medRxiv preprint doi: https://doi.org/10.1101/2020.11.09.20228114; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Genetic analyses of gynecological disease identify genetic relationships between uterine fibroids and endometrial cancer, and a novel endometrial cancer genetic risk region at the *WNT4* 1p36.12 locus

Authors

Pik Fang Kho[1, 2], Sally Mortlock[3], Endometrial Cancer Association Consortium, International Endometriosis Genetics Consortium, Peter A.W. Rogers[4], Dale R. Nyholt[2], Grant W. Montgomery[3], Amanda B. Spurdle[1], Dylan M. Glubb[1*], Tracy A. O'Mara[1*].

Affiliations

[1]Department of Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia.

[2]School of Biomedical Science, Faculty of Health, Queensland University of Technology, Brisbane, Queensland, Australia.

[3]The Institute for Molecular Bioscience, The University of Queensland, Brisbane, Queensland, Australia.

[4]Department of Obstetrics and Gynaecology, Gynaecology Research Centre, Royal Women's Hospital, University of Melbourne, Parkville, Victoria, Australia.

[*]These authors contributed equally to the work.

Corresponding Author

Dr Tracy O'Mara, PhD, Molecular Cancer Epidemiology Group, QIMR Berghofer Medical Research Institute, 300 Herston Road, Brisbane QLD Australia 4006. Phone: +61 7 3362 0389. Email: Tracy.OMara@qimrberghofer.edu.au

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.11.09.20228114; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Declarations

Funding

PFK is supported by an Australian Government Research Training Program PhD Scholarship and QIMR Berghofer Postgraduate Top-Up Scholarship, TAO'M, GWM and ABS are supported by NHMRC Investigator Fellowships (APP1173170, GNT1177194 and APP1177524).

This work was supported by National Health and Medical Research Council (NHMRC) Project Grants (APP1109286, GNT1026033, GNT1105321 and GNT1147846). Funding sources had no role in study design, data curation and analysis, data interpretation, report writing and submission for publication.

Conflict of interest/Competing interests

The authors declare no potential conflicts of interest.

Ethics approval

This work used summary-level GWAS meta-analysis results, and thus ethical approval was not required.

Consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and material (data transparency)

Summary-level GWAS meta-analysis results for PCOS, uterine fibroids and endometrial cancer that support the findings of this study are available at the NHGRI-EBI GWAS Catalog (https://www.ebi.ac.uk/gwas/downloads/summary-statistics). Other data generated during this study are included in this article and its supplementary information files or are available on reasonable request.

Code availability

Not applicable

medRxiv preprint doi: https://doi.org/10.1101/2020.11.09.20228114; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Abstract

Endometriosis, polycystic ovary syndrome (PCOS) and uterine fibroids have been proposed as endometrial cancer risk factors; however, disentangling their relationships with endometrial cancer is complicated due to shared risk factors and comorbidities. Using genome-wide association study (GWAS) data, we explored the relationships between these non-cancerous gynecological diseases and endometrial cancer risk by assessing genetic correlation, causal relationships and shared risk loci. We found significant genetic correlation between endometrial cancer and PCOS, and uterine fibroids. Adjustment for genetically predicted body mass index (a risk factor for PCOS, uterine fibroids and endometrial cancer) substantially attenuated the genetic correlation between endometrial cancer and PCOS but did not affect the correlation with uterine fibroids. Mendelian randomization analyses provided evidence of a causal relationship between only uterine fibroids and endometrial cancer. Gene-based analyses revealed risk regions shared between endometrial cancer and endometriosis, and uterine fibroids. Multi-trait GWAS analysis of endometrial cancer and the genetically correlated gynecological diseases identified a novel genome-wide significant endometrial cancer risk locus at 1p36.12, which replicated in an independent endometrial cancer dataset. Interrogation of functional genomic data at 1p36.12 revealed biologically relevant genes, including *WNT4* which is necessary for the development of the female reproductive system. In summary, our study provides genetic evidence for a causal relationship between uterine fibroids and endometrial cancer. It further provides evidence that the comorbidity of endometrial cancer, PCOS and uterine fibroids may partly be due to shared genetic architecture. Notably, this shared architecture has revealed a novel genome-wide risk locus for endometrial cancer.

Keywords

Endometrial cancer, endometriosis, polycystic ovary syndrome, uterine fibroids, genetic correlation, Mendelian randomization.

medRxiv preprint doi: https://doi.org/10.1101/2020.11.09.20228114; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Introduction

Endometriosis, polycystic ovary syndrome (PCOS) and uterine fibroids are three common non-cancerous gynecological diseases affecting 10-15% (Parasar et al. 2017), 6-9% (Azziz et al. 2011) and 5-69% (Stewart et al. 2017) of women of reproductive age, respectively; however, their prevalence is likely underestimated because of under diagnosis (Agarwal et al. 2019; De La Cruz and Buchanan 2017). Although these non-cancerous gynecological diseases primarily affect premenopausal women and endometrial cancer is largely a postmenopausal malignancy, many risk factors are shared with endometrial cancer (e.g. chronic estrogen exposure, inflammation, insulin resistance and obesity (Harris and Terry 2016; Li et al. 2019; Wise et al. 2016)), suggesting some shared biological relationship.

A number of studies have used observational data to assess associations between the three non-cancerous gynecological diseases and endometrial cancer risk, the findings of which have been heterogeneous (Harris and Terry 2016; Johnatty et al. 2020; Li et al. 2019; Wise et al. 2016). Indeed, the use of observational studies to evaluate these associations could be confounded by: (i) the failure to adequately account for potential confounders that are associated with risk of endometrial cancer and/or gynecological disease e.g. oral contraceptive use; (ii) the reliance of disease status classification on self-reported data which is subject to misclassification bias from asymptomatic undiagnosed cases; (iii) misdiagnosis of early stage endometrial cancer as uterine fibroids due to shared clinical presentation (Wise et al. 2016); (iv) detection bias in cohort studies as a result of an increased surveillance for endometrial cancer among patients with non-cancerous gynecological diseases; and (v) the comorbidity of non-cancerous gynecological diseases (Choi et al. 2017; Johnatty et al. 2020; Matalliotaki et al. 2018; Nagai et al. 2015; Uimari et al. 2011; Wise et al. 2007). Thus, it remains difficult to determine from observational studies the precise nature of the relationships between endometrial cancer and these non-cancerous gynecological diseases.

Genome-wide association study (GWAS) data have demonstrated genetic correlation between endometrial cancer and endometriosis (Masuda et al. 2020; Painter et al. 2018), and uterine fibroids (Masuda et al. 2020), which may partly explain the comorbidities of these diseases; whether these comorbidities are due to causal relationships or shared genetic etiology remains to be explained. In this study, we have used a variety of approaches to analyze GWAS data and elucidate relationships between endometrial cancer and non-

medRxiv preprint doi: https://doi.org/10.1101/2020.11.09.20228114; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

cancerous gynecological disease (summarized in **Supplementary Figure 1**). Firstly, we have performed genetic correlation analysis, using the largest currently available datasets to clarify the shared genetic risk between the non-cancerous gynecological diseases and endometrial cancer. As inherited genetic variants are less influenced by confounding inherent in observational studies, we have performed genetic causal inference analyses using gynecological disease-associated variants to investigate causal relationships. It is possible that these diseases may not be genetically correlated or causally related to endometrial cancer but share overlapping genetic risk regions. To assess this possibility, we have performed gene-based association analyses. Lastly, we have performed multi-trait GWAS, leveraging genetic correlation between endometrial cancer and non-cancerous gynecological diseases to discover novel GWAS risk loci.

Materials and Methods

GWAS data

GWAS summary data publicly available for PCOS (Day et al. 2018) (https://doi.org/10.17863/CAM.27720) and uterine fibroids (Gallagher et al. 2019) (ftp://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/GallagherCS_31649266_GCST009158), and via collaboration for endometriosis (Sapkota et al. 2017), were used for all analyses except Mendelian randomization. For PCOS and uterine fibroids, GWAS summary data from the 23andMe, Inc., cohort had been excluded because of restrictions related to data sharing agreements (Day et al. 2018; Gallagher et al. 2019). For Mendelian randomization analyses, risk estimates and respective standard errors of genome-wide significant variants were accessed from the largest published GWAS for each disease (Day et al. 2018; Gallagher et al. 2019; Rahmioglu et al. 2018). Details of studies and sample sizes used in each analysis are shown in **Supplementary Table 1.** Detailed descriptions of the quality control procedures and GWAS analysis can be found in the corresponding publications.

GWAS summary data for endometrial cancer were available from O'Mara et al. (2018). As the GWAS for endometrial cancer (O'Mara et al. 2018), endometriosis (Rahmioglu et al. 2018), and uterine fibroids (Gallagher et al. 2019) included participants from the UK Biobank (https://www.ukbiobank.ac.uk/), we re-analyzed the endometrial cancer dataset, excluding these participants to avoid sample overlap bias in the two sample Mendelian randomization analysis. This also allowed us to use the UK Biobank endometrial cancer dataset as part of

medRxiv preprint doi: https://doi.org/10.1101/2020.11.09.20228114; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

the replication set to confirm multi-trait GWAS results. This revised endometrial cancer GWAS meta-analysis consisted of 12,270 cases and 46,126 controls of European descent. Genetic variants with minor allele frequency (MAF) < 1% and imputation information score < 0.4 were excluded, leaving ~9 million genetic variants. The revised endometrial cancer GWAS data were used only in Mendelian randomization and multi-trait GWAS analyses, while the published endometrial cancer GWAS data (O'Mara et al. 2018) were used in all other analyses. Prior to genetic correlation analysis, genetic variants in the extended human major histocompatibility complex region (26–34 Mb on chromosome 6) were removed due to the complex linkage disequilibrium (LD) structure in this region.

