Identification of nine new susceptibility loci for endometrial cancer

Tracy A. O’Mara et al.#

Endometrial cancer is the most commonly diagnosed cancer of the female reproductive tract in developed countries. Through genome-wide association studies (GWAS), we have previously identified eight risk loci for endometrial cancer. Here, we present an expanded meta-analysis of 12,906 endometrial cancer cases and 108,979 controls (including new genotype data for 5624 cases) and identify nine novel genome-wide significant loci, including a locus on 12q24.12 previously identified by meta-GWAS of endometrial and colorectal cancer. At five loci, expression quantitative trait locus (eQTL) analyses identify candidate causal genes; risk alleles at two of these loci associate with decreased expression of genes, which encode negative regulators of oncogenic signal transduction proteins (SH2B3 (12q24.12) and NF1 (17q11.2)). In summary, this study has doubled the number of known endometrial cancer risk loci and revealed candidate causal genes for future study.

Correspondence and requests for materials should be addressed to T.A.O’M. (email: Tracy.OMara@qimrberghofer.edu.au) or to A.B.S. (email: Amanda.Spurdle@qimrberghofer.edu.au) or to D.J.T. (email: dt25@medschl.cam.ac.uk). #A full list of authors and their affiliations appears at the end of the paper.
Endometrial cancer accounts for ~7% of new cancer cases in women\(^1\) and is the most common invasive gynecological cancer in developed countries (http://gco.iarc.fr/today/home). Risk of endometrial cancer is approximately double for women who have a first degree relative with endometrial cancer\(^2\). Rare high-risk pathogenic variants in mismatch-repair genes, PTEN, and DNA polymerase genes\(^4\) explain a small proportion of endometrial cancers, and the eight previously published common endometrial cancer-associated single-nucleotide polymorphisms (SNPs) identified by genome-wide association studies (GWAS) studies\(^5\) together explain <5% of the familial relative risk (FRR).

Here, we conduct a meta-GWAS including 12,906 endometrial cancer cases and 108,979 country-matched controls of European ancestry from 17 studies identified via the Endometrial Cancer Association Consortium (ECAC), the Epidemiology of Endometrial Cancer Consortium (E2C2) and the UK Biobank and report a further nine genome-wide significant endometrial cancer genetic risk regions. Of these, nine regions on 12q24.12 was previously identified by meta-GWAS of endometrial and colorectal cancer\(^3\), eQTL and gene network analyses reveal candidate causal genes and pathways relevant for endometrial carcinogenesis.

### Results

**GWAS meta-analysis.** Details of genotyping for each study are found in Supplementary Data 1 and individual studies described in the Supplementary Information. Following standard quality control (QC) for each dataset (Supplementary Methods), genotypes were imputed using the 1000 Genomes Project v3 reference panel (combined with the UK10K reference panel for the WHI and UK Biobank studies). SNP-disease associations in each study were tested using logistic regression, adjusting for principal components, and risk estimates were combined using inverse-variance weighted fixed-effects meta-analysis. We found little evidence of genomic inflation in any dataset (\(\hat{\lambda}_{1000}=0.996–1.128\) or overall (\(\hat{\lambda}_{1000}=1.004\)) (Supplementary Fig. 1). Using linkage disequilibrium (LD) score regression, we estimate that 93% of the observed test statistic inflation is due to polygenic signal, as opposed to population stratification.

Seven of the eight published genome-wide significant endometrial cancer loci were confirmed with increased significance (Table 1, Fig. 1a), although the effect sizes for some loci were slightly attenuated compared with our previous analysis (comprising 7737 cases and 37,144 controls\(^7\), all also included in the current analysis). For example, the most significant SNP in this meta-analysis, rs1263761 in HNPP\(_B\), had an odds ratio (OR) = 1.15 (1.12–1.19; \(P = 3.2 \times 10^{-20}\)) compared with OR = 1.20 (1.15–1.25; \(P = 2.8 \times 10^{-19}\)) in our previous analysis\(^2\). The previously reported associations with intronic AKTI SNPs (rs2498796 OR = 1.17 (1.07–1.17); \(P = 3.6 \times 10^{-8}\)) did not reach genome-wide significance (rs2498796 OR = 1.07 (1.03–1.11); \(P = 6.3 \times 10^{-5}\)).

Excluding the 500 kb, either side of the risk loci previously reported at genome-wide significance for endometrial cancer alone, we found 125 SNPs with \(P < 5 \times 10^{-8}\). Using approximate conditional association testing with GCTA software\(^11\), these were reduced to eleven independent risk loci; eight newly reported regions, plus the 12q24.12 locus previously identified by a joint endometrial-colorectal cancer analysis\(^7\) (Table 2, Fig. 1a–i). The BFDP was ≤4% for all nine novel loci. The analysis was repeated with the restricted set of 8758 cases with endometrioid cancer, the most common histology (Fig. 1c); this identified one additional variant at 7p14.3 reaching genome-wide significance (rs9639594; Supplementary Data 2). However, given the sparse LD at this region and the fact that this is a single, imputed variant, further investigation of this region is required to confirm its association with endometrial cancer risk. No SNP reached genome-wide significance in an analysis restricted to the 1230 non-endometrioid cases (Fig. 1d) or in separate analyses of carcinosarcomas, serous, clear cell or mucinous carcinomas, for which statistical power is very limited (Supplementary Data 2, Supplementary Fig. 2).

For these nine newly reported endometrial cancer loci, a statistically significant difference in risk estimates by histological subgroup was observed only for the 2p16.1 locus; the risk was higher for non-endometrioid than for endometrioid cancer (rs148261157 OR = 1.64 (1.32–2.04) and OR = 1.25 (1.14–1.38), respectively, case-only \(P_I = 0.003\,\text{Table 2}\).

