nature genetics

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

We conducted a meta-analysis of three endometrial cancer genome-wide association studies (GWAS) and two follow-up phases totaling 7,737 endometrial cancer cases and 37,144 controls of European ancestry. Genome-wide imputation and meta-analysis identified five new 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*). We also found a second independent 8q24.21 signal (rs17232730). 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] allele that is protective in endometrial cancer suppressed gene expression *in vitro*, suggesting that regulation of the expression of *KLF5*, 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 (refs. 3,4), but highly penetrant germline mutations in mismatch-repair genes[5] and DNA polymerase genes[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 risk loci for endometrial cancer, tagged by rs11263763 in *HNF1B* (intron 1)[8] and rs727479 in *CYP19A1* (intron 4)[9].

To identify additional endometrial cancer risk loci, we reanalyzed data from our previous GWAS (the Australian National Endometrial Cancer Study (ANECS) and Studies of Epidemiology and Risk Factors in Cancer Heredity (SEARCH) data sets[10]) and conducted a meta-analysis with two further studies (**Supplementary Fig. 1**). The first study was an independent GWAS, the National Study of Endometrial Cancer (NSECG), including 925 endometrial cancer cases genotyped using the Illumina 660W Quad array, 1,286 cancer-free controls from the CORGI/SP1 GWAS[11,12] and 2,674 controls from the 1958 British 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 Illumina Infinium iSelect array designed by the Collaborative Oncological Gene-environment Study (COGS) initiative[14–17] (**Supplementary Table 1** and **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 analyses 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 Fig. 2**). A fixed-effects meta-analysis was conducted for all 2.3 million typed and well-imputed (info score >0.90) SNPs in a total of 6,542 endometrial cancer cases and 36,393 controls. The strongest associations were with SNPs in linkage disequilibrium (LD) with previously identified risk SNPs for endometrial cancer in *HNF1B*[8,10,18] and *CYP19A1* (refs. 9,19) (**Fig. 1** and **Table 1**). For fourteen 1.5-Mb regions containing at least one new SNP with $P_{meta} < 1 \times 10^{-5}$, we performed regional imputation using an additional reference panel that comprised high-coverage whole-genome sequencing data for 196 UK individuals (**Supplementary Table 2**).

We identified five new regions containing at least one endometrial cancer risk SNP with $P_{meta} < 1 \times 10^{-7}$ and genotyped the most strongly associated SNP in each region 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 of >98.5% between different genotyping platforms (**Supplementary Table 4**). All five SNPs were associated with endometrial cancer at genome-wide significance ($P < 5 \times 10^{-8}$; **Figs. 2** and **3**, and **Table 1**), and these associations remained highly significant when analysis was restricted to cases with the endometrioid subtype only. Endometrioid-only analysis did not identify any additional risk loci. Expression quantitative trait locus (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 500 kb of the lead SNP in likely enhancers and multiple potential regulatory targets (**Supplementary Fig. 3** and **Supplementary Table 6**). The most compelling candidates for future functional analysis are described below.

rs13328298 (odds ratio (OR) = 1.13, 95% confidence interval (CI) = 1.09–1.18; $P = 3.73 \times 10^{-10}$) at 6q22.31 lies in the long noncoding RNA *LOC643623*, 54 kb upstream of *HEY2* and 86 kb upstream of *NCOA7*.

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

Received 26 August 2015; accepted 8 April 2016; published online 2 May 2016; doi:10.1038/ng.3562

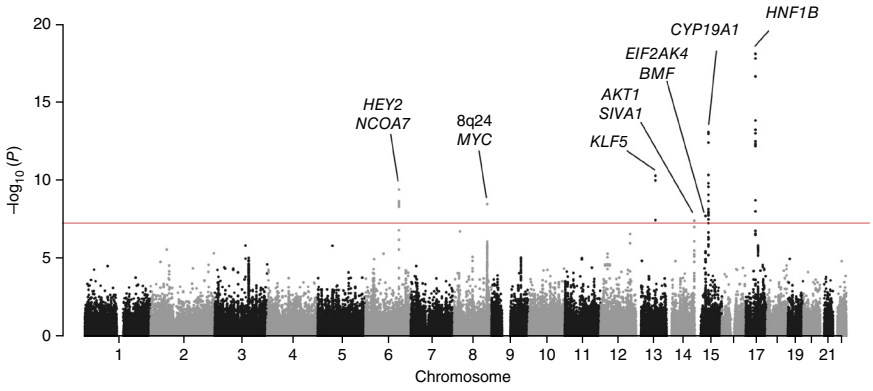

Figure 1 Manhattan plot from endometrial cancer meta-analysis. The Manhattan plot shows negative $\log_{10}$-transformed *P* values from meta-analysis over 22 autosomes. There are seven loci surpassing genome-wide significance (red horizontal line), including two known loci, 15q21 (*CYP19A1*) and 17q12 (*HNF1B*), and five new loci, 6q22 (*NCOA7* and *HEY2*), 8q24 (*MYC*), 13q22 (*KLF5*), 14q32 (*AKT1* and *SIVA1*) and 15q15 (*EIF2AK4* and *BMF*).

HEY2 encodes a helix-loop-helix transcriptional repressor in the Notch pathway, which maintains stem cells, and dysregulation has been associated with different cancers[22]. The NCOA7 protein modulates the activity of the estrogen receptor via direct binding[23].

