

RESEARCH ARTICLE

Open Access



# Polygenic risk score phenome-wide association study reveals an association between endometriosis and testosterone

Isabelle M. McGrath<sup>1\*</sup> , International Endometriosis Genetics Consortium, Grant W. Montgomery<sup>1</sup>  and Sally Mortlock<sup>1</sup> 

## Abstract

**Background** Endometriosis affects 1 in 9 women, yet it is poorly understood with long diagnostic delays, invasive diagnoses, and poor treatment outcomes. Characterised by the presence of endometrial-like tissue outside of the uterus, its main symptoms are pain and infertility. Endometriosis often co-occurs with other conditions, which may provide insights into the origins of endometriosis.

**Methods** Here a polygenic risk score phenome-wide association study of endometriosis was conducted in the UK Biobank to investigate the pleiotropic effects of a genetic liability to endometriosis. The relationship between the polygenic risk score for endometriosis and health conditions, blood and urine biomarkers and reproductive factors were investigated separately in females, males and females without an endometriosis diagnosis. The relationship between endometriosis and the blood and urine biomarkers was further investigated using genetic correlation and Mendelian randomisation approaches to identify causal relationships.

**Results** Multiple health conditions, blood and urine biomarkers and reproductive factors were associated with genetic liability to endometriosis in each group, indicating many endometriosis comorbidities are not dependent on the physical manifestation of endometriosis. Differences in the associated traits between males and females highlighted the importance of sex-specific pathways in the overlap of endometriosis with many other traits. Notably, an association of genetic liability to endometriosis with lower testosterone levels was identified. Follow-up analysis utilising Mendelian randomisation approaches suggested lower testosterone may be causal for both endometriosis and clear cell ovarian cancer.

**Conclusions** This study highlights the diversity of the pleiotropic effects of genetic risk to endometriosis irrespective of a diagnosis of endometriosis. A key finding was the identification of a causal effect of the genetic liability to lower testosterone on endometriosis using Mendelian randomisation.

**Keywords** Endometriosis, Testosterone, PheWAS, Mendelian randomisation, Genetic, Ovarian cancer

## Background

Endometriosis is a poorly understood common disease characterised by the growth of endometrial-like tissue outside of the uterus. The diagnostic delay for endometriosis is 7–11 years, which can be attributed to lack of disease awareness, variability in disease presentation, symptoms that overlap other conditions and the invasive

\*Correspondence:

Isabelle M. McGrath  
isabelle.mcgrath@uq.edu.au

<sup>1</sup> The Institute for Molecular Bioscience, The University of Queensland, Brisbane, QLD 4072, Australia



nature of the diagnostic technique: laparoscopic surgery. There is a genetic component to endometriosis, with heritability estimates of 47–51% [1, 2], and the most recent genome-wide association study revealed 42 loci associated with the disease, which explain up to 5.01% of disease variance [3].

Whilst the flagship symptoms of endometriosis are pelvic pain and infertility, it is now appreciated endometriosis patients will often exhibit disturbances to multiple bodily systems beyond the reproductive system. Recently, a wealth of epidemiological data has identified many conditions that are comorbid with endometriosis, multiple of which have evidence of a shared genetic architecture [3–6]. An understanding of the overlapping traits with endometriosis is critical for comprehensive health management of the patient, for developing predictive tools, and for elucidating the underlying biology of endometriosis.

This study expands on previous approaches to characterising the overlap of endometriosis with other traits. A phenome-wide association study (PheWAS) is a technique whereby the association of an array of traits with a particular genetic variant is determined. This approach can be extended to study the association of multiple traits with the genetic liability to a trait of interest through use of the polygenic risk score (PRS) for the trait of interest, a PRS-PheWAS. Compared to analysis with disease status, use of the PRS for cross-trait analysis does not require a cohort with the presence of a trait of interest accurately ascertained. This is advantageous as many endometriosis cases may be undiagnosed and thus be present in control cohorts given the heterogeneity in symptom severity and the invasive and lengthy diagnosis process for endometriosis. Further, as a PRS-PheWAS does not use disease status, it specifically looks for pleiotropic effects of the disease-associated genetic variants, meaning the effects of a genetic liability to endometriosis can be studied in individuals without the disease.

## Methods

A flow diagram summarising the workflow and methods is employed in Fig. 1.

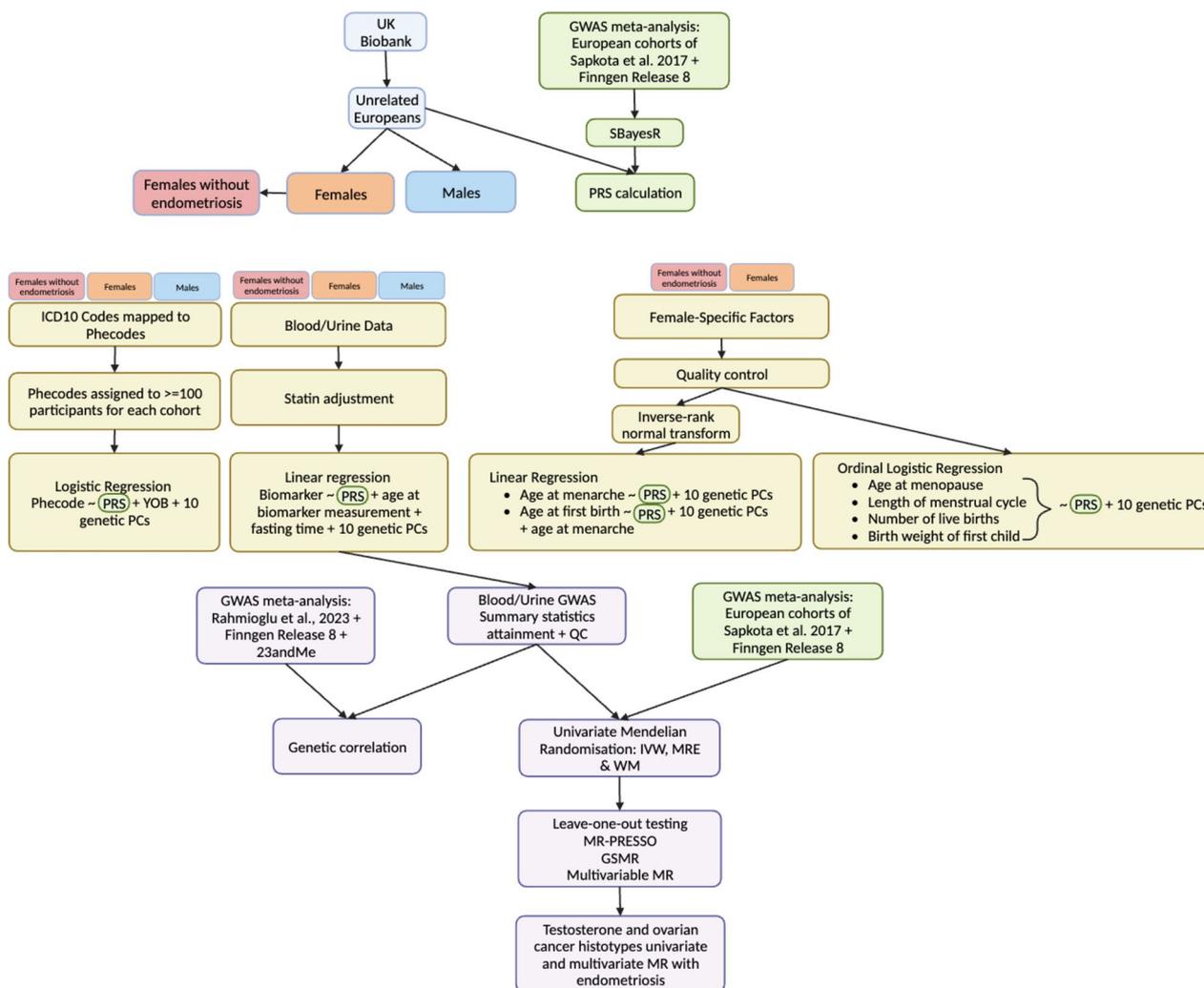
### Phenotype data from UK Biobank

The UK Biobank (UKB) is a large population database containing comprehensive health records and genetic data of approximately 500,000 individuals. The matched phenotype and genotype information allows use of the UKB for a PheWAS. Three groups of phenotype data were utilised: ICD10 diagnostic data, blood and urine biomarker data and female-specific factors. ICD10 diagnostic codes were mapped to phecodes and grouped into categories using the map from

