Supplemental Information

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 HT Cheng, John Pearson, Stephen Kazakoff, Nicola Waddell, Erling Høivik, Ellen L Goode, Rodney J Scott, Ian Tomlinson, Alison M Dunning, Douglas F Easton, Juliet D French, Helga B Salvesen, Pamela Pollock, Deborah J Thompson, Amanda B Spurdle, Stacey L Edwards

Supplemental Acknowledgements

Manuscript writing group: JNP, PP, DJT, ABS, SLE. Imputation: THTC, DJT. Statistical analyses and programming: JNP, THTC, DJT. Functional analysis and bioinformatics: JNP, SK, TO'M, KMH, HS, JDF, SLE. All other authors read and approved the manuscript. The QIMR Berghofer groups were supported by a Weekend to End Women's Cancer Walk Grant and NHMRC project grants 1058415 to SLE and 1031333 to ABS. ABS is supported by an NHMRC Senior Research Fellowship (1061779). DFE is a Principal Research Fellow of CR-UK. The funders have no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

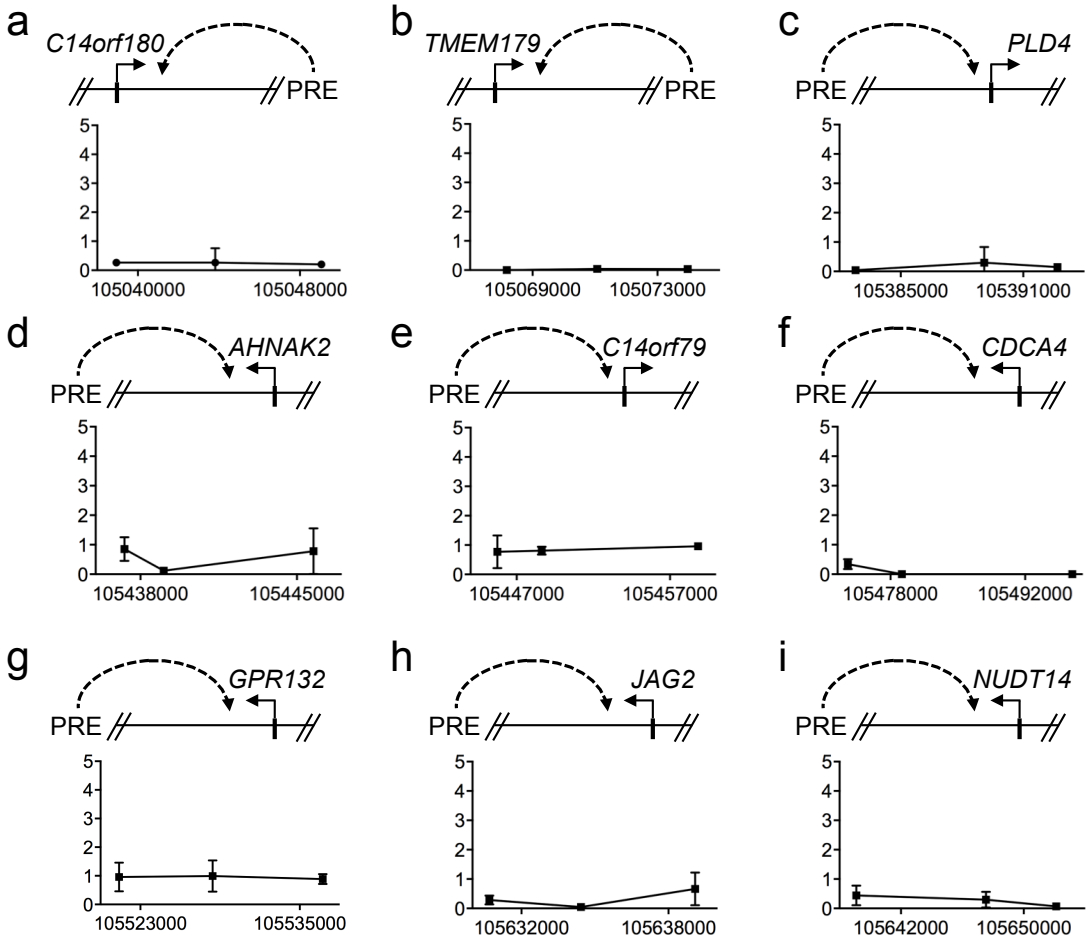

Figure S1. **Chromatin interactions at 14q32 in Ishikawa endometrial cancer cell lines**. 3C interaction profiles between the putative regulatory element (PRE; containing rs2498796, rs2498794 and rs2494737) and **(a)** *C14ORF180,* **(b)** *TMEM179,* **(c)** *PLD4,* **(d)** *AHNAK2,* **(e)** *C14ORF79,* **(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.

