Article

Genome-wide association study identifies a common variant associated with risk of endometrial cancer

MONTGOMERY, Stephen, DERMITZAKIS, Emmanouil

Abstract

Endometrial cancer is the most common malignancy of the female genital tract in developed countries. To identify genetic variants associated with endometrial cancer risk, we performed a genome-wide association study involving 1,265 individuals with endometrial cancer (cases) from Australia and the UK and 5,190 controls from the Wellcome Trust Case Control Consortium. We compared genotype frequencies in cases and controls for 519,655 SNPs. Forty seven SNPs that showed evidence of association with endometrial cancer in stage 1 were genotyped in 3,957 additional cases and 6,886 controls. We identified an endometrial cancer susceptibility locus close to HNF1B at 17q12 (rs4430796, P = 7.1 × 10(-10)) that is also associated with risk of prostate cancer and is inversely associated with risk of type 2 diabetes.

Reference

MONTGOMERY, Stephen, DERMITZAKIS, Emmanouil. Genome-wide association study identifies a common variant associated with risk of endometrial cancer. *Nature genetics*, 2011, vol. 43, no. 5, p. 451-4

PMID : 21499250
DOI : 10.1038/ng.812

nature genetics

Genome-wide association study identifies a common variant associated with risk of endometrial cancer

Amanda B Spurdle[1], Deborah J Thompson[2], Shahana Ahmed[3,29], Kaltin Ferguson[1,29], Catherine S Healey[3,29], Tracy O'Mara[1,4,29], Logan C Walker[1], Stephen B Montgomery[5], Emmanouil T Dermitzakis[5], The Australian National Endometrial Cancer Study Group[1,6], Paul Fahey[1], Grant W Montgomery[1], Penelope M Webb[1], Peter A Fasching[7], Matthias W Beckmann[8], Arif B Ekici[8], Alexander Hein[9,10], Diether Lambrechts[11,12], Lieve Coenegrachts[13], Ignace Vergote[13], Frederic Amant[13], Helga B Salvesen[14,15], Jone Trovik[14,15], Tormund S Njolstad[14,15], Harald Helland[14], Rodney J Scott[16–19], Katie Ashton[16,17], Tony Proietto[20], Geoffrey Otton[20], National Study of Endometrial Cancer Genetics Group[6,21], Ian Tomlinson[21], Maggie Gorman[21], Kimberley Howarth[21], Shirley Hodgson[22], Montserrat Garcia-Closas[23,24], Nicolas Wentzensen[23], Hannah Yang[23], Stephen Chanock[23], Per Hall[25], Kamila Czene[25], Jianjun Liu[26], Jingmei Li[25,26], Xiao-Ou Shu[27], Wei Zheng[27], Jirong Long[27], Yong-Bing Xiang[28], Mitul Shah[3], Jonathan Morrison[2], Kyriaki Michailidou[2], Paul D Pharoah[2,3], Alison M Dunning[3] & Douglas F Easton[2,3]

Endometrial cancer is the most common malignancy of the female genital tract in developed countries. To identify genetic variants associated with endometrial cancer risk, we performed a genome-wide association study involving 1,265 individuals with endometrial cancer (cases) from Australia and the UK and 5,190 controls from the Wellcome Trust Case Control Consortium. We compared genotype frequencies in cases and controls for 519,655 SNPs. Forty seven SNPs that showed evidence of association with endometrial cancer in stage 1 were genotyped in 3,957 additional cases and 6,886 controls. We identified an endometrial cancer susceptibility locus close to *HNF1B* at 17q12 (rs4430796, $P = 7.1 \times 10^{-10}$) that is also associated with risk of prostate cancer and is inversely associated with risk of type 2 diabetes.

Cancer of the uterine corpus, or endometrial cancer, is the most common invasive gynecological cancer in developed countries, with more than 280,000 cases annually worldwide[1]. The prognosis is considered favorable for the most common histological subtype, endometrioid endometrial cancer, which represents 80–90% of all endometrial cancers. Nevertheless, the disease is associated with substantial morbidity due to surgery and radiotherapy[2], and treatment is often complicated because most affected individuals present at older ages and with substantial co-morbidities.

Genome-wide association studies (GWAS) have successfully identified common genetic variants associated with modestly increased risks for numerous complex diseases, with more than 150 published loci at $P < 10^{-7}$ for common cancers[3]. These findings provide evidence that common variants are responsible, at least in part, for the increased familial risk of cancer[4,5]. A family history of endometrial

