17:43513191	-	-	-	-	-	3' UTR	PLEKHM1
rs62064654	chr17:43513896	-	-	-	-	-	3' UTR	PLEKHM1
rs34363898	chr17:43515846	-	-	-	-	-	intron	PLEKHM1
rs36114997	chr17:43515885	-	-	-	-	-	intron	PLEKHM1
rs17631303	chr17:43516402	-	-	-	-	-	intron	PLEKHM1
rs62065446	chr17:43567175	-	-	-	-	-	intron	PLEKHM1
rs1879586	chr17:43567337	-	-	-	-	-	intron	PLEKHM1
rs62065449	chr17:43569909	-	-	-	-	-	intergenic	
rs1879582	chr17:43570680	-	-	-	-	-	intergenic	
rs1879583	chr17:43570893	-	-	-	-	-	intergenic	
rs5026246	chr17:43572896	-	-	-	-	-	intergenic	

This table demonstrates JASPAR transcription factor binding site predictions (relative threshold=90%) for SNPs correlated with rs12942666 (R2>0.8).

Notes: JASPAR (http://jaspar.cgb.ki.se/) online output for predicted transcription factor binding site encompassing the SNP (TF Site), orientation of the predicted site (Strand), and DNA sequence identified (Motif), are noted alongside JASPAR scoring metrics.

Fields where no TF site was predicted are marked with (-). rs12946900 and rs2077606 scored highest in this analysis and are marked in **bold.**

JASPAR (http://jaspar.cgb.ki.se/) analysis was run using full set of matrix models in vertebrate database and Relative Score Profile

a The Relative Score is a measure of base-to-base identity of a query sequence (that contains the variant) compared to a reference/index sequence. It resembles the adjustable threshold parameter for influencing sensitivity of the query.

Supplementary Methods

Participating studies from the Ovarian Cancer Association Consortium (OCAC)

Forty-three individual OCAC studies contributed samples to the COGS project. Of these, nine studies were case only (GRR, HSK, LAX, ORE, PVD, SOC, RMH, SRO, UKR). The cases from these studies were pooled with case-control studies from the same geographic region and the two national Australian case-control studies were combined into a single study to create 34 case-control sets. Most studies frequency-matched cases and controls by age-group and race. Study characteristics are summarized in Supplementary Table S1.

Selection of candidate SNPs using *in silico* algorithms

After generating a list of 55 miRNAs reported to be deregulated in EOC tumors compared to normal tissue and a list of 665 genes that have been implicated in the pathogenesis of EOC (see Methods), we identified putative sites of miRNA:mRNA binding with the computational prediction algorithms TargetScan version 5.1[10] and PicTar[11]. These algorithms were selected based on their ability to predict experimentally validated mammalian miRNA:mRNA targets[59]. Both algorithms exploit a similar strategy for miRNA target prediction based on evolutionary conservation and complementarity to the 'seed' region, which typically comprises nucleotides 2-8 at the 5'end of the mature miRNA sequence. Briefly, TargetScan searches the 3'UTRs of mRNA for conserved segments of perfect complementarity to bases 2-7 of the miRNA and assigns a free energy score to the miRNA-target site interaction, given an internal database of miRNAs annotated from miRBase[60] and 3'UTR sequences annotated from the University of California Santa Cruz (UCSC) Genome Browser. PicTar, abbreviated from 'probabilistic identification of combinations of target sites', computes a maximum likelihood score that a given 3'UTR sequence is targeted by a fixed set of miRNAs by either perfect seed (i.e. complete Watson-Crick base-paired span of 7 nucleotides, starting at the first or second base of the miRNA) or imperfect seed complementarity (i.e. incomplete pairing that allows insertions or alterations in the miRNA sequence as long as its free energy of binding does not increase and contain G:U base-pairings).