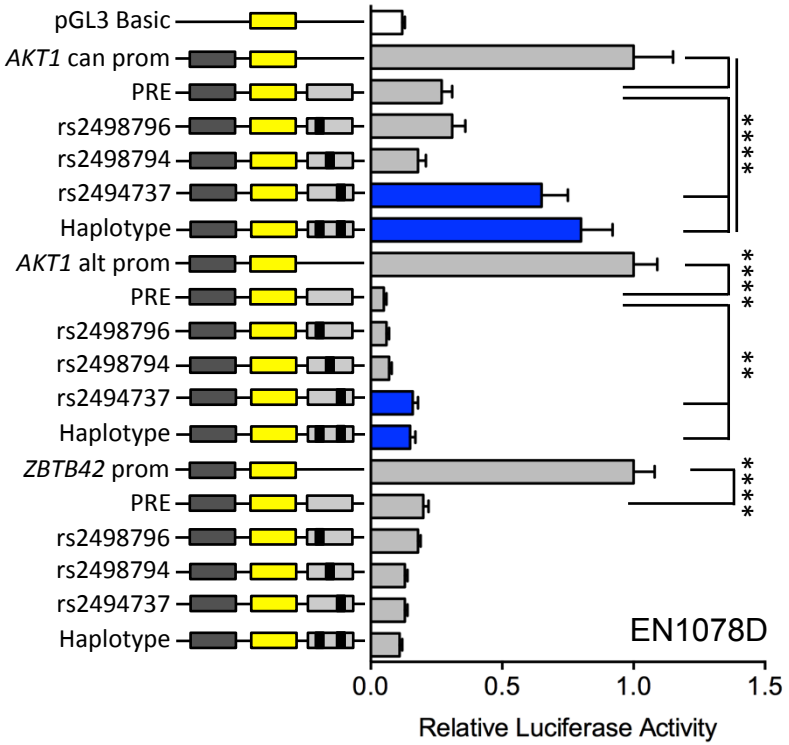

Figure S2. Luciferase reporter assays in EN-1078D endometrial cancer cells. The putative regulatory element (PRE) containing the major SNP alleles were cloned downstream of target gene promoter-driven luciferase constructs. *AKT1* can prom and *AKT1* alt prom denote a canonical and alternative *AKT1* 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).

