Identification and Validation of Novel Combinatorial Genetic Risk Factors for Endometriosis across Multiple UK and US Patient Cohorts

Authors: JM Sardell[1], S Das[1], GL Møller[1], M Sanna[1], K Chocian[1], K Taylor[1], AR Malinowski[1], C Stubberfield[1], A Rochlin[2], S Gardner[1 †]

Affiliations:
[1] PrecisionLife Ltd., Unit 8b Bankside, Hanborough Business Park, Long Hanborough OX29 8LJ, UK
[2] Complex Disorders Alliance, 2299 Summer St., Stamford, CT 06905, USA
[†] corresponding author

Abstract

Background Endometriosis affects about 10% of women usually of reproductive age. It often has severe negative impacts on patients' quality of life, but the average time to a definitive diagnosis remains 7-9 years, and there are few effective therapeutic options. Relatively little is known about the genetic drivers of the disease even though heritability of the disease is fairly high. A recent large genome wide association study (GWAS) meta-analysis identified 42 genomic loci associated with risk of endometriosis, but together these explain only 5% of disease variance.

Methods We used the PrecisionLife® combinatorial analytics platform to identify multi-SNP disease signatures significantly associated with endometriosis in a white European UK Biobank (UKB) cohort. We assessed the reproducibility of these multi-SNP disease signatures as well as 35 of the 42 SNPs identified by a recent meta-GWAS study in a multi-ancestry American endometriosis cohort from All of Us (AoU) after controlling for population structure.

Results We identified 1,709 disease signatures, comprising 2,957 unique SNPs in combinations of 2-5 SNPs, that were associated with increased prevalence of endometriosis in UKB. We observed a significant enrichment of these signatures (58-88%, $p<0.04$) that are also positively associated with endometriosis in the AoU cohort, including one 2-SNP signature that is individually significant. Reproducibility rates were greatest for higher frequency signatures, ranging from 80-88% for signatures with greater than 9% frequency ($p<0.01$) in AoU. Encouragingly, the disease signatures also show high reproducibility rates in non-white European AoU sub-cohorts (66-76% , $p<0.04$ for signatures with greater than 4% frequency).

A total of 195 unique SNPs mapping to 100 genes were identified in the high frequency reproducing signatures (>9%). Of these, 4 genes were previously identified in the endometriosis meta-GWAS study and 19 genes have a previous association with endometriosis in OpenTargets[1]. 77 novel genes were identified in this study.

We characterized 9 novel genes that occur at the highest frequency in reproducing signatures and that do not contain any SNPs linked to known GWAS genes, providing new evidence for links between endometriosis and autophagy and macrophage biology. Reproducibility rates, ranging between 73% to 85%. are especially strong for the signatures that contain these 9 genes independently of any SNPs mapping to the meta-GWAS genes. These genes also include several

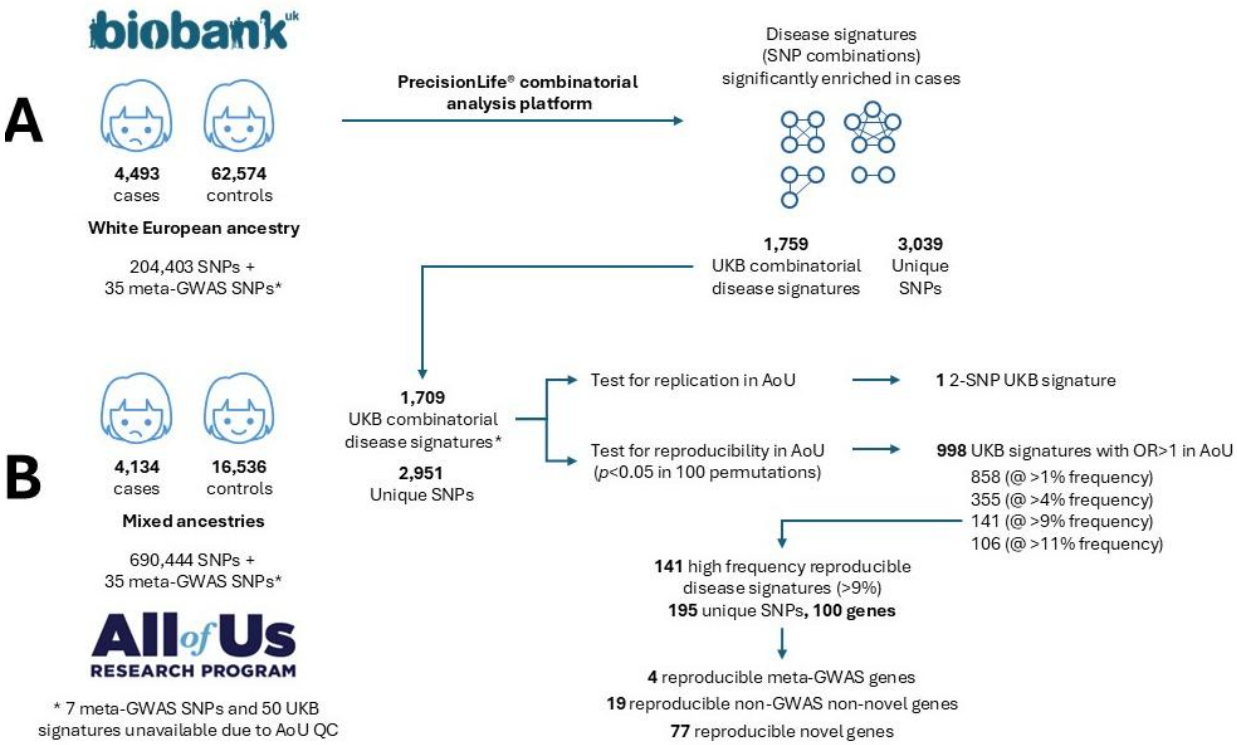
CONFIDENTIAL © PrecisionLife Ltd 2025 All rights reserved

targets novel to endometriosis with credible therapeutic discovery, repurposing and/or repositioning potential.