[https://phewascatalog.org/files/Phecode\\_map\\_v1\\_2\\_icd10\\_beta.csv.zip](https://phewascatalog.org/files/Phecode_map_v1_2_icd10_beta.csv.zip) and [https://phewascatalog.org/files/phecode\\_definitions1.2.csv.zip](https://phewascatalog.org/files/phecode_definitions1.2.csv.zip). Where an ICD10 code mapped to multiple phecodes, one combination was selected. Phecode categories included infectious disease, neoplasms, endocrine/metabolic, haematopoietic, mental disorders, neurological, sense organs, circulatory system, respiratory, digestive, genitourinary, dermatologic, musculoskeletal, congenital abnormalities, symptoms, injuries and poisonings and other. Numerical data from the blood biochemistry and urine assays were collated for each participant. In total, there were 34 blood/urine variables available. Blood/urine biomarker data were log transformed prior to PheWAS analysis. The female-specific factors age at menarche, age at menopause, length of menstrual cycle, number of live births, birth weight of first child and age at first birth were analysed in the PheWAS. For some UKB participants, there were multiple datapoints for these female-specific factors due to follow-up visits. For number of live births, the last recorded value was considered to account for any live births proceeding the enrolment visit. For all other female-specific factors, the first recorded value was considered to minimise the recall period. Quality control was conducted to remove extreme values that may reflect pathologies or values with less than 10 participants: age at menarche between 8 and 20 and age at first live birth between 14 and 43 years were considered, menopause before 40 or after 63 years was excluded, menstrual cycle length was restricted to 22–36 days and birth weight of first child between 3 and 12 pounds was considered. Four or more live births were collapsed into a single category.

### Development of endometriosis PRS weightings

Summary statistics from seven European cohorts included in the Sapkota et al. 2017 meta-analysis (14,926 cases; 189,715 controls) [7] of endometriosis were meta-analysed alongside endometriosis GWAS summary statistics obtained from FinnGen Release 8 (13,456 cases, 100,663 controls). The meta-analysis was conducted in METAL using the classical approach with genomic control for each cohort. Although a more recent, better powered GWAS has been published [3], the current dataset was selected to avoid sample overlap between the data used to generate the SNP weightings for the PRS and the testing cohorts for the PRS. A Bayesian method, SBayesR [8], as implemented in GCTB 2.02, was used for adjusting the GWAS summary statistics effect sizes. SBayesR was performed with default settings, in addition to the exclusion of the MHC region, and imputation of the sample size.



**Fig. 1** Flow diagram of methodology. PRS: polygenic risk score; PCs: principal components; GWAS: genome-wide association study; QC: quality control; MR: Mendelian randomisation. Created with Biorender.com

### Calculation of PRS

Firstly, two subgroups of the UKB genotype data were curated: unrelated European males ( $n=159,855$ ) and unrelated European females ( $n=188,221$ ). Specifically, female sex was initially determined using the genotype data provided by UKB and confirmed by only including participants with Sex=Female in the phenotype data. Males had Sex=Male and Sex Genetic=Male in the phenotype data. Ancestry was determined using genetic information [9].

PRS for three cohorts: females, males and a sensitivity cohort comprising of females without an endometriosis diagnosis ( $n=182,789$ ) (henceforth referred to as the sensitivity cohort) were calculated using plink1.9's score function [10] on the SBayesR weightings. The participants considered as endometriosis cases were

determined by entries in 132,122-0.0 and 132,123-0.0 (Date and Source of N80 first report) and was further refined by excluding individuals with the ICD10 diagnostic code N80.0: endometriosis of the uterus, and no other endometriosis diagnosis (N80.1-N80.9). N80.0 refers to adenomyosis, which is currently recognised as a distinct disease to endometriosis.

### Running PheWAS

In a PheWAS, the association of a given genotype with multiple phenotypes is tested. This differs from a typical GWAS, where for a given phenotype, the association with multiple genotypes is determined. Here, the association of endometriosis PRS with multiple phenotypes was tested, i.e. a PRS-PheWAS, essentially testing whether genetic liability to endometriosis affects the probability of

being diagnosed with other traits. PRS was converted to z-score for PRS-PheWAS.

The PRS-PheWAS was conducted using R's `glm` function for logistic regression for the phecodes. The first 10 genetic principal components (PCs) and age (calculated using year of birth) were used as covariates. PCs were calculated on genotype data filtered to minor allele frequency > 5%, genotype missingness < 5%, SNPs passing Hardy–Weinberg exact test with a  $p$  value threshold of  $1 \times 10^{-6}$  and pruned for linkage disequilibrium (window size 50 kb, step size 5 variants,  $r^2$  threshold 0.2) for males and females, in addition to the female sensitivity cohort set. PCA was performed using `plink2` PCA function with the `approx` flag [11]. Phecodes assigned to at least 100 participants were included in the PheWAS.

Linear regression using the `lm` function in R was used for the blood/urine data. Before running the PheWAS, certain biomarkers affected by statins were adjusted in individuals taking statins at enrolment. Adjustment for statin usage was possible for blood/urine biomarker data as statin usage at the time of biological sample collection is available in the UK Biobank. The codes for the medications recognised as statins were 1,141,146,234, 1,141,192,414, 1,140,910,632, 1,140,888,594, 1,140,864,592, 1,141,146,138, 1,140,861,970, 1,140,888,648, 1,141,192,410, 1,141,188,146, 1,140,861,958, 1,140,881,748 and 1,141,200,040. The adjustment method was based on a previously published GWAS of blood and urine biomarkers in the UKB [12]. Specifically, individuals who were taking statins at the first repeat visit, and not the enrolment visit, were identified. For each biomarker in these individuals, the ratio of on-statin biomarker value/pre-statin biomarker value was determined. The mean ratio across all individuals of each sex for each biomarker was determined. This was the sex-specific statin correction factor. Then to determine which biomarkers are affected by statin usage, participants taking statins at the first repeat visit only were utilised for a linear regression modelling the effect of each biomarker on the log ratio of pre-statin biomarker value to on-statin biomarker value. In this linear regression, covariates included Townsend Deprivation index, the first 10 PCs for all individuals of the corresponding cohort, age at enrolment and age difference between enrolment and first visit. For age, only the month and year were utilised. Traits that were significant ( $P < 0.05/34$  as 34 blood/urine traits) in this linear regression analysis for males and females separately were flagged for adjustment in individuals taking statins at enrolment (Additional file 1: Table S1). Specifically, for individuals taking statins at enrolment, their value for biomarkers affected by statins was divided by the biomarker-specific statin correction factor. In females, apolipoprotein B, C-reactive protein, cholesterol, LDL direct and triglycerides were adjusted by

the statin correction factor. In males, apolipoprotein B, C-reactive protein, cholesterol, direct bilirubin, LDL direct, microalbumin in urine, sodium in urine, testosterone and triglycerides were adjusted by the statin correction factor. In the PheWAS, fasting time, age at biomarker measurement (utilising month and year information) and the first 10 PCs were used as covariates.

For the female-specific factors, visual assessment for normality of a QQ plot on inverse-rank normal transformed data determined whether each variable should be treated as a continuous variable or an ordinal categorical variable. Age of menarche and age at first birth were treated as linear variables so tested with linear regression using R's `lm` function, whilst age at menopause, length of menstrual cycle, number of live births and birth weight of first child were treated as ordinal categorical variables and tested with ordinal logistic regression using the `polr` function from the MASS package in R. Linear variables were inverse-rank normal transformed. As ordinal logistic regression does not output a  $P$  value,  $P$  values were calculated by comparing the  $t$ -values to a standard normal distribution. The first 10 genetic PCs were included as covariates for all female-specific factors. Age at menarche was also a covariate for age at first live birth.

Phenotypes were declared significant if they passed a stringent Bonferroni threshold ( $P < 0.05/n$  traits), whereby each set of traits (phecodes, blood/urine biomarker and female-specific factors) were considered separately. Analyses were conducted for the male cohort, female cohort and the sensitivity cohort.