### Table 1: Meta-analysis results for previously identified genome-wide significant endometrial cancer risk loci

| Region | SNP | Position (bp)\(^a\) | Nearby gene(s) | Effect: other alleles | EAF Info | Allelic OR (95% CI) | \(P^2\) | BFDP (%) | Allelic OR (95% CI) | \(P^2\) | BFDP (%) |
|--------|-----|---------------------|----------------|----------------------|---------|---------------------|--------|----------|---------------------|--------|----------|
| 6p22.3 | rs1540828 | 21,648,854 | SOX4 | G | 0.52 | G | 1.15 (1.11, 1.19) | 4.2E-05 | <1% | 1.16 (1.11, 1.20) | 6.0E-01 | 1.17 |
| 6p22.3 | rs747716 | 125,687,226 | HEV2 | A | 0.57 | 1.00 (1.00, 1.00) | 2.9E-05 | <1% | 1.12 (1.08, 1.16) | 4.4E-01 | 1.12 |
| 8q24.2\(^b\) | rs53286446 | 128,433,617 | MYC | G | 0.58 | 0.99 (0.98, 1.00) | 3.1E-06 | <1% | 1.10 (1.06, 1.14) | 1.8E-01 | 1.10 |
| 8q24.2\(^b\) | rs4733631 | 128,587,032 | MYC | C | 0.12 | G | 1.18 (1.13, 1.24) | 2.4E-14 | <1% | 1.21 (1.15, 1.27) | 1.2E-01 | 1.21 |
| 8q24.2\(^b\) | rs309584729 | 128,161,956 | MYC | C | 0.98 | 0.97 (0.96, 0.98) | 1.4E-02 | <1% | 1.48 (1.28, 1.70) | 7.6E-01 | 1.48 |
| 13q22.1 | rs7898163 | 73,238,004 | KLF5 | C | 0.12 | G | 1.16 (1.12, 1.20) | 2.7E-06 | <1% | 1.17 (1.13, 1.22) | 4.9E-01 | 1.17 |
| 14q23.3\(^c\) | rs42984796 | 73,776,883 | AKT1 | A | 0.30 | 0.98 (0.97, 1.00) | 6.3E-07 | <1% | 1.09 (1.04, 1.13) | 3.2E-01 | 1.09 |
| 15q15.1 | rs952731 | 40,029,923 | E2F4AK | A | 0.42 | G | 1.09 (1.06, 1.13) | 6.7E-06 | <1% | 1.12 (1.08, 1.16) | 6.9E-01 | 1.12 |
| 15q21.2 | rs71608786 | 51,261,712 | CYP19A1 | G | 0.48 | 1.00 (1.01, 1.19) | 3.3E-03 | <1% | 1.12 (1.08, 1.16) | 2.3E-01 | 1.12 |
| 17q2 | rs71263761 | 73,777,784 | HNF8 | A | 0.52 | 0.98 (0.95, 1.02) | 1.5E-11 | <1% | 1.15 (1.11, 1.19) | 3.4E-01 | 1.15 |

\(\text{EAF: effect allele frequency among control subjects in the UK Biobank, Info: imputation quality score for the OncoArray set, G: genotyped SNPs, } P^2\): heterogeneity \(\chi^2\) statistic, BFDP: Bayes false discovery probability\(\text{.} \)

\(\text{\textbf{a}Position}\) with reference to build 38 of the reference genome

\(\text{\textbf{b}Regions}\) are from the conditional model containing all three SNPs. \(\text{\textbf{rs17488796}}\) OR = 1.07 (1.03–1.11); \(P = 6.3 \times 10^{-5}\), Bayes false discovery probability (BFDP) 98% in this meta-analysis, although the risk estimate direction is consistent with our original finding.
There was no evidence of secondary signals at any of these nine loci after conditioning on the most significant variant. There was no significant between-study heterogeneity (minimum Cochran Q-test $P_{het} = 0.04$, maximum $I^2 = 41\%$, Supplementary Fig. 3), and random-effects meta-analyses produced very similar results (Supplementary Data 2). Twenty-five additional independent loci showed moderately significant ($P < 1 \times 10^{-6}$) associations, nine with endometrial cancer overall, nine specifically with endometrioid histology, and seven with non-endometrioid histology (Supplementary Data 2).

Overlap with published GWAS associations. Using a 100:1 likelihood ratio, "credible causal risk" variants (ccrSNPs) were compiled for each of the nine new endometrial cancer risk loci (Supplementary Data 3). These included 239 variants located in non-coding regions, 2 missense variants (rs2278868 SKAP1 Gly161Ser and rs3184504 SH2B3 Trp262Arg), and 1 synonymous variant (rs1129506 EVI2A Ser23Ser). Comparing to the NHGRI-EBI catalog of published GWAS, 37 SNPs previously associated with a cancer, hormonal trait, or anthropometric phenotype fell within 500 kb of any one of the 37 novel endometrial cancer susceptibility SNP rs3184504 in EVI2A (Supplementary Data 3). These included 239 variants located in non-coding regions, 2 missense variants (rs2278868 SKAP1 Gly161Ser and rs3184504 SH2B3 Trp262Arg), and 1 synonymous variant (rs1129506 EVI2A Ser23Ser). Comparing to the NHGRI-EBI catalog of published GWAS, 37 SNPs previously associated with a cancer, hormonal trait, or anthropometric phenotype fell within 500 kb of any one of the 37 novel endometrial cancer susceptibility SNP rs3184504 in EVI2A (Supplementary Data 3).

eQTL analyses. LD score regression analyses using eQTL results from GTEx showed that endometrial cancer heritability exhibited the strongest evidence for enrichment for variants associated with genes specifically expressed in vaginal and uterine tissue, in line with prior assumptions, although none of the tissue-specific enrichments were significant (weighted regression with jackknife standard errors) after Bonferroni correction, adjusting for the number of tissues tested (Supplementary Fig. 4). eQTL analyses were performed using data from a variety of tissue sources (Supplementary Methods), including endometrial tumor and adjacent normal endometrium tissue from The Cancer Genome Atlas (TCGA), normal cycling endometrium and, in view of the GTEx enrichment results, vaginal and uterine tissue. Additionally, we assessed eQTLs from whole blood, which provided substantially increased power over solid tissue analyses due to increased sample size. eQTLs were detected at five of the nine novel loci (Supplementary Data 3, Supplementary Data 5, Supplementary Figs. 5–13, Table 2).