Genetic correlation between endometrial cancer and non-cancerous gynecological diseases

We used LD Score regression (Bulik-Sullivan et al. 2015) to estimate the genetic correlation between endometrial cancer and each non-cancerous gynecological disease. Genetic correlation analyses were restricted to common HapMap3 variants (MAF > 0.01). To reduce bias from potential residual confounding in genetic correlation analyses, including bias from unknown sample overlap, we used the estimated genetic covariance intercept, obtained without constraint. Genetic correlation values range from -1 to 1; positive values indicated that shared genetic variants have concordant effects across the genome, whereas negative values indicated divergent effects.

Obesity is a major risk factor for endometrial cancer, and is prevalent amongst women with PCOS and uterine fibroids (Ilaria and Marci 2018; Sam 2007). For genetic correlation analysis between endometrial cancer and PCOS or uterine fibroids, we thus additionally corrected for the effect of obesity, as measured by genetically predicted BMI. PCOS and uterine fibroids GWAS were conditioned using summary data from a large GWAS of BMI (Yengo et al. 2018) in GCTA-mtCOJO analysis (Zhu et al. 2018) before performing LD score regression analysis.

Genetic causal inference tests

We performed two-sample Mendelian randomization analysis to explore potential causal relationships between non-cancerous gynecological diseases and endometrial cancer. Independent (LD $r^2 < 0.01$) genetic variants associated with the non-cancerous gynecological diseases at genome wide significance ($P < 5 \times 10^{-8}$) were used as genetic instruments. The list of genetic instruments and the respective risk association estimates were extracted from the

medRxiv preprint doi: https://doi.org/10.1101/2020.11.09.20228114; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

largest GWAS of endometriosis (Rahmioglu et al. 2018), PCOS (Day et al. 2018) and uterine fibroids (Gallagher et al. 2019). We excluded independent genetic variants with ambiguous alleles and intermediate frequencies (i.e., variants with A/T or C/G alleles and minor allele frequency of more than 0.42), leaving 26 variants as genetic instruments for endometriosis, 14 for PCOS and 23 for uterine fibroids.

As the three non-cancerous gynecological diseases mostly affect premenopausal women and endometrial cancer primarily affects postmenopausal women, we performed a unidirectional Mendelian randomization analysis, assessing the effect of genetic predisposition to non-cancerous gynecological disease on endometrial cancer risk. We used inverse variance weighted (IVW) analysis as the primary analysis by regressing the genetic variant-endometrial cancer association on the genetic variant-non-cancerous gynecological disease association, weighted by inverse of their variance. This method has the most power to detect associations although it has a strong assumption of no heterogeneity (potentially resulting from pleiotropy) amongst genetic variants (Hemani et al. 2018); thus, this method assumes all genetic variants for the exposure of interest have a proportional effect on outcome risk.

We also performed several sensitivity analyses for Mendelian randomization that are more robust to heterogeneity amongst genetic variants: MR-Egger, weighted median, and weighted mode analysis. MR-Egger analysis regresses the genetic variant-outcome association on the genetic variant-exposure association, without constraining the regression intercept (Bowden et al. 2015). If the MR-Egger regression intercept is non-zero, it provides evidence that directional horizontal pleiotropy amongst genetic variants is driving the causal estimates (i.e. genetic variation influences the outcome through a pathway other than the exposure and indicates that the ratio of genetic variants with positive and negative pleiotropic effects is not balanced). The MR-Egger regression slope represents a valid effect estimate after adjustment for pleiotropic effects, provided the Instrument Strength Independent of Direct Effect (InSIDE) assumption is met (i.e. the association of a genetic variant with the exposure of interest is independent from its direct effect on outcome) (Bowden et al. 2015). We also performed weighted median (Bowden et al. 2016) and weighted mode (Hartwig et al. 2017) analyses, which are more robust to violation of the InSIDE assumption. Weighted median analysis relies on the assumption that more than 50% of the weights come from valid genetic instruments (Bowden et al. 2016), while weighted mode analysis relies on the assumption that most of the weights come from valid genetic instruments (Hartwig et al. 2017).

medRxiv preprint doi: https://doi.org/10.1101/2020.11.09.20228114; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Cochran's Q statistic was used to assess the heterogeneity in the effects of variants (a potential indicator of horizontal pleiotropy) (Bowden et al. 2018), and leave-one-out analysis was used to assess whether a single variant drives the causal association (Hemani et al. 2018).

Two-sample Mendelian randomization analysis was performed using the "TwoSampleMR" (Hemani et al. 2018) package in R. Unless stated otherwise, results with a Bonferroni-corrected p-value for testing the effects of the three non-cancerous gynecological diseases ($P < 0.05/3 = 0.017$) on endometrial cancer risk were considered statistically significant.

Gene-based association analysis

To identify genetic risk regions shared between the non-cancerous gynecological diseases and endometrial cancer, we performed gene-based analysis using the fast and flexible set-based association test (fastBAT) (Bakshi et al. 2016). fastBAT was used to perform an enrichment analysis on GWAS risk variants, located within 50kb of gene regions, for the non-cancerous gynecological cancers and endometrial cancer. A random sample of 10,000 unrelated participants from the UK Biobank was used as the reference panel in these analyses. We applied a false discovery rate (FDR) < 0.05 for the gene-based analysis, and adjacent risk-associated genes were considered a single risk region if within 1 Mb of each other.

Multi-trait analysis of GWAS (MTAG)

MTAG (Turley et al. 2018) was used to improve endometrial cancer risk loci discovery through joint analysis of endometrial cancer and non-cancerous gynecological diseases that showed evidence of genetic correlation with endometrial cancer (i.e. PCOS and uterine fibroids). GWAS summary statistics were used as input and bivariate LD score regression was used to account for sample overlap. Using pre-computed LD scores for Europeans, MTAG analysis was performed on common variants (MAF > 0.01). Alleles of genetic variants were aligned across GWAS, and only variants present across included studies were assessed by MTAG. The final number of included variants for MTAG was 4,734,443. Summary statistics were produced for each trait where effect sizes and standard error estimates could be interpreted as the output from a single-trait GWAS.

Replication of novel endometrial cancer GWAS risk loci

medRxiv preprint doi: https://doi.org/10.1101/2020.11.09.20228114; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

MTAG assumes that the variance-covariance matrix across traits is homogenous across the genome, but it is likely some variants are null for one trait and not null for another trait(s). Violation of this assumption could increase false positive discovery in MTAG (discussed in (Turley et al. 2018)). To address this issue, we tested the replication of novel genome-wide significant endometrial cancer risk variants from MTAG in an independent GWAS meta-analysis using data from the Finnish Biobank Study (FinnGen; https://www.finngen.fi/en) and the UK Biobank. Endometrial cancer GWAS summary statistics for 566 cases and 75,822 controls were downloaded directly from FinnGen (data freeze 2; http://r2.finngen.fi/). Quality control procedures for the FinnGen GWAS data are described in https://finngen.gitbook.io/documentation/. For UK Biobank, we performed an endometrial cancer GWAS using genotype and phenotype data obtained under the application number 25331. Endometrial cancer cases were defined based on ICD10 code (C54) in the data fields of 40006, 41270 and 41202. Controls were selected randomly from unrelated women participants (π < 0.1) with no history of any cancers. GWAS was performed on 1,866 cases and 18,660 controls using REGENIE (Mbatchou et al. 2020) to implement a logistic mixed model, adjusting for genotyping array and the top 10 principal components. A genetic relationship matrix was included in the model as a random effect to account for cryptic relatedness and population stratification. As recommended by REGENIE, we excluded genetic variants with MAF < 0.01, minor allele count below 100, genotype missingness above 10% and variants which deviated from Hardy-Weinberg equilibrium (P value < 1×10^{-15}). After quality control exclusion, a total of 9,789,172 SNPs remained in the GWAS analysis.

To create a replication set, the FinnGen and UK Biobank GWAS results were meta-analysed by a fixed-effect inverse variance weighted model using "meta" software in R. Novel endometrial cancer genome-wide significant variants identified by MTAG were considered to have replicated if they had the same effect direction, and a P-value < 0.05 for association in the replication set.

Identification of candidate target genes at the 1p36.12 endometrial cancer risk locus

We used previously generated promoter-associated HiChIP chromatin looping data from endometrial (one immortalized and three tumor) cell lines (O'Mara et al. 2019) to explore potential regulatory interactions between credible causal risk variants and gene promoters at

medRxiv preprint doi: https://doi.org/10.1101/2020.11.09.20228114; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

the 1p36.12 locus. Credible causal risk variants were defined as variants with P-value for association within two orders of magnitude of the lead variant P-value. We also explored the candidate target genes through overlap of credible causal variants with lead cis-eQTLs from GTEx v8 and the Blood eQTL Browser data (Munz et al. 2020).

