

Published in final edited form as:

Nat Genet. 2016 June ; 48(6): 667–674. doi:10.1038/ng.3562.

Five endometrial cancer risk loci identified through genome-wide association analysis

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

These authors contributed equally to this work.

Abstract

We conducted a meta-analysis of three endometrial cancer GWAS and two replication phases totaling 7,737 endometrial cancer cases and 37,144 controls of European ancestry. Genome-wide imputation and meta-analysis identified five novel risk loci of genome-wide significance at likely regulatory regions on chromosomes 13q22.1 (rs11841589, near *KLF5*), 6q22.31 (rs13328298, in *LOC643623* and near *HEY2* and *NCOA7*), 8q24.21 (rs4733613, telomeric to *MYC*), 15q15.1 (rs937213, in *EIF2AK4*, near *BMF*) and 14q32.33 (rs2498796, in *AKT1* near *SIVA1*). A second independent 8q24.21 signal (rs17232730) was found. Functional studies of the 13q22.1 locus showed that rs9600103 (pairwise $r^2=0.98$ with rs11841589) is located in a region of active chromatin that interacts with the *KLF5* promoter region. The rs9600103-T endometrial cancer protective allele suppressed gene expression *in vitro* suggesting that regulation of *KLF5*

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:http://www.nature.com/authors/editorial_policies/license.html#terms

Correspondence should be addressed to D.F.E. (dfe20@medschl.cam.ac.uk), I.T. (iant@well.ox.ac.uk) or A.B.S. (Amanda.Spurdle@qimrberghofer.edu.au).

Accession codes

Data access for participating studies were granted by their respective management groups i.e. Australian National Endometrial Cancer Study (ANECS), Queensland Institute of Medical Research Controls, Hunter Community Study (HCS), Studies of Epidemiology and Risk Factors in Cancer Heredity (SEARCH), Wellcome Trust Case-Control Consortium (WTCCC), National Study of Endometrial Cancer Genetics (NSEC), Endometrial Cancer Association Consortium (ECAC), Breast Cancer Association Consortium (BCAC) and Ovarian Cancer Association Consortium (OCAC). Genotype data are not freely accessible, but can be obtained by submitting an application to the respective management committees, institutions or data owners.

Author Contributions

A.B.S., D.F.E., A.M.D., G.W.M and P.M.W. obtained funding for the study; A.B.S. and D.F.E. designed the study; T.H.T.C, D.J.T, T.A.O'M, J.N.P., D.M.G, I.T and A.B.S. drafted the manuscript; T.H.T.C and D.J.T. conducted statistical analyses and genotype imputation; T.A.O'M, D.M.G, M.J.L., S.Y. and J.W., conducted bioinformatic analyses; T.A.O'M conducted eQTL analyses; S.F., A.L., J.D.F., L.F-M., D.C., S.L.E. performed functional assays; T.H.T.C., T.A.O'M. and J.N.P. performed additional genotyping by Kaspar and Fluidigm; T.A.O'M. co-ordinated the overall stage 2 genotyping, and associated data management; J.Dennis, J.P.T and K.M. co-ordinated quality control and data cleaning for the iCOGS control datasets; A.B.S. and T.A.O'M. co-ordinated the ANECS stage 1 genotyping; A.M.D., S.A., and C.S.H. co-ordinated the SEARCH stage 1 genotyping; I.T. and CHIBCHA funded and implemented the NSEC GWAS; I.T., L.M., M.G., and S.H. co-ordinated the National Study of the Genetics of Endometrial Cancer (NSEC), and collation of CORGI control GWAS data; A.B.S. and P.M.W. coordinated the Australian National Endometrial Cancer Study (ANECS); R.J.S., M.M^cE., J.A., and E.G.H co-ordinated collation of GWAS data for the Hunter Community Study; N. G. M., G.W.M., D.R.N., and A.K.H co-ordinated collation of GWAS data for the QIMR controls; P.D.P.P., D.F.E. and M.Shah coordinated Studies of Epidemiology and Risk Factors in Cancer Heredity (SEARCH); M.K.B. and Q.W. provided data management support for BCAC; The following authors designed and co-ordinated the baseline studies, and/or extraction of questionnaire and clinical information for studies: P.A.F., M.W.B., A.H., A.B.E, T. D., P. H., M. D., I.R., D.L., S.S., H.Z., and F.A., J.Depreeuw, S.C.D., E.L.G., B.L.F., S.J.W., H.B.S., J.T., T.S.N., H.M.J.W., R.J.S., K.A., T.P., and G.O., M.M, E. T., P.Hall, K.C., J.L, H.D, M.Dunlop, R.H., C.P., J.L.H., J.P., A.J.S., B.B., H.B., A.M., H.Brauch, A.Lindblom, J.C-C., F.J.C., G.G.G., V.N.K., A.C., J.M.C.; All authors provided critical review of the manuscript.

Competing Interests

The authors declare no competing financial interests.

expression, a gene linked to uterine development, is implicated in tumorigenesis. These findings provide enhanced insight into the genetic and biological basis of endometrial cancer.

Endometrial cancer is the fourth most common cancer in women in the United States¹ and Europe², and the most common cancer of the female reproductive system. The familial relative risk is ~23,4, but highly penetrant germline mutations in mismatch repair genes⁵, and DNA polymerases^{6,7} account for only a small proportion of the familial aggregation. Our previous GWAS and subsequent fine-mapping identified the only two reported genome-wide significant endometrial cancer risk loci, tagged by rs11263763 in *HNF1B* intron 18 and rs727479 in *CYP19A1* intron 49.

To identify additional endometrial cancer risk loci, we re-analysed data from our previous GWAS (ANECS, SEARCH datasets¹⁰) and conducted a meta-analysis with two further studies (Supplementary Figure 1). The first was an independent GWAS; the National Study of Endometrial Cancer (NSECG), including 925 endometrial cancer cases genotyped using the Illumina 660W array, 1,286 cancer-free controls from the CORGI/SP1 GWAS^{11,12} and 2,674 controls from the 1958 Birth Cohort¹³. The second study comprised 4,330 endometrial cancer cases and 26,849 controls from Europe, the United States and Australia, genotyped using a custom array designed by the Collaborative Oncological Gene-environment Study (COGS) initiative^{14–17} (Supplementary Table 1, Supplementary Note).

We first performed genome-wide imputation using 1000 Genomes Project data, allowing us to assess up to 8.6 million variants with allele frequency >1% across the different studies. Per-allele odds ratios and P-values for all SNPs in the GWAS and iCOGS were obtained using a logistic regression model. There was little evidence of systematic overdispersion of the test statistic ($\lambda_{GC}=1.002-1.038$, Supplementary Figure 2). A fixed-effects meta-analysis was conducted for all 2.3 million typed and well-imputed SNPs (info score>0.90) in a total of 6,542 endometrial cancer cases and 36,393 controls. The strongest associations were with SNPs in LD with previously identified endometrial cancer risk SNPs in *HNF1B*^{8,10,18} and *CYP19A1*^{9,19} (Figure 1, Table 1). For fourteen 1.5Mb regions containing at least one novel SNP with $P_{meta}<10^{-5}$, we performed regional imputation using an additional reference panel that comprised 196 high-coverage whole genome-sequenced UK individuals (Supplementary Table 2).

Five novel regions containing at least one endometrial cancer risk SNP with $P_{meta}<10^{-7}$ were identified and the most strongly associated SNP in each region was genotyped in an additional 1,195 NSECG endometrial cancer cases and 751 controls using competitive allele-specific PCR (KASPar, KBiosciences) and the Fluidigm BioMark System (Supplementary Table 3). Duplicate samples displayed concordance >98.5% between different genotyping platforms (Supplementary Table 4). All five SNPs were associated with endometrial cancer at genome-wide significance ($P<5\times 10^{-8}$, Table 1, Figure 2, Figure 3), and these associations remained highly significant when analysis was restricted to cases with endometrioid subtype only. Endometrioid-only analysis did not reveal any additional risk loci. eQTL analysis (Online Methods) in normal uterine tissue²⁰, and endometrial cancer tumor and adjacent normal tissue²¹ did not yield any SNPs robustly associated with the expression of nearby genes at the endometrial cancer risk loci (Supplementary Table 5).

However, for each risk locus, bioinformatic analysis including cell-type-specific expression and histone modification data identified correlated SNPs within 500kb in likely enhancers and multiple potential regulatory targets (Supplementary Table 6, Supplementary Figure 3). The most compelling candidates for future functional analysis are described below.

rs13328298 (OR=1.13, 95% CI:1.09–1.18, $P=3.73\times 10^{-10}$) on 6q22.31 lies in the long non-coding RNA *LOC643623*, 54kb upstream of *HEY2* and 86kb upstream of *NCOA7*. *HEY2* is a helix-loop-helix transcriptional repressor in the Notch pathway, which maintains stem cells, and dysregulation has been associated with different cancers²². *NCOA7* modulates the activity of the estrogen receptor via direct binding²³.

The second locus (rs4733613, OR=0.84, 95% CI:0.80–0.89, $P=3.09\times 10^{-9}$) is at 8q24.21. Stepwise conditional logistic regression identified another independent signal in this region, rs17232730 (pairwise $r^2=0.02$, $P_{\text{cond}}=1.29\times 10^{-5}$, Table 2). Both endometrial cancer SNPs lie further from *MYC* (784–846kb telomeric) than most of the other cancer SNPs in the region, including those for cancers of the bladder^{24,25}, breast^{15,26}, colorectum^{12,27}, ovary²⁸ and prostate^{29,30}. rs17232730 is in moderate LD with the ovarian cancer SNP rs10088218 ($r^2=0.43$), with both cancers sharing the same risk allele, but rs4733613 is not in LD ($r^2=0.02$) with any other cancer SNP in the region (Supplementary Figure 3). A role in tumorigenesis is implicated for several miRNAs in the region³¹. Of these, miR-1207-5p is reported to repress *TERT*, a locus also implicated in endometrial cancer risk³².

The lead SNP at 15q15 (rs937213; OR=0.90, 95% CI:0.86–0.93, $P=1.77\times 10^{-8}$) lies within an intron of *EIF2AK4*. *EIF2AK4* encodes a kinase that phosphorylates EIF2 α and downregulates protein synthesis during cellular stress³³. Another nearby gene, *BMF*, encodes an apoptotic regulator moderately to highly expressed in glandular endometrial tissue³⁴.