The second locus (rs4733613: OR = 0.84, 95% CI = 0.80–0.89; $P = 3.09 \times 10^{-9}$) is at 8q24.21. Stepwise conditional logistic regression identified another independent signal in this region, rs17232730 (pairwise $r^2 = 0.02$, $P_{cond} = 1.29 \times 10^{-5}$; **Table 2**). Both SNPs associated with endometrial cancer lie further from *MYC* (784–846 kb telomeric) than most of the other cancer-associated 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 rs10088218 SNP for ovarian cancer ($r^2 = 0.43$), with both cancers having the same risk allele, but rs4733613 is not in LD ($r^2 \leq 0.02$) with any other cancer-associated SNP in the region (**Supplementary Fig. 3**). A role in tumorigenesis is implicated for several microRNAs (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; $P = 1.77 \times 10^{-8}$) 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; $P = 3.55 \times 10^{-8}$) lies in intron 3 of the *AKT1* oncogene, 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 Fig. 3** and **Supplementary Table 6**). *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 in aggressive disease[37–39]. *SIVA1* encodes an apoptosis regulatory protein that inhibits p53 activity[40,41] and enhances epithelial–mesenchymal transition to promote the motility and invasiveness of epithelial cells[42]. *INF2* expression is reported to act as a pro-migratory signal in gastric cancer cells treated with mycophenolic acid[43].

The final newly identified risk SNP for endometrial cancer was rs11841589 (OR = 1.15, 95% CI = 1.11–1.21; $P = 4.83 \times 10^{-11}$) at 13q22.1, 163 kb and 445 kb downstream of the Krüppel-like factors *KLF5* and *KLF12*, respectively. KLF5 is a transcription factor associated with cell cycle regulation, and it has 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 3-kb LD block bounded by rs9600103 ($P = 8.70 \times 10^{-11}$) and rs11841589 (**Fig. 4a**). There was no residual association signal at this locus ($P_{cond} > 0.05$) after conditioning on rs11841589. Bioinformatic analysis suggested that the causal variant in the intergenic 13q22.1 locus may affect a regulatory element that modifies *KLF5* expression (**Supplementary Fig. 3**): rs9600103 overlaps a vertebrate conservation peak and a DNase I hypersensitivity site (DHS) in estrogen- and tamoxifen-treated Ishikawa cells in the Encyclopedia of DNA Elements (ENCODE)[50] (**Fig. 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 Fig. 4**).

We further investigated the epigenetic landscape of a 16-kb 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; providing a

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

							All histologies			Endometrioid histology		
Locus	SNP	Position (bp)[a]	Nearby gene(s)	EA[b]	OA[c]	EAF[d]	Allelic OR (95% CI)	*P*	I^2[e]	Allelic OR (95% CI)	*P*	I^2[e]
Newly identified 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 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

For all newly identified loci, the lead SNP was either directly genotyped or imputed with an information score of greater than 0.9. *HNF1B* and *CYP19A1* have been previously reported by Painter *et al.*[8] and Thompson *et al.*[9].
[a]Position is with reference to Build 37 of the reference genome. [b]Effect allele. [c]Other allele. [d]Effect allele frequency. [e]Heterogeneity I^2 statistic[55].