Results

We found endometrial cancer to be significantly genetically correlated with PCOS (r_G = 0.36, se = 0.12) and uterine fibroids (r_G = 0.24, se = 0.09) but not with endometriosis (**Table 1**). After adjusting for genetically predicted BMI, the genetic correlation between PCOS and endometrial cancer was no longer statistically significant, indicating that the initial genetic correlation was, at least partly, mediated by genetically predicted BMI (**Table 1**). In contrast, there was no material difference in the genetic correlation between uterine fibroids and endometrial cancer after adjusting for genetically predicted BMI (**Table 1**), consistent with a previous observation of no significant differences in BMI for endometrial cancer cases with or without uterine fibroids (Johnatty et al. 2020).

IVW Mendelian randomization analysis for the effects of genetic predisposition to the non-cancerous gynecological diseases on endometrial cancer provided evidence only for uterine fibroids (**Table 2, Figure 1**). Although sensitivity analyses were not statistically significant, the directionality of the associations between uterine fibroids and endometrial cancer were consistent with the IVW result (**Table 2, Figure 1**). The MR-Egger intercept did not significantly differ from zero (**Table 2**) providing no evidence for confounding by directional horizontal pleiotropy amongst genetic instruments. However, Cochran's Q statistics indicated evidence of heterogeneity between causal estimates based on individual variants (Cochran's Q statistics = 42.1, degrees of freedom = 22, P = 6×10^{-3}), suggesting that some variants may be associated with endometrial cancer risk through pathways other than uterine fibroids. Leave-one-out analysis showed that no single variant was driving the causal association revealed by the IVW analysis (**Supplementary Figure 2**).

While genetic correlation analysis assesses the average genetic concordance across the genome for two traits, it does not reveal common genomic regions that harbor trait-associated variation. Further, a lack of evidence for genetic correlation may reflect opposing pleiotropic effects across the genome. Thus, we performed gene-based analyses to identify common risk regions across endometrial cancer and the non-cancerous gynecological diseases. The initial

medRxiv preprint doi: https://doi.org/10.1101/2020.11.09.20228114; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

analysis revealed 24 genetic regions associated with endometrial cancer risk, 28 regions with endometriosis risk and 41 regions with uterine fibroids (**Supplementary Table 2**). No associations with PCOS passed FDR < 0.05, potentially reflecting a lack of power due to the small sample size of this cohort. We found four genetic risk regions (3q21.3, 9p21.3, 15q15.1 and 17q21.32), containing seven shared candidate susceptibility genes, were shared between endometriosis and endometrial cancer (**Table 3**). Three of these regions (9p21.3, 15q15.1 and 17q21.32) have independently been associated with the risks of endometrial cancer (O'Mara et al. 2018) and endometriosis through GWAS (Rahmioglu et al. 2018). The LD of lead risk variants at each gene was compared and only one region (17q21.32) demonstrated evidence of a shared genetic risk signal ($r^2 > 0.9$; **Table 3**). Additionally, we found two genetic risk regions (5p15.33 and 11p13), containing five shared candidate susceptibility genes, were shared between uterine fibroids and endometrial cancer (**Table 3**). 5p15.33 has been associated with uterine fibroids risk through GWAS (Gallagher et al. 2019) while 11p13 has independently associated with uterine fibroids and endometrial cancer risk in GWAS (Gallagher et al. 2019; O'Mara et al. 2018). The LD of lead risk variants at each gene was compared but there was no strong genetic correlation at either 5p15.33 or 11p13 ($r^2 \leq 0.4$; **Table 3**), suggesting that the genetic risk signals may be independent.

Incorporation of the two gynecological diseases genetically correlated with endometrial cancer (uterine fibroids and PCOS) in MTAG revealed ten genome-wide significant risk loci for endometrial cancer (**Table 4, Figure 2**). We observed an inflation of median test statistics in the MTAG result ($\lambda = 1.06$), which was likely due to a polygenic signal (LD score regression intercept = 0.98, se = 0.01) rather than population stratification. Two of the risk loci (5p15.33 and 1p36.12) were novel endometrial cancer genome-wide risk loci. We assessed both these risk loci in an independent endometrial cancer dataset and found that only the association at the 1p36.12 locus replicated (Table 4).

To identify candidate target genes at the replicated novel endometrial cancer GWAS risk locus (1p36.12), we intersected candidate causal variants with promoter-associated chromatin loops from four endometrial (immortalized and tumor) cell lines (O'Mara et al. 2019). We identified six candidate target genes through chromatin looping, including *WNT4* for which a candidate causal risk variant was revealed as a lead eQTL in lung tissue (**Supplementary Tables 3 and 4; Figure 3**). Additionally, we identified *CDC42* as a candidate target gene

medRxiv preprint doi: https://doi.org/10.1101/2020.11.09.20228114; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

through a candidate causal risk variant located in a chromatin looping anchor at its promoter (**Supplementary Table 3; Figure 3**). Furthermore, candidate causal risk variants were lead eQTLs for *CDC42* in peripheral blood (Westra et al. 2013) (**Supplementary Table 4)**, providing additional evidence for regulatory targeting.

Discussion

Using large-scale genome-wide datasets, we observed evidence of positive genetic correlation between endometrial cancer and PCOS, and uterine fibroids, but not endometriosis. The observed genetic correlation between endometrial cancer and PCOS was at least partly mediated by genetically predicted BMI, consistent with the role of BMI as a risk factor for both PCOS and endometrial cancer. Mendelian randomization analysis provided evidence for a causal relationship only between genetic predisposition to uterine fibroids and endometrial cancer risk. Gene-based analyses revealed several genetic risk regions shared between endometrial cancer and endometriosis, and uterine fibroids. This included one apparent joint genetic risk signal, for endometrial cancer and endometriosis at 17q21.32. Multi-trait GWAS analysis, including endometrial cancer and the genetically correlated gynecological diseases identified two novel genome-wide significant risk loci for endometrial cancer, one of which (1p36.12) replicated in an independent endometrial cancer dataset. Lastly, functional analyses highlighted *CDC42* and *WNT4* as candidate target genes at the 1p36.12 endometrial cancer risk locus.

Two previous studies have reported a positive genetic correlation between endometriosis and endometrial cancer (Masuda et al. 2020; Painter et al. 2018), but we found no evidence for such genetic correlation. This discrepancy may be related to: i) the smaller sample sets used by the prior studies; ii) the ethnicity studied (Masuda et al. (2020) analyzed a Japanese population); or iii) the different genetic correlation analysis approaches used. For example, unlike Painter et al. (2018), we used an unconstrained LD score regression intercept to account for potential residual confounding, resulting in a conservative estimate of genetic correlation. Indeed, we found the estimated genetic covariance intercept to be significantly different from zero, suggesting the presence of bias from population stratification and/or sample overlap. The null results from the genetic causal inference analyses of endometriosis and endometrial cancer are concordant with observational studies that observed no associations after controlling for ascertainment bias by excluding recent endometriosis diagnosis (Melin et al. 2007; Olson et al. 2002; Rowlands et al. 2011).

medRxiv preprint doi: https://doi.org/10.1101/2020.11.09.20228114; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Although there was minimal evidence for genetic correlation or a causal relationship between endometriosis and endometrial cancer, we identified four shared genetic risk regions, three of which (9p21.3, 15q15.1 and 17q21.32) have independently been associated with risk of both diseases through GWAS (O'Mara et al. 2018; Rahmioglu et al. 2018). Notably, the shared candidate susceptibility genes at 9p21.3 (C*DKN2B-AS1*), 15q15.1 (*BMF*) and 17q21.32 (*CBX1*, *MIR1203*, *SKAP1* and *SNX11)* have been previously identified as candidate target genes at endometrial cancer GWAS risk loci through promoter-associated chromatin looping studies (O'Mara et al. 2019). The remaining shared endometriosis and endometrial cancer risk region at 3q21.3 has not been independently identified by GWAS for either disease and may represent a novel GWAS risk locus for both in future studies. Indeed, this region was recently reported as an endometrial cancer risk region in a cross-cancer GWAS meta-analysis of endometrial, breast, ovarian and prostate cancer (Kar et al. 2020).

We found PCOS and endometrial cancer to be genetically correlated but no association was observed in genetic causal inference analyses, concordant with observational studies that account for the effect of obesity (Fearnley et al. 2010; Zucchetto et al. 2009). These findings are consistent with our observation of substantial attenuation in genetic correlation between PCOS and endometrial cancer after adjusting for genetic components of BMI.