At 14q42, the lead SNP rs2498796 (OR=0.89, 95% CI:0.85–0.93, $P=3.55\times 10^{-8}$) lies in intron 3 of oncogene *AKT1*, which is highly expressed in the endometrium³⁴. Several SNPs in LD with rs2498796 are bioinformatically linked with regulation of *AKT1* and four other nearby genes (*SIVA1*, *ZBTB42*, *ADSSL1* and *INF2*; Supplementary Table 6, Supplementary Figure 3). *AKT1* acts in the PI3K/AKT/MTOR intracellular signaling pathway, which affects cell survival and proliferation³⁵ and is activated in endometrial tumors³⁶, especially aggressive disease^{37–39}. *SIVA1* encodes an apoptosis regulatory protein that inhibits p53 activity^{40,41} and enhances epithelial–mesenchymal transition to promote motility and invasiveness of epithelial cells⁴². *INF2* expression is reported to act as a promigratory signal in gastric cancer cells treated with mycophenolic acid⁴³.

The final novel endometrial cancer SNP was rs11841589 (OR=1.15, 95% CI:1.11–1.21, $P=4.83\times 10^{-11}$) on chromosome 13q22.1, 163kb and 445kb downstream from Kruppel-like factors *KLF5* and *KLF12*, respectively. *KLF5* is a transcription factor associated with cell cycle regulation, and it plays a role in uterine development, homeostasis and tumorigenesis^{44–47}. Elevated *KLF5* levels are strongly correlated with activating *KRAS* mutations⁴⁸ and *KLF5* is targeted for degradation by the tumor suppressor *FBXW7*. Both *FBXW7* and *KRAS* are commonly mutated in endometrial cancer⁴⁹. rs11841589 was one

of a group of five highly correlated SNPs ($r^2 = 0.98$) surpassing genome-wide significance in a 3kb LD block bounded by rs9600103 ($P=8.70 \times 10^{-11}$) and rs11841589 (Figure 4a). There was no residual association signal at this locus ($P_{\text{cond}} > 0.05$) after conditioning for rs11841589. Bioinformatic analysis suggested that the causal variant at the intergenic 13q22.1 locus may affect a regulatory element that modifies *KLF5* expression (Supplementary Figure 3); rs9600103 overlaps a vertebrate conservation peak, and a DNaseI hypersensitivity site (DHS) in estrogen and tamoxifen-treated ENCODE50 Ishikawa cells (Figure 4a). In addition, in a Hi-C chromatin capture experiment in HeLa S3 cells⁵¹, a chromatin interaction loop was observed between a segment containing the *KLF5* promoter and the rs11841589/rs9600103 locus ($P=0.004$, Supplementary Figure 4).

We further investigated the epigenetic landscape of a 16kb region around rs11841589 and rs9600103 that contained the SNPs most strongly associated with endometrial cancer, by analysis of three endometrial cancer cell lines: Ishikawa (homozygous for the rs9600103-A and rs11841589-G high-risk alleles and provides a comparison with the ENCODE data); ARK-2 (homozygous for the low-risk T alleles at both SNPs); and AN3CA (a non-*KLF5* expressing line that is homozygous for the high-risk alleles) (Supplementary Figure 5). We conducted formaldehyde-assisted identification of regulatory elements (FAIRE, to identify regions of open chromatin), and chromatin immunoprecipitation (ChIP) using antibodies against H3K4Me2 (marker of transcription factor binding⁵²) and panH4Ac (marker of active chromatin). Although the anti-H4Ac ChIP did not display a consistent signal in the region, peaks in signals from FAIRE and anti-H3K4Me2 ChIP were specifically present in the *KLF5*-expressing lines and were co-located with the conservation peak and DHS from the ENCODE data at rs9600103, providing strong evidence for open chromatin and transcription factor binding at this site (Figure 4a). We then conducted chromatin conformation capture experiments for the *KLF5*-expressing Ishikawa endometrial cancer cells (Supplementary Figure 5) and found a significant interaction between the *NcoI* restriction fragment containing the rs11841589/rs9600103 risk loci SNPs and the promoter region of *KLF5* (Figure 4b).

The regulatory nature of the region around rs11841589/rs9600103 was investigated using allele-specific luciferase reporter assays in Ishikawa cells (Figure 4c). Paired t-tests were used to compare the relationships between fragments containing the rs11841589 and rs9600103 alleles, and the pGL3-Promoter reporter vector (no insert) control (Supplementary Table 7). Fragments containing the rs9600103-T, rs11841589-T and rs11841589-G alleles had activity significantly lower than that of the pGL3-Promoter control ($P = 0.014$). In contrast, the construct containing the rs9600103-A risk allele had luciferase expression similar to the pGL3-Promoter control ($P=0.23$) and significantly higher than that of the corresponding rs9600103-T protective allele ($P=0.02$). These results suggest that the endometrial cancer risk tagged by rs11841589 is at least partly due to a regulatory element containing rs9600103, which interacts with the *KLF5* promoter region, with the risk rs9600103-A allele likely associated with increased gene expression.

In summary, this meta-analysis identified five novel endometrial cancer risk loci at genome-wide significance, bringing the total number of common endometrial cancer risk loci identified by GWAS to seven (Figure 1). Together with other risk SNPs reaching study-wide

significance^{32,53,54}, these explain ~5.1% of the endometrial cancer familial relative risk. Novel endometrial cancer risk SNPs lie in likely enhancers predicted to regulate genes or miRNAs with known or suspected roles in tumorigenesis, and we specifically showed that a functional SNP at 13q22.1 may sit within a transcriptional repressor of *KLF5*. Our findings further clarify the genetic etiology of endometrial cancer, provide regions for functional follow-up, and add key information for future risk stratification models.

URLs

rmeta, <http://cran.r-project.org/web/packages/rmeta/>

The Cancer Genome Atlas (TCGA) <http://www.cancergenome.nih.gov/>

Online Methods

Cases and controls were matched as summarized in Supplementary Table 1. Each sample set is described in the Supplementary Note. Supplementary Figure 1 illustrates the overall study design.

Additional EC GWAS

The National Study of Endometrial Cancer Genetics (NSECG) consisted of 925 histologically confirmed endometrial cancer cases from the UK; 86% with endometrioid-only histology. Genotyping was done using Illumina 660W Quad arrays.

These cases were matched with 1,286 cancer-free controls from the UK1/CORGI12 and SP111 colorectal studies genotyped using Illumina Hap550, Hap300 and Hap240S arrays, and 1958 Birth Cohort55 controls from the Wellcome Trust Case Control Consortium (WTCCC2)¹³ genotyped using Illumina Infinium 1.2M arrays.

Original endometrial cancer GWAS

As described previously, cases with endometrioid histology were selected from two population studies; the UK Studies of Epidemiology and Risk factors in Cancer Heredity (SEARCH, n=681) and the Australian National Endometrial Cancer Study (ANECS, n=606), and genotypes generated using Illumina Infinium 610K arrays¹⁰. Compared with our previous study¹⁰, this meta-analysis analysed ANECS and SEARCH as two groups and included additional controls^{8,56}. SEARCH cases were compared with 2,501 controls from the National Blood Service (NBS) part of the WTCCC2 controls¹³. ANECS cases were compared to controls recruited as part of the Hunter Community Study⁵⁶ or Brisbane Adolescent Twin Study⁵⁷, genotyped using Illumina Infinium 610K arrays.

Phase 1 iCOGS genotyping

For the iCOGS genotyping stage, 4,330 women with a confirmed diagnosis of endometrial cancer and European ancestry were recruited via 11 studies in Western Europe, North America and Australia, collectively called the Endometrial Cancer Association Consortium (ECAC).

Healthy female controls with European ancestry and known age at sampling were selected from controls genotyped by the Breast Cancer Association Consortium (BCAC)¹⁵ or Ovarian Cancer Association Consortium (OCAC)¹⁶ iCOGS projects. Eight case-control groups were matched based on geographical location, and principal components analysis (PCA) conducted; individuals who clustered outside the main centroid in pairwise plots of the first four PCs were excluded (Supplementary Figure 7).

Cases and controls were genotyped on a custom Illumina Infinium iSelect array with 211,155 SNPs, designed by the Collaborative Oncological Gene-environment Study (iCOGS), a collaborative project involving four consortia. SNPs were included on this array based on promising regions of interest in previous breast, ovarian and prostate¹⁴ studies, and also the 1,483 top SNPs from our previous EC GWAS¹⁰ analysis. Cases and MoMaTEC controls were genotyped by Genome Quebec Innovation Center. BCAC and OCAC control samples were genotyped at four centres. Raw intensity data files for all consortia were sent to the COGS data co-ordination centre at the University of Cambridge for centralized genotype calling and quality control (QC), so that all case and control genotypes were called using the same procedure.

SNP genotyping arrays quality control

Genotype calling was done using Illumina's proprietary Gencall algorithm and Illuminus⁵⁸. Duplicate samples displayed >99% concordance. Standard QC measures applied to genotyping arrays are described in our original GWAS ¹⁰ and include: genotypic call rate <0.95; deviation from Hardy-Weinberg Equilibrium (HWE) at $P < 10^{-6}$; visual inspection of cluster plots for most significant SNPs. For iCOGS, all endometrial cancer cases and MoMaTEC controls were genotyped by Genome Quebec Innovation Center. BCAC and OCAC control samples were genotyped at four centres. Raw intensity data files for all consortia were sent to the COGS data co-ordination centre at the University of Cambridge for centralized genotype calling and QC, so that all case and control genotypes were called using the same procedure. Duplicate samples for QC showed a concordance of >99%. Samples were excluded based on the following measures: missingness >5%, heterozygosity rates $((N-O)/N) > 5$ S.D from the mean, X chromosome heterozygosity rate (PLINK F-score) >0.2, and pairwise identity by descent (IBD) >0.1875 (cut-off for second-degree relatives). PCA was conducted using Eigenstrat⁵⁹ software. Analysis was conducted using PLINK⁶⁰, and R packages GenABEL and SNPMatrix^{61,62}.

Phase 2 NSECG genotyping

A second genotyping phase consisted of assaying five SNPs with $P < 10^{-7}$ and IMPUTE info scores of >0.94 from the NSECG/ANECS/SEARCH/iCOGS meta-analysis; samples were NSECG cases and controls not previously been used in the NSECG GWAS or NSECG iCOGS. Genotyping was conducted using competitive allele-specific PCR (KASPar, KBiosciences) and the Fluidigm BioMarkTM HD System, using standard protocols. The genotyping call rate was >0.98 and there was a >0.985 concordance between different genotyping platforms (Supplementary Table 4). There was no significant deviation from HWE ($P > 0.05$). Genotyping primers are listed in Supplementary Table 8.