#### Genetic investigation of blood and urine biomarkers

The genetic relationship of biomarkers apolipoprotein A, apolipoprotein B, alanine aminotransferase, testosterone, bioavailable testosterone, triglycerides, HDL cholesterol, LDL cholesterol, albumin, calcium, sex hormone-binding globulin (SHBG) and urate with endometriosis was assessed. These traits were selected due to their significance or nominal significance in the PheWAS analysis and availability of appropriate GWAS summary statistics. SHBG was also included due to its previously reported strong correlation with testosterone. Bioavailable testosterone was also included, as most testosterone is bound to SHBG, and thus inactive. Summary statistics were downloaded from the GWAS Catalog (Table 1) [13]. GWAS summary statistics underwent quality control: missing SNP sample size was replaced with the published total sample size, rsIDs were updated to the rsID of the endometriosis summary statistics to ensure maximum SNP overlap, in the case of duplicated rsIDs the variant with the lowest  $P$  value was retained, and traits without SNP allele frequency had the allele frequencies of

**Table 1** GWAS Summary Statistics of Blood/Urine Biomarkers utilised for cross-trait analysis with endometriosis

Trait	Female-specific	Ancestry and sample size	Publication (GWAS Catalog Accession Number)
Apolipoprotein A	No	311,601 European ancestry individuals, 5550 African ancestry individuals, 6682 South Asian ancestry individuals	Sinnott-Armstrong, Tanigawa [12] GCST90019495
Apolipoprotein B	No	340,860 European ancestry individuals, 5962 African ancestry individuals, 7275 South Asian ancestry individuals	Sinnott-Armstrong, Tanigawa [12] GCST90019496
Alanine aminotransferase	No	342,387 European ancestry individuals, 6017 African ancestry individuals, 7325 South Asian ancestry individuals	Sinnott-Armstrong, Tanigawa [12] GCST90019492
Testosterone	Yes	230,454 European ancestry women	Ruth, Day [14] GCST90012112
Bioavailable Testosterone	Yes	188,507 European ancestry women	Ruth, Day [14] GCST90012102
Triglycerides (for Genetic Correlation)	No	1,320,016 European ancestry individuals	Graham, Clarke [16] GCST90239664
Triglycerides (for MR)	No	115,082 European ancestry individuals	Richardson, Sanderson [17] GCST90092992
Urate	No	342,087 European ancestry individuals, 6011 African ancestry individuals, 7328 South Asian ancestry individuals	Sinnott-Armstrong, Tanigawa [12] GCST90019524
HDL Cholesterol	No	313,372 European ancestry individuals, 5573 African ancestry individuals, 6689 South Asian ancestry individuals	Sinnott-Armstrong, Tanigawa [12] GCST90019510
LDL Cholesterol	No	341,875 European ancestry individuals, 6003 African ancestry individuals, 7319 South Asian ancestry individuals	Sinnott-Armstrong, Tanigawa [12] GCST90019512
Sex hormone-binding globulin	Yes	189,473 European ancestry women	Ruth, Day [14] GCST90012107
Albumin	No	313,032 European ancestry individuals, 5573 African ancestry individuals, 6687 South Asian ancestry individuals	Sinnott-Armstrong, Tanigawa [12] GCST90019493
Calcium	No	313,387 European ancestry individuals, 5576 African ancestry individuals, 6696 South Asian ancestry individuals	Sinnott-Armstrong, Tanigawa [12] GCST90019500

**MR Mendelian randomisation**

unrelated European females in the UKB appended. Two endometriosis datasets were utilised. In the first, utilised for genetic correlation, three endometriosis GWAS datasets were meta-analysed to maximise power: Rahmioglu 2023 (21,779 European ancestry cases, 449,087 European ancestry controls, 1713 Japanese ancestry cases, 1581 Japanese ancestry controls) [3], FinnGen release 8 (13,456 cases, 100,663 controls), and 23andMe, Inc. (4970 cases, 34,561 controls). SNP rsIDs were harmonised, SNPs with missing  $P$  or beta values were removed, SNPs with  $P < 0$  or  $P > 1$  were removed and in the case of duplicate SNPs, the SNP with the lowest  $p$  value was retained. The meta-analysis was conducted in METAL using the classical approach with genomic control for each cohort. The meta-analysis result underwent a secondary round of genomic control. As the allele frequency was not present in all three cohorts, the allele frequency from the largest cohort [3] was subbed in, restricting to

SNPs present in this GWAS. The sample size was approximated by the sum of the cohorts the SNP was present in. The second endometriosis dataset, detailed earlier in the methods for the PheWAS, was the meta-analysis of the European component of the 2017 endometriosis GWAS [7] with FinnGen release 8 (13,456 cases, 100,663 controls). Although less powered, this second dataset was necessary to avoid sample overlap with the biomarker GWAS, as this is known to bias estimates for Mendelian randomisation. These GWAS summary statistics are referred to as dataset 2. Follow-up analysis of testosterone [14] and ovarian cancer utilised ovarian cancer summary statistics published in Phelan, Kuchenbaecker [15].

Genetic correlation assesses the average genome-wide correlation in variant effects between traits. The genetic correlation between endometriosis and the blood/urine traits was estimated using LDSC (v1.0.1) and precomputed LD scores from the 1000 Genomes

European reference set. Using the `munge_sumstats.py` script, the alleles in the GWA summary statistics for each trait were crosschecked against HapMap3 SNPs used to estimate the LD scores using the `merge-alleles` function and chunk size 1,000,000. Genetic correlation was performed using the endometriosis dataset 1 GWAS summary statistics.

Mendelian randomisation analysis was performed to assess whether any of these traits could be linked to endometriosis via causality. Mendelian randomisation is a statistical technique that assesses causality through the use of genetic variants as instrumental variables (IVs) [18]. Initially, three models were applied: inverse-variance weighted (IVW), MR-Egger (MRE) and weighted median (WM). The IVW method assumes all variants satisfy the assumptions of MR. The MRE and WM methods are important sensitivity tests. MRE allows for an overall directional pleiotropic effect, which provides valid estimates if the pleiotropic effect on the outcome is independent of their effects on the exposure [19]. The WM method provides valid estimates if at least 50% of the variants are valid instrumental variables [20]. Heterogeneity and pleiotropy statistics were also calculated. Causality was assessed with endometriosis as both the exposure and outcome variable. SNPs were filtered using the MR Steiger directionality test to ensure they were more strongly associated with the exposure than the outcome. FDR-adjusted  $P$  values were calculated within each directional analysis (i.e. endometriosis as exposure, endometriosis as outcome), and within each MR method. Significant results (adjusted  $P < 0.05$ ) were further investigated using additional MR methods GSMR [21] and MR-PRESSO. GSMR and MR-PRESSO detect and remove outlier SNPs, so the result is unlikely to be biased by pleiotropy. As many blood biomarkers show strong phenotypic correlation and have many shared genetic risk loci, multivariable MR was conducted for testosterone with SHBG, apolipoprotein A with HDL-C and triglycerides with LDL-C and apolipoprotein B. Multivariable MR determines the causal effect of an exposure on the outcome, conditional on the other exposures in the model. This is an appropriate sensitivity test to run when potential confounders are known. The `TwoSampleMR` package in R was utilised for both univariate and multivariable MR, whilst GSMR was performed as implemented in GCTA v1.94.1. The LD reference dataset for GSMR was the QIMRHCS cohort [7]. Independent SNPs with genome-wide significance ( $P < 5 \times 10^{-8}$ ) were utilised as IVs. As many of the blood/urine biomarker summary statistics were derived from UK Biobank data, endometriosis dataset 2 summary statistics were utilised. All IVs were sufficiently powered: the  $F$ -statistics ( $\beta^2/se^2$ ) for each IV for every exposure were  $> 10$ .