We detected evidence of positive genetic correlation between uterine fibroids and endometrial cancer risk, consistent with observational studies (Fortuny et al. 2009; Rowlands et al. 2011; Wise et al. 2016). IVW Mendelian randomization analysis provided evidence of a causal relationship between genetic predisposition to uterine fibroids and endometrial cancer risk. However, Cochran's Q statistics showed evidence that variants used in the IVW analysis had heterogeneous effects, suggesting that not all variants that increase uterine fibroids risk are expected to increase endometrial cancer risk. Although results from subsequent sensitivity analyses that are robust to the presence of varying levels of pleiotropy were not statistically significant, they showed concordant effect directions with IVW result. It is important to note that the Mendelian randomization sensitivity analyses have lower power to detect causal relationships compared with IVW analysis. As genetic causal inference tests rely on the statistical power of GWAS used, future larger GWAS are required to provide more accurate causal estimates and thus greater confidence with regards to the nature of the relationship between uterine fibroids and endometrial cancer.

medRxiv preprint doi: https://doi.org/10.1101/2020.11.09.20228114; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Gene-based analysis revealed two genetic risk regions (5p15.33 and 11p13) that were shared by endometrial cancer and uterine fibroids. The 11p13 shared risk region has been associated with the risks of uterine fibroids (Gallagher et al. 2019) and endometrial cancer (O'Mara et al. 2018) in GWAS. From the gene-based analysis, we identified *WT1* and *WT1-AS* as candidate susceptibility genes for both uterine fibroids and endometrial cancer at 11p13. Consistent with this finding, we had previously established both genes as candidate targets of endometrial cancer risk GWAS variation through promoter-associated chromatin looping studies (O'Mara et al. 2019) and *WT1* had also been identified through chromatin looping as a candidate target of uterine fibroids risk variants (Rafnar et al. 2018). *WT1* encodes a transcription factor that is essential for urogenital development (reviewed by Roberts (2005)) and in the GTEx database of tissue gene expression it is most highly expressed in the uterus (https://gtexportal.org/home/). These observations suggest that alteration of uterine *WT1* expression by endometrial cancer and uterine fibroids genetic risk variation may affect susceptibility to these diseases.

The 5p15.33 region was found to associate with endometrial cancer risk through both the gene-based analysis and the multi-trait GWAS. However, the multi-trait GWAS association did not replicate in the independent endometrial cancer dataset, with discordant effect directions and non-overlapping confidence intervals. Previously, this region has associated with uterine fibroids risk in a GWAS (Gallagher et al. 2019), with endometrial cancer risk in a candidate locus study (Carvajal-Carmona et al. 2015) and in a cross-cancer GWAS meta-analysis of endometrial cancer and ovarian cancer (Glubb et al. 2021). The gene-based analysis at this region revealed three candidate risk genes that were shared between uterine fibroids and endometrial cancer. The most biologically relevant of these genes is *TERT*, which encodes telomerase reverse transcriptase and maintains chromosomal stability by elongating the telomere (Rubtsova et al. 2012). Relevantly, chromosomes in uterine fibroids (Bonatz et al. 1998; Rogalla et al. 1995) and in endometrial tumors (reviewed by Alnafakh et al. (2019)) have been shown to have shorter telomere length. Indeed, a recent Mendelian randomization study found genetically predicted telomere length to be strongly associated with endometrial cancer risk (Telomeres Mendelian Randomization et al. 2017).

The novel 1p36.12 endometrial cancer risk locus, revealed by the multi-trait GWAS, replicated in the independent endometrial cancer GWAS dataset. Genetic variation at this

medRxiv preprint doi: https://doi.org/10.1101/2020.11.09.20228114; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

region has associated with traits that are genetically correlated or causally related to endometrial cancer (e.g. heel bone mineral density (Morris et al. 2019), body mass index (Pulit et al. 2019), diabetes (Vujkovic et al. 2020), age at menarche (Kichaev et al. 2019) and ovarian cancer (Kuchenbaecker et al. 2015)). Furthermore, genetic variation at 1p36.12 has associated with endometriosis and the lead endometrial cancer risk variant from the multi-trait GWAS also represents a GWAS risk signal for pelvic organ prolapse (Olafsdottir et al. 2020). Promoter-associated chromatin looping data from endometrial cell lines highlighted seven candidate target genes, two of which (*CDC42* and *WNT4*) were supported by candidate causal risk variants that represent eQTLs for these genes. The lead candidate causal risk variant at 1p36.12 (rs3820282) and two candidate causal variants (rs61768001 & rs12037376) have previously been associated with expression of *CDC42* in blood and a long non-coding RNA (*LINC00339*) in blood and the endometrium (Mortlock et al. 2020). Semi-quantitative chromatin looping analysis in an endometrial cancer cell line demonstrated evidence of an interaction between a region containing rs3820282 and a ~15 kb region containing the promoter of *LINC00339* (Powell et al. 2016). However, the quantitative chromatin looping data from the HiChIP analysis of the normal immortalized and tumoral endometrial cell lines (O'Mara et al. 2019)), which also has much greater resolution (Lareau and Aryee 2018), did not provide evidence for a physical interaction between *LINC00339* and candidate causal endometrial cancer risk variants.

LINC00339, *CDC42* and *WNT4* all have biological function relevant to endometrial cancer. *LINC00339* has been found to promote oncogenesis in several different cancer types (Gao et al. 2020; Ye et al. 2020; Zhao et al. 2020), although not specifically endometrial cancer. *CDC42* encodes a small GTPase of the Rho-subfamily that regulates cell cycle, cell-cell adhesion, cell migration and cancer progression (Qadir et al. 2015). Notably, CDC42 binds to PAK6 (encoded by an endometrial cancer GWAS risk candidate target gene (O'Mara et al. 2019)) and this complex, which localizes to cell-cell adhesions, is correlated with epithelial colony escape (Morse et al. 2016). *WNT4* encodes a protein that activates WNT/β-catenin signaling and appears to be crucial for the development of the female reproductive system, including the uterus (reviewed in (Biason-Lauber and Konrad 2008)). Moreover, genes belonging to the WNT/β-catenin pathway are frequently mutated in cancer, including the gene encoding β-catenin which is mutated in ~25% of endometrial tumors (Kandoth et al. 2013). As with *CDC42*, there are also links between *WNT4* and other genes located at endometrial cancer GWAS risk loci, such as *WT1* and *RSPO1* (O'Mara et al. 2018). For

medRxiv preprint doi: https://doi.org/10.1101/2020.11.09.20228114; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

example, in the ovary there is evidence of *WNT4* regulation by proteins encoded by both of these genes (Biason-Lauber 2012; Gao et al. 2014) and RSPO protein activity potentiates WNT signaling (Bugter et al. 2021). There also appears to be some connection between *CDC42* and *WNT4*: both genes have been found to be differentially expressed in the endometrium during the menstrual cycle (Powell et al. 2016).

To reduce confounding inherent in the comorbidity observational studies of endometrial cancer and gynecological disease, prospective studies with long follow-up, large sample sizes and case identification using surgical confirmation would ideally be performed. Nevertheless, our study has demonstrated the utility of genetic causal inference analysis as a cost-effective alternative approach for unravelling relationships while reducing bias from unmeasured confounding. However, a limitation of our study is that the sample size of PCOS GWAS (the largest publicly available) was relatively small, reducing power to identify shared genetic risk regions or a causal relationship between PCOS and endometrial cancer. Consequently, these analyses should be revisited when more genome-wide significant variants are revealed in future PCOS GWAS.

In conclusion, our study has provided insights into the comorbidity of non-cancerous gynecological diseases and endometrial cancer by revealing shared genetic risk architecture, a potential causal relationship between uterine fibroids and endometrial cancer, and shared candidate risk regions and genes. Furthermore, our study has leveraged this shared genetic architecture to identify a novel risk locus for endometrial cancer, uncovering biologically relevant candidate target genes and furthering our understanding of endometrial cancer etiology.

Acknowledgements

This work was conducted using the UK Biobank Resource (application number 25331). We thank the research participants and employees of 23andMe for making this work possible. We thank the participants and investigators of FinnGen study. We thank the many individuals who participated in the Endometrial Cancer Association Consortium and the International Endometriosis Genetics Consortium studies, and the numerous institutions and their staff who supported recruitment. A full list of consortium members and acknowledgements can be found in the Supplementary Note.

medRxiv preprint doi: https://doi.org/10.1101/2020.11.09.20228114; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

References

Agarwal SK et al. (2019) Clinical diagnosis of endometriosis: a call to action Am J Obstet Gynecol 220:354 e351-354 e312 doi:10.1016/j.ajog.2018.12.039

Alnafakh RAA, Adishesh M, Button L, Saretzki G, Hapangama DK (2019) Telomerase and Telomeres in Endometrial Cancer Front Oncol 9:344 doi:10.3389/fonc.2019.00344

Azziz R, Dumesic DA, Goodarzi MO (2011) Polycystic ovary syndrome: an ancient disorder? Fertility and sterility 95:1544-1548 doi:10.1016/j.fertnstert.2010.09.032

Bakshi A, Zhu Z, Vinkhuyzen AA, Hill WD, McRae AF, Visscher PM, Yang J (2016) Fast set-based association analysis using summary data from GWAS identifies novel gene loci for human complex traits Sci Rep 6:32894 doi:10.1038/srep32894

Biason-Lauber A (2012) WNT4, RSPO1, and FOXL2 in sex development Semin Reprod Med 30:387-395 doi:10.1055/s-0032-1324722

Biason-Lauber A, Konrad D (2008) WNT4 and sex development Sex Dev 2:210-218 doi:10.1159/000152037

Bonatz G, Frahm SO, Andreas S, Heidorn K, Jonat W, Parwaresch R (1998) Telomere shortening in uterine leiomyomas American journal of obstetrics and gynecology 179:591-596 doi:10.1016/s0002-9378(98)70050-x

