A genetic perspective on the relationship between non-cancerous gynecological diseases and endometrial cancer

Authors

Pik Fang Kho¹,², Sally Mortlock³, Endometrial Cancer Association Consortium, International Endometriosis Genetics Consortium, Peter A.W. Rogers⁴, Dale R. Nyholt²⁵, Grant W. Montgomery³, Amanda B. Spurdle¹, Dylan M. Glubb¹*, Tracy A. O’Mara¹*.

Affiliations

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

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

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

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

⁵Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, 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NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
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 APP177524).

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.

Conflicts 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 and/or analyzed during this study are included in this article and its supplementary information files or are available on reasonable request.

Code availability
Not applicable
Abstract

The non-cancerous gynecological diseases, endometriosis, polycystic ovary syndrome (PCOS) and uterine fibroids, have been proposed as endometrial cancer risk factors; however, disentangling their relationships is complicated due to their shared risk factors and comorbidity. Using genome-wide association study (GWAS) summary data, we have explored the relationship between these non-cancerous gynecological diseases and endometrial cancer risk, assessing genetic correlations, causal relationships and shared genetic risk regions. Firstly, we found significant genetic correlation between endometrial cancer and PCOS ($r_G = 0.36$, $se = 0.12$, $P = 1.6 \times 10^{-3}$), and uterine fibroids ($r_G = 0.24$, $se = 0.09$, $P = 5.4 \times 10^{-3}$). Adjustment for genetically predicted body mass index (BMI; a risk factor for PCOS, uterine fibroids and endometrial cancer) substantially attenuated the genetic correlation between endometrial cancer and PCOS, but not uterine fibroids. Despite the observed genetic correlation, genetic causal inference tests (latent causal variable and Mendelian randomization analyses) did not support a causal relationship between any of the non-cancerous gynecological diseases and endometrial cancer. Gene-based association analysis revealed four shared endometriosis and endometrial cancer risk loci (9p21.3, 15q15.1, 17q21.32 and 3q21.3) and two shared uterine fibroid and endometrial cancer risk loci (5p15.33 and 11p13). In summary, we have shown that PCOS and uterine fibroids are genetically correlated with endometrial cancer, although the genetic architecture shared between endometrial cancer and PCOS likely relates to BMI. Furthermore, shared genetic risk regions, and thus potentially shared causal genes, were identified between the risk for endometrial cancer and endometriosis, and uterine fibroids.

Keywords

Endometrial cancer, endometriosis, polycystic ovary syndrome, uterine fibroids, genetic correlation, Mendelian randomization, causal relationship, gene-based analysis

Introduction

Endometriosis, polycystic ovary syndrome (PCOS) and uterine fibroids are three prevalent 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, the prevalence of these three diseases is likely to be underestimated because of underdiagnosis (Agarwal et al. 2019; De La Cruz and Buchanan 2017). Although the three non-cancerous gynecological diseases primarily affect premenopausal women and endometrial cancer is largely a postmenopausal malignancy, many risk factors are shared (e.g. chronic estrogen exposure, inflammation, insulin resistance and obesity (Harris and Terry 2016; Li et al. 2019; Wise et al. 2016)), supporting a relationship between them.

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 can be confounded by: (i) the failure to adequately account for potential confounders e.g. oral contraceptive use (a common treatment for endometriosis and PCOS, which is associated with reduced endometrial cancer risk); (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; (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 e.g. patients with uterine fibroids can concurrently suffer from endometriosis (Choi et al. 2017; Johnatty et al. 2020; Matalliotaki et al. 2018; Nagai et al. 2015; Uimari et al. 2011) or PCOS (Wise et al. 2007). Thus, it remains difficult to determine from observational studies if there is a true causal association between these non-cancerous gynecological diseases and endometrial cancer.

