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

Timothy HT Cheng[1,57], Deborah J Thompson[2,57], Tracy A O'Mara[3], Jodie N Painter[3], Dylan M Glubb[3], Susanne Flach[1], Annabelle Lewis[1], Juliet D French[3], Luke Freeman-Mills[1], David Church[1], Maggie Gorman[1], Lynn Martin[1], National Study of Endometrial Cancer Genetics Group (NSECG)[1, 4], Shirley Hodgson[5], Penelope M Webb[3], The Australian National Endometrial Cancer Study Group (ANECS)[3, 4], John Attia[6, 7], Elizabeth G Holliday[6, 7], Mark McEvoy[7], Rodney J Scott[6, 8-10], Anjali K Henders[3], Nicholas G Martin[3], Grant W Montgomery[3], Dale R Nyholt[3, 11], Shahana Ahmed[12], Catherine S Healey[12], Mitul Shah[12], Joe Dennis[2], Peter A Fasching[13, 14], Matthias W Beckmann[14], Alexander Hein[14], Arif B Ekici[15], Per Hall[16], Kamila Czene[16], Hatef Darabi[16], Jingmei Li[16], Thilo Dörk[17], Matthias Dürst[18], Peter Hillemanns[19], Ingo Runnebaum[18], Frederic Amant[20], Stefanie Schrauwen[20], Hui Zhao[21, 22], Diether Lambrechts[21, 22], Jeroen Depreeuw[20-22], Sean C Dowdy[23], Ellen L Goode[24], Brooke L Fridley[25], Stacey J Winham[24], 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 Otton[28], Tony Proietto[28], Tao Liu[29], Miriam Mints[30], Emma Tham[29, 31], RENDOCAS[4, 29], CHIBCHA Consortium[1, 4], Mulin Jun Li[32], Shun H Yip[32], Junwen Wang[32], Manjeet K Bolla[2], Kyriaki Michailidou[2], Qin Wang[2], Jonathan P Tyrer[12], Malcolm Dunlop[33, 34], Richard Houlston[35], Claire Palles[1], John L Hopper[36], AOCS Group[3, 4, 37], Julian Peto[38], Anthony J Swerdlow[35, 39], Barbara Burwinkel[40, 41], Hermann Brenner[42-44], Alfons Meindl[45], Hiltrud Brauch[44, 46, 47], Annika Lindblom[29], Jenny Chang-Claude[48, 49], Fergus J Couch[24, 50], Graham G Giles[36, 51, 52], Vessela N Kristensen[53-55], Angela Cox[56], Julie M Cunningham[50], Paul D P Pharoah[12], Alison M Dunning[12], Stacey L Edwards[3], Douglas F Easton[2, 12], Ian Tomlinson[1], Amanda B Spurdle[3]

[1] Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK.
[2] Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK.
[3] Department of Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia.
[4] A list of members and affiliations appears in the Supplementary Note.
[5] Department of Clinical Genetics, St George's, University of London, London, UK.
[6] Hunter Medical Research Institute, John Hunter Hospital, Newcastle, NSW, Australia.
[7] Centre for Clinical Epidemiology and Biostatistics, School of Medicine and Public Health, University of Newcastle, NSW, Australia.
[8] Hunter Area Pathology Service, John Hunter Hospital, Newcastle, NSW, Australia.
[9] Centre for Information Based Medicine, University of Newcastle, NSW, Australia.
[10] School of Biomedical Sciences and Pharmacy, University of Newcastle, Newcastle, NSW, Australia.
[11] Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, Australia.
[12] Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK.
[13] University of California at Los Angeles, Department of Medicine, Division of Hematology/Oncology, David Geffen School of Medicine, Los Angeles, CA, USA.
[14] Department of Gynecology and Obstetrics, University Hospital Erlangen, Friedrich-Alexander University Erlangen-Nuremberg, Erlangen, Germany.
[15] Institute of Human Genetics, University Hospital Erlangen, Friedrich-Alexander-University Erlangen-Nuremberg, Erlangen, Germany.
[16] Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden.
[17] Hannover Medical School, Gynaecology Research Unit, Hannover, Germany.
[18] Department of Gynaecology, Jena University Hospital - Friedrich Schiller University, Jena, Germany.
[19] Hannover Medical School, Clinics of Gynaecology and Obstetrics, Hannover, Germany.
[20] Department of Obstetrics and Gynecology, Division of Gynecologic Oncology, University Hospitals, KU Leuven - University of Leuven, Belgium.
[21] Vesalius Research Center, VIB, Leuven, Belgium.

[22] Laboratory for Translational Genetics, Department of Oncology, University Hospitals Leuven, Leuven, Belgium.
[23] Department of Obstetrics and Gynecology, Division of Gynecologic Oncology, Mayo Clinic, Rochester, MN, USA.
[24] Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA.
[25] Department of Biostatistics, University of Kansas Medical Center, Kansas City, KS, USA.
[26] Centre for Cancerbiomarkers, Department of Clinical Science, The University of Bergen, Norway.
[27] Department of Obstetrics and Gynecology, Haukeland University Hospital, Bergen, Norway.
[28] School of Medicine and Public Health, University of Newcastle, Newcastle, NSW, Australia.
[29] Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden.
[30] Department of Women's and Children's Health, Karolinska Institutet, Karolinska University Hospital, Stockholm, Sweden.
[31] Clinical Genetics, Karolinska University Hospital Solna, Stockholm, Sweden.
[32] Centre for Genomic Sciences, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong SAR, China.
[33] Colon Cancer Genetics Group, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK.
[34] MRC Human Genetics Unit, Western General Hospital Edinburgh, Edinburgh, UK.
[35] Division of Genetics and Epidemiology, Institute of Cancer Research, London, UK.
[36] Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Vic, Australia.
[37] Peter MacCallum Cancer Center, The University of Melbourne, Melbourne, Australia.
[38] London School of Hygiene and Tropical Medicine, London, UK.
[39] Division of Breast Cancer Research, Institute of Cancer Research, London, UK.
[40] Molecular Biology of Breast Cancer, Department of Gynecology and Obstetrics, University of Heidelberg, Heidelberg, Germany.
[41] Molecular Epidemiology Group, German Cancer Research Center (DKFZ), Heidelberg, Germany.
[42] Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany.
[43] Division of Preventive Oncology, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), Heidelberg, Germany
[44] German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany
[45] Department of Obstetrics and Gynecology, Division of Tumor Genetics, Technical University of Munich, Munich, Germany.
[46] Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Stuttgart, Germany.
[47] University of Tübingen, Tübingen, Germany
[48] Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany.
[49] University Cancer Center Hamburg (UCCH), University Medical Center Hamburg-Eppendorf, Hamburg, Germany
[50] Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, USA.
[51] Cancer Epidemiology Centre, Cancer Council Victoria, Melbourne, Vic, Australia.
[52] Department of Epidemiology and Preventive Medicine, Monash University, Melbourne, Vic, Australia.
[53] Department of Genetics, Institute for Cancer Research, The Norwegian Radium Hospital, Oslo, Norway.
[54] The K.G. Jebsen Center for Breast Cancer Research, Institute for Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway .
[55] Department of Clinical Molecular Oncology, Division of Medicine, Akershus University Hospital, Lørenskog, Norway.
[56] Sheffield Cancer Research, Department of Oncology, University of Sheffield, Sheffield, UK.
[57] These authors contributed equally to this work.

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).

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.

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* 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 States[1] and Europe[2], and the most common cancer of the female reproductive system. The familial relative risk is ~2[3,4], but highly penetrant germline mutations in mismatch repair genes[5], 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 1[8] and rs727479 in *CYP19A1* intron 4[9].

To identify additional endometrial cancer risk loci, we re-analysed data from our previous GWAS (ANECS, SEARCH datasets[10]) 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 Cohort[13]. 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 (λ_{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 ($P<5\times10^{-8}$, **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 tissue[20], and endometrial cancer tumor and adjacent normal tissue[21] 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, $P=3.73\times10^{-10}$) 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 cancers[22]. *NCOA7* modulates the activity of the estrogen receptor via direct binding[23].

The second locus (rs4733613, OR=0.84, 95%CI:0.80–0.89, $P=3.09\times10^{-9}$) is at 8q24.21. Stepwise conditional logistic regression identified another independent signal in this region, rs17232730 (pairwise r^2=0.02, $P_{cond}=1.29\times10^{-5}$, **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], ovary[28] 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 ($r^2\leq0.02$) with any other cancer SNP in the region (**Supplementary Figure 3**). A role in tumorigenesis is implicated for several miRNAs in the region[31]. Of these, miR-1207-5p is reported to repress *TERT*, a locus also implicated in endometrial cancer risk[32].

The lead SNP at 15q15 (rs937213; OR=0.90, 95%CI:0.86–0.93, $P=1.77\times10^{-8}$) lies within an intron of *EIF2AK4. EIF2AK4* encodes a kinase that phosphorylates EIF2α and downregulates protein synthesis during cellular stress[33]. Another nearby gene, *BMF,* encodes an apoptotic regulator moderately to highly expressed in glandular endometrial tissue[34].

At 14q42, the lead SNP rs2498796 (OR=0.89, 95%CI:0.85–0.93, $P=3.55\times10^{-8}$) lies in intron 3 of oncogene *AKT1,* which is highly expressed in the endometrium[34]. 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 proliferation[35] and is activated in endometrial tumors[36], 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 cells[42]. *INF2* expression is reported to act as a promigratory signal in gastric cancer cells treated with mycophenolic acid[43].

The final novel endometrial cancer SNP was rs11841589 (OR=1.15, 95%CI:1.11–1.21, $P=4.83\times10^{-11}$) 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, homoeostasis and tumorigenesis[44-47]. Elevated *KLF5* levels are strongly correlated with activating *KRAS* mutations [48] and KLF5 is targeted for degradation by the tumor suppressor FBXW7. Both *FBXW7* and *KRAS* are commonly mutated in endometrial cancer [49]. rs11841589 was one of a group of five highly correlated SNPs ($r^2 \geq 0.98$) surpassing genome-wide significance in a 3kb LD block bounded by rs9600103 ($P=8.70\times10^{-11}$) and rs11841589 (**Figure 4a)**. There was no residual association signal at this locus (P_{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 ENCODE[50] Ishikawa cells (**Figure 4a**). In addition, in a Hi-C chromatin capture experiment in Hela S3 cells[51], 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 binding[52]) 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 *Nco*I 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 ($P \leq 0.014$). 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/

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 (NSECG), 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.

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/A 10710, 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.

ANECS 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.

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 NSECG GWAS; I.T., L.M., M.G., and S.H. co-ordinated the National Study of the Genetics of Endometrial Cancer (NSECG), 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^{c}E., 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.

References

1. Siegel, R., Ma, J., Zou, Z. & Jemal, A. Cancer statistics, 2014. *CA Cancer J Clin* **64**, 9-29 (2014).
2. Ferlay, J. *et al.* Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012. *Eur J Cancer* **49**, 1374-1403 (2013).
3. Gruber, S.B. & Thompson, W.D. A population-based study of endometrial cancer and familial risk in younger women. Cancer and Steroid Hormone Study Group. *Cancer Epidemiol Biomarkers Prev* **5**, 411-417 (1996).
4. Win, A.K., Reece, J.C. & Ryan, S. Family history and risk of endometrial cancer: a systematic review and meta-analysis. *Obstet Gynecol* **125**, 89-98 (2015).
5. Barrow, E., Hill, J. & Evans, D.G. Cancer risk in Lynch Syndrome. *Fam Cancer* **12**, 229-240 (2013).
6. Church, D.N. *et al.* DNA polymerase epsilon and delta exonuclease domain mutations in endometrial cancer. *Hum Mol Genet* **22**, 2820-2828 (2013).
7. Palles, C. *et al.* Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. *Nat Genet* **45**, 136-144 (2013).
8. Painter, J.N. *et al.* Fine-mapping of the HNF1B multicancer locus identifies candidate variants that mediate endometrial cancer risk. *Hum Mol Genet* **24**, 1478-1492 (2015).
9. Thompson, D.J. *et al.* CYP19A1 fine-mapping and Mendelian randomization: estradiol is causal for endometrial cancer. *Endocr Relat Cancer* **23**, 77-91 (2016).
10. Spurdle, A.B. *et al.* Genome-wide association study identifies a common variant associated with risk of endometrial cancer. *Nat Genet* **43**, 451-454 (2011).
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* **40**, 631-637 (2008).
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* **39**, 984-988 (2007).
13. Wellcome Trust Case Control, C. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661-678 (2007).
14. Eeles, R.A. *et al.* Identification of 23 new prostate cancer susceptibility loci using the iCOGS custom genotyping array. *Nat Genet* **45**, 385-391, 391e381-382 (2013).
15. Michailidou, K. *et al.* Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat Genet* **45**, 353-361, 361e351-352 (2013).
16. Pharoah, P.D. *et al.* GWAS meta-analysis and replication identifies three new susceptibility loci for ovarian cancer. *Nat Genet* **45**, 362-370, 370e361-362 (2013).
17. Sakoda, L.C., Jorgenson, E. & Witte, J.S. Turning of COGS moves forward findings for hormonally mediated cancers. *Nat Genet* **45**, 345-348 (2013).
18. De Vivo, I. *et al.* Genome-wide association study of endometrial cancer in E2C2. *Hum Genet* **133**, 211-224 (2014).
19. Setiawan, V.W. *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* **18**, 242-247 (2009).
20. Consortium, G.T. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* **45**, 580-585 (2013).
21. Cancer Genome Atlas Research, N. *et al.* Integrated genomic characterization of endometrial carcinoma. *Nature* **497**, 67-73 (2013).
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* **31**, 461-466 (2007).
23. Shao, W., Halachmi, S. & Brown, M. ERAP140, a conserved tissue-specific nuclear receptor coactivator. *Mol Cell Biol* **22**, 3358-3372 (2002).

24. Kiemeney, L.A. *et al.* Sequence variant on 8q24 confers susceptibility to urinary bladder cancer. *Nat Genet* **40**, 1307-1312 (2008).
25. Rothman, N. *et al.* A multi-stage genome-wide association study of bladder cancer identifies multiple susceptibility loci. *Nat Genet* **42**, 978-984 (2010).
26. Easton, D.F. *et al.* Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* **447**, 1087-1093 (2007).
27. Whiffin, N. *et al.* Identification of susceptibility loci for colorectal cancer in a genome-wide meta-analysis. *Hum Mol Genet* **23**, 4729-4737 (2014).
28. Goode, E.L. *et al.* A genome-wide association study identifies susceptibility loci for ovarian cancer at 2q31 and 8q24. *Nat Genet* **42**, 874-879 (2010).
29. Eeles, R.A. *et al.* Identification of seven new prostate cancer susceptibility loci through a genome-wide association study. *Nat Genet* **41**, 1116-1121 (2009).
30. Gudmundsson, J. *et al.* Genome-wide association and replication studies identify four variants associated with prostate cancer susceptibility. *Nat Genet* **41**, 1122-1126 (2009).
31. Huppi, K., Pitt, J.J., Wahlberg, B.M. & Caplen, N.J. The 8q24 gene desert: an oasis of non-coding transcriptional activity. *Front Genet* **3**, 69 (2012).
32. Carvajal-Carmona, L.G. *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* **134**, 231-245 (2015).
33. Berlanga, J.J., Santoyo, J. & De Haro, C. Characterization of a mammalian homolog of the GCN2 eukaryotic initiation factor 2alpha kinase. *Eur J Biochem* **265**, 754-762 (1999).
34. Uhlen, M. *et al.* Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).
35. Cantley, L.C. The phosphoinositide 3-kinase pathway. *Science* **296**, 1655-1657 (2002).
36. Slomovitz, B.M. & Coleman, R.L. The PI3K/AKT/mTOR pathway as a therapeutic target in endometrial cancer. *Clin Cancer Res* **18**, 5856-5864 (2012).
37. Cohen, Y. *et al.* AKT1 pleckstrin homology domain E17K activating mutation in endometrial carcinoma. *Gynecol Oncol* **116**, 88-91 (2010).
38. Salvesen, H.B. *et al.* Integrated genomic profiling of endometrial carcinoma associates aggressive tumors with indicators of PI3 kinase activation. *Proc Natl Acad Sci U S A* **106**, 4834-4839 (2009).
39. Shoji, K. *et al.* The oncogenic mutation in the pleckstrin homology domain of AKT1 in endometrial carcinomas. *Br J Cancer* **101**, 145-148 (2009).
40. Du, W. *et al.* Suppression of p53 activity by Siva1. *Cell Death Differ* **16**, 1493-1504 (2009).
41. Wang, X. *et al.* Siva1 inhibits p53 function by acting as an ARF E3 ubiquitin ligase. *Nat Commun* **4**, 1551 (2013).
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* **108**, 12851-12856 (2011).
43. Dun, B. *et al.* Mycophenolic acid inhibits migration and invasion of gastric cancer cells via multiple molecular pathways. *PLoS One* **8**, e81702 (2013).
44. Davis, H. *et al.* FBXW7 mutations typically found in human cancers are distinct from null alleles and disrupt lung development. *J Pathol* **224**, 180-189 (2011).
45. Mutter, G.L. *et al.* Global expression changes of constitutive and hormonally regulated genes during endometrial neoplastic transformation. *Gynecol Oncol* **83**, 177-185 (2001).
46. Shi, H., Zhang, Z., Wang, X., Liu, S. & Teng, C.T. 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* **27**, 4807-4815 (1999).

47. Simmen, R.C. *et al.* The emerging role of Kruppel-like factors in endocrine-responsive cancers of female reproductive tissues. *J Endocrinol* **204**, 223-231 (2010).
48. Nandan, M.O. *et al.* Kruppel-like factor 5 mediates cellular transformation during oncogenic KRAS-induced intestinal tumorigenesis. *Gastroenterology* **134**, 120-130 (2008).
49. Forbes, S.A. *et al.* The Catalogue of Somatic Mutations in Cancer (COSMIC). *Curr Protoc Hum Genet* **Chapter 10**, Unit 10 11 (2008).
50. Consortium, E.P. *et al.* Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799-816 (2007).
51. Rao, S.S. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665-1680 (2014).
52. Wang, Y., Li, X. & Hu, H. H3K4me2 reliably defines transcription factor binding regions in different cells. *Genomics* **103**, 222-228 (2014).
53. Cheng, T.H. *et al.* Meta-analysis of genome-wide association studies identifies common susceptibility polymorphisms for colorectal and endometrial cancer near SH2B3 and TSHZ1. *Sci Rep* **5**, 17369 (2015).
54. O'Mara, T.A. *et al.* Comprehensive genetic assessment of the ESR1 locus identifies a risk region for endometrial cancer. *Endocr Relat Cancer* **22**, 851-861 (2015).

Figure legends

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 *Nco*I restriction fragment containing risk SNPs rs9600103 and rs11841589 (bait

fragment) and *Nco*I fragments across the *KLF5* promoter region, plotted against fragment position on chromosome 13. *Nco*I 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 $P< 5\times10^{-8}$ 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	*KLF5, KLF12*	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	*HEY2, NCOA7*	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	*MYC*	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	*EIF2AK, BMF*	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	*AKT1, SIVA1*	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	*HNF1B*	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	*CYP19A1*	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[55]. 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.*[8] and Thompson *et al*[9].

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

SNP	Position	EA	OA	EAF	Pairwise r^2 with		All histology meta-analysis		Conditioning on rs4733613		Conditioning on rs17232730	
					rs4733613	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.

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/CORGI[12] and SP1[11] colorectal studies genotypedusing Illumina Hap550, Hap300 and Hap240S arrays, and. 1958 Birth Cohort[55] 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 arrays [10]. Compared with our previous study [10], 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 controls [13]. ANECS cases were compared to controls recruited as part of the Hunter Community Study[56] or Brisbane Adolescent Twin Study[57], 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 Illuminia 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 prostate[14] studies, and also the 1,483 top SNPs from our previous EC GWAS[10] 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 Illumnus[58]. 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 Eigenstrat[59] software. Analysis was conducted using PLINK[60], 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 BioMark^{TM} 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 v2[63] and 1000 Genomes project (2012 release) as reference panel. For the first-pass genome-wide analysis we pre-phased chromosomes using SHAPEIT[64] 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≥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≥1% and info≥0.4 criteria, of which 5,437,135 had info≥0.7 and 2,333,040 had info≥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 accuracy[65]. All SNPs reported in this study had an info score ≥0.9 in all datasets.