Gene network analysis. Network analysis was performed using candidate causal genes identified in this study, in addition to candidate causal genes identified in previous studies (Supplementary Data 6). One major network was identified, containing 18 of the 25 candidate causal genes (Supplementary Fig. 4). Network hubs included CCND1, CTNNB1, and P53, which are encoded by genes that are somatically mutated in endometrial cancer. Analysis of the network revealed significant enrichment (Benjamini–Hochberg adjusted $P < 0.05$, hypergeometric test) in relevant pathways such as endometrial cancer signaling, adipogenesis, Wnt/beta-catenin signaling, estrogen-mediated S-phase entry, P53 signaling, and PI3K/AKT signaling (Supplementary Data 7).
Functional annotation of ccrSNPs. Next, ccrSNPs were mapped to epigenomic features from endometrial cancer cell lines (Supplementary Data 3, Supplementary Figs. 5–13). Chromatin immunoprecipitation (ChIP-seq) was used to map histone modifications indicative of promoters or enhancers (H3K4Me1, H3K4Me3, and H3K27Ac) in two endometrial cancer cell lines (Ishikawa and JHUEM-14). Mapping of DNase-I hypersensitivity sites (indicative of open chromatin) and ChIP-seq data for transcription factor binding sites from Ishikawa cells were accessed from ENCODE16. We also included mapping of H3K427Ac histone modifications for uterus and vagina from ENCODE. Overall, 73% of ccrSNPs overlapped at least one epigenomic feature, including at least one ccrSNP per novel risk region. This overlap was significantly greater than the overlap observed for these epigenomic features with ccrSNPs related to, for example, endometriosis17 (51%); Fisher’s exact P = 8.7 × 10−8) or schizophrenia18 (40%; Fisher’s exact P < 2.2 × 10−16). These findings indicate the relevance of the selected cell and tissue types for informing endometrial cancer biology and a role for the assessed epigenomic features in regulatory processes related to the ccrSNPs. Overlaps between ccrSNPs and epigenomic features increased significantly after stimulation with estrogen (50% versus 38% for unstimulated features; Fisher’s exact P = 5.6 × 10−3), emphasizing the importance of estrogen in endometrial cancer etiology.

Mendelian randomization analyses. This expanded meta-analysis allowed us to strengthen our previous Mendelian randomization findings19,20 that higher body mass index (BMI) (P = 1.7 × 10−11), two-sample inverse-variance weighted Mendelian randomization (MR) test, but not waist hip ratio (P = 0.71), is causal for endometrial cancer (Table 3) and that the protective effect of later menarche on endometrial cancer risk (OR = 0.82, 95% CI 0.77–0.87 per year of delayed menarche, P = 2.2 × 10−5) is partially mediated by the known relationship between lower BMI and later menarche, with a more modest protective effect of later menarche after adjusting for genetically predicted BMI (OR = 0.88, 95% CI 0.82–0.94, P = 3.8 × 10−4). The association between genetically predicted age at natural menopause and endometrial cancer did not reach statistical significance (OR = 1.03, 95% CI 1.00–1.06, P = 0.060). In contrast to the reported effects for breast and prostate cancer21,22, we found no evidence that genetically predicted adult height is associated with endometrial cancer (P = 0.90).

Genetic correlation analyses. Cross-trait LD score regression of 224 non-cancer traits available via the LD Hub interface23, identified significant genetic correlations between endometrial cancer and 14 traits. All of these are either a measure of obesity or are strongly and significantly (correlation-corrected jackknife P < 10−12) genetically correlated with BMI (i.e., age of menarche, type 2 diabetes, and years of schooling) (Supplementary Data 8), in line with the established relationship between obesity and endometrial cancer risk.

Discussion

In the largest GWAS meta-analysis assessing endometrial cancer risk, we discovered nine new genetic risk regions. We also confirmed the association of genetic variants with endometrial cancer risk at seven of the eight previously published genetic risk regions for this disease14–8. Using this larger GWAS-meta dataset, we were also able to confirm the previously published Mendelian randomization studies finding that higher BMI is causal for endometrial cancer risk20, and the protective effect of later age of menarche on endometrial cancer risk19. Genetic correlation analyses also indicated a relationship between endometrial cancer and obesity-related traits.

Candidate causal genes identified through eQTLs included CDCA8 (1p34.3), a putative ovarian cancer oncogene24, which encodes an essential regulator of mitosis and cell division25; RCNJ1 (11p13), encoding a calcium-binding protein that binds oncoproteins such as JAK226 and MYC25; WTI-AS (11p13), a long
non-coding RNA that regulates the WTI oncogene\textsuperscript{28,29}, SH2B3 (12p24.11) encoding a negative regulator of the oncogenic KIT and JAK2 signal transduction proteins\textsuperscript{30}, and tumor suppressor gene NF1 (17q11.2) encoding a negative regulator of RAS-mediated signal transduction\textsuperscript{31}, which acquires putative driver mutations in TCGA endometrial tumors (http://www.cbioportal.org/study?id=uccc_tcga). Notably, the highly significant eQTL associations between ccrSNPs and expression of SH2B3 (linear regression $P \geq 5.62 \times 10^{-20}$) and NF1 ($P \geq 1.32 \times 10^{-56}$) in blood revealed risk alleles to be associated with decreased gene expression for both loci, consistent with the role of these genes in tumor development.

Intersections of ccrSNPs with epigenomic marks mapped in endometrial cancer cell lines, uterine tissue, and vaginal tissue found more endometrial cancer ccrSNPs overlapped with these features than ccrSNPs for endometriosis\textsuperscript{17} or schizophrenia\textsuperscript{18}. These findings highlight the relevance of these tissues for functional studies of endometrial cancer biology. Given the established role of estrogen in endometrial carcinogenesis\textsuperscript{32}, it is perhaps not surprising that endometrial cancer ccrSNPs exhibited greater overlap with epigenomic features present after estrogen stimulation. However, this finding provides evidence that functional studies of endometrial cancer should be performed under these conditions.

Using LD score regression, we estimated that –28% of the approximately twofold FRR of endometrial cancer could be explained by variants, which can be reliably imputed from OncoArray genotypes. The common endometrial cancer variants identified to date together explain up to 6.8% of the FRR, including 2.7% contributed by the nine additional variants reported here; this may be an overestimate, given that the ORs for the new loci likely include some upwards bias (the so-called winner’s curse). In summary, we have doubled the number of endometrial cancer risk loci, explaining around one quarter (6.9%/28%) of the portion of the FRR attributable to common, readily-imputable SNPs. Furthermore, eQTL analyses have identified candidate causal genes and pathways related to tumor development for follow-up studies that will provide further insight into endometrial cancer biology.