Bowden J, Davey Smith G, Burgess S (2015) Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression Int J Epidemiol 44:512-525 doi:10.1093/ije/dyv080

Bowden J, Davey Smith G, Haycock PC, Burgess S (2016) Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator Genet Epidemiol 40:304-314 doi:10.1002/gepi.21965

Bowden J, Hemani G, Davey Smith G (2018) Invited Commentary: Detecting Individual and Global Horizontal Pleiotropy in Mendelian Randomization-A Job for the Humble Heterogeneity Statistic? Am J Epidemiol 187:2681-2685 doi:10.1093/aje/kwy185

Bugter JM, Fenderico N, Maurice MM (2021) Mutations and mechanisms of WNT pathway tumour suppressors in cancer Nat Rev Cancer 21:5-21 doi:10.1038/s41568-020-00307-z

Bulik-Sullivan B et al. (2015) An atlas of genetic correlations across human diseases and traits Nat Genet 47:1236-1241 doi:10.1038/ng.3406

Carvajal-Carmona LG et al. (2015) Candidate locus analysis of the TERT-CLPTM1L cancer risk region on chromosome 5p15 identifies multiple independent variants associated with endometrial cancer risk Hum Genet 134:231-245 doi:10.1007/s00439-014-1515-4

Choi EJ, Cho SB, Lee SR, Lim YM, Jeong K, Moon HS, Chung H (2017) Comorbidity of gynecological and non-gynecological diseases with adenomyosis and endometriosis Obstetrics & gynecology science 60:579-586 doi:10.5468/ogs.2017.60.6.579

Consortium GT (2013) The Genotype-Tissue Expression (GTEx) project Nat Genet 45:580-585 doi:10.1038/ng.2653

Day F et al. (2018) Large-scale genome-wide meta-analysis of polycystic ovary syndrome suggests shared genetic architecture for different diagnosis criteria PLoS Genet 14:e1007813 doi:10.1371/journal.pgen.1007813

De La Cruz MS, Buchanan EM (2017) Uterine Fibroids: Diagnosis and Treatment Am Fam Physician 95:100-107

Fearnley EJ, Marquart L, Spurdle AB, Weinstein P, Webb PM, Australian Ovarian Cancer Study G, Australian National Endometrial Cancer Study G (2010) Polycystic ovary

medRxiv preprint doi: https://doi.org/10.1101/2020.11.09.20228114; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

syndrome increases the risk of endometrial cancer in women aged less than 50 years: an Australian case-control study Cancer causes & control : CCC 21:2303-2308 doi:10.1007/s10552-010-9658-7

Fortuny J et al. (2009) Risk of endometrial cancer in relation to medical conditions and medication use Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology 18:1448-1456 doi:10.1158/1055-9965.EPI-08-0936

Gallagher CS et al. (2019) Genome-wide association and epidemiological analyses reveal common genetic origins between uterine leiomyomata and endometriosis Nat Commun 10:4857 doi:10.1038/s41467-019-12536-4

Gao F, Zhang J, Wang X, Yang J, Chen D, Huff V, Liu YX (2014) Wt1 functions in ovarian follicle development by regulating granulosa cell differentiation Hum Mol Genet 23:333-341 doi:10.1093/hmg/ddt423

Gao J, Dai C, Yu X, Yin XB, Zhou F (2020) Long noncoding RNA LINC00324 exerts protumorigenic effects on liver cancer stem cells by upregulating fas ligand via PU box binding protein FASEB J 34:5800-5817 doi:10.1096/fj.201902705RR

Glubb DM et al. (2021) Cross-Cancer Genome-Wide Association Study of Endometrial Cancer and Epithelial Ovarian Cancer Identifies Genetic Risk Regions Associated with Risk of Both Cancers Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology 30:217-228 doi:10.1158/1055-9965.EPI-20-0739

Harris HR, Terry KL (2016) Polycystic ovary syndrome and risk of endometrial, ovarian, and breast cancer: a systematic review Fertil Res Pract 2:14 doi:10.1186/s40738-016-0029-2

Hartwig FP, Davey Smith G, Bowden J (2017) Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption Int J Epidemiol 46:1985-1998 doi:10.1093/ije/dyx102

Hemani G et al. (2018) The MR-Base platform supports systematic causal inference across the human phenome Elife 7 doi:10.7554/eLife.34408

Ilaria S, Marci R (2018) From obesity to uterine fibroids: an intricate network Curr Med Res Opin 34:1877-1879 doi:10.1080/03007995.2018.1505606

Johnatty SE et al. (2020) Co-existence of leiomyomas, adenomyosis and endometriosis in women with endometrial cancer Scientific reports 10:3621 doi:10.1038/s41598-020-59916-1

Kandoth C et al. (2013) Integrated genomic characterization of endometrial carcinoma Nature 497:67-73 doi:10.1038/nature12113

Kar SP et al. (2020) Combining genome-wide studies of breast, prostate, ovarian and endometrial cancers maps cross-cancer susceptibility loci and identifies new genetic associations bioRxiv

Kichaev G et al. (2019) Leveraging Polygenic Functional Enrichment to Improve GWAS Power Am J Hum Genet 104:65-75 doi:10.1016/j.ajhg.2018.11.008

Kuchenbaecker KB et al. (2015) Identification of six new susceptibility loci for invasive epithelial ovarian cancer Nat Genet 47:164-171 doi:10.1038/ng.3185

Lareau CA, Aryee MJ (2018) hichipper: a preprocessing pipeline for calling DNA loops from HiChIP data Nat Methods 15:155-156 doi:10.1038/nmeth.4583

Li J et al. (2019) Impact of endometriosis on risk of ovarian, endometrial and cervical cancers: a meta-analysis Arch Gynecol Obstet 299:35-46 doi:10.1007/s00404-018-4968-1

medRxiv preprint doi: https://doi.org/10.1101/2020.11.09.20228114; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Masuda T et al. (2020) GWAS of five gynecologic diseases and cross-trait analysis in Japanese Eur J Hum Genet 28:95-107 doi:10.1038/s41431-019-0495-1

Matalliotaki C et al. (2018) Co-existence of benign gynecological tumors with endometriosis in a group of 1,000 women Oncology letters 15:1529-1532 doi:10.3892/ol.2017.7449

Mbatchou J et al. (2020) Computationally efficient whole genome regression for quantitative and binary traits bioRxiv

Melin A, Sparen P, Bergqvist A (2007) The risk of cancer and the role of parity among women with endometriosis Hum Reprod 22:3021-3026 doi:10.1093/humrep/dem209

Morris JA et al. (2019) An atlas of genetic influences on osteoporosis in humans and mice Nat Genet 51:258-266 doi:10.1038/s41588-018-0302-x

Morse EM, Sun X, Olberding JR, Ha BH, Boggon TJ, Calderwood DA (2016) PAK6 targets to cell-cell adhesions through its N-terminus in a Cdc42-dependent manner to drive epithelial colony escape J Cell Sci 129:380-393 doi:10.1242/jcs.177493

Mortlock S et al. (2020) Tissue specific regulation of transcription in endometrium and association with disease Hum Reprod 35:377-393 doi:10.1093/humrep/dez279

Munz M, Wohlers I, Simon E, Reinberger T, Busch H, Schaefer AS, Erdmann J (2020) Qtlizer: comprehensive QTL annotation of GWAS results Sci Rep 10:20417 doi:10.1038/s41598-020-75770-7

Nagai K et al. (2015) Disease history and risk of comorbidity in women's life course: a comprehensive analysis of the Japan Nurses' Health Study baseline survey BMJ open 5:e006360 doi:10.1136/bmjopen-2014-006360

O'Mara TA et al. (2018) Identification of nine new susceptibility loci for endometrial cancer Nat Commun 9:3166 doi:10.1038/s41467-018-05427-7

O'Mara TA, Spurdle AB, Glubb DM, Endometrial Cancer Association C (2019) Analysis of Promoter-Associated Chromatin Interactions Reveals Biologically Relevant Candidate Target Genes at Endometrial Cancer Risk Loci Cancers (Basel) 11 doi:10.3390/cancers11101440

Olafsdottir T et al. (2020) Genome-wide association identifies seven loci for pelvic organ prolapse in Iceland and the UK Biobank Commun Biol 3:129 doi:10.1038/s42003-020-0857-9

Olson JE, Cerhan JR, Janney CA, Anderson KE, Vachon CM, Sellers TA (2002) Postmenopausal cancer risk after self-reported endometriosis diagnosis in the Iowa Women's Health Study Cancer 94:1612-1618

Painter JN et al. (2018) Genetic overlap between endometriosis and endometrial cancer: evidence from cross-disease genetic correlation and GWAS meta-analyses Cancer medicine doi:10.1002/cam4.1445

Parasar P, Ozcan P, Terry KL (2017) Endometriosis: Epidemiology, Diagnosis and Clinical Management Curr Obstet Gynecol Rep 6:34-41 doi:10.1007/s13669-017-0187-1

Powell JE et al. (2016) Endometriosis risk alleles at 1p36.12 act through inverse regulation of CDC42 and LINC00339 Hum Mol Genet 25:5046-5058 doi:10.1093/hmg/ddw320

Pulit SL et al. (2019) Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry Hum Mol Genet 28:166-174 doi:10.1093/hmg/ddy327