Genome-wide association study (GWAS) data have demonstrated genetic overlap 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. However, whether these comorbidities are due to causal relationships or shared genetic etiology remains to be explained. In this study, we have further used GWAS data to disentangle associations between non-cancerous gynecological disease and endometrial cancer. As inherited genetic variants are non-modifiable, genetic approaches to explore...
relationships between phenotypes are less influenced by confounding inherent in observational studies. 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. To investigate causal relationships, we have assessed genetic causal inference through analyses that use genetic variants associated with the gynecological disease of interest as instruments. Lastly, it is possible that these diseases may not be causally related to endometrial cancer but rather develop through effects of disease-specific risk variants at shared risk loci. To assess this possibility, we have performed gene-based analysis of each GWAS to identify shared genetic risk loci.

**Materials and Methods**

**GWAS data**

GWAS summary data were publicly available for PCOS (Day et al. 2018) ([https://doi.org/10.17863/CAM.27720](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](ftp://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/GallagherCS_31649266_GCST009158)). For endometriosis, we accessed data (Sapkota et al. 2017) via collaboration. For PCOS and uterine fibroids, publicly available summary data were based on analyses excluding the 23andMe, Inc., cohort in the original publication, due to data sharing agreements with the company (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. Details of studies and sample sizes used in each analysis are shown in **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 derived 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 UK Biobank, we re-analyzed the endometrial cancer dataset, excluding these participants to avoid sample overlap bias in the two sample Mendelian randomization analysis. 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 filtered, leaving ~9 million genetic variants. This revised endometrial cancer GWAS was used only in Mendelian randomization analysis, while the published endometrial cancer GWAS (O’Mara et al. 2018) was used in all other analyses. Prior to genetic correlation and latent causal variable (LCV) analyses, 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 non-cancerous gynecological diseases and endometrial cancer**

We used LD Score regression (Bulik-Sullivan et al. 2015) to estimate the genetic correlation between each non-cancerous gynecological disease and endometrial cancer. Genetic correlation analyses were restricted to common HapMap3 variants (MAF > 0.01). To reduce bias from potential residual confounding in genetic correlation analyses, including 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.

Evidence of genetic correlation may reflect a causal relationship or sharing of genetic pathways. 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 of PCOS or uterine fibroids with endometrial cancer, 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 conducted LCV analysis using GWAS summary data to estimate the proportion of genetic components for non-cancerous gynecological diseases that also affect endometrial cancer risk, while correcting for heritability and genetic correlation between traits (O’Connor and Price 2018). This method is robust to confounding due to pleiotropy. The genetic causality proportion (GCP) derived from LCV analysis ranges from -1 to 1. GCP values close to 1 indicate that genetic components for non-cancerous gynecological disease may affect endometrial cancer risk, while GCP values close to -1 indicate that genetic components for endometrial cancer may affect non-cancerous gynecological disease risk.

To further explore potential causal relationships, we performed two-sample Mendelian randomization analysis using independent (LD $r^2 < 0.01$) genetic variants associated with the non-cancerous gynecological diseases at genome wide significance ($P < 5 \times 10^{-8}$) as instruments. The list of genetic instruments and the respective risk association estimates were extracted from the largest GWAS of endometriosis (Rahmioglu et al. 2018), PCOS (Day et al. 2018) and uterine fibroids (Gallagher et al. 2019). We filtered out 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 25 for uterine fibroids.

As non-cancerous gynecological diseases mostly affect premenopausal women and endometrial cancer primarily affects postmenopausal women, we performed a unidirectional 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; however 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 which are more robust to heterogeneity amongst genetic variants: MR-Egger, weighted median, and weighted mode analysis. MR-Egger analysis regresses genetic variant-outcome association on genetic variant-exposure association, without a constraint on the regression intercept (Bowden et al. 2015). If the MR-Egger regression intercept is non-zero, it provides evidence of horizontal pleiotropy amongst genetic variants (where a genetic variant belongs to more than one independent pathway that influences outcome risk). The MR-Egger regression slope represents valid effect estimates after controlling for pleiotropic effects, provided the Instrument Strength Independent of Direct Effect (InSIDE) assumption (where a genetic variant’s association with the exposure of interest is independent from its direct effect on outcome) is met (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).