## Results

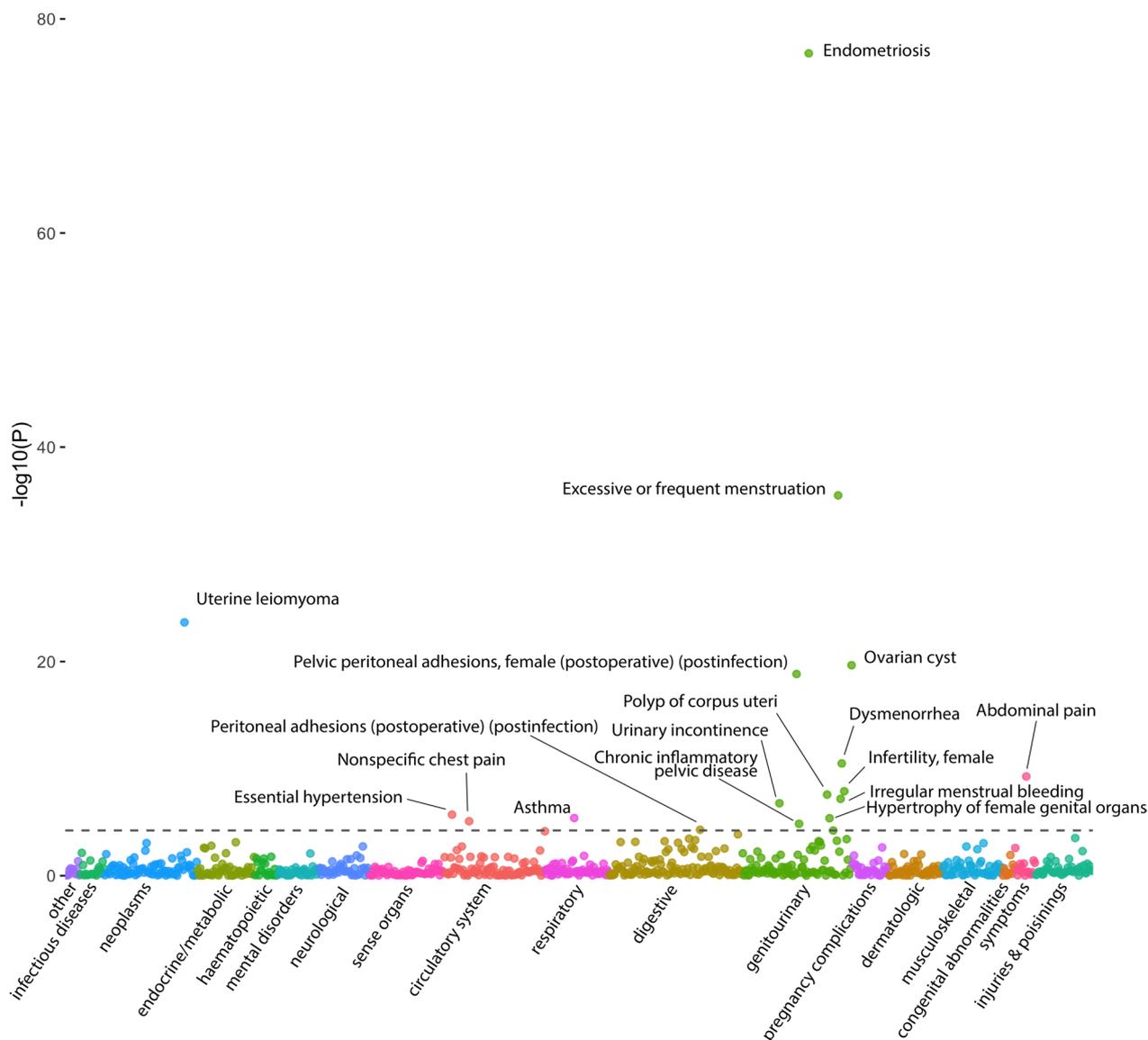
### PRS-PheWAS results

The PheWAS of phecodes tested association of multiple phenotypes with endometriosis PRS. Phecodes aggregate multiple ICD10 codes corresponding to a similar phenotype. There were 17, 11 and 2 significant phecodes in the female, sensitivity and male cohort PheWASs, respectively (Additional file 1: Tables S2-S4). All significant phecodes were positively associated with endometriosis PRS. In the female phecode PheWAS, the top associated phecode was 615: Endometriosis (Fig. 2). Other highly associated traits included excessive/frequent menstruation, uterine leiomyoma, ovarian cyst, and pelvic peritoneal adhesions. All phecodes significant in the female analysis replicated at least nominal significance in the female sensitivity cohort, except for chronic inflammatory pelvic disease ( $P = 0.064$ ). All 11 phecodes significant in the sensitivity cohort were also significant in the whole female cohort. The phecode for endometriosis was also significant in the sensitivity cohort. As the phecode definition of endometriosis includes the ICD10 code N80.0: endometriosis of the uterus, which here was excluded from the endometriosis definition, the endometriosis phecode in the female sensitivity cohort is representative of adenomyosis. In the male analysis, the two traits significantly associated with endometriosis PRS were abdominal pain and hyperplasia of prostate.

In the female-specific factor analysis, younger age at menopause, menarche and first live birth, and shorter length of menstrual cycle were significantly associated with endometriosis PRS (Table 2, Additional file 1: Table S5). This was replicated in the sensitivity cohort, although age at menopause was nominally significant ( $P = 0.047$ ) (Additional file 1: Table S6).

Multiple blood and urine biomarkers were associated with endometriosis PRS (Tables 3, Additional file 1: Tables S7-S9). All thirteen biomarkers significant in the female analysis replicated at the Bonferroni-corrected threshold or nominal significance in the female sensitivity cohort. In the male analysis five biomarkers were significant, four were also significant in the female analysis (triglycerides, HDL cholesterol, calcium, apolipoprotein A), whilst alkaline phosphatase had nominal significance in the female analysis.

The relationship of endometriosis with a subset of the blood/urine biomarkers was further explored. Genetic correlation analysis revealed apolipoprotein A, HDL cholesterol, triglycerides, testosterone, SHBG and alanine aminotransferase were all significantly correlated with endometriosis and passed the stringent Bonferroni-corrected  $P$  value threshold (Fig. 3, Additional file 1: Table S10).



**Fig. 2** Female PRS-PheWAS for endometriosis in the UK Biobank. In total, 841 phecodes were tested for their association with endometriosis PRS. Traits are grouped and colour-coded into categories. The *P* value threshold for significance (dotted line) is  $5.95 \times 10^{-5}$  (Bonferroni-corrected threshold). The *P* values were generated from logistic regression, with the first ten genetic principal components and age as covariates. Traits with significant *P* values are annotated

**Table 2** Female-specific factors associated with endometriosis PRS. Females in the UK Biobank were utilised. Estimate refers to the regression coefficient from linear regression (age at menarche and age at first live birth), or from ordinal logistic regression (age at menopause and length of menstrual cycle)

Biomarker	Estimate	SE	<i>P</i>
Age at menopause	-0.016	0.005	$2.64 \times 10^{-3}$
Length of menstrual cycle	-0.062	0.011	$3.38 \times 10^{-8}$
Age at menarche	-0.017	0.002	$4.72 \times 10^{-14}$
Age at first live birth	-0.021	0.002	$1.08 \times 10^{-13}$

SE Standard error, *P* *P* value

A negative causal effect of testosterone on endometriosis was supported by the IVW method ( $b = -0.20$ ,  $P = 2.12 \times 10^{-7}$ ,  $P_{adjusted} = 2.54 \times 10^{-6}$ ) Fig. 4, Additional file 1: Table S11). Although heterogeneity was present, this relationship was not driven by any individual SNP, as indicated by leave-one-out analysis (Additional file 1: Table S12A). Whilst MRE and WM were not significant, the direction of effect was concordant with the IVW result, and the unadjusted MRE result was significant ( $b = -0.14$ ,  $P = 0.049$ ,  $P_{adjusted} = 0.26$ ). Complementary tests MR-PRESSO and GSMR supported the result: the

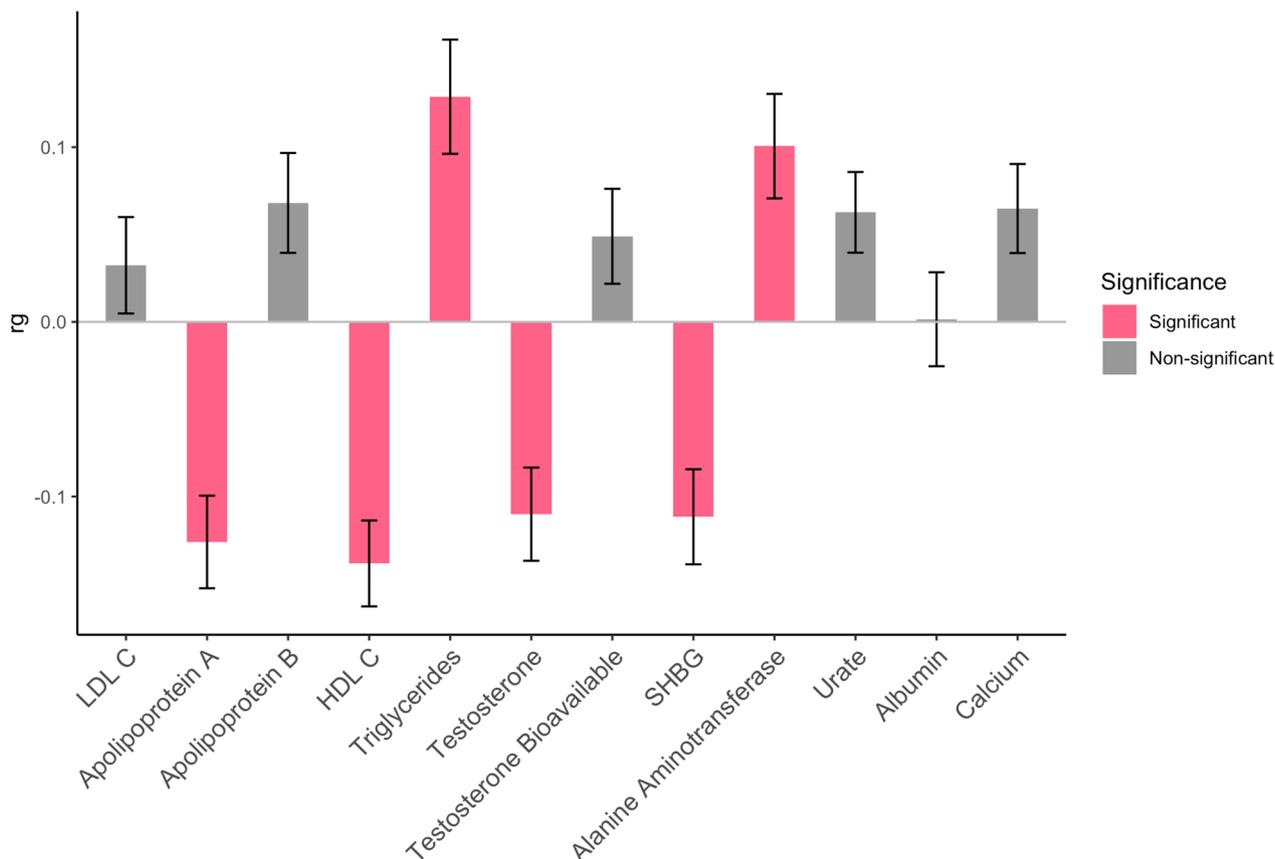
**Table 3** Significant blood/urine biomarkers associated with endometriosis PRS in females in the UK Biobank. Biomarkers were corrected for statin usage. Estimate refers to the regression coefficient from linear regression

