NIH-PA Author Manuscript

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,*], **Kaltin Ferguson**[1,*], **Catherine S. Healey**[3,*], **Tracy O'Mara**[1,4,*], **Logan C. Walker**[1], **Stephen B. Montgomery**[5], **Emmanouil T. Dermitzakis**[5], **Paul Fahey**[1], **Grant Montgomery**[1], **Penelope M. Webb**[1], **Peter A. Fasching**[7], **Matthias W. Beckmann**[8], **Arif B. Ekici**[8], **Alexander Hein**[9], **Diether Lambrechts**[10,11], **Lieve Coenegrachts**[12], **Ignace Vergote**[12], **Frederic Amant**[12], **Helga B. Salvesen**[13,14], **Jone Trovik**[13,14], **Tormund S Njolstad**[13,14], **Harald Helland**[13], **Rodney Scott**[15,16,17], **Katie Ashton**[15], **Tony Proietto**[18], **Geoffrey Otton**[18], **Ian Tomlinson**[19], **Maggie Gorman**[19], **Kimberley Howarth**[19], **Shirley Hodgson**[20], **Montserrat Garcia-Closas**[21,22], **Nicolas Wentzensen**[21], **Hannah Yang**[21], **Stephen Chanock**[21], **Per Hall**[23], **Kamila Czene**[23], **Jianjun Liu**[24], **Jingmei Li**[23,24], **Xiao-Ou Shu**[25], **Wei Zheng**[25], **Jirong Long**[25], **Yong-Bing Xiang**[26], **Mitul Shah**[3], **Jonathan Morrison**[2], **Kyriaki Michailidou**[2], **Paul P. Pharoah**[2,3], **Alison M. Dunning**[3], **Douglas F. Easton**[2,3], **The Australian National Endometrial Cancer Study Group**[1,6], and **National Study of Endometrial Cancer Genetics Group**[6,19]

[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, Musk Avenue, Kelvin Grove, Brisbane, Queensland, Australia [5]Department of Genetic Medicine and Development, University of Geneva Medical School, 1 rue Michel-Servet, Geneva 1211, Switzerland [7]University of California at Los Angeles, David Geffen School of Medicine, Department of Medicine, Division of Hematology and Oncology, Los Angeles, CA, USA [8]University Hospital Erlangen, Department of Gynecology and Obstetrics, Friedrich Alexander University Erlangen Nürnberg, Erlangen, Germany [9]Institute of Human Genetics, Friedrich Alexander University Erlangen Nürnberg, Erlangen, Germany [10]Vesalius Research Center, Vlaams Instituut voor Biotechnologie, Leuven, Belgium [11]Vesalius Research Center, University of Leuven, Belgium [12]Division of Gynecologic Oncology, Department of Obstetrics and Gynecology, University Hospitals Leuven, Belgium [13]Department of Obstetrics and Gynecology, Haukeland University Hospital, Bergen, Norway [14]Department of Clinical Medicine, The University of Bergen, Norway [15]The Centre for Information Based Medicine and the Discipline of Medical Genetics, School of Biomedical Sciences and Pharmacy, Faculty of Health, University of Newcastle, NSW, Australia and The Hunter Medical Research Institute, John Hunter Hospital NSW, Australia [16]Division of Genetics, Hunter Area Pathology Service, John Hunter Hospital, Newcastle. NSW, Australia [17]Hunter Centre for Gynaecological, John Hunter Hospital, NSW, Australia [18]School of Medicine and Public Health, Faculty of Health, University of Newcastle, NSW, Australia [19]Wellcome Trust Centre for Human Genetics, University of Oxford, Rosselvelt Drive, Oxford OX3 7BN, UK [20]Dept of Clinical Genetics, St George's Hospital Medical School, Cranmer Terrace, London SW17 0RE, UK [21]Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD, USA [22]Institute of Cancer Research, London, UK [23]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden [24]Human Genetics, Genome Institute of Singapore, Singapore, Singapore [25]Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University Medical Center, Nashville, USA [26]Department of Epidemiology, Shanghai Cancer Institute, Shanghai, China

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 undertook a genome-wide association study involving 1,265 endometrial cancer cases from Australia and the UK and 5,190 controls from the Wellcome Trust Case Control Consortium. Genotype frequencies in cases and controls were compared 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* on chromosome 17q (SNP rs4430796: P=7.1×10^{-10}), that is also associated with risk of prostate cancer and is inversely associated with type 2 diabetes.