[1]Division of Genetics and Population Health, Queensland Institute of Medical Research, Brisbane, Queensland, Australia. [2]Department of Public Health and Primary Care, University of Cambridge, Strangeways Research Laboratory, Cambridge, UK. [3]Department of Oncology, University of Cambridge, Strangeways Research Laboratory, Cambridge, UK. [4]Hormone Dependent Cancer Program, Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, Queensland, Australia. [5]Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva, Switzerland. [6]A full list of members is provided in the **Supplementary Note**. [7]University of California at Los Angeles, David Geffen School of Medicine, Department of Medicine, Division of Hematology and Oncology, Los Angeles, California, USA. [8]Institute of Human Genetics, Friedrich Alexander University Erlangen Nürnberg, Erlangen, Germany. [9]University Hospital Erlangen, Department of Gynecology and Obstetrics, Friedrich Alexander University Erlangen Nürnberg, Erlangen, Germany. [10]University Gynecologic Oncology Center, Comprehensive Cancer Centre, Erlangen Nürnberg, Erlangen, Germany. [11]Vesalius Research Center, Vlaams Instituut voor Biotechnologie, Leuven, Belgium. [12]Vesalius Research Center, University of Leuven, Leuven, Belgium. [13]Division of Gynecologic Oncology, Department of Obstetrics and Gynecology, University Hospitals Leuven, Leuven, Belgium. [14]Department of Obstetrics and Gynecology, Haukeland University Hospital, Bergen, Norway. [15]Department of Clinical Medicine, The University of Bergen, Bergen, Norway. [16]The Centre for Information Based Medicine and the Discipline of Medical Genetics, School of Biomedical Sciences and Pharmacy, Faculty of Health, University of Newcastle, New South Wales (NSW), Australia. [17]The Hunter Medical Research Institute, John Hunter Hospital, NSW, Australia. [18]Division of Genetics, Hunter Area Pathology Service, John Hunter Hospital, Newcastle, NSW, Australia. [19]Hunter Centre for Gynaecological, John Hunter Hospital, NSW, Australia. [20]School of Medicine and Public Health, Faculty of Health, University of Newcastle, NSW, Australia. [21]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK. [22]Department of Clinical Genetics, St. George's Hospital Medical School, London, UK. [23]Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, Maryland, USA. [24]Institute of Cancer Research, London, UK. [25]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. [26]Human Genetics, Genome Institute of Singapore, Singapore. [27]Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University Medical Center, Nashville, Tennessee, USA. [28]Department of Epidemiology, Shanghai Cancer Institute, Shanghai, China. [29]These authors contributed equally to this work. Correspondence should be addressed to A.B.S. (amanda.spurdle@qimr.edu.au).

Received 1 December 2010; accepted 24 March 2011; published online 17 April 2011; doi:10.1038/ng.812

Table 1 Association between genotype and endometrial cancer for three SNPs in *HNF1B*

Marker (alleles)	Chromosome (position)	Stage	Cases (*n*)	Controls (*n*)	Case MAF	Control MAF	Stage per-allele OR (95% CI)	Stage *P*	Combined per-allele OR (95% CI)	Combined *P*
rs4430796 (A/G)	17 (33,172,153)	Stage 1	1,262	5,179	0.43	0.48	0.79 (0.73–0.87)	3.06×10^{-7}		
		Stage 2 (all cases)	2,332	4,349	0.44	0.48	0.87 (0.81–0.94)	2.00×10^{-4}	0.84 (0.79–0.89)	7.11×10^{-10}
		Stage 2 (endometrioid cases)	1,786	4,349	0.43	0.48	0.84 (0.77–0.91)	2.56×10^{-5}	0.82 (0.77–0.87)	4.28×10^{-11}
		Stage 2 (non-endometrioid cases)	484	4,349	0.46	0.48	0.94 (0.82–1.08)	3.74×10^{-1}		
rs4239217 (A/G)	17 (33,173,100)	Stage 1	1,265	5,190	0.35	0.41	0.79 (0.72–0.86)	2.48×10^{-7}		
		Stage 2 (all cases)	2,342	4,307	0.37	0.40	0.89 (0.82–0.96)	2.00×10^{-3}	0.84 (0.80–0.90)	1.19×10^{-8}
		Stage 2 (endometrioid cases)	1,753	4,307	0.36	0.40	0.86 (0.79–0.94)	5.90×10^{-4}	0.83 (0.78–0.88)	1.27×10^{-9}
		Stage 2 (non-endometrioid cases)	588	4,307	0.39	0.40	0.97 (0.85–1.10)	6.18×10^{-1}		
rs7501939 (G/A)	17 (33,175,269)	Stage 1	1,263	5187	0.36	0.40	0.80 (0.73–0.88)	2.17×10^{-6}		
		Stage 2 (all cases)	3,337	5,234	0.37	0.40	0.90 (0.84–0.96)	1.00×10^{-3}	0.86 (0.82–0.91)	5.35×10^{-8}
		Stage 2 (endometrioid cases)	2,562	5,234	0.36	0.40	0.88 (0.82–0.94)	3.30×10^{-4}	0.85 (0.80–0.90)	7.57×10^{-9}
		Stage 2 (non-endometrioid cases)	690	5,234	0.38	0.40	0.95 (0.85–1.07)	4.29×10^{-1}		

MAF, minor allele frequency.

cancer is associated with increased risk of the disease[6–8]. Although germline mutations in mismatch repair genes confer a substantial risk of endometrial cancer in the context of Lynch syndrome, these are rare[9]. Candidate gene association studies have implicated *CYP19A1* as a common endometrial cancer susceptibility locus with modest effect[10,11], but no GWAS of endometrial cancer have been published to date.