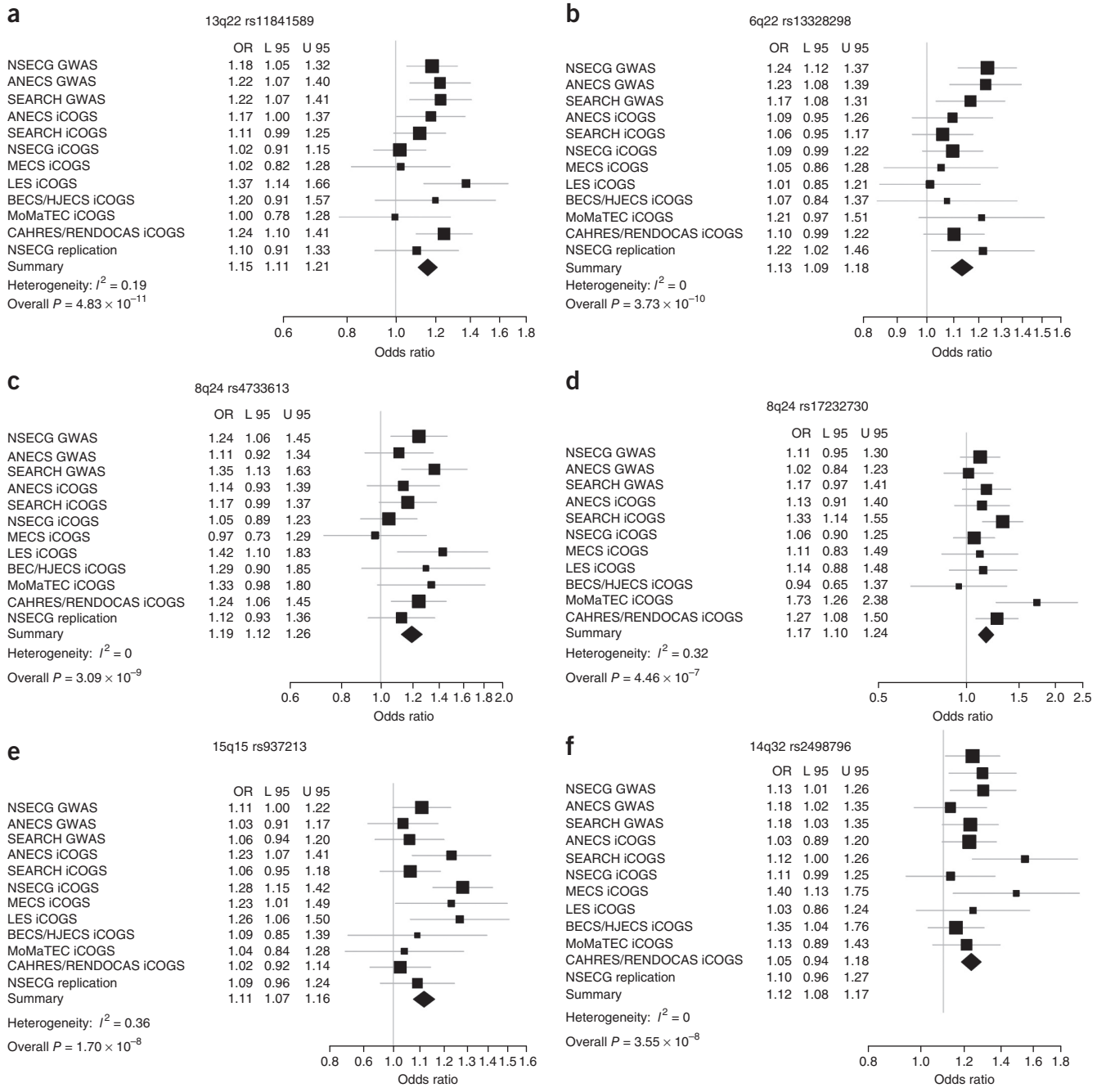

Figure 2 Forest plots for new endometrial cancer risk loci. The odds ratio and 95% confidence interval (L, lower; U, upper) for the loci in each study of the meta-analysis are listed and shown in the adjacent plot. The I^2 heterogeneity scores (all <0.4) suggest that there is no marked difference in effects between studies. (**a**) rs11841589 (13q22). (**b**) rs13328298 (6q22). (**c**) rs4733613 (8q24). (**d**) rs17232730 (8q24; pairwise $r^2 = 0.02$ with rs4733613). (**e**) rs937213 (15q15). (**f**) rs2498796 (14q32).

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 Fig. 5**). We conducted formaldehyde-assisted identification of regulatory elements (FAIRE; to identify regions of open chromatin) and chromatin immunoprecipitation (ChIP) using antibodies against dimethylation of histone H3 a lysine 4 (H3K4me2; a marker of transcription factor binding[52]) and pan acetylation of histone H4 (pan H4ac; a marker of active chromatin). Although ChIP analysis with antibody against H4ac did not yield a consistent signal in the region, peaks in signals from FAIRE and ChIP with antibody against H3K4me2 were specifically present in the *KLF5*-expressing lines and were colocalized 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 (**Fig. 4a**). We then conducted chromatin conformation capture (3C) experiments for the *KLF5*-expressing Ishikawa endometrial cancer cells (**Supplementary Fig. 5**) and found an interaction between the NcoI restriction fragment containing the rs11841589–rs9600103 risk locus SNPs and the promoter region of *KLF5* (**Fig. 4b**).

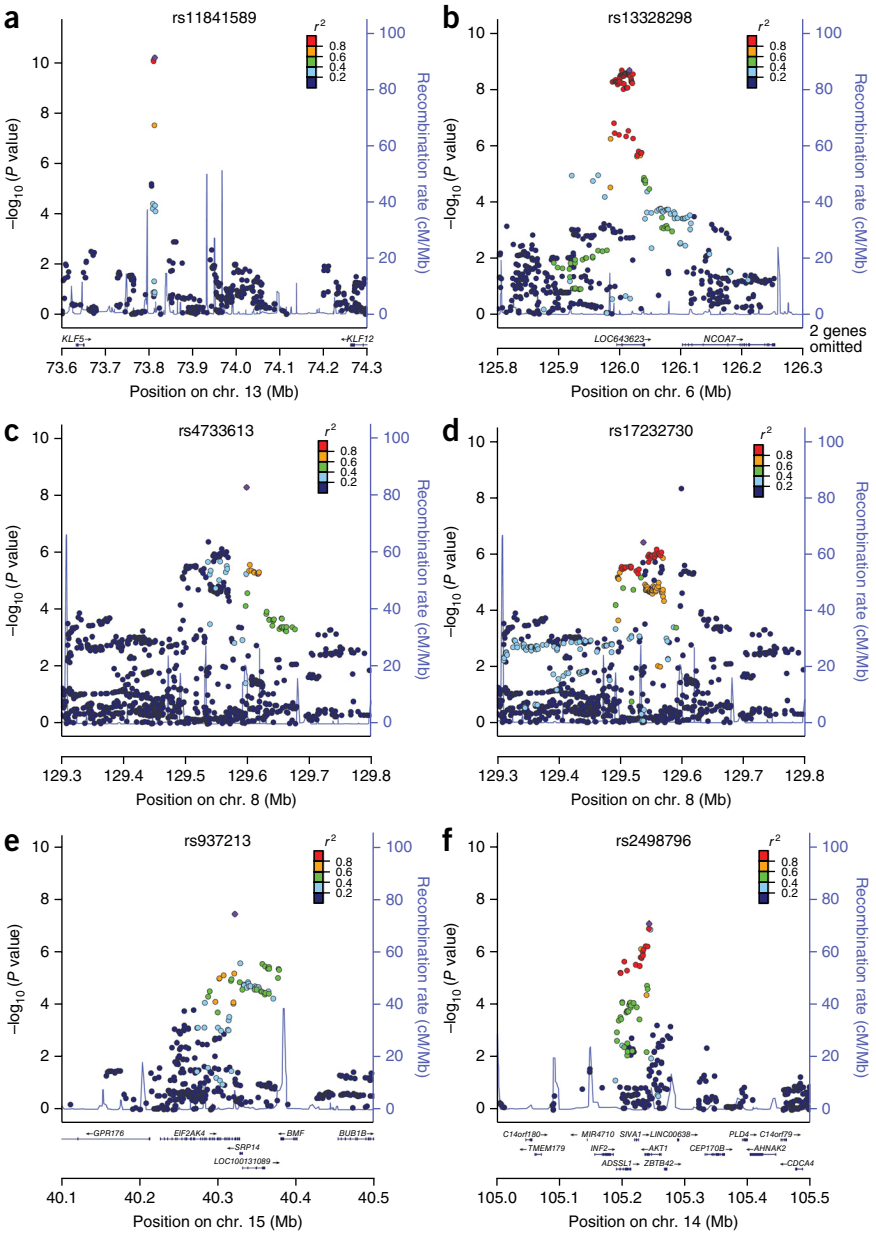

Figure 3 Regional association plots for the five new loci associated with endometrial cancer. (**a**–**f**) Negative $\log_{10}$-transformed *P* values from the meta-analysis and regional imputation for three GWAS and eight iCOGS groups are shown for SNPs at 13q22.1 (**a**), 6q22 (**b**), 8q24 (**c**,**d**), 15q15 (**e**) and 14q32.33 (**f**). The SNP with the lowest *P* value at each locus is labeled and represented by a purple diamond, and dot color indicates LD with the top SNP. The blue line shows the recombination rate. All plotted SNPs were either genotyped or had an IMPUTE info score of greater than 0.9 in all data sets. 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. Similar regional association plots with a larger number of SNPs using a less stringent info score cutoff are displayed in **Supplementary Figure 6**.