LCV and two-sample Mendelian randomization analysis were performed using the CTG-VL platform (Cuéllar-Partida et al. 2019) and the "TwoSampleMR" (Hemani et al. 2018) package in R, respectively. Unless stated otherwise, results with a Bonferroni-corrected p-value for testing the three non-cancerous gynecological diseases (P < 0.05/3 = 0.017) were considered statistically significant.

**Gene-based analysis**

While genetic correlation analysis assesses the average genetic concordance across the genome for two traits, it does not reveal common local genomic regions that harbor trait-associated variation. Further, a lack of evidence for genetic correlation may reflect opposing pleiotropic effects across the genome. 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 performs enrichment analysis on sets of GWAS variants located within 50kb of gene regions. 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 identified genes were combined into a single locus if within 1 Mb of each other.

Results
We found endometrial cancer was significantly genetically correlated with PCOS ($r_G = 0.36$, $se = 0.12$, $P = 1.6 \times 10^{-3}$) and uterine fibroids ($r_G = 0.24$, $se = 0.09$, $P = 5.4 \times 10^{-3}$) but not with endometriosis ($r_G = -0.02$, $se = 0.09$, $P = 0.83$) (Table 2). After adjusting for genetically predicted BMI, the genetic correlation between PCOS and endometrial cancer decreased and was no longer significant ($r_G = 0.19$, $se = 0.14$, $P = 0.17$), indicating that the initial genetic correlation was, at least partly, mediated by genetically predicted BMI (Table 2). In contrast, there was no material difference in the genetic correlation between uterine fibroids and endometrial cancer after adjusting for genetically predicted BMI (Table 2).

The LCV analysis provided no evidence to suggest genetic components of the non-cancerous gynecological diseases affect endometrial cancer risk (Table 3). Mendelian randomization analyses also provided no evidence of causal relationships between genetic predisposition to the non-cancerous gynecological diseases and endometrial cancer risk (Table 3, Figure 1). No MR-Egger intercepts differed from zero (Table 3), indicating a lack of evidence for confounding by horizontal pleiotropy amongst genetic instruments.

Gene-based analyses by fastBAT revealed 24 genetic regions associated with endometrial cancer risk, 28 regions with endometriosis risk and 41 regions with uterine fibroids, while no associations with PCOS passed FDR < 0.05 (Supplementary Table 1). We found four shared genetic risk regions (3q21.3, 9p21.3, 15q15.1 and 17q21.32) for endometriosis and endometrial cancer, containing seven shared candidate susceptibility genes (Table 4). Three of these regions (9p21.3, 15q15.1 and 17q21.32) have independently been associated with endometrial cancer and endometriosis risk through GWAS (OMara et al. 2018; Rahmioglu et al. 2018). The LD of lead risk variants at each gene was compared and one region (17q21.32) demonstrated evidence of a shared genetic risk signal ($r^2 > 0.9$; Table 4). Additionally, we
found two shared genetic risk regions (5p15.33 and 11p13) for uterine fibroids and endometrial cancer, containing five shared candidate susceptibility genes (Table 5). Through GWAS, one of these regions (5p15.33) has been associated with uterine fibroids risk (Gallagher et al. 2019) and the other (11p13) has independently associated with uterine fibroids and endometrial cancer risk (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 of these regions ($r^2 \leq 0.4$; Table 5), suggesting that these genetic risk signals may be independent.

**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 common risk factor for PCOS and endometrial cancer. Further genetic analyses provided no evidence for a causal relationship between the non-cancerous gynecological diseases and endometrial cancer but revealed several genetic risk regions shared between endometrial cancer and endometriosis, and uterine fibroids.

Previous studies have demonstrated biological links between endometriosis and endometrial cancer (reviewed in Bulun (2009)), but we found no evidence for genetic correlation. However, two previous studies have reported a positive genetic correlation (Masuda et al. 2020; Painter et al. 2018). These discrepancies with our study may be related to: i) the smaller sample sets used by these 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).

