A Common Variant at the 14q32 Endometrial Cancer Risk Locus Activates *AKT1* through YY1 Binding

Jodie N. Painter,[1] Susanne Kaufmann,[1] Tracy A. O'Mara,[1] Kristine M. Hillman,[1] Haran Sivakumaran,[1] Hatef Darabi,[2] Timothy H.T. Cheng,[3] John Pearson,[1] Stephen Kazakoff,[1] Nicola Waddell,[1] Erling A. Hoivik,[4,5] Ellen L. Goode,[6] Rodney J. Scott,[7,8,9,10] Ian Tomlinson,[3] Alison M. Dunning,[11] Douglas F. Easton,[11,12] Juliet D. French,[1] Helga B. Salvesen,[4,5] Pamela M. Pollock,[13] Deborah J. Thompson,[12] Amanda B. Spurdle,[1] and Stacey L. Edwards[1,*]

A recent meta-analysis of multiple genome-wide association and follow-up endometrial cancer case-control datasets identified a novel genetic risk locus for this disease at chromosome 14q32.33. To prioritize the functional SNP(s) and target gene(s) at this locus, we employed an in silico fine-mapping approach using genotyped and imputed SNP data for 6,608 endometrial cancer cases and 37,925 controls of European ancestry. Association and functional analyses provide evidence that the best candidate causal SNP is rs2494737. Multiple experimental analyses show that SNP rs2494737 maps to a silencer element located within *AKT1*, a member of the PI3K/AKT/MTOR intracellular signaling pathway activated in endometrial tumors. The rs2494737 risk *A* allele creates a YY1 transcription factor-binding site and abrogates the silencer activity in luciferase assays, an effect mimicked by transfection of YY1 siRNA. Our findings suggest YY1 is a positive regulator of *AKT1*, mediating the stimulatory effects of rs2494737 increasing endometrial cancer risk. Identification of an endometrial cancer risk allele within a member of the PI3K/AKT signaling pathway, more commonly activated in tumors by somatic alterations, raises the possibility that well tolerated inhibitors targeting this pathway could be candidates for evaluation as chemopreventive agents in individuals at high risk of developing endometrial cancer.

Introduction

Endometrial cancer (MIM: 608089) (cancer of the lining of the uterine corpus) is the fourth most diagnosed cancer in women in Europe and North America.[1,2] To date, analyses of multiple genome-wide association study (GWAS) and follow-up datasets, comprising up to 7,737 endometrial cancer cases and 37,144 controls, have identified seven risk loci at genome-wide significance for this disease, including *HNF1B* (MIM: 189907),[3,4] *CYP19A1* (MIM: 107910),[5] and novel loci on chromosomes 13q22.1, 6q22.31, 8q24.21, 15q15.1, and 14q32.33.[6] The lead SNP at the 14q32.33 locus, rs2498796, represents a single association signal located in the region of the *AKT1* (MIM: 164730) oncogene.[6] *AKT1* is a member of the P13K/AKT/MTOR intracellular signaling pathway affecting cell survival and proliferation.[7] This gene is of particular interest for endometrial cancer as increased PI3K/AKT/MTOR signaling is a common occurrence in endometrial tumors, and in aggressive subtypes in particular.[8] Somatic alterations in one or more members of the PI3K/AKT/MTOR signaling pathway are common, with *PTEN* (MIM: 601728) as the most frequently altered gene.[9] Moreover, high *PIK3CA* (MIM: 171834) copy number and elevated levels of phosphorylated AKT have been associated with aggressive disease.[8,10,11] Our previous bioinformatic analysis indicated that rs2498796 and other SNPs in high linkage disequilibrium (LD) with this SNP might also regulate other nearby genes *SIVA1* (MIM: 605567), *ZBTB42* (MIM: 613915), *ADSSL1* (MIM: 612498), and *INF2* (MIM: 610982).[6] Here, we detail in silico fine-mapping and bioinformatic investigation of an expanded set of genotyped and imputed SNPs at 14q32.33, derived from the meta-analysis dataset described above, and multiple laboratory analyses to identify the functional SNP(s) and target gene(s) increasing endometrial cancer risk at this locus.

Material and Methods

Previously, meta-analysis of data for 7,737 endometrial cancer cases and 37,144 controls of European ancestry from three GWAS datasets (ANECS, SEARCH, and NSECG) and two follow-up datasets (iCOGs and NSECG Phase 2) identified rs2498796 (OR = 1.12 for the minor A allele, 95% CI:1.07–1.17, p value = 3.55×10^{-8}) as the top SNP representing a single association signal at the 14q32.33 endometrial cancer risk locus.[6] For the current

[1]QIMR Berghofer Medical Research Institute, Brisbane, QLD 4006, Australia; [2]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm 17177, Sweden; [3]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK; [4]Centre for Cancer Biomarkers, Department of Clinical Science, The University of Bergen, N5020 Bergen, Norway; [5]Department of Obstetrics and Gynecology, Haukeland University Hospital, N5021 Bergen, Norway; [6]Department of Health Sciences Research, Mayo Clinic, Rochester, MN 55905, USA; [7]Hunter Medical Research Institute, John Hunter Hospital, Newcastle, NSW 2305, Australia; [8]Pathology North (Newcastle) John Hunter Hospital, Newcastle, NSW 2305, Australia; [9]Centre for Information Based Medicine, University of Newcastle, NSW 2308, Australia; [10]School of Biomedical Sciences and Pharmacy, University of Newcastle, NSW 2308, Australia; [11]Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge CB1 8RN, UK; [12]Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge CB1 8RN, UK; [13]Institute of Health and Biomedical Innovation and School of Biomedical Science, Queensland University of Technology at the Translation Research Institute, Brisbane 4102, Australia
*Correspondence: stacey.edwards@qimrberghofer.edu.au
http://dx.doi.org/10.1016/j.ajhg.2016.04.012.

study we employed an in silico fine-mapping approach[12] previously used to fine-map other endometrial cancer risk loci,[4,5,13,14] focussing on the 1Mb region surrounding rs2498796 (bases 104,743,220-105,743,220; NCBI build 37/hg19 assembly). The current analysis utilized genotyped and imputed SNP data for the three GWAS (ANECS, SEARCH, and NSECG) and the iCOGs follow-up datasets and included a total of 6,608 endometrial cancer cases and 37,925 controls (details of these datasets can be found in[4,5]). The Cheng et al. analysis had included a total of 420 genotyped and imputed SNPs with minor allele frequencies (MAF) $\geq$1% and information scores $\geq$0.9 per dataset within the focal region.[6] To expand the search for potentially functional SNPs, we considered all genotyped and imputed SNPs (N = 2,922) with MAF $\geq$1% and information scores $\geq$0.4 per dataset. As described previously[4], regional imputation to the 1,000 Genomes v3 2012 release was conducted separately for each of the four datasets, based on inference panels of SNPs typed for each dataset, using IMPUTE v2.[15] Association testing was performed separately for each dataset using frequentist tests with a logistic regression model in SNPTEST v2.[16], and standard fixed effects meta-analysis using the beta estimates and standard errors per dataset conducted using METAL.[17] The regional association plot was created using LocusZoom.[18] Log-likelihood tests were used to determine the most likely causal SNPs by comparing the log-likelihoods obtained from the meta-analysis of our top SNPs ($p < 10^{-6}$) with that of the most significantly associated SNP. SNPs with odds of 100:1 or better of being the top SNP were prioritized as potential causal candidates for bioinformatic and functional analyses.[4,19,20] LD between SNPs was calculated from European Phase 3 1000 Genomes data and accessed from the National Cancer Institute LDlink tool.[21]

Bioinformatic Analysis

Bioinformatic analyses on SNPs prioritized by the log-likelihood tests were performed using publically available datasets from ENCODE[22], which includes information such as the location of promoter and enhancer histone marks, open chromatin, bound proteins and altered motifs for the Ishikawa endometrial cancer cell line. Data from Hnisz et al.[23] and PreSTIGE[24] was accessed to identify the location of likely enhancers and their gene targets in a cell-specific context.

Expression Analyses

Expression quantitative trait locus (eQTL) analyses were conducted using uterine tissue-specific data (N = 70) generated by the Genotype-Tissue Expression Project (GTEx)[25], and SNP (Affymetrix 6.0 arrays), RNA-seq and copy number (CNV) data for endometrial carcinoma samples (N = 526) and normal tissue samples adjacent to endometrial carcinoma (N = 29) obtained from restricted (SNP and RNA-Seq) and public (CNV) data portals of The Cancer Genome Atlas (TCGA).[26] For the TCGA data, to investigate the expression of all *AKT1* isoforms, including unannotated transcripts, unprocessed RNA-Seq FASTQ files were adapter trimmed using cutadapt (v1.8.1) and aligned to the Ensembl[27] GRCh37 reference (version 70) using STAR[28] (v2.4.2a). RNA-SeQC[29] (v1.1.8.1) was used to assess sequencing quality for all aligned data. Gene and transcript counts were estimated using RSEM[30] (v1.2.22). Genotypes for *AKT1* region SNPs present in the 1000 Genomes v3 2012 dataset which were not present on the Affymetrix 6.0 arrays were imputed using MaCH[31,32] and minimac[33,34] software. eQTL analyses were performed on transcripts expressed in >80% of samples using Kruskal-Wallis tests adjusting for copy number and sequencing method, with Bonferroni corrected p values < 0.006 (0.05/8 transcripts per SNP) considered statistically significant.

Cell Lines

Endometrial cancer cell lines Ishikawa and EN-1078D (both heterozygous for the SNPs under investigation) were grown in DMEM medium with 10% FCS and antibiotics. Cell lines were maintained under standard conditions, routinely tested for *Mycoplasma* and short tandem repeat (STR) profiled to confirm cell line identity.

Chromatin Conformation Capture

Chromatin conformation capture (3C) libraries were generated using *Nco*I as described previously.[35] 3C interactions were quantitated by real-time PCR (qPCR) using primers designed within restriction fragments (Table S1). All qPCR was performed on a RotorGene 6000 using MyTaq HS DNA polymerase (Bioline) with the addition of 5 mM of Syto9, annealing temperature of 66°C and extension of 30 s. 3C analyses were performed in three independent 3C libraries from each cell line with each experiment quantified in duplicate. BAC clones covering the 14q32 region were used to create artificial libraries of ligation products in order to normalize for PCR efficiency. Data were normalized to the signal from the BAC clone library and, between cell lines, by reference to a region within *GAPDH* (MIM: 138400).[35] All qPCR products were electrophoresed on 2% agarose gels, gel purified, and sequenced to verify the 3C product.

Electromobility Shift Assays

Gel shift assays were performed with Ishikawa and EN-1078D nuclear lysates and biotinylated oligonucleotide duplexes (Table S2). Nuclear lysates were prepared using the NE-PER nuclear and cytoplasmic extraction reagents (Thermo Fisher Scientific) as per the manufacturer's instructions. Total protein concentrations in nuclear lysates were determined by Bradford's method. Duplexes were prepared by combining sense and antisense oligonucleotides in NEBuffer2 (New England Biolabs) and heat annealing at 80°C for 10 min and slow cooling to 25°C for 1 hr. Binding reactions were performed in binding buffer (10% [vol/vol] glycerol, 20 mM HEPES [pH 7.4], 1 mM DTT, protease inhibitor cocktail [Roche], 0.75 μg poly[dI:dC] [Sigma-Aldrich]) with 7.5 μg of nuclear lysate. For competition assays, binding reactions were pre-incubated with 1 pmol of competitor duplex (Table S3) at 25°C for 10 min before the addition of 10 fmol of biotinylated oligo duplex and a further incubation at 25°C for 15 min. For gel-supershift assays, 5 μg of rabbit polyclonal YY1 antibody (Santa Cruz H-414) or C/EBP antibody (Santa Cruz sc-150) was added immediately before probe addition. The rabbit pre-immune IgG (Santa Cruz sc-2027) was used as a negative control. Reactions were separated on 10% (WT/vol) Tris-Borate-EDTA (TBE) polyacrylamide gels (Bio-Rad) in TBE buffer at 160 V for 40 min. Duplex-bound complexes were transferred onto Zeta-Probe positively-charged nylon membranes (Bio-Rad) by semi-dry transfer at 25 V for 20 min then cross-linked onto the membranes under 254 nm ultra-violet light for 10 min. Membranes were processed with the LightShift Chemiluminescent EMSA kit (Thermo Fisher Scientific) as per the manufacturer's instructions. Chemiluminescent signals were visualized with the C-DiGit blot scanner (LI-COR).

Table 1. Association of the Top Candidate Causal SNPs at Chromosome 14q32.33 with Endometrial Cancer Risk

SNP	Position (Build 19)	Minor allele	Common allele	MAF[a]	OR (95% CI)[b]	p value	r^2 to rs2498796	Likelihood ratio[c]
rs2498794	105245251	G	A	0.48	1.13 (1.09–1.17)	8.7×10^{-9}	0.43	1
rs2494737	105246325	A	T	0.30	1.13 (1.08–1.17)	2.5×10^{-7}	0.83	26
rs2498796	105243220	A	G	0.30	1.11 (1.07–1.16)	1.2×10^{-6}	–	120

[a]Minor (risk-increasing) allele frequency.
[b]OR for the effect allele.
[c]Ratio of the likelihood of rs2498794 to the likelihood of this SNP.