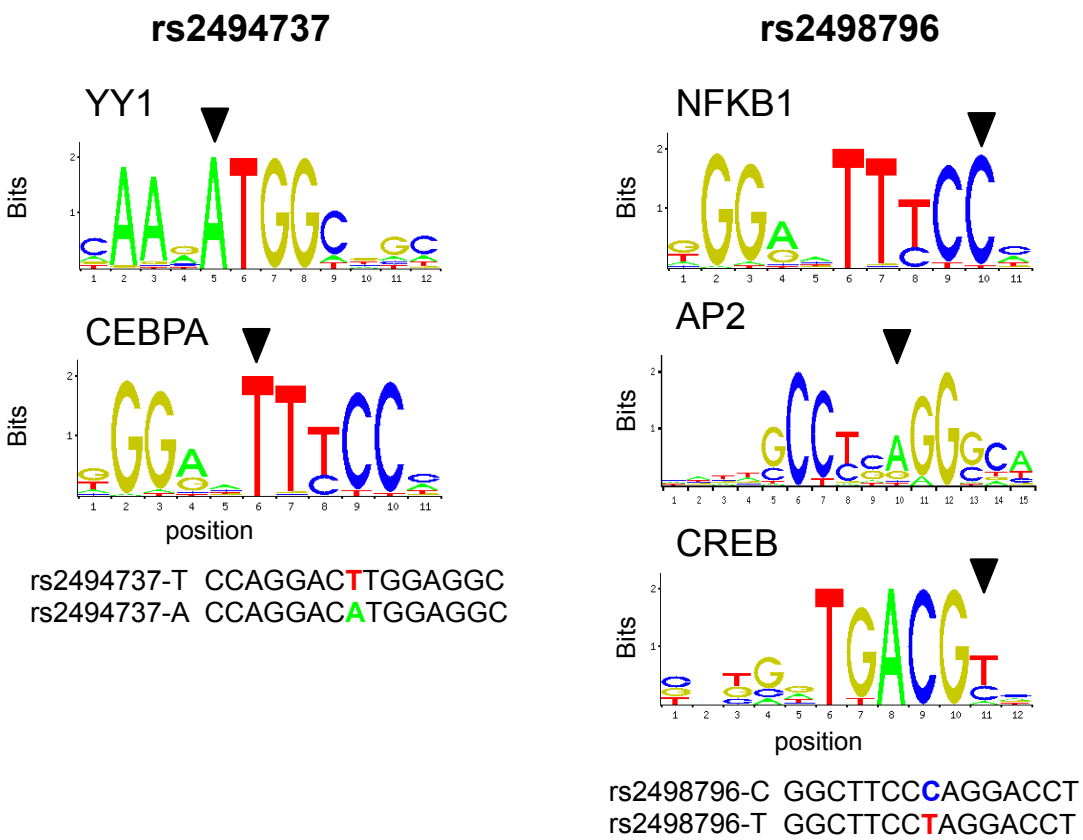

Figure S3. **Transcription factor binding.** Position weight matrix (PWM) of YY1, CEBPA, NFKB1, AP2 and CREB from JASPAR, with homology to the risk-associated alleles of rs2494737 and rs2498796 colored below. Predicted SNP changes are indicated by black arrowheads.