Despite the limited evidence for genetic correlation and causal relationships 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). The shared candidate susceptibility genes at 9p21.3 (CDKN2B-AS1), 15q15.1 (BMF) and 17q21.32 (CBXI, MIR1203, SKAP1 and SNX11) have been previously identified as candidate target genes at endometrial cancer GWAS risk loci through chromatin looping studies in endometrial cell lines (O'Mara et al. 2019). The remaining shared endometriosis and endometrial cancer risk region at 3q21.3 has not been previously identified by GWAS and may represent a future GWAS risk locus for both diseases. This region was recently reported as an endometrial cancer risk region in a cross-cancer meta-analysis of endometrial, breast, ovarian and prostate cancer (Kar et al. 2020). However, larger GWAS for endometriosis and endometrial cancer, with more statistical power, will be required to detect genome-wide significant associations at this region.

The null associations identified between PCOS and endometrial cancer risk from genetic causal inference analyses is concordant with observational studies that have found no association after accounting for the effect of obesity (Fearnley et al. 2010; Zucchetto et al. 2009). Indeed, 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.

Although we detected a 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), no evidence for a causal association was found. fastBAT analysis revealed two genetic risk regions (5p15.33 and 11p13) that may be shared by both diseases. The 5p15.33 region has been associated with uterine fibroids in GWAS (Gallagher et al. 2019) and with endometrial cancer in a candidate locus study (Carvajal-Carmona et al. 2015). Three
candidate risk genes were shared between uterine fibroids and endometrial cancer at 5p15.33. The most biologically relevant of these 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. Telomere shortening could induce genomic instability and thus abnormal cellular growth. Indeed, a Mendelian randomization study found genetically predisposed telomere length strongly predicted endometrial cancer risk (Telomeres Mendelian Randomization et al. 2017). The 11p13 shared risk region has been associated with both uterine fibroids (Gallagher et al. 2019) and endometrial cancer (O’Mara et al. 2018) risk in GWAS. WT1 and WT1-AS were revealed as shared candidate susceptibility genes for uterine fibroids and endometrial cancer at this region. We have previously identified WT1 and WT1-AS as candidate targets of endometrial cancer risk GWAS variants through chromatin looping studies (O’Mara et al. 2019). WT1 has 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 is most highly expressed in the uterus (https://gtexportal.org/home/). These observations suggest that alteration of uterine WT1 expression by risk variation associated with endometrial cancer and uterine fibroids may affect susceptibility to these diseases.

Despite the genetic causal inference test analyses providing no evidence for causal relationships, we cannot discard the possibility that non-cancerous gynecological diseases have modest causal effects on endometrial cancer. Nevertheless, our findings indicate that the shared comorbidities are more likely to be related through shared genetic risk elements and thus women who have had non-cancerous gynecological disease may be at a greater risk of developing endometrial cancer. Future research will be needed to clarify such relationships.

Comorbidities may complicate our interpretation of the relationships between the non-cancerous gynecological diseases and endometrial cancer revealed in observational studies. To reduce this confounding, prospective studies with long follow-up, large sample sizes and case identification using surgical confirmation would ideally be required. A strength of our
study is that genetic causal inference analysis is a cost-effective alternative approach that allows assessment of potential causal relationships while reducing bias from unmeasured confounding. Another strength is that we have demonstrated the use of gene-based analysis to detect genetic risk regions below a genome-wide significance threshold. A limitation of the gene-based analysis is that many nearby genes are likely to share GWAS variants, thus the role of the identified candidate susceptibility genes requires functional follow-up. Further, we have limited ability to identify candidate genetic risk regions and causal relationships for PCOS given the smaller dataset available. These analyses should be revisited when more genome-wide significant variants are revealed in future PCOS GWAS.

In conclusion, our study has revealed genetic risk variation shared between non-cancerous gynecological diseases and endometrial cancer, both across the genome and at local genomic regions, providing insights into the shared etiology of these diseases.

Acknowledgement

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 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.