Plasmid Construction and Reporter Assays

Promoter-driven luciferase reporter constructs were generated by the insertion of PCR amplified fragments containing *AKT1* canonical (can), *AKT1* alternative (alt), or *ZBTB42* promoters into the *Mlu*I and *Bgl*II sites of pGL3-Basic. A 2537 bp fragment containing a putative regulatory element (PRE) identified by bioinformatic analysis was generated by PCR and cloned into *Bam*HI and *Sal*I sites of the modified pGL3-promoter constructs (Table S4). The minor (risk-increasing) alleles of individual SNPs were introduced into the PRE sequences by overlap extension PCR or gBlocks (Integrated DNA Technologies). Sequencing of all constructs confirmed variant incorporation (Australian Genome Research Facility). Ishikawa and EN-1078D cells were transfected with equimolar amounts of luciferase reporter plasmids and 50 ng of pRL-SV40 transfection control plasmid with Lipofectamine 2000. The total amount of transfected DNA was kept constant at 600 ng for each construct by the addition of pUC19 as a carrier plasmid. Luciferase activity was measured 24 hr post-transfection by the Dual-Glo Luciferase Assay System. To correct for any differences in transfection efficiency or cell lysate preparation, we normalized *Firefly* luciferase activity to *Renilla* luciferase and measured the activity of each construct relative to the promoter alone construct, which had a defined activity of 1. Statistical significance was tested by log transforming the data and performing two-way ANOVA, followed by Dunnett's multiple comparisons test in GraphPad Prism.

siRNA Silencing for Reporter Assays

Two Silencer Select siRNAs against YY1 (siYY1; s224779) and a Silencer Select nontargeting siRNA (siCON; 4390843) were purchased from Life Technologies (Thermo Fisher Scientific). For silencing, Ishikawa cells were co-transfected with the relevant luciferase reporter plasmids and 100 nM of either YY1 or non-targeting siRNAs with Lipofectamine 2000. Luciferase assays were performed as described above after 72 hr. qPCR was performed as described previously[36] to validate YY1 knockdown.

Chromatin Immunoprecipitation

Ishikawa cells were cross-linked with 1% formaldehyde at 37°C for 10 min, rinsed once with ice-cold PBS containing 5% BSA and once with PBS, and harvested in PBS containing 1X protease inhibitor cocktail (Roche). Harvested cells were centrifuged for 2 min at 3,000 rpm. Cell pellets were resuspended in 0.35 mL of lysis buffer (1% SDS, 10 mM EDTA, 50 mM Tris-HCl, pH 8.1, 1X protease inhibitor cocktail) and sonicated 3 times for 15 sec at 70% duty cycle (Branson SLPt) followed by centrifugation at 13,000 rpm for 15 min. Supernatants were collected and diluted in dilution buffer (1% Triton X-100, 2 mM EDTA, 150 mM NaCl, 20 mM Tris-HCl, pH 8.1). Two micrograms of antibody was pre-bound for 6 hr to protein G Dynabeads (Life Technologies) and then added to the diluted chromatin for overnight immunoprecipitation. The magnetic bead-chromatin complexes were collected and washed six times in RIPA buffer (50 mM HEPES [pH 7.6], 1 mM EDTA, 0.7% Na deoxycholate, 1% NP-40, 0.5 M LiCl), then twice with TE buffer. To reverse the cross-linking, the magnetic bead complexes were incubated overnight at 65°C in elution buffer (1% SDS, 0.1 M NaHCO3). DNA fragments were purified using a QIAquick Spin Kit (Qiagen). For qPCR, 2.0 uL from a 100 uL immunoprecipitated chromatin extraction and 40 cycles of amplification were used. All PCR products were sequenced by Sanger sequencing (AGRF). Antibodies used were anti-NFκB p50 (06-886), anti-YY1 (sc-1703-X), and control IgG (sc-2027). ChIP primers are listed in Table S5.

Results

Association and Likelihood Testing at the 14q32.33 Endometrial Cancer Risk Locus Prioritizes Three SNPs for Follow-Up

A total of 2,922 SNPs with MAF ≥1% and information scores ≥0.4 per endometrial cancer dataset were included in the fine-mapping analysis, representing 76.6% of the SNPs with a MAF ≥1% in the 1000 Genomes 2012 reference panel in this region (hg19 chr14: 104,743,220-105,743,220; Table S6). Considering SNPs correlated ($r^2 > 0.2$) with rs2498796, the previously reported top endometrial cancer risk SNP at this locus,[6] coverage was good, with >94% of correlated SNPs in the same 1000 Genomes reference panel represented in each dataset.

Association and log-likelihood tests prioritized two SNPs for bioinformatic and functional follow-up: rs2498794 (OR = 1.13, 95% CI 1.09–1.17, p value 8.7×10^{-9}) and rs2494737 (OR = 1.13, 95% CI 1.08–1.17, p value 2.5×10^{-7}; Table 1 and Table S7). No other SNP was significant at $p < 1 \times 10^{-4}$ in analyses conditioning on rs2498794 or rs2494737 (r^2 to each other 0.54, Table S7), confirming the single association signal at this locus (Figure 1). As SNP rs2498796 (p value 1.2×10^{-6} in the current analysis, r^2 to rs2498794 0.43, and to rs2494737 0.83; Table 1) was the original endometrial cancer risk SNP reported for this locus, it was also included in the bioinformatic and functional analyses detailed below. Neither rs2498794 nor rs2494737 had been reported in our previous genome-wide analysis because of the more stringent imputation threshold used in that study.

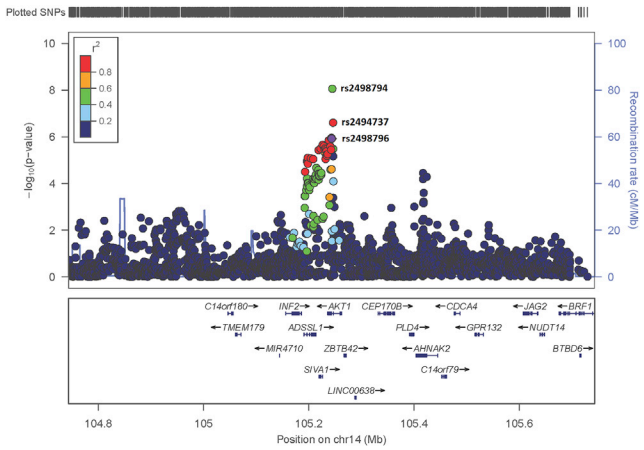

Figure 1. Regional Association Plot for the 14q32.33 Endometrial Cancer Risk Locus
The location (Build 19) and $-\log_{10}$ p value of the original top SNP at this locus, rs2498796,[6] is shown in purple; all other SNPs are shown in colors corresponding to their r^2 (linkage disequilibrium) value with rs2498796.

The Top Candidate SNPs Fall within a Putative Regulatory Element that Frequently Interacts with *AKT1* and *ZBTB42* Promoter Regions

Analysis of *cis* enhancer-gene interactions using data from Hnisz et al.[23] and PreSTIGE[24] identified *AKT1*, *ZBTB42*, *SIVA1*, *ADSSL1*, and *INF2* as potential candidate target genes of a PRE located in the region containing the top candidate SNPs (Figure 2). To determine the target gene(s) of the PRE, we performed chromosome conformation capture (3C) using an anchor primer within the PRE and primers within restriction fragments spanning all protein coding gene promoters within 2Mb of the PRE. The results showed that the PRE frequently interacted with a canonical and alternative promoter of *AKT1* and the *ZBTB42* promoter in both Ishikawa and EN-1078D endometrial cancer cells (Figures 3A and 3B). To assess any potential impact of SNP rs2494737 on chromatin looping, we performed allele-specific 3C in heterozygous Ishikawa cell lines. A primer was designed to incorporate the rs2494737 into the 3C PCR products, which were then Sanger sequenced. The sequence profiles indicate that the cancer risk- and non-risk-associated rs2494737 alleles form loops with the *AKT1* and *ZBTB42* promoters with equal efficiencies (Figure S1). No significant interactions were detected between the PRE and other flanking genes including *SIVA1*, *ADSSL1*, *INF2*, and *CEPB170B* (Figure 3A and Figure S2).

SNP rs2494737 Affects the Regulatory Capability of the PRE on *AKT1* Promoter Regions

The regulatory capability of the PRE, combined with the effects of candidate SNPs, was examined in luciferase

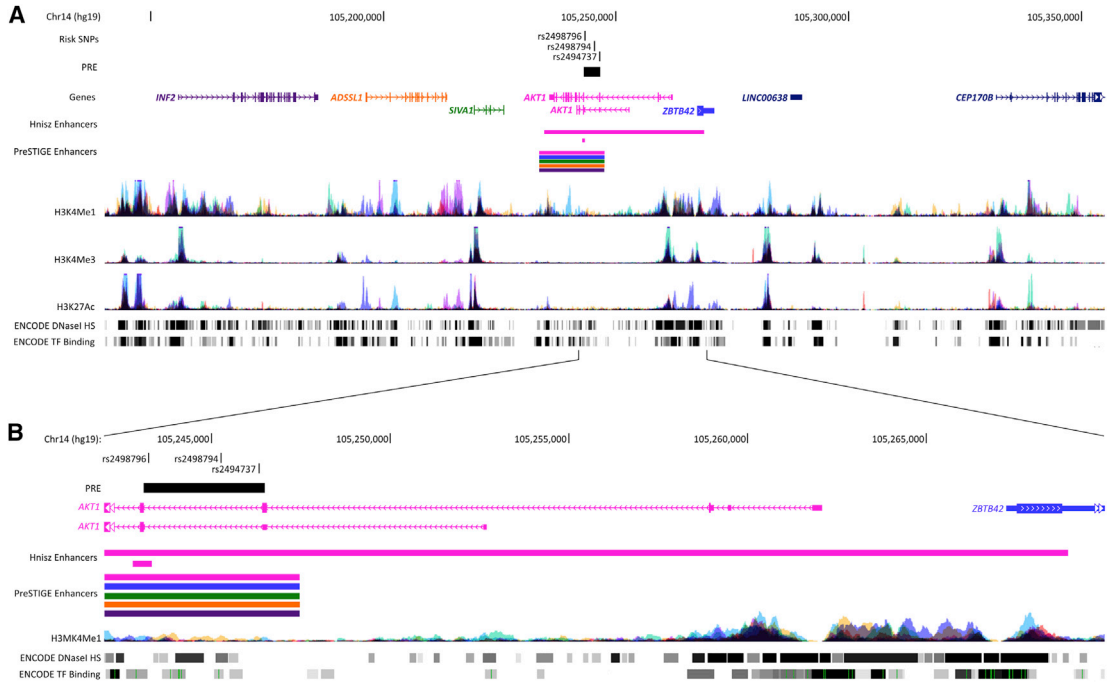

Figure 2. Regulatory Landscape at the 14q32.33 Endometrial Cancer Risk Locus
(A) The location of the candidate SNPs are represented by black ticks, and the PRE is shown as a black box. Gene structures are depicted with exons (vertical boxes) joined by introns (lines). The subset of enhancers predicted in Hnisz et al.[23] and PreSTIGE[24] which overlap the candidate causal SNPs are shown as colored bars, where the color matches its predicted gene target. Regions showing histone binding (H3K4Me1, indicative of regulatory regions; H3K4Me3, indicative of promoters; and H3K27Ac, indicative of active enhancers), *DNAseI* hypersensitivity (indicative of open chromatin, with darker shading indicating stronger experimental signal) and transcription factor (TF) binding in multiple ENCODE cell lines are indicated at the bottom of the panel.
(B) Zoomed-in view of the location of candidate SNPs, PRE, and nearby gene promoter regions.

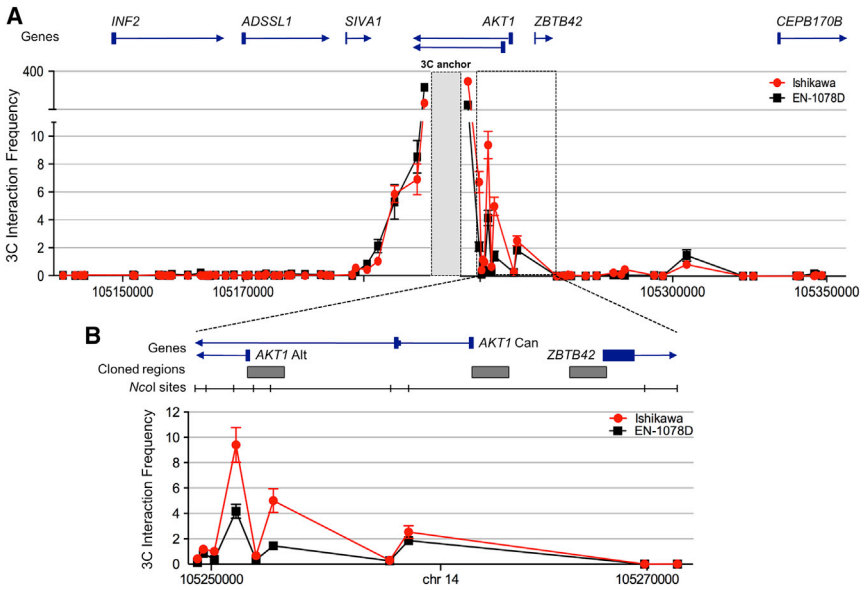

Figure 3. Candidate Causal SNPs Are Located within a PRE that Interacts with the *AKT1* and *ZBTB42* Promoter Regions

(A) 3C interaction profiles between the PRE and local genes in Ishikawa and EN-1078D endometrial cancer cell lines. The 3C anchor (which contains the PRE) is shown as a grey box and significant interactions are outlined.

(B) Zoomed-in view of significant interactions. *AKT1* Can and *AKT1* Alt denote a canonical and alternative *AKT1* promoter (prom) region, respectively. 3C libraries were generated with *Nco*I. Graphs represent three biological replicates. Error bars represent SD. Regions cloned into reporter gene constructs are shown as grey horizontal boxes.

reporter assays in Ishikawa and EN-1078D cell lines. PRE constructs containing the reference (common, protective) alleles of the three candidate SNPs significantly reduced their associated target gene promoter activities, suggesting that the PRE can act as a transcriptional silencer (Figure 4 and Figure S3). Inclusion of the minor (risk-increasing) allele of rs2494737 significantly increased the canonical and alternative *AKT1* promoter activities in both cell lines, but had no effect on the *ZBTB42* promoter. In contrast, inclusion of the minor (risk-increasing) alleles of SNPs rs2498796 and rs2498794 had no significant effects on *AKT1* or *ZBTB42* promoter activities (Figure 4 and Figure S3).