Association testing

Association testing was done using SNPTEST v2[66] 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 >500kb 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 $P_{cond}<10^{-4}$. Regional association plots (**Figure 1, Supplementary Figure 6**) were created using LocusZoom[67].

Meta-analysis

Inverse variance, fixed effects meta-analysis of the 11 groups (three GWAS, eight iCOGS groups) was conducted using GWAMA[68]. 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 HaploregV2[71], RegulomeDB[72] and data from ENCODE[50] 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 ΔΔCT method described by Livak and Schmittgen[75], 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 al[76]. Briefly, cross-linking was done on a rocker at room temperature. 1% formaldehyde was added to ~10^8 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, thenspun 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 ΔΔCt method[31] 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 10^8 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 ΔΔCt 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 *Nco*I fragments spanning 76 kb of the *KLF5* promoter region. BAC clones (RP11-81D9, RP11-179I20) covering the region were digested with *Nco*I, ligated with T4 ligase and used 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 ΔΔCt 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 activity[78]. 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).

Online Methods-Only References

55. Power, C. & Elliott, J. Cohort profile: 1958 British birth cohort (National Child Development Study). *Int J Epidemiol* **35**, 34-41 (2006).
56. McEvoy, M. *et al.* Cohort profile: The Hunter Community Study. *Int J Epidemiol* **39**, 1452-1463 (2010).
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* **16**, 40-53 (1999).
58. Teo, Y.Y. *et al.* A genotype calling algorithm for the Illumina BeadArray platform. *Bioinformatics* **23**, 2741-2746 (2007).
59. Price, A.L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**, 904-909 (2006).
60. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-575 (2007).
61. Aulchenko, Y.S., Ripke, S., Isaacs, A. & van Duijn, C.M. GenABEL: an R library for genome-wide association analysis. *Bioinformatics* **23**, 1294-1296 (2007).
62. Clayton, D. & Leung, H.T. An R package for analysis of whole-genome association studies. *Hum Hered* **64**, 45-51 (2007).
63. Howie, B., Marchini, J. & Stephens, M. Genotype imputation with thousands of genomes. *G3 (Bethesda)* **1**, 457-470 (2011).
64. Delaneau, O., Zagury, J.F. & Marchini, J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat Methods* **10**, 5-6 (2013).
65. Timpson, N.J. *et al.* A rare variant in APOC3 is associated with plasma triglyceride and VLDL levels in Europeans. *Nat Commun* **5**, 4871 (2014).
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* **39**, 906-913 (2007).
67. Pruim, R.J. *et al.* LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26**, 2336-2337 (2010).
68. Magi, R. & Morris, A.P. GWAMA: software for genome-wide association meta-analysis. *BMC Bioinformatics* **11**, 288 (2010).
69. Higgins, J.P. & Thompson, S.G. Quantifying heterogeneity in a meta-analysis. *Stat Med* **21**, 1539-1558 (2002).
70. Huedo-Medina, T.B., Sanchez-Meca, J., Marin-Martinez, F. & Botella, J. Assessing heterogeneity in meta-analysis: Q statistic or I2 index? *Psychol Methods* **11**, 193-206 (2006).
71. Ward, L.D. & Kellis, M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res* **40**, D930-934 (2012).
72. Boyle, A.P. *et al.* Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res* **22**, 1790-1797 (2012).
73. Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934-947 (2013).
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* **24**, 1-13 (2014).
75. Livak, K.J. & Schmittgen, T.D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**, 402-408 (2001).
76. Giresi, P.G., Kim, J., McDaniell, R.M., Iyer, V.R. & Lieb, J.D. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Res* **17**, 877-885 (2007).
77. Ghoussaini, M. *et al.* Evidence that breast cancer risk at the 2q35 locus is mediated through IGFBP5 regulation. *Nat Commun* **4**, 4999 (2014).

78. Lewis, A. *et al.* A polymorphic enhancer near GREM1 influences bowel cancer risk through differential CDX2 and TCF7L2 binding. *Cell Rep* **8**, 983-990 (2014).

SUPPLEMENTARY NOTE

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

Timothy HT Cheng[1,57], Deborah J Thompson[2,57], Tracy A O'Mara[3], Jodie N Painter[3], Dylan M Glubb[3], Susanne Flach[1], Annabelle Lewis[1], Juliet D French[3], Luke Freeman-Mills[1], David Church[1], Maggie Gorman[1], Lynn Martin[1], National Study of Endometrial Cancer Genetics Group (NSECG)[1, 4], Shirley Hodgson[5], Penelope M Webb[3], The Australian National Endometrial Cancer Study Group (ANECS)[3, 4], John Attia[6, 7], Elizabeth G Holliday[6, 7], Mark McEvoy[7], Rodney J Scott[6, 8-10], Anjali K Henders[3], Nicholas G Martin[3], Grant W Montgomery[3], Dale R Nyholt[3, 11], Shahana Ahmed[12], Catherine S Healey[12], Mitul Shah[12], Joe Dennis[2], Peter A Fasching[13, 14], Matthias W Beckmann[14], Alexander Hein[14], Arif B Ekici[15], Per Hall[16], Kamila Czene[16], Hatef Darabi[16], Jingmei Li[16], Thilo Dörk[17], Matthias Dürst[18], Peter Hillemanns[19], Ingo Runnebaum[18], Frederic Amant[20], Stefanie Schrauwen[20], Hui Zhao[21, 22], Diether Lambrechts[21, 22], Jeroen Depreeuw[20-22], Sean C Dowdy[23], Ellen L Goode[24], Brooke L Fridley[25], Stacey J Winham[24], 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 Otton[28], Tony Proietto[28], Tao Liu[29], Miriam Mints[30], Emma Tham[29, 31], RENDOCAS[4, 29], CHIBCHA Consortium[1, 4], Mulin Jun Li[32], Shun H Yip[32], Junwen Wang[32], Manjeet K Bolla[2], Kyriaki Michailidou[2], Qin Wang[2], Jonathan P Tyrer[12], Malcolm Dunlop[33, 34], Richard Houlston[35], Claire Palles[1], John L Hopper[36], AOCS Group[3, 4, 37], Julian Peto[38], Anthony J Swerdlow[35, 39], Barbara Burwinkel[40, 41], Hermann Brenner[42-44], Alfons Meindl[45], Hiltrud Brauch[44, 46, 47], Annika Lindblom[29], Jenny Chang-Claude[48, 49], Fergus J Couch[24, 50], Graham G Giles[36, 51, 52], Vessela N Kristensen[53-55], Angela Cox[56], Julie M Cunningham[50], Paul D P Pharoah[12], Alison M Dunning[12], Stacey L Edwards[3], Douglas F Easton[2, 12], Ian Tomlinson[1], Amanda B Spurdle[3]

[1] Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK.
[2] Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK.
[3] Department of Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia.
[4] A list of members and affiliations appears in the Supplementary Note.
[5] Department of Clinical Genetics, St George's, University of London, London, UK.
[6] Hunter Medical Research Institute, John Hunter Hospital, Newcastle, NSW, Australia.
[7] Centre for Clinical Epidemiology and Biostatistics, School of Medicine and Public Health, University of Newcastle, NSW, Australia.
[8] Hunter Area Pathology Service, John Hunter Hospital, Newcastle, NSW, Australia.
[9] Centre for Information Based Medicine, University of Newcastle, NSW, Australia.
[10] School of Biomedical Sciences and Pharmacy, University of Newcastle, Newcastle, NSW, Australia.
[11] Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, Australia.
[12] Centre for Cancer Genetic Epidemiology, Dept of Oncology, University of Cambridge, Cambridge, UK.
[13] University of California at Los Angeles, Department of Medicine, Division of Hematology/Oncology, David Geffen School of Medicine, Los Angeles, CA, USA.
[14] Department of Gynecology and Obstetrics, University Hospital Erlangen, Friedrich-Alexander University Erlangen-Nuremberg, Erlangen, Germany.
[15] Institute of Human Genetics, University Hospital Erlangen, Friedrich-Alexander-University Erlangen-Nuremberg, Erlangen, Germany.
[16] Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden.
[17] Hannover Medical School, Gynaecology Research Unit, Hannover, Germany.
[18] Department of Gynaecology, Jena University Hospital - Friedrich Schiller University, Jena, Germany.
[19] Hannover Medical School, Clinics of Gynaecology and Obstetrics, Hannover, Germany.
[20] Department of Obstetrics and Gynecology, Division of Gynecologic Oncology, University Hospitals, KU Leuven - University of Leuven, Belgium.
[21] Vesalius Research Center, VIB, Leuven, Belgium.

Detailed Description of the Case and Control Sample Sets

A summary of the studies included in the GWAS and both additional genotyping phases is shown in **Supplementary Table 1**, with additional details provided below. **Supplementary Figure 1** provides a flow diagram of the overall study design. All studies were of women of European ancestry. All studies have the relevant IRB approval in each country in accordance with the principles embodied in the Declaration of Helsinki, and informed consent was obtained from all participants. A total of 7,737 cases and 37,144 controls were included in this analysis. Cases and controls were matched based on geographical location and case-control clustering in principal components analysis (PCA) (**Supplementary Figure 1**).

Endometrial cancer case and control GWAS Sample Sets:

Quality control (QC) was applied to all GWAS sets, following standard QC approaches detailed in Spurdle et al[1]. Also see online methods.

NSECG

National Study of the Genetics of Endometrial Cancer (NSECG) cases were identified from collaborating clinicians throughout the UK from 2008 to 2013, taking care not to recruit from centres involved in SEARCH. Inclusion criteria were adenocarcinomas of the uterus presenting at 70 years of age or younger. Almost all cases were incident and sampled within 6 months of diagnosis. Peripheral blood was collected from each participant and DNA extracted using standard methods and the participants completed the associated questionnaire. Tumour histology was confirmed from routine hospital reports and further details of histopathology and other tumor pathology characteristic was abstracted from these clinical pathology reports. 925 samples were genotyped using the Ilumina 660W Quads in the GWAS scan, 965 samples were genotyped in the phase 1 additional genotyping using iCOGS arrays, and a further 1195 were genotyped using KASPar and Fluidigm genotyping for the second phase. There was no overlap in samples used and all cases were of European ancestry.

ANECS

The Australian National Endometrial Cancer Study (ANECS) is an Australian population-based case-control family study of cancer of the uterine corpus[2]. Women aged 18-79, newly diagnosed with histologically confirmed primary cancer of the endometrium between July 2005 and December 2007 were identified through major hospitals nationally, and also from state-based cancer registries. Excluding women who could not be contacted (mostly due to death, illness or failure to contact), case participation rate was 63%. Participants completed a detailed questionnaire providing clinical and epidemiological information, including ethnicity of all four grandparents. Information on tumor pathology characteristics was abstracted in standardized format from clinical pathology reports for all patients. 606 ANECS samples all of endometrioid-only histology were used for the original endometrial cancer case-control GWAS and a further 538 were genotyped using iCOGS for the first additional genotyping phase.

SEARCH

The Studies of Epidemiology and Risk factors in Cancer Heredity (SEARCH) is an ongoing population-based study with cases ascertained through the Eastern Cancer Registration and Information Centre (http://www.ecric.org.uk). All women diagnosed with endometrial cancer between the ages of 18-69 years (average age of diagnosis 58 years) from August 2001 to September 2007 were eligible for inclusion. Approximately 54% of eligible patients have enrolled in the study. Women taking part in the study were asked to provide a 20ml blood sample for DNA analysis, and to complete a comprehensive epidemiological questionnaire. Controls were also drawn from SEARCH (http://ccge.medschl.cam.ac.uk/search/), but had no prior history of cancer at the time of recruitment. They were female, also between the ages of 18-69 at the time of recruitment and matched to cases in geographical profile. Approximately 35% of eligible controls enrolled in the study. All participants reported Caucasian ethnicity. Information on tumor pathology characteristics was provided by the

Eastern Cancer Registration and Information Centre and was derived from clinical pathology reports for all patients. 681 SEARCH samples with endometroid-only histology were used in the original GWAS and a further 773 non-overlapping cases were used in the iCOGS analysis.

UK1/CORGI

The UK1 Colorectal Tumour Gene Identification (CoRGI) is a GWAS for colorectal neoplasia[3]. The 894 controls matched with the NSECG cases were spouses or partners unaffected by cancer and without a personal family history (to second degree relative level) of colorectal neoplasia. Known dominant polyposis syndromes, HNPCC/Lynch syndrome or bi-allelic MUTYH mutation carriers were excluded. All cases and controls were of white UK ethnic origin. Genotyping was done on the Illumina Hap550 arrays.

Scotland Phase 1

Scotland Phase1 is a colorectal cancer GWAS[4] with 1012 cancer-free population controls Known dominant polyposis syndromes, HNPCC/Lynch syndrome or bi-allelic MUTYH mutation carriers were excluded. Control subjects were sampled from the Scottish population NHS registers, matched by age (±5 years), gender and area of residence within Scotland. A subset of 392 controls from this dataset were matched with the NSECG GWAS cases and these were chosen based on case-control clustering on PCA. Genotyping was done on the Illumina Hap300 and Hap 240S arrays.

QIMR

The Queensland Institute of Medical Research (QIMR) control sample is a subsection of subjects recruited as part of the Brisbane Adolescent Twin Study[5,6]. Twins were recruited from schools in Brisbane, Australia and surrounding areas of southeast Queensland and were examined close to their 12th birthday. Blood was obtained from all twins and most parents. Parents were asked the ancestry of all eight great-grandparents of the twins. More than 95% of great-grandparents were identified as being of northern European ancestry, mainly from Britain and Ireland. This analysis used genotype data from parents and siblings only, extracted from an existing Illumina 610K BeadChip genome-wide association scan[7] and recalled using the Illuminus algorithm. After QC, 1846 QIMR controls were available for inclusion in the analysis.

HCS

The Hunter Community Study (HCS) is a population-based cohort study consisting of men and women aged 55-85 years of age in Newcastle, New South Wales, Australia[8]. Participants were randomly selected from the NSW State electoral roll (listing on the electoral roll is compulsory in Australia) and contacted between December 2004 and December 2007. Non-English speaking persons and those living in a residential aged-care facility were ineligible for participation in the study. Participants were asked to complete five self-report questionnaires as well as attend the HCS data collection centre so clinical measures could be obtained. In total, 44.5% of eligible controls agreed to participate in this study. Genotype data for this study were extracted from an existing Illumina 610K BeadChip genome-wide association study scan and recalled using the Illuminus algorithm. After QC, 1237 HCS controls were available for inclusion in the analysis.

WTCCC

Controls utilized were genotyped as part of the Wellcome Trust Case Control Consortium (WTCCC2)[9]. These controls are drawn from two sources: 2,674 controls from the 1958 Birth Cohort (1958BC), a population-based study in the United Kingdom of individuals born in 1 week in 1958[10]; and 2,501 controls identified through the UK National Blood Service (NBS)[9]. 1958BC controls were matched with NSECG cases and the NBS controls were matched with SEARCH cases.

Phase 1 additional genotyping - iCOGS Case Sample Sets:

All samples in this phase were genotyped as part of the Collaborative Oncological Gene-environment Study (iCOGS) initiative on a custom Ilumina Infinium iSelect array. Cases from ANECS and SEARCH and NSECG were recruited as detailed above, and are non-overlapping.

BECS
The Bavarian Endometrial Cancer Cases and Controls Study (BECS) is a single-center case-control study, conducted between 2002 and 2008, with the aim of investigating genetic and epidemiological risk factors for endometrial cancer. Cases were either incident cases referred to the University Hospital Erlangen by surrounding practitioners (66% of the case sample set), or prevalent cases that were outpatients in follow-up care approached within 6.2 (±4.6 SD) years after treatment for primary endometrial cancer in the same hospital (34% of the case sample set). Epidemiological information was collected by a structured questionnaire completed during an interview and clinical data for the cases was obtained from clinical health records.

CAHRES
Details of the population selection process have been published previously for the Cancer Hormone Replacement Epidemiology Study (CAHRES)[11]. Formerly known as the Singapore and Sweden Breast/Endometrial Cancer Study (SASBAC), this population based case-control study was conducted among Swedish women aged 50-74 years, who were residing in Sweden between January 1st 1994 and December 31st 1995. Endometrial cancer cases were identified through the nation-wide cancer registries in Sweden. All participants provided detailed questionnaire information. For endometrial cancer, histological specimens were reviewed and re-classified by the study pathologist. All participants reported Caucasian ethnicity.

HJECS
The Hannover-Jena Endometrial Cancer Study (HJECS), a hospital-based case-control study, included 250 German women, aged 31-89 years, who were recruited either at the Friedrich Schiller University of Jena or at Hannover Medical School after having been diagnosed with histologically confirmed primary incident endometrial carcinoma between 2004 and 2010. Epidemiological data were obtained from questionnaires, and information on tumor stage and histology was obtained from pathology and clinical reports. Over 98% were of German descent. Interviews were conducted at either the Friedrich Schiller University of Jena or at Hannover Medical School, and peripheral blood was collected for the extraction of DNA from white blood cells.

LES
The Leuven Endometrial Study (LES) is a hospital based case-control study. Eligible cases, identified by active surveillance of electronic patient files at the Leuven University Hospital, were white women aged 27-80 years diagnosed with endometrial cancer. Clinical data for endometrial cancer patients were recorded during interview at the time of diagnosis, and from pathology reports. All medical records were reviewed by trained abstractors and pathology reports compatible with primary, invasive, epithelial endometrial adenocarcinoma of all stages (I –IV) and all grades were consulted. Participation rates exceeded 95% for cases.

MECS
The Mayo Endometrial Cancer Study (MECS) includes a clinic-based prospective collection of primary endometrial cases diagnosed from 2008 to 2011 and seen at Mayo Clinic Rochester with primary endometrial cancer diagnosed at age 18 and older. DNA was

isolated from white blood cells using a Qiagen isolation kit. DNA concentration was measured with picogreen. Clinical data were abstracted from electronic medical records and supplemented with a risk factor questionaire. Control data were obtained from Mayo Clinic OCAC controls (MAY) and BCAC controls (MCBCS).

MoMaTEC

Molecular Markers in Treatment of Endometrial Cancer (MoMaTEC) cases were recruited from an unselected patient population primarily treated for endometrial carcinoma at Haukeland University Hospital, Bergen during 2001-2009. This is the referral hospital for Hordaland county; the area is demographically well defined, with about 450,000 inhabitants, representing approximately 10% of the Norwegian population and with a similar incidence rate and prognosis as the total Norwegian population of endometrial cancers[12–14]. Clinical Information for cases regarding age, FIGO stage, histologic subtype, grade and prognosis was extracted from medical records. DNA was extracted from peripheral blood samples.

NECS

The Newcastle Endometrial Cancer Study (NECS) includes histologically confirmed endometrial cancer cases consecutively recruited from 1992 up to 2005 at the Hunter Centre for Gynaecological Cancer, John Hunter Hospital, Newcastle, New South Wales, Australia[15]. The final analysis included 194 endometrial cancer patients. Data on reproductive and environmental risk factors including ethnicity was collected using self reported questionnaires. Information regarding recurrence, stage, grade and histology of endometrial cancer was collected from medical records. Patients presenting at this hospital-based site were captured by ANECS recruitment from 2005 onwards.

RENDOCAS

The Registry of Endometrial Cancer in Sweden (RENDOCAS) is a hospital based case-control study. Patients (n=520) who underwent surgery for endometrial cancer at Karolinska University hospital Solna, Sweden between 2008 and 2011 were included in the study. For each patient, the following was collected: blood and tumor samples; detailed family history and formulation of a pedigree where all suspected cancer cases were verified in medical records/pathology report if possible; questionnaire covering relevant environmental factors underlying endometrial cancer.

Phase 1 additional genotyping - iCOGS Control Sample Sets:

As indicated in **Supplementary Table 1**, iCOGS endometrial cancer case sample sets were matched with controls from the same countries and also clustered with cases in PCA (**Supplementary Figure 2**). Controls were genotyped using the same iCOGS array and data were largely drawn from healthy controls participating in the Breast Cancer Association Consortium (BCAC)[16] part of the iCOGS initiative. Additional controls were from the Mayo Clinic via the Ovarian Cancer Association Consortium (OCAC)[17], and Norwegian female controls recruited in Bergen for use in the MoMaTEC case-control genotyping studies.