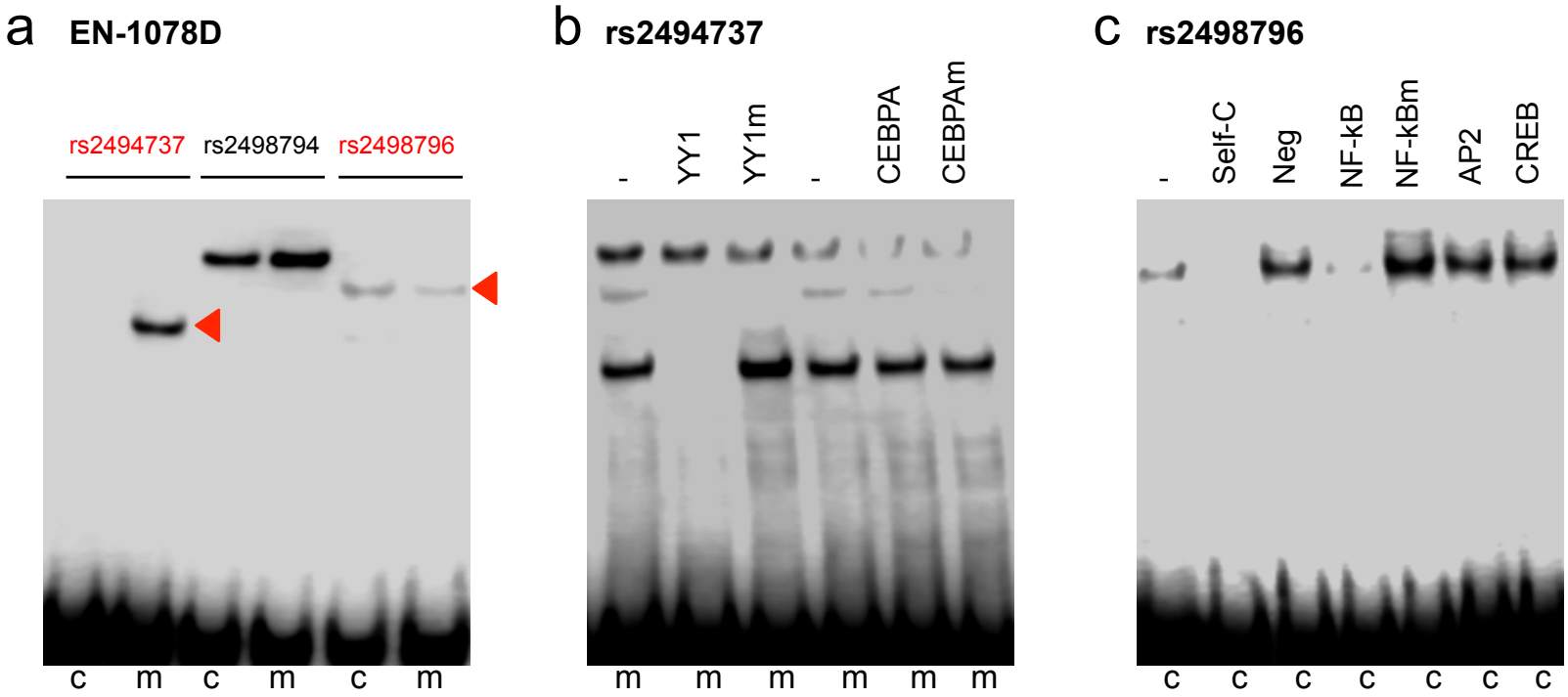

Figure S4. EMSAs for candidate causal SNPs to detect allele-specific binding of nuclear proteins. (a) Oligonucleotides were incubated with EN-1078D nuclear extracts. Red arrowheads show bands of different mobility or intensity detected between the common (c) and minor (m) alleles for the three candidate 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 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.

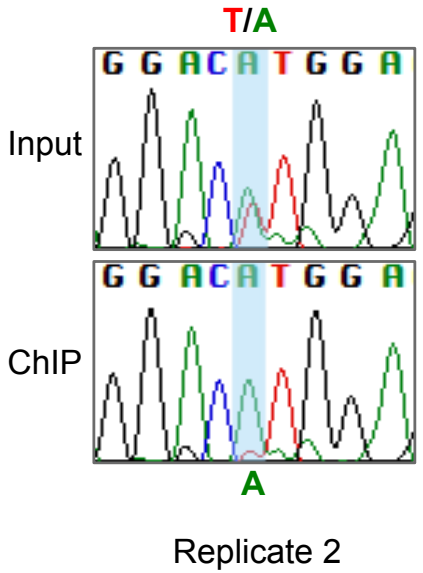

Figure S5. YY1 transcription factor binding *in vivo*. Sanger sequencing of the PCR fragment generated using primers flanking SNP rs2494737 in heterozygous Ishikawa endometrial cancer cells following YY1 ChIP-qPCR and the input DNA controls. Primers are listed in **Table S5.**

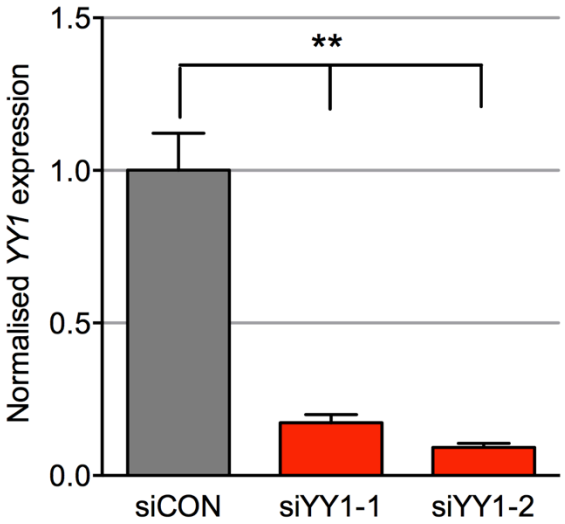

Figure S6. TaqMan real-time PCR assays confirming knockdown of *YY1* in Ishikawa cells. siCON is a nontargeting negative control and siYY1-1 and siYY1-2 are two independent siRNAs targeting *YY1*. Error bars denote the standard error of the mean from three experiments performed in duplicate. Statistical significance was determined by a paired t-test (**P<0.01)