We have conducted a GWAS using cases with endometrial cancer from Australia and the UK. To reduce the potential effects of disease heterogeneity, we selected cases with endometrioid histology for genotyping using the Human 610K array on the Illumina Infinium platform. We extracted control data for SNPs included on the 610K platform from existing Illumina 1.2M genome-wide scan data for controls of European-ancestry from two UK population-based studies genotyped by the Wellcome Trust Case Control Consortium[12]. After applying standard quality control measures (Online Methods), we analyzed data for 519,655 SNPs from 1,265 cases and 5,190 controls. We compared genotype frequencies between cases and controls using a 1-degree-of-freedom Cochran-Armitage trend test. The test statistic inflation factor (λ) was 1.04 after adjustment for population stratification using the principal components approach[13] (**Supplementary Fig. 1**). Following a review of cluster plots to eliminate likely artifactual associations, 130 SNPs were significant at $P < 10^{-4}$ compared to the ~52 SNPs expected by chance. Next, we eliminated redundant SNPs (Online Methods). In total, we selected 49 SNPs for follow up in stage 2, 47 of which passed genotype quality control (**Supplementary Table 1**). We collated data from up to 3,957 endometrial cancer cases and 6,886 controls of European ancestry from ten centers, with genotypes generated specifically for this study or data derived from existing GWAS scans (**Supplementary Table 2**).

After combining results from stage 1 and 2 (**Supplementary Table 1**), three SNPs were significant at the $P < 10^{-7}$ level. All three lie in a region of linkage disequilibrium (pairwise r^2 = 0.68–0.90 in Europeans) encompassing the first four exons of *HNF1B* at 17q12 (**Table 1**). The most significant association was for rs4430796 (odds ratio (OR) per G allele = 0.84, 95% CI 0.79–0.89, $P = 7.1 \times 10^{-10}$). There was no significant heterogeneity in the per-allele OR between stages 1 and 2 (stage 1 OR = 0.79, 95% CI 0.73–0.87; stage 2 OR = 0.87, 95% CI 0.81–0.94; heterogeneity $P = 0.11$) or among the five stage 2 studies in which it was genotyped (heterogeneity $P = 0.75$) (**Fig. 1**). There was no significant deviation from the multiplicative, per-allele model ($P = 0.74$). Restricting stage 2 of the analysis to cancers with an endometrioid histology (1,786 of 2,332 cases genotyped for this SNP) slightly strengthened the effect (stage 2 OR = 0.84, 95% CI 0.77–0.91, overall $P = 4.3 \times 10^{-11}$; **Table 1**). The three *HNF1B* SNPs were not significantly associated with non-endometrioid disease (**Table 1**), although the numbers of cases with this type of disease were small. In addition, we genotyped 832 cases and 2,049 controls of Chinese ancestry from Shanghai (**Supplementary Table 2**) for rs11651755, a surrogate for rs4430796 (HapMap $r^2 = 1.0$ in the CHB population). The estimated odds ratio for rs11651755 in the Chinese samples was 0.96 (95% CI 0.84–1.09, $P = 0.55$).

The only SNP outside the *HNF1B* region to reach $P < 10^{-5}$ was rs673604 on 1p34, located 29 kb centromeric to *SFPQ* (combined OR = 1.21, 95% CI 1.12–1.32, $P = 5.9 \times 10^{-6}$). Although highly significant in stage 1 ($P = 6.1 \times 10^{-7}$), this SNP showed only weak evidence for association in stage 2 ($P = 0.041$). This stage 2 result was similar when restricted to cases with endometrioid histology ($P = 0.049$, combined OR = 1.23, 95% CI 1.12–1.34, $P = 4.6 \times 10^{-6}$) or non-endometrioid histology (stage 2 OR = 1.07, 95% CI 0.88–1.30, $P = 0.50$) (**Supplementary Tables 3** and **4**). Further large studies will be required to determine whether this association is genuine. Results for the remaining SNPs investigated in stage 2 were also little different when analyses were restricted to cases of endometrioid or non-endometrioid subtype; three SNPs showed significant differences in frequencies between cases with and without endometrioid histology ($P < 0.05$), which is in line with what would be expected by chance (**Supplementary Tables 3** and **4**).

Multiple studies have independently reported the G allele of rs4430796, associated with decreased risk of endometrial cancer in this study, to be associated with a decreased risk of prostate cancer[14–16]. A recent meta-analysis estimated an OR per G allele of 0.79 (95% CI 0.76–0.83) for prostate cancer[17] but found no association of the same allele with breast, lung, colorectal or pancreatic cancers or

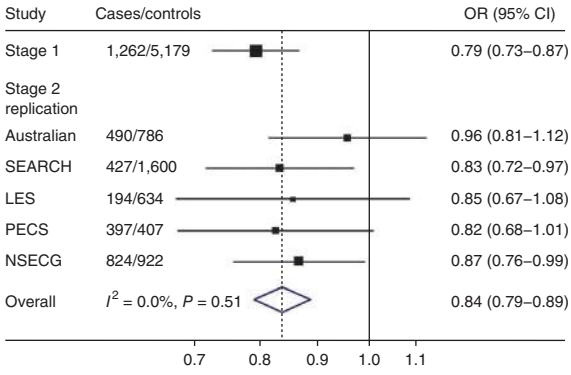

Figure 1 Forest plot showing the association between rs4430796 and endometrial cancer for each component of the study. Study abbreviations are as given in **Supplementary Table 2**. The SEARCH and Australian stage 2 replication sets excluded samples included in stage 1. The solid line denotes the null, and the dashed line indicates the overall odds ratio (OR) estimate. I^2 measures the heterogeneity in effect size between studies[36].