Endometrioid and non-endometrioid histology analysis

Cases were defined as having endometrioid subtype based on pathology report of endometrioid histology only. Non-endometrioid subtypes included carcinosarcoma, clear cell, serous, mucinous, and tumours of mixed histology (any combination). 6,635 (86%) of the 7,737 endometrial cancer cases displayed endometrioid-only histology and association testing and meta-analysis was also conducted using endometrioid-only histology cases. The results of this analysis for the novel risk loci are shown in **Table 1**. Endometrioid-only phase 1 meta-analysis (n=5,590) found only novel risk loci that were identified in the all histologies analysis (**Table 1**, **Supplementary Table 2**). Analysis of endometrial cancer cases that displayed non-endometrioid histology (in Phase 1, or the final meta-analysis) found no SNPs near genome-wide significance, as expected given the limited statistical power.

Ishikawa and ECC-1 cells

Results from Ishikawa and ECC-1 cells are listed separately in publicly available ENCODE[18] tier 3 data but STR-profiling has shown that these two cell lines are very similar[19]. Based on the results presented by Korch *et al.*, the International Cell Line Authentication Committee (ICLAC) recommended in the 2013 Database of Cross-Contaminated or Misidentified Cell Lines that ECC-1 be re-identified as Ishikawa cells. Our i*n vitro* functional analysis for the 13q22 locus made use of our supply of Ishikawa cells for FAIRE and ChIP experiments and ECC-1 cells for luciferase reporter assays. Both cell lines displayed identical genotypes for rs9600103 and rs11841589, similar *KLF5* expression levels, and 20x sequencing using the Ion AmpliSeq™ Comprehensive Cancer Panel confirmed that variants in Ishikawa and ECC-1 are 90% concordant (based on 3,004 exonic SNVs in 409 cancer-related genes). In line with these findings and ICLAC recommendations, we have presented functional work on these cells as Ishikawa cells.

ECAC Study Collaborators

The ANECS Group comprises*:* AB Spurdle, PM Webb, J Young (QIMR Berghofer Medical Research Institute); Consumer representative: L McQuire; Clinical Collaborators: NSW: S Baron-Hay, D Bell, A Bonaventura, A Brand, S Braye, J Carter, F Chan, C Dalrymple, A Ferrier (deceased), G Gard, N Hacker, R Hogg, R Houghton, D Marsden, K McIlroy, G Otton, S Pather, A Proietto, G Robertson, J Scurry, R Sharma, G Wain, F Wong; Qld: J Armes, A Crandon, M Cummings, R Land, J Nicklin, L Perrin, A Obermair, B Ward; SA: M Davy, T Dodd, J Miller, M Oehler, S Paramasivum, J Pierides, F Whitehead; Tas: P Blomfield, D Challis; Vic: D Neesham, J Pyman, M Quinn, R Rome, M Weitzer; WA: B Brennan, I Hammond, Y Leung, A McCartney (deceased), C Stewart, J Thompson; Project Managers: S O'Brien, S Moore; Laboratory Manager: K Ferguson; Pathology Support: M Walsh; Admin Support: R Cicero, L Green, J Griffith, L Jackman, B Ranieri; Laboratory Assistants: M O'Brien, P Schultz; Research Nurses: B Alexander, C Baxter, H Croy, A Fitzgerald, E Herron, C Hill, M Jones, J Maidens, A Marshall, K Martin, J Mayhew, E Minehan, D Roffe, H Shirley, H Steane, A Stenlake, A Ward, S Webb, J White.

CHIBCHA (study of hereditary cancer in Europe and Latin America) collaborators include: Ma. Magdalena Echeverry de Polanco, Mabel Elena Bohórquez, Rodrigo Prieto, Angel Criollo, Carolina Ramírez, Ana Patricia Estrada, Jhon Jairo Suárez (Grupo de Citogenética Filogenia y Evolución de Poblaciones, Universidad del Tolima, Colombia); Augusto Rojas Martinez (Center for Research and Development in Health Sciences, Universidad Autónoma de Nuevo León, Monterrey, Mexico); Silvia Rogatto, Samuel Aguiar Jnr, Ericka Maria Monteiro Santos (Department of Urology, School of Medicine, UNESP - São Paulo State University, Botucatu, Brazil); Monica Sans, Valentina Colistro, Pedro C. Hidalgo, Patricia Mut (Department of Biological Anthropology, College of Humanities and Educational Sciences, University of the Republic, Magallanes, Montevideo, Uruguay); Angel Carracedo, Clara Ruiz Ponte, Ines Quntela Garcia (Fundacion Publica Galega de Medicina Xenomica, CIBERER, Genomic Medicine Group-University of Santiago de Compostela, Hospital Clinico, Santiago de Compostela, Galicia, Spain); Sergi Castellvi-Bel (Department of Gastroenterology, Institut de Malalties Digestives i Metabòliques, Hospital Clínic, Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas, IDIBAPS, University of Barcelona, Barcelona, Catalonia, Spain); Manuel Teixeira (Department of Genetics, Portuguese Oncology Institute, Rua Dr, António Bernardino de Almeida, Porto, Portugal).

The NSECG Group comprises: Ian Tomlinson (Oxford University); M Adams, A Al-Samarraie, S Anwar, R Athavale, S Awad, A Bali, A Barnes, G Cawdell, S Chan, K Chin, P Cornes, M Crawford, J Cullimore, S Ghaem-Maghami, R Gornall, J Green, M Hall, M Harvey, J Hawe, A Head, J Herod, M Hingorani, M Hocking, C Holland, T Hollingsworth,J Hollingworth, T Ind, R Irvine, C Irwin, M Katesmark, S Kehoe, G Kheng-Chew, K Lankester, A Linder, D Luesley, C B-Lynch,V McFarlane, R Naik, N Nicholas, D Nugent, S Oates, A Oladipo, A Papadopoulos, S Pearson, D Radstone, S Raju, A Rathmell, C Redman, M Rymer, P Sarhanis, G Sparrow, N Stuart, S Sundar, A Thompson, S Tinkler, S Trent, A Tristram, N Walji, R Woolas.

RENDOCAS investigators include: Annika Lindblom, Gerasimos Tzortzatos, Miriam Mints, Emma Tham, Ofra Castro, Kristina Gemzell-Danielsson.

SEARCH collaborators include: Helen Baker, Caroline Baynes, Don Conroy, Bridget Curzon, Patricia Harrington, Sue Irvine, Craig Luccarini, Rebecca Mayes, Hannah Munday, Barbara Perkins, Daisy Pharoah, Radka Platte, Anne Stafford and Judy West.

BCAC and OCAC Study Collaborators (for control samples):

The Australian Ovarian Cancer Study Group comprises: R Stuart-Harris; NSW- F Kirsten, J Rutovitz, P Clingan, A Glasgow, A Proietto, S Braye, G Otton, J Shannon, T Bonaventura, J Stewart, S Begbie, M Friedlander, D Bell, S Baron-Hay, A Ferrier (deceased), G Gard, D Nevell, N Pavlakis, S Valmadre, B Young, C Camaris, R Crouch, L Edwards, N Hacker, D Marsden, G Robertson, P Beale, J Beith, J Carter, C Dalrymple, R Houghton, P Russell, L Anderson, M Links, J Grygiel, J Hill, A Brand, K Byth, R Jaworski, P Harnett, R Sharma, G Wain; QLD- D Purdie, D Whiteman, B Ward, D Papadimos, A Crandon, M Cummings, K Horwood. A Obermair, L Perrin, D Wyld, J Nicklin; SA- M Davy, MK Oehler, C Hall, T Dodd, T Healy, K Pittman, D Henderson, J Miller, J Pierdes, A Achan; TAS- P Blomfield, D Challis, R McIntosh, A Parker; VIC- B Brown, R Rome, D Allen, P Grant, S Hyde, R Laurie M Robbie, D Healy, T Jobling, T Manolitsas, J McNealage, P Rogers, B Susil, E Sumithran, I Simpson, I Haviv, K Phillips, D Rischin, S Fox, D Johnson, S Lade, P Waring, M Loughrey, N O'Callaghan, B Murray, L Mileshkin, P Allan; V Billson, J Pyman, D Neesham, M Quinn, A Hamilton, C Underhill, R Bell, LF Ng, R Blum, V Ganju; WA- I Hammond, A McCartney (deceased), C Stewart, Y Leung, M Buck, N Zeps (WARTN); AOCS Management Group- DDL Bowtell, AC Green, G Chenevix-Trench, A deFazio, D Gertig, PM Webb.

BSUCH collaborator: Peter Bugert

ESTHER collaborators: Volker Arndt, Heiko Müller, Christa Stegmaier

GENICA Network collaborators: Wing-Yee Lo, Christina Justenhoven, Ute Hamann, Thomas Brüning, Beate Pesch, Yon-Dschun Ko, Sylvia Rabstein, Anne Lotz, Christina Baisch, Hans-Peter Fischer, Volker Harth.

Supplementary Acknowledgements

The authors thank the many individuals who participated in this study and the numerous institutions and their staff who have supported recruitment.

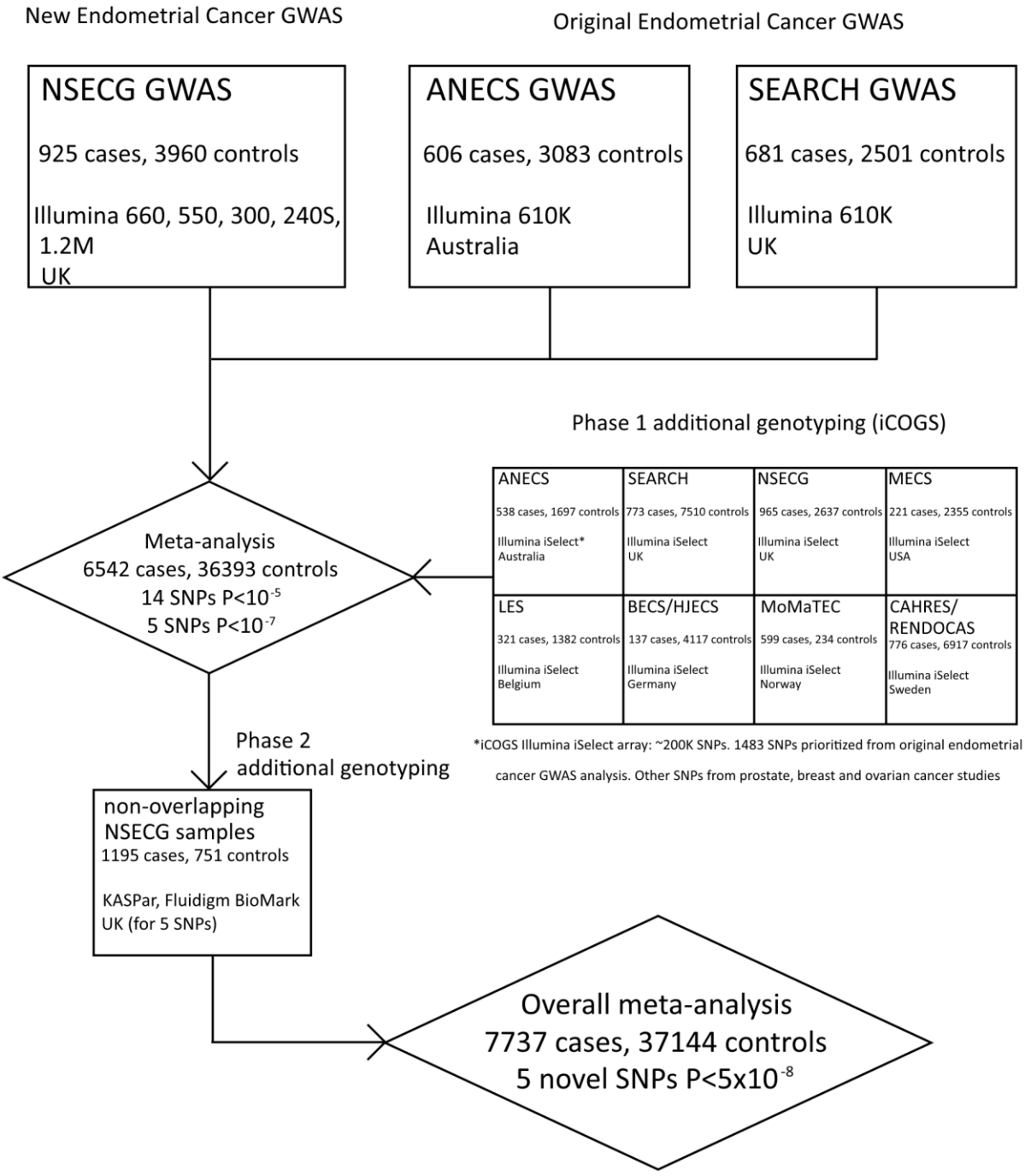
ANECS thanks members of the Molecular Cancer Epidemiology and Cancer Genetic laboratories at QIMR Berghofer Medical Research Institute for technical assistance, and the ANECS research team for assistance with the collection of risk factor information and blood samples. ANECS also gratefully acknowledges the cooperation of the following institutions: NSW: John Hunter Hospital, Liverpool Hospital, Mater Misericordiae Hospital (Sydney), Mater Misericordiae Hospital (Newcastle), Newcastle Private Hospital, North Shore Private Hospital, Royal Hospital for Women, Royal Prince Alfred Hospital, Royal North Shore Hospital, Royal Prince Alfred Hospital, St George Hospital; Westmead Hospital, Westmead Private Hospital; Qld: Brisbane Private Hospital, Greenslopes Hospital, Mater Misericordiae Hospitals, Royal Brisbane and Women's Hospital, Wesley Hospital, Queensland Cancer Registry; SA: Adelaide Pathology Partners, Burnside Hospital, Calvary Hospital, Flinders Medical Centre, Queen Elizabeth Hospital, Royal Adelaide Hospital, South Australian Cancer Registry; Tas: Launceston Hospital, North West Regional Hospitals, Royal Hobart Hospital; Vic: Freemasons Hospital, Melbourne Pathology Services, Mercy Hospital for Women, Royal Women's Hospital, Victorian Cancer Registry; WA: King Edward Memorial Hospital, St John of God Hospitals Subiaco & Murdoch, Western Australian Cancer Registry.
SEARCH thanks the SEARCH research team for recruitment, and also acknowledges the assistance of the Eastern Cancer Registration and Information Centre for subject recruitment.
BECS thanks Reiner Strick, Silke Landrith and Sonja Oeser for their logistic support during the study.
CAHRES (formerly known as SASBAC) thanks Li Yuqing from the Genome Institute of Singapore for contributions to this study, and also acknowledges previous input to SASBAC resource creation by Anna Christensson, Boel Bissmarck, Kirsimari Aaltonen, Karl von Smitten, Nina Puolakka, Christer Halldén, Lim Siew Lan and Irene Chen, Lena U. Rosenberg, Mattias Hammarström, and Eija Flygare.
HJECS thanks Wen Zheng, Hermann Hertel, and Tjoung-Won Park-Simon at Hannover Medical School for their contribution to sample recruitment.
LES gratefully acknowledges Helena Soenen, Gilian Peuteman and Dominiek Smeets for their technical assistance.
MECS thanks Tom Sellers, Catherine Phelan, Andrew Berchuck, and Kimberly Kalli, Amanda von Bismarck, Luisa Freyer and Lisa Rogmann.
MoMaTEC thanks Britt Edvardsen, Ingjerd Bergo and Mari Kyllsø Halle for technical assistance and Inger Marie Aksnes and Tor Audun Hervig at the Blood bank, Haukeland University Hospital for assistance with control recruitment.
NECS thanks staff at the University of Newcastle and the Hunter Medical Research Institute.
NSECG thank Ella Barclay and Lynn Martin for their contribution, and acknowledge the invaluable help of the National Cancer Research Network with the collection of study participants.
QIMR Berghofer thanks Margie Wright, Lisa Bowdler, Sara Smith, Megan Campbell and Scott Gordon for control sample collection and data processing, Kerenaftali Klein for statistical advice and Brendan Ryan for assistance with the figures.
RENDOCAS thanks Berith Wejderot, Sigrid Sahlen, Tao Liu, Margareta Ström, Maria Karlsson, and Birgitta Byström for their contribution to the study.
BSUCH thanks the Medical Faculty, Mannheim, the Diemtmar Hopp Foundation and the German Cancer Research Center.
MCCS was made possible by the contribution of many people, including the original investigators and the diligent team who recruited the participants and who continue working on follow up. We would like to express our gratitude to the many thousands of Melbourne residents who continue to participate in the study.

The UKBGS thank Breakthrough Breast Cancer and the Institute of Cancer Research for support and funding, and the Study participants, Study staff, and the doctors, nurses and other health care staff and data providers who have contributed to the Study. The ICR acknowledges NHS funding to the NIHR Biomedical Research Centre.
In addition, the iCOGS study would not have been possible without the contributions of: Andrew Berchuck (OCAC), Rosalind A. Eeles, Ali Amin Al Olama, Zsofia Kote-Jarai , Sara Benlloch (PRACTICAL), Georgia Chenevix-Trench, Antonis Antoniou, Lesley McGuffog, Fergus Couch and Ken Offit (CIMBA), Andrew Lee, and Ed Dicks, Craig Luccarini and the staff of the Centre for Cancer Genetic Epidemiology Laboratory (Cambridge), Javier Benitez, Anna Gonzalez-Neira and the staff of the CNIO genotyping unit, Jacques Simard and Daniel C. Tessier, Francois Bacot, Daniel Vincent, Sylvie LaBoissière and Frederic Robidoux and the staff of the McGill University and Génome Québec Innovation Centre, Stig E. Bojesen, Sune F. Nielsen, Borge G. Nordestgaard, and the staff of the Copenhagen DNA laboratory, Sharon A. Windebank, Christopher A. Hilker, Jeffrey Meyer and the staff of Mayo Clinic Genotyping Core Facility.