We investigated the regulatory nature of the region around rs11841589–rs9600103 using allele-specific luciferase reporter assays in Ishikawa cells (**Fig. 4c**). We used paired *t* tests to compare the activities of fragments containing the rs11841589 and rs9600103 alleles with that of 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 activity similar to that of the pGL3-Promoter control ($P = 0.23$) and significantly higher than that of the construct with 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 that interacts with the *KLF5* promoter region, with the rs9600103[A] risk allele likely associated with increased gene expression.

In summary, this meta-analysis identified five new endometrial cancer risk loci at genome-wide significance, bringing the total number of common risk loci for endometrial cancer identified by GWAS to seven (**Fig. 1**). Together with other risk-associated SNPs reaching study-wide significance[32,53,54], these loci explain ~5.1% of the familial relative risk for endometrial cancer. These new endometrial cancer risk SNPs lie in likely enhancers predicted to regulate gene expression for proteins or miRNAs with known or suspected roles in tumorigenesis, and we specifically showed that a functional SNP at 13q22.1 may lie within a transcriptional repressor of *KLF5*. Our findings further clarify the genetic etiology of endometrial cancer, provide

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

					Pairwise r^2 with		All-histology meta-analysis		Conditioning on rs4733613		Conditioning on rs17232730	
SNP	Position (bp)[a]	EA[b]	OA[c]	EAF[d]	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[e]	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

[a]Position with respect to Build 37 of the reference genome. [b]Effect allele. [c]Other allele. [d]Effect allele frequency. [e]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.