Conclusion Although using much smaller, less well-characterized datasets than the previous whole genome meta-GWAS study, combinatorial analysis has provided important new insights into the genetics and biology of endometriosis. The finding of 77 novel gene associations that have high frequency and reproduce in an independent, ancestrally diverse dataset demonstrates that combinatorial analysis can identify biologically relevant genes that are overlooked by GWAS approaches. Several of these novel genes will are credible targets for drug discovery and repurposing, as shown by the examples highlighted.

The broad reproducibility of results across datasets and ancestries suggests that combinatorial disease signatures can be used to identify different mechanistic etiologies that have the potential to inform precision medicine-based approaches and generate new clinical treatments for this complex disease.

Graphical Abstract



Graphical abstract. A. Discovery of novel combinatorial genetic associations with endometriosis in a White European patient population from UK Biobank (UKB). B. Analysis of the reproducibility of the UKB disease signatures and 35 of the 42 SNPs identified by a recent meta-GWAS analysis in a mixed ancestry US population from All of Us and identification of 77 novel, high frequency genes associated with endometriosis in both populations.

Keywords: Endometriosis, genetics, reproducibility, combinatorial analytics, women's health

CONFIDENTIAL © PrecisionLife Ltd 2025 All rights reserved

Introduction

Endometriosis Background

Endometriosis is a debilitating and, in many cases, progressive chronic disease that impacts approximately 10% of reproductive age women[2,3]. It typically causes severe pain in afflicted patients and occurs in up to 30–50% of women suffering from infertility[4]. Despite its high prevalence and severe pathophysiology, the average time to diagnosis of endometriosis from onset of symptoms is almost 9 years in the UK, with many affected women scheduling 10+ additional medical consultations before receiving a diagnosis[5].

The long delay between onset of symptoms, diagnosis, and treatment occurs in part because the mechanisms underlying the development of endometriosis and its associated symptoms, such as chronic pain and infertility, are poorly understood[6]. The wide variety of symptoms experienced by those with endometriosis, including chronic pelvic pain, dysmenorrhea, infertility, gastrointestinal, and urological problems, contributes to a high rate of misdiagnosis and missed diagnosis of the disease[7].

Currently, the gold standard for endometriosis diagnosis is surgical confirmation of the presence of endometrial lesions (laparotomy/laparoscopy/transvaginal hydrolaparoscopy)[8]. However, these surgeries are invasive, painful and carry a degree of risk, especially with deep infiltrating disease[9]. They may fail to detect lesions, especially superficial peritoneal and filmy adhesion forms[10], and even when lesions are detected, the severity of different symptoms experienced by patients does not always correlate with the extent of disease pathology observed[6,11]. While the sensitivity and accuracy of non-invasive imaging methods such as MRI and ultrasound for diagnosing the disease are improving, they continue to vary enormously depending on the technique, type, size, and location of lesions, and experience of the operator[12].

Successful treatment of endometriosis is also challenging. Invasive surgery to remove lesions and hormonal therapies (e.g., the contraceptive pill) often have limited symptomatic impact for endometriosis patients[13,14]. Hormone therapies also have significant side effects and are inappropriate for women aiming to conceive[15,16].

The pressing need for more effective treatment options and more accurate, ideally non-invasive, diagnostic tools for endometriosis highlights the importance of obtaining better insight into the biology of this poorly understood disease. A major challenge is that endometriosis is a multi-factorial disease, believed to be caused by complex interactions between genetic, hormonal, and environmental factors[17]. Several hypotheses have been proposed to explain the development of endometriosis, including retrograde menstruation, alterations in the peritoneal fluid, and bacterial infection. However many of these have contradicting evidence, and none are definitive[18].

Endometriosis is also often co-associated with other chronic disorders associated with pain, including migraine, irritable bowel syndrome, interstitial cystitis, and fibromyalgia[19,20]. This may potentially be due to shared underlying biological mechanisms in pain hypersensitivity and inflammatory pathways.

CONFIDENTIAL © PrecisionLife Ltd 2025 All rights reserved

A recent genome-wide association study (GWAS) meta-analysis based on combined whole genome datasets comprised of 60,674 endometriosis cases and 701,926 controls identified 42 genomic loci significantly associated with prevalence of endometriosis including several genes linked to pain perception and maintenance[21]. However, these signals explained only 5% of the disease variance, indicating that much remains unknown about the biology of this complex disease.

Combinatorial Analytics for Complex Diseases

Combinatorial analysis enables hypothesis-free identification of combinations of genetic variants ('disease signatures') that are significantly over- or under-enriched in patients with a disease or other specific phenotype. These studies can be of a case-control design or can use a quantitative trait approach and can find both disease risk and disease resilience factors. The signatures found capture both the linear and non-linear (e.g., epistatic) interactions between multiple genomic loci that reflect the complex disease biology involved in heterogeneous chronic diseases.

Combinatorial disease risk signatures improve our understanding of complex diseases beyond the single SNP associations identified by GWAS[22]. They also create opportunities for clinically actionable diagnostic tests and targeted trials of drug candidates that may provide clinical benefit to selected patient cohorts based on their specific disease drivers[23].

The PrecisionLife® combinatorial analytics platform has been used in multiple studies to identify key genes and genetic mechanisms associated with disease risk and resilience including amyotrophic lateral sclerosis (ALS)[24], myalgic encephalomyelitis / chronic fatigue syndrome (ME/CFS)[25], and long COVID[26]. A significant enrichment of disease signatures from an original UK population long COVID study were shown to be positively correlated with increased risk of long COVID in an ancestrally heterogeneous American cohort from an All of Us (AoU) dataset[27], with up to 83% reproducibility of signatures and 92% reproducibility of genes between the cohorts[28]. These results represented the first demonstration of reproducible genetic signals for both long COVID and ME/CFS, as GWAS approaches have to date offered no validated insights into the biology of these two complex diseases even in much larger populations.

Aims of Study

We first ran combinatorial analysis on a cohort of endometriosis patients and controls from a UK Biobank (UKB) dataset, specifically including 35 of the 42 meta-GWAS SNPs in the analysis. This analysis identified 1,709 disease signatures significantly associated with increased risk of endometriosis, each comprising between 2 to 5 SNP-genotypes. We then sought to demonstrate significant enrichment of reproducing disease associations for these signatures in an independent, diverse-ancestry American endometriosis cohort from All of Us (AoU) for which we had overlapping genotype data. We wished to evaluate the relative reproducibility of the disease associated signatures in sub-cohorts of self-reported Black/African American and Hispanic/Latino patients.