The Risk Allele of SNP rs2494737 Binds the YY1 Transcription Factor

We used bioinformatic analyses and functional studies to examine DNA-protein interactions for the three candidate SNPs. In silico prediction tools including HaploReg[37] and Alibaba2[38] predicted all three SNPs to alter transcription factor (TF) binding (Table S8 and Figure S4). We performed electrophoretic mobility shift assays (EMSAs) to assess binding of TFs to the common (protective) and minor (risk-increasing) alleles of each of these SNPs and showed allele-specific protein binding for rs2494737 and rs2498796 (Figure 5A and Figure S5). Competition with TF binding sites suggested that YY1 binds to the minor (risk-increasing) allele of rs2494737 and NF-κB binds to the common allele of rs2498796 (Figure 5B and Figure S5). No other predicted TFs were able to compete for binding at either site, including CEBPA, AP2, and CREB (Figure S5). Supershift assays using anti-YY1 antiserum indicated that the protein binding the minor allele of rs2494737 is likely to be YY1 (Figure S6). Chromatin immunoprecipitation (ChIP) in heterozygous Ishikawa cells confirmed occupancy of YY1 binding in vivo and showed it is preferentially recruited to the minor *A* (risk-increasing) allele of rs2494737 (Figures 5C and 5D and Figure S7). The importance of YY1 binding was confirmed in cotransfection assays that showed that two independent siRNAs against YY1 repressed the promoter activation in the presence of the minor *A* allele of rs2494737 (Figure 5E and Figure S8). We found no evidence of CEBPA binding to the rs2494737 site or NF-κB binding to the rs2498796 site in vivo.

Gene-Expression Analysis in Uterine Tissue

Association between SNPs in the *AKT1* region and *AKT1* mRNA expression was investigated in both normal and endometrial tumor tissue. In the GTEx dataset, with the three candidate SNPs as input, rs2497896 was associated with increased *AKT1* expression in normal uterine tissue (sample N = 70, p = 0.01, Figure S9), but no eQTL effect was detected for rs2494737 or rs2498794, suggesting a stochastic effect due to the reasonably small sample size as these SNPs are in moderate to high LD with each other. Performing an eQTL search for *AKT1* in uterine tissue returned no results. However, including all GTEx tissues revealed all three candidate SNPs, and others in moderate to high LD, to be highly significantly associated with *AKT1* expression in thyroid tissue (N = 278; rs2494737 = 3.6×10^{-14}, rs2498796 = 5.10×10^{-25}, and rs2498794=6.1×10^{-19}; Table S7), indicating these SNPs are eQTLs for *AKT1* in some cellular contexts. In the TCGA datasets, no SNP in the *AKT1* region was associated with differential *AKT1* expression of any isoform in normal endometrial tissue (N = 29) or in endometrial tumors (N = 526; Figure S10).

Discussion

In the largest association study for endometrial cancer to date, a recent meta-analysis of five GWAS and follow-up

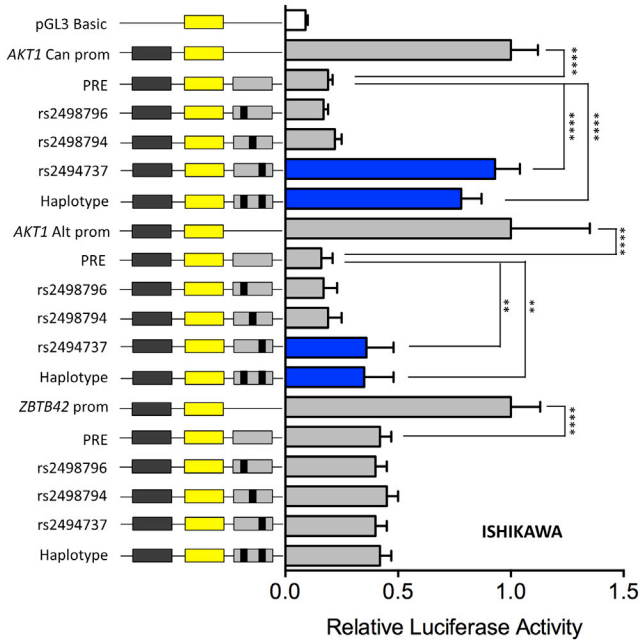

Figure 4. The Risk Allele of SNP rs2494737 Enhances *AKT1* Promoter Activity

Luciferase reporter assays following transient transfection of Ishikawa endometrial cancer cell lines. The putative regulatory element (PRE) containing the major SNP alleles were cloned downstream of target gene promoter-driven luciferase constructs. *AKT1* can and *AKT1* alt denote a canonical and alternative *AKT1* promoter (prom) region, respectively. Minor (risk-increasing) SNP alleles were engineered into the constructs and are designated by the rs ID of the corresponding SNP. Haplotype denotes a construct that contains the minor alleles of rs2498796 and rs2494737. Error bars denote 95% confidence intervals from three independent experiments performed in duplicate. P values were determined by 2-way ANOVA followed by Dunnett's multiple comparisons test (**$p < 0.01$, ****$p < 0.0001$).

datasets revealed the presence of one multi-variant haplotype at the 14q32.33 chromosomal locus associated with the risk of this cancer.[6] In consideration of the fact that our genotyping platforms were not specifically designed for fine-mapping of this region, we conducted in silico fine-mapping of the 14q32 region using SNPs with imputation scores down to 0.4. We identified two SNPs as most likely to be the causal SNPs increasing endometrial cancer risk in this region: SNPs rs2498794 and rs2494737, in moderate and high LD, respectively, with the original hit at this locus rs2498796.[6] Multiple laboratory analyses then confirmed that rs2494737 has a functional impact on the *AKT1* oncogene, a gene of potential biological relevance to endometrial cancer risk as other PI3K pathway mutations have been detected in precursor lesions of complex atypical endometrial hyperplasia.[39]

Our fine-mapping, together with multiple lines of bioinformatic and experimental evidence indicate that rs2494737 is the functional SNP most likely to be relevant for endometrial cancer at the 14q32 risk locus. However, additional bioinformatic analyses indicated multiple regulatory elements across the region that contained several less significantly risk-associated SNPs. Additionally, our SNP coverage of the region was not complete, although ~98% of SNPs in at least moderate LD ($r^2 > 0.2$), and 100% of SNPs in high LD ($r^2 > 0.8$), with rs2498796 in the 1000 Genomes 2012 panel were also present in our datasets, and imputed to high quality scores (>0.71). Therefore, we cannot rule out the possibility that additional SNPs exert effects on *AKT1* expression via alternative mechanisms. For example, at the well-characterized 8q24 risk locus, multiple risk-associated enhancers interact with *MYC* in a tissue-specific manner.[40] Furthermore, a few recent studies have indicated that risk-associated SNPs might also influence epigenetic features,[41,42] adding yet another layer of complexity to the control of gene expression.

Publicly available enhancer data from multiple cell types indicate that the region harboring rs2494737 might target a number of genes in the 14q32 region, some of which are highly plausible endometrial cancer candidate genes. Our 3C analyses show that in endometrial tumor cells the rs2494737 region specifically targets *AKT1* and *ZBTB42*, while luciferase assays showed that the rs2494737 minor *A* (risk-increasing) allele affects only *AKT1*, increasing the activity of both the canonical and an alternative promoter. Therefore, we expect the causal risk allele to result in increased expression of one or more *AKT1* isoforms in vivo. Although we observed no significant effect of the rs2494737 minor allele on overall or isoform-specific *AKT1* expression in normal uterine or endometrial tumor tissue, there are multiple possible explanations. One reason could be that the risk allele affects *AKT1* expression in endometrial epithelial cells that represent only a fraction of the total cells in a normal uterine sample, which is composed of substantially more endometrial stromal cells as well as underlying myometrium. Any effect on expression might also occur only in specific cellular contexts. Further, the lack of association in the normal tissue sample sets examined might also be due to low power, with only 47% and 23% power to detect an effect of a SNP (MAF 0.3) explaining even 5% of the variance in *AKT1* expression in the GTEx and TCGA datasets, respectively. We had 99.9% power to detect the same effect in the larger (N = 526) endometrial tumor dataset, although here any eQTL effect might be difficult to detect due to the overall increase in *AKT1* expression seen in endometrial tumor cells in general. The apparent discrepancy between eQTL results and our in vitro findings is not unprecedented: functional SNPs in *CCND1* (MIM: 168461)[36] and *MYC* (MIM: 190080)[43] show no association with gene expression in human tumor cells, although one *MYC* region SNP (rs6983267) has been demonstrated to have a functional effect in vivo.[43]

AKT1, a serine/threonine kinase highly expressed in the endometrium,[44] regulates many processes including cell metabolism, proliferation, survival, growth, and angiogenesis[45] and is already of considerable interest as a potential therapeutic target for endometrial cancer.[9,46] Activation

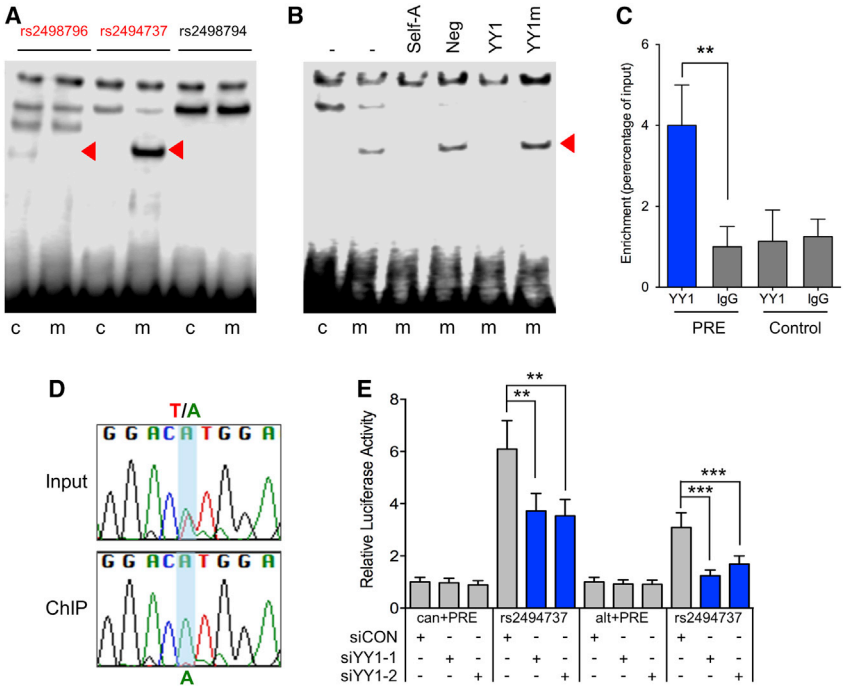

Figure 5. The Risk Allele of rs2494737 Demonstrates Allele-Specific YY1 Binding

(A) EMSAs to detect allele-specific binding of nuclear proteins. Oligonucleotides were incubated with Ishikawa nuclear extracts. Red arrowheads show bands of different mobility detected between the common (C) and minor (risk-increasing) (m) alleles for the candidate causal SNPs.

(B) Oligonucleotides for SNP rs2494737 were incubated with Ishikawa nuclear extracts. Red arrowhead indicates the band that was competed for complex formation on the minor (m) allele. Competitor oligonucleotides are listed above each panel and were used at 100-fold molar excess: (–) no competitor; (Neg) a non-specific competitor; (YY1) consensus binding site; (YY1m) an identical oligonucleotide but with a mutated binding site.

(C) ChIP–qPCR on SNP rs2494737 in heterozygous Ishikawa cell lines. ChIP assays were performed with YY1 antibody or non-immune IgG, a region 3.2kb upstream of the predicted YY1-binding site served as a negative control (Control). Graphs represent two biological replicates. Error bars denote SD. P-values were determined with a two-tailed t-test (**$p < 0.01$).

(D) Sanger sequencing of the PCR fragment generated using primers flanking SNP rs2494737 following YY1 ChIP-qPCR and the input DNA controls.

(E) Luciferase assays in Ishikawa cells shows the effect of YY1 siRNA silencing on the activity of the *AKT1* canonical (can) and alternative (alt) promoter regions with the PRE containing the reference T allele (can+PRE; alt+PRE) or the risk A allele (rs2494737). Error bars denote 95% confidence intervals from three independent experiments performed in duplicate. P values were determined by two-way ANOVA followed by Dunnett's multiple comparisons test (**$p < 0.01$, ***$p < 0.001$). The level of YY1 silencing is shown in Figure S8.

of the PI3K/AKT/mTOR intracellular signaling pathway, of which AKT1 is a member, occurs in numerous cancers[9,45] and up to 80% of endometrial tumors.[47] This pathway activation has been linked to somatic mutations and copy-number alterations in various PI3K/AKT/mTOR pathway genes, including inactivating mutations and deletion of the *PTEN* tumor-suppressor gene and activating mutations or amplifications in the *PIK3R1* (MIM: 171833) and/or *PIK3CA* genes.[48] *AKT1* mutations are rare, with an oncogenic c.49G>A (p.Glu17Lys) (NM_005163.2) mutation occurring in only ~2% of endometrial tumors,[10,11] and activation is thought to result from the concomitant loss or activation of upstream pathway proteins.[48]

Our results indicate a possible additional mechanism whereby the presence of a common SNP allele results in increased *AKT1* transcriptional activity mediated through YY1 and results in an increased risk of endometrial cancer. YY1 is found at elevated levels in numerous cancers, including breast (MIM: 114480), prostate (MIM: 176807), and cervical cancers (MIM: 603956),[49] and was recently demonstrated to be over-expressed particularly in early stage (I and II) endometrial tumors, indicating this transcription factor could be a molecular marker of early tumor development.[50] Of note, YY1 knockdown using small inferring RNA (siRNA) and small hairpin RNA (shRNA) reduced YY1 protein levels, decreased cell proliferation and reduced cell motility of the AN3CA endometrial cancer cell line, while siYY1 injected directly into xenograft tumors in mice delayed endometrial tumor growth.[50] Although these data would suggest YY1 is a potential therapeutic target, transcription factors are notoriously hard to target with small molecules. The data presented here suggests that elevated levels of YY1 are oncogenic in part through upregulation of *AKT1* expression, which is a signaling pathway that is more amenable to drug targeting.[51]

Activation of AKT1 requires translocation to the plasma membrane followed by phosphorylation of the Thr308 and Ser473 residues: high levels of p-AKT1 are a marker of poor prognosis in endometrial and other cancers.[8,52,53] A large number of inhibitors targeting mTOR and/or PI3K have been tested in early clinical trials in multiple tumor types, however, toxicity issues have complicated their ongoing development and many have not been taken forward into large phase III trials. Several AKT inhibitors are also in development, and initial clinical activity recently reported in several different solid tumor types including breast, lung, and gynaecological tumors carrying the *AKT1* c.49G>A (p.Glu17Lys) hotspot mutation.[54,55] There is a current emphasis on reducing systemic toxicities by optimizing scheduling as well as evaluating nanoparticles to target these agents to tumors and reduce systemic exposure,[56] nonetheless it is unlikely that AKT inhibitors developed for treatment of metastatic disease will have an acceptable toxicity profile to be used as chemopreventive agents.