Biomarker	Estimate	SE	P
Triglycerides	0.0140	0.0020	4.56E-12
Calcium	0.0013	0.0002	2.56E-08
Alanine aminotransferase	0.1396	0.0291	1.62E-06
Urate	0.7117	0.1525	3.07E-06
HDL cholesterol	-0.0042	0.0009	6.37E-06
Oestradiol	9.8021	2.3176	2.35E-05
Gamma glutamyltransferase	0.3345	0.0794	2.53E-05
Albumin	0.0265	0.0064	3.55E-05
Apolipoprotein B	0.0022	0.0005	4.35E-05
Testosterone	-0.0060	0.0016	2.61E-04
LDL direct	0.0065	0.0019	8.74E-04
Apolipoprotein A	-0.0022	0.0007	8.80E-04
Total protein	0.0323	0.0100	1.23E-03

SE Standard error, P P value

MR-PRESSO method was significant, even after adjustment for outliers ( $b = -0.18, P = 2.14 \times 10^{-6}$ ), and GSMR was also significant ( $b = -0.14, P = 8.41 \times 10^{-7}$ ). As the majority of testosterone is bound to SHBG and inactive, we also considered the effect of bioavailable testosterone on endometriosis. The IVW method for bioavailable testosterone was also significant with a negative causal effect ( $b = -0.16, P = 8.04 \times 10^{-3}$ ) (Fig. 4). Although there was significant heterogeneity in the IVW model, the result was robust to leave-one-out analysis (Additional file 1: Table S12B). MR-PRESSO and GSMR supported a negative causal effect of bioavailable testosterone on endometriosis (MR-PRESSO outlier adjusted:  $b = -0.18, P = 1.53 \times 10^{-3}$ , GSMR:  $b = -0.20, P = 7.75 \times 10^{-7}$ ). In a multivariable model including SHBG, the effect of overall testosterone on endometriosis was retained ( $b = -0.21, P = 5.03 \times 10^{-5}$ ).

A causal effect of testosterone on ovarian cancer risk has previously been reported using MR approaches [14]. Likewise, a causal effect of endometriosis on ovarian cancer has also been reported [22]. Therefore, we investigated the causal pathways between these three



**Fig. 3** Genetic correlation of blood and urine traits with endometriosis. Genetic correlation (rg) was determined using LDSC. Traits were considered significantly genetically correlated with endometriosis if  $P < 4.17 \times 10^{-3}$ . Error bars represent standard errors

traits. Firstly, we evaluated the causal effect of testosterone on ovarian cancer histotypes, as the previously reported analysis was limited to overall ovarian cancer. A negative causal effect of bioavailable testosterone was identified on one histotype: clear cell carcinoma ovarian cancer with the IVW method ( $b = -0.44$ ,  $P = 6.61 \times 10^{-3}$ ,  $P_{adjusted} = 0.046$ ), in the absence of heterogeneity (Fig. 4, Additional file 1: Table S13). The unadjusted  $P$  values and direction of effect for the WM ( $b = -0.63$ ,  $P = 0.0432$ ) and MRE ( $b = -0.66$ ,  $P = 0.0430$ ) methods supported this result for bioavailable testosterone (Fig. 4, Additional file 1: Table S13). The IVW result was not driven by an individual SNP (Additional file 1: Table S12C). Both MR-PRESSO ( $b = -0.41$ ,  $P = 0.014$ , no outliers detected) and GSMR ( $b = -0.48$ ,  $P = 6.48 \times 10^{-4}$ ) also supported the causal effect. Total testosterone also had concordant directions of effect on clear cell carcinoma (Additional file 1: Table S13). When considering the effect of bioavailable testosterone on clear cell ovarian cancer in a multivariable model with endometriosis, the causal effect of bioavailable testosterone was no longer significant ( $P = 0.073$ ), whilst endometriosis retained its causative effect ( $b = 0.78$ ,  $P = 1.80 \times 10^{-13}$ ) (Fig. 4, Additional file 1: Table S14).

## Discussion

In this study, we use a PheWAS approach to identify conditions, female-specific traits, and blood/urine biomarkers associated with the genetic liability to endometriosis. A PheWAS approach differs from epidemiological and genomic studies through the integrated analysis of phenotype and genotype data. This approach enabled investigation of the effects of genetic liability to a disease in the absence of disease: utilisation of a female cohort without endometriosis diagnoses and a male cohort suggested that association of many traits with endometriosis cannot solely be attributed to the physical presence of endometriosis. Using Mendelian randomisation approaches, we also identify a possible causal effect of the genetic liability to lower testosterone on endometriosis and clear cell ovarian cancer, which, following further validation, may have important clinical implications for both endometriosis and ovarian cancer.

Given many endometriosis diagnoses are likely missed in the UK Biobank participants [23], the exclusion of individuals with an endometriosis diagnosis from the sensitivity cohort was likely imperfect. In addition to the endometriosis sensitivity cohort, males were used as a high-confidence endometriosis-free group to investigate the effects of genetic liability to endometriosis in the absence of the physical presence of endometriosis. There were two phecodes associated with the endometriosis PRS in males: abdominal pain and hyperplasia of prostate. The relationship with prostate hyperplasia is intriguing, as although the cause of benign prostate hyperplasia is unclear, inflammation and proliferation, also characteristic features of endometriosis, are key [24]. Given these individuals did not have a history of endometriosis, the association of abdominal pain with the endometriosis PRS must be explained by factors beyond presence of the lesion. A shared genetic background has previously been identified between endometriosis and multiple pain traits [3, 5]. Multiple other traits previously identified to have a shared genetic background with endometriosis were significant in the female analysis but not the male analysis. This may be explained through sex-specific hormonal-related pathways being involved in the overlap of endometriosis with these traits. Pain is an incentive to seeking an endometriosis diagnosis, so in addition to the endometriosis genetic risk signals capturing variants associated with lesion growth, gynaecological healthcare-seeking factors such as sensitivity to pain signals may also be captured. Therefore, the association of abdominal pain in males with the endometriosis PRS could be explained by the endometriosis risk variants being enriched for pain sensitivity signals. However, this explanation does not negate pleiotropic effects of the genetic variants on both endometriosis and pain sensitivity, or pain-causing effects of the endometriosis lesions, given most individuals report a reduction in pain in the short term following surgical removal of the lesions [25]. Disentangling the genetic effects of the risk signals for endometriosis on lesion characteristics and endometriosis symptoms will be of interest for future studies. Further, the association of traits with the endometriosis PRS in males and in the female sensitivity cohort suggests epidemiological studies should look for symptoms and traits in relatives of

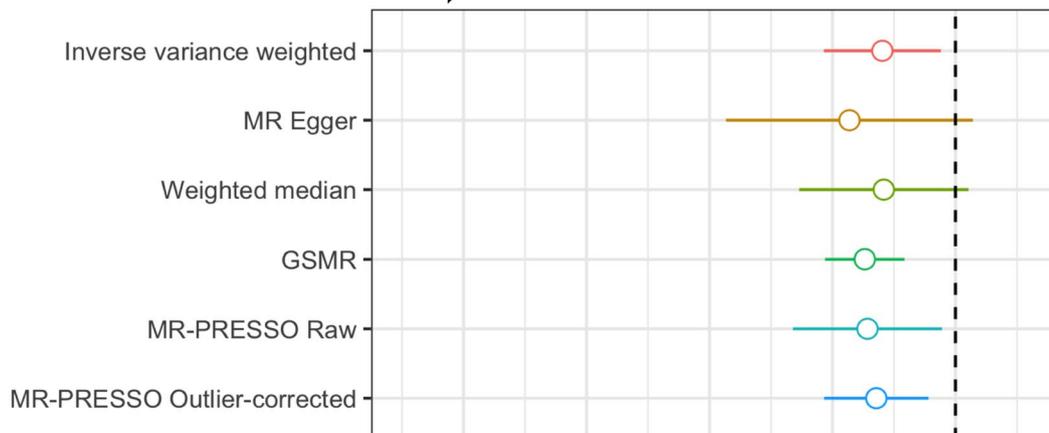
(See figure on next page.)