Cancer of the uterine corpus, 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 favourable for the most common histological subtype, endometrioid endometrial cancer, representing 80–90% of all endometrial cancers. Nevertheless, the disease associated with significant morbidity due to surgery and radiotherapy[2], and treatment is often complicated because most patients present at older ages and with significant 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 hits 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]. Family history of endometrial 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 the 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 undertaken to date.

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

After combining results from stage 1 and stage 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 range 0.68 – 0.90 in Europeans) encompassing the first four exons of *HNF1B* on chromosome 17q12 (Table 1, http://www.hapmap.org.uk). The most significant association was for rs4430796 (OR per G-allele =0.84, CI=0.79–0.89, $P=7.1\times10^{-10}$). There was no significant heterogeneity in the per-allele OR between stages 1 and 2 (stage 1 OR=0.79, CI=0.73–0.87; stage 2 OR=0.87, 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, CI=0.77–0.91, overall $P=4.3\times10^{-11}$; Table 1). The three *HNF1B* SNPs were not significantly associated with non-endometrioid disease (Table 1), although the numbers of cases were small. In addition, 832 cases and 2,049 controls of Chinese ancestry from Shanghai (Supplementary Table 2) were genotyped for rs11651755, a surrogate for rs4430796 (Hapmap r^2=1.0 in the CHB population). The estimated effect for rs11651755 in this Asian study was OR=0.96 (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 29kB centromeric of *SFPQ* (combined OR=1.21, CI=1.12–1.32, $P=5.9\times10^{-6}$). Although highly significant in stage 1 ($P=6.1\times10^{-7}$), it 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, CI=1.12–1.34; $P=4.6\times10^{-6}$) or non-endometrioid

histology (stage 2 OR=1.07, 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), in line with what would be expected by chance (Supplementary Table 3 and 4).

Multiple studies have independently reported the G allele of SNP 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 (0.76–0.83) for prostate cancer[17], but found no association of the same allele with breast, lung, colorectal or pancreatic cancers or 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 (1.06–1.15)[15], and OR 1.14 (1.08–1.20)[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 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 is recorded, adjusting for BMI did not materially alter the risk estimate (OR=0.87 (0.78–0.97) versus OR=0.88 (0.79–0.99); n=3055).

To provide a more comprehensive analysis of SNPs in the *HNF1B* region, we identified variants in the region using resequencing data in individuals with 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 A and B considered to act as transcription 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 characterized by congenital aplasia of the uterus and upper vagina due to anomalous development of the Mullerian ducts[26], and *HNF1B* mutations/deletions are associated with uterine abnormalities due to 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* over-expression 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 compared to benign tissue[34]. Analysis of several lymphocyte-derived gene expression datasets (See Supplementary Fig. 2) reveals 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*[35] but none of the others revealed any evidence of association with endometrial cancer in this study (Supplementary Table 7). Further common low penetrant 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 highlights the value of the agnostic GWAS approach for identifying previously unexplored biological pathways and novel molecular targets for prevention.

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 1265 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 (SEARCH, n=666) in the United Kingdom. Control samples were genotyped as part of the Wellcome Trust Case Control Consortium (WTCCC2[12]). The final stage 1 control set included 5190 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 Caucasian cases from ANECS and SEARCH cases and female controls from these studies (Supplementary Table 2, Supplementary text).

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[37]. Analyses were restricted to 519,655 SNPs meeting the following criteria; call rate 95% if MAF 5%, (or 99% if MAF<5%), 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). Duplicate concordance was 99.998%.

Genotypes were available for 1344 endometrial cancer cases. A subset of individuals were 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 2 probable sibling pairs and 26 duplicate pairs, identified as close relatives by identity-by-state probabilities >0.85; 8 individuals with >15% non-European

ancestry as indicated from computing IBS 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 (see Statistical Methods), leaving 1265 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. Raw data was reviewed for sample sets with evidence for departure from Hardy-Weinberg equilibrium (HWE) using the Chi-squared test (1 df), 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 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).