A promising alternative might be the re-positioning of the type 2 diabetes drug metformin. This drug has multiple mechanisms of action targeting both metabolism, by decreasing circulating glucose levels, as well as altering intracellular signaling by activating AMPK.[57] Activation of AMPK has been shown to inhibit mTOR, a downstream effector of PI3K/AKT signaling. Metformin is currently being evaluated in the adjuvant treatment of endometrial cancer[58] as well as large chemoprevention trials (e.g. the Diabetes Prevention Program Outcomes Study). It would be interesting to determine the outcome analyses from these large-scale chemoprevention trials if patients were genotyped and retrospectively stratified based on their germline *AKT1* risk alleles. Perhaps the ability of metformin to blunt mTOR signaling would be reflected in a greater decrease in endometrial cancer incidence in the *AKT1* SNP carriers treated with metformin.

Although our data indicate that *AKT1* is the likely target gene, it is possible that these SNPs also exert functional effects through long-range control of other genes under different conditions of cell activation or in other cell types, including nearby *SIVA1*, *ZBTB42*, *ADSSL1*, and *INF2*. Notably, SIVA1 is reported to activate and suppress apoptosis, a process dysregulated in cancer. Among other roles, SIVA1 can inhibit p53 tumor suppressor functions and is mutated in up to 90% of aggressive endometrial tumors.[26,59] ZBTB42 (zinc finger and BTB domain containing 42) is a poorly characterized member of the C_2H_2 zinc finger protein family[60]. It is highly expressed in subsynaptic nuclei in skeletal muscles underlying the neuromuscular junctions,[60] and might be involved in muscle development.[61] ADSSL1 (Adenylosuccinate Synthase Like 1) is a muscle isozyme that is selectively deleted in carcinogen-induced mouse lung adenocarcinomas.[62] While *INF2* (Inverted Formin 2) encodes a member of the diaphanous-related formin family, which is involved in remodelling the actin and microtubule cytoskeltons.[63] Mutations in this gene are reported to cause a form of autosomal-dominant focal and segmental glomerulosclerosis and Charcot-Marie-Tooth disease.[64,65]

In conclusion, we have identified a common SNP allele associated with endometrial cancer risk that functions to increase *AKT1* expression through YY1-mediated transcription. Identification of an endometrial cancer risk allele within a member of the PI3K/AKT signaling pathway, more commonly activated in tumors by somatic alterations, raises the possibility that well-tolerated inhibitors targeting this pathway could be candidates for evaluation as chemopreventive agents in individuals at high risk of developing endometrial cancer.

Supplemental Data

Supplemental Data includes ten figures, eight tables, and Supplemental Experimental Procedures and can be found with this article online at http://dx.doi.org/10.1016/j.ajhg.2016.04.012.

Acknowledgments

The QIMR Berghofer groups were supported by a Rio Tinto Ride to Conquer Cancer (RTCC)/Weekend to End Women's Cancers (WEWC) Grant and NHMRC project grants 1058415 to S.L.E. and 1031333 to A.B.S. A.M.D acknowledges a CRUK grant C8197/A16565. I.T. acknowledges core funding to the Wellcome Trust Centre for Human Genetics from the Wellcome Trust (090532/Z/09/Z). A.B.S. is supported by an NHMRC Senior Research Fellowship (1061779). The funders have no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Received: January 10, 2016
Accepted: April 19, 2016
Published: June 2, 2016

Web Resources

1000 Genomes, http://www.1000genomes.org
ENCODE, https://www.encodeproject.org/
GEO, http://www.ncbi.nlm.nih.gov/geo/
GTEx Portal, http://www.gtexportal.org/home/
National Cancer Institute LDlink tool, http://analysistools.nci.nih.gov/LDlink/
OMIM, http://www.omim.org/
PreSTIGE, http://genetics.case.edu/prestige/
QTL Genetic Power Calculator, http://pngu.mgh.harvard.edu/~purcell/gpc/
The Cancer Genome Atlas, http://cancergenome.nih.gov/

References

1. Ferlay, J.S.I., Ervik, M., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D.M., Forman, D., and Bray, F. (2013). GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 (Lyon, France: International Agency for Research on Cancer).
2. Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D.M., Forman, D., and Bray, F. (2015). Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int. J. Cancer *136*, E359–E386.
3. Spurdle, A.B., Thompson, D.J., Ahmed, S., Ferguson, K., Healey, C.S., O'Mara, T., Walker, L.C., Montgomery, S.B., Dermitzakis, E.T., Fahey, P., et al.; Australian National Endometrial Cancer Study Group; National Study of Endometrial Cancer Genetics Group (2011). Genome-wide association study identifies a common variant associated with risk of endometrial cancer. Nat. Genet. *43*, 451–454.
4. Painter, J.N., O'Mara, T.A., Batra, J., Cheng, T., Lose, F.A., Dennis, J., Michailidou, K., Tyrer, J.P., Ahmed, S., Ferguson, K., et al.; National Study of Endometrial Cancer Genetics Group (NSECG); CHIBCHA Consortium; Australian National Endometrial Cancer Study Group (ANECS); RENDOCAS; Australian Ovarian Cancer Study (AOCS); GENICA Network (2015). Fine-mapping of the HNF1B multicancer locus identifies candidate variants that mediate endometrial cancer risk. Hum. Mol. Genet. *24*, 1478–1492.
5. Thompson, D.J., O'Mara, T.A., Glubb, D.M., Painter, J.N., Cheng, T., Folkerd, E., Doody, D., Dennis, J., Webb, P.M., Gorman, M., et al. (2015). CYP19A1 fine-mapping and Mendelian

randomisation: estradiol is causal for endometrial cancer. Endocr. Relat. Cancer *23*, 77–91.

6. Cheng, T.H., Thompson, D.J., O'Mara, T.A., Painter, J.N., Glubb, D.M., Flach, S., Lewis, A., French, J.D., Freeman-Mills, L., Church, D., et al.; National Study of Endometrial Cancer Genetics Group (NSECG); Australian National Endometrial Cancer Study Group (ANECS); RENDOCAS; CHIBCHA Consortium; AOCS Group (2016). Five endometrial cancer risk loci identified through genome-wide association analysis. Nat. Genet. http://dx.doi.org/10.1038/ng.3562.
7. Cantley, L.C. (2002). The phosphoinositide 3-kinase pathway. Science *296*, 1655–1657.
8. Salvesen, H.B., Carter, S.L., Mannelqvist, M., Dutt, A., Getz, G., Stefansson, I.M., Raeder, M.B., Sos, M.L., Engelsen, I.B., Trovik, J., et al. (2009). Integrated genomic profiling of endometrial carcinoma associates aggressive tumors with indicators of PI3 kinase activation. Proc. Natl. Acad. Sci. USA *106*, 4834–4839.
9. Slomovitz, B.M., and Coleman, R.L. (2012). The PI3K/AKT/mTOR pathway as a therapeutic target in endometrial cancer. Clin. Cancer Res. *18*, 5856–5864.
10. Shoji, K., Oda, K., Nakagawa, S., Hosokawa, S., Nagae, G., Uehara, Y., Sone, K., Miyamoto, Y., Hiraike, H., Hiraike-Wada, O., et al. (2009). The oncogenic mutation in the pleckstrin homology domain of AKT1 in endometrial carcinomas. Br. J. Cancer *101*, 145–148.
11. Cohen, Y., Shalmon, B., Korach, J., Barshack, I., Fridman, E., and Rechavi, G. (2010). AKT1 pleckstrin homology domain E17K activating mutation in endometrial carcinoma. Gynecol. Oncol. *116*, 88–91.
12. Barrett, J.H., Taylor, J.C., Bright, C., Harland, M., Dunning, A.M., Akslen, L.A., Andresen, P.A., Avril, M.F., Azizi, E., Bianchi Scarrà, G., et al.; GenoMEL Consortium (2015). Fine mapping of genetic susceptibility loci for melanoma reveals a mixture of single variant and multiple variant regions. Int. J. Cancer *136*, 1351–1360.
13. Carvajal-Carmona, L.G., O'Mara, T.A., Painter, J.N., Lose, F.A., Dennis, J., Michailidou, K., Tyrer, J.P., Ahmed, S., Ferguson, K., Healey, C.S., et al.; National Study of Endometrial Cancer Genetics Group (NSECG); Australian National Endometrial Cancer Study Group (ANECS); RENDOCAS; Australian Ovarian Cancer Study (AOCS); GENICA Network (2015). 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.
14. O'Mara, T.A., Glubb, D.M., Painter, J.N., Cheng, T., Dennis, J., Attia, J., Holliday, E.G., McEvoy, M., Scott, R.J., Ashton, K., et al.; Australian National Endometrial Cancer Study Group (ANECS); National Study of Endometrial Cancer Genetics Group (NSECG); RENDOCAS; AOCS Group (2015). Comprehensive genetic assessment of the ESR1 locus identifies a risk region for endometrial cancer. Endocr. Relat. Cancer *22*, 851–861.
15. Howie, B.N., Donnelly, P., and Marchini, J. (2009). A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. *5*, e1000529.
16. Marchini, J., Howie, B., Myers, S., McVean, G., and Donnelly, P. (2007). A new multipoint method for genome-wide association studies by imputation of genotypes. Nat. Genet. *39*, 906–913.
17. Willer, C.J., Li, Y., and Abecasis, G.R. (2010). METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics *26*, 2190–2191.
18. Pruim, R.J., Welch, R.P., Sanna, S., Teslovich, T.M., Chines, P.S., Gliedt, T.P., Boehnke, M., Abecasis, G.R., and Willer, C.J. (2010). LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics *26*, 2336–2337.
19. Edwards, S.L., Beesley, J., French, J.D., and Dunning, A.M. (2013). Beyond GWASs: illuminating the dark road from association to function. Am. J. Hum. Genet. *93*, 779–797.
20. Udler, M.S., Tyrer, J., and Easton, D.F. (2010). Evaluating the power to discriminate between highly correlated SNPs in genetic association studies. Genet. Epidemiol. *34*, 463–468.
21. Machiela, M.J., and Chanock, S.J. (2015). LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. Bioinformatics *31*, 3555–3557.
22. Birney, E., Stamatoyannopoulos, J.A., Dutta, A., Guigó, R., Gingeras, T.R., Margulies, E.H., Weng, Z., Snyder, M., Dermitzakis, E.T., Thurman, R.E., et al.; ENCODE Project Consortium; NISC Comparative Sequencing Program; Baylor College of Medicine Human Genome Sequencing Center; Washington University Genome Sequencing Center; Broad Institute; Children's Hospital Oakland Research Institute (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature *447*, 799–816.
23. Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-André, V., Sigova, A.A., Hoke, H.A., and Young, R.A. (2013). Super-enhancers in the control of cell identity and disease. Cell *155*, 934–947.
24. Corradin, O., Saiakhova, A., Akhtar-Zaidi, B., Myeroff, L., Willis, J., Cowper-Sal lari, R., Lupien, M., Markowitz, S., and Scacheri, P.C. (2014). 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.
25. Consortium, G.T.; GTEx Consortium (2013). The Genotype-Tissue Expression (GTEx) project. Nat. Genet. *45*, 580–585.
26. Kandoth, C., Schultz, N., Cherniack, A.D., Akbani, R., Liu, Y., Shen, H., Robertson, A.G., Pashtan, I., Shen, R., Benz, C.C., et al.; Cancer Genome Atlas Research Network (2013). Integrated genomic characterization of endometrial carcinoma. Nature *497*, 67–73.
27. Flicek, P., Amode, M.R., Barrell, D., Beal, K., Billis, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fitzgerald, S., et al. (2014). Ensembl 2014. Nucleic Acids Res. *42*, D749–D755.
28. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.
29. DeLuca, D.S., Levin, J.Z., Sivachenko, A., Fennell, T., Nazaire, M.D., Williams, C., Reich, M., Winckler, W., and Getz, G. (2012). RNA-SeQC: RNA-seq metrics for quality control and process optimization. Bioinformatics *28*, 1530–1532.
30. Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics *12*, 323.
31. Li, Y., Willer, C., Sanna, S., and Abecasis, G. (2009). Genotype imputation. Annu. Rev. Genomics Hum. Genet. *10*, 387–406.