**Methods**

**Study samples.** Analyses were based on 13 studies of endometrial cancer, of which four studies contributed case samples to more than one genotyping project. Data were also included from the E2C2 consortium of 45 separate studies. All
The OncoArray study. The “OncoArray” genotyping chip contains 533,631 variants, around half of which were selected to provide a “GWAS backbone,” with the remaining variants selected on the basis of prior evidence of association with cancer or a cancer-related trait. The OncoArray chip was used to genotype 5061 endometrial cancer cases from ten studies in Australia, Belgium, Germany, Sweden, UK, and USA. Genotyping was carried out at two sites: the Cancer Institute Disease Research (CIDR; nine studies) and The University of Melbourne (one study). Details of the genotype calling are given in Amos et al. SNP-wise QC was conducted using genotype data from all consortia participating in the OncoArray experiment\(^1\), SNPs with call rate <95% in any of the consortia, SNPs not in Hardy–Weinberg equilibrium (HWE) (P < 10\(^{-7}\) in controls and P < 10\(^{-12}\) in cases) and SNPs with concordance <98% among 5280 duplicate pairs of samples were excluded, leaving 483,972 SNPs. Prior to imputation, SNPs with minor allele frequency (MAF) <1% and call rate <98% in any consortium were also excluded, as were SNPs that could not be linked to the 1000 Genomes Project reference panel or for which the MAF differed significantly from the European reference panel frequency. A further 1128 SNPs were excluded after review of cluster plots, hence 469,364 SNPs were used in the imputation. The 5061 OncoArray-genotyped endometrial cancer cases were country-matched to controls who had been genotyped in an identical process as part of the Breast Cancer Association Consortium\(^2\). Samples with call rate <95%, with excessively low or high heterozygosity or with an estimated proportion of European ancestry <80% (based on a principal components analysis of 2318 informative markers and with reference to the HapMap populations) were excluded, as were suspected males and individuals who were XO or XXY.

Duplicates and close relatives were identified from estimated genomic kinship matrices. Pairwise comparisons were made among all samples genotyped as part of the OncoArray, iCOGS, or ANECS/SEARCH/NSECG GWASs genotyping projects. Where pairs of duplicates or close relatives were identified between projects, the sample with the more recent genotyping was retained, hence the numbers of cases included were from the ANECS/SEARCH/NSECG GWASs and iCOGS projects are lower than in the original publications. For case–control pairs from within the same project, the case was preferentially retained, and for case–case or control–control pairs, the sample with the higher call rate was used. Following these exclusions, OncoArray genotypes from 4710 cases and 19,438 controls were included in the analyses.

All OncoArray samples (along with all samples from the ANECS/SEARCH/NSECG GWASs and iCOGS project) were imputed using the October 2014 1000 Genomes Project v3 reference panel or for which the MAF differed significantly from the European reference panel frequency. A further 1128 SNPs were excluded after review of cluster plots, hence 469,364 SNPs were used in the imputation. The 288 cases from six population-based case–control studies within the Women’s Health Initiative were genotyped using five different arrays (Supplementary Data 1) and were each separately imputed using the combined 1000 Genome Project v3 and UK10K reference panels using “minimac2” software\(^9\), following standard quality measures and the exclusion of SNPs with a MAF <1%. Five controls for each case were selected randomly, matched on study. Data were also included from the first phase of UK Biobank genotyping, comprising the UK Cancer Registry–confirmed endometrial cancer cases (as of October 2016) and 62,853 cancer-free female controls. Samples were genotyped using Affymetrix UK BiLEVE Axiom array and Affymetrix UK Biobank Axiom array and imputed to the combined 1000 Genome Project v3 and UK10K reference panels using SHAPEIT\(^{10}\) and IMPUTER\(^{41}\).

No analyses to identify duplicates or relatives between samples from the E2C2, WHI, or UK Biobank studies, and any other study were carried out. However, given the sampling frame of these studies, it is very unlikely that there would have been any meaningful sample overlap.

After QC exclusions, the analysis included 12,906 endometrial cancer cases (3613 of which have not been included in any previous publication) and 108,979 controls. Analyses were also carried out specifically for endometrial cancer of endometrioid histology (8758 cases) and endometrial cancer with non-endometrioid histology (1230 cases). Exploratory analyses for specific non-endometrioid histologies (serous carcinoma, carcinoisaroma, clear cell carcinoma, and mucinous carcinoma) included a small number of cases of mixed histotype, where the major component was non-endometrioid. The UK Biobank data did not include information about histology.

All participating studies were approved by research ethics committees from QIMR Berghofer Medical Research Institute, University-Clinic Erlangen, Karolinska Institutet, UZ Leuven, The Mayo Clinic, The Hunter New England Health District, The Regional Committees for Medical and Health Research Ethics Norway, and the UK National Research Ethics Service (04/Q0803/148 and 05/MRE05/1). All participants provided written, informed consent.

Statistical analyses. Per-allele ORs and the s.e. of the logORs were computed using logistic regression for each of the ANECS, SEARCH, NSECG, WHI, and UK Biobank GWASs, for the two E2C2 GWASs and, by country, for the iCOGS and OncoArray studies, giving a total of 17 strata. Case-only analyses were used to assess heterogeneity in SNP effects by histology (endometrioid histology versus non-endometrioid histology). In the OncoArray analysis, potential population stratification was adjusted for using the first nine principal components; these were estimated using data for 33,661 uncorrelated SNPs with MAF >0.05 and pairwise r\(^2\) >0.1 (including 2318 SNPs specifically selected as informative for continental ancestry) using purpose-written software (http://cge.medschl.cam.ac.uk/software/pcalc). Other studies were similarly adjusted for their relevant principal components.