Genome-wide and regional imputation

Genome-wide imputation for all SNP array generated data was conducted using IMPUTE v263 and 1000 Genomes project (2012 release) as reference panel. For the first-pass genome-wide analysis we pre-phased chromosomes using SHAPEIT64 to improve the computational speed. Imputation was carried out separately for the each of the three GWAS studies (for each GWAS study the cases and controls were imputed together as a single dataset, using only SNPs which passed QC in both cases and controls) and for the iCOGS study (all studies within iCOGS were imputed together). SNPs with MAF<0.1% were removed from all studies prior to imputation. Genome-wide imputation produced 9,594,066 SNPs with MAF \geq 1% and info \geq 0.4 in at least one of the three GWAS and eight iCOGS groups. Of these, 8,308,423 SNPs met these criteria in all studies. The iCOGS genotyping array (~200,000 SNPs) is aimed at capturing previously prioritised cancer SNPs and not genome-wide coverage, but nonetheless 8,631,871 SNPs met MAF \geq 1% and info \geq 0.4 criteria, of which 5,437,135 had info \geq 0.7 and 2,333,040 had info \geq 0.9.

Regional imputation of regions of interest (1.5Mb region around SNPs with meta-analysis $P < 10^{-5}$) used both 1000 Genomes 2012 release and 196 high-coverage, whole genome-sequenced UK individuals as reference panels as a means to improve imputation accuracy⁶⁵. All SNPs reported in this study had an info score \geq 0.9 in all datasets.

Association testing

Association testing was done using SNPTEST v266 employing frequentist tests with a logistic regression model for each of the 11 groups as matched in Supplementary Table 1. There was little evidence of systematic over-dispersion of the test statistic from the quantile-quantile plots (Supplementary Figure 2) and the genomic inflation λ_{GC} , calculated using all genotyped SNPs passing QC for the three GWAS. For iCOGS, 105,000 SNPs after LD-pruning ($r^2 < 0.2$) and > 500 kb from the 1,483 EC prioritized SNPs on the iCOGS were used. λ_{GC} was between 1.002 and 1.038 for each study. Conditional logistic regression analysis was conducted for each locus of genome-wide significance using SNPTEST to look for the presence of multiple independent association signals. This was done in a stepwise manner, first conditioning for the most significant SNP and subsequently for any SNPs that remained significant at $P_{\text{cond}} < 10^{-4}$. Regional association plots (Figure 1, Supplementary Figure 6) were created using LocusZoom⁶⁷.

Meta-analysis

Inverse variance, fixed effects meta-analysis of the 11 groups (three GWAS, eight iCOGS groups) was conducted using GWAMA⁶⁸. The per allele effect size of each SNP in a particular study is represented by β (the log-odds ratio) and its standard error. Inter-study differences are represented by the I^2 heterogeneity score^{69,70}. Forest plots of the genome-wide significant loci (Figure 2) representing risk effects across different studies were made using rmeta. A random-effects meta-analysis was also performed for SNPs with $I^2 < 0.3$. The results of the second replication phase (NSECG replication) were meta-analyzed in a 12-way meta-analysis for the top 5 SNPs yielding a total of 7,737 EC cases and 37,144 controls. 6,635 (86%) of the EC cases had endometrioid-only histology and association testing and meta-analysis were also conducted with just these samples.

Bioinformatic analysis and functional annotation of genome-wide significant risk loci

The five novel genome-wide significant loci and SNPs in LD ($r^2 > 0.7$ in European 1000 Genomes) were annotated using HaploregV271, RegulomeDB72 and data from ENCODE50 in Supplementary Table 6. This includes information such as promoter and enhancer histone marks, DHS, bound proteins, altered motifs, GENCODE and dbSNP annotations, RegulomeDB score and PhastCons conservation scores.

Bioinformatic analysis in Supplementary Figure 3 used datasets described by Hnisz *et al.* 73 and Corradin *et al.* 74 to identify likely enhancers in a cell-specific context for the risk loci. Enhancer-gene interactions are predicted by identifying ‘super-enhancers’ (regions containing neighbouring H3K27Ac modifications) from 86 cell and tissue types and then the expressed transcript with transcription start site closest to the centre of the super-enhancer was assigned as the target gene. PresTIGE pairs cell-type specific H3K4Me1 and gene expression data from 13 cell types to identify likely enhancer-gene interactions.

Endometrial-tissue expression quantitative trait loci (eQTL) analysis for associated SNPs using GTEx and TCGA data

Publicly available data generated by the Genotype-Tissue Expression Project (GTEx)²⁰ and The Cancer Genome Atlas (TCGA) were accessed to examine tissue-specific eQTLs. For GTEx, expression and genotype data were generated from 70 normal uteri from post-mortem biopsies, using an Affymetrix Expression array and Illumina Omni 5M SNP array. GTEx provided processed results, evaluating association between genotype and expression data. The expression levels are represented as a rank normalized score. TCGA genotype and copy number variation (CNV) data were derived from Affymetrix 6.0 SNP arrays. Expression data were from RNAseq arrays (Illumina HiSeq and Illumina GA) for 458 endometrial cancer tissues and 30 adjacent normal endometrial tissues. Association analyses for TCGA datasets were performed as follows. Genes within 500kb flanking our SNPs of interest were selected for analysis. Since there may be significant variation in tumour tissue copy number, somatic CNVs were taken into account by regressing gene expression to average copy number spanning the gene. Residual unexplained variance in gene expression was then regressed on the genotype of the lead SNP at each locus, using genotyped or imputed data. Statistical comparisons were subject to Bonferroni correction for number of tests (number of sample sets, and number of genes assessed).

DNA and RNA extraction from cell lines

Cell lines were from the laboratory of Dr David Church, acquired as gifts from Brittia Weigelt (currently at Memorial Sloan Kettering, USA), and Konstantin Dedes (University of Zurich, Austria), were routinely tested for mycoplasma contamination. Somatic mutation data generated previously matches that reported in publicly available resources and the literature, where available. Cells were snap frozen with dry ice after centrifugation, and DNA and RNA extracted using DNeasy and RNeasy minikits (Qiagen). Nucleic acids were quantified using Nanodrop 2000 (ThermoScientific) spectrophotometry.

Quantification of *KLF5* expression in endometrial cancer cell lines

Extracted RNA was treated with DNase 1, and complimentary DNA (cDNA) was reverse transcribed from RNA using the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems). TaqMan Gene Expression Assays were used for *KLF5* and *GAPDH* (details available from authors). The absolute expression of *KLF5* was quantified using qRT-PCR using the ABI 7900HT cycler (Applied Biosystems), and the critical threshold was manually set at 0.2. Relative expression was calculated using the $\Delta\Delta C_T$ method described by Livak and Schmittgen⁷⁵, with *GAPDH* as an endogenous control.

Formaldehyde-assisted identification of regulatory elements (FAIRE)

Formaldehyde-Assisted Isolation of Regulatory Elements (FAIRE) was conducted using the method adapted from Giresi et al⁷⁶. Briefly, cross-linking was done on a rocker at room temperature. 1% formaldehyde was added to ~10⁸ cells for 5 minutes, and 115mM glycine added to inhibit cross-linking. For each cell line, a non-crosslinked control was prepared in parallel for all remaining steps. After two rinses with 4°C phosphate buffered saline solution (PBS), cells were suspended in successive buffers: Lysis buffer I (50mM HEPES-KOH, 140mM NaCl, 1mM EDTA, 10% glycerol, 0.5% NP-40, 0.25% tritonX-100); lysis buffer II (10mM tris-HCl, 200mM NaCl, 1mM EDTA, 0.5mM EGTA); lysis buffer III (10mM tris-HCl, 2100mM NaCl, 1mM EDTA, 0.1% sodium deoxycholate, 0.5% N-lauroylsarcosine). Cells were incubated on a rocker at 4°C for 10 minutes in each lysis buffer, then spun down at 1300 g for 5 minutes, and the supernatant removed. The cells were then sonicated using the Bioruptor in seven to fifteen 30-second cycles to generate fragments 100-1000 bp in size, and gel electrophoresis in 1% agarose used to confirm DNA fragment sizes. The DNA was extracted with a standard phenol/chloroform method and ethanol-precipitated. 50ng of DNA from paired crosslinked and non-crosslinked cells was analyzed in duplicate by SYBR-green quantitative PCR (qPCR) using primers at ~1kb intervals in the 13q22.1 region downstream of *KLF5* (Supplementary Table 8). The ΔC_T method³¹ was used to normalize results to the input DNA from non-crosslinked cells and then expressed relative to the Rhodopsin promoter as negative control. For each experiment there were two replicates for the crosslinked cells and non-crosslinked controls, each performed on two occasions.

Cross-linked Chromatin immunoprecipitation (ChIP)

About 10⁸ cells were cross-linked using 1% formaldehyde for 10 minutes. Glycine was used to stop the cross-linking, cells were then rinsed twice in PBS, and cell scrapers used to detach cells adhered to the Petri dish surface. Cells were then resuspended in lysis buffer (1% sodium dodecyl sulfate (SDS), 10mM EDTA (Ambion), 50mM Tris-HCl (Ambion)) incubated for 10 minutes, and then sonicated using the Bioruptor (Diagenode) in 7 to 15 30-second cycles to generate fragments 1000-1500 bp in size. Gel electrophoresis in 1% agarose confirmed the size of the DNA fragments. The fragmented DNA was then diluted ten times to the immuno-precipitation dilution buffer (1% tritonX-100, 2nM EDTA, 20mM Tris-HCl, 150mM sodium chloride and each cell line was separated into four tubes: input chromatin, no-antibody-control and one tube for each antibody. 5ul of anti-dimethyl-histone H3 Lys4 (Millipore 07-030) and anti-acetyl-histone H4 (Millipore 06-866) were added to the antibody tubes and, along with the no-antibody-control, incubated overnight at 4°C for

immunoprecipitation. The input chromatin was kept refrigerated at 4°C until the reverse cross-linking of day 2. Phenylmethylsulfonyl fluoride and protease inhibitor was added to the lysis buffer and IP dilution buffer to deactivate proteases, while sodium butyrate was added to these solutions to inhibit histone deacetylases. 5ul of protein A Dynabeads was added to each tube and incubated for 4 hours. A series of washes were done using Tris/Sucrose/EDTA (TSE) I (1% tritonX-100, 2mM EDTA, 20mM Tris-HCl, 150mM NaCl, 0.1% SDS), TSE II (1% tritonX-100, 2mM EDTA, 20mM Tris-HCl, 500mM NaCl, 0.1% SDS), Buffer III (0.25M lithium chloride, 1mM EDTA, 10mM Tris-HCl, 1% tergitol-type NP-40, 1% sodium deoxycholate) and tris-EDTA (1X). 300ul of extraction solution (1% SDS 0.1M sodium bicarbonate) was added and Dynabeads were removed after a 30 minute incubation. Then 0.7 M NaCl was added and reverse cross-linking occurred overnight at 65°C. DNA was purified using the QIAquick PCR purification kit (Qiagen). 1ul of DNA was analyzed in duplicate or triplicate by SYBR green qPCR as above and the Ct method was used to identify areas with enrichment. For each experiment there were two replicates for each antibody along with the input and no-antibody control, each performed on two occasions. Primers used are listed in Supplementary Table 8.