melanoma[17]. The same SNP allele has also been identified by GWAS to be associated with an increased risk of type 2 diabetes (per G allele OR = 1.10 (95% CI 1.06–1.15) (ref. 15) and OR = 1.14 (95% CI 1.08–1.20) (ref. 18)). There is evidence of an inverse correlation between type 2 diabetes and prostate cancer risk[19], but the association between *HNF1B* SNPs and prostate cancer does not appear to be mediated by history of diabetes[20]. Increased body mass index (BMI) is a major risk factor for both type 2 diabetes and endometrial cancer, and there is a positive correlation between type 2 diabetes and endometrial cancer risk[21–24]. However, the opposite direction of the effects of rs4430796 on endometrial cancer and type 2 diabetes risk indicates that the association between rs4430796 and endometrial cancer risk is not mediated through BMI or type 2 diabetes. In addition, in the subset of endometrial case-control studies where BMI was recorded, adjusting for BMI did not materially alter the risk estimate (OR = 0.87 (95% CI 0.78–0.97) compared to OR = 0.88 (95% CI 0.79–0.99); n = 3,055).

To provide a more comprehensive analysis of SNPs in the *HNF1B* region, we identified variants in the region using re-sequencing data in individuals of European ancestry from the 1000 Genomes Project and performed association analyses for all SNPs using genotypes imputed from the stage 1 data. We identified 20 SNPs associated with endometrial cancer at $P < 10^{-5}$, the most significant being rs11651755 (**Supplementary Table 5**). All these SNPs are strongly correlated with rs4430796 ($r^2 > 0.45$). It is plausible therefore that one or more of these variants is functionally associated with endometrial cancer risk.

The common haplotypes formed by the 33 SNPs from the 1000 Genomes Project are well tagged by the six SNPs genotyped in stage 1 of our study (rs757210, rs4430796, rs4239217, rs7501939, rs3760511 and rs1762642). For example, the rare allele of rs11651755 appears to always occur on the same haplotype as the rare allele of rs4430796. We found no evidence of specific haplotype effects at this locus; the haplotypes carrying the common A allele at rs4430796 were all more frequent in cases than controls, and the reverse was true for three of four haplotypes carrying the protective G allele (**Supplementary Table 6**).

HNF1B (also known as *TCF2*, *LFB3 MODY5* and *VHNF1*) encodes a member of the homeodomain-containing superfamily of transcription factors. The gene encodes three isoforms, with isoforms A and B considered to act as transcriptional activators and isoform C as a transcriptional repressor[25]. Abrogating mutations in *HNF1B* result in diabetes phenotypes, including maturity-onset diabetes of the young subtype 5 (MODY5) as well as renal cysts. Of relevance to endometrial cancer, microdeletions encompassing *HNF1B* have been reported in Mayer-Rokitansky-Kuster-Hauser syndrome, which is characterized by congenital aplasia of the uterus and upper vagina due to anomalous development of the Mullerian ducts[26], and *HNF1B* mutations or deletions are associated with uterine abnormalities caused by incomplete Mullerian duct fusion and Mullerian duct aplasia[27]. Human embryonic gene expression studies have shown that *HNF1B* expression occurs during early development of the human urogenital tract, with expression maintained in Wolffian duct derivatives but not in Mullerian duct deriviatives[28]. In contrast, *HNF1B* overexpression has been reported to be a biomarker of clear-cell carcinoma of the pancreas[29] and of clear-cell carcinoma of the ovary and its probable precursor ovarian endometriosis[30–33]. There is also evidence to suggest that *HNF1B* isoform usage may be altered in prostate cancer tissue, with upregulated *HNF1B* isoform B expression in prostate cancer tissue as compared to benign tissue[34]. Analysis of several lymphocyte-derived gene expression datasets (**Supplementary Fig. 2**) identified significant associations between rs4430796 genotype and *HNF1B* expression in individuals of European ancestry but not for individuals of African ancestry. These observations suggest that *HNF1B* may underlie the observed association with endometrial cancer risk, but that rs4430796 is unlikely to be the causal SNP driving the association.

GWAS have so far identified 29 prostate cancer loci in addition to *HNF1B* (ref. 35), but none of the other loci showed any evidence of association with endometrial cancer in this study (**Supplementary Table 7**). Further common low penetrance endometrial cancer loci are likely to be identifiable through larger collaborative GWAS and follow-up studies. The independent discovery of a common risk allele for both prostate cancer and endometrial cancer indicates some shared etiology between these two diseases that had not previously been recognized and also highlights the value of the agnostic GWAS approach for identifying previously unexplored biological pathways and new molecular targets for prevention.