**Fig. 4** Causal relationships between testosterone, endometriosis and clear cell carcinoma ovarian cancer determined by Mendelian randomisation. A causal effect of testosterone and bioavailable testosterone on endometriosis is supported by multiple methods. A causal effect of bioavailable testosterone on clear cell carcinoma ovarian cancer was identified using univariate approaches, but adjustment for endometriosis using multivariable MR attenuated the effect and it was non-significant. MR: Mendelian randomisation, GSMR: generalised summary Mendelian randomisation, SHBG: sex hormone-binding globulin, IVW: inverse-variance weighted. MR-PRESSO: Mendelian Randomization Pleiotropy Residual Sum and Outlier

Testosterone → Endometriosis



Bioavailable Testosterone → Endometriosis



Bioavailable Testosterone → Clear Cell Carcinoma

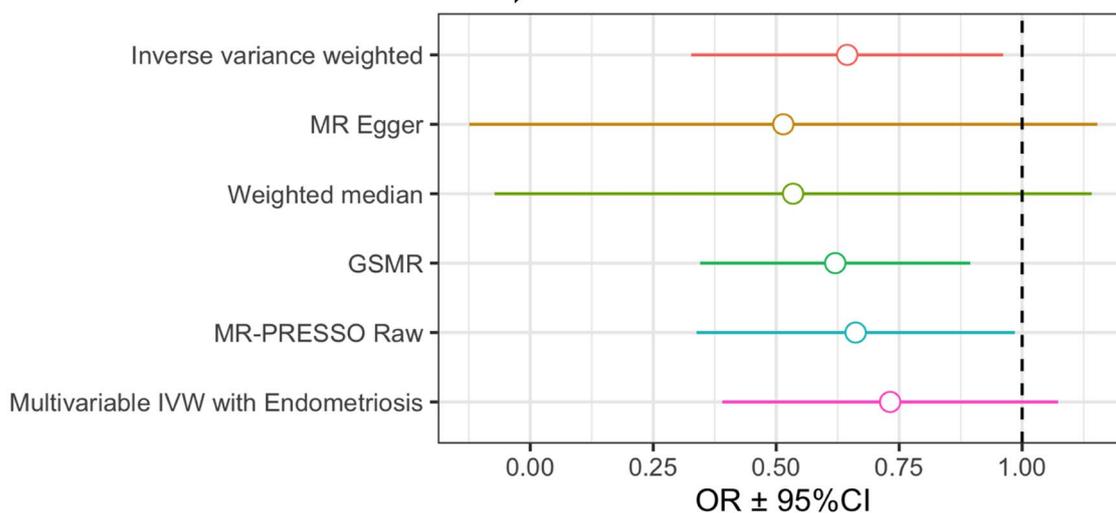


Fig. 4 (See legend on previous page.)

individuals with endometriosis, which could prove valuable disease-predictive factors.

The traits in the PRS-PheWAS analysis identified to be associated with endometriosis genetic risk are directly dependent on the characteristics of the cohorts used to generate the GWAS summary statistics used to calculate the PRSs. Endometriosis is a highly heterogeneous condition, with a large spectrum in symptom intensity. Further, women face a multitude of barriers in accessing a diagnosis including the trivialisation of their symptoms, the non-specificity of their symptoms, cost of seeking healthcare and the invasiveness of the gold-standard diagnosis (laparoscopic surgery). Therefore, women with more severe symptoms and fewer barriers to seeking healthcare are more likely to be diagnosed, and thus present in the case sample of a GWAS study. Likewise, there are likely many undiagnosed cases in the control cohort, reducing the power of the GWAS. If there is variability in the genetic architecture of endometriosis that is correlated with healthcare-seeking factors, the results of the PRS-PheWAS analysis may be restricted to a subset of endometriosis cases.

A key finding of this study was the suggestion of a causal effect of genetically predicted lower testosterone on endometriosis using MR. This effect was consistent across multiple models, and when using multiple measures of testosterone: overall testosterone and bioavailable testosterone. Although not all MR models were significant, the direction of effect was consistent between models. A limitation of this approach to modelling the relationship with the MR techniques utilised is that a linear effect is assumed: meaning the model suggests lower testosterone is causative, and higher testosterone is preventative of endometriosis. However, this may be a simplification of the relationship if testosterone only exerts its causative/preventative effect in one of these directions. The testosterone SNPs used for MR were female-specific which is important given there is no genetic correlation between testosterone in males and females [14]; however, they were generated from a GWAS of mostly postmenopausal adults. Given endometriosis onsets in the early reproductive years, assessment of the causative effect using adolescent, childhood and/or prenatal testosterone-associated SNPs should be assessed if the genetic control of testosterone differs at these stages.

Nevertheless, a role for alterations to the hypothalamic-pituitary-gonadal (HPG) axis, and a role for testosterone, specifically lower prenatal testosterone in endometriosis, has been previously discussed [26–28]. Differences in anogenital distance, a proxy for prenatal testosterone, between endometriosis cases and controls implicates lower prenatal testosterone in endometriosis cases [29]. In females, testosterone is synthesised from

cholesterol in the adrenal gland and ovaries, and the primary regulator of testosterone levels in females is this steroid biosynthesis pathway [30]. In the ovary, testosterone is converted to oestradiol by aromatase, which is under the control of follicle-stimulating hormone (FSH). Elevated FSH has been reported in endometriosis cases (although some studies find no difference), alongside increased oestradiol only in the menstrual fluid (not in circulation) and increased aromatase activity in the eutopic endometrium [27]. In the PheWAS, oestradiol levels were positively correlated with endometriosis PRS; however, this should be considered cautiously as most women in the UKB are postmenopausal. In contrast, a reduced ovarian oestrogen to testosterone ratio, reduced FSH and reduced ovarian aromatase activity are observed in polycystic ovarian syndrome (PCOS). Alternate alleles of SNPs in strong LD upstream of *FSHB* are alternatively associated with endometriosis and PCOS: i.e. one haplotype confers risk to endometriosis, the other to PCOS [31].

There are a few possible mechanisms of lower testosterone impacting endometriosis risk. In support of Sampson's retrograde menstruation hypothesis for endometriosis, lower prenatal and postnatal testosterone cause earlier menarche, shorter menstrual cycles and thicker endometrial lining, increasing the exposure to menstruation [27]. Oestrogens have inflammatory effects, whilst androgens have anti-inflammatory effects, so a high oestrogen:testosterone ratio could promote inflammation in response to ectopic endometrial tissue [27]. The Müllerian remnants are another hypothesis for endometriosis, whereby misplaced stem cells are activated by a stimulus [32, 33]. Low testosterone could contribute to the activation of these stem cells by facilitating a high inflammatory environment, and/or through facilitating deposition of these stem cells [27]. Testosterone is a controller of *HOXA10* expression [34], which is a key player in the development of the female reproductive tract, and has been implicated in endometriosis [35]. Low testosterone could also contribute to pain in endometriosis, through mechanisms such as the inverse association with inflammation, and/or through links with  $\beta$ -endorphin levels within the central nervous system [28]. Importantly, a role for prenatal testosterone and early disturbance of the HPG axis in endometriosis would imply a developmental origin for endometriosis [28].