32. Li, Y., Willer, C.J., Ding, J., Scheet, P., and Abecasis, G.R. (2010). MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. Genet. Epidemiol. *34*, 816–834.
33. Fuchsberger, C., Abecasis, G.R., and Hinds, D.A. (2014). minimac2: faster genotype imputation. Bioinformatics.
34. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J., and Abecasis, G.R. (2012). Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat. Genet. *44*, 955–959.
35. Ghoussaini, M., Edwards, S.L., Michailidou, K., Nord, S., Cowper-Sal Lari, R., Desai, K., Kar, S., Hillman, K.M., Kaufmann, S., Glubb, D.M., et al.; Australian Ovarian Cancer Management Group; Australian Ovarian Cancer Management Group (2014). Evidence that breast cancer risk at the 2q35 locus is mediated through IGFBP5 regulation. Nat. Commun. *4*, 4999.
36. French, J.D., Ghoussaini, M., Edwards, S.L., Meyer, K.B., Michailidou, K., Ahmed, S., Khan, S., Maranian, M.J., O'Reilly, M., Hillman, K.M., et al.; GENICA Network; kConFab Investigators (2013). Functional variants at the 11q13 risk locus for breast cancer regulate cyclin D1 expression through long-range enhancers. Am. J. Hum. Genet. *92*, 489–503.
37. Ward, L.D., and Kellis, M. (2015). HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. Nucleic Acids Res. *44*, D877–D881.
38. Grabe, N. (2002). AliBaba2: context specific identification of transcription factor binding sites. In Silico Biol. (Gedrukt) *2*, S1–S15.
39. Hayes, M.P., Wang, H., Espinal-Witter, R., Douglas, W., Solomon, G.J., Baker, S.J., and Ellenson, L.H. (2006). PIK3CA and PTEN mutations in uterine endometrioid carcinoma and complex atypical hyperplasia. Clin. Cancer Res. *12*, 5932–5935.
40. Ahmadiyeh, N., Pomerantz, M.M., Grisanzio, C., Herman, P., Jia, L., Almendro, V., He, H.H., Brown, M., Liu, X.S., Davis, M., et al. (2010). 8q24 prostate, breast, and colon cancer risk loci show tissue-specific long-range interaction with MYC. Proc. Natl. Acad. Sci. USA *107*, 9742–9746.
41. Voisin, S., Almén, M.S., Zheleznyakova, G.Y., Lundberg, L., Zarei, S., Castillo, S., Eriksson, F.E., Nilsson, E.K., Blüher, M., Böttcher, Y., et al. (2015). Many obesity-associated SNPs strongly associate with DNA methylation changes at proximal promoters and enhancers. Genome Med. *7*, 103.
42. Bell, C.G., Finer, S., Lindgren, C.M., Wilson, G.A., Rakyan, V.K., Teschendorff, A.E., Akan, P., Stupka, E., Down, T.A., Prokopenko, I., et al.; International Type 2 Diabetes 1q Consortium (2010). Integrated genetic and epigenetic analysis identifies haplotype-specific methylation in the FTO type 2 diabetes and obesity susceptibility locus. PLoS ONE *5*, e14040.
43. Sur, I.K., Hallikas, O., Vähärautio, A., Yan, J., Turunen, M., Enge, M., Taipale, M., Karhu, A., Aaltonen, L.A., and Taipale, J. (2012). Mice lacking a Myc enhancer that includes human SNP rs6983267 are resistant to intestinal tumors. Science *338*, 1360–1363.
44. Uhlén, M., Fagerberg, L., Hallström, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., Asplund, A., et al. (2015). Proteomics. Tissue-based map of the human proteome. Science *347*, 1260419.
45. Vivanco, I., and Sawyers, C.L. (2002). The phosphatidylinositol 3-Kinase AKT pathway in human cancer. Nat. Rev. Cancer *2*, 489–501.
46. Salvesen, H.B., Haldorsen, I.S., and Trovik, J. (2012). Markers for individualised therapy in endometrial carcinoma. Lancet Oncol. *13*, e353–e361.
47. Cheung, L.W., Hennessy, B.T., Li, J., Yu, S., Myers, A.P., Djordjevic, B., Lu, Y., Stemke-Hale, K., Dyer, M.D., Zhang, F., et al. (2011). High frequency of PIK3R1 and PIK3R2 mutations in endometrial cancer elucidates a novel mechanism for regulation of PTEN protein stability. Cancer Discov. *1*, 170–185.
48. Fabi, F., and Asselin, E. (2014). Expression, activation, and role of AKT isoforms in the uterus. Reproduction *148*, R85–R95.
49. Castellano, G., Torrisi, E., Ligresti, G., Malaponte, G., Militello, L., Russo, A.E., McCubrey, J.A., Canevari, S., and Libra, M. (2009). The involvement of the transcription factor Yin Yang 1 in cancer development and progression. Cell Cycle *8*, 1367–1372.
50. Yang, Y., Zhou, L., Lu, L., Wang, L., Li, X., Jiang, P., Chan, L.K., Zhang, T., Yu, J., Kwong, J., et al. (2013). A novel miR-193a-5p-YY1-APC regulatory axis in human endometrioid endometrial adenocarcinoma. Oncogene *32*, 3432–3442.
51. Courtney, K.D., Corcoran, R.B., and Engelman, J.A. (2010). The PI3K pathway as drug target in human cancer. J. Clin. Oncol. *28*, 1075–1083.
52. Gungorduk, K., Ertas, I.E., Sahbaz, A., Ozvural, S., Sarica, Y., Ozdemir, A., Sayhan, S., Gokcu, M., Yilmaz, B., Sanci, M., et al. (2014). Immunolocalization of ERK1/2 and p-AKT in normal endometrium, endometrial hyperplasia, and early and advanced stage endometrioid endometrial adenocancer and their prognostic significance in malignant group. Eur. J. Obstet. Gynecol. Reprod. Biol. *179*, 147–152.
53. Zhu, Z., Yu, W., Fu, X., Sun, M., Wei, Q., Li, D., Chen, H., Xiang, J., Li, H., Zhang, Y., et al. (2015). Phosphorylated AKT1 is associated with poor prognosis in esophageal squamous cell carcinoma. J. Exp. Clin. Cancer Res. *34*, 95.
54. Hyman, D.M., Smyth, L., Bedard, P.L., Oza, A., Dean, E., Armstrobg, A., Lima, J., Bando, H., Kabos, P., Perez-Fidalgo, J.A., et al. (2015). AZD5363, a catalytic pan-Akt inhibitor, in Akt1 E17K mutation positive advanced solid tumors . Molecular Targets and Cancer Therapeutics Meeting Boston Clinical trial number NCT01226316.
55. Eathiraj, S., Schwartz, B., Yu, Y., Wick, M.J., Hall, T., Chai, F., Sachdev, J., and Abbadessa, G. (2015). Targeting PI3K pathway dependent endometrial tumors with allosteric AKT inhibitors, ARQ 092 and ARQ 751 Molecular Targets and Cancer Therapeutics Meeting Boston.
56. Cadzow, L., Lam, M.H., Wang, H., DeWitt, D., Pucci, V., Mo, J.-R., Lewies-Clark, E., Ferguson, H., Gindy, M., Low, S., et al. (2015). Accurins improve the pharmacokinetics, pharmacodynamics, tolerability and anti-tumor activity of the AKT inhibitor MK-2206 (Molecular Targets and Cancer Therapeutics Meeting Boston).
57. Rena, G., Pearson, E.R., and Sakamoto, K. (2013). Molecular mechanism of action of metformin: old or new insights? Diabetologia *56*, 1898–1906.
58. Porta, C., Paglino, C., and Mosca, A. (2014). Targeting PI3K/Akt/mTOR Signaling in Cancer. Front. Oncol. *4*, 64.
59. Wang, X., Zha, M., Zhao, X., Jiang, P., Du, W., Tam, A.Y., Mei, Y., and Wu, M. (2013). Siva1 inhibits p53 function by acting as an ARF E3 ubiquitin ligase. Nat. Commun. *4*, 1551.
60. Devaney, S.A., Mate, S.E., Devaney, J.M., and Hoffman, E.P. (2011). Characterization of the ZBTB42 gene in humans and mice. Hum. Genet. *129*, 433–441.

61. Patel, N., Smith, L.L., Faqeih, E., Mohamed, J., Gupta, V.A., and Alkuraya, F.S. (2014). ZBTB42 mutation defines a novel lethal congenital contracture syndrome (LCCS6). Hum. Mol. Genet. *23*, 6584–6593.
62. Miller, J.C., Blake, D.C., Jr., and Herzog, C.R. (2009). Adenylosuccinate synthetase 1 gene is a novel target of deletion in lung adenocarcinoma. Mol. Carcinog. *48*, 1116–1122.
63. Chesarone, M.A., DuPage, A.G., and Goode, B.L. (2010). Unleashing formins to remodel the actin and microtubule cytoskeletons. Nat. Rev. Mol. Cell Biol. *11*, 62–74.
64. Barua, M., Brown, E.J., Charoonratana, V.T., Genovese, G., Sun, H., and Pollak, M.R. (2013). Mutations in the INF2 gene account for a significant proportion of familial but not sporadic focal and segmental glomerulosclerosis. Kidney Int. *83*, 316–322.
65. Boyer, O., Nevo, F., Plaisier, E., Funalot, B., Gribouval, O., Benoit, G., Huynh Cong, E., Arrondel, C., Tete, M.J., Montjean, R., et al. (2011). INF2 mutations in Charcot-Marie-Tooth disease with glomerulopathy. N Engl J Med *365*, 2377–2388.

The American Journal of Human Genetics, Volume *98*

Supplemental Data

A Common Variant at the 14q32 Endometrial Cancer Risk Locus Activates *AKT1* through YY1 Binding

Jodie N. Painter, Susanne Kaufmann, Tracy A. O'Mara, Kristine M. Hillman, Haran Sivakumaran, Hatef Darabi, Timothy H.T. Cheng, John Pearson, Stephen Kazakoff, Nicola Waddell, Erling A. Hoivik, Ellen L. Goode, Rodney J. Scott, Ian Tomlinson, Alison M. Dunning, Douglas F. Easton, Juliet D. French, Helga B. Salvesen, Pamela M. Pollock, Deborah J. Thompson, Amanda B. Spurdle, and Stacey L. Edwards

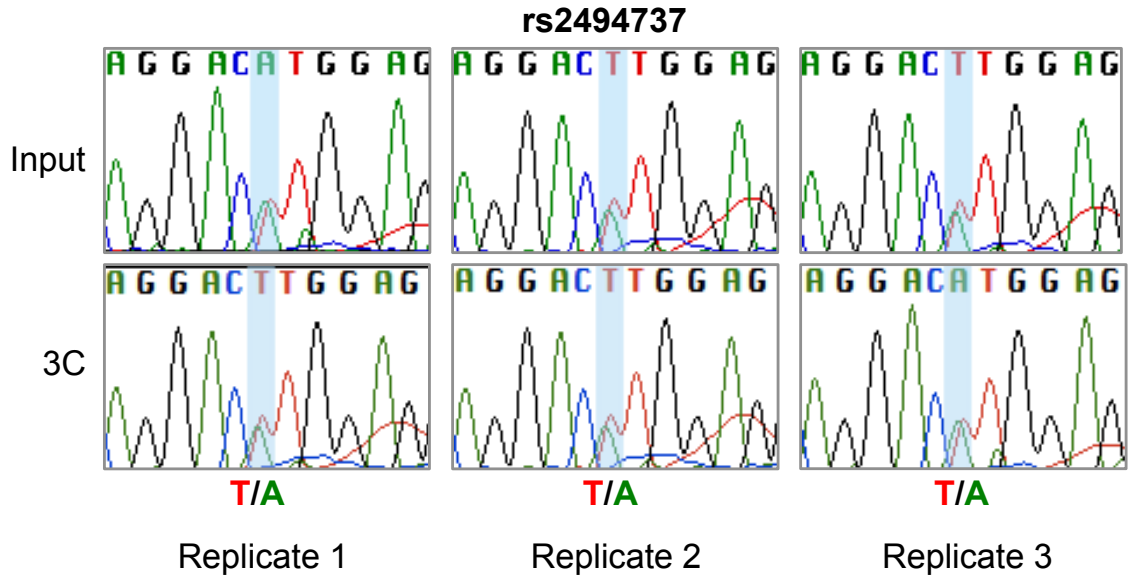

Figure S1. **Allele-specific 3C in Ishikawa endometrial cancer cell lines**. 3C followed by sequencing for the rs2494737-containing region in heterozygous Ishikawa endometrial cancer cells. Chromatograms represent three independent 3C libraries generated and sequenced. 3C libraries were generated with *Nco*I, with the anchor primer designed to incorporate the SNP into 3C PCR products.