Of a total of 3246 unique miRSNPs that were identified via this selection strategy, 1102 SNPs obtained adequate design scores when using Illumina's Assay Design Tool. Of note, the majority (n=1085, 98.5%) of the 1102 SNPs resided in predicted sites of miRNA binding (and therefore represent miRSNPs), while the remainder (n=17) are tagSNPs ($r^2 > 0.80$) for miRSNPs that were not designable or

had poor to moderate design scores). Ninety nine of the 1102 SNPs failed during custom assay development, leaving a total of 1,003 SNPs that were successfully designed and genotyped.

Molecular Studies

Several assays were used to evaluate the putative role in EOC development of every known protein-coding gene in a one megabase (MB) region centered around the most statistically significant SNP at 17q21.31. A 1 MB region was chosen based on previous recommendations[1]. We also evaluated associations between genotype and putative target genes. Data for each gene are illustrated as part of functional maps (Fig. 2).

mRNA expression analyses

Cell culture and RNA extraction for mRNA expression studies

Normal ovarian surface epithelial cells (OSECs) and fallopian tube secretory epithelial cells (FTSECs) were collected with informed consent under the approval of the University College Hospital (UCH), Ethics Committee in London. Primary cell lines were established in culture and expression of lineage-specific markers confirmed by immunofluorescence. Epithelial ovarian cancer (EOC) cells lines were obtained from the ATCC, or were a kind donation from Dr G Mills at MD Anderson. EOC cells were grown in the recommended media for each cell line. All cell lines used in these studies were confirmed to be free of mycoplasma.

RNA for qPCR was extracted from cell cultures harvested at ~80% confluency using the QIAgen RNeasy Kit, according to manufacturers instructions. On-column DNaseI digests were performed. RNA for RNAseq was extracted with a GE Healthcare Illustra RNAspin mini kit with on-column DNaseI digestion. Concentrations were determined with a nanodrop Spectrophotometer.

mRNA expression by qPCR

Expression analysis was performed on 73 normal OSECs/FTSECs, 108 LCL and 50 EOC cell lines. For each cell line 500 ng of RNA was reverse transcribed using SuperScript III First-Strand Synthesis System (Invitrogen). The cDNA was diluted to 10ng/ul and 12.5 ng was used in target specific amplification prior to real-time PCR using TaqMan PreAmp Master Mix Kit (Applied Biosystems) following Fluidigm's Specific Target Amplification Protocol. An aliquot of 1.25 ul of the 25 ul pre-amplified cDNA was added to each chip. Each cDNA sample was run in triplicate and each experiment included PCR negative controls as well as no template controls from the cDNA reactions. 96.96 Dynamic Array Integrated Fluidic Circuits (Fluidigm) were loaded with 96 pre-amplified cDNA samples and 96 TaqMan gene expression probes (Applied Biosystems) using the BioMark HD System (Fluidigm). The assay for each gene was selected to detect the most common transcripts and also the largest number of transcripts (Supplementary Table S3).

Expression data were analyzed using the comparative ΔΔCt method. The mean Ct value for each gene was normalized to the average Ct value for glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and β-Actin controls, and these ΔCt values were then normalized to the highest expressing cancer cell line. The difference between the average ΔΔCT values for EOC and OCPT tissues was determined for each gene, and fold change calculated as $2^{-(\Delta\Delta Ct \text{ cancer} - \Delta\Delta Ct \text{ normal})}$. For pairwise analysis of EOC and normal cell lines, differences in the relative expression between EOC and OCPT lines for each gene was tested using a non-parametric Wilcoxon rank-sum test. For eQTL analysis, genotype specific gene expression was assessed using a nonparametric Jonckheere-Terpstra test. Statistical analysis was performed using R 2.14.1. To analyze relative expression across all genes at each locus, expression of all genes at each locus was calculated relative to the first gene in the region with detectable expression in the 1847 EOC cell line. Gene expression is displayed on a linear scale as box and whisker plots, whiskers indicate the 10th-90th percentile (Supplementary Table S3).

mRNA expression by RNAseq

RNAseq was performed on 6 cell lines (2 OSEC, 2 FTSEC and 2 serous EOC cell lines) using polyA selection and libraries were prepared with an Illumina Tru-Seq RNA sample prep kit. The libraries were barcoded with 4 samples multiplexed per lane of an Illumina HiSeq 2000 using 50bp paired end reads. The average number of aligned reads per sample was 46.5 million reads. The BAM files were uploaded onto the UCSC browser. The color indicates the strand; forward strand=blue, reverse strand=red. (Supplementary Table S3).