Chromatin conformation capture (3C)

Experiments were performed as described in Ghoussaini *et al.*⁷⁷, using the *KLF5*-expressing Ishikawa endometrial cancer cell line from ATCC. The cell line was authenticated using a short tandem repeat (STR) profiling, and routinely tested for mycoplasma contamination (QIMR Berghofer in-house Support Services). Briefly, Ishikawa cell lines were crosslinked with 1% formaldehyde for 10 mins, quenched with 125mM glycine, washed with PBS and collected by scraping. Cells were lysed for 30 min on ice in 10mM Tris-HCl, pH 7.5, 10mM NaCl, 0.2% Igepal with protease inhibitors and homogenized in a Dounce homogenizer. Nuclei were pelleted and resuspended in 1ml 1.2X restriction buffer (NEB 3.1) with 0.3% SDS for 1h at 37°C. 2% Triton X-100 was added then 1000U NcoI was added 3 times over 24h at 37°C with shaking. The enzyme was inactivated, and digested DNA diluted 8X before ligation with 4000U of T4 DNA ligase overnight at 16°C. Crosslinks were reversed by proteinase K digestion at 65°C overnight, and the DNA purified by phenol-chloroform extraction and ethanol precipitation. The final DNA pellet was dissolved in 10mM Tris (pH 7.5) and purified through Amicon Ultra 0.5 ml columns (Millipore). 3C interactions were quantified by SYTO9 qPCR (performed on a RotorGene 6000) using primers designed to amplify across ligated NcoI restriction fragments with one constant primer within the risk fragment (including rs11841589 and rs9600103) and a series of test primers within NcoI fragments spanning 76 kb of the *KLF5* promoter region. BAC clones (RP11-81D9, RP11-179I20) covering the region were digested with NcoI, ligated with T4 ligase and used to determine PCR efficiency. 3C analyses were performed on three independent 3C libraries, with each data point in duplicate. Data were normalized to the signal from the BAC clone library and from a non-interacting chromosomal region using the Ct method with incorporated individual primer pair efficiencies.

Luciferase reporter assays

For luciferase reporter assays, the regions chr:13 73,810,509-73,813,452 around rs9600103 and chr13:73,813,268-73,816,290 around rs11841589 were cloned into the pGL3-Promoter vector (Promega) to test for regulatory effects in Ishikawa cells. Ishikawa cells were selected because they express *KLF5*, showed evidence of a DHS, FAIRE and H3K4Me2 enrichment at rs9600103 and were readily transfectable. Site-directed mutagenesis was used so both the high- and low-risk alleles of rs9600103 and rs11841589 were tested. After sequencing to verify the correct insert sequences, cells were transiently co-transfected using lipofectamine with the appropriate pGL3-Promoter constructs, and the Renilla luciferase pGL4.75 vector (Promega) as control for transfection efficiency. After 48 hours, luciferase activity was measured (Dual-Glo Luciferase Assay System, Promega), and after subtracting background from lipofectamine-only controls, firefly luciferase activity from the putative enhancer regions was normalized to the Renilla luciferase values for each sample. Levels of firefly luciferase activity were compared with a control plasmid consisting of an empty pGL3, and also a noncoding 2.2-kb stretch of plasmid sequence from the pENTR1A plasmid (Invitrogen) cloned into the pGL3-Promoter vector previously used as a length of DNA with no regulatory activity⁷⁸. Luciferase activity experiments had three or four replicates, each performed on three occasions (total of 11 assays). Primers used are listed in Supplementary Table 8.

ANOVA found significant differences in luciferase levels ($P < 0.0001$, $F: 11.6$) but no significant differences between replicates conducted on different days ($P = 0.91$, $F: 0.09$). There were no significant differences between the pENTR1A control and the empty pGL3-Promoter vector ($P = 0.085$); pGL3-Promoter vector was used as control. We conducted paired t-tests for all comparisons using the average of biological repeats, between the pGL3 no insert, rs9600103-A, rs9600103-T, rs11841589-G and rs11841589-T fragments (Supplementary Table 7, results unadjusted for multiple comparisons).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Authors

Timothy HT Cheng^{#1}, Deborah J Thompson^{#2}, Tracy A O'Mara³, Jodie N Painter³, Dylan M Glubb³, Susanne Flach¹, Annabelle Lewis¹, Juliet D French³, Luke Freeman-Mills¹, David Church¹, Maggie Gorman¹, Lynn Martin¹, National Study of Endometrial Cancer Genetics Group (NSECG)^{1,4}, Shirley Hodgson⁵, Penelope M Webb³, The Australian National Endometrial Cancer Study Group (ANECS)^{3,4}, John Attia^{6,7}, Elizabeth G Holliday^{6,7}, Mark McEvoy⁷, Rodney J Scott^{6,8,9,10}, Anjali K Henders³, Nicholas G Martin³, Grant W Montgomery³, Dale R Nyholt^{3,11}, Shahana Ahmed¹², Catherine S Healey¹², Mitul Shah¹², Joe Dennis², Peter A Fasching^{13,14}, Matthias W Beckmann¹⁴, Alexander Hein¹⁴, Arif B Ekici¹⁵, Per Hall¹⁶, Kamila Czene¹⁶, Hatem Darabi¹⁶, Jingmei Li¹⁶, Thilo Dörk¹⁷, Matthias Dürst¹⁸, Peter Hillemanns¹⁹, Ingo Runnebaum¹⁸, Frederic Amant²⁰, Stefanie Schrauwen²⁰, Hui Zhao^{21,22}, Diether Lambrechts^{21,22}, Jeroen Depreeuw^{20,21,22}, Sean C Dowdy²³,

Ellen L Goode²⁴, Brooke L Fridley²⁵, Stacey J Winham²⁴, Tormund S Njølstad^{26,27}, Helga B Salvesen^{26,27}, Jone Trovik^{26,27}, Henrica MJ Werner^{26,27}, Katie Ashton^{6,9,10}, Geoffrey Otton²⁸, Tony Proietto²⁸, Tao Liu²⁹, Miriam Mints³⁰, Emma Tham^{29,31}, RENDOCAS^{4,29}, CHIBCHA Consortium^{1,4}, Mulin Jun Li³², Shun H Yip³², Junwen Wang³², Manjeet K Bolla², Kyriaki Michailidou², Qin Wang², Jonathan P Tyrer¹², Malcolm Dunlop^{33,34}, Richard Houlston³⁵, Claire Palles¹, John L Hopper³⁶, AOCs Group^{3,4,37}, Julian Peto³⁸, Anthony J Swerdlow^{35,39}, Barbara Burwinkel^{40,41}, Hermann Brenner^{42,44}, Alfons Meindl⁴⁵, Hiltrud Brauch^{44,46,47}, Annika Lindblom²⁹, Jenny Chang-Claude^{48,49}, Fergus J Couch^{24,50}, Graham G Giles^{36,51,52}, Vessela N Kristensen^{53,55}, Angela Cox⁵⁶, Julie M Cunningham⁵⁰, Paul D P Pharoah¹², Alison M Dunning¹², Stacey L Edwards³, Douglas F Easton^{2,12}, Ian Tomlinson¹, and Amanda B Spurdle³

Affiliations

- ¹Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK
- ²Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK
- ³Department of Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia
- ⁴A list of members and affiliations appears in the Supplementary Note
- ⁵Department of Clinical Genetics, St George's, University of London, London, UK
- ⁶Hunter Medical Research Institute, John Hunter Hospital, Newcastle, NSW, Australia
- ⁷Centre for Clinical Epidemiology and Biostatistics, School of Medicine and Public Health, University of Newcastle, NSW, Australia
- ⁸Hunter Area Pathology Service, John Hunter Hospital, Newcastle, NSW, Australia
- ⁹Centre for Information Based Medicine, University of Newcastle, NSW, Australia
- ¹⁰School of Biomedical Sciences and Pharmacy, University of Newcastle, Newcastle, NSW, Australia
- ¹¹Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, Australia
- ¹²Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK
- ¹³University of California at Los Angeles, Department of Medicine, Division of Hematology/Oncology, David Geffen School of Medicine, Los Angeles, CA, USA
- ¹⁴Department of Gynecology and Obstetrics, University Hospital Erlangen, Friedrich-Alexander University Erlangen-Nuremberg, Erlangen, Germany
- ¹⁵Institute of Human Genetics, University Hospital Erlangen, Friedrich-Alexander-University Erlangen-Nuremberg, Erlangen, Germany

- ¹⁶Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden
- ¹⁷Hannover Medical School, Gynaecology Research Unit, Hannover, Germany
- ¹⁸Department of Gynaecology, Jena University Hospital - Friedrich Schiller University, Jena, Germany
- ¹⁹Hannover Medical School, Clinics of Gynaecology and Obstetrics, Hannover, Germany
- ²⁰Department of Obstetrics and Gynecology, Division of Gynecologic Oncology, University Hospitals, KU Leuven - University of Leuven, Belgium
- ²¹Vesalius Research Center, VIB, Leuven, Belgium
- ²²Laboratory for Translational Genetics, Department of Oncology, University Hospitals Leuven, Leuven, Belgium
- ²³Department of Obstetrics and Gynecology, Division of Gynecologic Oncology, Mayo Clinic, Rochester, MN, USA
- ²⁴Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA
- ²⁵Department of Biostatistics, University of Kansas Medical Center, Kansas City, KS, USA
- ²⁶Centre for Cancerbiomarkers, Department of Clinical Science, The University of Bergen, Norway
- ²⁷Department of Obstetrics and Gynecology, Haukeland University Hospital, Bergen, Norway
- ²⁸School of Medicine and Public Health, University of Newcastle, Newcastle, NSW, Australia
- ²⁹Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden
- ³⁰Department of Women's and Children's Health, Karolinska Institutet, Karolinska University Hospital, Stockholm, Sweden
- ³¹Clinical Genetics, Karolinska University Hospital Solna, Stockholm, Sweden
- ³²Centre for Genomic Sciences, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong SAR, China
- ³³Colon Cancer Genetics Group, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK
- ³⁴MRC Human Genetics Unit, Western General Hospital Edinburgh, Edinburgh, UK
- ³⁵Division of Genetics and Epidemiology, Institute of Cancer Research, London, UK
- ³⁶Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Vic, Australia