CONFIDENTIAL © PrecisionLife Ltd 2025 All rights reserved

We also sought to calculate the relative reproducibility statistics respectively for meta-GWAS SNPs, combinatorial disease signatures containing meta-GWAS SNPs, and disease signatures made from entirely novel sets of SNPs that had not previously been associated by GWAS or in the literature with endometriosis. We evaluated reproducibility across multiple self-identified race/ethnicity cohorts to assess the applicability of disease signatures to diverse patient populations. Finally, we annotated novel genes from a biological perspective to inform the broader understanding of the biological drivers of endometriosis.

Materials and Methods

Generation of UK Biobank Endometriosis Cohort

We used a cohort of endometriosis patients and healthy controls from UKB to identify disease signatures that are significantly associated with elevated prevalence of endometriosis in a case-control design. Cases were initially defined as any UKB participant with self-reported diagnosis (Data field 20002: data-code 1402) or ICD-10/9 codes (N80.*) of endometriosis. We further removed all patients with adenomyosis (ICD-10 code N80.0 - 'endometriosis of uterus').

To minimize the risk of detecting false positive disease signatures that are indirectly associated with disease due to population substructure rather than disease biology, we filtered the patient cohort to include only patients identified as having 'white British' genetic ancestry. We were unable to perform combinatorial analyses on patients with alternative ancestries due to their low representation and small sample sizes in UKB.

196,188 potential controls matched the following criteria:

- Females with white British genetic ancestry
- Exclude all patients with self-reported diagnosis (Data field 20002: data-code 1402) or ICD-10/9 codes (N80.*) for endometriosis

We randomly selected a subset of controls to produce a 1:10 case-control ratio relative to the number of UKB patients with all forms of endometriosis (N80.*).

The bulk of the study was hypothesis-free. However we also wanted to test the relative reproducibility of the loci identified by the Rahmioglu et al. (2023) meta-GWAS study[21]. Of the 42 significant SNPs identified by the meta-GWAS only 7 were present on the genotyping array used by UKB. To allow us to capture the effects of these missing loci in our combinatorial analysis, we incorporated genotypes for the missing meta-GWAS SNPs using imputed SNP genotype data provided by UKB.

We would not typically conduct combinatorial analysis on imputed genotype data as the systematic genotyping errors associated with imputation compound when considering combinations of SNP genotypes, especially across ancestries. For example, a 2% average misgenotyping error rate for individual SNPs implies that nearly 10% of samples will likely be misgenotyped for at least one of the SNPs in a five-SNP disease signature. However, it was considered that the UKB imputation error rate is lowest for white British samples, and that inclusion of the 35 imputed meta-GWAS SNPs was very unlikely to result in compounded error in this study as they comprise a very small fraction (0.02%) of the total dataset.

CONFIDENTIAL © PrecisionLife Ltd 2025 All rights reserved

In any combinatorial analysis, only a small fraction of the total possible number of disease signatures can be evaluated due to computational and statistical power limitations, and it would be inefficient to devote limited resources to evaluating signatures that could not be further validated using the available datasets. The original intent of our study design was to validate the UKB disease signatures in the Copenhagen Hospital Biobank (CHB)[29]. We had therefore filtered the UKB dataset to only include SNPs that were also present in CHB, which removed 7 of the meta-GWAS SNPs. Unfortunately, this resource was ultimately not available to us for this validation study so instead we used the All of Us (AoU) resource, with the limitations that this imposed on the diminished overlap of SNP genotypes.

We also removed SNPs in the low-recombination MHC region of chromosome 6 and conducted LD pruning in PLINK 1.9 (--indep-pairwise 50 2 0.2). This step maximizes opportunities for discovering unique signal by reducing the risk of the combinatorial analytics algorithm redundantly testing effectively equivalent disease signatures comprised of SNPs in linkage disequilibrium. Any meta-GWAS SNPs removed during pruning were added back to the dataset.

The above steps resulted in a UKB cohort with 4,493 cases, 62,574 controls, and 204,423 SNPs used for disease signature discovery.

The distribution of ages differs considerably between cases and controls, with a stronger skew towards older age UKB participants in controls (Supplementary Figure 1). This finding implies that older UKB participants were less likely to have received an ICD-10 code for endometriosis and that a subset of the 'controls' in our dataset are likely misphenotyped and may represent missed diagnoses due to historic clinical practices. If that is the case, it would result in decreased associations between genetic variants and disease and reduced reproducibility of the results (see Limitations of the Analysis).

Combinatorial Analysis

We ran a combinatorial analysis to identify disease signatures that are significantly enriched in endometriosis cases relative to controls in the UKB discovery cohort. The general methodology behind the hypothesis-free combinatorial analytics approach is described in more detail elsewhere[25,26].

Briefly, the PrecisionLife platform uses a fully deterministic approach to construct and rank disease signatures in 'layers' of increasing combinatorial complexity (i.e., adding new SNP-genotypes to the top-ranked disease signatures identified in previous layer). Statistical validation is performed by comparing the properties of disease signatures and their associated 'networks' identified by the analysis to the properties of signatures and networks identified in combinatorial analyses using randomly permuted versions of the dataset (i.e., ones constructed by randomly shuffling the case-control assignments to remove any biological link between genotype and phenotype). Typically 1,000 cycles of fully random permutation are used.

Generation of All of Us Endometriosis Cohort

We identified a second cohort of female patients diagnosed with endometriosis, along with female controls matched for self-identified ethnicity / race in the All of Us Curated Data Repository (CDR) version 8 (accessed on the 24th of July 2025)[27]. AoU CDR v8 contains data for

CONFIDENTIAL © PrecisionLife Ltd 2025 All rights reserved

over 865,000 American participants including 414,840 patients with whole genome sequencing (WGS) data and 354,400 patients with electronic health record data. Notably, it encompasses a diverse population with a wide age range of 18 to 90 years and substantial representation of non-European ancestry groups, which are frequently underrepresented in genomic research.