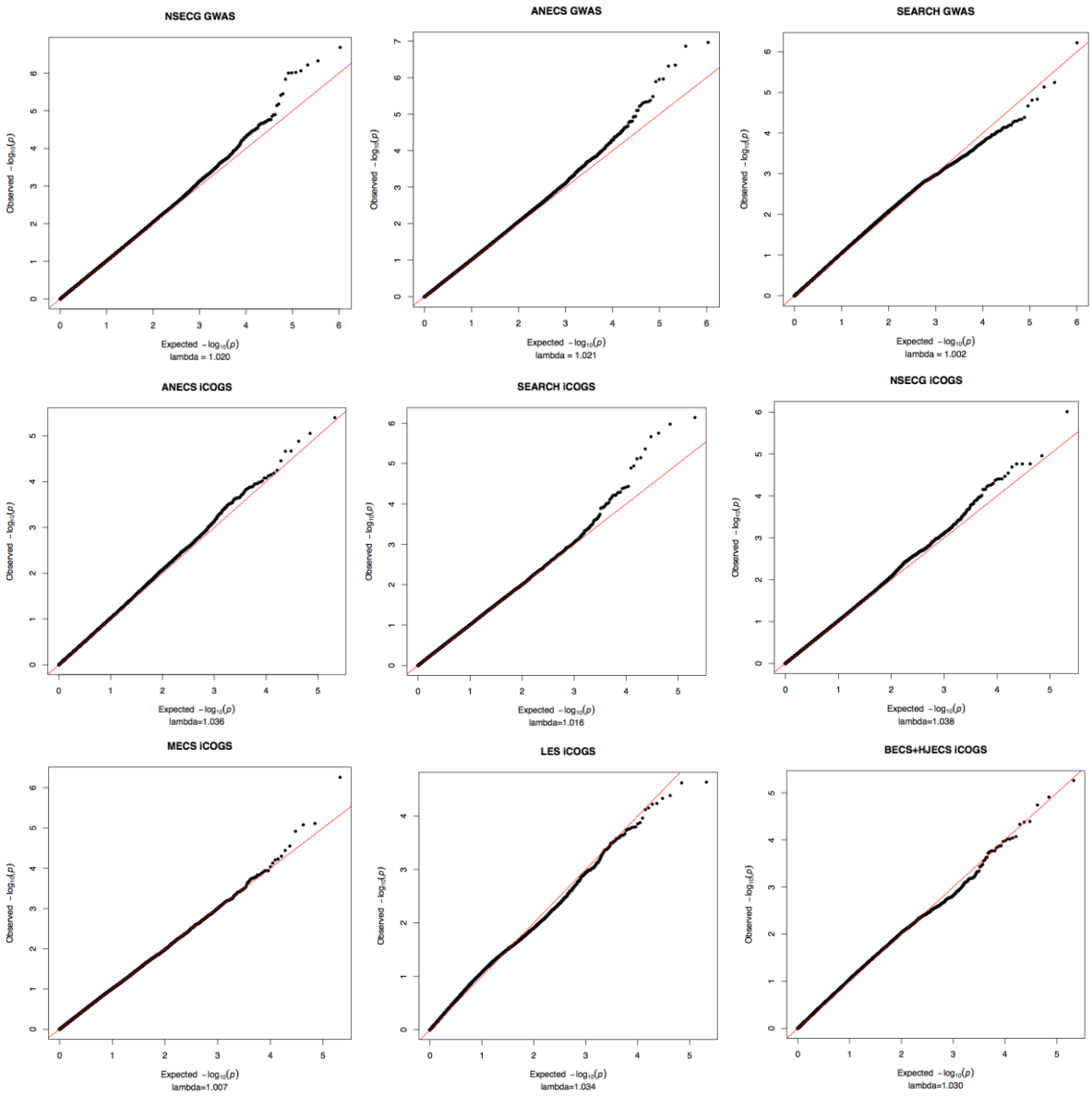
Acknowledgments of Funding to BCAC/OCAC control groups

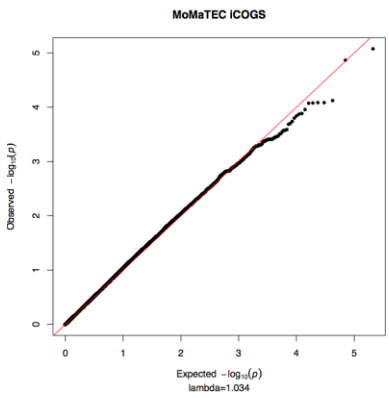
The ESTHER study was funded by the Baden-Württemberg state Ministry of Science, Research and Arts (Stuttgart, Germany), the Federal Ministry of Education and Research (Berlin, Germany) and the Federal Ministry of Family Affairs, Senior Citizens, Women and Youth (Berlin, Germany).
The GENICA was funded by the Federal Ministry of Education and Research (BMBF) Germany grants 01KW9975/5, 01KW9976/8, 01KW9977/0 and 01KW0114, the Robert Bosch Foundation, Stuttgart, Deutsches Krebsforschungszentrum (DKFZ), Heidelberg, Institute for Prevention and Occupational Medicine of the German Social Accident Insurance, Institute of the Ruhr University Bochum (IPA), Germany, as well as the Department of Internal Medicine, Evangelische Kliniken Bonn gGmbH, Johanniter Krankenhaus, Bonn, Germany.
Financial support for the KARBAC study was provided through the regional agreement on medical training and clinical research (ALF) between Stockholm County Council and Karolinska Institutet, as well as the Swedish Cancer Society.
The MARIE study was supported by the Deutsche Krebshilfe e.V. [70-2892-BR I], the Hamburg Cancer Society and the German Cancer Research Center
MAY was supported by R01-CA122443, P50-CA136393, the Fred C. and Katherine B. Andersen Foundation, and the Mayo Foundation.
MCBCS recognizes funding from the Breast Cancer Research Foundation (BCRF), the David F. and Margaret T. Grohne Family Foundation, and the Ting Tsung and Wei Fong Chao Foundation
MCCS recruitment was funded by VicHealth and Cancer Council Victoria, and its follow-up has been continuously supported by infrastructure provided by Cancer Council Victoria.
UKBGS was funded by Breakthrough Breast Cancer and the Institute of Cancer Research, which acknowledges NHS funding to the NIHR Biomedical Research Centre

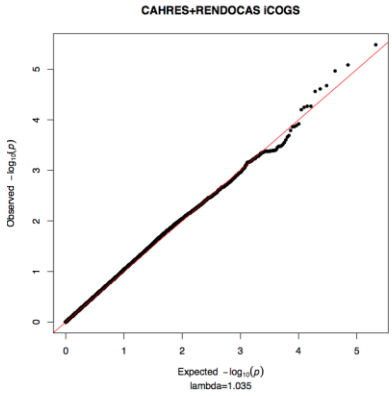
Supplementary Figure 1: Endometrial cancer meta-analysis flow diagram.
This schematic figure illustrates the endometrial cancer case-control meta-analysis study design. The new NSECG GWAS was meta-analysed with a re-analysis of the original endometrial cancer GWAS (ANECS and SEARCH) and the iCOGS phase 1 genotypes (eight groups). This meta-analysis of 6,542 cases and 36,393 controls yielded 14 regions with SNPs $P<10^{-5}$, of which five regions had SNPs $P<10^{-7}$. These five SNPs were brought forward to the additional genotyping in a second phase, involving non-overlapping NSECG cases and controls, and were confirmed as novel genome-wide significant risk loci ($P<5\times10^{-8}$) in the overall meta-analysis of 7,737 cases and 37,144 controls.



Supplementary Figure 2: Quantile-quantile plots of the ranked trend test statistics for three GWAS and eight iCOGS groups.
The $-\log_{10}$ transformed observed P-values (y-axis) were plotted against the expected P-values under the null hypothesis (x-axis). The red line denotes the expectation under no deviation from the null hypothesis. The QQ-plots show little evidence of genomic inflation and the λ_{GC} for each study are: NSECG GWAS 1.020, ANECS GWAS 1.021, SEARCH GWAS 1.002, ANECS iCOGS 1.036, SEARCH iCOGS 1.016, NESCG iCOGS 1.038, MECS iCOGS 1.007, LES iCOGS 1.034, BECS iCOGS 1.030, MoMaTEC iCOGS 1.034, CAHRES RENDOCAS iCOGS 1.037. For the three GWAS, all genotyped SNPs passing QC are displayed. For iCOGS, 105,000 SNPs after LD-pruning (r^2<0.2) and >500kb from the 1,483 endometrial cancer prioritized SNPs on the iCOGS are displayed.

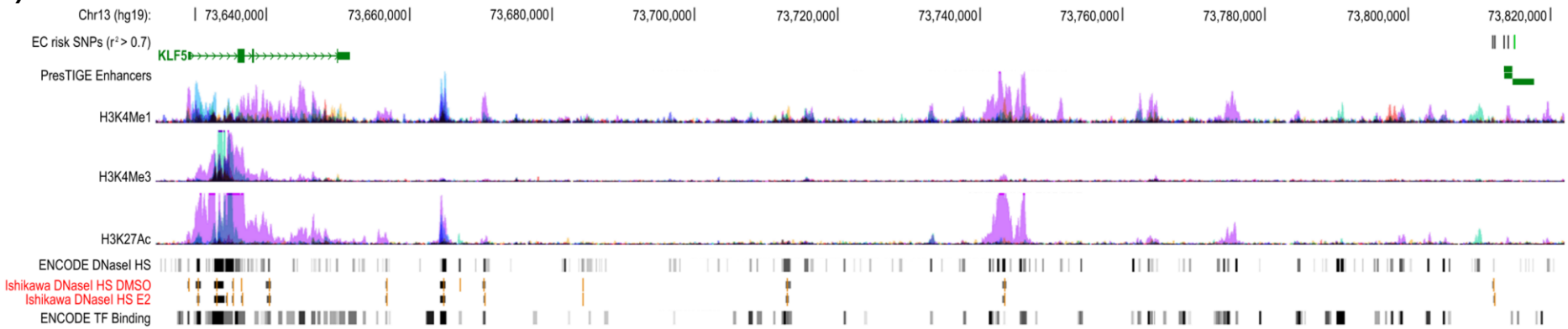



MoMaTEC iCOGS
Observed $-\log_{10}(p)$
Expected $-\log_{10}(p)$
lambda=1.034

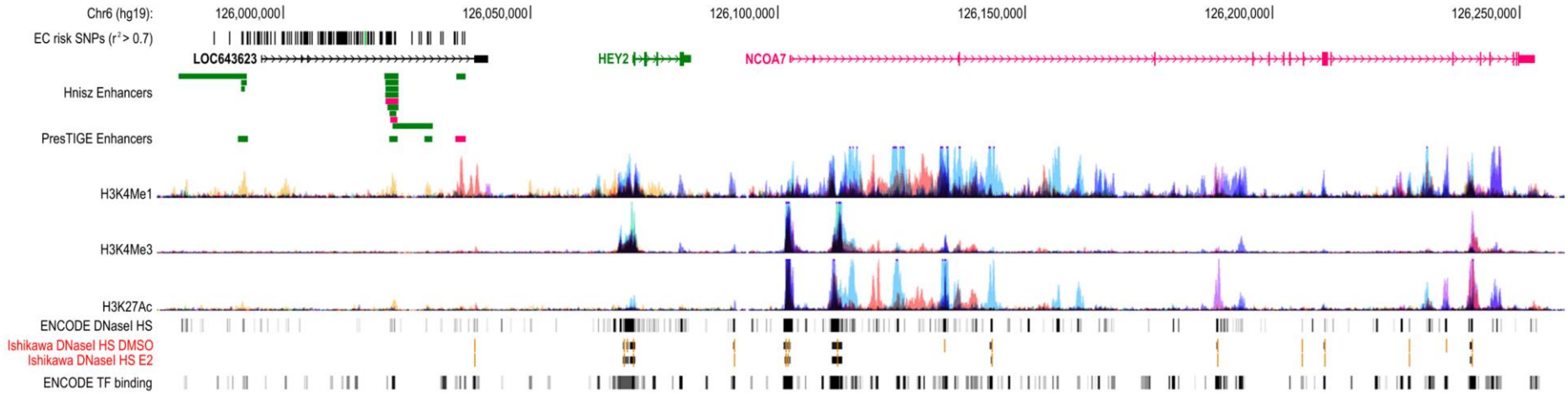

CAHRES+RENDOCAS iCOGS
Observed $-\log_{10}(p)$
Expected $-\log_{10}(p)$
lambda=1.035

Supplementary Figure 3: Genetic landscape of novel endometrial cancer associated regions.
Plots for novel risk loci at a) 13q22.1, b) 6q22.31, c) 8q24.21, d) 15q15 and e) 14q32. SNPs in strong LD ($r^2 > 0.7$) with the lead endometrial cancer risk SNP have been plotted for each region and the lead SNP denoted in green. The second, independently associated SNP found at 8q24.21 after conditioning on the lead SNP is denoted in red (c). Previously reported cancer risk SNPs identified by GWAS at 8q24 are shown in blue (c), none of which are in LD (r2≤0.02) with endometrial cancer risk SNPs. Likely enhancers identified by Hnisz *et al.*[20] and PresTIGE[21] that overlap endometrial cancer risk associated SNPs are depicted as colored bars, where the color of the likely enhancer matches the schematic of its predicted target gene, as determined by correlations with gene expression. As described in Online Methods Hnisz *et al*. predicted enhancer-gene interactions by identifying 'super-enhancers' (regions containing neighboring 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. Additional tracks include: Histone modifications associated with promoters (H3K4Me3) and enhancers (H3K4Me1 and H3K27Ac) from seven ENCODE Project cell types; DNaseI hypersensitivity sites (DHS) and transcription factor (TF) binding identified in 125 and 91 ENCODE Project cell types, respectively; DHS identified in Ishikawa endometrial cancer cells using DMSO vehicle and under estrogen (E2) stimulation are shown; transcription factor binding regions in Ishikawa cells that encompass endometrial cancer risk SNP loci are also displayed. For all risk loci, endometrial cancer risk associated SNPs co-locate with at least one enhancer predicted by cell-type specific analysis, implicating the following genes/transcripts as worthy of investigation: a) *KLF5*; b) *HEY2*, *NCOA7*; c) *MYC*, MIR1204, MIR1205, MIR1207, MIR1208; d) *BMF*, *GPR176*, *SRP14, LOC100131089*; e) *AKT*, *ADSSL1*, *INF2*, *ZBTB42, SIVA1*. For four loci (a, c, d and e), likely enhancers overlap with at least one region displaying evidence of regulatory activity (DHS and/or TF binding) in the single endometrial cancer cell line (Ishikawa) assayed by ENCODE.

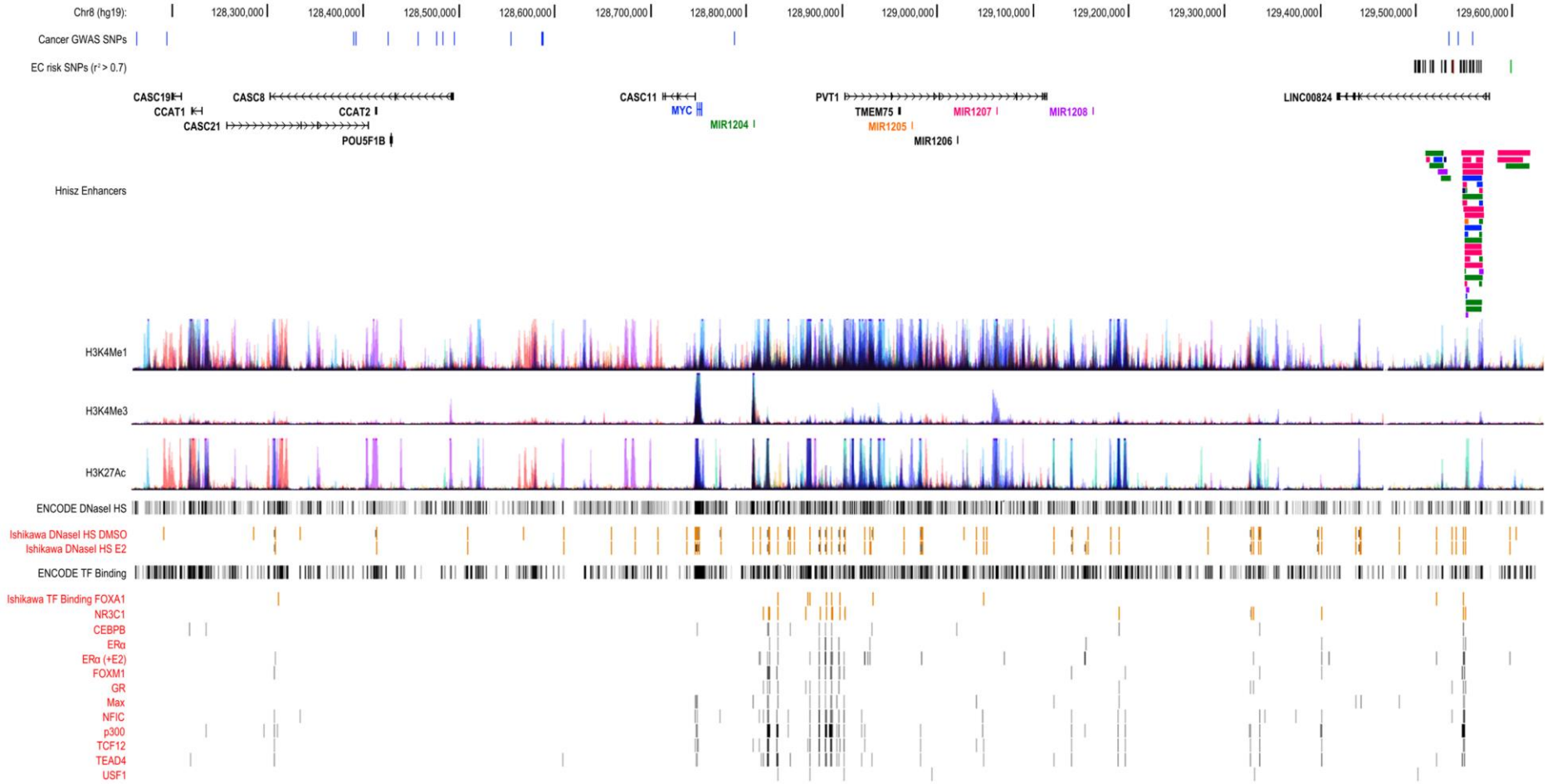
a)



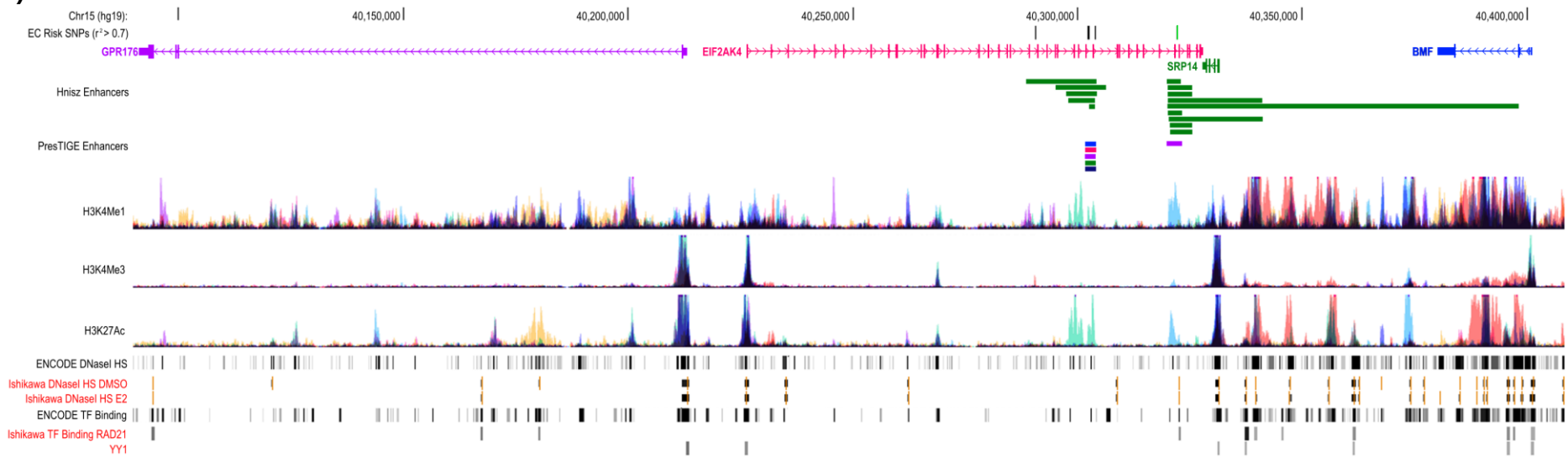
b)


Chr6 (hg19):
126,000,000
126,050,000
126,100,000
126,150,000
126,200,000
126,250,000
EC risk SNPs (r² > 0.7)
LOC643623
HEY2
NCOA7
Hnisz Enhancers
PresTIGE Enhancers
H3K4Me1
H3K4Me3
H3K27Ac
ENCODE DNaseI HS
Ishikawa DNaseI HS DMSO
Ishikawa DNaseI HS E2
ENCODE TF binding

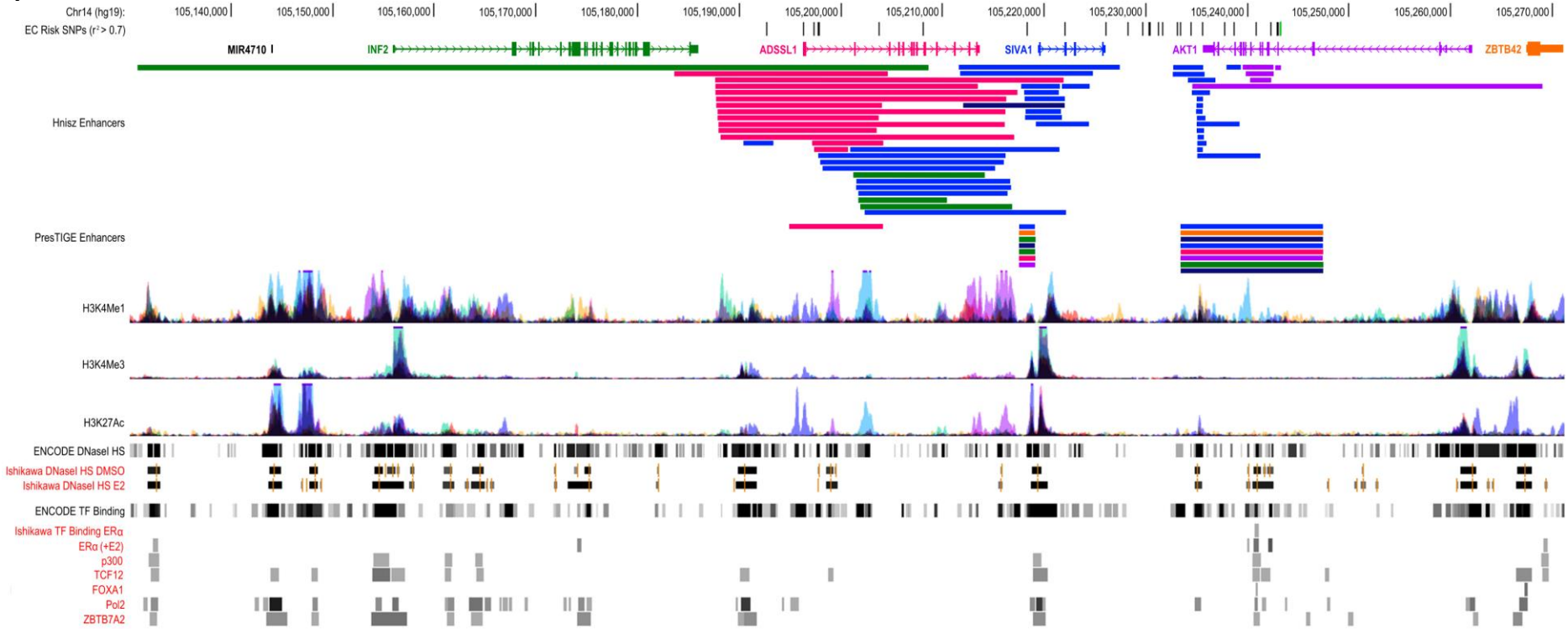
c)