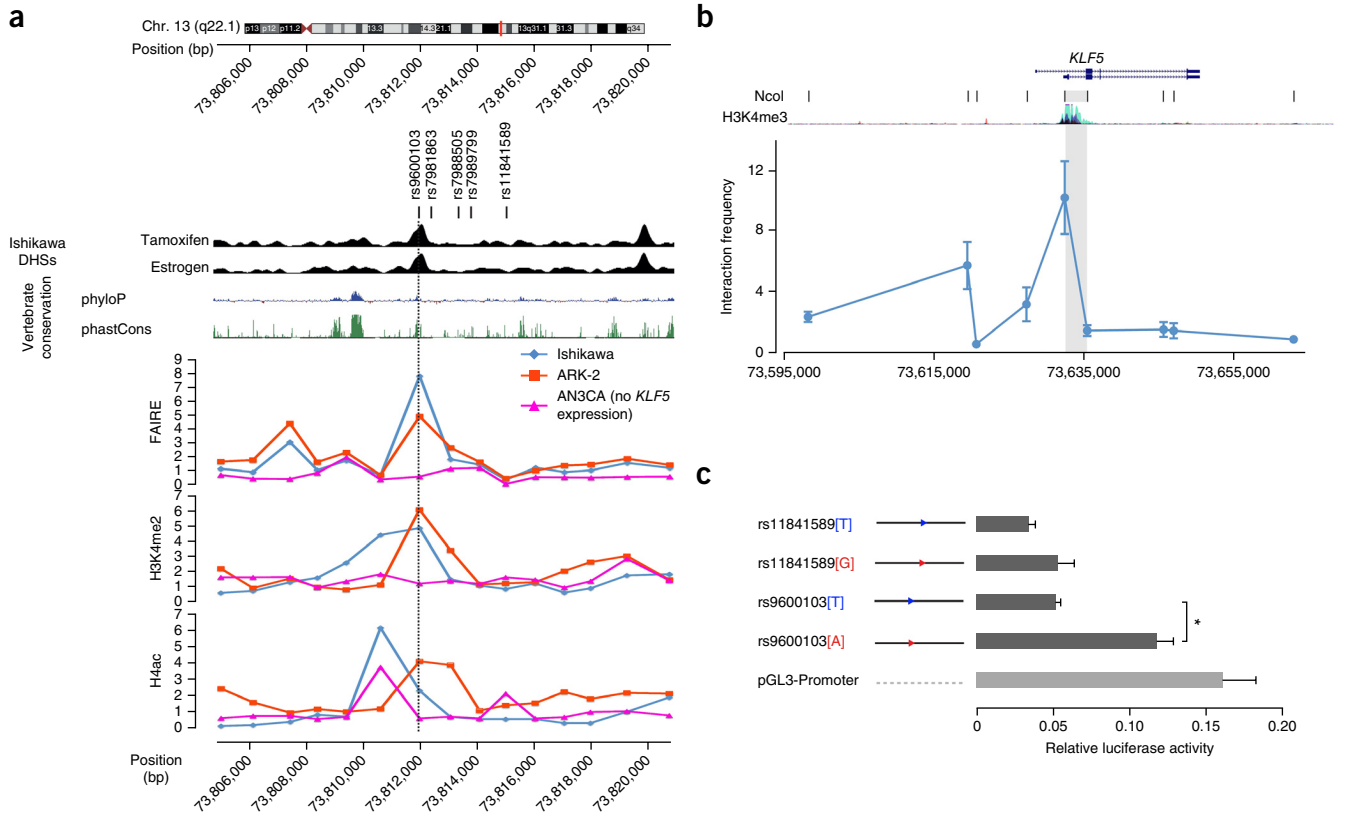

Figure 4 The 13q22.1 endometrial cancer susceptibility locus. (**a**) Diagram showing the 16-kb region around rs11841589, rs9600103 and correlated SNPs rs7981863, rs7988505 and rs7989799 (vertical black bars), DHS density signal in estrogen- and tamoxifen-treated Ishikawa cells in ENCODE (**Supplementary Note**) and conservation in 100 vertebrates. FAIRE and ChIP assays for H3K4me2 and H4ac in endometrial cancer cell lines ARK-2 (rs9600103 genotype TT), Ishikawa (rs9600103 genotype AA) and AN3CA (rs9600103 genotype AA) show evidence for enrichment of histone modifications. The vertical dotted line represents the position of rs9600103. (**b**) 3C experiment for *KLF5*-expressing Ishikawa cells. Relative interaction frequencies are shown for an NcoI restriction fragment containing risk SNPs rs9600103 and rs11841589 (bait fragment) and NcoI fragments across the *KLF5* promoter region, plotted against fragment position on chromosome 13. NcoI restriction sites (black vertical bars) are displayed below the schematic of the *KLF5* transcripts. The profile for trimethylation of histone H3 at lysine 4 (H3K4me3), indicative of promoters, from multiple ENCODE cell lines is also shown. The graph represents three biological replicates. Error bars, s.d. An interaction was seen with the fragment containing a *KLF5* transcriptional start site (fragment shaded in gray). (**c**) Luciferase reporter assays to analyze the activity of 3-kb fragments containing either rs9600103 or rs11841589 using the pGL3-Promoter vector in Ishikawa cells. Blue arrowheads represent the low-risk alleles, and red arrowheads represent the high-risk alleles. Error bars, s.e.m. (three biological replicates each with four technical replicates). 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**).

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/.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Data access was granted by the respective management groups of the participating studies, that is, the Australian National Endometrial Cancer Study (ANECS), the Queensland Institute of Medical Research Controls, the Hunter Community Study (HCS), the Studies of Epidemiology and Risk Factors in Cancer Heredity (SEARCH), the Wellcome Trust Case Control Consortium (WTCCC), the National Study of Endometrial Cancer Genetics (NSECG), the Endometrial Cancer Association Consortium (ECAC), the Breast Cancer Association Consortium (BCAC) and the 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.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank the many individuals who participated in this study and the numerous institutions and their staff that supported recruitment, detailed in full in the **Supplementary Note**.

The iCOGS endometrial cancer analysis was supported by an NHMRC project grant (1031333) to A.B.S., D.F.E. and A.M.D. A.B.S., P.M.W., G.W.M. and D.R.N. are supported by the NHMRC Fellowship scheme. A.M.D. was supported by the Joseph Mitchell Trust. I.T. is supported by Cancer Research UK and the Oxford Comprehensive Biomedical Research Centre. T.H.T.C. is supported by the Rhodes Trust and the Nuffield Department of Medicine. Funding for iCOGS infrastructure came from the European Community's Seventh Framework Programme under grant agreement 223175 (HEALTH-F2-2009-223175) (COGS), Cancer Research UK (C1287/A10118, C1287/A10710, C12292/A11174, C1281/A12014, C5047/A8384, C5047/A15007, C5047/A10692 and C8197/A16565), the US

National Institutes of Health (R01 CA128978, U19 CA148537, U19 CA148065 and U19 CA148112), the US Department of Defense (W81XWH-10-1-0341), the Canadian Institutes of Health Research (CIHR) for the CIHR Team in Familial Risks of Breast Cancer, the Susan G. 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 (339435), Cancer Council Queensland (4196615) and Cancer Council Tasmania (403031 and 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 (552402 and 1031333). Control data were generated by the 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 UK Medical Research Council grant G0000934 and Wellcome Trust grant 068545/Z/02; funding for this project was provided by the Wellcome Trust under award 085475. NSECG was supported by the European Union's Framework Programme 7 CHIBCHA grant and 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 toward the costs of establishing HCS.

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 the US Public Health Service (R01 CA122443, P30 CA15083 and P50 CA136393), 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, the 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, Jennie Thomas and the Hunter Medical Research Institute. RENDOCAS was supported through the regional agreement on medical training and clinical research (ALF) between the Stockholm County Council and Karolinska Institutet (20110222, 20110483, 20110141 and DF 07015), Swedish Labor Market Insurance (100069) and the Swedish Cancer Society (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.

BCAC is funded by Cancer Research UK (C1287/A10118 and C1287/A12014). 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 Note**.

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.H.Y. and J.W. conducted bioinformatic analyses. T.A.O'M. conducted eQTL analyses. S.F., A. Lewis, J.D.F., L.F.-M., D.C. and 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. coordinated the overall stage 2 genotyping and associated data management. J. Dennis, J.P.T. and K.M. coordinated quality control and data cleaning for the iCOGS control data sets. A.B.S. and T.A.O'M. coordinated the ANECS stage 1 genotyping. A.M.D., S.A. and C.S.H. coordinated the SEARCH stage 1 genotyping. I.T. and CHIBCHA funded and implemented the NSECG GWAS. I.T., L.M., M.G. and S.H. coordinated NSECG and collation of CORGI control GWAS data. A.B.S. and P.M.W. coordinated ANECS. R.J.S., M. McEvoy, J.A. and E.G.H. coordinated collation of GWAS data for HCS. N.G.M., G.W.M., D.R.N. and A.K.H. coordinated collation of GWAS data for the QIMR controls. P.D.P.P., D.F.E. and M.S. coordinated SEARCH. M.K.B. and Q.W. provided data management support for BCAC. The following authors designed and coordinated 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. Hillemanns, M. Dürst, I.R., D.L., S.S., H.Z., 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., G.O., T.L., M. Mints, 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. Brenner, A.M., H. Brauch, A. Lindblom, J.C.-C., F.J.C., G.G.G., V.N.K., A.C. and J.M.C. All authors provided critical review of the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

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. & Cancer and Steroid Hormone Study Group. A population-based study of endometrial cancer and familial risk in younger women. *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 ε and δ 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 Consortium. 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, e1–e2 (2013).
15. Michailidou, K. *et al.* Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat. Genet.* **45**, 353–361, e1–e2 (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, e1–e2 (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. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).
21. Cancer Genome Atlas Research Network. 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 2α kinase. *Eur. J. Biochem.* **265**, 754–762 (1999).
34. Uhlén, M. *et al.* 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. USA* **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. USA* **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 Krüppel-like factors in endocrine-responsive cancers of female reproductive tissues. *J. Endocrinol.* **204**, 223–231 (2010).
48. Nandan, M.O. *et al.* Krüppel-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. ENCODE Project Consortium. 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).