Analyses were carried out using SNPTES\(^{42}\) for the ANECS, SEARCH, and NSECG GWASs, using ProBA\(^{39}\) for the E2C2 GWASs, and using in-house software for the iCOGS, OncoArray, WHI, and UK Biobank studies. We assessed residual population stratification by computing the test statistic inflation.
adjusted to a sample size of 1000 cases and 1000 controls ($\lambda_{1000}$), both overall and with each strata, using 33,278 uncorrelated SNPs ($R^2 < 0.1$). The overall $\lambda_{1000}$ was 1.004, with stratum-specific $\lambda_{1000}$’s between 0.996 and 1.128 (observed for the smallest strata, the German iCOGS dataset; Supplementary fig. 1).

The estimated ORs from the different studies were combined in a fixed-effects inverse-variance weighted meta-analysis using the “meta” software44. For each variant, results from any strata for which the imputation information score was $< 0.4$, the MAF $< 0.005$ or the OR $> 0.333$ were excluded. Following the meta-analysis, SNPs with valid results in fewer than two of the strata, or with between-strata heterogeneity $P < 5 \times 10^{-10}$ were also excluded, leaving 11.7 million SNPs. A random-effects meta-analysis was also carried out.

Using the conventional $5 \times 10^{-8}$ genome-wide significance threshold, all SNPs lying within 500 kb of a significant SNP were initially considered as part of that locus. Approximate conditional analysis in the GCTA program11,19 with an LD reference panel of 4000 OncoArray-controls was then used to look for additional independently associated SNPs within each locus. Only uncorrelated $R^2 < 0.05$ secondary signals were included. The loci lost evidence of significant signals after conditioning on the most strongly associated SNP was the previously published 8q24 locus (Table 1). For each locus, the set of credible causal risk SNPs (csSNPs) was defined as those variants within 500 kb of the most significant SNP and for which the likelihood from the association analysis was no less than one hundredth the likelihood of the most significant SNP (i.e., odds of < 1:100). A BFDP for each significant SNP was estimated on the basis of a maximum plausible OR of 1.5 and a prior probability of association of 0.000149.

The proportion of the FRR of endometrial cancer due to the identified variants was estimated using a log-additive model, where $p_j$, $P_j$, and $\beta$ are the MAF, logOR, and $\log(p_j)$ OR, respectively for variant $j$, and $\lambda = 2$ is the reported FRR of endometrial cancer. The effect estimates used were those estimated in the current study, both for the new loci and for the loci replicated from previous studies.

Proportion FRR = $\frac{1}{\ln(2)} \sum p_j (1 - p_j) (\beta_j^* - \hat{r}^2_j)$.

The proportion of the endometrial cancer FRR that can be explained by all SNPs is given by the frailty-scale heritability, $h^2_F$, divided by $2\ln(2)$. This was estimated using the LD score regression40, based on the full set of meta-analysis summary estimates, restricted to those SNPs present on the MapMarker v3 database with MAF $> 1%$ and imputation quality $R^2 > 0.9$ in the OncoArray imputation using the 1000 Genomes Phase 3 reference panel. The frailty-scale heritability (as opposed to the observed-scale heritability) was obtained by replacing the total sample, $N$, for each study with an effective sample size $N_0$ for SNP $j$, which effectively weights each SNP according to its frequency and the variance of the effect estimate, i.e.,

$N_j = \frac{1}{\ln(2)} \frac{1}{2p_j (1 - p_j)} \hat{r}^2_j$.

Cross-trait LD score regression via the LD Hub interface (28 September 2017, v1.4.1) was used to estimate the genetic correlations between endometrial cancer and 224 traits from 24 analyses41.

Theusal effects of five anthropometric or reproductive factors on the risks of endometrial cancer were estimated using two-sample statistic summary inverse-variance weighted meta-analysis. Instrumental variables for each factor consisted of the most recent set of published GWAS-significant SNPs for that trait; 77 SNPs for body mass index (BMI)49, 47 SNPs for waist:hip ratio50, 81 SNPs for adult height51,52, 54 SNPs for age at natural menopause53, and 368 SNPs for age at menarche. A multivariable MR adjusting for the effects of the 368 menarche SNPs on BMI (a potential mediator) was used to test the effect of menarche on endometrial cancer, not via BMI54.

Cell culture. Ishikawa and JHUEM-14 cells were a gift from Prof PM Pollock (Queensland University of Technology). Cell lines were authenticated using STR profiling and confirmed to be negative for mycoplasma contamination. Ishikawa cells were cultured in Dulbecco’s modified Eagle’s medium (DMEM; Life Technologies #11965-065) with 10% fetal bovine serum (FBS) and antibiotics (100 IU/ml penicillin and 100 g/ml streptomycin). JHUEM-14 cells were cultured in DMEM/F12 medium (Life Technologies #11320-033) with 10% FBS and antibiotics.

Cell fixing and chromatin shearing. Ishikawa and JHUEM-14 cells were plated on to 10-cm tissue culture dishes in phenol red-free DMEM (Sigma-Aldrich #D1145-0250) supplemented with 10% calf serum, 0.5% glucose,雪山d 10% charcoal-dextran-stripped FBS. Cells were incubated in 5% CO2 at 37°C and harvested by scraping with a rubber policeman. Cells were then washed twice with PBS and resuspended in 300 μl of lysis buffer (1% NP-40, 0.5% sodium deoxycholate, 1 mM EDTA, 10 mM Tris-HCl, pH 8.1). The samples were centrifuged at 2000xg at 4°C. Cells were resuspended in 300–750 μl of lysis buffer (1% SDS, 10 mM EDTA, 50 mM Tris-HCl, pH 8.1, with complete protease inhibitor cocktail (Sigma-Aldrich #1183614501)). Chromatin was then sonicated for 10 cycles (10 s) and JHUEM-14 cells for 20 cycles using the highest power setting of a Branson Digital Sonifier SLPI. After chromatin digestion, agarose gel electrophoresis, samples were centrifuged for 10 min at 4°C.