Chr8 (hg19):
128,300,000
128,400,000
128,500,000
128,600,000
128,700,000
128,800,000
128,900,000
129,000,000
129,100,000
129,200,000
129,300,000
129,400,000
129,500,000
129,600,000
Cancer GWAS SNPs
EC risk SNPs (r² > 0.7)
CASC19
CCAT1
CASC8
CCAT2
CASC21
POU5F1B
CASC11
MYC
MIR1204
PVT1
TMEM75
MIR1207
MIR1208
MIR1205
MIR1206
LINC00824
Hnisz Enhancers
H3K4Me1
H3K4Me3
H3K27Ac
ENCODE DNaseI HS
Ishikawa DNaseI HS DMSO
Ishikawa DNaseI HS E2
ENCODE TF Binding
Ishikawa TF Binding FOXA1
NR3C1
CEBPB
ERα
ERα (+E2)
FOXM1
GR
Max
NFIC
p300
TCF12
TEAD4
USF1

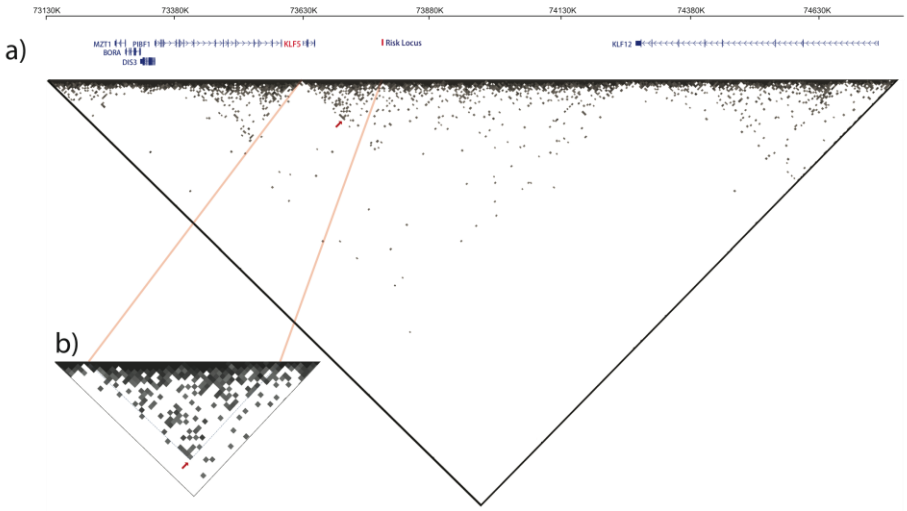
d)


Chr15 (hg19):
40,150,000
40,200,000
40,250,000
40,300,000
40,350,000
40,400,000
EC Risk SNPs (r² > 0.7)
GPR176
EIF2AK4
SRP14
BMF
Hnisz Enhancers
PresTIGE Enhancers
H3K4Me1
H3K4Me3
H3K27Ac
ENCODE DNaseI HS
Ishikawa DNaseI HS DMSO
Ishikawa DNaseI HS E2
ENCODE TF Binding
Ishikawa TF Binding RAD21
YY1

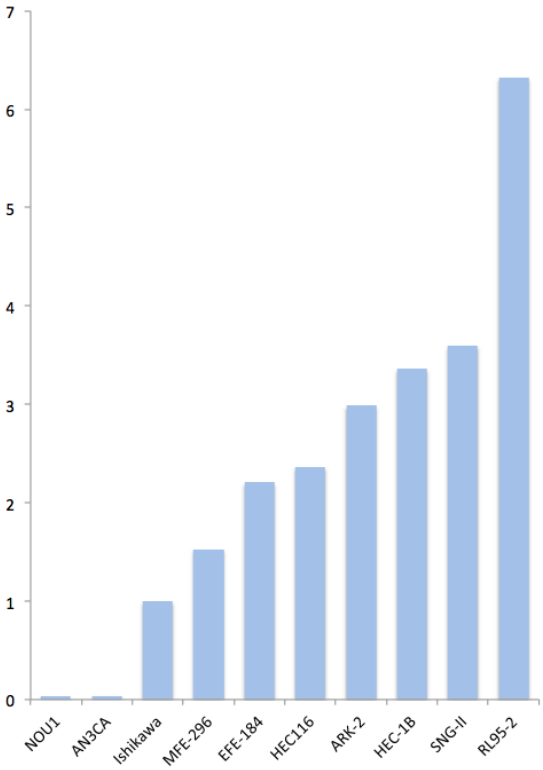
e)


Chr14 (hg19):
105,140,000
105,150,000
105,160,000
105,170,000
105,180,000
105,190,000
105,200,000
105,210,000
105,220,000
105,230,000
105,240,000
105,250,000
105,260,000
105,270,000
EC Risk SNPs (r² > 0.7)
MIR4710
INF2
ADSSL1
SIVA1
AKT1
ZBTB42
Hnisz Enhancers
PresTIGE Enhancers
H3K4Me1
H3K4Me3
H3K27Ac
ENCODE DNaseI HS
Ishikawa DNaseI HS DMSO
Ishikawa DNaseI HS E2
ENCODE TF Binding
Ishikawa TF Binding ERα
ERα (+E2)
p300
TCF12
FOXA1
Pol2
ZBTB7A2

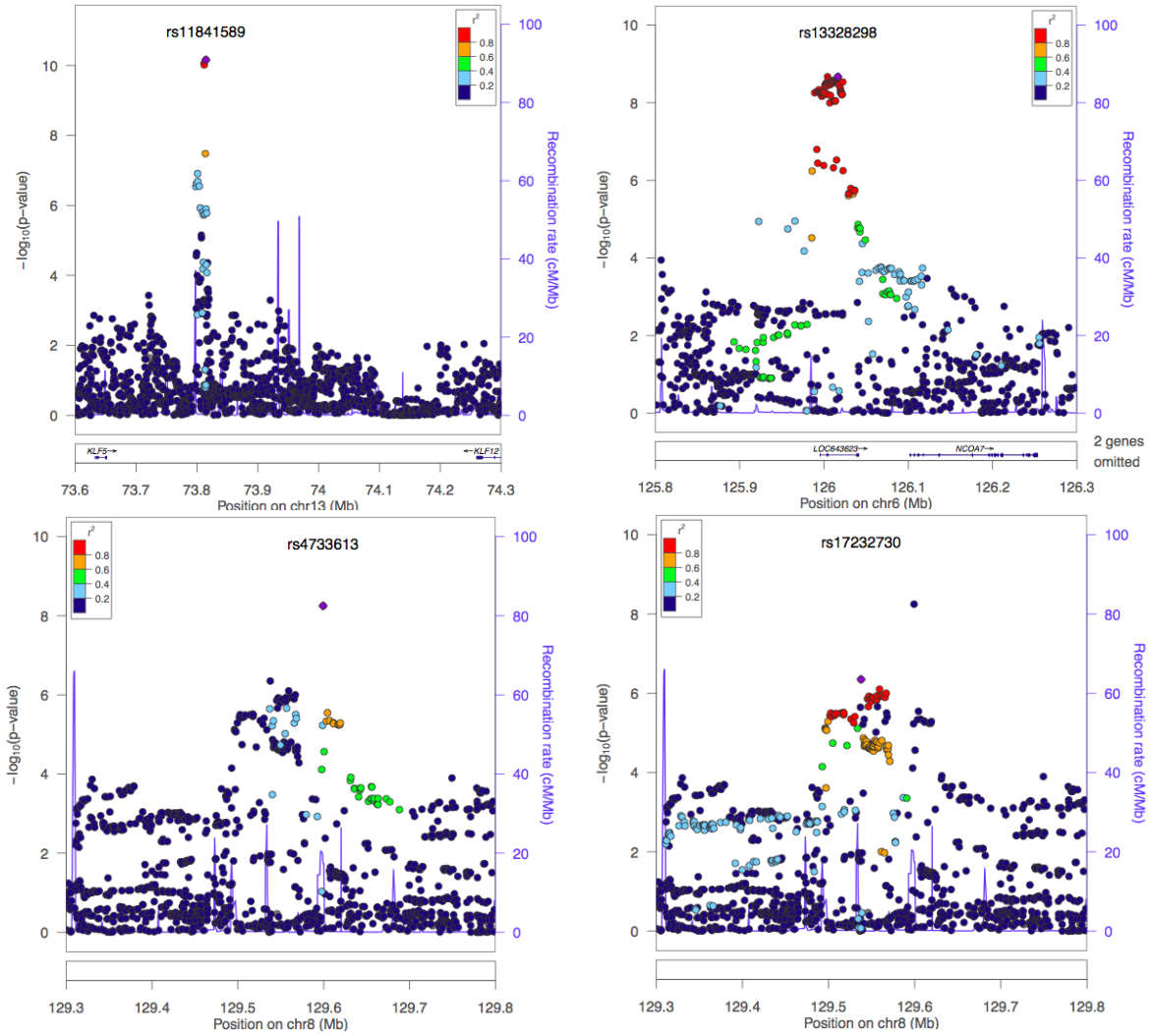
Supplementary Figure 4: Hi-C chromatin capture of 13q22 locus in Hela S3 cells. a) 5Kb KR normalized contact matrix in Hi-C experiment for HeLa S3 cells was used to represent the interaction pattern between *KLF5* and risk locus rs11841589/rs9600103[22]. A loop was anchored at the *KLF5* promoter and the risk locus (see b) for the small topologically associated domain and the red arrow for loop anchor), which indicated distal *cis*-regulatory element within the risk locus. The interaction between the rs11841589/rs9600103 risk locus with the *KLF5* promoter was the strongest interaction observed out of the 262 protein-coding genes on chromosome 13 (P=0.004). The color scheme in the contact matrix is KR normalized score, with the black indicating a strong interaction.

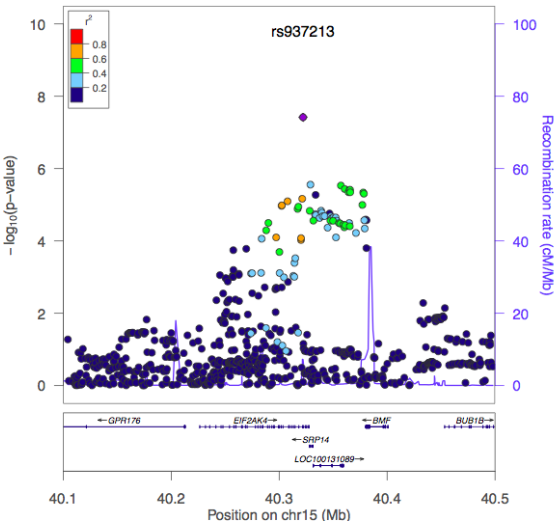


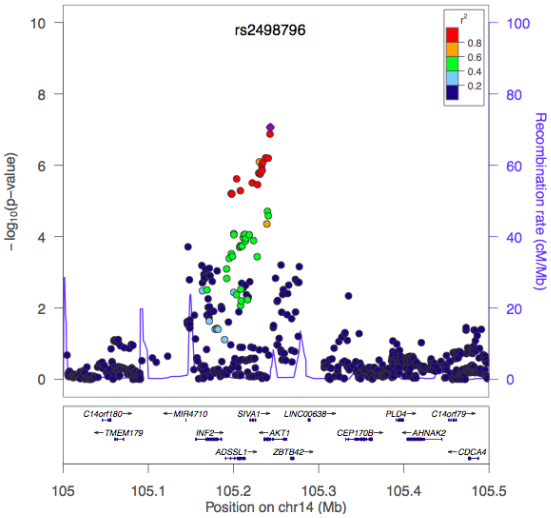
Supplementary Figure 5: Quantification of KLF5 expression in endometrial cancer cell lines. Expression of *KLF5* in 11 endometrial cancer cell lines as described in **Online methods** using qRT-PCR, expression levels on the x-axis are relative to *KLF5* expression in Ishikawa cells using the ddCT method.



Supplementary Figure 6: 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 NSECG, ANECS, SEARCH, 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 labelled 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. Compared with **Figure 2**, more SNPs are displayed in these plots. SNPs with info scores of more than 0.6 in iCOGS and more than 0.9 in NSECG, ANECS, and SEARCH are included.




rs937213
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
GPR176
EIF2AK4
BMF
BUB1B
SRP14
LOC100131089
Position on chr15 (Mb)


rs2498796
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
C14orf180
MIR4710
SIVA1
LINC00638
PLD4
C14orf79
TMEM179
INF2
AKT1
CEP170B
AHNAK2
ADSSL1
ZBTB42
CDCA4
Position on chr14 (Mb)

Supplementary Figure 7: Principal Components Analysis (PCA) of three GWAS and eight iCOGS studies.
Plots of the first two principal components (PCs) in each study. Endometrial cancer cases are represented by blue dots, whereas controls are in black. Samples were excluded if they clustered away from the centroid in the first four PCs and these are plots of the samples used in the analysis.

References

1. Spurdle, A. B. *et al.* Genome-wide association study identifies a common variant associated with risk of endometrial cancer. *Nat. Genet.* **43,** 451–454 (2011).
2. Spurdle, A., Webb, P. & Australian National Endometrial Cancer Study. Re: Excess of early onset multiple myeloma in endometrial cancer probands and their relatives suggests common susceptibility. *Gynecol. Oncol.* **109,** 153; author reply 154 (2008).
3. Tomlinson, I. *et al.* A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. *Nat. Genet.* **39,** 984–988 (2007).
4. 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.* **40,** 631–637 (2008).
5. 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.* **16,** 40–53 (1999).
6. Zhu, G. *et al.* A major quantitative-trait locus for mole density is linked to the familial melanoma gene CDKN2A: a maximum-likelihood combined linkage and association analysis in twins and their sibs. *Am. J. Hum. Genet.* **65,** 483–492 (1999).
7. Painter, J. N. *et al.* Genome-wide association study identifies a locus at 7p15.2 associated with endometriosis. *Nat. Genet.* **43,** 51–54 (2011).
8. McEvoy, M. *et al.* Cohort profile: The Hunter Community Study. *Int. J. Epidemiol.* **39,** 1452–1463 (2010).
9. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447,** 661–678 (2007).
10. Power, C. & Elliott, J. Cohort profile: 1958 British birth cohort (National Child Development Study). *Int. J. Epidemiol.* **35,** 34–41 (2006).
11. Weiderpass, E. *et al.* Risk of endometrial cancer following estrogen replacement with and without progestins. *J. Natl. Cancer Inst.* **91,** 1131–1137 (1999).

12. Wik, E. *et al.* Deoxyribonucleic acid ploidy in endometrial carcinoma: a reproducible and valid prognostic marker in a routine diagnostic setting. *Am. J. Obstet. Gynecol.* **201,** 603.e1–7 (2009).
13. Salvesen, H. B., Iversen, O. E. & Akslen, L. A. Prognostic significance of angiogenesis and Ki-67, p53, and p21 expression: a population-based endometrial carcinoma study. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **17,** 1382–1390 (1999).
14. Salvesen, H. B. *et al.* Integrated genomic profiling of endometrial carcinoma associates aggressive tumors with indicators of PI3 kinase activation. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 4834–4839 (2009).
15. Ashton, K. A. *et al.* The influence of the Cyclin D1 870 G>A polymorphism as an endometrial cancer risk factor. *BMC Cancer* **8,** 272 (2008).
16. Michailidou, K. *et al.* Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat. Genet.* **45,** 353–361 (2013).
17. Pharoah, P. D. P. *et al.* GWAS meta-analysis and replication identifies three new susceptibility loci for ovarian cancer. *Nat. Genet.* **45,** 362–370, 370e1–2 (2013).
18. ENCODE Project Consortium *et al.* Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447,** 799–816 (2007).
19. Korch, C. *et al.* DNA profiling analysis of endometrial and ovarian cell lines reveals misidentification, redundancy and contamination. *Gynecol. Oncol.* **127,** 241–248 (2012).
20. Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155,** 934–947 (2013).
21. 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.* **24,** 1–13 (2014).
22. Rao, S. S. P. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159,** 1665–1680 (2014).

Supplementary Table 1: Endometrial cancer case and control sample sets

	Study		Case sampling frame	Control sampling frame	Genotyping Platform	Endometrial Cancer Cases	Endometoid histology	Controls
	New Endometrial Cancer GWAS							
1	NSECG	UK National Study of Endometrial Cancer Genetics	UK; population based cases		Illumina660WQuads	925	795	
	UK1-CORGI	UK Colorectal Tumour Gene Identification Consortium		England; spouses and partners of cases with no personal or family history of colorectal neoplasia	Illumina Hap550			894
	SP1	Scotland1		Scotland; cancer free controls from NHS registers	Illumina HumanHap300 Illumina HumanHap240S			392
	BC58	UK 1958 Birth Cohort		UK; population based controls, born within one week in 1958	Illumina 1.2M			2,674
	Original Endometrial Cancer GWAS							
2	ANECS	Australian National Endometrial Cancer Study	Australia; population based cases		Illumina 610K	606	606	
	QIMR	Queensland Institute of Medical Research		Australia; parents of participants in adolescent twin study	Illumina 610K			1,846
	HCS	Hunter Community Study		Australia; population-based controls	Illumina 610K			1,237
3	SEARCH	UK Studies of Epidemiology and Risk factors in Cancer Heredity	England; population based cases via cancer registries, age <69		Illumina 610K	681	681	
	NBS			UK; population based controls identified through National Blood Service	Illumina 1.2M			2,501
	iCOGS Genotyping Sample Sets (Phase 1 additional genotyping)							
4	ANECS	Australian National Endometrial Cancer Study	Australia; population based cases		Illumina Infinium iSelect	373	217	
	NECS	Newcastle Endometrial Cancer Study	Australia; hospital-based cases		Illumina Infinium iSelect	165	116	
	ABCFS	Australian Breast Cancer Family Study		Australia; from electoral rolls	Illumina Infinium iSelect			443
	AOCS	Australian Ovarian Cancer Study		Australia; population-based, from electoral rolls	Illumina Infinium iSelect			817
	MCCS	Melbourne Collaborative Cohort Study		Australia; random sample from initial cohort	Illumina Infinium iSelect			437
5	SEARCH	UK Studies of Epidemiology and Risk factors in Cancer Heredity	England; population based cases	England; population based controls	Illumina Infinium iSelect	773	620	7,510
6	NSECG	National Study of Endometrial Cancer Genetics	England; population based cases		Illumina Infinium iSelect	965	839	
	BBCS	British Breast Cancer Study		UK; friend, sister-in-law, daughter-in-law or other non-blood relative of breast cancer case	Illumina Infinium iSelect			1,353
	SBCS	Sheffield Breast Cancer Study		UK; women attending Sheffield Mammography Screening, with no breast lesion	Illumina Infinium iSelect			835
	UKBGS	UK Breakthrough Generations Study		UK; women without breast cancer selected from BGS cohort	Illumina Infinium iSelect			449
7	MECS	Mayo Endometrial Cancer Study	USA; Clinic based cases		Illumina Infinium iSelect	221	163	
	MCBCS	Mayo Clinic Breast Cancer Study		USA; Cancer-free women presenting for general medical examination	Illumina Infinium iSelect			1,762
	MAY	Mayo Clinic Ovarian Cancer Case-Control Study		USA; Cancer-free women presenting for general medical examination	Illumina Infinium iSelect			593
8	LES	Leuven Endometrial Cancer Study	Belgium; hospital based cases		Illumina Infinium iSelect	321	219	
	LMBC	Leuven Multidisciplinary Breast Centre		Belgium; controls from blood donors	Illumina Infinium iSelect			1,382
9	BECS/HJECS	Bavarian Endometrial Cancer Study/Hannover-Jena Endometrial Cancer Study	Germany; hospital-based cases, population-based cases		Illumina Infinium iSelect	137	112	
	BBCC	Bavarian Breast Cancer Cases and Controls		Germany; healthy women >55yrs from newspaper advertisement	Illumina Infinium iSelect			441
	BSUCH	Breast Cancer Study of the University Clinic Heidelberg		Germany; female blood donors	Illumina Infinium iSelect			920
	ESTHER	ESTHER Breast Cancer Study		Germany; random sample from routine health check-up	Illumina Infinium iSelect			486
	GC-HBOC	German Consortium for Hereditary Breast & Ovarian Cancer		Germany; KORA study	Illumina Infinium iSelect			138
	GENICA	Gene Environment Interaction and Breast Cancer in Germany		Germany; random address sample	Illumina Infinium iSelect			420
	MARIE	Mammary Carcinoma Risk Factor Investigation		Germany; randomly drawn from population registries	Illumina Infinium iSelect			1,712
10	MoMaTEC	Molecular Markers in Treatment of Endometrial Cancer	Norway; population based cases	Norway; female blood donors	Illumina Infinium iSelect	599	505	
	NBCS	Norwegian Breast Cancer Study		Norway; attendees at Norwegian Breast Cancer Screening Program	Illumina Infinium iSelect			234
11	CAHRES	Cancer Hormone Replacement Epidemiology	Sweden; population based cases	Sweden; population based controls	Illumina Infinium iSelect	543	512	1,345
	RENDOCAS	Registry of Endometrial Cancer in Sweden	Sweden; hospital based cases		Illumina Infinium iSelect	233	205	
	KARBAC	Karolinska Breast Cancer Study		Sweden; blood donors	Illumina Infinium iSelect			585
	pKARMA	Karolinska Mammography Project for Risk Prediction of Breast Cancer		Sweden; cancer-free participants of mammography screening	Illumina Infinium iSelect			4,987
	NSECG genotyping (Phase 2 additional genotyping)							
12	NSECG	National Study of Endometrial Cancer Genetics	UK; population based cases		KASPar, Fluidigm	1,195	1,045	
	UK1/CORGI	UK Colorectal Tumour Gene Identification Consortium		England; spouses and partners of cases with no personal or family history of colorectal neoplasia	KASPar, Fluidigm			751
	OVERALL META-ANALYSIS TOTAL					**7,737**	**6,635**	**37,144**