URLs. Australian National Endometrial Cancer Study (ANECS), http://www.anecs.org.au/index.html; Gene Expression Variation (Genevar), http://www.sanger.ac.uk/resources/software/genevar; HapMap, http://hapmap.ncbi.nlm.nih.gov/; The Studies of Epidemiology and Risk factors in Cancer Heredity (SEARCH), http://www.srl.cam.ac.uk/search/Homepage.htm; Wellcome Trust Case Control Consortium (WTCCC), http://www.wtccc.org.uk/; R, http://www.r-project.org/; Stata, http://www.stata.com/.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturegenetics/.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

This work was supported by the National Health and Medical Research Council (NHMRC; ID 552402), the Wellcome Trust and by Cancer Research UK grants C1287/A10118, C490/A1021, C8197/A10865 and C8197/A10123. A.B.S. and P.M.W. are NHMRC Senior Research Fellows, and G.M. is an NHMRC Senior Principle Research Fellow. T.O. is supported by an Australian Postgraduate Award, an Institute of Health and Biomedical Innovation PhD Top-Up and a Smart State PhD Award. L.C.W. is a John Gavin Postdoctoral Fellow (Genesis Oncology Trust, New Zealand). D.F.E. is a Principal Research Fellow of Cancer Research UK. A.M.D. is supported by the Joseph Mitchell Trust. I.T. is supported by Cancer Research UK and the Oxford Comprehensive Biomedical Research Centre. P.A.F. was partly funded by the Dr. Mildred Scheel Stiftung of the Deutsche Krebshilfe (German Cancer Aid).

This study makes use of data generated by the Wellcome Trust Case Control Consortium (WTCCC) 2. A full list of the investigators who contributed to the generation of the data is available from the WTCCC website. We acknowledge use of DNA from the British 1958 Birth Cohort collection, funded by the Medical Research Council grant G0000934 and the Wellcome Trust grant 068545/Z/02. Funding for this project was provided by the Wellcome Trust under award 085475.

We thank the study participants and collaborators and the research teams involved in the design and implementation of the individual studies included (see **Supplementary Note** for full list of collaborators and specific acknowledgments). Australian National Endometrial Cancer Study (ANECS) recruitment was supported by project grants from the National Health and Medical Research Council of Australia (ID#339435), The Cancer Council Queensland (ID#4196615) and Cancer Council Tasmania (ID#403031 and ID#457636). The Bavarian Endometrial Cancer Study (BECS) was partly funded by the Erlanger Leistungsbezogene Anschubfinanzierung und Nachwuchsförderung Fond (ELAN fund) of the University of Erlangen. The Leuven Endometrium Study (LES) was supported by the Verelst Foundation for endometrial cancer. Molecular Markers in Treatment of Endometrial Cancer (MoMaTEC) received financial support from a Helse Vest Grant, the University of Bergen, Melzer Foundation, the Norwegian Cancer Society (Harald Andersens legat), the Research Council of Norway and Haukeland University Hospital. The Newcastle Endometrial Cancer Study (NECS) acknowledges contributions from the University of Newcastle, the National Broadcasting Network (NBN) Children's Cancer Research Group, Jennie Thomas and the Hunter Medical Research Institute. The National Study of Endometrial Cancer Genetics (NSECG) Group was supported principally by Cancer Research UK and by funds from the Oxford Comprehensive Biomedical Research Centre, with core infrastructure support to the Wellcome Trust Centre for Human Genetics, Oxford provided by grant 075491/Z/04. The Polish Endometrial Cancer Study (PECS) was funded by the intramural research program at the US National Cancer Institute, Division of Cancer Epidemiology and Genetics in the Hormonal and Reproductive Epidemiology Branch. The Singapore and Swedish Breast/Endometrial Cancer Study (SASBAC) was supported by funding from the Agency for Science, Technology and Research of Singapore (A*STAR), the US National Institutes of Health (NIH) and the Susan G. Komen Breast Cancer Foundation. The Shanghai Endometrial Cancer Genetic Study (SECGS) was supported by grants from the National Cancer Institute of the United States Public Health Service (RO1 CA 092585 and R01 CA90899, R01 CA64277). SEARCH is funded by a program grant from Cancer Research UK (C490/A10124).