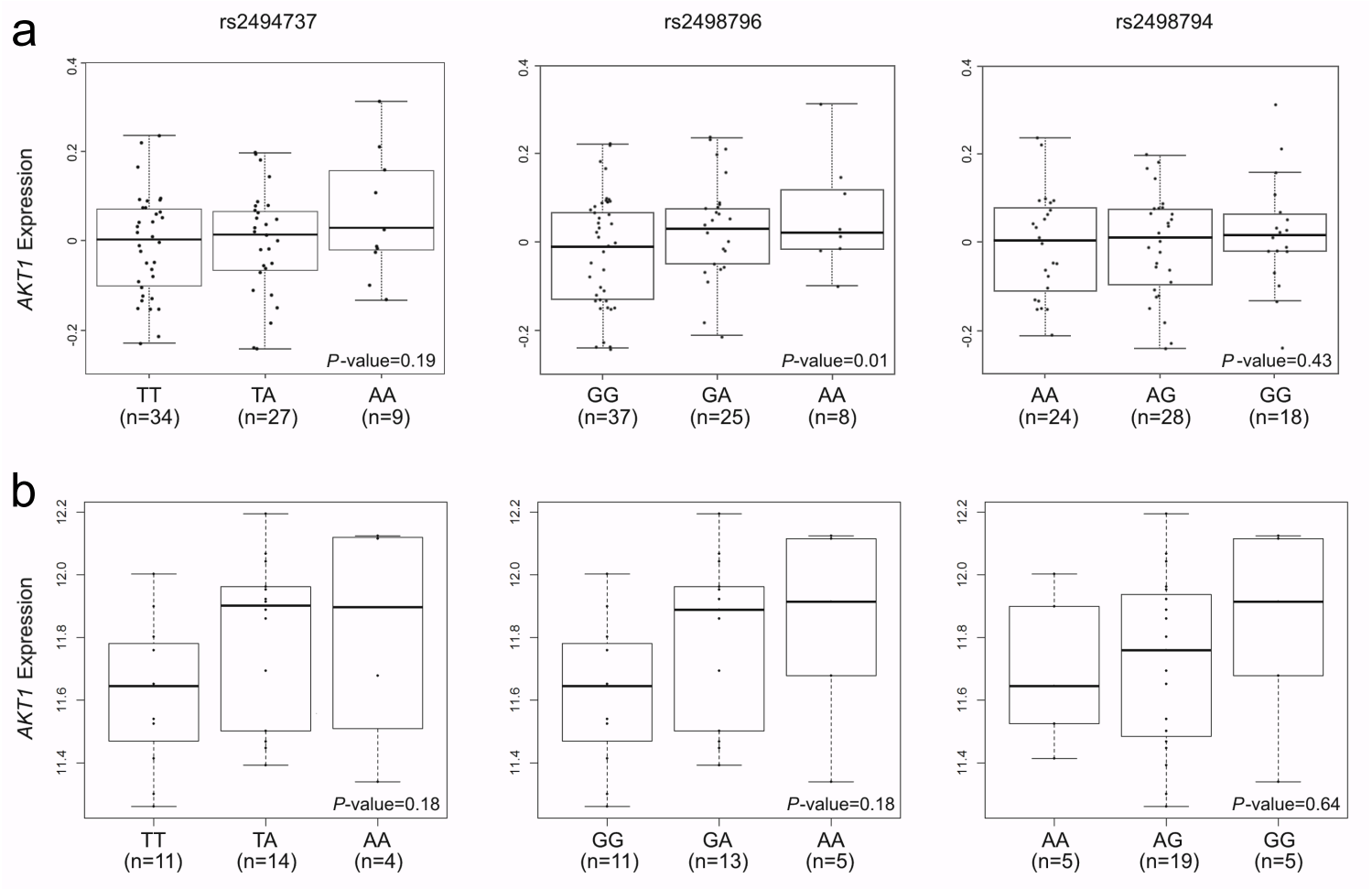

Figure S7. Associations of candidate causal SNPs with overall expression of *AKT1* in uterine samples from the **(A)** GTEx[1] database and **(B)** TCGA[2] dataset. The x-axis of each plot corresponds to the three observed SNP genotypes and the y-axis represents either log2-normalized gene expression 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.