FAIREseq analysis

For each OSEC and FTSEC cell line, pairs of biological replicate 15 cm culture dishes containing cells at 80%-90% confluency were cross-linked in 1% formaldehyde. Cells were harvested and lysed in a Tris-buffered 1% SDS lysis buffer containing protease inhibitors. Lysates were sonicated using a QSONICA Model Q125 Ultra Sonic Processor to shear chromatin to 200bp-1kb fragments. Insoluble cell material was removed through centrifugation, and 4x 50µl aliquots from each supernatant were designated as INPUT and 4x 50µl aliquots for FAIRE processing. INPUT samples were incubated overnight at 65° C to reverse cross-linking. All samples were purified through 2 rounds of phenol-chloroform extraction followed by a round of chloroform extraction. DNA was recovered through ethanol precipitation and resulting material from the 4x aliquots of FAIRE and INPUT combined respectively for each biological replicate.

Genome wide sequencing was performed on the Illumina HiSeq platform using 50 bp single read runs. For the FAIRE samples, the 4 samples were barcoded and multiplexed in each lane and 6 lanes were run (equivalent to 1.5 lanes per sample) with an average number of aligned reads per sample of 127 million reads. For the INPUT samples, 2 samples were multiplexed in one lane and 2 lanes were run (equivalent to 1.0 lane per sample) with an average number of aligned reads per sample of 85.3 million reads. The raw data files were uploaded onto the UCSC browser, and a minimum quality score of q30 used.

In silico analysis of mRNA expression in tumor and normal tissue

Affymetrix U133A based gene expression profiling data was obtained from the TCGA website. The data set consisted of 568 primary serous ovarian tumor samples and 8 normal fallopian tube samples. Robust Multi-array analysis was used to normalize and calculate signal intensity for the entire set. The boxplot function in R was used to compare the ovarian samples and fallopian tube samples for 22 probesets mapping to 11 unique genes. A p-value for each comparison was generated using R.

Due to the absence of some probesets on the U133A gene chip and to attempt to verify the results obtained when using the TCGA dataset, a second set of Affymetrix U133Plus based gene expression profiling data was obtained from Gene Expression Omnibus (GEO), series GSE18520. The data set

consisted of 10 normal ovarian surface epithelial brushings and 53 microdissected advanced stage high-grade serous ovarian tumors. Robust multi-array analysis was used to normalize and calculate signal intensity for the entire set. The boxplot function in R was used to compare normal ovary to the ovarian tumor samples for 35 probesets mapping to 14 unique genes. A P-value for each comparison was generated using R. (Supplementary Table S3).

Copy number analysis using TCGA data

Serous ovarian cancer samples for 481 tumors with log2 copy number data were analyzed using the cBio portal for analysis of TCGA data. For each gene in a region the classes of copy number; homozygous deletion, heterozygous loss, diploid, gain, and amplification were queried individually using the advanced onco query language (OQL) option. The frequency of gain and amplification were combined as "gain", and homozygous deletion and heterozygous loss were combined as "loss".

Analysis of copy number vs mRNA expression using TCGA data

Serous ovarian cancer samples for 316 complete tumors (those with CNA, mRNA and sequencing data) were analyzed. Graphs were generated using the cBio portal for analysis of TCGA data and the setting were mRNA expression data Z-score (all genes) with the Z-score threshold of 2 (default setting) and putative copy number alterations (GISTIC). The Z-score is the number of standard deviations away from the mean of expression in the reference population. GISTIC is an algorithm that attempts to identify significantly altered regions of amplification or deletion across sets of patients. (Supplementary Fig. 5)

DNA Methylation

Snap frozen invasive ovarian cancer tumor samples from genotyped cases from the MAY site were obtained during surgery at the Mayo Clinic, reviewed by an experienced gynecologic pathologist (Dr. Gary Keeney), and stored in liquid nitrogen.