Chromatin immunoprecipitation and sequencing. Samples were diluted 10-fold in 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl (pH 8.1), and 150 mM NaCl with complete protease inhibitor cocktail. Magna ChiP protein A/G magnetic beads (EMD Millipore #16-663) were added to 500 μl of diluted chromatin and incubated with 5 μg of antibody overnight at 4°C. Antibodies against H3K4me1 (Abcam #ab68953), H3K4me3 (Abcam #ab8580), and H3K27ac (Abcam #ab4729) were used (Supplementary Table 1). The next day supernatant was removed and the beads washed three times with the following ice-cold buffers: RIPA 150 (0.1% SDS, 1% Triton X-100, 1 mM EDTA, 50 mM Tris-HCl (pH 8.1), 150 mM NaCl, 0.1% sodium deoxycholate), RIPA 500 (0.1% SDS, 1% Triton X-100, 1 mM EDTA, 50 mM Tris-HCl (pH 8.1), 500 mM NaCl, 0.1% sodium deoxycholate), LiCl RIPA (500 mM LiCl, 1% NP-40, 1% deoxycholate, 1 mM EDTA, 50 mM Tris-HCl (pH 8.1)), and TE buffer. Chromatin was then eluted by incubating beads overnight at 60°C with 100 μl of elution buffer (1% SDS, 100 mM NaHCO3) and 0.5 mg/ml proteinase K. The next day beads were incubated at 95°C for 10 min and supernatant removed. Chromatin was purified using the Help-It Quick Spin kit (QIAEN) and eluted from columns with 50 μl of QIAGEN EB buffer. DNA was quantified using a Qubit dsDNA HS Assay kit (ThermoFischer Scientific).

Samples from two biological replicates for each treatment were sent to the Australian Genome Research Facility (Melbourne, Australia) for library preparation and sequencing (Illumina HiSeq 2500) with 50 bp reads. Mapping and analysis of Chip-seq reads were performed using the ENCODE analysis pipeline, histone Chip-seq Unified Array (GRCh37), with DANexus software tools (https://danexus.com). Replicated peaks across biological replicates were used for downstream analyses.

eQTL analyses. Summary eQTL results for non-cancer tissue were obtained using uterine (N = 70) and vaginal (N = 79) tissue-specific data generated by the Genotype-Tissue Expression Project (GTEx)15, an endometrium eQTL dataset (N = 123) provided by Fung et al.14, and a blood eQTL dataset (males and females; N = 53115).

Data from endometrial cancer tumors and adjacent normal endometrial tissue were accessed from The Cancer Genome Atlas16. Patient germ line SNP genotypes (Affymetrix 6.0 arrays) and tissue expression RNA-seq data were downloaded through the controlled access portal, while epidemiological and tumour genomic copy-number data were downloaded through the Express Access portal. RNA-seq data were aligned and expression quantified to reads per kilobase per million (RPKM) as described in Painter et al.10 and quality control performed on the germ line SNP genotypes as per Carvaljo-Carmona et al.16. Complete genotype, RNA-seq, and copy-number data were available for 277 genetically European patients (218 with endometrioid histology, 29 with adjacent normal tissue).

Germ line genotypes underwent further quality control before imputation to the 1000 Genomes Phase 3v5 reference panel by Eagle v2.3.6, using the Michigan Imputation Server27. Briefly, subjects were removed for genotype missingness $>$10% and SNPs were removed for missingness $>$10%, MAF $<$5%, and HWE P-value $<$5 x $10^{-8}$. SNPs were also removed if they were indels or non-biallelic variants, were ambiguous SNPs with a MAF $>$40%, were not matched to the reference panel, had a MAF difference with the reference panel of $>$20%, or were duplicates.

Genes with a median expression level of 0 RPKM across samples were removed, and the RPKM values of each gene were log2-transformed and samples were quantile normalized. The expression of the genes located within a 2-Mb window surrounding the ccsSNP at each of the newly identified risk loci were extracted from the expression dataset.

The associations between ccsSNPs and gene expression in all endometrial cancer tumor tissues, endometrioid endometrial cancer tissues only, and adjacent normal endometrial tissue, were evaluated using linear regression models using the MatrixEQTL program in R 3.2.5, adjusting for sequencing platform. Tumor tissue expression was also adjusted for copy-number variation, as previously described in Li et al.49. A false discovery rate of $<$0.2% was used to report eQTL results from all datasets, except for the endometrium eQTL dataset where we used a P-value $<$0.01.

Candidate causal gene network analysis. Candidate causal genes identified in our previous studies and from the eQTL results in the current study (Supplementary Table 6) were analyzed using Ingenuity Pathway Analysis (QIAGEN;
accessed on 23 March 2018 and available at www.qiagen.com/ingenuity) to define gene networks and enrichment of genes in canonical signaling pathways.

**Data availability.** OncoArray germ line genotype data for the ECAC studies and E2C2 germ line genotype data have been deposited through the database of Genotypes and Phenotypes (dbGaP; accession number phs000893.v1.p1). Meta-GWAS summary statistics are available from the authors by request. Genotype data for non-cancer controls were provided by the Breast Cancer Association Consortium (BCAC) by application to the BCAC Data Access Coordination Committee (http://bcac.cge.medschl.cam.ac.uk/). ChiP-seq data are available from the Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE113818.

Received: 25 December 2017 Accepted: 2 July 2018
Published online: 09 August 2018