Qadir MI, Parveen A, Ali M (2015) Cdc42: Role in Cancer Management Chem Biol Drug Des 86:432-439 doi:10.1111/cbdd.12556

Rafnar T et al. (2018) Variants associating with uterine leiomyoma highlight genetic background shared by various cancers and hormone-related traits Nat Commun 9:3636 doi:10.1038/s41467-018-05428-6

medRxiv preprint doi: https://doi.org/10.1101/2020.11.09.20228114; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Rahmioglu N et al. (2018) Large-scale genome-wide association meta-analysis of endometriosis reveals 13 novel loci and genetically-associated comorbidity with other pain conditions bioRxiv:406967 doi:10.1101/406967

Roberts SG (2005) Transcriptional regulation by WT1 in development Curr Opin Genet Dev 15:542-547 doi:10.1016/j.gde.2005.08.004

Rogalla P, Rohen C, Hennig Y, Deichert U, Bonk U, Bullerdiek J (1995) Telomere repeat fragment sizes do not limit the growth potential of uterine leiomyomas Biochem Biophys Res Commun 211:175-182 doi:10.1006/bbrc.1995.1793

Rowlands IJ, Nagle CM, Spurdle AB, Webb PM, Australian National Endometrial Cancer Study G, Australian Ovarian Cancer Study G (2011) Gynecological conditions and the risk of endometrial cancer Gynecologic oncology 123:537-541 doi:10.1016/j.ygyno.2011.08.022

Rubtsova MP, Vasilkova DP, Malyavko AN, Naraikina YV, Zvereva MI, Dontsova OA (2012) Telomere lengthening and other functions of telomerase Acta Naturae 4:44-61

Sam S (2007) Obesity and Polycystic Ovary Syndrome Obes Manag 3:69-73 doi:10.1089/obe.2007.0019

Sapkota Y et al. (2017) Meta-analysis identifies five novel loci associated with endometriosis highlighting key genes involved in hormone metabolism Nat Commun 8:15539 doi:10.1038/ncomms15539

Stewart EA, Cookson CL, Gandolfo RA, Schulze-Rath R (2017) Epidemiology of uterine fibroids: a systematic review BJOG 124:1501-1512 doi:10.1111/1471-0528.14640

Telomeres Mendelian Randomization C et al. (2017) Association Between Telomere Length and Risk of Cancer and Non-Neoplastic Diseases: A Mendelian Randomization Study JAMA Oncol 3:636-651 doi:10.1001/jamaoncol.2016.5945

Turley P et al. (2018) Multi-trait analysis of genome-wide association summary statistics using MTAG Nat Genet 50:229-237 doi:10.1038/s41588-017-0009-4

Uimari O, Jarvela I, Ryynanen M (2011) Do symptomatic endometriosis and uterine fibroids appear together? Journal of human reproductive sciences 4:34-38 doi:10.4103/0974-1208.82358

Vujkovic M et al. (2020) Discovery of 318 new risk loci for type 2 diabetes and related vascular outcomes among 1.4 million participants in a multi-ancestry meta-analysis Nat Genet 52:680-691 doi:10.1038/s41588-020-0637-y

Westra HJ et al. (2013) Systematic identification of trans eQTLs as putative drivers of known disease associations Nat Genet 45:1238-1243 doi:10.1038/ng.2756

Wise LA, Palmer JR, Stewart EA, Rosenberg L (2007) Polycystic ovary syndrome and risk of uterine leiomyomata Fertility and sterility 87:1108-1115 doi:10.1016/j.fertnstert.2006.11.012

Wise LA, Sponholtz TR, Rosenberg L, Adams-Campbell LL, Kuohung W, LaValley MP, Palmer JR (2016) History of uterine leiomyoma and risk of endometrial cancer in black women Cancer causes & control : CCC 27:545-552 doi:10.1007/s10552-016-0728-3

Ye H et al. (2020) The SP1-Induced Long Noncoding RNA, LINC00339, Promotes Tumorigenesis in Colorectal Cancer via the miR-378a-3p/MED19 Axis Onco Targets Ther 13:11711-11724 doi:10.2147/OTT.S277254

Yengo L et al. (2018) Meta-analysis of genome-wide association studies for height and body mass index in approximately 700000 individuals of European ancestry Hum Mol Genet 27:3641-3649 doi:10.1093/hmg/ddy271

Zhao H, Xiao H, Lu Y, Liu S, Wang C (2020) Long noncoding RNA LINC00339 promotes the oncogenicity of gastric cancer by regulating SRY-box 9 expression via sponging of microRNA-539 Cell Cycle 19:1143-1157 doi:10.1080/15384101.2020.1749404

medRxiv preprint doi: https://doi.org/10.1101/2020.11.09.20228114; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Zhu Z et al. (2018) Causal associations between risk factors and common diseases inferred from GWAS summary data Nat Commun 9:224 doi:10.1038/s41467-017-02317-2

Zucchetto A et al. (2009) Hormone-related factors and gynecological conditions in relation to endometrial cancer risk Eur J Cancer Prev 18:316-321

Tables

Table 1. Genetic correlation between non-cancerous gynecological diseases and endometrial cancer

Non-cancerous gynecological disease	Covariate	r_G	SE	P-value
Endometriosis	-	-0.02	0.09	0.83
PCOS	-	**0.36**	**0.12**	**1.6×10^{-3}**
PCOS	BMI	0.19	0.14	0.17
Uterine fibroids	-	**0.24**	**0.09**	**5.4×10^{-3}**
Uterine fibroids	BMI	0.23	0.10	0.02

r_G: genetic correlation, SE: standard error. Results passing **Bonferroni correction (P < 0.017) are bolded**

medRxiv preprint doi: https://doi.org/10.1101/2020.11.09.20228114; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Table 2. Genetic causal inference results for effects of non-cancerous gynecological diseases on endometrial cancer

Gynecological disease	Genetic causal inference analysis	Beta	SE	P-value
Endometriosis	IVW	0.09	0.09	0.34
	MR-Egger	0.44	0.30	0.15
	MR-Egger (intercept)	-0.03	0.03	0.23
	Weighted median	0.11	0.08	0.15
	Weighted mode	0.13	0.11	0.26
PCOS	IVW	-0.05	0.04	0.26
	MR-Egger	0.07	0.21	0.75
	MR-Egger (intercept)	-0.02	0.03	0.57
	Weighted median	-0.08	0.06	0.21
	Weighted mode	-0.19	0.12	0.13
Uterine fibroids	IVW	**0.17**	**0.07**	**0.01**
	MR-Egger	0.15	0.15	0.32
	MR-Egger (intercept)	0.00	0.01	0.91
	Weighted median	0.07	0.07	0.35
	Weighted mode	0.01	0.12	0.96

medRxiv preprint doi: https://doi.org/10.1101/2020.11.09.20228114; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Table 3. Shared candidate endometrial cancer and non-cancerous gynecological diseases risk regions

Region	Gene	$TopSNP_{EC}$ (p-value)	$TopSNP_{Gyne}$ (p-value)	LD between TopSNPs (r^2)*	$fastBAT_{EC}$ p-value	$fastBAT_{EC}$ FDR	$fastBAT_{Gyne}$ p-value	$fastBAT_{Gyne}$ FDR
Endometrial cancer and endometriosis								
3q21.3	*RUVBL1*	rs872267 (9.95×10^{-7})	rs4857864 (5.35×10^{-6})	0.47	8.49×10^{-6}	8.01×10^{-3}	8.26×10^{-5}	0.043
9p21.3	*CDKN2B-AS1*	rs568447 (2.68×10^{-6})	rs6475610 (1.73×10^{-9})	1.0×10^{-4}	7.26×10^{-5}	0.036	9.02×10^{-9}	3.55×10^{-5}
15q15.1	*BMF*	rs28371998 (6.81×10^{-9})	rs7183386 (7.23×10^{-6})	0.36	1.64×10^{-7}	5.74×10^{-4}	4.93×10^{-6}	9.35×10^{-3}
17q21.32	*CBX1*	rs7225865 (1.34×10^{-8})	rs10445377 (3.20×10^{-7})	0.99	3.35×10^{-6}	4.32×10^{-3}	2.63×10^{-5}	0.025
17q21.32	*MIR1203*	rs4794505 (9.50×10^{-9})	rs10445377 (3.20×10^{-7})	0.99	5.88×10^{-7}	1.44×10^{-3}	4.93×10^{-6}	9.35×10^{-3}
17q21.32	*SKAP1*	rs882380 (4.66×10^{-9})	rs10445377 (3.20×10^{-7})	0.96	5.80×10^{-6}	5.93×10^{-3}	7.93×10^{-6}	0.013
17q21.32	*SNX11*	rs17681336 (1.17×10^{-8})	rs10445377 (3.20×10^{-7})	0.99	1.55×10^{-6}	3.17×10^{-3}	4.51×10^{-6}	9.35×10^{-3}
Endometrial cancer and uterine fibroids								
5p15.33	*CLPTM1L*	rs2736100 (5.17×10^{-6})	rs72709458 (6.07×10^{-15})	0.28	3.93×10^{-7}	1.20×10^{-3}	1.00×10^{-9}	1.53×10^{-6}

medRxiv preprint doi: https://doi.org/10.1101/2020.11.09.20228114; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Region	Gene	$TopSNP_{EC}$ (p-value)	$TopSNP_{Gyne}$ (p-value)	LD between TopSNPs (r^2)*	$fastBAT_{EC}$ p-value	$fastBAT_{EC}$ FDR	$fastBAT_{Gyne}$ p-value	$fastBAT_{Gyne}$ FDR
5p15.33	*MIR4457*	rs2736100 (5.17×10^{-6})	rs72709458 (6.07×10^{-15})	0.28	1.50×10^{-6}	3.17×10^{-3}	8.01×10^{-11}	1.63×10^{-7}
5p15.33	*TERT*	rs2736100 (5.17×10^{-6})	rs72709458 (6.07×10^{-15})	0.28	1.72×10^{-6}	3.25×10^{-3}	1.50×10^{-11}	4.06×10^{-8}
11p13	*WT1*	rs10835920 (1.33×10^{-8})	rs11031731 (2.04×10^{-21})	0.20	1.86×10^{-5}	0.015	2.64×10^{-16}	3.22×10^{-12}
11p13	*WT1-AS*	rs10835920 (1.33×10^{-8})	rs11031762 (9.95×10^{-14})	0.40	1.91×10^{-6}	3.34×10^{-3}	1.04×10^{-12}	4.24×10^{-9}