Infinium HumanMethylation450 BeadChip analyses were performed on extracted tumor DNA from fresh frozen samples with >70% tumor content (n=227 genotyped MAY high grade serous cases) and normal ovarian tissue (n=7) by the Mayo Clinic Genotyping Shared Facility using the recommended Illumina protocol. Tumor DNA samples (1 ug) were bisulfite modified using the Zymo EZ96 DNA Methylation Kit (Zymo Research, Orange, CA) according to the manufacturer's protocol. BeadChips were imaged on an

Illumina BeadArray iScan reader and analyzed by the GenomeStudio Methylation Module. Analysis included control probes for assessing sample-independent and -dependent performance. Methylation status of the target CpG sites was determined by comparing the ratio of fluorescent signal from the methylated allele to the sum from the fluorescent signal from both methylated and unmethylated alleles. Two batches of tumor samples were processed (n=121 and n=106) with separate QC samples and analysis. The quality of bi-sulfite modification and the performance of the CpG probes were assessed using CEPH control, whole-genome amplified negative control and placental positive control samples. The mean intra-class correlation for the QC samples across the two batches of samples was 0.99, 0.96, and 0.90, respectively. The intra-class correlation for duplicate samples was > 0.99. Based on a plate effect observed within each batch of samples, a correction was applied by fitting models with a fixed plate effect for each maker with the unstandardized residual saved. The logit transformation of the probe mean was added back onto the residual before back transforming to get on 0 to 1 scale. Age-adjusted methylation beta values (% methylated) were compared between HGS tumor samples (n=106) and normal tissue (n=7) that were in a common batch using the non-parametric Kruskal-Wallis test (Supplementary Table S4).

Methylation QTL (mQTL) analysis

Due to the large number of CpG probes for each gene, we prioritized probes for genes within each region based on methylation-expression correlations using 43 MAY fresh frozen tumors with both types of data (Illumina methylation arrays and Agilent expression arrays). Histology- and age-adjusted methylation beta values (% methylated) were correlated with RNA expression using Spearman correlations. The methylation probe (CpG) within 20 kilobases of a particular gene that was most inversely correlated with RNA expression of that gene was reported (cis-negative).

Association between selected CpGs for each gene and genotypes at key SNPs in 17q21.31 were evaluated for the high-grade serous histological subtype cases (n=227). Normalized methylation beta values were adjusted for methylation batch (part 1 or 2) and age. SNP genotype was treated as a continuous variable (additive genetic model) coded in terms of the number of minor alleles present (Supplementary Table S4), with association a with the adjusted methylation beta value completed using Spearman's correlation.

***In silico* functional assessment of SNPs strongly correlated (r^2>0.80) with rs12942666**

We calculated the LD between rs12942666 and SNPs within 1Mb using genotype data from our study population and calls imputed from the phase 1 haplotype data from the January 2012 release of the 1000 genome project. We then evaluated the potential functional role of each correlated SNP (r^2>0.80) using the following *in silico* SNP annotation tools, ANNOVAR[23], SNPinfo[24], and SNPnexus[25]. These tools determine whether the location of a SNP (in relation to adjacent genes) may have a functional influence on regulatory regions, splicing, protein function, or on candidate miRNA genes and their targets. Results from these tools were combined for each SNP to generate a comprehensive annotation list. Collectively, these approaches revealed possible functional significance for approximately 50 of the SNPs. As summarized in Supplementary Table S2, we identified non-synonymous SNPs, SNPs that may affect splicing, transcription factor and/or miRNA binding, methylation, and other regulatory processes.

59 Alexiou, P., Maragkakis, M., Papadopoulos, G.L., Reczko, M. & Hatzigeorgiou, A.G. Lost in translation: an assessment and perspective for computational microRNA target identification. *Bioinformatics* **25**, 3049-55 (2009).

60. Griffiths-Jones, S., Saini, H.K., van Dongen, S. & Enright, A.J. miRBase: tools for microRNA genomics. *Nucleic Acids Res* **36**, D154-8 (2008).