References


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


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

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


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
doi:10.1016/j.fertnstert.2006.11.012
### TABLES

#### Table 1. GWAS data analyzed in this study

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>NCases</th>
<th>NControls</th>
<th>Data available*</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Endometrial cancer(^a)</td>
<td>12,906</td>
<td>108,979</td>
<td>Published summary data</td>
<td>O'Mara et al. (2018)</td>
</tr>
<tr>
<td>Endometriosis</td>
<td>14,926</td>
<td>189,715</td>
<td>Published summary data</td>
<td>Sapkota et al. (2017)</td>
</tr>
<tr>
<td>Endometriosis</td>
<td>58,115</td>
<td>733,480</td>
<td>Only top variant statistics</td>
<td>Rahmioglu et al. (2018)</td>
</tr>
<tr>
<td>PCOS(^b)</td>
<td>4,890</td>
<td>20,405</td>
<td>Published summary data</td>
<td>Day et al. (2018)</td>
</tr>
<tr>
<td>PCOS(^c)</td>
<td>10,074</td>
<td>103,164</td>
<td>Only top variant statistics</td>
<td>Day et al. (2018)</td>
</tr>
<tr>
<td>Uterine fibroids(^b)</td>
<td>20,406</td>
<td>223,918</td>
<td>Published summary data</td>
<td>Gallagher et al. (2019)</td>
</tr>
<tr>
<td>Uterine fibroids(^c)</td>
<td>35,474</td>
<td>267,505</td>
<td>Only top variant statistics</td>
<td>Gallagher et al. (2019)</td>
</tr>
</tbody>
</table>

*Datasets with full summary statistics available were used in genetic correlation, LCV and gene-based analysis. Datasets with only top variant statistics available were used in Mendelian randomization analyses. \(^a\)UK Biobank samples removed from dataset to minimize bias in Mendelian randomization analysis, resulting in 12,270 endometrial cancer cases and 46,126 controls. \(^b\)23andMe samples removed from publicly available dataset due to data sharing issue. \(^c\)23andMe samples included.
Table 2. Genetic correlation results for non-cancerous gynecological diseases and endometrial cancer obtained from LD score regression

<table>
<thead>
<tr>
<th>Non-cancerous gynecological diseases</th>
<th>Covariate</th>
<th>r_G</th>
<th>SE</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Endometriosis</td>
<td>-</td>
<td>-0.02</td>
<td>0.09</td>
<td>0.83</td>
</tr>
<tr>
<td>PCOS</td>
<td>-</td>
<td><strong>0.36</strong></td>
<td>0.12</td>
<td><strong>1.6×10^{-3}</strong></td>
</tr>
<tr>
<td>PCOS</td>
<td>BMI</td>
<td>0.19</td>
<td>0.14</td>
<td>0.17</td>
</tr>
<tr>
<td>Uterine fibroids</td>
<td>-</td>
<td><strong>0.24</strong></td>
<td>0.09</td>
<td><strong>5.4×10^{-3}</strong></td>
</tr>
<tr>
<td>Uterine fibroids</td>
<td>BMI</td>
<td>0.23</td>
<td>0.1</td>
<td>0.02</td>
</tr>
</tbody>
</table>

r_G: genetic correlation, SE: standard error, P: P value. Results passing Bonferroni correction (P < 0.017) are in bold.
Table 3. Genetic causal inference results for non-cancerous gynecological diseases and endometrial cancer obtained from LCV and two-sample Mendelian randomization analyses