1275 case samples (585 ANECS and 680 SEARCH) were genotyped in stages 1 and 2, allowing us to check between-stage genotyping concordance. For two SNPs (rs4862110 and rs3019885), both of which had 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 due to 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 1 discordancy per sample. The 1275 samples genotyped in both stages were included only in the stage 1 analysis.

Stage 2 genotype frequencies were compared between cases and controls using the 1 degree of freedom trend test and the 2 degree of freedom genotype test, with odds ratios and 95% confidence intervals 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 analysed 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.

The pattern of LD in the 50kb around rs4430796 was examined using genotypes obtained from the 1000 Genomes Project (August 2010 release)[39] No SNPs outside a 15.3kb block had an r^2>0.2 with rs4430796. This region contained 42 SNPs, 6 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 3 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 (R, http://www.r-project.org/; Stata, http://www.stata.com/). All statistical tests were 2-sided.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Ferlay, J., et al. IARC. CancerBase No. 10 [Internet]. Vol. 2010. International Agency for Research on Cancer Lyon; France: 2010. GLOBOCAN 2008, Cancer Incidence and Mortality Worldwide.

2. Varol N, et al. Ten-year review of hysterectomy morbidity and mortality: can we change direction? Aust N Z J Obstet Gynaecol. 2001; 41:295–302. [PubMed: 11592544]

3. Hindorff, LA.; Junkins, HA.; Hall, PN.; Mehta, JP.; Manolio, TA. A Catalog of Published Genome-Wide Association Studies. National Human Genome Research Institute; 2010.

4. Goldgar DE, Easton DF, Cannon-Albright LA, Skolnick MH. Systematic population-based assessment of cancer risk in first-degree relatives of cancer probands. J Natl Cancer Inst. 1994; 86:1600–8. [PubMed: 7932824]

5. Hemminki K, Vaittinen P, Kyyronen P. Age-specific familial risks in common cancers of the offspring. Int J Cancer. 1998; 78:172–5. [PubMed: 9754648]

6. Gruber SB, Thompson WD. A population-based study of endometrial cancer and familial risk in younger women. Cancer and Steroid Hormone Study Group. Cancer Epidemiol Biomarkers Prev. 1996; 5:411–7. [PubMed: 8781735]

7. Lucenteforte E, et al. Family history of cancer and the risk of endometrial cancer. Eur J Cancer Prev. 2009; 18:95–9. [PubMed: 19337055]

8. Haidopoulos D, et al. Risk factors in women 40 years of age and younger with endometrial carcinoma. Acta Obstet Gynecol Scand. 2010; 89:1326–30. [PubMed: 20846065]

9. Hampel H, et al. Screening for Lynch syndrome (hereditary nonpolyposis colorectal cancer) among endometrial cancer patients. Cancer Res. 2006; 66:7810–7. [PubMed: 16885385]

10. Setiawan VW, 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. 2009; 18:242–7. [PubMed: 19124504]

11. Low YL, et al. Multi-variant pathway association analysis reveals the importance of genetic determinants of estrogen metabolism in breast and endometrial cancer susceptibility. PLoS Genet. 2010; 6:e1001012. [PubMed: 20617168]

12. Genome-wide association study of 14, 000 cases of seven common diseases and 3, 000 shared controls. Nature. 2007; 447:661–78. [PubMed: 17554300]

13. Price AL, et al. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006; 38:904–9. [PubMed: 16862161]

14. Gudmundsson J, et al. Genome-wide association and replication studies identify four variants associated with prostate cancer susceptibility. Nat Genet. 2009; 41:1122–6. [PubMed: 19767754]

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. 2007; 39:977–83. [PubMed: 17603485]

16. Thomas G, et al. Multiple loci identified in a genome-wide association study of prostate cancer. Nat Genet. 2008; 40:310–5. [PubMed: 18264096]

17. Elliott KS, et al. Evaluation of association of HNF1B variants with diverse cancers: collaborative analysis of data from 19 genome-wide association studies. PLoS One. 2010; 5:e10858. [PubMed: 20526366]

18. Voight BF, et al. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. Nat Genet. 2010; 42:579–89. [PubMed: 20581827]

19. Kasper JS, Liu Y, Giovannucci E. Diabetes mellitus and risk of prostate cancer in the health professionals follow-up study. Int J Cancer. 2009; 124:1398–403. [PubMed: 19058180]