- ³⁷Peter MacCallum Cancer Center, The University of Melbourne, Melbourne, Australia
- ³⁸London School of Hygiene and Tropical Medicine, London, UK
- ³⁹Division of Breast Cancer Research, Institute of Cancer Research, London, UK
- ⁴⁰Molecular Biology of Breast Cancer, Department of Gynecology and Obstetrics, University of Heidelberg, Heidelberg, Germany
- ⁴¹Molecular Epidemiology Group, German Cancer Research Center (DKFZ), Heidelberg, Germany
- ⁴²Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany
- ⁴³Division of Preventive Oncology, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), Heidelberg, Germany
- ⁴⁴German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany
- ⁴⁵Department of Obstetrics and Gynecology, Division of Tumor Genetics, Technical University of Munich, Munich, Germany
- ⁴⁶Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Stuttgart, Germany
- ⁴⁷University of Tübingen, Tübingen, Germany
- ⁴⁸Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany
- ⁴⁹University Cancer Center Hamburg (UCCH), University Medical Center Hamburg-Eppendorf, Hamburg, Germany
- ⁵⁰Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, USA
- ⁵¹Cancer Epidemiology Centre, Cancer Council Victoria, Melbourne, Vic, Australia
- ⁵²Department of Epidemiology and Preventive Medicine, Monash University, Melbourne, Vic, Australia
- ⁵³Department of Genetics, Institute for Cancer Research, The Norwegian Radium Hospital, Oslo, Norway
- ⁵⁴The K.G. Jebsen Center for Breast Cancer Research, Institute for Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway
- ⁵⁵Department of Clinical Molecular Oncology, Division of Medicine, Akershus University Hospital, Lørenskog, Norway
- ⁵⁶Sheffield Cancer Research, Department of Oncology, University of Sheffield, Sheffield, UK

Acknowledgments

The authors thank the many individuals who participated in this study and the numerous institutions and their staff who supported recruitment, detailed in full in the Supplementary Text.

The iCOGS endometrial cancer analysis was supported by NHMRC project grant [ID#1031333] to ABS, DFE and AMD. ABS, PW, GWM, and DRN are supported by the NHMRC Fellowship scheme. AMD was supported by the Joseph Mitchell Trust. IT is supported by Cancer Research UK and the Oxford Comprehensive Biomedical Research Centre. THTC is supported by the Rhodes Trust and the Nuffield Department of Medicine. Funding for the iCOGS infrastructure came from: the European Community's Seventh Framework Programme under grant agreement no 223175 [HEALTH-F2-2009-223175] [COGS], Cancer Research UK [C1287/A10118, C1287/A10710, C12292/A11174, C1281/A12014, C5047/A8384, C5047/A15007, C5047/A10692, C8197/A16565], the National Institutes of Health [CA128978] and Post-Cancer GWAS initiative [1U19 CA148537, 1U19 CA148065 and 1U19 CA148112 - the GAME-ON initiative], the Department of Defence [W81XWH-10-1-0341], the Canadian Institutes of Health Research [CIHR] for the CIHR Team in Familial Risks of Breast Cancer, Komen Foundation for the Cure, the Breast Cancer Research Foundation, and the Ovarian Cancer Research Fund.

ANES recruitment was supported by project grants from the NHMRC [ID#339435], The Cancer Council Queensland [ID#4196615] and Cancer Council Tasmania [ID#403031] and ID#457636]. SEARCH recruitment was funded by a programme grant from Cancer Research UK [C490/A10124]. Stage 1 and stage 2 case genotyping was supported by the NHMRC [ID#552402, ID#1031333]. Control data were generated by the Wellcome Trust Case Control Consortium (WTCCC), and a full list of the investigators who contributed to the generation of the data is available from the WTCCC website. We acknowledge use of DNA from the British 1958 Birth Cohort collection, funded by the Medical Research Council grant G0000934 and the Wellcome Trust grant 068545/Z/02 - funding for this project was provided by the Wellcome Trust under award 085475. NSECG was supported by the EU FP7 CHIBCHA grant, Wellcome Trust Centre for Human Genetics Core Grant 090532/Z/09Z, and CORGI was funded by Cancer Research UK. Recruitment of the QIMR Berghofer controls was supported by the NHMRC. The University of Newcastle, the Gladys M Brawn Senior Research Fellowship scheme, The Vincent Fairfax Family Foundation, the Hunter Medical Research Institute and the Hunter Area Pathology Service all contributed towards the costs of establishing the Hunter Community Study.

The Bavarian Endometrial Cancer Study (BECS) was partly funded by the ELAN fund of the University of Erlangen. The Hannover-Jena Endometrial Cancer Study was partly supported by the Rudolf Bartling Foundation. The Leuven Endometrium Study (LES) was supported by the Verelst Foundation for endometrial cancer. The Mayo Endometrial Cancer Study (MECS) and Mayo controls (MAY) were supported by grants from the National Cancer Institute of United States Public Health Service [R01 CA122443, P30 CA15083, P50 CA136393, and GAME-ON the NCI Cancer Post-GWAS Initiative U19 CA148112], the Fred C and Katherine B Andersen Foundation, the Mayo Foundation, and the Ovarian Cancer Research Fund with support of the Smith family, in memory of Kathryn Sladek Smith. MoMaTEC received financial support from a Helse Vest Grant, the University of Bergen, Melzer Foundation, The Norwegian Cancer Society (Harald Andersens legat), The Research Council of Norway and Haukeland University Hospital. The Newcastle Endometrial Cancer Study (NECS) acknowledges contributions from the University of Newcastle, The NBN Children's Cancer Research Group, Ms Jennie Thomas and the Hunter Medical Research Institute. RENDOCAS was supported through the regional agreement on medical training and clinical research (ALF) between Stockholm County Council and Karolinska Institutet [numbers: 20110222, 20110483, 20110141 and DF 07015], The Swedish Labor Market Insurance [number 100069] and The Swedish Cancer Society [number 11 0439]. The Cancer Hormone Replacement Epidemiology in Sweden Study (CAHRES, formerly called The Singapore and Swedish Breast/Endometrial Cancer Study; SASBAC) was supported by funding from the Agency for Science, Technology and Research of Singapore (A*STAR), the US National Institutes of Health and the Susan G. Komen Breast Cancer Foundation.

The Breast Cancer Association Consortium (BCAC) is funded by Cancer Research UK [C1287/A10118, C1287/A12014]. The Ovarian Cancer Association Consortium (OCAC) 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], and the UK National Institute for Health Research Biomedical Research Centres at the University of Cambridge.

Additional funding for individual control groups is detailed in the Supplementary Text.

Abbreviations

CI	confidence interval
GWAS	genome-wide association study
LD	linkage disequilibrium

OR	odds ratio
kb	kilobase
Mb	megabase
PCA	principal components analysis
DHS	DNase1 hypersensitivity site

References

1. Siegel R, Ma J, Zou Z, Jemal A. Cancer statistics, 2014. *CA Cancer J Clin.* 2014; 64:9–29. [PubMed: 24399786]
2. Ferlay J, et al. Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012. *Eur J Cancer.* 2013; 49:1374–1403. [PubMed: 23485231]
3. Gruber SB, Thompson WD. A population-based study of endometrial cancer and familial risk in younger women. *Cancer and Steroid Hormone Study Group. Cancer Epidemiol Biomarkers Prev.* 1996; 5:411–417. [PubMed: 8781735]
4. Win AK, Reece JC, Ryan S. Family history and risk of endometrial cancer: a systematic review and meta-analysis. *Obstet Gynecol.* 2015; 125:89–98. [PubMed: 25560109]
5. Barrow E, Hill J, Evans DG. Cancer risk in Lynch Syndrome. *Fam Cancer.* 2013; 12:229–240. [PubMed: 23604856]
6. Church DN, et al. DNA polymerase epsilon and delta exonuclease domain mutations in endometrial cancer. *Hum Mol Genet.* 2013; 22:2820–2828. [PubMed: 23528559]
7. Palles C, et al. Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. *Nat Genet.* 2013; 45:136–144. [PubMed: 23263490]
8. Painter JN, et al. Fine-mapping of the HNF1B multicancer locus identifies candidate variants that mediate endometrial cancer risk. *Hum Mol Genet.* 2015; 24:1478–1492. [PubMed: 25378557]
9. Thompson DJ, et al. CYP19A1 fine-mapping and Mendelian randomization: estradiol is causal for endometrial cancer. *Endocr Relat Cancer.* 2016; 23:77–91. [PubMed: 26574572]
10. Spurdle AB, et al. Genome-wide association study identifies a common variant associated with risk of endometrial cancer. *Nat Genet.* 2011; 43:451–454. [PubMed: 21499250]
11. Tenesa A, et al. Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21. *Nat Genet.* 2008; 40:631–637. [PubMed: 18372901]
12. Tomlinson I, et al. A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. *Nat Genet.* 2007; 39:984–988. [PubMed: 17618284]
13. Wellcome Trust Case Control C. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature.* 2007; 447:661–678. [PubMed: 17554300]
14. Eeles RA, et al. Identification of 23 new prostate cancer susceptibility loci using the iCOGS custom genotyping array. *Nat Genet.* 2013; 45:385–391. 391e381–382. [PubMed: 23535732]
15. Michailidou K, et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat Genet.* 2013; 45:353–361. 361e351–352. [PubMed: 23535729]
16. Pharoah PD, et al. GWAS meta-analysis and replication identifies three new susceptibility loci for ovarian cancer. *Nat Genet.* 2013; 45:362–370. 370e361–362. [PubMed: 23535730]
17. Sakoda LC, Jorgenson E, Witte JS. Turning of COGS moves forward findings for hormonally mediated cancers. *Nat Genet.* 2013; 45:345–348. [PubMed: 23535722]
18. De Vivo I, et al. Genome-wide association study of endometrial cancer in E2C2. *Hum Genet.* 2014; 133:211–224. [PubMed: 24096698]