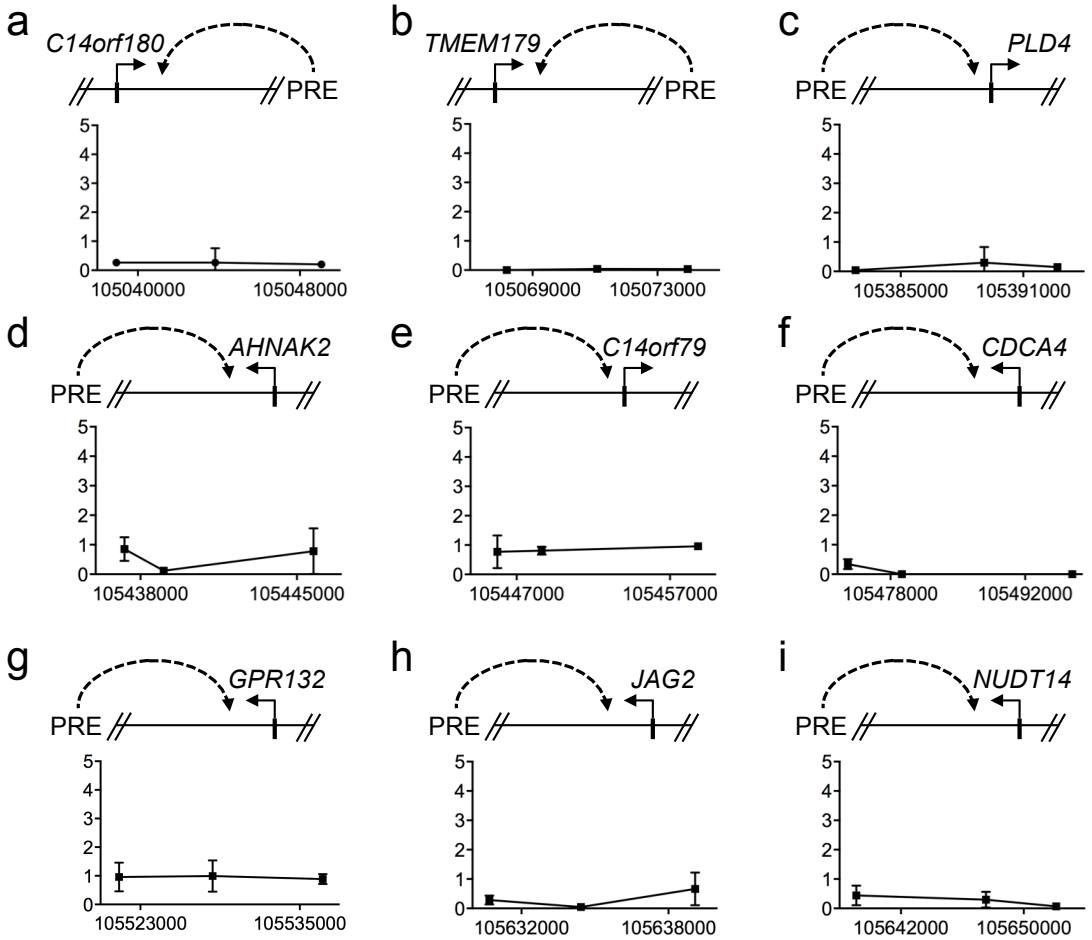

Figure S2. **Chromatin interactions at 14q32 in Ishikawa endometrial cancer cell lines**. 3C interaction profiles between the putative regulatory element (PRE; containing rs2498796, rs2498794 and rs2494737) and **(a)** *C14ORF180,* **(b)** *TMEM179,* **(c)** *PLD4,* **(d)** *AHNAK2,* **(e)** *C14ORF79,* **(f)** *CDCA4,* **(g)** *GPR132,* **(h)** *JAG2* and **(i)** *NUDT14* promoter regions. 3C libraries were generated with *Nco*I, with the anchor point set at the PRE. A physical map of the region interrogated by 3C is shown above. Graph represents three independent replicates. Error bars denote SD.

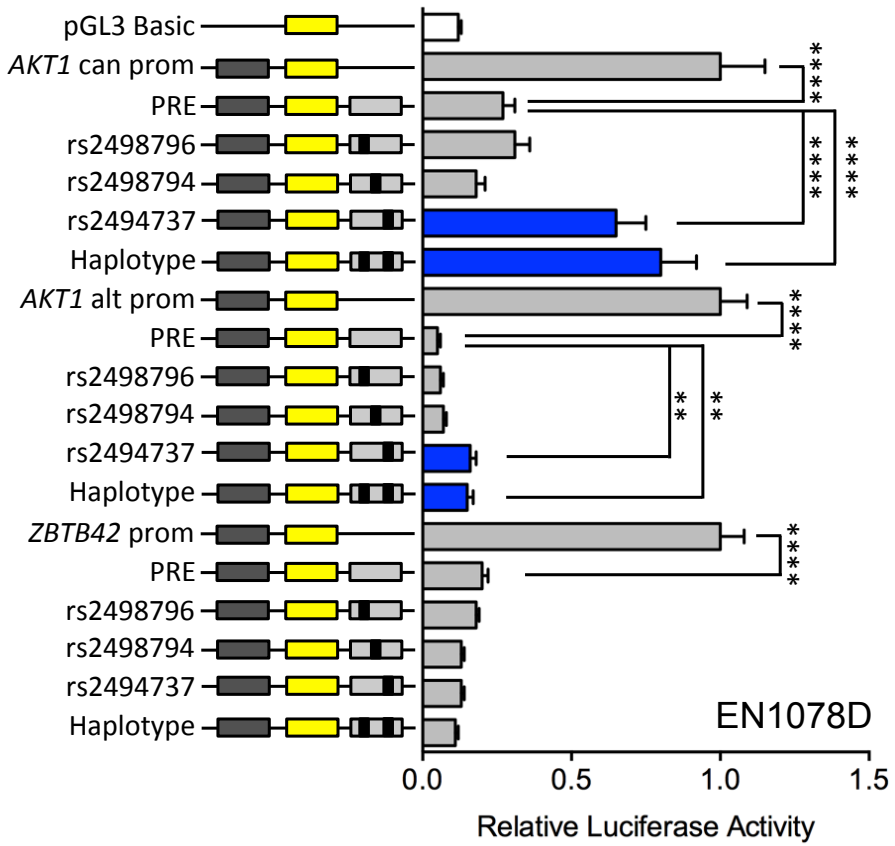

Figure S3. Luciferase reporter assays in EN-1078D endometrial cancer cells. The putative regulatory element (PRE) containing the major SNP alleles were cloned downstream of target gene promoter-driven luciferase constructs. *AKT1* can prom and *AKT1* alt prom denote a canonical and alternative *AKT1* promoter (prom) region, respectively. Minor SNP alleles were engineered into the constructs and are designated by the rs ID of the corresponding SNP. Haplotype denotes a construct that contains the minor alleles of rs2498796 and rs2494737. Error bars denote 95% confidence intervals from three independent experiments performed in duplicate. P-values were determined by 2-way ANOVA followed by Dunnett's multiple comparisons test (**P<0.01, ****P<0.0001).

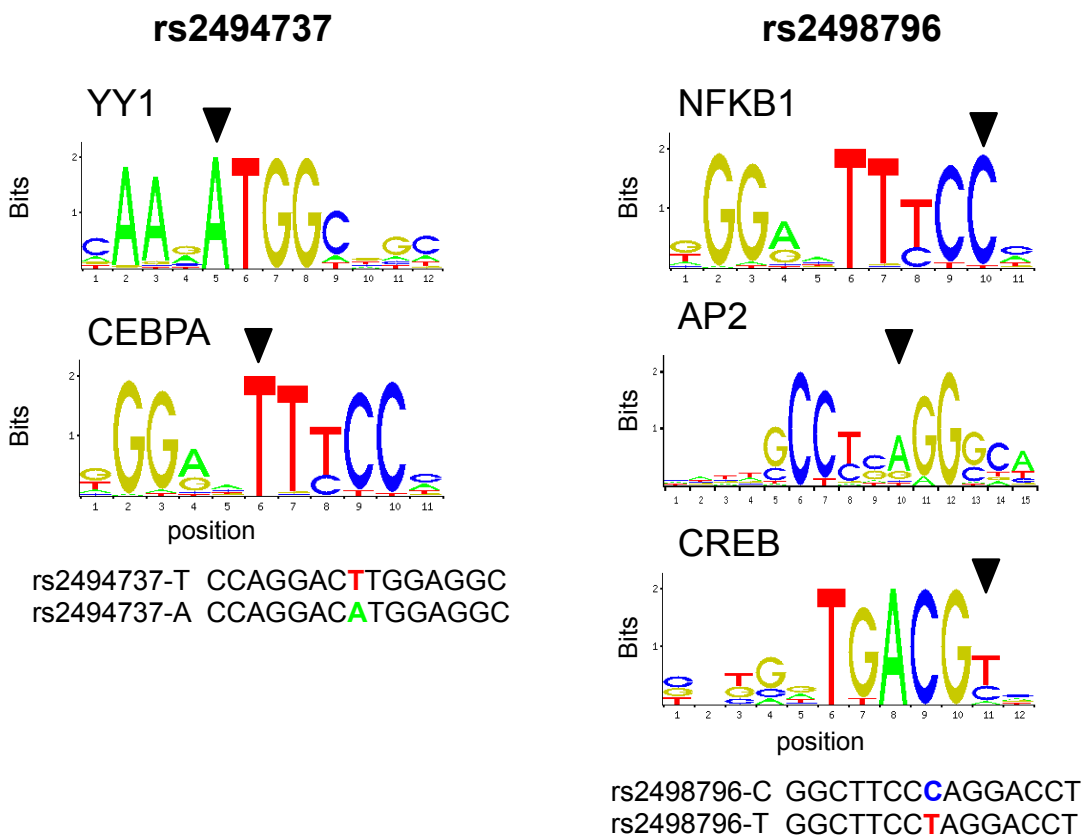

Figure S4. **Transcription factor binding.** Position weight matrix (PWM) of YY1, CEBPA, NFKB1, AP2 and CREB from JASPAR, with homology to the risk-associated alleles of rs2494737 and rs2498796 colored below. Predicted SNP changes are indicated by black arrowheads.

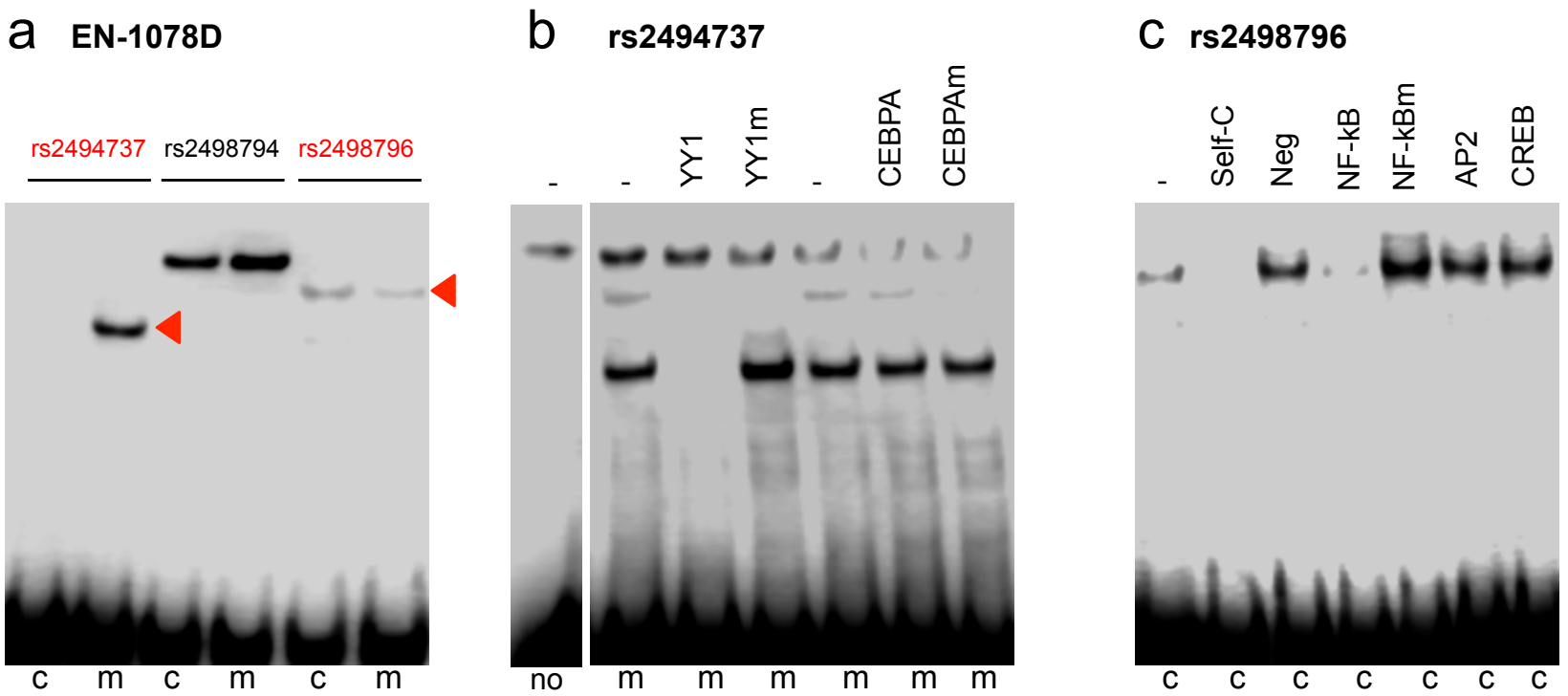

Figure S5. EMSAs for candidate causal SNPs to detect allele-specific binding of nuclear proteins. (a) Oligonucleotides were incubated with EN-1078D nuclear extracts. Red arrowheads show bands of different mobility or intensity detected between the common (c) and minor (m) alleles for the three candidate causal SNPs. Oligonucleotides for SNPs rs2494737 **(b)** and rs2498796 **(c)** were incubated with Ishikawa nuclear extracts. Competitor oligonucleotides are listed above each panel and were used at 100-fold molar excess: (no) no oligonucleotide; (-) no competitor; YY1 consensus binding site; YY1m, an identical oligonucleotide but with a mutated binding site (independent replicate of Figure 4a); CEPBA consensus binding site; CEBPAm, an identical oligonucleotide but with a mutated binding site; NF-kB consensus binding site; NF-kBm, an identical oligonucleotide but with a mutated binding site; AP2 and CREB consensus binding sites. Negative control (Neg) denotes a non-specific competitor.

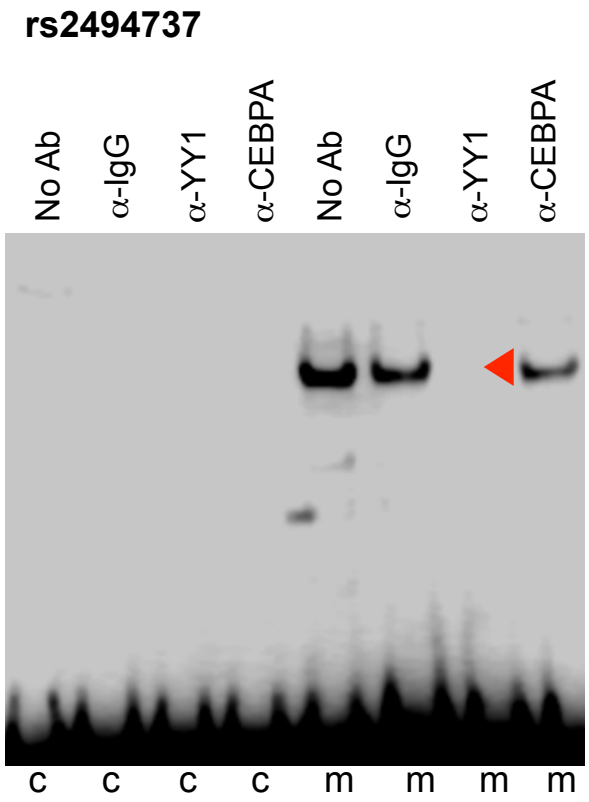

Figure S6. EMSA-supershift for candidate causal SNP rs2494737. Oligonucleotide duplexes for the common (c) or minor (m) alleles of SNP rs2494737 and antibodies against YY1 or CEPBA were incubated with Ishikawa nuclear extracts. Rabbit IgG was used as a negative control. The red arrowhead denotes the YY1 supershifted complex.