TopSNP: lead variant for gene from fastBAT analysis; EC: endometrial cancer; Gyne: non-cancerous gynecological diseases; LD: linkage disequilibrium; FDR: false discovery rate.

*LD was estimated using the EUR 1000Genomes reference panel.

medRxiv preprint doi: https://doi.org/10.1101/2020.11.09.20228114; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Table 4. Genome-wide significant endometrial cancer risk loci identified using MTAG

Region	SNP	EA	OA	EAF	Original GWAS			MTAG			Replication	
					Beta	SE	P-value	Beta	SE	P-value	Beta	P-value
Known risk loci												
6p22.3	rs1740828	A	G	0.49	-0.135	0.018	2.47×10^{-14}	-0.111	0.015	3.23×10^{-13}	-	-
8q24.21	rs10089519[#]	G	A	0.67	0.093	0.017	3.89×10^{-8}	0.096	0.015	7.61×10^{-11}	-	-
8q24.21	rs72724795[#]	G	T	0.12	0.150	0.024	3.40×10^{-10}	0.126	0.021	1.53×10^{-9}	-	-
11p13	rs10835917	C	T	0.36	0.089	0.016	3.99×10^{-8}	0.084	0.014	2.09×10^{-9}	-	-
12q24.12	rs3184504	C	T	0.52	0.099	0.016	3.59×10^{-10}	0.100	0.014	5.61×10^{-13}	-	-
13q22.1	rs7981863	T	C	0.27	-0.154	0.018	9.78×10^{-18}	-0.103	0.016	4.28×10^{-11}	-	-
15q21.2	rs12595627	C	T	0.67	0.120	0.017	1.35×10^{-12}	0.085	0.015	6.73×10^{-9}	-	-
17q12	rs11263761	A	G	0.54	0.141	0.016	2.29×10^{-18}	0.095	0.014	1.49×10^{-11}	-	-
*Novel risk loci**												
1p36.12	rs3820282	T	C	0.17	0.078	0.021	2.66×10^{-4}	0.134	0.018	2.74×10^{-13}	0.12	4.36×10^{-3}
5p15.33	rs7713218	G	A	0.52	-0.077	0.016	3.33×10^{-6}	-0.09	0.014	3.63×10^{-10}	0.02	0.62

EA: effect allele; OA: other allele; EAF: effect allele frequency; SE: standard error; Beta, Se and P-values were estimated from MTAG analysis of endometrial cancer, PCOS and uterine fibroids.
[#]LD between rs10089519 and rs72724795 is 0.02.
*Located >1Mb from known endometrial cancer risk loci.

medRxiv preprint doi: https://doi.org/10.1101/2020.11.09.20228114; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

medRxiv preprint doi: https://doi.org/10.1101/2020.11.09.20228114; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Figure legends

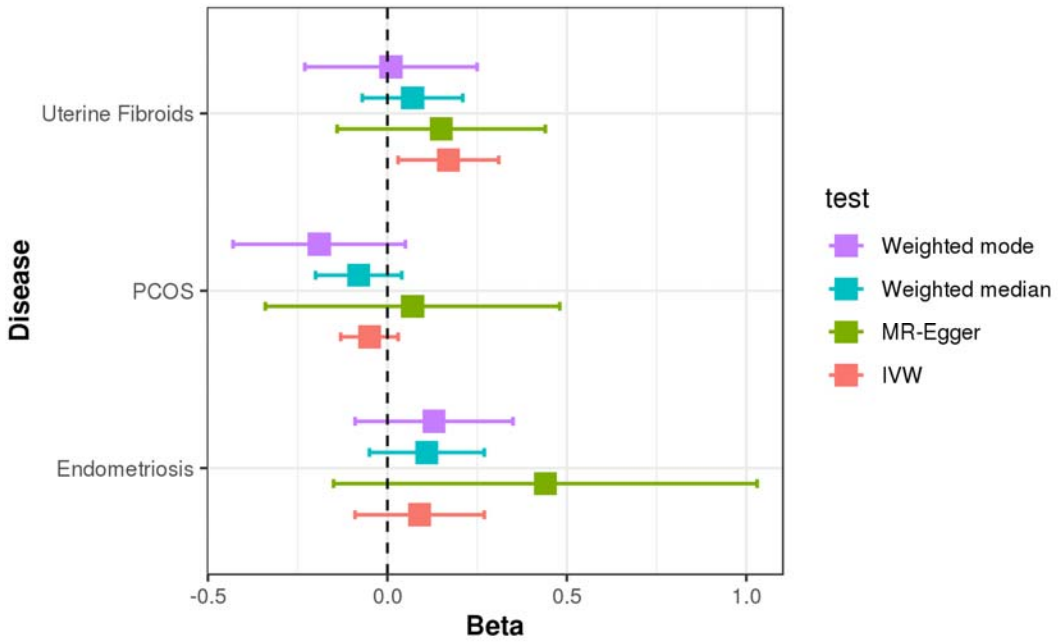


Figure 1. Association between genetic predisposition to non-cancerous gynecological diseases and endometrial cancer, obtained from two-sample Mendelian randomization analysis. The boxes represent the risk of endometrial cancer (beta) per standard deviation increment in genetic predisposition to non-cancerous gynecological disease. Error bars represent 95% confidence intervals.

medRxiv preprint doi: https://doi.org/10.1101/2020.11.09.20228114; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

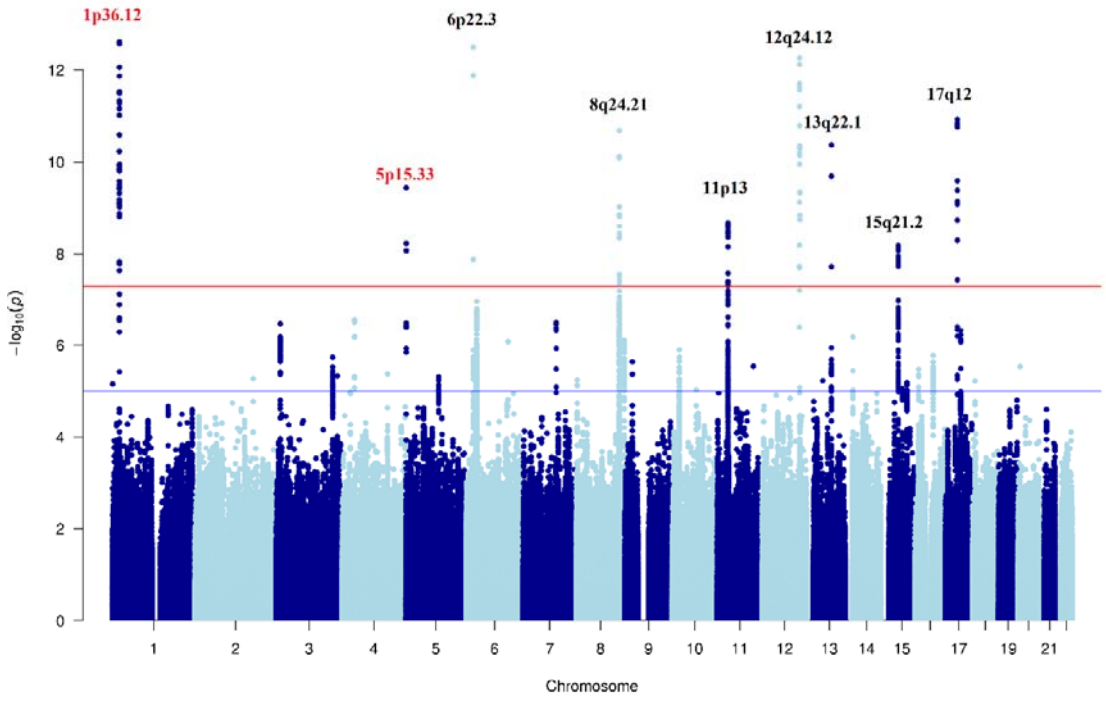


Figure 2. Manhattan plot of MTAG result for endometrial cancer risk. Known endometrial cancer GWAS risk loci are marked in black, and novel genome-wide significant risk loci that are located >1Mb from known endometrial cancer risk loci in red.