Supplementary Table 2: Meta-analysis after regional imputation for risk loci P<10^{-5} identified by meta-analysis of GWAS and iCOGS datasets

Locus	Nearby gene(s)	Top SNP in meta-analysis*									Top SNP after regional imputation**								
		SNP	Position	EA	OA	Allelic OR (95%CI)	P	RE OR (95%CI)***	RE P	I^2	SNP	Position	EA	OA	Allelic OR (95%CI)	P	RE OR (95%CI)***	RE P	I^2
13q22	*KLF5, KLF12*	rs11841589	73,814,891	G	T	1.16 (1.11-1.21)	6.89E-11	-	-	0.25	rs11841589	73,814,891	G	T	1.16 (1.11-1.21)	6.89E-11	-	-	0.25
6q22	*NCOA7, HEY2*	rs2747714	126,007,620	A	G	1.13 (1.08-1.17)	3.35E-09	-	-	0.00	**rs13328298**	126,016,580	G	A	1.13 (1.08-1.17)	2.78E-09	-	-	0.00
8q24	*MYC*	rs4733613	129,599,278	G	C	0.84 (0.79-0.89)	5.64E-09	-	-	0.02	rs4733613	129,599,278	G	C	0.84 (0.79-0.89)	5.64E-09	-	-	0.02
15q15	*EIF2AK, BMF*	rs937213	40,322,124	T	C	0.90 (0.86-0.93)	3.81E-08	0.90 (0.85-0.94)	8.35E-06	0.41	rs937213	40,322,124	T	C	0.90 (0.86-0.93)	3.81E-08	0.90 (0.85-0.94)	8.35E-06	0.41
14q32	*AKT1, SIVA1*	rs3001371	105,242,831	C	T	0.89 (0.85-0.93)	1.33E-07	-	-	0.00	**rs2498796**	105,243,220	G	A	0.89 (0.85-0.93)	8.66E-08	-	-	0.02
12q24	*SH2B3, ATXN2*	rs3184504	111,884,608	C	T	1.11 (1.07-1.15)	2.16E-07	-	-	0.00	rs3184504	111,884,608	C	T	1.11 (1.07-1.15)	2.16E-07	-	-	0.00
17q21	*SKAP1*	rs3944039	46,205,070	C	T	1.10 (1.06-1.15)	1.63E-06	1.10(1.04-1.17)	0.001364	0.50	**rs1452666**	46,307,750	C	T	1.11 (1.06-1.15)	9.49E-07	1.10(1.04-1.17)	7.41E-04	0.48
2p16	*BCL11A*	rs6732518	60,708,597	T	C	1.11 (1.06-1.16)	2.26E-06			0.00	**rs7579014**	60,707,894	A	G	1.11 (1.07-1.16)	5.34E-07			0.00
12p12	*SSPN*	rs17467365	26,431,039	C	T	1.27 (1.15-1.41)	2.82E-06	1.29 (1.13-1.46)	9.97E-05	0.33	rs17467365	26,431,039	C	T	1.27 (1.15-1.41)	2.82E-06	1.29 (1.13-1.46)	9.97E-05	0.33
3q13	*LSAMP*	rs4378954	115,650,448	C	T	1.16 (1.09-1.24)	7.20E-06	1.17 (1.08-1.27)	2.05E-04	0.31	rs4378954	115,650,448	C	T	1.16 (1.09-1.24)	7.20E-06	1.17 (1.08-1.27)	2.05E-04	0.31
9q31	*KLF4*	rs4978670	110,826,546	A	G	0.91 (0.88-0.95)	7.33E-06	-	-	0.20	rs4978670	110,826,546	A	G	0.91 (0.88-0.95)	7.33E-06	-	-	0.20
8q21	*STAU2*	rs4237005	74,391,780	G	A	0.88 (0.83-0.93)	8.02E-06	-	-	0.00	rs4237005	74,391,780	G	A	0.88 (0.83-0.93)	8.02E-06	-	-	0.00
11q13	*CCDC88B, RPS6KA4*	rs71456310	64,128,494	C	T	1.14 (1.08-1.21)	8.69E-06	-	-	0.00	**rs12808002**	64,110,932	C	T	1.14 (1.08-1.20)	2.18E-06	-	-	0.00
3q21	*EEFSEC*	rs3021461	128,095,652	G	A	0.90 (0.86-0.94)	9.96E-06	0.91 (0.85-0.96)	0.001909	0.41	rs3021461	128,095,652	G	A	0.90 (0.86-0.94)	9.96E-06	0.91 (0.85-0.96)	0.001909	0.41

Positions in build 37; EA, effect allele; OA, other allele; P, P-value; I^2, heterogeneity I^2 statistic.

*Top SNPs from meta-analysis were genotyped in iCOGS, and genotyped or imputed in the GWAS datasets.

** Imputed SNPs (in bold) are only included if information score was >0.9 in all datasets.

*** Random effects (RE) meta-analysis was done for SNPs with I^2 >0.3. OR and P-values are reported.

Supplementary Table 3: Overall meta-analysis including additional genotyping from Phase 2

	Locus	Nearby gene(s)	SNP	Position			Risk estimates from GWAS+iCOGs meta-analysis					Risk estimates from Phase 2 NSECG genotyping				Overall P
					EA	OA	Case EAF	Control EAF	Allelic OR (95%CI)	P	I^2	Case EAF	Control EAF	Allelic OR (95%CI)	P	
	13q22	*KLF5, KLF12*	rs11841589	73,814,891	G	T	0.757	0.729	1.16 (1.11-1.21)	6.89E-11	0.25	0.741	0.723	1.10 (0.91-1.33)	0.323	4.83E-11
	6q22	*NCOA7, HEY2*	rs13328298	126,016,580	G	A	0.600	0.573	1.13 (1.08-1.17)	2.78E-09	0.00	0.616	0.569	1.22 (1.02-1.46)	0.032	3.73E-10
	8q24	*MYC*	rs4733613	129,599,278	G	C	0.856	0.877	0.84 (0.79-0.89)	5.64E-09	0.02	0.853	0.867	0.89 (0.74-1.08)	0.219	3.09E-09
	15q15	*EIF2AK, BMF*	rs937213	40,322,124	T	C	0.553	0.582	0.90 (0.86-0.93)	3.81E-08	0.41	0.549	0.570	0.92 (0.81-1.05)	0.524	1.77E-08
	14q32	*AKT1, SIVA1*	rs2498796	105,243,220	G	A	0.677	0.705	0.89 (0.85-0.93)	8.66E-08	0.02	0.681	0.702	0.91 (0.79-1.04)	0.177	3.55E-08

Positions in build 37; EA, Effect allele; OA, Other allele; EAF, effect allele frequency; I^2, heterogeneity I^2 statistic.

All SNPs are either typed or imputed with information score >0.9 in all arrays including iCOGS; phase 2 NSECG genotyping was performed using KASPar (rs4733613, rs937213, rs2498796) or Fluidigm (rs13328298, rs11841589)

Supplementary Table 4: Genotyping concordance rates for different platforms in quality control duplicates

Genotyping		Overlapping		% Concordance
Platform1	Platform2	Samples	SNPs	
KASPar	Fluidigm	491	2	0.988
KASPar	Illumina 660W	23	6	0.986
KASPar	Illumina iSelect	9	3	1.000

% Concordance: number of concordant genotypes were counted and divided by the total. Duplicate samples were excluded from all analysis.

Supplementary Table 5: Endometrial tissue eQTL: association between GWAS risk loci genotypes and transcript levels of nearby genes

Locus	GWAS SNP	Gene	GTEx Normal Uterus P n=70	TCGA Adjacent Normal RNASeq P n=30	TCGA Tumor RNASeq P n=458
13q22.1	rs11841589	*KLF5*	0.88	0.27	0.02
		KLF12	0.39	0.72	0.32
		PIBF1	0.80	0.83	0.96
		DIS3	0.52	0.17	0.36
		BORA	0.71	-	-
6q22.31	rs13328298	*HEY2*	0.14	0.73	0.01
		NCOA7	0.05	0.31	0.98
		HDDC2	-	0.12	0.24
		TPD52L1	0.83	0.005	0.14
		HINT3	0.35	0.57	0.41
		TRMT11	-	0.84	0.29
8q24.21	rs4733613	*MYC*	0.98	0.16	0.38
		PVT1	0.12	0.45	0.45
		POU5F1B	0.26	0.79	0.85
	rs17232730	*MYC*	0.51	0.17	0.23
		PVT1	0.65	0.75	0.87
		POU5F1B	0.05	0.64	0.15
15q15.1	rs937213	*EIF2AK4*	0.10	0.73	0.11
		SRP14	0.40	0.53	0.07
		BMF	0.90	0.27	0.61
		BUB1B	0.44	0.003	0.62
		PAK6	-	0.06	0.96
		GPR176	0.32	0.37	0.29
		THBS1	0.61	0.93	0.48
		FSIP1	0.42	0.62	0.48
		C15orf56	-	0.26	0.15
		PLCB2	0.52	0.38	0.05
		C15orf52	1.00	0.69	0.13
		PHGR1	0.51	-	-
		DISP2	0.12	0.58	0.19
		KNSTRN	-	-	-
		IVD	0.63	0.87	0.94
		BAHD1	0.15	0.52	0.03
		CHST14	0.86	0.01	0.57
14q32.33	rs2498796	*AKT1*	0.01	0.09	0.12
		SIVA1	0.06	0.02	0.50
		ZBTB42	0.49	0.15	0.008
		INF2	0.98	0.85	0.20
		ADSSL1	0.01	0.95	0.63
		PLD4	0.74	0.05	0.86
		TMEM179	-	-	-
		AHNAK2	0.29	0.27	0.61
		C14orf79	0.62	0.16	0.76
		CDCA4	0.44	0.71	0.76
		GPR132	0.26	0.46	0.66
		C14orf180	0.58	0.92	0.96
		JAG2	0.88	0.14	0.13
		NUDT14	0.89	0.10	0.43
		CEP170B	0.55	-	-
		BRF1	0.72	0.06	0.14
		BTBD6	0.89	0.20	0.93

GTEx, Genotype Tissue Expression Project; TCGA, The Cancer Genome Atlas (TCGA).

"-" denotes expression data not present in this database. See Online Methods.

After Bonferroni correction for number of tests (number of SNPs investigated = 6, number of sample sets=3, and number of genes assessed - up to 48), no associations were considered statistically signficant.

SNPs within 500kb of the top SNP were also analysed and we did not find any SNPs in these loci that were robustly associated with the expression of nearby genes.

Supplementary Table 6: Functional annotation of SNPs in LD with GWAS risk loci (r^2>0.7 in 1000 Genomes EUR) from Haploreg, RegulomeDB and ENCODE

Locus	Position (hg19)	LD	SNP	Ref	Alt	EUR AF	Histone marks		DHS	Bound proteins	Altered Motifs
							Promoter	Enhancer			
13q22	73,811,879	0.98	rs9600103	A	T	0.28					AP-1, Elf3, GR,PRDM1,SIX5
13q22	73,812,141	0.98	rs7981863	C	T	0.28					Rad21,SMC3
13q22	73,813,435	0.98	rs7988505	C	G	0.27					CDP,Foxa, Nanog,Sox
13q22	73,813,436	0.99	rs7989799	T	A	0.27					
13q22	73,813,982	0.75	rs9592895	T	C	0.33		NHEK			DMRT7,SIX5
13q22	73,814,891	-	**rs11841589**	G	T	0.28					Pou2f2
6q22	125,985,934	0.72	rs1832937	G	A	0.55					Sox
6q22	125,988,964	0.96	rs1832938	G	C	0.56					Hand1,Pou2f2, STAT
6q22	125,991,507	0.87	rs12527010	G	A	0.54		H1			9 altered motifs
6q22	125,991,715	0.97	rs9491471	C	T	0.56		H1	HRPEpiC		4 altered motifs
6q22	125,992,553	0.86	rs1418637	A	G	0.59		H1		MAFF,MAFK	5 altered motifs
6q22	125,992,810	0.97	rs1343120	A	G	0.56					4 altered motifs
6q22	125,993,202	0.97	rs4897151	T	G	0.56					Nr2f2
6q22	125,994,080	0.96	rs6940748	C	T	0.57			RWPE1		4 altered motifs
6q22	125,994,708	0.97	rs1935772	T	C	0.56					AP-1,Sin3Ak-20
6q22	125,994,753	0.97	rs202061058	C	CCAA	0.56					BRCA1,HNF1, STAT
6q22	125,995,134	0.97	rs6904069	G	A	0.56					VDR
6q22	125,995,467	0.84	rs2226158	G	A	0.53			HAc		LBP-1
6q22	125,995,503	0.93	rs2211418	G	A	0.55			FibroP,HAc		
6q22	125,995,533	0.97	rs2211419	A	G	0.56			FibroP		4 altered motifs
6q22	125,995,549	0.97	rs2211420	T	C	0.56			FibroP		STAT,Tel2
6q22	125,996,185	0.97	rs1418951	A	G	0.56					CTCF
6q22	125,996,475	0.93	rs1935773	A	G	0.55		H1	HRPEpiC		RREB-1,Sp4
6q22	125,996,661	0.93	rs1935774	C	T	0.55					CEBPB,CTCF,p300
6q22	125,997,436	0.93	rs1832979	A	G	0.55					
6q22	125,997,444	0.97	rs1832980	T	G	0.56					6 altered motifs

		6q22	125,998,186	0.97	rs6569435	T	C	0.56					
		6q22	125,998,286	0.96	rs11443856	T	TC	0.56					4 altered motifs
		6q22	125,999,674	0.86	rs2326292	G	A	0.59		H1			Crx,Gsc,Otx2
		6q22	125,999,768	0.97	rs1418642	G	A	0.56		H1			4 altered motifs
		6q22	125,999,854	0.97	rs1418641	C	T	0.56		H1			4 altered motifs
		6q22	125,999,866	0.97	rs1418640	G	A	0.56		H1			
		6q22	125,999,940	0.96	rs1418639	C	T	0.56		H1			HEN1
		6q22	126,000,162	0.97	rs4895798	A	G	0.56		H1			4 altered motifs
		6q22	126,000,599	0.97	rs8180614	G	C	0.56		H1			
		6q22	126,001,064	0.97	rs1954360	A	G	0.56		H1	Jurkat, SK-N-MC		
		6q22	126,001,423	0.97	rs1954361	C	G	0.56					SF1,STAT
		6q22	126,001,568	0.97	rs9321050	A	G	0.56					Arid5a,CEBPA, HDAC2
		6q22	126,002,400	0.97	rs4897152	A	G	0.56					4 altered motifs
		6q22	126,002,774	0.97	rs1935979	A	G	0.56					Ik-1
		6q22	126,003,403	0.97	rs4897153	A	G	0.56					GR
		6q22	126,003,603	0.96	rs201652751	C	CCT	0.56					p300
		6q22	126,004,124	0.96	rs9401843	T	C	0.56					4 altered motifs
		6q22	126,004,193	0.82	rs80303782	C	T	0.52			H1-hESC	POL24H8	HMG-IY,Pax-4
		6q22	126,004,194	0.86	rs76407388	A	G	0.53			H1-hESC	POL24H8	Pax-4
		6q22	126,004,197	0.94	rs28629380	A	G	0.56			H1-hESC	POL24H8	Pax-4
		6q22	126,004,468	0.94	rs5879788	T	TC,TT	0.56					
		6q22	126,004,720	0.97	rs1739367	G	T	0.56				FOXA1,GATA3	5 altered motifs
		6q22	126,004,883	0.97	rs1777194	G	A	0.56					18 altered motifs
		6q22	126,004,935	0.97	rs2747724	G	A	0.56					Obox6
		6q22	126,005,197	0.96	rs2797154	A	G	0.56					
		6q22	126,005,310	0.97	rs1739352	T	C	0.56					4 altered motifs
		6q22	126,005,767	0.97	rs983543	G	A	0.56					12 altered motifs
		6q22	126,006,861	0.95	rs1777195	A	C	0.44					7 altered motifs
		6q22	126,007,018	0.97	rs1418948	C	T	0.44					Ets
		6q22	126,007,401	0.95	rs1777197	G	A	0.43					11 altered motifs
		6q22	126,007,409	0.95	rs1739366	T	C	0.43					7 altered motifs
		6q22	126,007,416	0.93	rs1777198	C	T	0.44					4 altered motifs
		6q22	126,007,620	0.97	rs2747714	A	G	0.44					Ets
		6q22	126,007,719	0.94	rs2747715	A	T	0.44					5 altered motifs
		6q22	126,007,996	0.97	rs78229684	C	T	0.44					5 altered motifs
		6q22	126,008,435	0.98	rs2747717	A	G	0.44					Pou2f2,Pou3f3,Sox
		6q22	126,009,109	0.98	rs2747718	C	A	0.44					6 altered motifs