19. Setiawan VW, et al. Two estrogen-related variants in CYP19A1 and endometrial cancer risk: a pooled analysis in the Epidemiology of Endometrial Cancer Consortium. *Cancer Epidemiol Biomarkers Prev.* 2009; 18:242–247. [PubMed: 19124504]
20. Consortium GT. The Genotype-Tissue Expression (GTEx) project. *Nat Genet.* 2013; 45:580–585. [PubMed: 23715323]
21. Cancer Genome Atlas Research, N. Integrated genomic characterization of endometrial carcinoma. *Nature.* 2013; 497:67–73. [PubMed: 23636398]
22. Katoh M, Katoh M. Integrative genomic analyses on HES/HEY family: Notch-independent HES1, HES3 transcription in undifferentiated ES cells, and Notch-dependent HES1, HES5, HEY1, HEY2, HEYL transcription in fetal tissues, adult tissues, or cancer. *Int J Oncol.* 2007; 31:461–466. [PubMed: 17611704]
23. Shao W, Halachmi S, Brown M. ERAP140, a conserved tissue-specific nuclear receptor coactivator. *Mol Cell Biol.* 2002; 22:3358–3372. [PubMed: 11971969]
24. Kiemeny LA, et al. Sequence variant on 8q24 confers susceptibility to urinary bladder cancer. *Nat Genet.* 2008; 40:1307–1312. [PubMed: 18794855]
25. Rothman N, et al. A multi-stage genome-wide association study of bladder cancer identifies multiple susceptibility loci. *Nat Genet.* 2010; 42:978–984. [PubMed: 20972438]
26. Easton DF, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature.* 2007; 447:1087–1093. [PubMed: 17529967]
27. Whiffin N, et al. Identification of susceptibility loci for colorectal cancer in a genome-wide meta-analysis. *Hum Mol Genet.* 2014; 23:4729–4737. [PubMed: 24737748]
28. Goode EL, et al. A genome-wide association study identifies susceptibility loci for ovarian cancer at 2q31 and 8q24. *Nat Genet.* 2010; 42:874–879. [PubMed: 20852632]
29. Eeles RA, et al. Identification of seven new prostate cancer susceptibility loci through a genome-wide association study. *Nat Genet.* 2009; 41:1116–1121. [PubMed: 19767753]
30. Gudmundsson J, et al. Genome-wide association and replication studies identify four variants associated with prostate cancer susceptibility. *Nat Genet.* 2009; 41:1122–1126. [PubMed: 19767754]
31. Huppi K, Pitt JJ, Wahlberg BM, Caplen NJ. The 8q24 gene desert: an oasis of non-coding transcriptional activity. *Front Genet.* 2012; 3:69. [PubMed: 22558003]
32. Carvajal-Carmona LG, et al. Candidate locus analysis of the TERT-CLPTM1L cancer risk region on chromosome 5p15 identifies multiple independent variants associated with endometrial cancer risk. *Hum Genet.* 2015; 134:231–245. [PubMed: 25487306]
33. Berlanga JJ, Santoyo J, De Haro C. Characterization of a mammalian homolog of the GCN2 eukaryotic initiation factor 2alpha kinase. *Eur J Biochem.* 1999; 265:754–762. [PubMed: 10504407]
34. Uhlen M, et al. Proteomics. Tissue-based map of the human proteome. *Science.* 2015; 347:1260419. [PubMed: 25613900]
35. Cantley LC. The phosphoinositide 3-kinase pathway. *Science.* 2002; 296:1655–1657. [PubMed: 12040186]
36. Slomovitz BM, Coleman RL. The PI3K/AKT/mTOR pathway as a therapeutic target in endometrial cancer. *Clin Cancer Res.* 2012; 18:5856–5864. [PubMed: 23082003]
37. Cohen Y, et al. AKT1 pleckstrin homology domain E17K activating mutation in endometrial carcinoma. *Gynecol Oncol.* 2010; 116:88–91. [PubMed: 19853286]
38. Salvesen HB, et al. Integrated genomic profiling of endometrial carcinoma associates aggressive tumors with indicators of PI3 kinase activation. *Proc Natl Acad Sci U S A.* 2009; 106:4834–4839. [PubMed: 19261849]
39. Shoji K, et al. The oncogenic mutation in the pleckstrin homology domain of AKT1 in endometrial carcinomas. *Br J Cancer.* 2009; 101:145–148. [PubMed: 19491896]
40. Du W, et al. Suppression of p53 activity by Siva1. *Cell Death Differ.* 2009; 16:1493–1504. [PubMed: 19590512]
41. Wang X, et al. Siva1 inhibits p53 function by acting as an ARF E3 ubiquitin ligase. *Nat Commun.* 2013; 4:1551. [PubMed: 23462994]

42. Li N, et al. Siva1 suppresses epithelial-mesenchymal transition and metastasis of tumor cells by inhibiting stathmin and stabilizing microtubules. *Proc Natl Acad Sci U S A*. 2011; 108:12851–12856. [PubMed: 21768358]
43. Dun B, et al. Mycophenolic acid inhibits migration and invasion of gastric cancer cells via multiple molecular pathways. *PLoS One*. 2013; 8:e81702. [PubMed: 24260584]
44. Davis H, et al. FBXW7 mutations typically found in human cancers are distinct from null alleles and disrupt lung development. *J Pathol*. 2011; 224:180–189. [PubMed: 21503901]
45. Mutter GL, et al. Global expression changes of constitutive and hormonally regulated genes during endometrial neoplastic transformation. *Gynecol Oncol*. 2001; 83:177–185. [PubMed: 11606070]
46. Shi H, Zhang Z, Wang X, Liu S, Teng CT. Isolation and characterization of a gene encoding human Kruppel-like factor 5 (IKLF): binding to the CAAT/GT box of the mouse lactoferrin gene promoter. *Nucleic Acids Res*. 1999; 27:4807–4815. [PubMed: 10572182]
47. Simmen RC, et al. The emerging role of Kruppel-like factors in endocrine-responsive cancers of female reproductive tissues. *J Endocrinol*. 2010; 204:223–231. [PubMed: 19833720]
48. Nandan MO, et al. Kruppel-like factor 5 mediates cellular transformation during oncogenic KRAS-induced intestinal tumorigenesis. *Gastroenterology*. 2008; 134:120–130. [PubMed: 18054006]
49. Forbes SA, et al. The Catalogue of Somatic Mutations in Cancer (COSMIC). *Curr Protoc Hum Genet*. 2008; Chapter 10:Unit 10.11. [PubMed: 18428421]
50. Consortium EP, et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*. 2007; 447:799–816. [PubMed: 17571346]
51. Rao SS, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*. 2014; 159:1665–1680. [PubMed: 25497547]
52. Wang Y, Li X, Hu H. H3K4me2 reliably defines transcription factor binding regions in different cells. *Genomics*. 2014; 103:222–228. [PubMed: 24530516]
53. Cheng TH, et al. Meta-analysis of genome-wide association studies identifies common susceptibility polymorphisms for colorectal and endometrial cancer near SH2B3 and TSHZ1. *Sci Rep*. 2015; 5:17369. [PubMed: 26621817]
54. O'Mara TA, et al. Comprehensive genetic assessment of the ESR1 locus identifies a risk region for endometrial cancer. *Endocr Relat Cancer*. 2015; 22:851–861. [PubMed: 26330482]
55. Power C, Elliott J. Cohort profile: 1958 British birth cohort (National Child Development Study). *Int J Epidemiol*. 2006; 35:34–41. [PubMed: 16155052]
56. McEvoy M, et al. Cohort profile: The Hunter Community Study. *Int J Epidemiol*. 2010; 39:1452–1463. [PubMed: 20056765]
57. McGregor B, et al. Genetic and environmental contributions to size, color, shape, and other characteristics of melanocytic naevi in a sample of adolescent twins. *Genet Epidemiol*. 1999; 16:40–53. [PubMed: 9915566]
58. Teo YY, et al. A genotype calling algorithm for the Illumina BeadArray platform. *Bioinformatics*. 2007; 23:2741–2746. [PubMed: 17846035]
59. Price AL, et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*. 2006; 38:904–909. [PubMed: 16862161]
60. Purcell S, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007; 81:559–575. [PubMed: 17701901]
61. Aulchenko YS, Ripke S, Isaacs A, van Duijn CM. GenABEL: an R library for genome-wide association analysis. *Bioinformatics*. 2007; 23:1294–1296. [PubMed: 17384015]
62. Clayton D, Leung HT. An R package for analysis of whole-genome association studies. *Hum Hered*. 2007; 64:45–51. [PubMed: 17483596]
63. Howie B, Marchini J, Stephens M. Genotype imputation with thousands of genomes. *G3 (Bethesda)*. 2011; 1:457–470. [PubMed: 22384356]
64. Delaneau O, Zagury JF, Marchini J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat Methods*. 2013; 10:5–6. [PubMed: 23269371]
65. Timpson NJ, et al. A rare variant in APOC3 is associated with plasma triglyceride and VLDL levels in Europeans. *Nat Commun*. 2014; 5:4871. [PubMed: 25225788]

66. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet.* 2007; 39:906–913. [PubMed: 17572673]
67. Pruim RJ, et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics.* 2010; 26:2336–2337. [PubMed: 20634204]
68. Magi R, Morris AP. GWAMA: software for genome-wide association meta-analysis. *BMC Bioinformatics.* 2010; 11:288. [PubMed: 20509871]
69. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med.* 2002; 21:1539–1558. [PubMed: 12111919]
70. Huedo-Medina TB, Sanchez-Meca J, Marin-Martinez F, Botella J. Assessing heterogeneity in meta-analysis: Q statistic or I2 index? *Psychol Methods.* 2006; 11:193–206. [PubMed: 16784338]
71. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.* 2012; 40:D930–934. [PubMed: 22064851]
72. Boyle AP, et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.* 2012; 22:1790–1797. [PubMed: 22955989]
73. Hnisz D, et al. Super-enhancers in the control of cell identity and disease. *Cell.* 2013; 155:934–947. [PubMed: 24119843]
74. Corradin O, et al. Combinatorial effects of multiple enhancer variants in linkage disequilibrium dictate levels of gene expression to confer susceptibility to common traits. *Genome Res.* 2014; 24:1–13. [PubMed: 24196873]
75. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods.* 2001; 25:402–408. [PubMed: 11846609]
76. Giresi PG, Kim J, McDaniel RM, Iyer VR, Lieb JD. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Res.* 2007; 17:877–885. [PubMed: 17179217]
77. Ghossaini M, et al. Evidence that breast cancer risk at the 2q35 locus is mediated through IGFBP5 regulation. *Nat Commun.* 2014; 4:4999. [PubMed: 25248036]
78. Lewis A, et al. A polymorphic enhancer near GREM1 influences bowel cancer risk through differential CDX2 and TCF7L2 binding. *Cell Rep.* 2014; 8:983–990. [PubMed: 25131200]