Previously a causative effect of lower testosterone on ovarian cancer was reported using MR [14]. Analysis of the major ovarian cancer histotypes indicated this causative effect was restricted to clear cell carcinoma; however, this effect was attenuated and non-significant when considered in a multivariable MR model

with endometriosis. This suggests the effect of lower testosterone on clear cell carcinoma may be partially mediated through endometriosis; however, confidence intervals were large, so validation in larger datasets is necessary. Endometriosis has been previously identified as a cause of multiple histotypes of ovarian cancer: most strongly clear cell carcinoma and endometrioid ovarian cancer [22]. As lower testosterone did not show a causative effect on endometrioid ovarian cancer, testosterone may play a role in determining the histotype of ovarian cancer resulting from the endometriosis. Interestingly, polycystic ovarian syndrome, for which high testosterone is a characteristic feature and identified as causative using MR [14], has been determined as preventative for endometrioid ovarian cancer using MR techniques [36].

Four female-specific factors were associated with the endometriosis PRS: earlier age at menopause, shorter length of menstrual cycle, earlier age at menarche and earlier age at first live birth. A significant genetic correlation of endometriosis with these traits has previously been reported [3]. Shorter length of menstrual cycle and earlier age at menarche in endometriosis, also observed in epidemiological data, mean a greater exposure to menstruation, in support of Sampson's theory of retrograde menstruation for endometriosis [37, 38]. The association of earlier age at first live birth, a proxy for fertility, with endometriosis PRS likely owes to the progressive nature of fertility issues. Age at first birth remained significant in the female sensitivity cohort. This may be explained by the presence of undiagnosed endometriosis cases in this cohort, and/or the direct effects of some endometriosis risk loci on infertility. Well-designed epidemiological studies of the relationship between age of menopause and endometriosis are lacking [39]. One study reported an increased risk of early natural menopause (< 45 years) in endometriosis patients [40]. Potential mechanisms include effects of endometrioma (endometriosis on the ovary) on ovarian function, effects of surgical excision of endometrioma [41] and effects of endometriosis-related traits such as reduced body mass index [39]. The attenuation of the effect of endometriosis PRS on age at menopause when endometriosis cases were excluded may point to effects of lesion presence on age at menopause, but further validation is needed.

The association of blood biomarkers with the endometriosis PRS provides promising evidence these biomarkers could be useful in predicting endometriosis. However, as the biomarker measurements from the UKB utilised here were measured in women older than the typical age for seeking an endometriosis diagnosis,

their prediction accuracy should be assessed in a younger cohort, prior to surgical excision of the endometriosis lesions. Given most women in the UKB are postmenopausal, this is particularly relevant for biomarkers that are strongly affected by menopause, such as oestrogen. The association of various lipid-related biomarkers aligns with the association of endometriosis with cardiovascular traits in the phecode PheWAS and has been previously reported in epidemiological data [42]. Triglycerides show significant differences between endometriotic and normal endometrium of endometriosis patients [43]. The absence of any causal relationships between endometriosis and most blood/urine biomarkers may suggest pleiotropic genetic effects and/or non-genetic factors may be responsible for the overlap. One limitation of the genetic analysis is that for most traits the GWAS summary statistics were not female-specific, nor specific to young reproductive aged women. Timely and female-specific genetic associations may be necessary to reveal causal relationships and the true magnitude of genetic overlap with endometriosis. Further, not all GWAS summary statistics were derived from a purely European ancestry sample; however, the non-European component is small, and thus any bias is expected to be small.

## Conclusions

We have performed a comprehensive PRS-PheWAS for endometriosis. Associations of traits with genetic liability to disease, rather than disease presence, has provided interesting insights into the comorbidity of these traits with endometriosis and has ramifications for co-treatment of these diseases. Validation of a causal effect of lower testosterone on endometriosis using MR, and the finding of an effect of lower testosterone on clear cell carcinoma, prompts further investigation into the developmental origins of endometriosis and the malignant transformation of endometriosis into ovarian cancer in relation to testosterone.

## Abbreviations

FSH	Follicle-stimulating hormone
GSMR	Generalised summary Mendelian randomisation
HPG	Hypothalamic-pituitary-gonadal
IVs	Instrumental variables
IWV	Inverse-variance weighted
MRE	MR-Egger
MR-PRESSO	Mendelian Randomization Pleiotropy Residual Sum and Outlier
PCOS	Polycystic ovarian syndrome
PCs	Principal components
PheWAS	Phenome-wide association study
PRS	Polygenic risk score
SHBG	Sex hormone-binding globulin
UKB	UK Biobank
WM	Weighted median

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12916-023-03184-z>.

**Additional file 1: Table S1.** Statin Adjustment Factors and Linear Regression Analysis of Effect of Statin Usage on Biomarker Levels. **Table S2.** PRS-PheWAS of phecodes in females utilising endometriosis PRS. **Table S3.** PheWAS of phecodes in females without endometriosis utilising endometriosis PRS. **Table S4.** PheWAS of phecodes in males utilising endometriosis PRS. **Table S5.** PheWAS of female-specific factors in females utilising endometriosis PRS. **Table S6.** PheWAS of female-specific factors in females without endometriosis utilising endometriosis PRS. **Table S7.** PRS-PheWAS of blood/urine biomarkers in females utilising endometriosis PRS. **Table S8.** PRS-PheWAS of blood/urine biomarkers in females without endometriosis utilising endometriosis PRS. **Table S9.** PRS-PheWAS of blood/urine biomarkers in males utilising endometriosis PRS. **Table S10.** Genetic Correlation (rg) between endometriosis and blood/urine biomarkers. **Table S11.** Mendelian randomisation of endometriosis and blood/urine biomarkers. **Tables S12.** A-C Leave one out analyses. **S13.** Mendelian randomisation ovarian cancer and testosterone. **S14.** Multivariable Mendelian randomisation.

### Acknowledgements

Summary statistics from the endometriosis GWAS used in this study contain data from 23andMe and FinnGen. We want to acknowledge the participants and investigators of the FinnGen study and would like to thank the research participants and employees of 23andMe, Inc. for making this work possible. This research has been conducted using the UK Biobank Resource under Application Number 54861. Individual-level phenotype information (including age and IDC10 diagnosis) and genotype data for genetic risk assessment was accessed under Application Number 54861. We thank the UK Biobank participants and the UKB research teams for their generous contributions to generating an important research resource. This work uses data provided by patients and collected by the NHS as part of their care and support. Copyright © 2023, NHS England. Re-used with the permission of the UK Biobank. All rights reserved.

### Authors' contributions

IMM, GWM and SM contributed to the study conception and design. Data analysis was performed by IMM and was interpreted by IMM, GWM and SM. IEGC generated, and provided access to, the endometriosis GWA meta-analysis. The first draft of the manuscript was written by IMM, and IMM, GWM and SM commented on previous versions of the manuscript. IMM, GWM and SM read and approved the final manuscript.

### Authors' Twitter handles

Twitter handles: @SallyMortlock (Sally Mortlock); @IsabelleMcGr (Isabelle M. McGrath).

### Funding

Open Access funding enabled and organised by CAUL and its Member Institutions. GWM was supported by NHMRC Fellowship GNT1177194. SM was supported by Medical Research Future Fund Research Grant MRF1199785.

### Availability of data and materials

Data can be accessed from UK Biobank (<https://www.ukbiobank.ac.uk/>) as per their published data access procedures. Summary data for the FinnGen Endometriosis GWAS is available from their results portal ([https://www.finnngen.fi/en/access\\_results](https://www.finnngen.fi/en/access_results)). Publicly available GWAS summary data is cited in Table 1 or detailed in text. Endometriosis GWAS summary statistics, where not publicly available, are available on request. Any additional data supporting the conclusions of this article are included within the article and its additional files.

### Declarations

#### Ethics approval and consent to participate

This study employed data from the UK Biobank and summary statistics from additional sources and were approved by the UK Biobank and the Human

Research Ethics Committee of The University of Queensland (Project 2020/HE002852).

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare that they have no competing interests.