<table>
<thead>
<tr>
<th>Trait 1</th>
<th>Trait 2</th>
<th>Mendelian randomization method</th>
<th>Beta</th>
<th>SE</th>
<th>P</th>
<th>GCP</th>
<th>SE</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Endometriosis</td>
<td>Endometrial</td>
<td>IVW</td>
<td>0.09</td>
<td>0.09</td>
<td>0.34</td>
<td>0.06</td>
<td>0.54</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>Cancer</td>
<td>MR-Egger</td>
<td>0.44</td>
<td>0.30</td>
<td>0.15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>MR-Egger (intercept)</td>
<td>-0.03</td>
<td>0.03</td>
<td>0.23</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Weighted median</td>
<td>0.11</td>
<td>0.08</td>
<td>0.14</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Weighted mode</td>
<td>0.13</td>
<td>0.11</td>
<td>0.26</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCOS</td>
<td>Endometrial</td>
<td>IVW</td>
<td>-0.05</td>
<td>0.04</td>
<td>0.26</td>
<td>0.18</td>
<td>0.5</td>
<td>0.78</td>
</tr>
<tr>
<td></td>
<td>Cancer</td>
<td>MR-Egger</td>
<td>0.07</td>
<td>0.21</td>
<td>0.75</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>MR-Egger (intercept)</td>
<td>-0.02</td>
<td>0.03</td>
<td>0.57</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Weighted median</td>
<td>-0.08</td>
<td>0.06</td>
<td>0.22</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Weighted mode</td>
<td>-0.19</td>
<td>0.12</td>
<td>0.13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uterine fibroids</td>
<td>Endometrial</td>
<td>IVW</td>
<td>0.08</td>
<td>0.07</td>
<td>0.22</td>
<td>0.53</td>
<td>0.31</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>Cancer</td>
<td>MR-Egger</td>
<td>0.31</td>
<td>0.15</td>
<td>0.06</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>MR-Egger (intercept)</td>
<td>-0.02</td>
<td>0.01</td>
<td>0.12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Weighted median</td>
<td>0.06</td>
<td>0.07</td>
<td>0.37</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Weighted mode</td>
<td>0.05</td>
<td>0.11</td>
<td>0.66</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Beta: effect size in natural log OR; SE: standard error; P: P-value; GCP: genetic causality proportion
Table 4. Shared candidate risk genes at shared endometrial cancer and endometriosis risk regions

<table>
<thead>
<tr>
<th>Gene</th>
<th>Region</th>
<th>TopSNP&lt;sub&gt;EC&lt;/sub&gt; (p-value)</th>
<th>TopSNP&lt;sub&gt;Endom&lt;/sub&gt; (p-value)</th>
<th>LD between TopSNPs (r&lt;sup&gt;2&lt;/sup&gt;)*</th>
<th>fastBAT&lt;sub&gt;EC&lt;/sub&gt; p-value</th>
<th>fastBAT&lt;sub&gt;EC&lt;/sub&gt; FDR</th>
<th>fastBAT&lt;sub&gt;Endom&lt;/sub&gt; p-value</th>
<th>fastBAT&lt;sub&gt;Endom&lt;/sub&gt; FDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>RUVBL1</td>
<td>3q21.3</td>
<td>rs872267 (9.95E-07)</td>
<td>rs4857864 (5.35E-06)</td>
<td>0.47</td>
<td>8.49E-06</td>
<td>8.01E-03</td>
<td>8.26E-05</td>
<td>4.25E-02</td>
</tr>
<tr>
<td>CDKN2B-AS1</td>
<td>9p21.3</td>
<td>rs568447 (2.68E-06)</td>
<td>rs6475610 (1.73E-09)</td>
<td>1.0E-04</td>
<td>7.26E-05</td>
<td>3.56E-02</td>
<td>9.02E-09</td>
<td>3.55E-05</td>
</tr>
<tr>
<td>BMF</td>
<td>15q15.1</td>
<td>rs28371998 (6.81E-09)</td>
<td>rs7183386 (7.23E-06)</td>
<td>0.36</td>
<td>1.64E-07</td>
<td>5.74E-04</td>
<td>4.93E-06</td>
<td>9.35E-03</td>
</tr>
<tr>
<td>CBX1</td>
<td>17q21.32</td>
<td>rs7225865 (1.34E-08)</td>
<td>rs10445377 (3.20E-07)</td>
<td>0.99</td>
<td>3.35E-06</td>
<td>4.32E-03</td>
<td>2.63E-05</td>
<td>2.48E-02</td>
</tr>
<tr>
<td>MIR1203</td>
<td>17q21.32</td>
<td>rs4794505 (9.50E-09)</td>
<td>rs10445377 (3.20E-07)</td>
<td>0.99</td>
<td>5.88E-07</td>
<td>1.44E-03</td>
<td>4.93E-06</td>
<td>9.35E-03</td>
</tr>
<tr>
<td>SKAPI</td>
<td>17q21.32</td>
<td>rs882380 (4.66E-09)</td>
<td>rs10445377 (3.20E-07)</td>
<td>0.96</td>
<td>5.80E-06</td>
<td>5.93E-03</td>
<td>7.93E-06</td>
<td>1.25E-02</td>
</tr>
<tr>
<td>SNX11</td>
<td>17q21.32</td>
<td>rs17681336 (1.17E-08)</td>
<td>rs10445377 (3.20E-07)</td>
<td>0.99</td>
<td>1.55E-06</td>
<td>3.17E-03</td>
<td>4.51E-06</td>
<td>9.35E-03</td>
</tr>
</tbody>
</table>