**References**

1. Ferlay, J. et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBCAN 2012. *Int. J. Cancer* **136**, E359–E386 (2015).
2. Win, A. K., Reece, J. C. & Ryan, S. Family history and risk of endometrial cancer: a systematic review and meta-analysis. *Obstet. Gynecol.** **125**, 89–98 (2015).
3. J ohnatty, S. et al. Family history of cancer predicts endometrial cancer risk independently of Lynch syndrome: implications for genetic counselling. *Gynecol. Oncol.* **147**, 381–387 (2017).
4. S pinell, A. B., Bowman, M. A., Shamsani, J. & Kirk, J. Endometrial cancer gene panels: clinical diagnostic vs research germline DNA testing. *Mod. Pathol.* **30**, 1048–1068 (2017).
5. Chen, M. M. et al. GWAS meta-analysis of 16,852 women identifies new susceptibility loci for endometrial cancer. *Hum. Mol. Genet.* **25**, 2612–2620 (2016).
6. Cheng, T. H. et al. Five endometrial cancer risk loci identified through genome-wide association analysis. *Nat. Genet.* **48**, 667–674 (2016).
7. Painter, J. N. et al. Fine-mapping of the HNF1B multicancer locus identifies candidate variants that mediate endometrial cancer risk. *Hum. Mol. Genet.* **24**, 1478–1492 (2015).
8. Thompson, D. J. et al. CYP19A1 fine-mapping and Mendelian randomization: estradiol is causal for endometrial cancer. *Endocr. Relat. Cancer* **23**, 77–91 (2016).
9. Cheng, T. H. et al. Meta-analysis of genome-wide association studies identifies common susceptibility polymorphisms for colorectal and endometrial cancer near SH2B3 and TSHZ1. *Am. J. Hum. Genet.* **98**, 1159–1169 (2016).
10. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
11. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).
12. Cancer Genome Atlas Research Network et al. Integrated genomic characterization of endometrial carcinoma. *Nature* **497**, 67–73 (2013).
13. Fung, J. N. et al. The genetic regulation of transcription in human endometrial tissue. *Hum. Reprod.* **32**, 893–904 (2017).
14. W estra, H. J. et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat. Genet.* **45**, 1238–1243 (2013).
15. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2013).
16. Sapkota, Y. et al. Meta-analysis identifies five novel loci associated with endometriosis highlighting key genes involved in hormone metabolism. *Nat. Commun.* **8**, 15537 (2017).
17. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **471**, 291–295 (2011).
18. B uggess, S., Dudbridge, F. & Thompson, S. G. Combining information on multiple instrumental variables in Mendelian randomization: comparison of allele score and summarized data methods. *Stat. Med.* **35**, 1880–1906 (2016).
19. Locke, A. E. et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206 (2015).
20. Shungin, D. et al. New genetic loci link adipose and insulin biology to body fat distribution. *Nature* **518**, 187–196 (2015).
21. Wood, A. R. et al. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.* **46**, 1173–1186 (2014).
Acknowledgements

We thank the many individuals who participated in this study and the numerous institutions and their staff who supported recruitment, detailed in full in the Supplemental Information. The iCOGS and OncoArray endometrial cancer analysis were supported by NHMRC project grants (ID#103133 and ID#1109286) to A.B.S., D.F.E., A.M.D., D.J.T., and L.T. A.B.S. (APPP16777), P.M.W., and T.A.O. M. (APPP111126) are supported by the NHMRC Fellowship scheme. A.M.D. was supported by the Joseph Mitchell Trust. I.T. is supported by Cancer Research UK and the Oxford Comprehensive Biomedical Research Center. Funding for the iCOGS infrastructure came from: the European Community’s Seventh Framework Programme under grant agreement no. 223175 (HEALTH-F2-2009-223175) (COGS), Cancer Research UK (C1287/A10118, C1287/A10170, C1292/ A11174, C1281/ A12014, C12047/A13834, C12047/A15007, C1497/A16565), the National Institutes of Health (CA128978) and Post-Cancer GWAS initiative (1U19 CA148537, 1U19 CA148065 and 1U19 CA148112—the GAME-ON initiative), the Department of Defence (W81XWH-10-1-0341), the Canadian Institutes of Health Research (CIHR) for the CIHR Team in Familial Risks of Breast Cancer, Komen Foundation for the Cure, the Breast Cancer Research Foundation, and the Ovarian Cancer Research Fund. OncoArray genotyping of ECAC cases was performed with the generous assistance of the Ovarian Cancer Association Consortium (OCAC). We particularly thank the efforts of Cathy Phelan. The OCAC OncoArray genotyping project was funded through grants from the US National Institutes of Health (CA1X01HG007491-01 (Christopher I. Amos), U19-CA148112 (Thomas A. Sellers), R01-CA149429 (Catherine M. Phelan), and R01-CA148059 (Marc T. Goodman); Canadian Institutes of Health Research (MOP-86727 (Linda E. Kelemen)); and the Ovarian Cancer Research Programme of the NCI. The program has been conducted using the UK Biobank Resource under application 2682012000081. OncoArray genotyping of the BCAC controls was funded by Genome Canada Grant GPH-129344, NIHR Grant U19 CA148065, and Cancer UK Grant C1287/A16563. ANECS recruitment was supported by project grants from the NHMRC (ID#339435), the Cancer Council Queensland (ID#4196615), and Cancer Council Tasmania (ID#403031 and ID#457536). SEARCH recruitment was funded by a programme grant from Cancer Research UK (C490/A10214). Stage 1 and stage 2 case genotyping was supported by the NHMRC (ID#552402, ID#103133). Control data were generated by the Wellcome Trust Case Control Consortium (WTCCC), and a full list of the investigators who contributed to the generation of the data is available from the WTCCC website. We acknowledge use of DNA from the British 1958 Birth Cohort collection, funded by the Medical Research Council grant G0000934 and the Wellcome Trust grant 088545/Z/02—funding for this project was provided by the Wellcome Trust under award 085745. NSECG was supported by the EU FP7 CHIRCHA grant, Wellcome Trust Centre for Human Genetics Core Grant 090532/Z/09/Z, and CORGI was funded by Cancer Research UK. We thank Nick Martin, Dale Nyholt, and Anjali Henders for access to GWAS data from QMIR Controls. Recruitment of the QMIR controls was supported by the NHMRC. The University of Newcastle, the Gladys M Brawn Senior Research Fellowship scheme, the Vincent Fairfax Family Foundation, the Hunter Medical Research Institute, and the Hunter Area Pathology Service provided pathology reports and the costs of establishing the Hunter Community Study. The Bavarian Endometrial Cancer Study (BECS) was partly funded by the ELAN fund of the University of Erlangen. The Hannover-Jena Endometrial Cancer Study was partly supported by the Rudolf Bartling Foundation. The Leuven Endometrium Study (LES) was supported by the Verstel Foundation for endometrial cancer. The Mayo Endometrial Cancer Consortium, Mayo Endometrial Cancer Control, Mayo Endometrial Cancer Projects, Mayo Endometrial Cancer Study, the National Cancer Institute of United States Public Health Service (R01 CA122443, P50 CA15083, P50 CA16393, and GAME-ON the NCIC Post-GWAS Initiative U19 CA148112), the Fred C and Katherine B Andersen Foundation, the Mayo Foundation, and the Ovarian Cancer Research Fund with support of the Smith family, in memory of Kathryn Sladek Smith. MoMuTec received financial support from a Hele Vest Grant, the University of Bergen, Melzer Foundation, The Norwegian Cancer Society (Harald Andersen legat), The Research Council of Norway and Haukeland University Hospital. The Newcastle Endometrial Cancer Study (NECS) acknowledges contributions from the University of Newcastle, The NBN Children’s Cancer Research Group, Ms. Jennie Thomas, and the Hunter Medical Research Institute. RENDOCAS was supported through the regional agreement and clinical trials using the iCOGS and OncoArray control datasets, A.B.S. and T.A.O. M. co-ordinated the iCOGS and OncoArray case cohorts. The Cancer Research UK (C490/A10124) was supported by funding from the Agency for Science, Technology and Research of Singapore (A*STAR), the US National Institutes of Health, and the Susan G. Komen Breast Cancer Foundation. The WHI program is funded by the National Heart, Lung, and Blood Institute, the US National Institutes of Health, and the US Department of Health and Human Services (HHSN268201000046C, HHSN26820100001C, HHSN268201010002C, HHSN268201100003C, and HHSN272101100045C). This work was also funded by NCI U19 CA148065-01. The Nurses’ Health Study (NHS) is supported by the NIH, NIH Grant Number U01 CA116107, P01 CA87969, 1R01 CA44949, 1R01 CA134958, and 2R01 CA82838. We thank the participants and staff of the Nurses’ Health Study for their valuable contributions. We also note the following state cancer registries for their help: AL, AZ, AR, CA, CO, CT, DE, FL, GA, IL, ID, IN, KY, LA, ME, MI, MN, MO, MS, MT, NC, ND, OH, OK, OR, PA, RI, SC, TN, TX, VA, WA, and WY. We assume full responsibility for analyses and interpretation of these data. We also thank Channing Division of Network Medicine, Department of Medicine, Brigham and Women’s Hospital, and Harvard Medical School. Finally, we also acknowledge Patti Soule and the Nurses Health Study. The Canadian Ovarian Cancer Study was supported by NCI, NIH Grant Number R01CA83846. The Fred Hutchinson Cancer Research Center (FHCRC) is supported by NCI, NIH Grant Number ROI1CA105212, RO1 CA87538, RO1 CA75977, RO1 CA86363, RO1 HD23166, R35 CA39797, K05 CA2002, and funds from the Fred Hutchinson Cancer Research Center. The Multietnic Cohort Study (MIEC) is supported by the NIH, NIH Grants Number CA154281, CA178908, and 2R01 CA82838. The California Teachers Study (CTS) is supported by NCI, NIH Grant Number 2R01 CA80838, R01 CA91019, and R01 CA7398, and contract 97-10500 from the California Breast Cancer Fund. The Polish Endometrial Cancer Study (PECS) is supported by the Intramural Research Program of the NCI. The Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial (PLCO) is supported by the Extramural and the Intramural Research Programs of the NCI. The WHI program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services through contracts, HHSN268201600001C, HHSN268201600002C, HHSN268201600003C, and HHSN268201600004C. This manuscript was prepared in collaboration with investigators of the WHI, and has been reviewed and Approved by the Women’s Health Initiative (WHI). WHI investigators are listed at https://www.whi.org/researchers/Documents%20%20Write%20a%20Paper/WHI%20Investigator%20Short%20List.pdf.