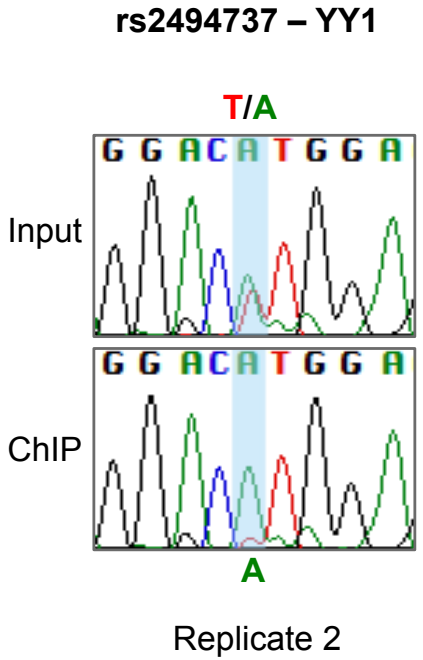

Figure S7. YY1 transcription factor binding *in vivo*. Sanger sequencing of the PCR fragment generated using primers flanking SNP rs2494737 in heterozygous Ishikawa endometrial cancer cells following YY1 ChIP-qPCR and the input DNA controls. Primers are listed in **Table S5.**

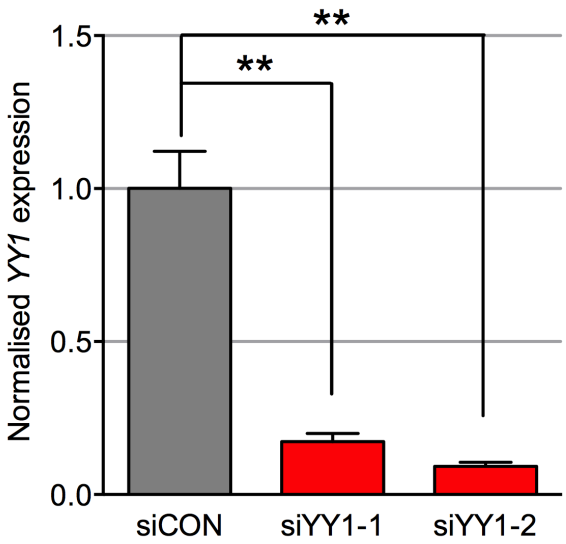

Figure S8. TaqMan real-time PCR assays confirming knockdown of *YY1* in Ishikawa cells. siCON is a nontargeting negative control and siYY1-1 and siYY1-2 are two independent siRNAs targeting *YY1*. Error bars denote the standard error of the mean from three experiments performed in duplicate. Statistical significance was determined by a paired t-test (**P<0.01)

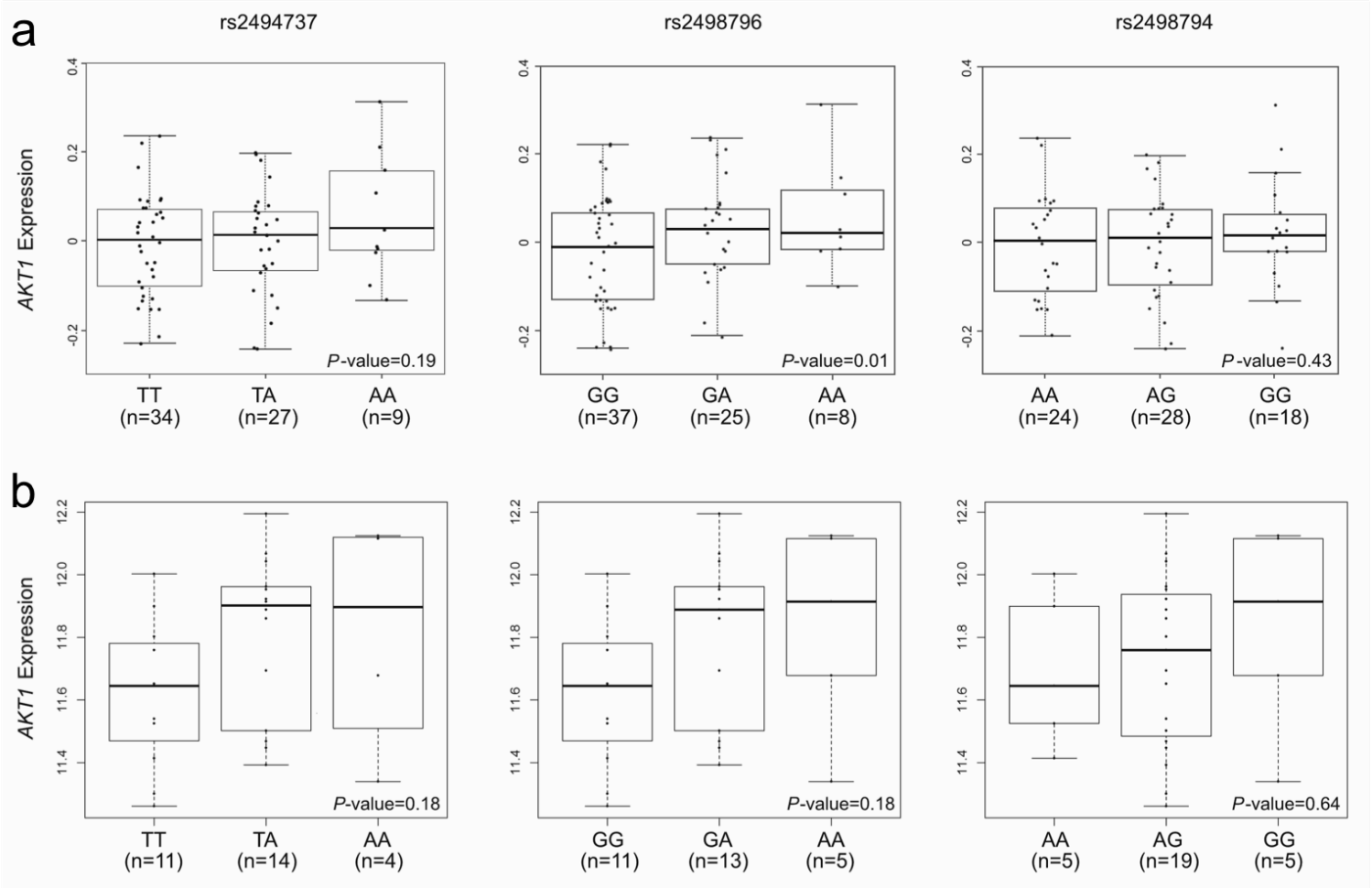

Figure S9. Associations of candidate causal SNPs with overall expression of *AKT1* in uterine samples from the **(A)** GTEx[1] database and **(B)** TCGA[2] dataset. The x-axis of each plot corresponds to the three observed SNP genotypes and the y-axis represents either log2-normalized gene expression values (GTEx) or RSEM gene expression values (TCGA). For the TCGA data, prior to the eQTL, analyses the expression data were adjusted to account for copy-number at the *AKT1* locus, and the three candidate SNPs were imputed with the following RSQR quality scores: rs2494737=0.57, rs2498796=0.72 and rs2498794=0.46.

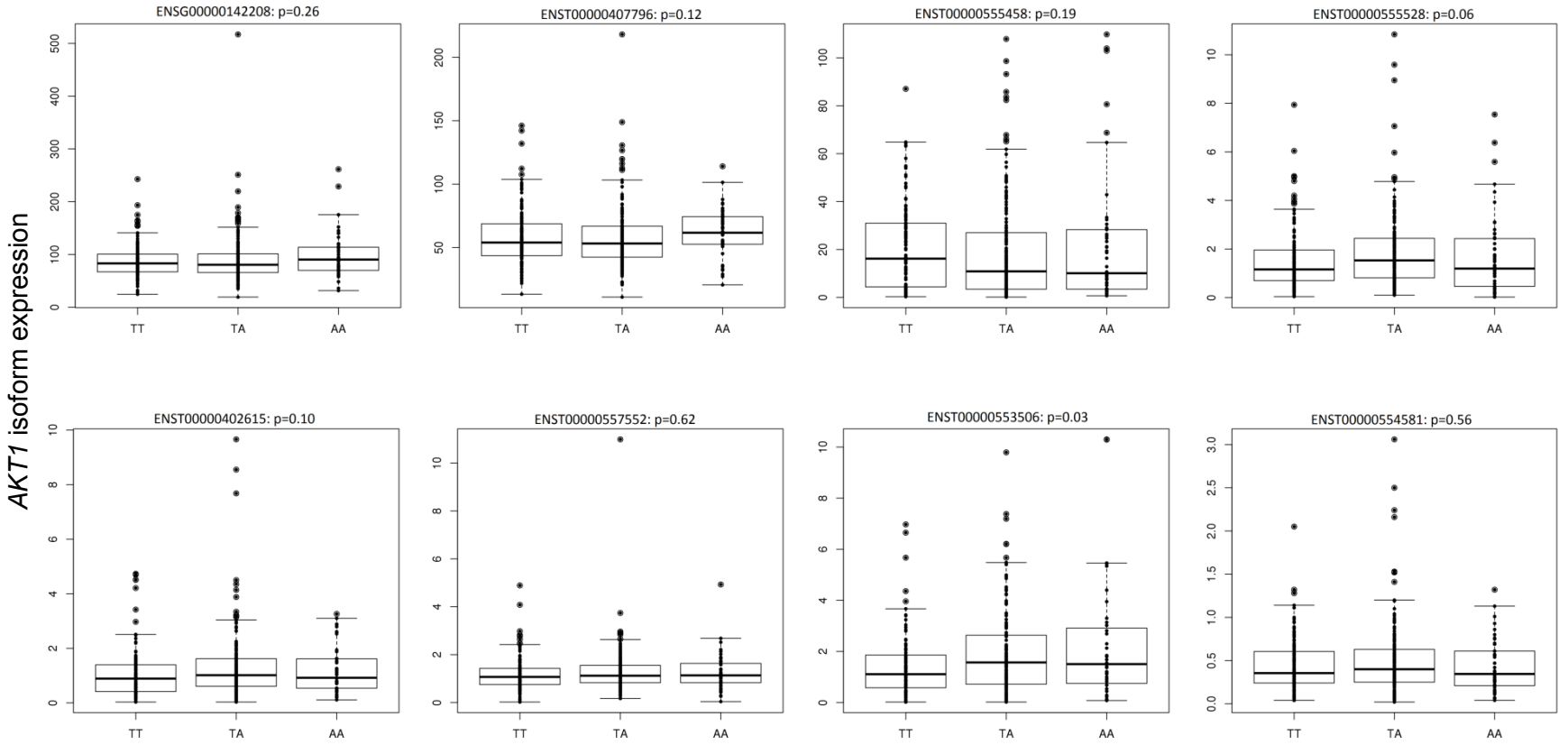

Figure S10. Associations of SNP rs2494737 with expression of *AKT1* isoforms in endometrial tumour samples from the TCGA dataset.[2] The *x*-axis of each plot corresponds to the three SNP genotypes and the *y*-axis represents the RSEM gene expression values for each isoform in 526 unique samples. Data was generated using single-read or paired-end RNA sequencing where SNP data was available. The main *AKT1* isoform is ENSG00000142208. Isoform ENST00000555380, corresponding to the 'alt' promoter examined in luciferase assays detailed above, was not expressed in the TCGA tumour or adjacent normal tissue datasets. Only isoforms for which expression was detected in >80% of samples are shown. Genotype at the risk SNP rs2494737 was not associated with differential expression for any isoform once multiple testing was taken into account (where the Bonferroni corrected *P*-value for a significant association was 0.05/8 transcripts=0.006). Results for SNP rs2494737 in adjacent normal tissue, and for SNPs rs2498794 and rs2498796 in tumour and adjacent normal tissue were all not significant (data not shown).

Table S1. Oligonucleotides used in 3C assays.