Timothy H T 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)[4], Shirley Hodgson[5], Penelope M Webb[3], The Australian National Endometrial Cancer Study Group (ANECS)[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 M J 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], CHIBCHA Consortium[4], Mulin Jun Li[32], Shun H Yip[32], Junwen Wang[24,33], Manjeet K Bolla[2], Kyriaki Michailidou[2], Qin Wang[2], Jonathan P Tyrer[12], Malcolm Dunlop[34,35], Richard Houlston[36], Claire Palles[1], John L Hopper[37], AOCS Group[4], Julian Peto[38], Anthony J Swerdlow[36,39], 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[37,51,52], 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]Oxford Centre for Cancer Gene Research, 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, Queensland, Australia. [4]A list of members 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, New South Wales, Australia. [7]Centre for Clinical Epidemiology and Biostatistics, School of Medicine and Public Health, University of Newcastle, Newcastle, New South Wales, Australia. [8]Hunter Area Pathology Service, John Hunter Hospital, Newcastle, New South Wales, Australia. [9]Centre for Information-Based Medicine, University of Newcastle, Newcastle, New South Wales, Australia. [10]School of Biomedical Sciences and Pharmacy, University of Newcastle, Newcastle, New South Wales, Australia. [11]Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, Queensland, Australia. [12]Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK. [13]Division of Hematology/Oncology, Department of Medicine, University of California, Los Angeles, David Geffen School of Medicine, Los Angeles, California, 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]Gynaecology Research Unit, Hannover Medical School, Hannover, Germany. [18]Department of Gynaecology, Jena University Hospital, Friedrich Schiller University, Jena, Germany.

[19]Clinics of Gynaecology and Obstetrics, Hannover Medical School, Hannover, Germany. [20]Division of Gynecologic Oncology, Department of Obstetrics and Gynecology, University Hospitals, KU Leuven, University of Leuven, Leuven, Belgium. [21]Vesalius Research Center, VIB, Leuven, Belgium. [22]Laboratory for Translational Genetics, Department of Oncology, University Hospitals Leuven, Leuven, Belgium. [23]Division of Gynecologic Oncology, Department of Obstetrics and Gynecology, Mayo Clinic, Rochester, Minnesota, USA. [24]Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota, USA. [25]Department of Biostatistics, University of Kansas Medical Center, Kansas City, Kansas, USA. [26]Centre for Cancerbiomarkers, Department of Clinical Science, University of Bergen, Bergen, Norway. [27]Department of Obstetrics and Gynecology, Haukeland University Hospital, Bergen, Norway. [28]School of Medicine and Public Health, University of Newcastle, Newcastle, New South Wales, 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]Department of Clinical Genetics, Karolinska University Hospital Solna, Stockholm, Sweden. [32]Centre for Genomic Sciences, Li Ka Shing Faculty of Medicine, University of Hong Kong, Hong Kong, China. [33]Department of Biomedical Informatics, Arizona State University, Scottsdale, Arizona, USA. [34]Colon Cancer Genetics Group, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK. [35]MRC Human Genetics Unit, Western General Hospital Edinburgh, Edinburgh, UK. [36]Division of Genetics and Epidemiology, Institute of Cancer Research, London, UK. [37]Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, University of Melbourne, Melbourne, Victoria, 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]Division of Tumor Genetics, Department of Obstetrics and Gynecology, 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, Minnesota, USA. [51]Cancer Epidemiology Centre, Cancer Council Victoria, Melbourne, Victoria, Australia. [52]Department of Epidemiology and Preventive Medicine, Monash University, Melbourne, Victoria, Australia. [53]Department of Genetics, Institute for Cancer Research, Norwegian Radium Hospital, Oslo, Norway. [54]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).

ONLINE METHODS

Study overview. Cases and controls were matched as summarized in **Supplementary Table 1**. Each sample set is described in the **Supplementary Note**. The overall study design is illustrated in **Supplementary Figure 1**.

Additional endometrial cancer 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 performed using Illumina 660W Quad arrays.

These cases were matched with 1,286 cancer-free controls from the UK1/CORGI[12] and SP1 (ref. 11) colorectal studies genotyped using Illumina Hap550, Hap300 and Hap240S arrays and 1958 Birth Cohort[55] controls from the Wellcome Trust Case Control Consortium 2 (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 were generated using Illumina Infinium 610K arrays[10]. In comparison with our previous study[10], this meta-analysis analyzed 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 on the basis of geographical location, and principal-components analysis (PCA) was conducted; individuals who clustered outside the main centroid in pairwise plots of the first four principal components were excluded (**Supplementary Fig. 7**).

Cases and controls were genotyped on a custom Illumina Infinium iSelect array with 211,155 SNPs, designed by the Collaborative Oncological Gene-environment Study (iCOGS), a collaborative project involving four consortia. SNPs were included on this array on the basis of promising regions of interest in previous breast, ovarian and prostate[14] studies and also the 1,483 top SNPs from our previous endometrial cancer GWAS[10] analysis. Cases and MoMaTEC controls were genotyped by the Génome Québec Innovation Center. BCAC and OCAC control samples were genotyped at four centers. Raw intensity data files for all consortia were sent to the COGS data coordination center at the University of Cambridge for centralized genotype calling and quality control, so that all case and control genotypes were called using the same procedure.