The Breast Cancer Association Consortium (BCAC) is funded by Cancer Research UK (C1287/A10118, C1287/A12014). The Ovarian Cancer Association Consortium (OCAC) is supported by a grant from the Ovarian Cancer Research Fund thanks to donations from the family and friends of Kathryn Sladek Smith (PDP/RPCI07), and the UK National Institute of Medical Research, University of Cambridge. This research has been conducted using the UK Biobank Resource under applications 5122 and 9797. We gratefully acknowledge the TCGA endometrial cancer consortium for providing samples, tissues, data processing, and making data and results available. Additional funding for individual control groups is detailed in the Supplementary Information.

Author contributions

T.A.O., D.J.T., A.B.S., and D.F.E. designed the study; T.A.O., D.J.T., D.M.G., and A.B.S. drafted the manuscript; T.A.O. and D.J.T. conducted statistical analyses and genotype imputation; T.A.O. and D.M.G. conducted bioinformatic analyses; T.A.O., J.F., and G.W.M. conducted eQTL analyses; D.M.G. and T.A.O. performed and analyzed ChiP-seq experiments; T.A.O. co-ordinated the iCOGS and OncoArray case genotyping, and associated data management; I.D., J.F., and K.M. co-ordinated quality control and de-identified the iCOGS and OncoArray control datasets; A.B.S. and T.A.O. co-ordinated the ANECS stage 1 genotyping; M.D. and C.S.H. co-ordinated the stage 1 genotyping; I.T. and CHIRCHA funded and implemented the NSECG GWAS; I.T., L.M., M.G., A.I., and S.H. co-ordinated the National Study of Endometrial Cancer Genetics (NSECG), and collation of CORGI control GWAS data; A. B. and D. J. T. co-ordinated the Translational Endometrial Cancer (TAC) study; R.S., M.M.C.E., I.A., and E.G.H. co-ordinated collation of GWAS data for the Hunter Community Study; P.D.P.P., D.F.E., and M.S. co-ordinated Studies of Epilepsy and Risk Factors in Cancer Heredity (SEARCH); M.K.B. provided data.
The authors declare no competing interests.

Competing interests:

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/
Leuven 3000, Belgium. 

Department of Population Health, QIMR Berghofer Medical Research Institute, Brisbane 4006 QLD, Australia.

Department of Health Science Research, Mayo Clinic, Division of Biomedical Statistics and Informatics, Rochester 55905 MN, USA.

Department of Environmental Medicine, Karolinska Institutet, Division of Nutritional Epidemiology, Stockholm 171 77, Sweden.

Department of Epidemiology, Shanghai Cancer Institute, Renji Hospital, Shanghai Jiaotong University School of Medicine, State Key Laboratory of Oncogene and Related Genes, Shanghai, China. These authors contributed equally: Amanda B. Spurdle, Deborah J. Thompson.