AUTHOR CONTRIBUTIONS

A.B.S., D.F.E., G.M. and P.M.W. obtained funding for the study. A.B.S. and D.F.E. designed the study. A.B.S. and D.J.T. drafted the manuscript. P.F. and K.M. conducted preliminary analysis, and D.F.E. and D.J.T. conducted the final statistical analyses. A.B.S. and P.M.W. coordinated the ANECS. P.D.P. and D.F.E. coordinated Studies of Epidemiology and Risk Factors in Cancer Heredity (SEARCH). A.B.S., K.F. and T.O. coordinated the ANECS stage 1 genotyping. A.M.D., S.A. and C.S.H. coordinated the SEARCH stage 1 genotyping. L.C.W., S.B.M. and E.T.D. conducted analyses to assess correlations between genotype and gene expression. J.M. provided data management and bioinformatics support. T.O. and K.F. coordinated the ANECS and other Brisbane-based stage 2 genotyping and assisted with data management. S.A., C.S.H. and A.M.D. coordinated the stage 2 genotyping of the SEARCH samples. M.S. coordinated overall management of data for SEARCH samples. D.L., P.H., K.C., J. Liu, J. Li, I.T., K.H., M.G.-C., N.W., H.Y., S.C., X.-O.S. and J. Long coordinated the stage 2 genotyping, or extraction of existing genotype data, for the LES, SASBAC, NSECG, PECS and SECGS samples. The following authors coordinated the baseline studies and/or extraction of questionnaire and clinical information for studies included in stage 2 analysis: BECS (P.A.F., M.W.B., A.H. and A.B.E.); LES (D.L., L.C., I.V. and F.A.); MoMaTEC (H.B.S., J.T., H.H. and T.S.N.); NECS (R.J.S., K.A., T.P. and G.O.); NSECG (I.T., K.H., M.G. and S.H.); PECS (M.G.-C., H.Y. and N.W.); SASBAC (P.H., K.C., and J. Li); SECGS (X.-O.S. and W.Z. (principal investigators), J. Long (principal study geneticist) and Y.-B.X. (site principal investigator at the Shanghai Cancer Institute)). All authors provided critical review of the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/naturegenetics/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. Ferlay, J. *et al.* GLOBOCAN 2008, cancer incidence and mortality worldwide. in *CancerBase No. 10, Vol. 2010* (International Agency for Research on Cancer Lyon, France, 2010).
2. Varol, N. *et al.* Ten-year review of hysterectomy morbidity and mortality: can we change direction? *Aust. N. Z. J. Obstet. Gynaecol.* **41**, 295–302 (2001).
3. Hindorff, L.A., Junkins, H.A., Hall, P.N., Mehta, J.P. & Manolio, T.A. *A Catalog of Published Genome-Wide Association Studies. Vol. 2010* (National Human Genome Research Institute, Bethesda, Maryland, USA, 2010).
4. Goldgar, D.E., Easton, D.F., Cannon-Albright, L.A. & Skolnick, M.H. Systematic population-based assessment of cancer risk in first-degree relatives of cancer probands. *J. Natl. Cancer Inst.* **86**, 1600–1608 (1994).
5. Hemminki, K., Vaittinen, P. & Kyyronen, P. Age-specific familial risks in common cancers of the offspring. *Int. J. Cancer* **78**, 172–175 (1998).
6. Gruber, S.B. & Thompson, W.D. A population-based study of endometrial cancer and familial risk in younger women. Cancer and Steroid Hormone Study Group. *Cancer Epidemiol. Biomarkers Prev.* **5**, 411–417 (1996).
7. Lucenteforte, E. *et al.* Family history of cancer and the risk of endometrial cancer. *Eur. J. Cancer Prev.* **18**, 95–99 (2009).
8. Haidopoulos, D. *et al.* Risk factors in women 40 years of age and younger with endometrial carcinoma. *Acta Obstet. Gynecol. Scand.* **89**, 1326–1330 (2010).
9. Hampel, H. *et al.* Screening for Lynch syndrome (hereditary nonpolyposis colorectal cancer) among endometrial cancer patients. *Cancer Res.* **66**, 7810–7817 (2006).
10. Setiawan, V.W. *et al.* Two estrogen-related variants in *CYP19A1* and endometrial cancer risk: a pooled analysis in the Epidemiology of Endometrial Cancer Consortium. *Cancer Epidemiol. Biomarkers Prev.* **18**, 242–247 (2009).
11. Low, Y.L. *et al.* Multi-variant pathway association analysis reveals the importance of genetic determinants of estrogen metabolism in breast and endometrial cancer susceptibility. *PLoS Genet.* **6**, e1001012 (2010).
12. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–678 (2007).
13. Price, A.L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
14. Gudmundsson, J. *et al.* Genome-wide association and replication studies identify four variants associated with prostate cancer susceptibility. *Nat. Genet.* **41**, 1122–1126 (2009).
15. Gudmundsson, J. *et al.* Two variants on chromosome 17 confer prostate cancer risk, and the one in *TCF2* protects against type 2 diabetes. *Nat. Genet.* **39**, 977–983 (2007).
16. Thomas, G. *et al.* Multiple loci identified in a genome-wide association study of prostate cancer. *Nat. Genet.* **40**, 310–315 (2008).
17. Elliott,, K.S. *et al.* Evaluation of association of *HNF1B* variants with diverse cancers: collaborative analysis of data from 19 genome-wide association studies. *PLoS ONE* **5**, e10858 (2010).
18. Voight, B.F. *et al.* Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat. Genet.* **42**, 579–589 (2010).
19. Kasper, J.S., Liu, Y. & Giovannucci, E. Diabetes mellitus and risk of prostate cancer in the health professionals follow-up study. *Int. J. Cancer* **124**, 1398–1403 (2009).
20. Stevens, V.L. *et al. HNF1B* and *JAZF1* genes, diabetes, and prostate cancer risk. *Prostate* **70**, 601–607 (2010).
21. Hemminki, K., Li, X., Sundquist, J. & Sundquist, K. Risk of cancer following hospitalization for type 2 diabetes. *Oncologist* **15**, 548–555 (2010).
22. Hjartåker, A., Langseth, H. & Weiderpass, E. Obesity and diabetes epidemics: cancer repercussions. *Adv. Exp. Med. Biol.* **630**, 72–93 (2008).
23. Noto, H., Osame, K., Sasazuki, T. & Noda, M. Substantially increased risk of cancer in patients with diabetes mellitus: a systematic review and meta-analysis of epidemiologic evidence in Japan. *J. Diabetes Complications* **24**, 345–353 (2010).
24. Rosato, V. *et al.* Metabolic syndrome and endometrial cancer risk. *Ann. Oncol.* published online, doi:10.1093/annonc/mdq464 (11 October 2010).
25. Bach, I. & Yaniv, M. More potent transcriptional activators or a transdominant inhibitor of the HNF1 homeoprotein family are generated by alternative RNA processing. *EMBO J.* **12**, 4229–4242 (1993).
26. Bernardini, L. *et al.* Recurrent microdeletion at 17q12 as a cause of Mayer-Rokitansky-Kuster-Hauser (MRKH) syndrome: two case reports. *Orphanet J. Rare Dis.* **4**, 25 (2009).
27. Oram, R.A. *et al.* Mutations in the hepatocyte nuclear factor-1β (*HNF1B*) gene are common with combined uterine and renal malformations but are not found with isolated uterine malformations. *Am. J. Obstet. Gynecol.* **203**, 364 e1–364.e5 (2010).
28. Kato, N. & Motoyama, T. Expression of hepatocyte nuclear factor-1β in human urogenital tract during the embryonic stage. *Anal. Quant. Cytol. Histol.* **31**, 34–40 (2009).
29. Kim, L. *et al.* Clear cell carcinoma of the pancreas: histopathologic features and a unique biomarker: hepatocyte nuclear factor-1β. *Mod. Pathol.* **21**, 1075–1083 (2008).
30. Kato, N., Sasou, S. & Motoyama, T. Expression of hepatocyte nuclear factor-1β (HNF-1β) in clear cell tumors and endometriosis of the ovary. *Mod. Pathol.* **19**, 83–89 (2006).
31. Kato, N. & Motoyama, T. Hepatocyte nuclear factor-1β (HNF-1β) in human urogenital organs: its expression and role in embryogenesis and tumorigenesis. *Histol. Histopathol.* **24**, 1479–1486 (2009).
32. Tsuchiya, A. *et al.* Expression profiling in ovarian clear cell carcinoma: identification of hepatocyte nuclear factor-1 β as a molecular marker and a possible molecular target for therapy of ovarian clear cell carcinoma. *Am. J. Pathol.* **163**, 2503–2512 (2003).
33. Mahata, P. Biomarkers for epithelial ovarian cancers. *Genome Inform.* **17**, 184–193 (2006).
34. Harries, L.W., Perry, J.R., McCullagh, P. & Crundwell, M. Alterations in *LMTK2, MSMB* and *HNF1B* gene expression are associated with the development of prostate cancer. *BMC Cancer* **10**, 315 (2010).
35. Liu, H., Wang, B. & Han, C. Meta-analysis of genome-wide and replication association studies on prostate cancer. *Prostate* **71**, 209–224 (2011).
36. Higgins, J.P., Thompson, S.G., Deeks, J.J. & Altman, D.G. Measuring inconsistency in meta-analyses. *Br. Med. J.* **327**, 557–560 (2003).