		6q22	126,009,214	0.98	rs2747719	C	T	0.44				Pax-5,TCF4
		6q22	126,009,398	0.98	rs2797158	G	A	0.44				BATF	AP-1,Hbp1, Mrg1::Hoxa9
		6q22	126,009,458	0.98	rs2747720	A	G	0.44			GM12878	BATF	4 altered motifs
		6q22	126,009,527	0.98	rs2747721	A	G	0.44			GM12878	BATF	9 altered motifs
		6q22	126,009,557	0.97	rs2797159	G	A	0.44			GM12878		
		6q22	126,009,629	0.98	rs2747722	A	G	0.44					4 altered motifs
		6q22	126,010,086	0.99	rs35069021	AT	A	0.44					5 altered motifs
		6q22	126,010,116	0.99	rs2797160	A	G	0.44					6 altered motifs
		6q22	126,010,789	0.92	rs2797161	A	G	0.42					GZF1,Smad4
		6q22	126,010,790	0.92	rs2747723	C	T	0.42					RFX5,Smad4
		6q22	126,010,904	0.99	rs34649676	CA	C	0.44					CIZ,HNF1,Pou5f1
		6q22	126,011,079	0.99	rs1739368	C	T	0.44					5 altered motifs
		6q22	126,011,231	0.88	rs1739370	C	T	0.41					8 altered motifs
		6q22	126,011,291	0.99	rs1739371	G	A	0.44					
		6q22	126,011,325	0.99	rs1739372	G	A	0.44					4 altered motifs
		6q22	126,011,381	0.99	rs2797162	G	T	0.44					Pou3f2,p300
		6q22	126,011,509	0.99	rs1739373	A	G	0.44					CHOP::CEBPalpha, Pou2f2,Sp4
		6q22	126,011,825	0.97	rs1739374	C	T	0.44					10 altered motifs
		6q22	126,011,995	0.97	rs1777183	A	C	0.44					8 altered motifs
		6q22	126,012,013	0.97	rs1739375	T	C	0.44					7 altered motifs
		6q22	126,012,084	0.97	rs1739376	C	G	0.44					4 altered motifs
		6q22	126,012,236	0.97	rs1739377	T	C	0.44					Arid5b,HDAC2,Pou5f1
		6q22	126,012,262	0.99	rs1739378	C	A	0.44					Hsf,TEF-1
		6q22	126,012,397	0.97	rs2747725	T	G	0.44					
		6q22	126,012,593	0.97	rs1739379	T	C	0.44					Pou6f1
		6q22	126,012,858	0.97	rs1739380	T	C	0.44					5 altered motifs
		6q22	126,013,155	0.97	rs1630556	A	G	0.44				SETDB1	10 altered motifs
		6q22	126,013,614	0.97	rs1777182	T	A	0.44					6 altered motifs
		6q22	126,014,157	0.99	rs1739347	C	T	0.44					Ascl2,Myc,Myf
		6q22	126,014,573	0.99	rs1739348	T	C	0.44					PU.1
		6q22	126,014,907	0.88	rs1612274	C	A	0.41				POL2	4 altered motifs
		6q22	126,014,916	0.98	rs1612249	C	A,T	0.44					
		6q22	126,014,984	0.99	rs1739349	G	C	0.44					Ik-3,Pax-4,Sox
		6q22	126,015,057	0.99	rs1578793	A	G	0.44					4 altered motifs
		6q22	126,015,469	0.99	rs1578794	T	C	0.44					4 altered motifs
		6q22	126,015,954	0.99	rs6927161	T	C	0.44					Foxc1,GATA,Mef2

		6q22	126,016,003	0.99	rs6904992	G	A	0.44					Myc,Osr
		6q22	126,016,499	0.99	rs12717178	G	A	0.44					6 altered motifs
		6q22	126,016,580	-	**rs13328298**	G	A	0.44					14 altered motifs
		6q22	126,016,951	0.99	rs6933302	T	C	0.44					PRDM1,YY1
		6q22	126,017,029	0.99	rs6933471	T	G	0.44					DMRT5,HNF1,Irf
		6q22	126,017,141	0.98	rs6910786	A	T	0.44					9 altered motifs
		6q22	126,017,155	0.99	rs6910933	C	G	0.44					4 altered motifs
		6q22	126,017,481	0.99	rs6934435	T	G	0.44					16 altered motifs
		6q22	126,017,551	0.99	rs201940333	T	TC	0.44					10 altered motifs
		6q22	126,017,691	0.99	rs1777226	C	A	0.44			H1-hESC		BAF155,SP1,WT1
		6q22	126,017,808	0.99	rs1739354	G	C	0.44					CTCF,ERalpha-a,RXRA
		6q22	126,018,114	0.99	rs1739355	A	G	0.44					
		6q22	126,018,270	0.99	rs1777225	T	C	0.44					E2A,LBP-1
		6q22	126,019,527	0.96	rs1777224	T	C	0.45					BATF,E2F,Irf
		6q22	126,019,655	0.96	rs1739357	T	G	0.45					EBF,SP1
		6q22	126,019,736	0.98	rs1739358	G	A	0.44					
		6q22	126,019,738	0.95	rs200274442	A	ACG	0.43					4 altered motifs
		6q22	126,019,738	0.98	rs77678056	A	G	0.44					
		6q22	126,019,768	0.98	rs78602343	T	C	0.43					Foxp1,Nkx3
		6q22	126,020,703	0.98	rs1739362	A	T	0.44					Foxj2,Maf
		6q22	126,020,980	0.98	rs1739363	G	A	0.44		H1	H1-hESC, CD20+	EBF1	Pou3f3
		6q22	126,021,030	0.98	rs1777222	C	T	0.44		H1	H1-hESC, CD20+	EBF1	Pax-4
		6q22	126,021,277	0.98	rs984040	T	C	0.44					Foxp1
		6q22	126,021,328	0.98	rs984041	A	T	0.44					Pou2f2
		6q22	126,021,435	0.98	rs926853	A	T	0.44					GATA
		6q22	126,021,780	0.96	rs926854	A	G	0.43		H1	RPTEC		4 altered motifs
		6q22	126,021,782	0.96	rs926855	A	G	0.43		H1	RPTEC		HP1-site-factor,Mef2
		6q22	126,022,383	0.91	rs1739364	G	A	0.44		H1	4 cell types	JUND,POL2,POL24H8	
		6q22	126,022,602	0.81	rs1777220	G	T	0.41		H1			
		6q22	126,025,819	0.7	rs9491500	A	G	0.39	H1				4 altered motifs
		6q22	126,027,318	0.71	rs6939865	A	C	0.4	H1				11 altered motifs
		6q22	126,027,833	0.71	rs144251392	G	A	0.4	H1				
		6q22	126,029,043	0.74	rs1268092	C	T	0.41		H1	H1-hESC		5 altered motifs
		6q22	126,029,235	0.74	rs1268093	G	A	0.41		H1			5 altered motifs
		6q22	126,029,682	0.77	rs1269176	T	A	0.41		H1			7 altered motifs

		6q22	126,031,682	0.77	rs9491503	G	A	0.41					En-1
		6q22	126,032,104	0.73	rs6569436	A	G	0.42			RWPE1		Fox,HP1-site-factor,Nrf-2
		6q22	126,034,540	0.73	rs6569437	T	G	0.41			HAc		6 altered motifs
		6q22	126,034,563	0.73	rs6939969	C	T	0.41			HAc		GR,Pax-4,RP58
		6q22	126,035,041	0.76	rs1268066	C	T	0.41		HBMEC,HNPCE		GATA2	Hoxa7
		6q22	126,036,184	0.76	rs1343121	C	T	0.41			GM12878		4 altered motifs
		6q22	126,036,621	0.73	rs1268067	T	C	0.41					4 altered motifs
		8q24	129,599,278	-	**rs4733613**	C	G	0.85		HepG2			
		8q24	129,497,735	0.71	rs77156523	T	A	0.09					Znf143
		8q24	129,499,432	0.78	rs77241108	G	A	0.09					4 altered motifs
		8q24	129,501,028	0.79	rs113834169	CT	C	0.09		HSMM, NHLF, HepG2	Th1	USF2,USF1	4 altered motifs
		8q24	129,501,760	0.79	rs76384007	C	T	0.09		HepG2		MAFK	Hand1,RXRA,VDR
		8q24	129,502,569	0.79	rs10505515	A	G	0.09		HepG2	HRPEpiC		Foxa,Mef2
		8q24	129,503,700	0.79	rs17805799	T	C	0.09					10 altered motifs
		8q24	129,506,822	0.79	rs77683961	G	A	0.09		HepG2			
		8q24	129,507,019	0.79	rs75706585	C	T	0.09					Nr2e3
		8q24	129,508,906	0.79	rs17231703	G	C	0.09					
		8q24	129,514,190	0.79	rs17805924	A	C	0.09		HepG2			SP1
		8q24	129,517,137	0.79	rs17231857	T	C	0.09					4 altered motifs
		8q24	129,517,554	0.79	rs17231948	A	G	0.09					Ets
		8q24	129,517,938	0.79	rs76388889	T	C	0.09					Cdx,DMRT1
		8q24	129,526,064	0.81	rs17232172	C	T	0.09		HepG2, H1, NHEK			E2F,Smad3
		8q24	129,529,129	0.81	rs1878378	T	A	0.09					HNF4
		8q24	129,530,492	0.81	rs1400485	A	G	0.09					Hoxb13
		8q24	129,537,624	0.88	rs79593486	A	C	0.1		NHLF			10 altered motifs
		8q24	129,537,746	-	**rs17232730**	G	C	0.09		NHLF	8 cell types		
		8q24	129,545,257	0.76	rs1516973	C	T	0.08		HepG2, HSMM			
		8q24	129,546,233	0.76	rs1607120	G	A	0.08		NHLF	7 cell types		6 altered motifs
		8q24	129,546,982	0.76	rs10086718	T	C	0.08					11 altered motifs
		8q24	129,548,654	0.76	rs74866331	T	C	0.08					TLX1::NFIC
		8q24	129,549,261	0.76	rs1516977	A	G	0.08				ERALPHA_A	7 altered motifs
		8q24	129,550,041	0.76	rs10112057	G	A	0.08			Th1,Th2		4 altered motifs
		8q24	129,551,752	0.76	rs10088873	G	A	0.08		HSMM, HepG2			Pbx-1
		8q24	129,552,140	0.76	rs10098999	C	T	0.08		4 cell types	4 cell types	HNF4A,HNF4G,RXRA	4 altered motifs

		8q24	129,555,724	0.72	rs2138636	C	T	0.08		NHLF, HSMM	9 cell types		5 altered motifs
		8q24	129,555,928	0.73	rs17807904	G	A	0.08		HSMM	Th1,Adult_CD4_Th0		Myc,RBP-Jkappa,Zfx
		8q24	129,556,514	0.73	rs10087367	G	A	0.08					7 altered motifs
		8q24	129,559,228	0.73	rs10098821	C	T	0.08					7 altered motifs
		8q24	129,560,314	0.72	rs10102835	C	T	0.08					17 altered motifs
		8q24	129,562,824	0.73	rs76076434	G	A	0.08					5 altered motifs
		8q24	129,563,362	0.72	rs78830272	T	C	0.08					5 altered motifs
		8q24	129,565,545	0.73	rs75370373	G	T	0.08		HepG2, K562	K562, Jurkat, Th2	E2F6,MAX	Egr-1,SETDB1,Znf143
		8q24	129,566,898	0.73	rs7007074	C	T	0.08		HepG2, K562, NHLF			11 altered motifs
		15q15	40,290,475	0.71	rs199884855	TG	T	0.58					19 altered motifs
		15q15	40,302,243	0.72	rs3816900	C	T	0.55		HSMM			FXR,HNF1,RORalpha1
		15q15	40,302,441	0.72	rs12592831	G	A	0.55		HSMM	BE2_C		Myf
		15q15	40,303,842	0.72	rs2412464	G	A	0.55		HSMM			Sin3Ak-20
		15q15	40,322,124	-	**rs937213**	T	C	0.44			FibroP		7 altered motifs
		14q32	105,192,685	0.74	rs72715972	G	A	0.32	HepG2		HMEC, Fibrobl, HL-60		5 altered motifs
		14q32	105,196,230	0.79	rs80097179	A	C	0.31		K562		4 bound proteins	Znf143
		14q32	105,197,354	0.78	rs60876857	C	G	0.32		K562			Roaz
		14q32	105,197,756	0.79	rs60798007	G	C	0.31		HepG2	40 cell types	9 bound proteins	9 altered motifs
		14q32	105,197,846	0.77	rs57098433	G	A	0.32		HepG2	30 cell types	8 bound proteins	NF-E2,PLZF
		14q32	105,203,678	0.79	rs7160733	G	A	0.32		Huvec, HepG2			SP1,ZEB1
		14q32	105,208,057	0.8	rs4983384	C	T	0.32					5 altered motifs
		14q32	105,218,333	0.76	rs8006580	G	A	0.33		HepG2, GM12878, HMEC	GM06990	SETDB1	Hic1
		14q32	105,222,037	0.79	rs1132975	C	T	0.33	HepG2	GM12878			RXRA
		14q32	105,226,075	0.79	rs45607139	C	T	0.33					17 altered motifs
		14q32	105,228,216	0.8	rs4983549	G	A	0.33					7 altered motifs
		14q32	105,229,646	0.85	rs7158655	T	C	0.32	HepG2				13 altered motifs
		14q32	105,230,225	0.84	rs144188214	TTTTA	T	0.32	HepG2				34 altered motifs
		14q32	105,230,391	0.74	rs28368454	T	C	0.33		HepG2			HMG-IY,NRSF
		14q32	105,231,196	0.79	rs66464514	A	G	0.33	HepG2				RP58,Sox,TAL1

	14q32	105,231,640	0.77	rs4983550	T	G	0.35	HepG2				Hltf,Sox
	14q32	105,233,095	0.85	rs2498804	C	A	0.32	HepG2	H1, K562	HepG2	POL2	Nkx2,Nkx3
	14q32	105,233,408	0.85	rs2498803	G	A	0.32		HepG2, K562	H7-hESC	7 bound proteins	CCNT2, NF-kappaB, UF1H3BETA
	14q32	105,233,421	0.85	rs2494730	T	A	0.32		HepG2, K562	H7-hESC	7 bound proteins	6 altered motifs
	14q32	105,234,442	0.78	rs2498802	G	C	0.34		HepG2	Fibrobl		5 altered motifs
	14q32	105,235,558	0.79	rs2498801	T	C	0.34	HepG2	K562, H1	6 cell types	CTCF,AP2ALPHA,POL2	5 altered motifs
	14q32	105,237,680	0.82	rs2494731	G	C	0.34				SETDB1	Pax-5,RXRA
	14q32	105,238,591	0.87	rs55839843	CCTGAG	C	0.32			PrEC		4 altered motifs
	14q32	105,240,784	0.89	rs2494733	C	G	0.33				ERALPHA_A,CTCF	8 altered motifs
	14q32	105,242,228	0.85	rs2498797	T	C	0.34			4 cell types	ERALPHA_A,POL2,CMYC	6 altered motifs
	14q32	105,242,831	0.98	rs3001371	C	T	0.31				ZNF263	RREB-1
	14q32	105,242,966	0.77	rs2494735	T	C	0.37				ZNF263	5 altered motifs
	14q32	105,243,220	-	**rs2498796**	G	A	0.31					4 altered motifs
	LD, r^2 with lead SNP; Ref,reference allele; alt, alternative allele; EUR AF, reference allele frequency in Europeans; DHS, DNase1 hypersensitivity site; TFBS, Transcription											
	GWAS hits are noted in bold											
	RegulomeDB scores:											
		1b: eQTL + TF binding + any motif + DNase Footprint + DNase peak										
		1c: eQTL + TF binding + matched TF motif + DNase peak										
		1d: eQTL + TF binding + any motif + DNase peak										
		1e: eQTL + TF binding + matched TF motif										
		1f: eQTL + TF binding / DNase peak										
		2a: TF binding + matched TF motif + matched DNase Footprint + DNase peak										
		2b: TF binding + any motif + DNase Footprint + DNase peak										
		2c: TF binding + matched TF motif + DNase peak										
		3a: TF binding + any motif + DNase peak										
		3b: TF binding + matched TF motif										
		4: TF binding + DNase peak										
		5: TF binding or DNase peak										
		6: other										
	Phastcons derived from hg19_phastConsElements46way.txt using annovar											

GENCODE genes	dbSNP annotation	Regulome DB score	PhastCons
10kb 3' of RNY1P8		5	
11kb 3' of RNY1P8			
12kb 3' of RNY1P8			
12kb 3' of RNY1P8			
13kb 3' of RNY1P8			
13kb 3' of RNY1P8			
RP11-624M8.1			
RP11-624M8.1		5	
RP11-624M8.1		6	
RP11-624M8.1		5	Score=451;Name=lod=91
RP11-624M8.1		5	
RP11-624M8.1		6	
RP11-624M8.1			
RP11-624M8.1		5	
RP11-624M8.1			
RP11-624M8.1			
RP11-624M8.1		5	
RP11-624M8.1		5	
RP1-293L8.2		5	
RP1-293L8.2		5	
RP1-293L8.2		5	
RP1-293L8.2	intronic	5	
RP1-293L8.2	intronic	3a	
RP1-293L8.2	intronic	5	
RP1-293L8.2	intronic	5	
RP1-293L8.2	intronic	5	

RP1-293L8.2	intronic		
RP1-293L8.2	intronic		
RP1-293L8.2	intronic	5	
RP1-293L8.2	intronic	5	
RP1-293L8.2	intronic	4	
RP1-293L8.2	intronic	3a	
RP1-293L8.2	intronic	3a	
RP1-293L8.2	intronic	5	
RP1-293L8.2	intronic		
RP1-293L8.2	intronic	5	Score=497;Name=lod=139
RP1-293L8.2	intronic		
RP1-293L8.2	intronic		
RP1-293L8.2	intronic	6	
RP1-293L8.2	intronic	6	
RP1-293L8.2	intronic		
RP1-293L8.2	intronic		
RP1-293L8.2	intronic		Score=272;Name=lod=17
RP1-293L8.2	intronic		
RP1-293L8.2	intronic		
RP1-293L8.2	intronic	3a	
RP1-293L8.2	intronic		
RP1-293L8.2	intronic	3a	
RP1-293L8.2	intronic	5	
RP1-293L8.2	intronic		
RP1-293L8.2	intronic		
RP1-293L8.2	intronic		
RP1-293L8.2	intronic	6	
RP1-293L8.2	intronic	6	
RP1-293L8.2	intronic		
RP1-293L8.2	intronic	6	
RP1-293L8.2	intronic	6	
RP1-293L8.2	intronic	6	
RP1-293L8.2	intronic		
RP1-293L8.2	intronic	6	
RP1-293L8.2	intronic		
RP1-293L8.2	intronic	6	
RP1-293L8.2	intronic	6	

RP1-293L8.2	intronic		
RP1-293L8.2	intronic	5	
RP1-293L8.2	intronic	3a	
RP1-293L8.2	intronic	4	
RP1-293L8.2	intronic	3a	
RP1-293L8.2	intronic		
RP1-293L8.2	intronic	6	
RP1-293L8.2	intronic		
RP1-293L8.2	intronic		
RP1-293L8.2	intronic		
RP1-293L8.2	intronic	6	
RP1-293L8.2	intronic	6	
RP1-293L8.2	intronic	6	
RP1-293L8.2	intronic		
RP1-293L8.2	intronic		
RP1-293L8.2	intronic		
RP1-293L8.2	intronic	6	
RP1-293L8.2	intronic	5	
RP1-293L8.2	intronic	5	
RP1-293L8.2	intronic	5	
RP1-293L8.2	intronic	6	
RP1-293L8.2	intronic		
RP1-293L8.2	intronic		
RP1-293L8.2	intronic	6	
RP1-293L8.2	intronic		
RP1-293L8.2	intronic	5	
RP1-293L8.2	intronic	5	
RP1-293L8.2	intronic		
RP1-293L8.2	intronic	6	
RP1-293L8.2	intronic		
RP1-293L8.2	intronic	3a	
RP1-293L8.2	intronic	4	
RP1-293L8.2	intronic	5	
RP1-293L8.2	intronic	5	
RP1-293L8.2	intronic	5	
RP1-293L8.2	intronic	6	

RP1-293L8.2	intronic		
RP1-293L8.2	intronic		
RP1-293L8.2	intronic		
RP1-293L8.2	intronic		
RP1-293L8.2	intronic	6	
RP1-293L8.2	intronic	6	
RP1-293L8.2	intronic	6	
RP1-293L8.2	intronic	6	
RP1-293L8.2	intronic	6	
RP1-293L8.2	intronic	5	
RP1-293L8.2	intronic		
RP1-293L8.2	intronic	6	
RP1-293L8.2	intronic		
RP1-293L8.2	intronic	6	
RP1-293L8.2	intronic	6	
RP1-293L8.2	intronic		
RP1-293L8.2	intronic		
RP1-293L8.2	intronic		
RP1-293L8.2	intronic		
RP1-293L8.2	intronic		
RP1-293L8.2	intronic	4	
RP1-293L8.2	intronic	4	
RP1-293L8.2	intronic		
RP1-293L8.2	intronic		
RP1-293L8.2	intronic		
RP1-293L8.2	intronic	5	Score=321;Name=lod=27
RP1-293L8.2	intronic	5	Score=321;Name=lod=27
RP1-293L8.2	intronic	4	Score=346;Name=lod=34
RP1-293L8.2	intronic	5	
RP1-293L8.2	intronic	6	
RP1-293L8.2	intronic	6	
RP1-293L8.2	intronic		
RP1-293L8.2	intronic	5	Score=608;Name=lod=393
RP1-293L8.2	intronic		Score=343;Name=lod=33
RP1-293L8.2	intronic	6	