Endometriosis cases were identified using the AoU Cohort Browser by selecting all females with whole genome sequencing (WGS) data who had a diagnosis of endometriosis based on ICD9 codes 617.1-9 or ICD-10 codes N80.1-9 and N80.A-D (see Supplementary Table 1). These ICD code criteria are referred to as 'Source Concepts' in AoU. This resulted in 4,134 endometriosis cases.

The AoU control cohort was generated by selecting females with WGS data who do not have any evidence of endometriosis, either based on ICD-9/ICD-10 codes (AoU Source Concepts) or related AoU Standard Concepts (see Supplementary Tables 2 and 3). We also excluded individuals who have a history of procedures such as laparoscopy or any symptomatic phenotypes potentially consistent with undiagnosed endometriosis such as pelvic pain, dysmenorrhea or infertility reported in EHR or surveys (see Supplementary Tables 2 and 3). We used the sex-imputation functionality of PLINK to confirm that all selected case and controls were imputed as genetically female. Applying these criteria, our maximum control population included 191,331 individuals. As in UKB, we observed a similar skew towards higher age among controls than cases in AoU (Supplementary Figure 2).

For each patient, we took whole genome sequence (WGS) data from AoU and extracted all SNP genotypes included in the unfiltered UKB genotype dataset. We then applied the same QC protocol that we used for the UKB cohort. This resulted in 690,444 SNPs available for signature lookups.

To reduce potential confounding effects arising from population substructure, we matched controls to the endometriosis cases based on self-reported race/ethnicity to generate a dataset with a case:control ratio of 1:4. We employed the probabilistic stratified sampling approach described in Sardell *et al.* (2025)[28] to select the subset of 16,536 controls that most closely matches the distribution of demographic subgroups in cases. The demographic breakdown of the AoU endometriosis dataset is included in Supplementary Table 4.

We used principal component analysis (PCA) to model any remaining population substructure within the AoU study cohort. Prior to performing PCA, we first removed all SNPs that are associated with the sex chromosomes, that fall within the MHC region on chromosome 6, or that have minor allele frequency less than 0.05. We then conducted LD-pruning in PLINK 1.9 (--indep-pairwise 50 5 0.2) before generating genetic principal components (PCs) using the --pca command in PLINK 1.9. We selected the top 4 PCs for use in our analyses based on the variance explained by the associated eigenvalues (Supplementary Table 5).

Reproducibility of Disease Associations for Combinatorial Signatures in All of Us

We evaluated the degree to which endometriosis disease signatures identified in a UKB cohort reproduce in AoU, using an approach similar to the one previously used to validate combinatorial disease signatures for long COVID[28]. 50 of the disease signatures identified in UKB could not be evaluated in AoU because one or more of their component SNP genotypes were not included in the latter dataset following QC. These were excluded from the analysis.

CONFIDENTIAL © PrecisionLife Ltd 2025 All rights reserved

First, the association between each disease signature and disease status in AoU was assessed via a logistic regression with absence/presence of the disease signature as the dependent variable alongside the top 4 genetic PCs as covariates. The latter were included as covariates to control for the effects of population substructure on signature frequency.

Using the results of these logistic regressions, we first tested whether any individual signatures are significantly associated with increased risk of endometriosis in AoU, after applying Bonferroni[30] and Benjamini-Hochberg[31] false discovery rate (FDR) adjustments to account for testing of multiple signatures. We then repeated this analysis individually for the output of each 'layer' of the UKB combinatorial analysis, where 'layer' denotes the number of SNP-genotypes in a signature. We also assessed whether any of the 'lead' SNPs identified by the meta-GWAS[21] are significantly associated with endometriosis in our AoU cohort.

Statistically validating replication of disease associations for individual signatures is challenging due to the limited statistical power associated with the relatively small size of the AoU validation cohort. Multi-SNP disease signatures, especially with 4 or 5 SNPs, also by definition have much lower frequency than their individual component SNPs, requiring larger size cohorts to achieve equivalent statistical power. This difference in statistical power is further magnified when the number of disease signatures from combinatorial analysis is orders of magnitude larger than the number of significant loci detected in the meta-GWAS, as is expected when the disease biology reflects a complex network of interacting genes. This necessitates a much more severe FDR adjustment when assessing statistical significance of the former relative to the latter.

Therefore, we also assessed the evidence for broad reproducibility of positive disease associations among the set of endometriosis signatures in AoU. This analysis tests whether the proportion of disease signatures that are positively correlated with endometriosis in AoU is significantly greater than expected for a set of signatures unlinked to disease. This observation suggests that the set of signatures are reflective of disease biology, even if most individual signatures cannot achieve a level of statistical validation due to the small sample size.

First, we counted the fraction of signatures that have a positive coefficient (odds ratio > 1) in the logistic regression. We call this the 'reproducibility rate' for a set of signatures. To test whether the observed enrichment is statistically significant, we randomly reassigned case-control status to the patient cohort while maintaining a fixed case-control ratio, generated logistic regression results in the randomized cohort, and calculated its observed reproducibility rate.

Repeating this process a total of 100 times allowed us to generate a distribution of reproducibility rates under the null hypothesis where disease signatures are unlinked to case-control status. The *p*-value associated with the observed reproducibility data is the number of permutations where the observed reproducibility rate is equal to or greater than the observed reproducibility rate in the AoU endometriosis cohort. This permutation-based approach for calculating *p*-values is most appropriate because the signatures are non-independent due to shared component SNP-genotypes and therefore violate the assumptions of independent observations required by standard statistical tests.

CONFIDENTIAL © PrecisionLife Ltd 2025 All rights reserved

Our previous analysis of long COVID[28] showed that the reproducibility rates for combinatorial disease signatures were positively correlated with the frequency of those signatures within the population. This likely reflects the reduced statistical power for rare signatures, which are more likely to randomly occur at higher frequency in controls than cases due to random sampling even when they are biologically associated with disease. Therefore, we assessed the reproducibility rate for endometriosis disease signatures after applying a set of filters for signature frequency in the AoU cohort.

Cross-ancestry Analysis

We also tested whether the disease signatures identified in the white British UKB cohort reproduced among AoU participants who have self-reported 'Black' and/or 'African American' race/ethnicity as well as patients with self-reported 'Hispanic', 'Latino', and/or 'Latina' race/ethnicity. These cohorts are not fully mutually exclusive, as they represent responses to different questions within AoU.