ONLINE METHODS

Samples. Stage 1 and 2 sample sets are described in **Supplementary Table 2** and in the **Supplementary Note**. The final stage 1 case sample set included 1,265 endometrioid endometrial cancer cases with self-reported European ancestry from the Australian National Endometrial Cancer Study (ANECS, $n = 599$) or the Studies of Epidemiology and Risk factors in Cancer Heredity study (SEARCH, $n = 666$) in the UK. Control samples were genotyped as part of the Wellcome Trust Case Control Consortium (WTCCC2)[12]. The final stage 1 control set included 5,190 controls with valid genotype data available at the time of analysis. Stage 2 encompassed a total of 3,957 cases and 6,886 controls from nine countries, including additional cases of European-ancestry from ANECS and SEARCH and female controls from these studies (**Supplementary Table 2** and **Supplementary Note**).

Genotyping and quality control. Genotypes for stage 1 cases were generated using an Illumina Infinium 610K array and called using the Illumina GenCall algorithm. Controls were genotyped using an Illumina Infinium 1.2M array as part of WTCCC2 and called using the Illuminus algorithm using genotypes that were successfully called with posterior probability >0.95 (ref. 37). Analyses were restricted to the 519,655 SNPs meeting the following criteria: call rate ≥95% if minor allele frequency (MAF) ≥ 5% (or call rate ≥ 99% if MAF < 5%), Hardy-Weinberg equilibrium (HWE) $P > 10^{-12}$ (cases) or HWE $P > 10^{-7}$ with no difference in frequency between the two WTCCC2 control groups at $P < 10^{-6}$ (controls). The duplicate concordance was 99.998%.