Received: 13 August 2023 Accepted: 20 November 2023

Published online: 05 December 2023

### References

- Treloar SA, O'Connor DT, O'Connor VM, Martin NG. Genetic influences on endometriosis in an Australian twin sample. *Fertil Steril*. 1999;71(4):701–10.
- Saha R, Pettersson HJ, Svedberg P, Olovsson M, Bergqvist A, Marions L, et al. Heritability of endometriosis. *Fertil Steril*. 2015;104(4):947–52.
- Rahmioglu N, Mortlock S, Ghiasi M, Moller PL, Stefansdottir L, Galarneau G, et al. The genetic basis of endometriosis and comorbidity with other pain and inflammatory conditions. *Nat Genet*. 2023;55(3):423–36.
- McGrath IM, Montgomery GW, Mortlock S. Insights from Mendelian randomization and genetic correlation analyses into the relationship between endometriosis and its comorbidities. *Hum Reprod Update*. 2023;29(5):655–74.
- McGrath IM, International Endometriosis Genetics Consortium, Montgomery GW, Mortlock S. Genomic characterisation of the overlap of endometriosis with 76 comorbidities identifies pleiotropic and causal mechanisms underlying disease risk. *Hum Genet*. 2023;142(9):1345–60.
- Yang F, Wu Y, Hockey R, Doust J, Mishra GD, Montgomery GW, et al. Evidence of shared genetic factors in the etiology of gastrointestinal disorders and endometriosis and clinical implications for disease management. *Cell Rep Med*. 2023;4(11):101250.
- Sapkota Y, Steinhorsdottir V, Morris AP, Fassbender A, Rahmioglu N, De Vivo I, et al. Meta-analysis identifies five novel loci associated with endometriosis highlighting key genes involved in hormone metabolism. *Nat Commun*. 2017;8:15539.
- Lloyd-Jones LR, Zeng J, Sidorenko J, Yengo L, Moser G, Kemper KE, et al. Improved polygenic prediction by Bayesian multiple regression on summary statistics. *Nat Commun*. 2019;10(1):5086.
- Yengo L, Sidorenko J, Kemper KE, Zheng Z, Wood AR, Weedon MN, et al. Meta-analysis of genome-wide association studies for height and body mass index in ~700,000 individuals of European ancestry. *Hum Mol Genet*. 2018;27(20):3641–9.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81(1):559–75.
- Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*. 2015;4(1):7.
- Sinnott-Armstrong N, Tanigawa Y, Amar D, Mars N, Benner C, Aguirre M, et al. Genetics of 35 blood and urine biomarkers in the UK Biobank. *Nat Genet*. 2021;53(2):185–94.
- Sollis E, Mosaku A, Abid A, Buniello A, Cerezo M, Gil L, et al. The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. *Nucleic Acids Res*. 2023;51(D1):D977–85.
- Ruth KS, Day FR, Tyrrell J, Thompson DJ, Wood AR, Mahajan A, et al. Using human genetics to understand the disease impacts of testosterone in men and women. *Nat Med*. 2020;26(2):252–8.
- Phelan CM, Kuchenbaecker KB, Tyrer JP, Kar SP, Lawrenson K, Winham SJ, et al. Identification of 12 new susceptibility loci for different histotypes of epithelial ovarian cancer. *Nat Genet*. 2017;49(5):680–91.
- Graham SE, Clarke SL, Wu KH, Kanoni S, Zajac GJM, Ramdas S, et al. The power of genetic diversity in genome-wide association studies of lipids. *Nature*. 2021;600(7890):675–9.
- Richardson TG, Sanderson E, Palmer TM, Ala-Korpela M, Ference BA, Davey Smith G, et al. Evaluating the relationship between circulating lipoprotein lipids and apolipoproteins with risk of coronary heart disease:

- a multivariable Mendelian randomisation analysis. *PLoS Med.* 2020;17(3): e1003062.
18. Smith GD, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol.* 2003;32(1):1–22.
  19. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol.* 2013;37(7):658–65.
  20. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genet Epidemiol.* 2016;40(4):304–14.
  21. Zhu Z, Zheng Z, Zhang F, Wu Y, Trzaskowski M, Maier R, et al. Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nat Commun.* 2018;9(1):224.
  22. Mortlock S, Corona RI, Kho PF, Pharoah P, Seo JH, Freedman ML, et al. A multi-level investigation of the genetic relationship between endometriosis and ovarian cancer histotypes. *Cell Rep Med.* 2022;3(3): 100542.
  23. Blass I, Sahar T, Shraibman A, Ofer D, Rappoport N, Linal M. Revisiting the risk factors for endometriosis: a machine learning approach. *J Pers Med.* 2022;12(7):1114.
  24. Devlin CM, Simms MS, Maitland NJ. Benign prostatic hyperplasia – what do we know? *BJU Int.* 2021;127(4):389–99.
  25. Leonardi M, Gibbons T, Armour M, Wang R, Glanville E, Hodgson R, et al. When to do surgery and when not to do surgery for endometriosis: a systematic review and meta-analysis. *J Minim Invasive Gynecol.* 2020;27(2):390–407.e3.
  26. Dinsdale N, Nepomnaschy P, Crespi B. The evolutionary biology of endometriosis. *Evol Med Public Health.* 2021;9(1):174–91.
  27. Dinsdale NL, Crespi BJ. Endometriosis and polycystic ovary syndrome are diametric disorders. *Evol Appl.* 2021;14(7):1693–715.
  28. Crespi BJ, Evans SF. Prenatal origins of endometriosis pathology and pain: reviewing the evidence of a role for low testosterone. *J Pain Res.* 2023;16:307–16.
  29. Sanchez-Ferrer ML, Mendiola J, Jimenez-Velazquez R, Canovas-Lopez L, Corbalan-Biyang S, Hernandez-Penalver AI, et al. Investigation of anogenital distance as a diagnostic tool in endometriosis. *Reprod Biomed Online.* 2017;34(4):375–82.
  30. Sinnott-Armstrong N, Naqvi S, Rivas M, Pritchard JK. GWAS of three molecular traits highlights core genes and pathways alongside a highly polygenic background. *Elife.* 2021;10: e58615.
  31. McGrath IM, Mortlock S, Montgomery GW. Genetic regulation of physiological reproductive lifespan and female fertility. *Int J Mol Sci.* 2021;22(5):2556.
  32. Von Recklinghausen F. Adenomyomas and cystadenomas of the wall of the uterus and tube: their origin as remnants of the wolffian body. *Wien Klin Wochenschr.* 1896;8:530.
  33. Longo LD. Classic pages in obstetrics and gynecology. Aberrant portions of the mullerian duct found in an ovary: William Wood Russell Johns Hopkins Hospital bulletin, vol. 10, pp. 8–10, 1899. *Am J Obstet Gynecol.* 1979;134(2):225–6.
  34. Cermik D, Selam B, Taylor HS. Regulation of HOXA-10 expression by testosterone in vitro and in the endometrium of patients with polycystic ovary syndrome. *J Clin Endocrinol Metab.* 2003;88(1):238–43.
  35. Zanatta A, Rocha AM, Carvalho FM, Pereira RM, Taylor HS, Motta EL, et al. The role of the Hoxa10/HOXA10 gene in the etiology of endometriosis and its related infertility: a review. *J Assist Reprod Genet.* 2010;27(12):701–10.
  36. Harris HR, Cushing-Haugen KL, Webb PM, Nagle CM, Jordan SJ, Australian Ovarian Cancer Study G, et al. Association between genetically predicted polycystic ovary syndrome and ovarian cancer: a Mendelian randomization study. *Int J Epidemiol.* 2019;48(3):822–30.
  37. Treloar SA, Bell TA, Nagle CM, Purdie DM, Green AC. Early menstrual characteristics associated with subsequent diagnosis of endometriosis. *Am J Obstet Gynecol.* 2010;202(6):534.e1–6.
  38. Sampson JA. Peritoneal endometriosis due to the menstrual dissemination of endometrial tissue into the peritoneal cavity. *Am J Obstet Gynecol.* 1927;14:422–69.
  39. Alio L, Angioni S, Arena S, Bartiromo L, Bergamini V, Berlanda N, et al. Endometriosis: seeking optimal management in women approaching menopause. *Climacteric.* 2019;22(4):329–38.
  40. Thombre Kulkarni M, Shafir A, Farland LV, Terry KL, Whitcomb BW, Eliassen AH, et al. Association between laparoscopically confirmed endometriosis and risk of early natural menopause. *JAMA Netw Open.* 2022;5(1): e2144391.
  41. Tan Z, Gong X, Li Y, Hung SW, Huang J, Wang CC, et al. Impacts of endometrioma on ovarian aging from basic science to clinical management. *Front Endocrinol (Lausanne).* 2022;13:1073261.
  42. Nahar K, Khanam NN, Chowdhury AA, Khan NJ, Mohamed Z. Association of dyslipidemia with endometriosis: a case control study. *Mymensingh Med J.* 2023;32(1):118–24.
  43. Adamyan LV, Starodubtseva N, Borisova A, Stepanian AA, Chagovets V, Salimova D, et al. Direct mass spectrometry differentiation of ectopic and eutopic endometrium in patients with endometriosis. *J Minim Invasive Gynecol.* 2018;25(3):426–33.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