medRxiv preprint doi: https://doi.org/10.1101/2020.11.09.20228114; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

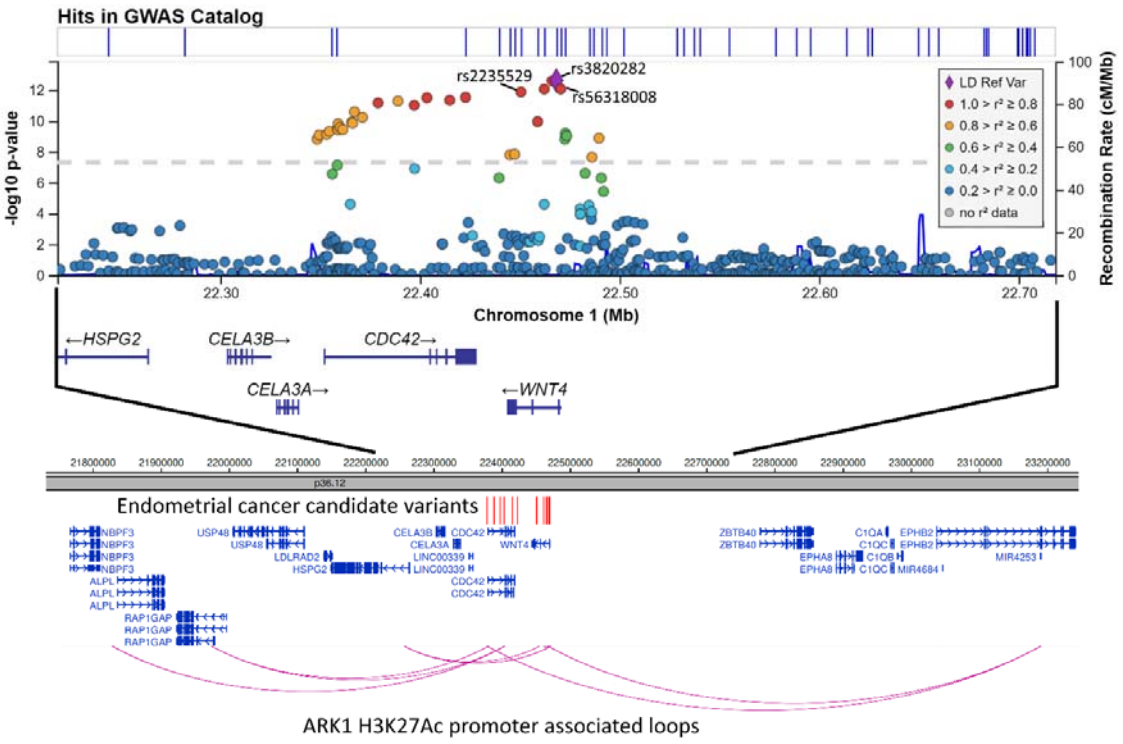
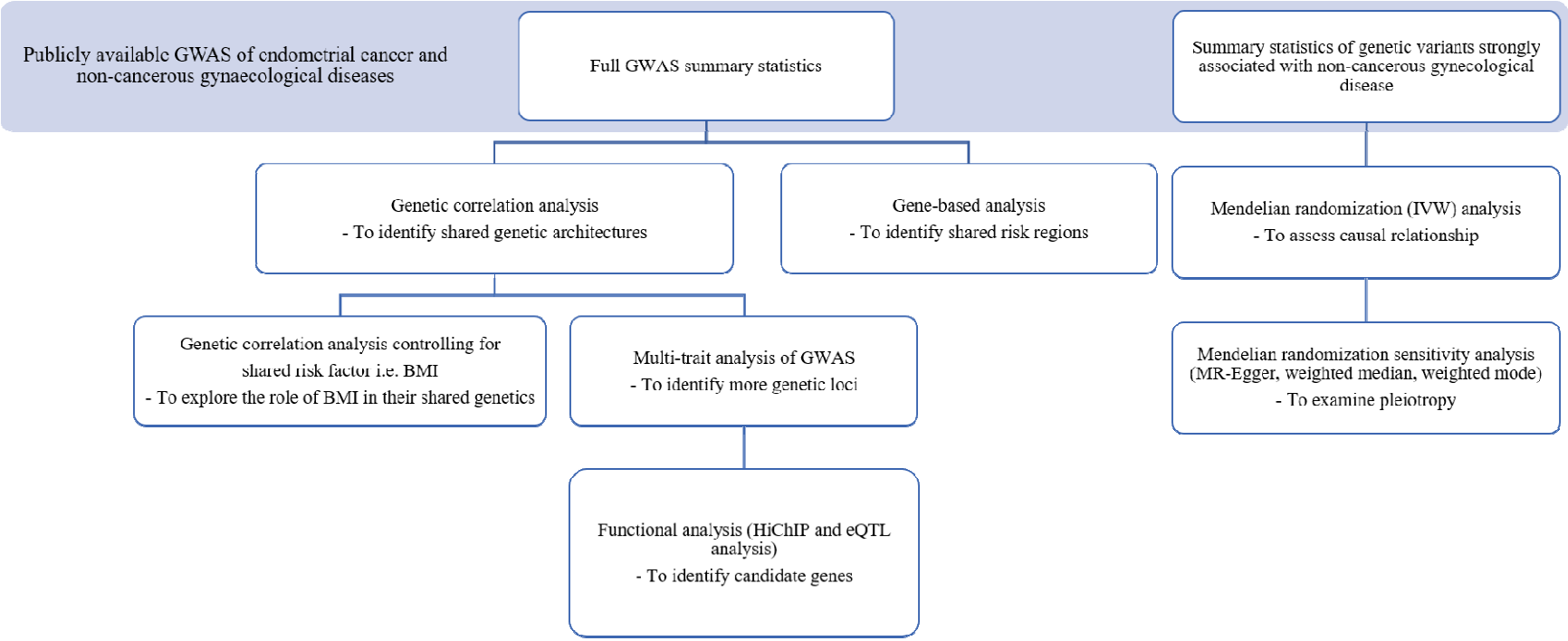


Figure 3. The upper panel depicts a regional association plot for the 1p36.12 novel endometrial cancer risk locus. Genetic variants at the locus are plotted by their genomic position (hg19) and MTAG -$\log_{10}$(P) for association with endometrial cancer risk is on the left y-axis. Recombination rate (cM/Mb) is on the right y-axis and plotted as blue lines. The color of the circles indicates the level of linkage disequilibrium between each variant and the lead variant, rs3820282 (purple diamond), from the 1000 Genomes 2014 EUR reference panel (see legend, inset). The lower panel shows promoter-associated chromatin looping at 1p36.12 identified from HiChIP analysis of the ARK-1 endometrial cancer cell line. Promoter-associated loops that intersect with candidate causal variants (shown as red vertical lines) are shown as purple arcs.

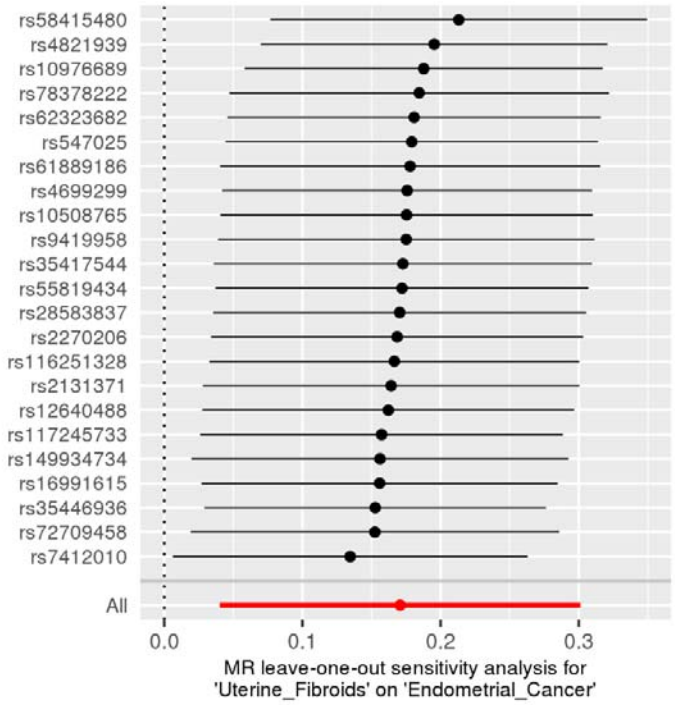
Supplementary Figure



Supplementary Figure 1. Study design to explore the relationships between non-cancerous gynecological diseases and endometrial cancer. Genetic analyses including genome-wide genetic analysis, gene-based analysis, Mendelian randomization analysis, multi-trait analysis of GWAS and GWAS functional analysis were used in this study.

medRxiv preprint doi: https://doi.org/10.1101/2020.11.09.20228114; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

medRxiv preprint doi: https://doi.org/10.1101/2020.11.09.20228114; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .



Supplementary Figure 2. Leave-one-out sensitivity analysis plot for association between genetic predisposition to uterine fibroids and endometrial cancer. Each black dot in the forest plot represents the IVW estimates after excluding the corresponding variant. The plot highlighted in red represents the IVW estimate for all variants.