Genotypes were available for 1,344 cases with endometrial cancer. A subset of individuals was identified for exclusion as follows: two individuals with probable Turner's syndrome and two males based on genotypes for markers on the X and Y chromosomes; samples with call rate <97% ($n = 14$); samples with heterozygosity <0.65 or >0.68 ($n = 11$); the sample with the lower call rate from two probable sibling pairs and 26 duplicate pairs, identified as close relatives by identity-by-state probabilities >0.85; eight individuals with >15% non-European ancestry as indicated from computing identity-by-state scores between participants and individuals in HapMap and multidimensional scaling. Twenty-one cases were also excluded from the final analysis of stage 1 because of unresolved discrepancies between their stage 1 and stage 2 genotypes, leaving a total of 1,265 cases. The WTCCC2 control data (5,190 individuals) had been cleaned for a previous study[38] to remove probable close relatives and individuals with >15% estimated non-European ancestry, low or high heterozygosity (<0.65 or >0.68) or call rate <97%.

Genotyping for stage 2 was performed as indicated in **Supplementary Table 2**. All studies complied with quality-control standards by including ≥2 no DNA template controls per 384-well assay plate, ≥2% of samples in duplicate, genotyping call rate >95% and ≥98% concordance between duplicated samples for each SNP assay. The raw data were reviewed for sample sets with evidence for departure from HWE using the χ^2 test (1 d.f.) as a marker of poor genotyping quality. Plates or studies with HWE $P < 0.0001$ were automatically excluded from combined analysis.

Statistical methods. Stage 1 genotype frequencies were compared between cases and controls using the 1-degree-of-freedom Cochran-Armitage trend (per-allele) test. Population stratification was adjusted for using the first three principal components of the genomic kinship matrix, as estimated using 28,494 uncorrelated SNPs ($r^2 < 0.1$). The inflation factor was computed from the lower 90% of the χ^2 statistics.

The 49 SNPs genotyped in stage 2 were chosen from the top 200 stage 1 SNPs after assessing genotyping quality using the cluster plots. In the case of correlated SNPs within a region, multiple logistic regression was used to select the best candidate(s) for stage 2. For the most strongly associated SNP (rs4239217), we included a second correlated SNP (rs4430796).

We genotyped 1,275 case samples (585 ANECS and 680 SEARCH) in stages 1 and 2, allowing us to check between-stage genotyping concordance. For two SNPs (rs4862110 and rs3019885), both of which appeared to be highly significant in stage 1, the concordance was very poor (86.2% and 86.7%, respectively, similar in the ANECS and SEARCH sets). These discrepancies appeared to be because of poor genotyping of these SNPs on the Illumina 610K platform. Therefore, we removed both SNPs from the analyses. After exclusion of these two SNPs, 21 samples showed poor overall concordance and were excluded from the final analysis of stage 1, with no important differences from the original analysis. For the remaining samples, there was a maximum of one discordance per sample. The 1,275 samples genotyped in both stages were included only in the stage 1 analysis.

We compared stage 2 genotype frequencies between cases and controls using the 1-degree-of-freedom trend test and the 2-degree-of-freedom genotype test, with ORs and 95% CIs estimated using unconditional logistic regression stratified by study. Heterogeneity between studies was expressed using the I^2 statistic. Results from stages 1 and 2 were combined using a fixed-effects analysis. Stage 2 data were also analyzed separately for the subgroups of cases with or without endometrioid histology, and the genotype frequencies in these two groups were compared in a case-only analysis.

We examined the pattern of linkage disequilibrium in the 50 kb around rs4430796 using genotypes obtained from the 1000 Genomes Project (August 2010 release)[39]. No SNPs outside a 15.3-kb block had an $r^2 > 0.2$ with rs4430796. This region contained 42 SNPs, six of which had been genotyped in stage 1 of our study. Non-genotyped SNPs were imputed for all stage 1 samples using the 1000 Genomes Project data as a reference panel. Imputed genotype dosages were compared between cases and controls, adjusting for the first three principal components of the genomic kinship matrix. Haplotype frequencies based on the genotyped SNPs were compared between cases and controls.

Analyses were performed in R (including GenABEL[40] and SNPMatrix[41]), ProbABEL[42], MACH[43], Haploview[44] and Stata (see URLs). All statistical tests were two-sided.

37. Teo, Y.Y. *et al.* A genotype calling algorithm for the Illumina BeadArray platform. *Bioinformatics* **23**, 2741–2746 (2007).
38. Turnbull, C. *et al.* Genome-wide association study identifies five new breast cancer susceptibility loci. *Nat. Genet.* **42**, 504–507 (2010).
39. Durbin, R.M. *et al.* A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).
40. Aulchenko, Y.S., Ripke, S., Isaacs, A. & van Duijn, C.M. GenABEL: an R library for genome-wide association analysis. *Bioinformatics* **23**, 1294–1296 (2007).
41. Clayton, D. & Leung, H.T. An R package for analysis of whole-genome association studies. *Hum. Hered.* **64**, 45–51 (2007).
42. Aulchenko, Y.S., Struchalin, M.V. & van Duijn, C.M. ProbABEL package for genome-wide association analysis of imputed data. *BMC Bioinformatics* **11**, 134 (2010).
43. Li, Y., Willer, C., Sanna, S. & Abecasis, G. Genotype imputation. *Annu. Rev. Genomics Hum. Genet.* **10**, 387–406 (2009).
44. Barrett, J.C., Fry, B., Maller, J. & Daly, M.J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263–265 (2005).