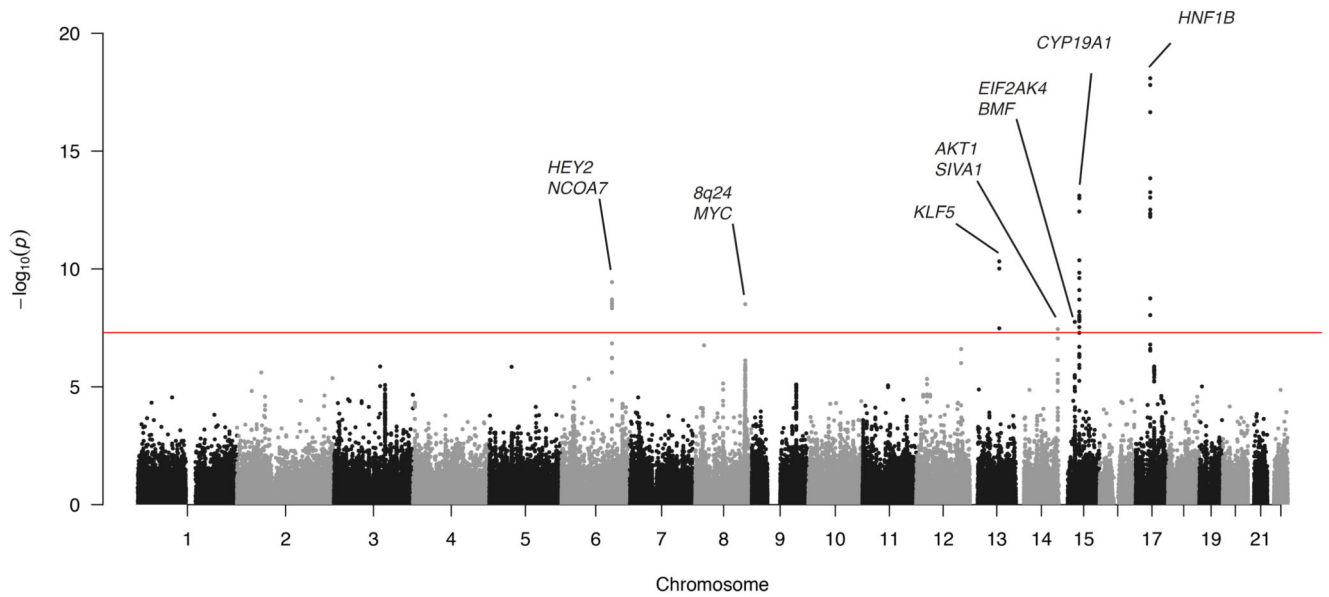


Figure 1. Endometrial cancer meta-analysis Manhattan plot

Manhattan plot of $-\log_{10}$ -transformed P-values from meta-analysis of 22 autosomes. There are seven loci surpassing genome-wide significance including two known loci: 15q21 (*CYP19A1*) and 17q12 (*HNF1B*) and five novel loci: 6q22 (*NCOA7*, *HEY2*), 8q24 (*MYC*), 13q22 (*KLF5*), 14q32 (*AKT1*, *SIVA1*), 15q15 (*EIF2AK4*, *BMF*).

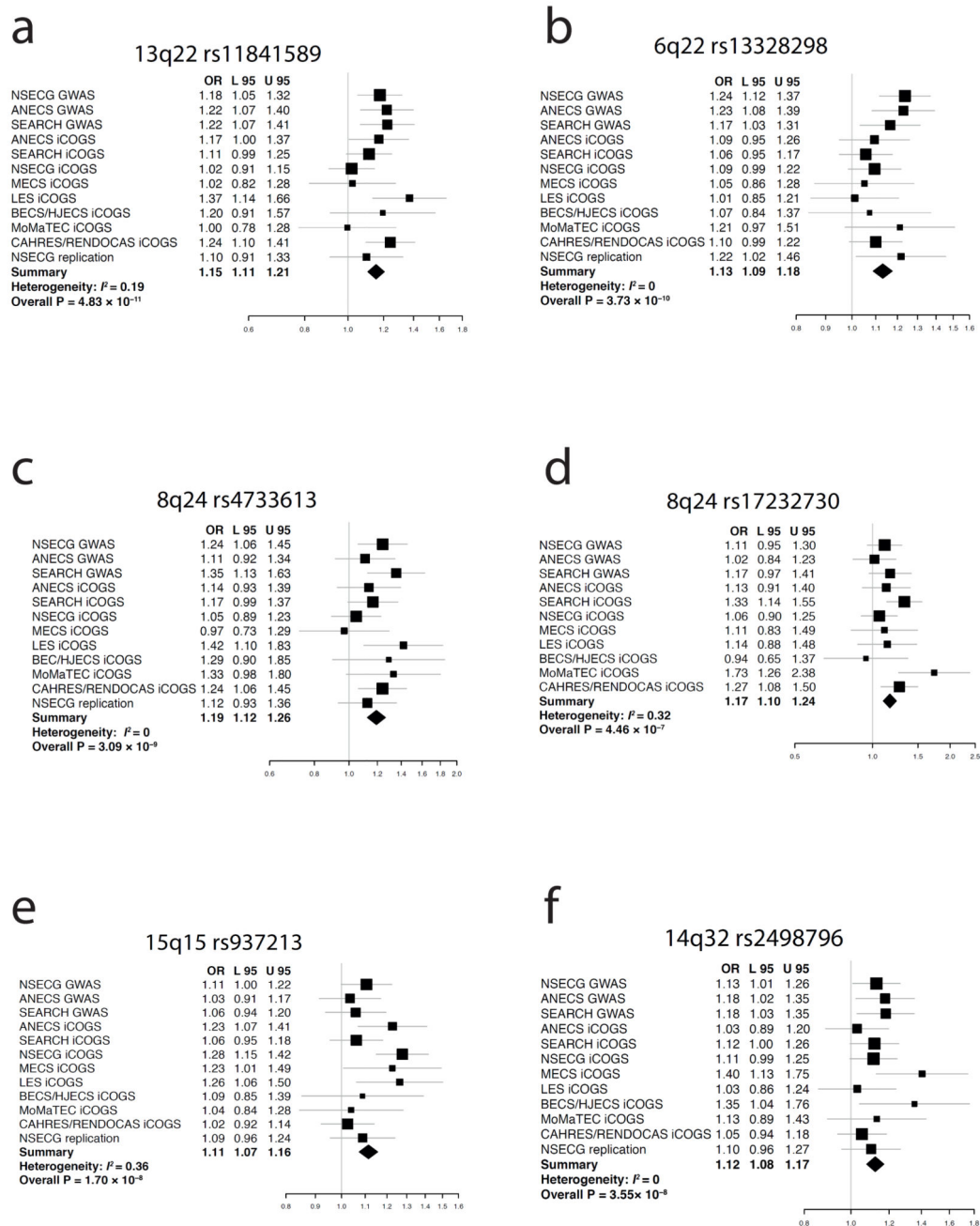


Figure 2. Forest plots of novel endometrial cancer risk loci

The odds ratio and 95% confidence intervals of each study of the meta-analysis are listed and shown in the adjacent plot. The I^2 heterogeneity scores (all <0.4) suggest that there is no marked difference in effects between studies. The SNPs represented are: a) rs11841589 (13q22), b) rs13328298 (6q22), c) rs4733613 (8q24), d) rs17232730 (8q24, pairwise $r^2 = 0.02$ with rs4733613), e) rs937213 (15q15) and f) rs2498796 (14q32).