We again included genetic PCs as covariates to control for indirect relationships between signature frequency and disease prevalence resulting from population substructure. We conducted separate PCAs for each sub-cohort using the approach described above for the whole cohort. We then selected the first two PCs as covariates for each ancestry-specific analysis based on the variance explained by the PCA eigenvalues (see Supplementary Table 5).

SNP - Gene Annotation

SNPs identified in disease signatures were mapped to genes using an annotation cascade process against the human reference genome (GRCh38). SNPs that lie within coding regions of gene(s) were assigned directly to the corresponding gene(s). Remaining SNPs that lie within 2 kb upstream or 0.5 kb downstream of any gene(s) were mapped to the closest gene(s) within this region. Additional gene assignments for identified SNPs using publicly available eQTL data and/or chromatin interaction data for ovary and uterine tissues[32,33].

The mapped genes were annotated using data from over 50 public data sources (see Supplementary Table 6), and their biological relevance analyzed to develop a deeper understanding of the gene target and its potential mechanism of action link to endometriosis phenotypes. All genes were screened for matches against existing drugs in preclinical and clinical development using GlobalData[34]. This was used to identify, evaluate and prioritize existing endometriosis drugs and potential drug repurposing candidates that can be mapped to specific mechanistic patient subgroups identified in this analysis.

Reproducibility of Meta-GWAS Disease Associations in All of Us

For comparison, we also assessed the replicability and reproducibility of 35 of the 42 significant SNPs from the Rahmioglu *et al.* meta-GWAS[21] (hereafter referred to as 'meta-GWAS SNPs'). We used the --logistic command in PLINK 1.9 to generate GWAS results for the AoU cohort, including as covariates the same top 4 genetic PCs used in the reproducibility study for the combinatorial signatures. Although the GWAS implementation in PLINK is not state-of-the-art, it offers an approximate estimate of the degree of reproducible signal from the meta-GWAS SNPs.

CONFIDENTIAL © PrecisionLife Ltd 2025 All rights reserved

A disease association for a GWAS SNP is considered to replicate if its logistic regression additive model *p*-value is statistically significant ($p < 0.05$ after Bonferroni FDR adjustment for multiple tests). A disease association for a GWAS SNP is considered to reproduce if the coefficient from the logistic regression additive model is greater than 0 (i.e., odds ratio > 1).

Results

Results of Combinatorial Analysis of UKB Cohort

Running combinatorial analysis on the UKB population and including the 35 meta-GWAS SNPs, we identified 1,759 multi-SNP signatures significantly associated with endometriosis (UKB disease signatures), containing 3,039 unique SNPs. 1,709 of these signatures could also be assessed in AoU. These include:

- 196 combinations of two SNP-genotypes
- 403 combinations of three SNP-genotypes
- 440 combinations of four SNP-genotypes
- 670 combinations of five SNP-genotypes

Together, these 1,709 combinations include a total of 2,946 unique SNPs.

We then evaluated the frequency and disease association of these UKB disease signatures in AoU. The most common UKB disease signature occurs in 20.1% of AoU participants, and 20% of UKB disease signatures (342/1709) occur in fewer than 1% of AoU participants. Figure 1 presents a histogram of UKB disease signatures by frequency.

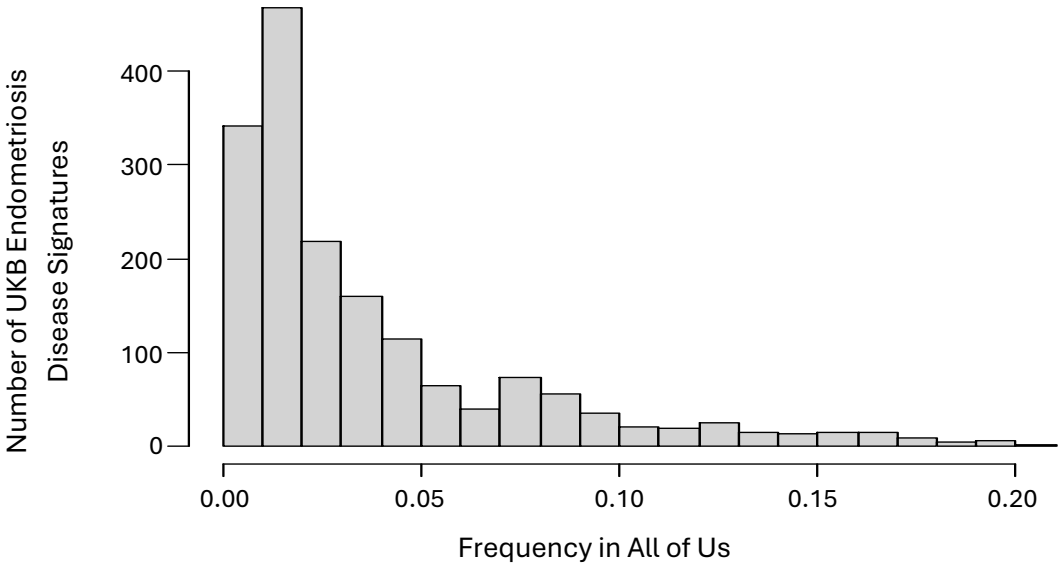


Figure 1. Distribution of UKB endometriosis disease signatures by frequency in AoU

Statistical Validation of Individual Disease Signature in All of Us

When considering the full set of 1,709 disease signatures, none were significantly associated with increased disease risk in AoU after applying Bonferroni or Benjamini-Hochberg FDR

CONFIDENTIAL © PrecisionLife Ltd 2025 All rights reserved

adjustments. However, when we tested just the 196 disease signatures comprised of 2 SNP-genotypes resulting in a less severe FDR correction, one signature was significantly associated with increased disease risk in AoU under both FDR adjustment approaches.

This signature, which has a Benjami-Hochberg adjusted *p*-value = 0.038 and logistic regression odds ratio = 1.21 in AoU, is comprised of the heterozygous genotypes for rs11751190 (G/A) and rs1888328 (T/G). The former SNP is located within an intron of the gene *SYNE1* while the latter SNP is located in intronic regions of multiple lncRNAs. *SYNE1* was also linked to a meta-GWAS SNP, but the *SYNE1* SNP in this signature is not the same as the lead *SYNE1* SNP in the meta-GWAS (rs71575922). No disease signatures were statistically significant when we individually assess 3-, 4-, or 5-SNP genotype signatures.