20. Stevens VL, et al. HNF1B and JAZF1 genes, diabetes, and prostate cancer risk. Prostate. 2010; 70:601–7. [PubMed: 19998368]

21. Hemminki K, Li X, Sundquist J, Sundquist K. Risk of cancer following hospitalization for type 2 diabetes. Oncologist. 2010; 15:548–55. [PubMed: 20479278]

22. Hjartaker A, Langseth H, Weiderpass E. Obesity and diabetes epidemics: cancer repercussions. Adv Exp Med Biol. 2008; 630:72–93. [PubMed: 18637486]

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. 2010; 24:345–53. [PubMed: 20656522]

24. Rosato V, et al. Metabolic syndrome and endometrial cancer risk. Ann Oncol. 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. 1993; 12:4229–42. [PubMed: 7900999]

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. 2009; 4:25. [PubMed: 19889212]

27. Oram RA, et al. Mutations in the hepatocyte nuclear factor-1beta (HNF1B) gene are common with combined uterine and renal malformations but are not found with isolated uterine malformations. Am J Obstet Gynecol. 2010; 203:364 e1–5. [PubMed: 20633866]

28. Kato N, Motoyama T. Expression of hepatocyte nuclear factor-1beta in human urogenital tract during the embryonic stage. Anal Quant Cytol Histol. 2009; 31:34–40. [PubMed: 19320191]

29. Kim L, et al. Clear cell carcinoma of the pancreas: histopathologic features and a unique biomarker: hepatocyte nuclear factor-1beta. Mod Pathol. 2008; 21:1075–83. [PubMed: 18536653]

30. Kato N, Sasou S, Motoyama T. Expression of hepatocyte nuclear factor-1beta (HNF-1beta) in clear cell tumors and endometriosis of the ovary. Mod Pathol. 2006; 19:83–9. [PubMed: 16258507]

31. Kato N, Motoyama T. Hepatocyte nuclear factor-1beta(HNF-1beta) in human urogenital organs: its expression and role in embryogenesis and tumorigenesis. Histol Histopathol. 2009; 24:1479–86. [PubMed: 19760597]

32. Tsuchiya A, et al. Expression profiling in ovarian clear cell carcinoma: identification of hepatocyte nuclear factor-1 beta as a molecular marker and a possible molecular target for therapy of ovarian clear cell carcinoma. Am J Pathol. 2003; 163:2503–12. [PubMed: 14633622]

33. Mahata P. Biomarkers for epithelial ovarian cancers. Genome Inform. 2006; 17:184–93. [PubMed: 17503391]

34. Harries LW, Perry JR, McCullagh P, Crundwell M. Alterations in LMTK2, MSMB and HNF1B gene expression are associated with the development of prostate cancer. BMC Cancer. 2010; 10:315. [PubMed: 20569440]

35. Liu H, Wang B, Han C. Meta-analysis of genome-wide and replication association studies on prostate cancer. Prostate. 2010

36. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ. 2003; 327:557–60. [PubMed: 12958120]

37. Teo YY, et al. A genotype calling algorithm for the Illumina BeadArray platform. Bioinformatics. 2007; 23:2741–6. [PubMed: 17846035]

38. Turnbull C, et al. Genome-wide association study identifies five new breast cancer susceptibility loci. Nat Genet. 2010; 42:504–7. [PubMed: 20453838]

39. Durbin RM, et al. A map of human genome variation from population-scale sequencing. Nature. 2010; 467:1061–73. [PubMed: 20981092]

40. Aulchenko YS, Ripke S, Isaacs A, van Duijn CM. GenABEL: an R library for genome-wide association analysis. Bioinformatics. 2007; 23:1294–6. [PubMed: 17384015]

41. Clayton D, Leung HT. An R package for analysis of whole-genome association studies. Hum Hered. 2007; 64:45–51. [PubMed: 17483596]

42. Aulchenko YS, Struchalin MV, van Duijn CM. ProbABEL package for genome-wide association analysis of imputed data. BMC Bioinformatics. 2010; 11:134. [PubMed: 20233392]

43. Li Y, Willer C, Sanna S, Abecasis G. Genotype imputation. Annu Rev Genomics Hum Genet. 2009; 10:387–406. [PubMed: 19715440]

44. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics. 2005; 21:263–5. [PubMed: 15297300]
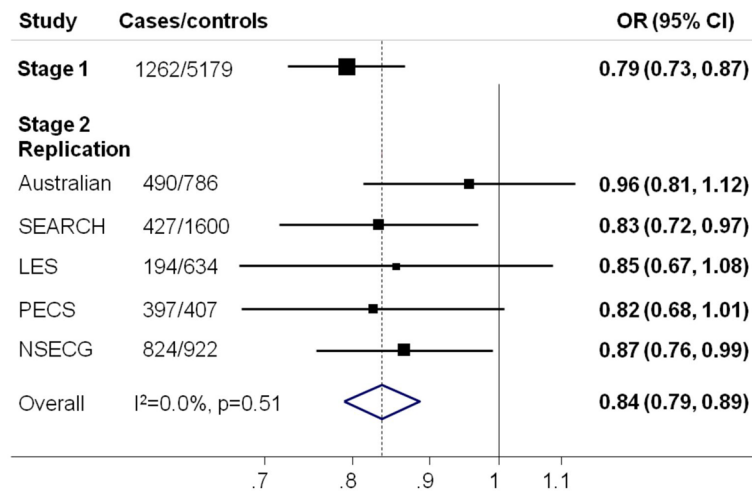
Figure 1.
Forest plot showing the association between SNP rs4430796 and endometrial cancer for each component of the study. OR=odds ratio. Study abbreviations are as given in Supplementary Table 2. SEARCH and Australian Stage2 replication sets exclude samples included in Stage 1. The solid line denotes the null, and the dashed line indicates the overall OR estimate. I^2 measures the heterogeneity in effect size between studies[36]

Table 1

Association between genotype and endometrial cancer for three SNPs in the *HNF1B* gene

Marker	Chromosome *Position*	Stage	Case N	Control N	Case MAF	Control MAF	Stage per-allele OR (95% CI)	Stage *P*-value	Combined per-allele OR (95% CI)	Combined *P*-value
rs4430796 A/G	17 *33,172,153*	Stage 1	1262	5179	0.43	0.48	0.79 (0.73–0.87)	3.06E-07		
		Stage 2 all cases	2332	4349	0.44	0.48	0.87 (0.81–0.94)	2.00E-04	0.84 (0.79–0.89)	7.11E-10
		Stage 2 endometrioid cases	*1786*	*4349*	*0.43*	*0.48*	*0.84 (0.77–0.91)*	*2.56E-05*	*0.82 (0.77–0.87)*	*4.28E-11*
		Stage 2 non-endometrioid cases	*484*	*4349*	*0.46*	*0.48*	*0.94 (0.82–1.08)*	*3.74E-01*		
rs4239217 A/G	17 *33,173,100*	Stage 1	1265	5190	0.35	0.41	0.79 (0.72–0.86)	2.48E-07		
		Stage 2 all cases	2342	4307	0.37	0.40	0.89 (0.82–0.96)	2.00E-03	0.84 (0.80–0.90)	1.19E-08
		Stage 2 endometrioid cases	*1753*	*4307*	*0.36*	*0.40*	*0.86 (0.79–0.94)*	*5.90E-04*	*0.83 (0.78–0.88)*	*1.27E-09*
		Stage 2 non-endometrioid cases	*588*	*4307*	*0.39*	*0.40*	*0.97 (0.85–1.10)*	*6.18E-01*		
rs7501939 G/A	17 *33,175,269*	Stage 1	1263	5187	0.36	0.40	0.80 (0.73–0.88)	2.17E-06		
		Stage 2 all cases	3337	5234	0.37	0.40	0.90 (0.84–0.96)	1.00E-03	0.86 (0.82–0.91)	5.35E-08
		Stage 2 endometrioid cases	*2562*	*5234*	*0.36*	*0.40*	*0.88 (0.82–0.94)*	*3.30E-04*	*0.85 (0.80–0.90)*	*7.57E-09*
		Stage 2 non-endometrioid cases	*690*	*5234*	*0.38*	*0.40*	*0.95 (0.85–1.07)*	*4.29E-01*		

*
Minor Allele Frequency