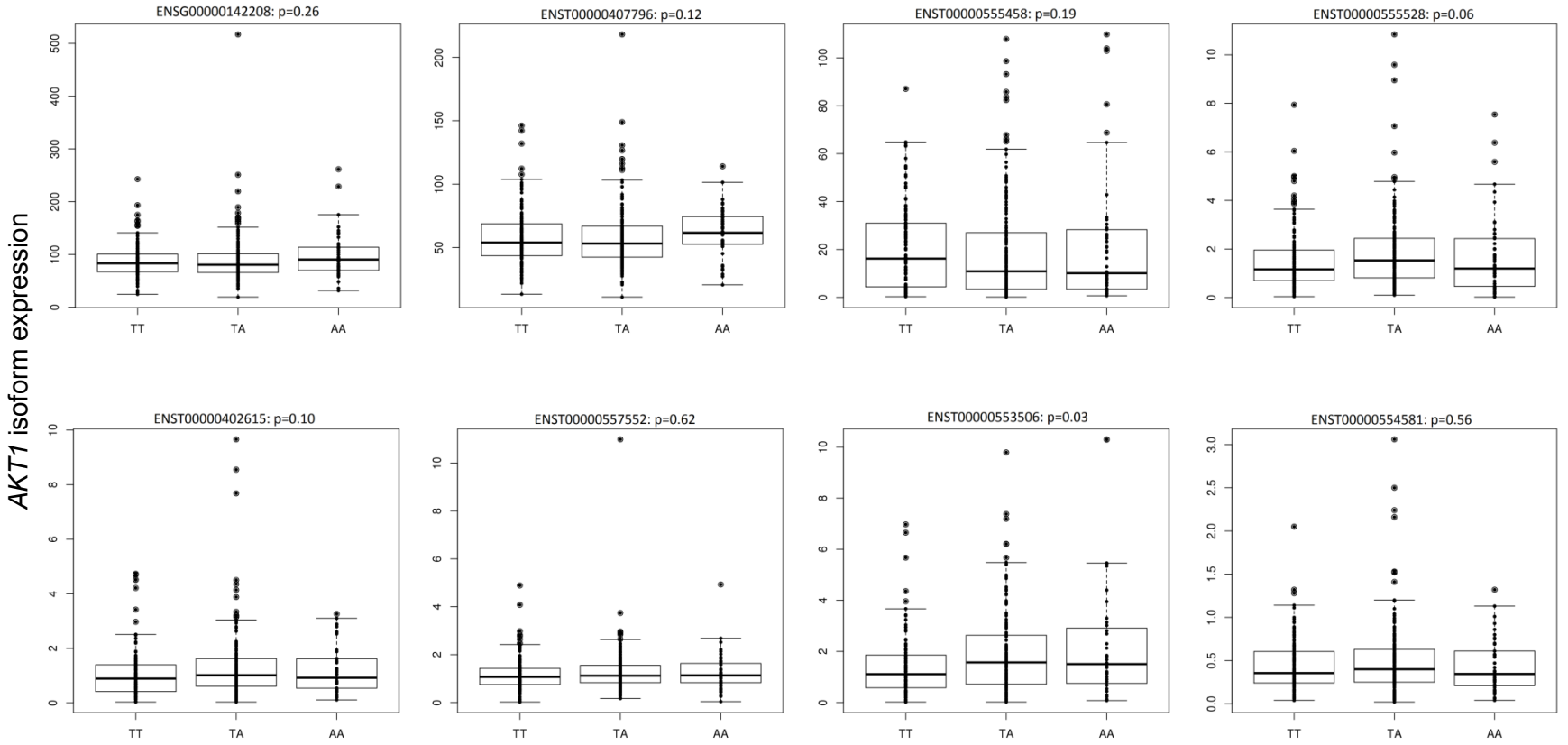

Figure S8. Associations of SNP rs2494737 with expression of *AKT1* isoforms in endometrial tumour samples from the TCGA dataset.[2] The *x*-axis of each plot corresponds to the three SNP genotypes and the *y*-axis represents the RSEM gene expression values for each isoform in 526 unique samples. Data 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
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	CAACAGTCAAGCTGTCCAAGTAGGACCAGGAGAGC
GPR132F3	GGCAAGCTGAATCCCTCACCGTAAACC
JAG2F1	ACACCTTCCCAGTAGGGACCAGGAGAGC
JAG2F2	GAACATACTTTCCCTGCAGCGTGCAGC
JAG2F3	GGAAGCAGTGACCCTGACCTGAGATGG
NUDT14F1	GGAAGCAGTGACCCTGACCTGAGATGG
NUDT14F2	GGAAGCTGTCCTGGCAGGAGGAGACC
NUDT14F3	GCTCCCTGCTGAAGTAGTCCTTACCTGTCACGCGTGCAGC

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	++
		CEPBA	+
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.