3C Primer (*Nco*I)	Sequence (5' to 3')
PRE bait	CGCTACAGGTAAGGAATAAAGCCACAGCAGG
Allele-specific PRE bait	CCTTAGGACTCAGCCTGGAGACTCCCACC
Fragment 1	GCTGGCAGAGAGAAGCTGTGTATAAGCCTGG
Fragment 2	CCTGTGTGCACATAGCTCAGGGTTCTGC
Fragment 3	GCACAGTGTCTGGTTTCCTCCACTCAGC
Fragment 4	GCTGAGAAGTGGAGTGGGATAAGACGATGATAGG
Fragment 5	GGTGTGGGCGTTCTGAGAGAAATCCTCC
Fragment 6	TGCACAGACATGAGTGGCCTGAGAACG
Fragment 7	CAGTCTCACCCTGAATCAGAGCCGTCC
Fragment 8	GGAGGATTCTGGTGACGAGCTCCTGG
Fragment 9	TGACACATGCTGGAGGCTCAAAGGAGC
Fragment 10	GAGCTCCCACACTGTGCTGTGGAAGG
Fragment 11	ATGTAGGCATTCGGATGGAGGTGCTGG
Fragment 12	GCCTCTGCTCGTGTTTTCTGCCTTTGC
Fragment 13	ACCAGGAGGTCTTTGCCTCCCTGTTCC
Fragment 14	GTGAGCTGCTCCCCGGTGTCTGC
Fragment 15	CACTTCCTCCAGGGTGCATTCCTGG
Fragment 16	GACGCGCACACAAGTTCCATGTGC
Fragment 17	GACAGAGCACAACTCTATGTGGGCGTCC
Fragment 18	GTTGAGGTTCAGGCTCTTCTTGGCATCG
Fragment 19	GTTTAGCCACTACTTGTCTGTGGCCTTGTGG
Fragment 20	GCAGGGTTTCCCCACGTAGTCATGG
Fragment 21	GACACGTTCAGCACCATGAAGGCTTTCC
Fragment 22	CATGTCCCCAGGAAGTCTGTGAGGAGACC
Fragment 23	CTGCCAGTGTCCACCACAGCTCTGC
Fragment 24	GTTGATGTGATGGCCAAGTTTCAGCTGC
Fragment 25	GCTGCACCTGAATCACTAACTCAGTGTGAGC
Fragment 26	ATCAGGTTCTTGCTTCAGAGCAGGGAGG
Fragment 27	CGAGGAGGCCAGACCTGCTTTGTCC
Fragment 28	CACATCACATCGTCTGCCTGTCTGTGC

Fragment 29	AGACATGCAGTCCGCTAACGCTGTGG
Fragment 30	CACTCAAGGCAGGTGTTCTGCACCATCC
Fragment 31	GCACTCACTCTGTCTTTCCTGCCTCATGG
Fragment 32	GATTCCCACAGCAAAGGCATCCAAGG
Fragment 33	AGGTAGGGAAACTGAGAGAAGGGAAGCCTATCC
Fragment 34	GCAGGAAACAAGGCCAAAGAGGCACC
Fragment 35	GAACACCCTTGGGGCACACCTGATACTAGG
Fragment 36	GGCTTGAGAGGGTGCAGGGATACATATCG
Fragment 37	GGATCCGTGACCCTCACTTTCCTTGTGC
Fragment 38	ACGTGCACTTTCCCCACAGCACAGC
Fragment 39	TGCCTCCGGTGTGAAGAGGTGATGC
Fragment 40	GTGTGATTTACCTGGTGCCGCTTGTGC
Fragment 41	GAACTGGCTCCATAGACCCAAAGCAAGC
Fragment 42	GCAGAAAGTAGGTAGAGGCCAGGAGGAAATGG
Fragment 43	GGTCTGTCTCATTCACTGCCCTACCCAGG
Fragment 44	AAGCAGCATCCTCAGAGCAGCTGGTCC
Fragment 45	TCCTCACGTGTGCACATCACCTTATAGTCACC
Fragment 46	CACACAAGCCACTGTCACCTGCTGTCG
Fragment 47	CCACCCGCTGCACATGTTTCAGACC
Fragment 48	GACCCTTAACCCTGTGACACTGCACCTATCC
Fragment 49	GGACCACATGGACAGTCACAGGCAGC
Fragment 50	AGGTGACCCTCAGAGGCAGATCATGACC
Fragment 51	AGTGCTGGCCTCTCAATCCCTGACACC
Fragment 52	GGAACTCCCGTTGGAGATGAGGAAGTAAGG
Fragment 53	GCCTTCCAGGAAAGCCAGGAGAGAGG
Fragment 54	CCCTAACCTGATGCACCAGCTGACAGG
Fragment 55	GTTGGCCAAATGAATGAACCAGATTCAGACC
Fragment 56	GTTGTGGTCTCCACATTCTATTCATGTTCGAGG
Fragment 57	CAGCTGACTGCTAGAGCTGTCGTGGAAGC
Fragment 58	CAATCCTGGCTGTCCCAGCTCTCAGG
Fragment 59	CTCACTCAGTGGAGCTTCAGTATCTGCACTTCC
Fragment 60	GGATGAACCCACACATTCCCTTCACTGC
Fragment 61	CTAATTCAGATGGCAATTTGATCACTGCTGTCC

Fragment 62	GACATCACACCATGTTCTGGCTGTAAGAATGG
Fragment 63	AGTGTGGGTGAGCACTGTCCAATCTGAGG
Fragment 64	TCGGAGCTGTGTTGTGAGCCACTAGTAATCC
Fragment 65	ATCTAGGCTCAAGTGGTGGCTGTTGGTGG
Fragment 66	CGTAGGCTTTGAAGATGCTTGTTTCAGAAACG
Fragment 67	AGAAGGGATGATATGCTCGGAATAACTGGAGG
Fragment 68	TTCCACTATGACCCTCAGCGAGTGTTTTCC
Fragment 69	TGAGATGTGCATGGCTGCTGGAATGG
Fragment 70	AGGAAAGGCTTTGAGGCAGGTGGTCC
Fragment 71	CACTCCCTCACTCCATTCATACCTCCACTTCC
Fragment 72	GCTGTAGAGGCCTCCTGGAGGCTTTGC
Fragment 73	CACGCCCAAGGTCTTCAGCTTTGAGG
Fragment 74	CCGAGTTTCTGCACCTGTCAGTGGAGC
Fragment 75	GAGAAGCCTCTAGGGCAGGTGCACAGG
Fragment 76	CTCGACTGTTCCCAAGGGCTCATGG
Fragment 77	CCAGGACTTCATGGCCCAGTGTCTGC
Fragment 78	TCAGAGGGGACAGAGATGAGTCTGATGACG
C14ORF180F1	CAACATAACATGACTGGCTGTGGCACTGG
C14ORF180F2	CCAGCCTAGCAGGAATGGATTCGTTACTCC
C14ORF180F3	GTGTAACTGGAGGCCTCGTGACAGATGG
TMEM179F1	GGTTTGGCAACATGGGTGCAGATGACG
TMEM179F2	GGATTAGTGGTCTCATGGATTAATGGGTTGCC
TMEM179F3	GCCTTTGTAAGCACATGTTGATCAGTCACTGG
PLD4F1	GGAAAGCTTCCTGCATAATCACAGCTTCATTACC
PLD4F2	CCAGAGAGTCACACAGCCTCCAGCTAGTCC
PLD4F3	GCTCTTATCTGCCTCCTGTGGCAAGTGC
AHNAK2F1	ACAGGAAGGAGACGCTGGCACAGAGC
AHNAK2F2	GAGGTGCCACTTAAGGCTCCAAGCAGG
AHNAK2F3	CCTCTGTGTGGTGCCCAAGCTAGATCC
C14ORF79F1	CCTCTGTGTGGTGCCCAAGCTAGATCC
C14ORF79F2	CTGAGACAGTCCTAGATGCTCCCACCTCACC

C14ORF79F3	GTAGGAGGTAACAAGGACCTGAGACTGAGCTGG
CDCA4F1	GCCTTAGGGATCACACCCATTCCTTGG
CDCA4F2	CGAGACCAGCCTGGACAACATAGTGAGACC
CDCA4F3	GCTGGTCTCAAACTCCTGAACTCAAGTGATCC
GPR132F1	CAGGGGACTCTGTTCTTGATCTGCTCTGAGG
GPR132F2	CAACAGTCAAGCTGTCCCAGTAGGGACCAGGAGAGC
GPR132F3	GGCAAGCTGAATCCCTCACCGTAAACC
JAG2F1	ACACCTTCCCAGTAGGGACCAGGAGAGC
JAG2F2	GAACATACTTTCCCTGCAGCGTGCAGC
JAG2F3	GGAAGCAGTGACCCTGACCTGAGATGG
NUDT14F1	GGAAGCAGTGACCCTGACCTGAGATGG
NUDT14F2	GGAAGCTGTCCTGGCAGGAGGAGACC
NUDT14F3	GCTCCCTGCTGAAGCTGTCCTCAGCTGCAGCGTGCAGC

Table S2. Oligonucleotides used in EMSAs.

SNP	allele[a]	Sequence (5' to 3')[b]
rs2498796	com	^{BIO}CACCCACCAGGTCCTGGGAAGCCCCATCTCT
	min	^{BIO}CACCCACCAGGTCCTAGGAAGCCCCATCTCT
rs2498794	com	^{BIO}AGACCTGCCTGAGACAGATCCCAGAGGCCTG
	min	^{BIO}AGACCTGCCTGAGACGGATCCCAGAGGCCTG
rs2494737	com	^{BIO}TTGCCAGCCCAGGACTTGGAGGCTCCAGGGG
	min	^{BIO}TTGCCAGCCCAGGACATGGAGGCTCCAGGGG

[a] com: common allele, min: minor allele
[b] BIO: 5' biotinylation (present on both the sense and antisense strands of the duplex)

Table S3. EMSA competitor duplexes and their target DNA binding proteins.

Competition Target	Sequence (5' to 3')
YY1 consensus FOR	CGCTCCCCGGCCATCTTGGCGGCTGGT
YY1 consensus REV	ACCAGCCGCCAAGATGGCCGGGGAGCG
YY1 mutated FOR (YY1m)	CGCTCCGCGATTATCTTGGCGGCTGGT
YY1 mutated REV (YY1m)	ACCAGCCGCCAAGATAATCGCGGAGCG
NFkB consensus FOR	AGTTGAGGGGACTTTCCCAGGC
NFkB consensus REV	GCCTGGGAAAGTCCCCTCAACT
NFkB mutated FOR (NFkBm)	AGTTGAATTGACTTTGCCAGGC
NFkB mutated REV (NFkBm)	GCCTGGCAAAGTCAATTCAACT
AP2 consensus FOR	GATCGAACTGACCGCCCGCGGCCCGT
AP2 consensus REV	ACGGGCCGCGGGCGGTCAGTTCGATC
CEBP consensus FOR	TGCAGATTGCGCAATCTGCA
CEBP consensus REV	TGCAGATTGCGCAATCTGCA
Negative Control FOR (Neg)	TGCAGAGACTAGTCTCTGCA
Negative Control REV (Neg)	TGCAGAGACTAGTCTCTGCA

Table S4. Oligonucleotides used in cloning luciferase constructs.

Primer	Sequence (5' to 3')
AKT CAN promoter FOR	ACGCGTGTCACTTTACAGACGGGGAAACTGAGG
AKT CAN promoter REV	AGATCTGGAAATGCCCCAAGTACTTAGCAGG
AKT ALT promoter FOR	ACGCGTTCTAGGTGGCTTCAGTGTGAGACC
AKT ALT promoter FOR	AGATCTATGGGGACAGCACACAGTGC
ZBTB42 promoter FOR	ACGCGTAGGGCTGTGATCCAGGCAGG
ZBTB42 promoter REV	AGATCTCCGAGCTCCTCTCCGGTCG
PRE WT FOR	GGATCCCTCAAGAATGATGGCACCTTCATTGG
PRE WT REV	GTCGACGTGAGTGGAGTGTGTAGCCGCTGG

Table S5. Oligonucleotides used for ChIP analyses.

Primer name	Sequence (5' to 3')
SNPrs2494737FOR	AGGACTCAGCCTGGAGACTCC
SNPrs2494737REV	TCTCGGGATTCAGATTTGGG
SNPrs2498796FOR	TTCATCAGCTGGCACTCTGC
SNPrs2498796REV	GTAGAGTGTCTGAGCTGGAACAGG
NegControlFOR	CACAACAGGATCTTATGCGTGG
NegControlREV	CAGTCCCTGCTCATGATCTTGC

Table S6. Numbers of SNPs included in the *AKT1* fine-mapping region on chromosome 14 (bases 104,743,220-105,743,220) compared to SNPs present in the 1000Genomes 2012 reference panel. Linkage disequlibrium was calculated to the top hit previously published for this locus, which was drawn from a subset of risk SNPs selected on the basis of info score >0.9.[3]

SNP category	1000Genomes 2012 release	*AKT1* fine-mapping region	% of 1000G SNPs included in the fine-mapping dataset
SNPs with MAF ≥1% in the 1000Genomes 2012 release[a]	3813	2922	76.6%
LD to rs2498796:			
≥0.8	26	26	100%
0.6-0.799	31	30	96.7%
0.4-0.599	16	16	100%
0.2-0.399	18	17	94.4%
<0.2/NA	3721	2832	76.1%

[a] Minor allele frequencies (MAF) calculated for Europeans only (85 CEU individuals)

Table S7. See excel file.

Table S8. Predicted effects of candidate casual variants on transcription factor binding motifs.

rsID	Position (hg19; chr14)	TFBS[a]	Motif change[b]
rs2494737	105246325	YY1	++
		CEBPA	+
rs2498796	105243220	NF-kappaB	-
		AP2	-
		CREB	-
rs2498794	105245251	BCL	--
		GATA1	-
		AP1	+

[a] Altered transcription factor binding site (TFBS) determined by HaploRegv3[4] or AliBaba2[5] (TRANSFAC and JASPAR matrices).
[b] Degree of change to motif for minor allele: + increased agreement with consensus, - decreased.

Supplementary References

1. GTEx Consortium (2013). The Genotype-Tissue Expression (GTEx) project. Nat Genet 45, 580-585.
2. Cancer Genome Atlas Research, N., Kandoth, C., Schultz, N., Cherniack, A.D., Akbani, R., Liu, Y., Shen, H., Robertson, A.G., Pashtan, I., Shen, R., et al. (2013). Integrated genomic characterization of endometrial carcinoma. Nature 497, 67-73.
3. Cheng, T., Thompson, D.J., O'Mara, T.A., Painter, J.N., Glubb, D.M., Flach, S., Lewis, A., French, J.D., Freeman-Mills, L., Church, D., et al. (2016). Five endometrial cancer risk loci identified through genome-wide association analysis. Nat Genet (under re-review).
4. Ward, L.D. and Kellis, M. (2015). HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. Nucleic Acids Res (doi: 10.1093/nar/gkv1340).
5. Grabe, N. (2002). AliBaba2: context specific identification of transcription factor binding sites. In Silico Biol 2, S1-15.