RP1-293L8.2	intronic	6	
RP1-293L8.2	intronic	3a	
RP1-293L8.2	intronic	5	
RP1-293L8.2	intronic	5	
RP1-293L8.2	intronic	4	
RP1-293L8.2	intronic	6	
RP1-293L8.2	intronic	3a	
22kb 5' of RP11-89M16.1			
RP11-89M16.1			
RP11-89M16.1		6	
RP11-89M16.1		3a	
RP11-89M16.1		3a	
RP11-89M16.1		5	
RP11-89M16.1		6	
RP11-89M16.1		5	
RP11-89M16.1		5	
RP11-89M16.1			
RP11-89M16.1		5	
RP11-89M16.1		5	
RP11-89M16.1			
RP11-89M16.1			
RP11-89M16.1			
RP11-89M16.1			
RP11-89M16.1			
RP11-89M16.1		3b	
RP11-89M16.1		4	
RP11-89M16.1			
RP11-89M16.1		5	
RP11-89M16.1		6	
RP11-89M16.1		5	
RP11-89M16.1		2b	Score=669;Name=lod=696
RP11-89M16.1		5	
RP11-89M16.1			
RP11-89M16.1		4	

RP11-89M16.1		4	
RP11-89M16.1		5	
RP11-89M16.1			
RP11-89M16.1			
RP11-89M16.1		6	
RP11-89M16.1		6	
RP11-89M16.1		6	
RP11-89M16.1		3a	
RP11-89M16.1		5	Score=309;Name=lod=24
EIF2AK4	intronic		
EIF2AK4	intronic		
EIF2AK4	intronic	5	
EIF2AK4	intronic		
EIF2AK4	intronic	5	
ADSSL1	intronic	2b	
ADSSL1	missense	4	
ADSSL1	intronic	5	
ADSSL1	intronic	4	
ADSSL1	intronic	4	Score=317;Name=lod=26
ADSSL1	intronic		
ADSSL1	intronic	1f	
1.1kb 5' of *SIVA1*		4	Score=498;Name=lod=141
SIVA1	synonymous	6	
SIVA1		2b	
SIVA1		6	
SIVA1		5	
SIVA1		6	
SIVA1		6	
SIVA1		6	

SIVA1			
SIVA1		1b	
SIVA1		4	
SIVA1		1b	
SIVA1		1f	
127bp 5' of RP11-982M15.2		4	
AKT1	intronic	1f	
AKT1	intronic		
AKT1	intronic	2b	
AKT1	intronic	2b	
AKT1	intronic	1b	
AKT1	intronic	4	
AKT1	intronic	5	
Factor Binding Sites			

Supplementary Table 7: Pairwise t-test P-values for 13q22 luciferase assays

	rs11841589-T	rs11841589-G	rs9600103-T	rs9600103-A	pGL3
rs11841589-T	-	0.058	0.128	0.045	0.008
rs11841589-G	-	-	0.979	0.039	0.013
rs9600103-T	-	-	-	0.018	0.014
rs9600103-A	-	-	-	-	0.226
pGL3	-	-	-	-	-

P-values from paired t-test comparisons using average of biological replicates.
Results unadjusted for multiple comparisons.

Supplementary Table 8: Primers and oligos used in experimental procedures

	Locus	Primer name	Forward sequence	Reverse sequence
SYBR-Green PCR	13q22	13q22_n2_rs10047671	TGCCATGGTTATTCTGAGCTTC	ACCATCTTCATTTTGTGTGGCT
	13q22	13q22_n1_rs9564955	ATGCAATGCCACAATCTCGG	CTGGGCAATGTAAAGGGACC
	13q22	13q22_0_rs56207966	TGCCACCCACCATACTTACT	AACCCAGGAGTCAGAGGTTG
	13q22	13q22_1_rs78441730	AGCCCATCTCCTCCATGAAG	ACTGGAAGTGAAGGTGGAGT
	13q22	13q22_2_rs6562748	GGCCAGGCTTGATCTCCTAA	CTCTCTTGGACCTGGCCAC
	13q22	13q22_3_rs7996454	TCTTGGGCTCGAACAATCCT	GTGACAGAGTGAGAGCCTGT
	13q22	13q22_4_rs9600103	CTGTGGTAGAGATTGATGATGGT	ACTCACTCCGGGAAAGAAGA
	13q22	13q22_5_rs12429143	AATGGTTTGGTTCAGGGTTCT	CATGATCGTGCCACTCCAG
	13q22	13q22_6_rs9592895	TGTCTGATGGCTGAAAGAAATCC	GGTGGGAAAGGTGCAAATGA
	13q22	13q22_7_rs11841589	GTTGGCCAGTGTGTGCTC	GAGACAGCGTTTCACCATGT
	13q22	13q22_8_rs7995365	CCCTGCAGGCTATAGTTATTGAG	TGCCTATCTTCAGTAATGACCAA
	13q22	13q22_9_rs8001970	AGGCCACAGACATCATGGAA	TTCCAGAAGAATCCCTGGGG
	13q22	13q22_10_rs9543288	GTGAAGGTTACAGTGAGCCG	TGTGGAAATTTTAGGCCTCTGA
	13q22	13q22_11_rs7328804	CATCACTGTGGCAAGGGAAC	GGCCTATGAAGTGTGAAGCG
	13q22	13q22_12_rs186045575	GGGCCTTGCACTAAAAGATGT	ACAATCCGCGCTATAATGCC
	12p13	GAPDH	GAGCCTCGAGGAGAAGTTCC	ACGACTGAGATGGGGAATTG
	3q22	RHO	TAACTTGTGGGGGAACGAAC	ATGGGCCTCTGTGCTATGTC
Cloning and Luciferase assay	13q22	Luc_rs9600103	GGCTGAGCTCAAACTCTTGG	GGGGCATAAAATTAACTTGACA
	13q22	Luc_rs11841589	GGAGACAGGTATGTTGAAAGCTC	ACAGATTTGTGCCACCACAC
	13q22	SDM_rs9600103_A	GAAAAATCATTTCCCAGTTCTCTATCCATTCATATTGTTGTGG	CCACAACAATATGAATGGATAGAGAACTGGGAAATGATTTTTC
	13q22	SDM_rs9600103_T	GAAAAATCATTTCCCTGTTCTCTATCCATTCATATTGTTGTGG	CCACAACAATATGAATGGATAGAGAACAGGGAAATGATTTTTC
	13q22	SDM_rs11841589_G	CATTTTATATACATCTTAACAGATTCTAGTCAGGCACAGTGGC	GCCACTGTGCCTGACTAGAATCTGTTAAGATGTATATAAAATG
	13q22	SDM_rs11841589_T	CATTTTATATACATCTTAACATATTCTAGTCAGGCACAGTGGC	GCCACTGTGCCTGACTAGAATATGTTAAGATGTATATAAAATG
KASPar genotyping	8q25	KASP_rs4733613	GAAGGTGACCAAGTTCATGCTGTGAGTCAATTAAACCTTTTTCTCAGC	TTGTATTAGTCTGCTCTCACACTGCTAAT
			GAAGGTCGGAGTCAACGGATTGTGAGTCAATTAAACCTTTTTCTCAGG	TCTGCTCTCACACTGCTAATAAAGACATA
	15q15	KASP_937213	GAAGGTGACCAAGTTCATGCTGCCAAATATGTCTTCTCATGGCCTTA	GCCTTTGCTTTCATCACCTCTAGCAA
			GAAGGTCGGAGTCAACGGATTCCAAATATGTCTTCTCATGGCCTTG	TTCATCACCTCTAGCAACCTTACTTCTTT
	14q32	KASP_rs2498796	GAAGGTGACCAAGTTCATGCTCATACCACCCACCAGGTCCTG	GGAGAGAGGAAGAGATGGGGCTT
			GAAGGTCGGAGTCAACGGATTCATACCACCCACCAGGTCCTA	AGAAACTGAGGCTTGGAGAGAGGAA
			Specific Target Amplification / Locus Specific Primer	Allele Specific Primer 1&2
Fluidigm genotyping	13q22	rs11841589	CAAAGTGCTGGGATTACAGGT	GCCACTGTGCCTGACTAGAATC
			TGGCCAGTGTGTGCTCAAGA	GCCACTGTGCCTGACTAGAATA
	6q22	rs13328298	ACAAATGGCCAAGAGATACATGAAA	GAAAGGTGGTCAACATCGCTAATT
			ACCTGCCTTTTTAGCAGGTATAAGGT	GAAAGGTGGTCAACATCGCTAATC
	15q15	rs937213	GTGCTAACAGAAGAGAATGGGAAA	GCCAAATATGTCTTCTCATGGCCTTA
			TGCTTTCATCACCTCTAGCAACCT	CCAAATATGTCTTCTCATGGCCTTG

Fig. 1

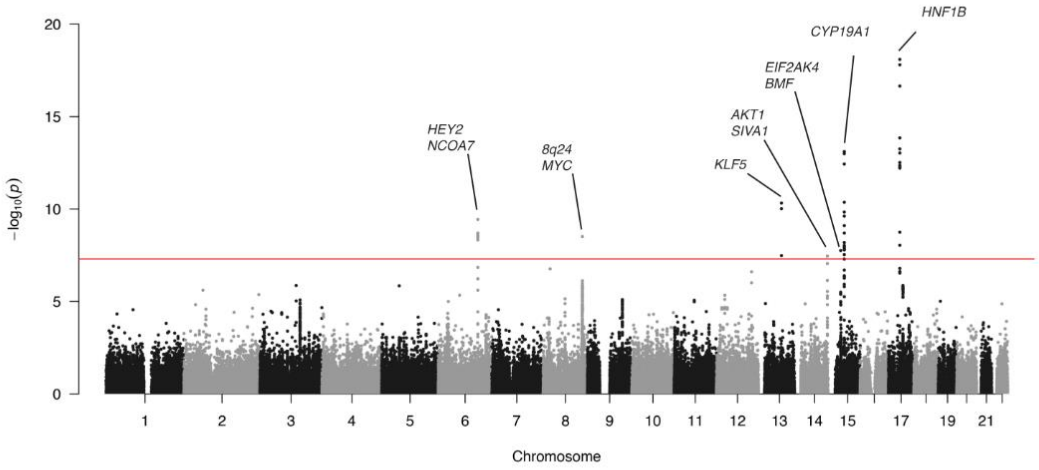

HNF1B
CYP19A1
EIF2AK4
BMF
AKT1
SIVA1
KLF5
HEY2
NCOA7
8q24
MYC
−log10(p)
Chromosome

Fig. 2

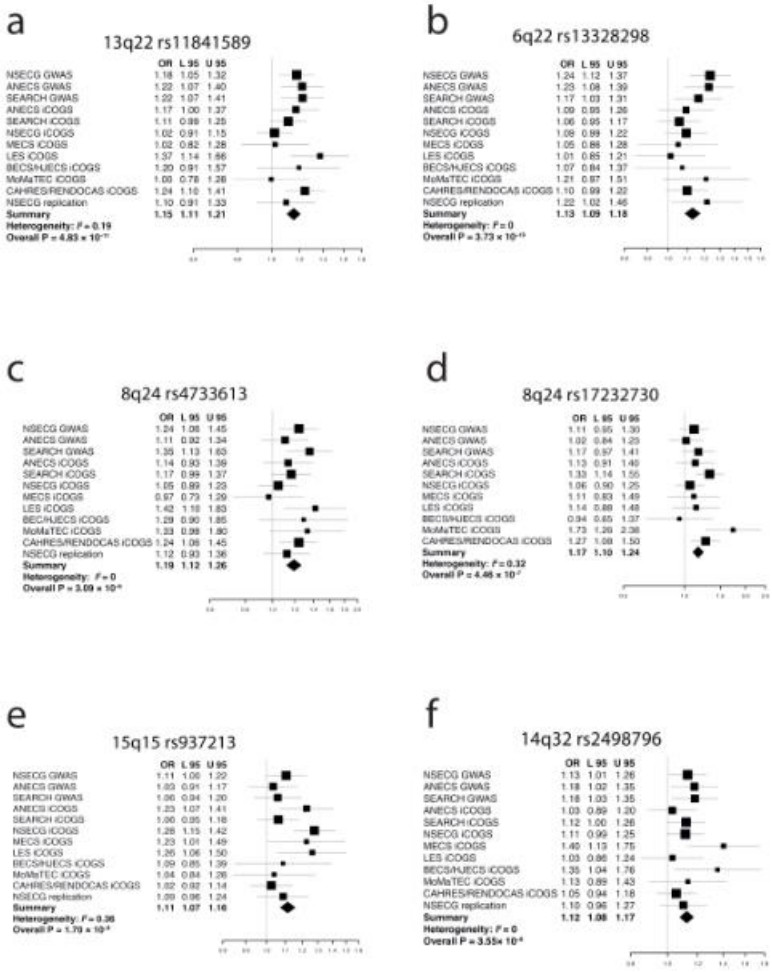

a
13q22 rs11841589
b
6q22 rs13328298
c
8q24 rs4733613
d
8q24 rs17232730
e
15q15 rs937213
f
14q32 rs2498796

Fig. 3

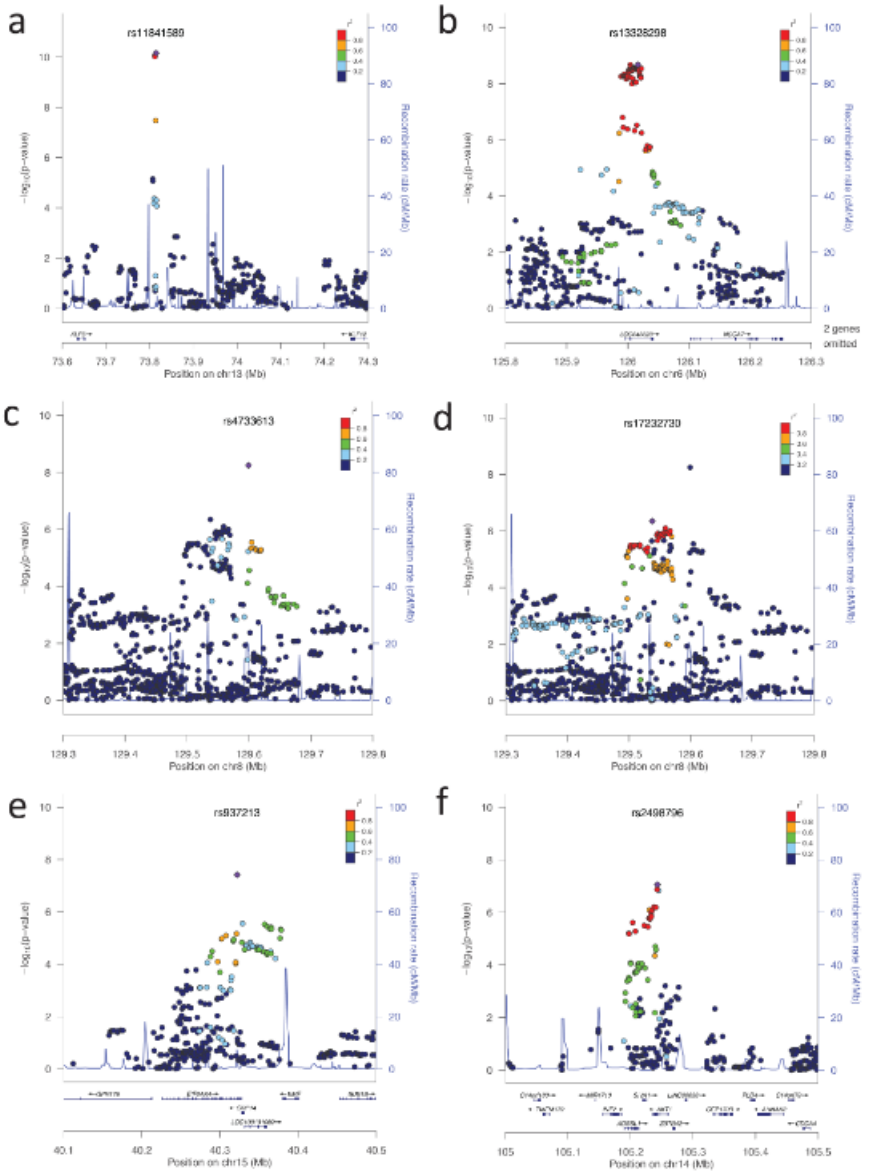
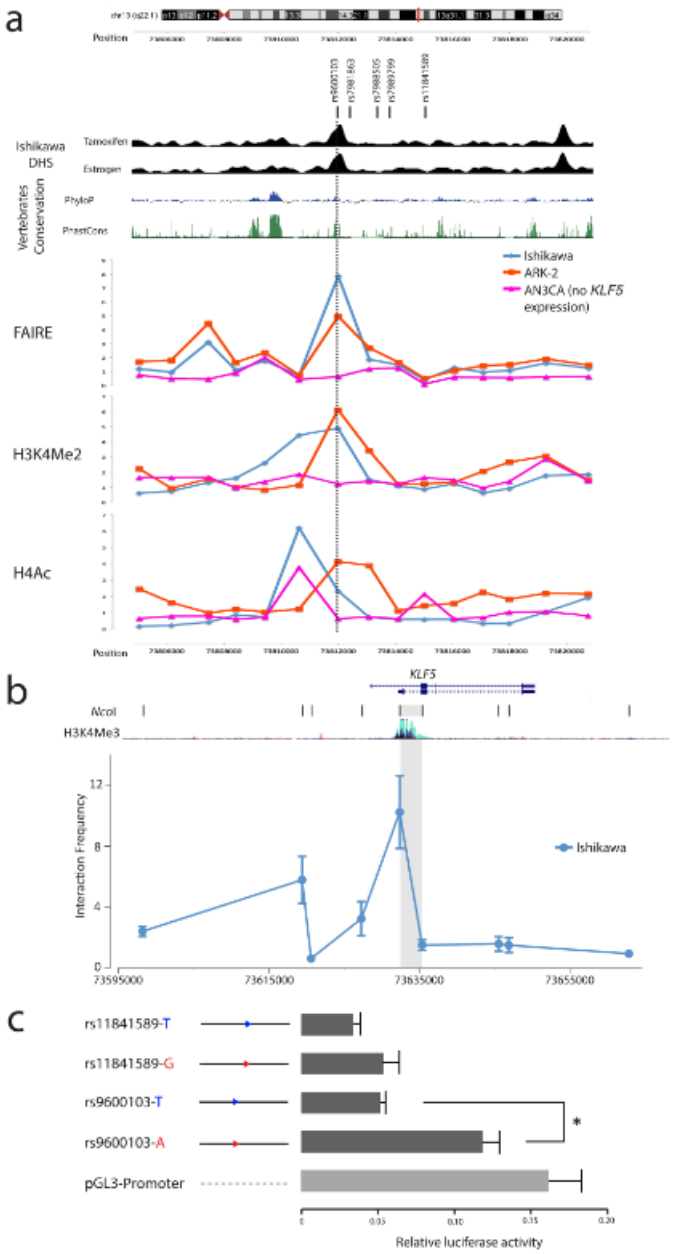

a
rs11841589
b
rs13328298
2 genes omitted
c
rs4733613
d
rs17232730
e
rs937213
f
rs2498796
Position on chr13 (Mb)
Position on chr6 (Mb)
Position on chr8 (Mb)
Position on chr15 (Mb)
Position on chr14 (Mb)
Recombination rate (cM/Mb)

Fig. 4


a
Position
rs9600103
rs2981863
rs17969505
rs17969399
rs11841589
Ishikawa DHS
Tamoxifen
Estrogen
Vertebrates Conservation
PhyloP
PhastCons
Ishikawa
ARK-2
AN3CA (no KLF5 expression)
FAIRE
H3K4Me2
H4Ac
b
KLF5
NcoI
H3K4Me3
Interaction Frequency
Ishikawa
73595000
73615000
73635000
73655000
c
rs11841589-T
rs11841589-G
rs9600103-T
rs9600103-A
pGL3-Promoter
*
Relative luciferase activity