Reproducibility of Endometriosis Disease Associations in AoU Cohort

Evaluating all 1,709 endometriosis disease signatures, we observed a significant enrichment of signatures that also have odds ratios > 1 in AoU (58.5%). This reproducibility rate trends higher when the signatures are filtered by increasingly greater frequency (Figure 2). For example, reproducibility rates increased to 62.8% when assessing the 1,367 signatures with frequency >1%. The reproducibility rate further increased to 68.1% for the 521 signatures with frequency >4%, 80.1% for the 176 signatures with frequency >9%, and 88.3% for the 120 signatures with frequency >11%. These frequency cutoffs were chosen based on inflection points in reproducibility rates across 1% frequency intervals (e.g., signatures with frequency between 8%-9% vs. 9%-10%). We confirmed that all results are statistically significant, i.e., similar reproducibility rates are observed in fewer than 5% of random permutations (p<0.05 Table 1).

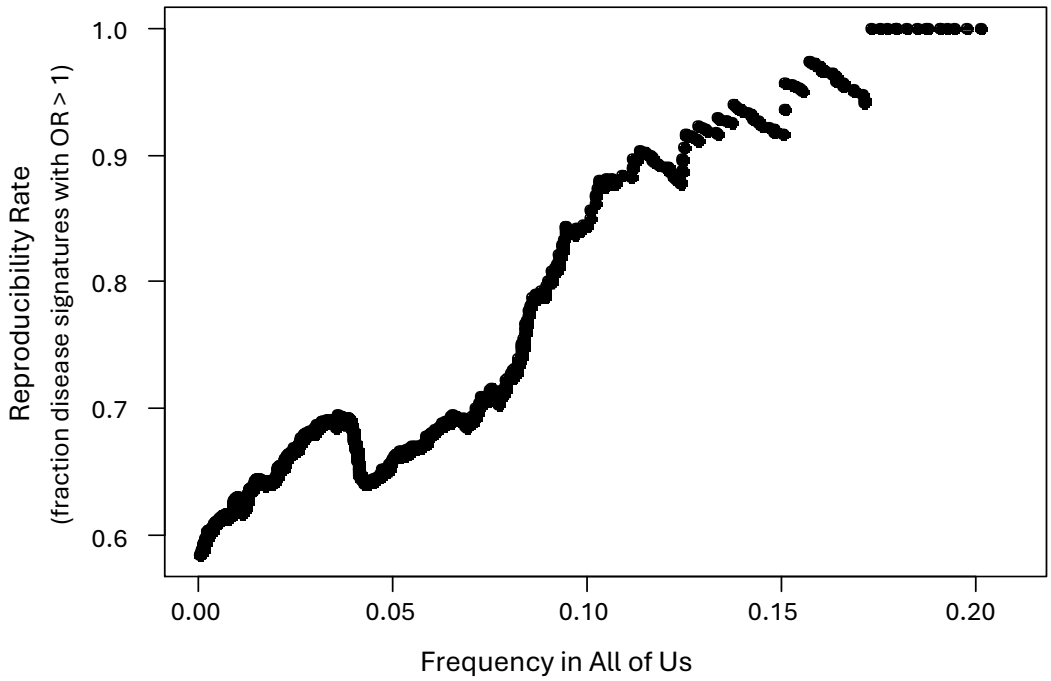


Figure 2. Reproducibility rates for UKB endometriosis disease signatures in AoU under different minimum frequency filters.

CONFIDENTIAL © PrecisionLife Ltd 2025 All rights reserved

Table 1. Reproducibility statistics in AoU for UKB endometriosis disease signatures under different signature frequency thresholds. Bold indicates statistical significance, *p*-values reported as < 0.01 means that the observed reproducibility was reported in 0/100 permutations.

Minimum signature frequency	# signatures tested	% (#) signatures with odds ratio > 1 in AoU	% signatures with odds ratio > 1 in permutations mean (range)	*p*-value
11%	120	88.3% (106)	49.5% (14.9% - 83.5%)	**<0.01**
9%	176	80.1% (141)	49.3% (15.2% - 78%)	**<0.01**
4%	521	68.1% (355)	49.7% (24.7% - 70.7%)	**0.04**
1%	1,367	62.8% (859)	50.4% (33.4% - 64.6%)	**0.02**
none	1,709	58.5% (999)	50.2% (39.4% - 62%)	**0.04**

UKB disease signatures that reproduce in AoU exhibit a range of odds ratios – across all the reproducing signatures regardless of frequency, 46% have odds ratios greater than 1.1 in AoU, 6% have odds ratios greater than 1.3, the maximum odds ratio is 2.37, and the mean odds ratio is 1.12 (Table 2 and Supplementary Figure 3).

We observed a narrower range of odds ratios and lower mean odds ratios after applying frequency filters, as expected given the trade-off between effect size and feature frequency in disease genetics (Supplementary Figure 3). However, we continue to observe a relatively large proportion of reproducing signatures with moderate odds ratios (> 1.1). Among reproducing signatures that occur in more than 1% of individuals, the maximum odds ratio is 1.49, the mean odds ratio is 1.09, and the proportion of signatures with odds ratio greater 1.1 is 40%. Among reproducing signatures that occur in more than 9% of individuals, the maximum odds ratio is 1.21, the mean odds ratio is 1.08, and 40% of signatures have odds ratios greater than 1.1.

In comparison, the mean odds ratio for the set of meta-GWAS SNPs that reproduce in AoU (33 of 35) is 1.06 and just 15% have odds ratios greater than 1.1.