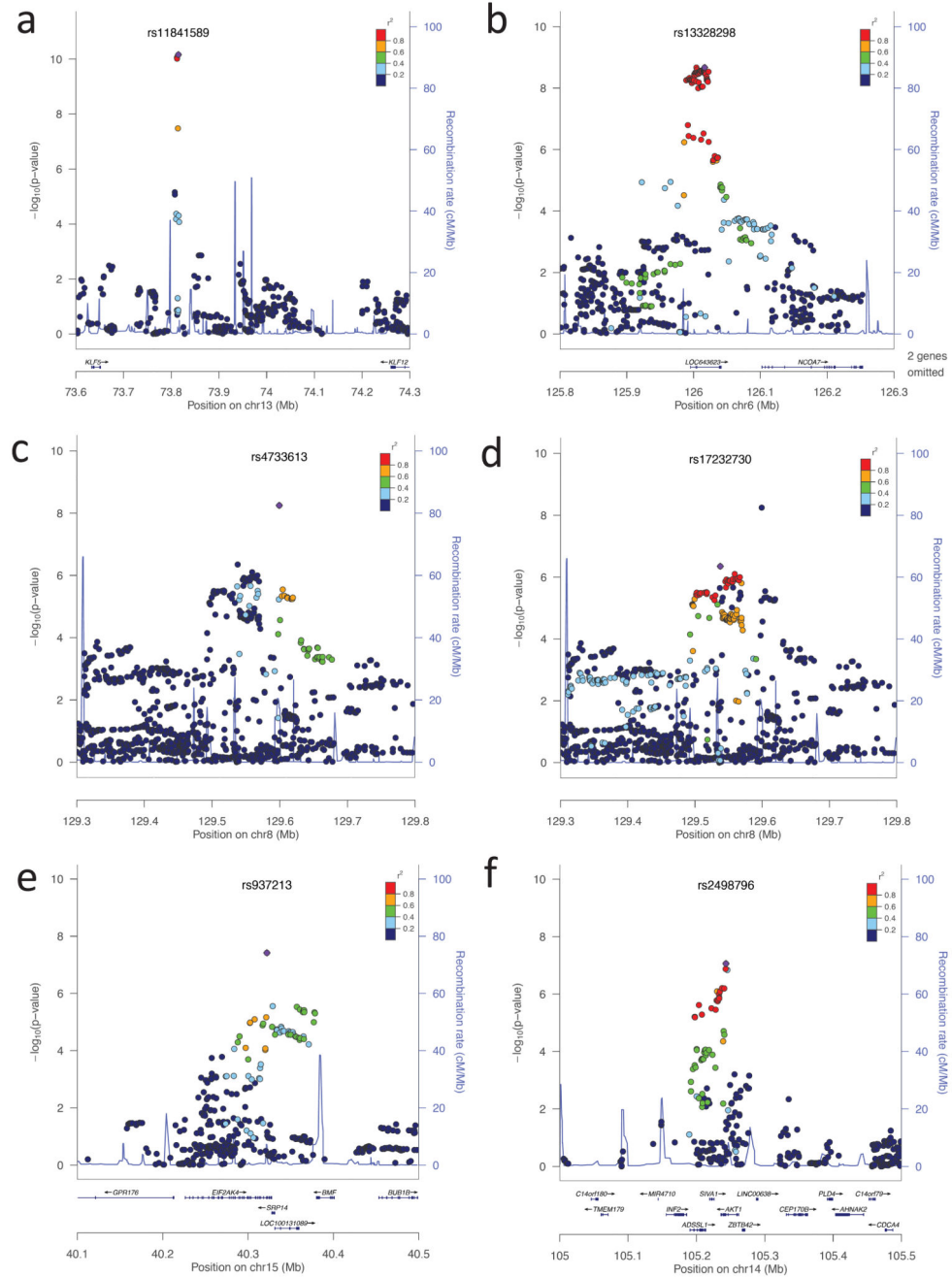


Figure 3. Regional association plots for the five novel loci associated with endometrial cancer. The $-\log_{10}$ P-values from the meta-analysis and regional imputation for three GWAS and eight iCOGS groups are shown for SNPs at: a) 13q22.1, b) 6q22, c) & d) 8q24, e) 15q15 and f) 14q32.33. The SNP with the lowest P-value at each locus is labeled and marked as a purple diamond, and the dot color represents the LD with the top SNP. The blue line shows recombination rates in cM/Mb. All plotted SNPs are either genotyped or have an IMPUTE info score of more than 0.9 in all datasets. Although genome-wide significant results for the 14q32.33 locus rely on imputed data, it should be noted that there is strong support from

nearby genotyped markers. Supplementary Figure 6 displays similar regional association plots with a larger number of SNPs using a less stringent info score cut-off.

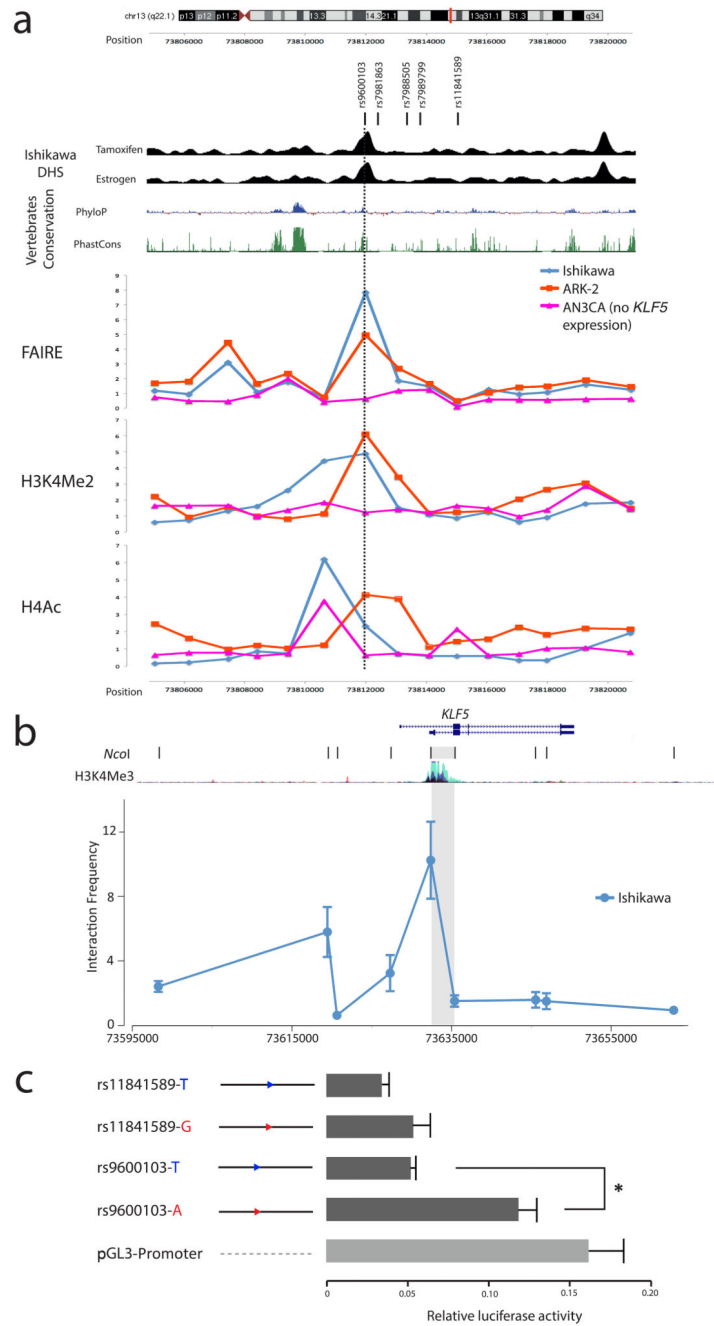


Figure 4. The 13q22.1 endometrial cancer susceptibility locus

a) Diagram showing the 16kb region around rs11841589, rs9600103 and correlated SNPs rs7981863, rs7988505 and rs7989799 (black marks), DNaseI hypersensitivity site (DHS) density signal in estrogen- and tamoxifen-treated ENCODE Ishikawa cells (Supplementary Note), and 100 vertebrates conservation. Vertical dotted line represents the position of rs9600103. FAIRE and ChIP assays for H3K4Me2 and H4Ac in endometrial cancer cell lines ARK-2 (rs9600103-TT), Ishikawa (rs9600103-AA) and AN3CA (rs9600103-AA) show evidence for enrichment of histone modifications.

b) 3C experiment for *KLF5*-expressing Ishikawa cells. Relative interaction frequencies between an *NcoI* restriction fragment containing risk SNPs rs9600103 and rs11841589 (bait fragment) and *NcoI* fragments across the *KLF5* promoter region, plotted against fragment position on chromosome 13. *NcoI* restriction sites are displayed below the schematic of *KLF5* transcripts. H3K4Me3 binding, indicative of promoters, from multiple ENCODE cell lines are also shown. The graph represents three biological replicates. Error bars represent standard deviation. A significant interaction was seen with the fragment containing a *KLF5* transcriptional start site (fragment shaded in grey).

c) Luciferase reporter assays to analyze the activity of 3kb fragments containing either rs9600103 or rs11841589 using the pGL3-Promoter vector in Ishikawa cells. Green arrows represent the low-risk alleles, and red arrows the high-risk alleles. Error bars represent the standard error of the mean (n=3). Luciferase activity for the rs9600103-A risk allele was more than double that of the rs9600103-T protective allele (P=0.018). There was no significant difference in luciferase activity between the rs11841589 alleles (Supplementary Table 7).

Table 1
Risk loci associated with endometrial cancer at $P < 5 \times 10^{-8}$ in the meta-analysis.

Locus	SNP	Position	Nearby gene(s)	EA	OA	EAF	All histologies			Endometrioid histology		
							Allelic OR (95%CI)	P	I^2	Allelic OR (95%CI)	P	I^2
Novel GWAS loci												
13q22.1	rs11841589	73,814,891	<i>KLF5, KLF12</i>	G	T	0.74	1.15 (1.11-1.21)	4.83×10^{-11}	0.19	1.16 (1.10-1.21)	6.01×10^{-10}	0.00
6q22.31	rs13328298	126,016,580	<i>HEY2, NCOA7</i>	G	A	0.58	1.13 (1.09-1.18)	3.73×10^{-10}	0.00	1.15 (1.11-1.20)	1.02×10^{-11}	0.00
8q24.21	rs4733613	129,599,278	<i>MYC</i>	G	C	0.87	0.84 (0.80-0.89)	3.09×10^{-9}	0.00	0.84 (0.79-0.89)	7.70×10^{-9}	0.09
15q15.1	rs937213	40,322,124	<i>EIF2AK, BMF</i>	T	C	0.58	0.90 (0.86-0.93)	1.77×10^{-8}	0.36	0.90 (0.86-0.94)	2.22×10^{-7}	0.30
14q32.33	rs2498796	105,243,220	<i>AKT1, SIVA1</i>	G	A	0.70	0.89 (0.85-0.93)	3.55×10^{-8}	0.00	0.88 (0.85-0.92)	4.22×10^{-8}	0.00
Previously reported GWAS loci												
17q12	rs11263763	36,103,565	<i>HNF1B</i>	A	G	0.54	1.20 (1.15-1.25)	2.78×10^{-19}	0.37	1.20 (1.15-1.25)	6.51×10^{-17}	0.52
15q21	rs2414098	51,537,806	<i>CYP19A1</i>	C	T	0.62	1.17 (1.13-1.23)	4.51×10^{-13}	0.00	1.18 (1.13-1.23)	2.48×10^{-13}	0.00

Positions in build 37; EA, Effect allele; OA, Other allele; EAF, effect allele frequency; I^2 , heterogeneity I^2 statistic. For all novel loci, the lead SNP was either directly genotyped or imputed with an information score of more than 0.9. *HNF1B* and *CYP19A1* have been previously reported by Painter *et al.* and Thompson *et al.*

Table 2
Conditional analysis of 8q24 locus showing two independent association signals.

SNP	Position	EA	OA	EAF	Pairwise r^2 with rs4733613	rs17232730	All histology meta-analysis		Conditioning on rs4733613		Conditioning on rs17232730	
							Allelic OR (95%CI)	P	Allelic OR (95%CI)	P	Allelic OR (95%CI)	P
rs4733613	129,599,278	G	C	0.87	-	0.02	0.84 (0.79-0.89)	5.64×10^{-9}	-	-	0.86 (0.81-0.91)	2.32×10^{-7}
rs17232730	129,537,746	G	C	0.88	0.02	-	1.17 (1.10-1.24)	4.46×10^{-7}	1.14 (1.08-1.22)	1.29×10^{-5}	-	-
rs10088218*	129,543,949	G	A	0.87	0.02	0.43	1.14 (1.07-1.20)	1.65×10^{-5}	1.12 (1.05-1.18)	2.92×10^{-4}	1.01 (0.91-1.12)	0.818

Positions in build 37; EA, Effect allele; OA, Other allele; EAF, effect allele frequency.

* rs10088218 is associated with ovarian cancer (all subtypes), with the association being more significant for cancers of serous histology. rs10088218-G is the risk allele for both endometrial cancer and ovarian cancer.