SNP genotyping array quality control. Genotype calling was performed using Illumina's proprietary Gencall algorithm and Illumnus[58]. Duplicate samples displayed >99% concordance. Standard quality control measures applied to genotyping arrays are described in our original GWAS[10] and included genotypic call rate <0.95; deviation from Hardy–Weinberg equilibrium at $P < 1 \times 10^{-6}$; and visual inspection of cluster plots for most significant SNPs. For iCOGS, all endometrial cancer cases and MoMaTEC controls were genotyped by the Génome Québec Innovation Center. BCAC and OCAC control samples were genotyped at four centers. Raw intensity data files for all consortia were sent to the COGS data coordination center at the University of Cambridge for centralized genotype calling and quality control, so that all case and control genotypes were called using the same procedure. Duplicate samples for quality control showed a concordance of >99%. Samples were excluded on the basis of 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 (cutoff for second-degree relatives). PCA was conducted using EIGENSTRAT[59] software. Analysis was conducted using PLINK[60] and the R packages GenABEL and SNPMatrix[61,62].

Phase 2 NSECG genotyping. A second genotyping phase consisted of assaying five SNPs with $P < 1 \times 10^{-7}$ and IMPUTE info scores >0.94 from the NSECG–ANECS–SEARCH–iCOGS meta-analysis; samples were NSECG cases and controls not previously used in the NSECG GWAS or NSECG iCOGS. Genotyping was conducted using competitive allele-specific PCR (KASPar, KBiosciences) and the Fluidigm BioMark 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 Hardy–Weinberg equilibrium ($P > 0.05$). Sequences for 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 the 1000 Genomes Project (2012 release) as reference panel. For the first-pass genome-wide analysis, we prephased chromosomes using SHAPEIT[64] to improve the computational speed. Imputation was carried out separately for each of the three GWAS (for each GWAS, the cases and controls were imputed together as a single data set, using only SNPs that passed quality control 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 before imputation. Genome-wide imputation produced 9,594,066 SNPs with MAF ≥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 prioritized cancer-associated SNPs and not genome-wide coverage, but nonetheless 8,631,871 SNPs met the criteria of MAF ≥1% and info ≥0.4, of which 5,437,135 had info ≥0.7 and 2,333,040 had info ≥0.9.

Regional imputation of regions of interest (1.5-Mb region around SNPs with meta-analysis $P < 1 \times 10^{-5}$) used both 1000 Genomes Project 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 data sets.

Association testing. Association testing was carried out 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 overdispersion of the test statistic from the quantile–quantile plots (**Supplementary Fig. 2**) and the genomic inflation λ_{GC}, calculated using all genotyped SNPs passing quality control for the three GWAS. For iCOGS, 105,000 SNPs after LD pruning (r^2 <0.2) and >500 kb from the 1,483 endometrial cancer 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 carried out in a stepwise manner, first conditioning for the most significant SNP and subsequently for any SNPs that remained significant at $P_{cond} < 1 \times 10^{-4}$. Regional association plots (**Fig. 1** and **Supplementary Fig. 6**) were created using LocusZoom[67].

Meta-analysis. Inverse variance, fixed-effects meta-analysis of the 11 groups (3 GWAS and 8 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-transformed odds ratio) and its standard error. Between-study differences are represented by the I^2 heterogeneity score[69,70]. Forest plots of the genome-wide significant loci (**Fig. 2**) representing risk effects across different studies were generated 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 subjected to meta-analysis in a 12-way meta-analysis for the top five SNPs yielding a total of 7,737 endometrial cancer cases and 37,144 controls. 6,635 (86%) of the endometrial cancer 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 new genome-wide significant loci and SNPs in LD (r^2 >0.7 in the European 1000 Genomes Project population) were annotated using HaploRegV2 (ref. 71), RegulomeDB[72] and data from ENCODE[50] in **Supplementary Table 6**. These include information such as promoter and enhancer histone marks, DHSs, bound proteins, altered motifs, GENCODE and dbSNP annotations, RegulomeDB score and phastCons conservation scores.

The bioinformatic analysis in **Supplementary Figure 3** used data sets described by Hnisz *et al.*[73] and Corradin *et al.*[74] to identify likely enhancers in a cell-type-specific context for the risk loci. Enhancer–gene interactions are predicted by identifying 'super-enhancers' (regions containing neighboring H3K27ac modifications) from 86 cell and tissue types, and the expressed transcript with the transcription start site closest to the center of the super-enhancer was then 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 analysis for associated SNPs using GTEx and TCGA data. Publicly available data generated by the Genotype-Tissue Expression (GTEx) Project[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 RNA-seq arrays (Illumina HiSeq and Illumina Genome Analyzer) for 458 endometrial cancer tissues and 30 adjacent normal endometrial tissues. Association analyses for TCGA data sets were performed as follows. Genes within 500 kb flanking our SNPs of interest were selected for analysis. Because there may be significant variation in tumor 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 from the laboratory of D. Church, acquired as gifts from B. Weigelt (currently at the Memorial Sloan Kettering Cancer Center) and K. Dedes (University of Zurich), were routinely tested for mycoplasma contamination. Somatic mutation data generated previously match those reported in publicly available resources and the literature, where available. Cells were snap frozen with dry ice after centrifugation, and DNA and RNA were extracted using DNeasy and RNeasy mini kits (Qiagen). Nucleic acids were quantified using Nanodrop 2000 (Thermo Scientific) spectrophotometry.

Quantification of *KLF5* expression in endometrial cancer cell lines. Extracted RNA was treated with DNase I, and 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 upon request). The absolute expression of *KLF5* was quantified using qRT–PCR on the ABI 7900HT cycler (Applied Biosystems), and the critical threshold was manually set at 0.2. Relative expression was calculated using the $\Delta\Delta C_t$ method described by Livak and Schmittgen[75], with *GAPDH* as an endogenous control.