Table 2. Odds ratio (OR) statistics for different reproducing UKB endometriosis signature sets and meta-GWAS SNPs in AoU

Minimum signature frequency	# reproducing signatures	OR > 1.1 # (%)	OR > 1.3 # (%)	Mean OR	Max OR
11%	106	43 (41%)	0 (0%)	1.08	1.21
9%	141	57 (40%)	0 (0%)	1.08	1.21
4%	355	130 (37%)	0 (0%)	1.08	1.30
1%	859	345 (40%)	11 (1%)	1.09	1.49
none	999	458 (46%)	60 (6%)	1.12	2.37
meta-GWAS SNPs	33 SNPs	5 (15%)	0 (0%)	1.06	1.29

CONFIDENTIAL © PrecisionLife Ltd 2025 All rights reserved

Reproducibility in Other Ancestries

Reproducibility rates for all combinatorial UKB disease signatures in self-identified 'Black / African American' and 'Hispanic / Latino' cohorts from AoU are similar to reproducibility rates for the whole cohort (Table 3). We also observe similar reproducibility rates across cohorts for signatures with greater than 1% frequency and greater than 4% frequency. Reproducibility rates for signatures with greater than 9% frequency are lower than the reproducibility rates for the full cohort but still demonstrate strong enrichment of reproducing signatures.

Table 3. Reproducibility statistics under different signature frequency thresholds for UKB endometriosis disease signatures in self-identified 'Black / African American' and 'Hispanic' AoU cohorts compared to All Ancestries

Minimum signature frequency	# signatures tested	Reproducibility Rate % (#) All Ancestries	Reproducibility Rate % (#) Black / African American	Reproducibility Rate % (#) Hispanic / Latino
11%	120	88.3% (106)	72.5% (87)	75.8% (91)
9%	176	80.1% (141)	70.0% (123)	72.1% (127)
4%	521	68.1% (355)	69.3% (361)	65.8% (343)
1%	1367	62.8% (859)	63.2% (864)	59.9% (819)
none	1709	58.5% (999)	59.4% (1015)	57.9% (990)

Replicability and Reproducibility of Meta-GWAS Results

35 of the 42 meta-GWAS SNPs were present in the dataset used in our combinatorial analysis. The identified risk alleles for six of these SNPs (14%), annotated to the genes *CDKN2-BAS1*, *FRMD7*, *LINC00629*, *WNT4*, *MLLT10*, and *KDR*, are also significantly associated with increased risk of endometriosis in AoU based on GWAS results after applying a Bonferroni FDR correction (see Supplementary Table 7). An additional four SNPs, annotated to the genes *CD109*, *ACTL9*, *GDAP1*, and *VEZT* significantly replicate in AoU if we apply a Benjamini-Hochberg FDR correction, corresponding to 29% replication.

The GWAS associations for 21 (60%) of the meta-GWAS SNPs in AoU are not even nominally significant ($p > 0.05$), including 9 SNPs (26%) with p-values > 0.5 in AoU. This lack of statistical replication likely reflects the much smaller sample size of the cohorts for AoU vs. the meta-GWAS, as 94% (33/35) of the meta-GWAS SNPs reproduce with odds ratios greater than 1.

20 percent of the reproducing UKB disease signatures (336/1709) contain at least one of the lead meta-GWAS SNPs, with 17 of the 35 meta-GWAS SNPs present in at least one disease signature (Table 4). Disease signatures assigned to 11 of these 17 SNPs have reproducibility rates greater than 60% in the full AoU cohort, and the overall reproducibility rate for these 336 signatures is 74% (247/336)

If we expand the definition of meta-GWAS signatures to include any signature containing a SNP that is annotated to one of the meta-GWAS genes, the number of meta-GWAS signatures increases to 644 (38% of signatures) with a reproducibility rate of 75%. The number of

CONFIDENTIAL © PrecisionLife Ltd 2025 All rights reserved

associated meta-GWAS genes increases by four, including two genes where the meta-GWAS SNP was not detected in the combinatorial analysis and two where the meta-GWAS SNP was absent from the dataset. The increase in meta-GWAS signatures is largely due to the inclusion of 261 signatures that contain one of three SNPs assigned to *SYNE1* that differ from the lead meta-GWAS SNP. The 80% reproducibility rate for the 261 signatures associated with these alternative *SYNE1* SNPs is higher than the 74% reproducibility rate for signatures associated with the lead *SYNE1* meta-GWAS SNP.

Table 4. Reproducibility of UKB disease signatures linked to meta-GWAS SNPs and their annotated genes

Meta-GWAS gene	Lead meta-GWAS SNP	# signatures with lead meta-GWAS SNP	Reproducibility rate of meta-GWAS SNP signatures in AoU (OR>1)	# Signatures assigned to meta-GWAS gene	Reproducibility rate of meta-GWAS gene signatures in AoU (OR>1)
ABO	rs507666	2	100%	2	100%
ACTL9	rs2967684	2	100%	2	100%
ASTN2	rs10983311	0	n/a	3	67%
CD109	rs4540228	4	100%	4	100%
CDKN2B-AS1	rs10122243	4	25%	30	53%
CEP112	rs7214750	0	n/a	1	100%
DLEU1	rs7334326	1	100%	3	67%
DNM3	rs2421985	0	n/a	3	100%
FRMD7	rs5933091	2	100%	2	100%
GDAP1	rs10090060	1	100%	2	100%
GREB1	rs11674184	3	33%	19	53%
IGF1	rs10860864	3	33%	3	33%
FSHB/ARL14EP	rs3858429	34	56%	35	57%
KDR	rs1903068	73	45%	73	45%
MLLT10	rs10828249	2	50%	2	50%
RIN3	rs57281976	0	n/a	2	50%
RNLS	rs7907732	1	100%	1	100%
SKAP1	rs66683298	14	64%	15	67%
SYNE1	rs71575922	34	74%	295	80%
WNT4/CDC42	rs10917151	156	92%	156	92%
7p12.3	rs55909142	9	67%	10	60%

Reproduced Novel Gene Findings

The combinatorial analysis identified novel associations between SNPs and endometriosis in two contexts: novel SNPs that occur in the same disease signature as at least one meta-GWAS SNP ('meta-GWAS signatures') and novel SNPs that occur in disease signatures containing no meta-GWAS SNPs ('non-meta-GWAS signatures'). SNPs can also occur in a mix of meta-GWAS and non-meta-GWAS signatures.

Higher frequency disease signatures exhibit the strongest evidence for a consistent link to endometriosis in the AoU replication analysis. A total of 195 unique SNPs mapping to 100 genes were identified among the 141 disease signatures that have frequency >9% and that also replicate in AoU (odds ratio > 1). Of these, 4 genes were previously identified in the endometriosis meta-GWAS study and 19 additional genes have a previous association with endometriosis in OpenTargets[35] but were not identified by the meta-GWAS.