TopSNP: lead variant for gene from fastBAT analysis; EC: endometrial cancer; Endom: endometriosis; LD: linkage disequilibrium; FDR: false discovery rate.

*LD was estimated using the EUR 1000Genomes reference panel.
Table 5. Shared candidate risk genes at endometrial cancer and uterine fibroids risk regions

<table>
<thead>
<tr>
<th>Gene</th>
<th>Region</th>
<th>TopSNP&lt;sub&gt;EC&lt;/sub&gt; (p-value)</th>
<th>TopSNP&lt;sub&gt;UF&lt;/sub&gt; (p-value)</th>
<th>LD between TopSNPs (r&lt;sup&gt;2&lt;/sup&gt;)*</th>
<th>fastBAT&lt;sub&gt;EC&lt;/sub&gt; p-value</th>
<th>fastBAT&lt;sub&gt;EC&lt;/sub&gt; FDR</th>
<th>fastBAT&lt;sub&gt;UF&lt;/sub&gt; p-value</th>
<th>fastBAT&lt;sub&gt;UF&lt;/sub&gt; FDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLPTM1L</td>
<td>5p15.33</td>
<td>rs2736100 (5.17E-06)</td>
<td>rs72709458 (6.07E-15)</td>
<td>0.28</td>
<td>3.93E-07</td>
<td>1.20E-03</td>
<td>1.00E-09</td>
<td>1.53E-06</td>
</tr>
<tr>
<td>MIR4457</td>
<td>5p15.33</td>
<td>rs2736100 (5.17E-06)</td>
<td>rs72709458 (6.07E-15)</td>
<td>0.28</td>
<td>1.50E-06</td>
<td>3.17E-03</td>
<td>8.01E-11</td>
<td>1.63E-07</td>
</tr>
<tr>
<td>TERT</td>
<td>5p15.33</td>
<td>rs2736100 (5.17E-06)</td>
<td>rs72709458 (6.07E-15)</td>
<td>0.28</td>
<td>1.72E-06</td>
<td>3.25E-03</td>
<td>1.50E-11</td>
<td>4.06E-08</td>
</tr>
<tr>
<td>WT1</td>
<td>11p13</td>
<td>rs10835920 (1.33E-08)</td>
<td>rs11031731 (2.04E-21)</td>
<td>0.20</td>
<td>1.86E-05</td>
<td>1.52E-02</td>
<td>2.64E-16</td>
<td>3.22E-12</td>
</tr>
<tr>
<td>WT1-AS</td>
<td>11p13</td>
<td>rs10835920 (1.33E-08)</td>
<td>rs11031762 (9.95E-14)</td>
<td>0.40</td>
<td>1.91E-06</td>
<td>3.34E-03</td>
<td>1.04E-12</td>
<td>4.24E-09</td>
</tr>
</tbody>
</table>

TopSNP: lead variant for gene from fastBAT analysis; EC: endometrial cancer; UF: uterine fibroids; LD: linkage disequilibrium; FDR: false discovery rate.

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

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