Formaldehyde-assisted identification of regulatory elements. FAIRE was conducted using a method adapted from Giresi *et al.*[76]. Briefly, cross-linking was performed on a rocker at room temperature. 1% formaldehyde was added to ~1×10^8 cells for 5 min, and 115 mM glycine was added to inhibit cross-linking. For each cell line, a non-cross-linked control was prepared in parallel for all remaining steps. After two rinses with 4 °C PBS, cells were suspended in successive buffers: lysis buffer I (50 mM HEPES-KOH, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40 and 0.25% Triton X-100); lysis buffer II (10 mM Tris-HCl, 200 mM NaCl, 1 mM EDTA and 0.5 mM EGTA); lysis buffer III (10 mM Tris-HCl, 2.1 M NaCl, 1 mM EDTA, 0.1% sodium deoxycholate and 0.5% *N*-lauroylsarcosine). Cells were incubated on a rocker at 4 °C for 10 min in each lysis buffer, then spun down at 1,300*g* for 5 min, and the supernatant was removed. The cells were then sonicated using the Bioruptor in 7 to 15 30-s cycles to generate fragments 100–1,000 bp in size, and gel electrophoresis in 1% agarose was used to confirm DNA fragment sizes. The DNA was extracted with a standard phenol–chloroform method and ethanol precipitated. 50 ng of DNA from paired cross-linked and non-cross-linked cells was analyzed in duplicate by SYBR Green qPCR using primers at ~1-kb intervals in the 13q22.1 region downstream of *KLF5* (**Supplementary Table 8**). The $\Delta\Delta Ct$ method[31] was used to normalize results to the input DNA from non-cross-linked cells and then expressed relative to the rhodopsin promoter as negative control. For each experiment, there were two replicates for the cross-linked cells and non-cross-linked controls, each performed on two occasions.

Cross-linked chromatin immunoprecipitation. About 1×10^8 cells were cross-linked using 1% formaldehyde for 10 min. Glycine was used to stop the cross-linking, cells were then rinsed twice in PBS and cell scrapers were used to detach cells adhered to the Petri dish surface. Cells were then resuspended in lysis buffer (1% SDS, 10 mM EDTA (Ambion) and 50 mM Tris-HCl (Ambion)), incubated for 10 min and then sonicated using the Bioruptor (Diagenode) in 7 to 15 30-s cycles to generate fragments 1,000–1,500 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 immunoprecipitation dilution buffer (1% Triton X-100, 2 nM EDTA, 20 mM Tris-HCl and 150 mM NaCl), and each cell line was separated into four tubes: input chromatin, no-antibody control and one tube for each antibody. Five microliters of antibody to H3K4me2 (Millipore, 07-030) and acetylated histone H4 (Millipore, 06-866) was 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 on day 2. Phenylmethylsulfonyl fluoride and protease inhibitors were added to the lysis buffer and immunoprecipitation dilution buffer to deactivate proteases, and sodium butyrate was added to these solutions to inhibit histone deacetylases. Five microliters of protein A Dynabeads was added to each tube, and beads were incubated for 4 h. A series of washes were performed using Tris-sucrose-EDTA (TSE) I (1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl, 150 mM NaCl and 0.1% SDS), TSE II (1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl, 500 mM NaCl and 0.1% SDS), Buffer III (0.25 M lithium chloride, 1 mM EDTA, 10 mM Tris-HCl, 1% tergitol-type NP-40 and 1% sodium deoxycholate) and Tris-EDTA (1×). Three hundred microliters of extraction solution (1% SDS 0.1 M sodium bicarbonate) was added, and Dynabeads were removed after a 30-min 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). One microliter of DNA was analyzed in duplicate or triplicate by SYBR Green qPCR as above, and the $\Delta\Delta C_t$ 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. Sequences for the primers used are listed in **Supplementary Table 8**.

Chromatin conformation capture. Experiments were performed as described in Ghoussaini *et al.*[77], using the *KLF5*-expressing Ishikawa endometrial cancer cell line from the American Type Culture Collection. The cell line was authenticated using short-tandem-repeat (STR) profiling, and routinely tested for mycoplasma contamination (QIMR Berghofer in-house Support Services). Briefly, Ishikawa cells were cross-linked with 1% formaldehyde for 10 min, the reaction was quenched with 125 mM glycine, and cells were washed with PBS and collected by scraping. Cells were lysed for 30 min on ice in 10 mM Tris-HCl, pH 7.5, 10 mM NaCl and 0.2% Igepal with protease inhibitors and homogenized in a Dounce homogenizer. Nuclei were pelleted and resuspended in 1 ml of 1.2× restriction buffer (NEB 3.1) with 0.3% SDS for 1 h at 37 °C. 2% Triton X-100 was added, and 1,000 U NcoI was then added three times over 24 h at 37 °C with shaking. The enzyme was inactivated, and digested DNA was diluted 8× before ligation with 4,000 U of T4 DNA ligase overnight at 16 °C. Cross-links were reversed by proteinase K digestion

doi:10.1038/ng.3562

at 65 °C overnight, and DNA was purified by phenol–chloroform extraction and ethanol precipitation. The final DNA pellet was dissolved in 10 mM Tris (pH 7.5) and purified through Amicon Ultra 0.5-ml columns (Millipore). 3C interactions were quantified by SYTO9 qPCR (performed on a RotorGene 6000) using primers designed to amplify across ligated NcoI restriction fragments with one constant primer within the risk fragment (including rs11841589 and rs9600103) and a series of test primers within NcoI fragments spanning 76 kb of the *KLF5* promoter region. BAC clones (RP11-81D9 and RP11-179I20) covering the region were digested with NcoI, ligated with T4 ligase and used to determine PCR efficiency. 3C analyses were performed on three independent 3C libraries, with each data point in duplicate. Data were normalized to the signal from the BAC clone library and from a non-interacting chromosomal region using the $\Delta\Delta C_t$ 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 chr. 13: 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 cotransfected using Lipofectamine with the appropriate pGL3-Promoter constructs, and the *Renilla* luciferase pGL4.75 vector (Promega) as control for transfection efficiency. After 48 h, 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 vector 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 3 or 4 replicates, each performed on three occasions (total of 11 assays). The sequences for the 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).

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. Mägi, 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., Sánchez-Meca, J., Marín-Martínez, F. & Botella, J. Assessing heterogeneity in meta-analysis: *Q* statistic or I^2 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–D934 (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).