CONFIDENTIAL © PrecisionLife Ltd 2025 All rights reserved

This finding of 77 novel gene associations that have high frequency and reproduce in an independent, ancestrally diverse dataset confirms that combinatorial analysis can identify biologically relevant genes that are overlooked by GWAS approaches even when the GWAS uses much larger and higher quality datasets. Many of these novel genes occur in meta-GWAS signatures which exhibit strong reproducibility rates overall, but it is challenging to separate the contribution of meta-GWAS genes vs. novel genes to this reproducibility.

To validate novel genes associated with endometriosis, we focused on the set of 25 SNPs that were found in 25 or more disease signatures (termed 'core' SNPs). This criterion reflects the network geometry-based approach utilized in the PrecisionLife combinatorial analytics platform, which more strongly supports SNPs that are repeatedly linked to disease status in different combinatorial contexts. 20 of these 25 core SNPs were annotated to genes. 6 of those 20 genes overlap with genes identified by the meta-GWAS (the meta-GWAS SNPs annotated to *WNT4/CDC42*, *CDKN2B-AS1*, and *FSHB/ARL14EP*, plus 3 SNPs associated with *SYNE1* including the lead meta-GWAS SNP). Reproducibility statistics for these genes are included in Table 4 and Supplementary Table 7).

Of the remaining 14 annotated core SNPs, 9 most frequently occur (>50%) in non-meta-GWAS disease signatures (i.e., signatures that contain no component SNPs annotated to a meta-GWAS gene). We refer to these 9 SNPs as 'novel core SNPs'. Importantly, all 9 novel core SNPs exhibit a strong enrichment of disease signatures that reproduce in AoU (Table 5). Reproducibility rates for the novel core SNPs are especially strong for non-meta-GWAS signatures, ranging between 73% to 85%. Reproducibility of these non-meta-GWAS signatures unambiguously demonstrates the capability of combinatorial analytics to identify novel signal missed by GWAS.

Five additional core SNPs predominantly co-occur with meta-GWAS SNPs (in >80% of signatures), and as such we cannot evaluate whether reproducibility or lack thereof is due to the core SNP or the meta-GWAS SNP. Of these meta-GWAS signatures, three show high rates of reproducibility, while two show no evidence of reproducibility (Table 5).

Table 5. Reproducibility rates in AoU endometriosis cohort for gene-annotated non-meta-GWAS SNPs that occur in 25 or more disease signatures ('core' SNPs). 'Non-meta GWAS signatures' are UKB disease signatures that contain no SNPs assigned to a meta-GWAS gene.

Gene	# signatures	Reproducibility rate of all signatures in AoU % (#)	# non-meta-GWAS signatures	Reproducibility rate of non-meta-GWAS signatures in AoU % (#)
Occur independently of meta-GWAS signatures				
AC105942.1	25	72% (18)	19	79% (15)
ATG16L1	40	78% (31)	39	77% (30)
CDC14B	47	75% (35)	47	75% (35)
COLEC12	33	85% (28)	33	85% (28)
NEDD4L	33	61% (20)	20	85% (17)
RBMS3	48	73% (35)	48	73% (35)
TMEM71	33	85% (28)	33	85% (28)
VSTM1	50	74% (37)	50	74% (37)
Z82186.1	35	74% (26)	19	74% (14)

CONFIDENTIAL © PrecisionLife Ltd 2025 All rights reserved

Predominantly co-occur with meta-GWAS SNPs				
ATP13A4	27	70% (19)	5	100% (5)
ATP8A2	60	35% (21)	1	0% (0)
CADM1	125	95% (119)	5	20% (1)
PPP2R5C	60	37% (22)	0	n/a
SHISA6	157	83% (130)	1	0% (0)

Discussion

Combinatorial analysis identified 1,709 combinatorial disease signatures that are associated with significantly increased prevalence of endometriosis in a UKB white British cohort. The UKB endometriosis disease signatures exhibit a significant enrichment of reproducing disease signal in AoU. That is, we observe significantly more UKB disease signatures that are also positively correlated with increased prevalence of endometriosis in AoU than expected by chance. The reproducibility rate increases after applying filters based on the frequency of disease signatures in AoU, a pattern previously observed in a reproducibility analysis for long COVID combinatorial disease signatures[28].

These UKB disease signatures include several wholly novel genes that were associated with endometriosis in both UKB and AoU, as well as many of the same SNPs and genes identified by a large meta-GWAS study[21] and many genes previously linked to endometriosis via non-genetic association studies. Over a third of the disease signatures contain at least one SNP assigned to a gene identified in the endometriosis meta-GWAS, and these exhibit high rates of reproducibility. The broader biological context captured by these combinatorial disease signatures provide potential to improve our understanding of disease biology and clinical development tools for novel therapeutics.

Novel Insights into the Biology of Endometriosis

We identified 9 novel genes associated with the set of core SNP genotypes that feature most prominently in non-meta GWAS disease signatures, all of which universally exhibit high degrees of reproducibility in AoU. This subset includes multiple genes associated with autophagy and macrophage biology as well as others that could hold significant potential for novel therapeutics and targeted drug repurposing/repositioning.

Autophagy

Autophagy, a cellular process responsible for degradation and recycling of intracellular components[36], has previously been implicated in the pathophysiology of endometriosis. Suppressed autophagic activity has been observed in ectopic endometrial tissue, facilitating the survival, immune evasion, and proliferation of ectopic cells (reviewed in [37,38,39]). Among the novel core genes identified in our study, *ATG16L1* and *NEDD4L* are of particular interest due to their central roles in autophagy.

ATG16L1 encodes a core autophagy protein that forms a complex with ATG5 and ATG12, acting as an E3-like ligase essential for substrate recognition and autophagosome biogenesis[40]. While *ATG16L1* has not previously been associated with endometriosis, knockdown of interacting

CONFIDENTIAL © PrecisionLife Ltd 2025 All rights reserved

autophagy genes *ATG5* and *ATG7* has been shown to impair decidualization in human endometrial stromal cells[41], suggesting a role in endometrial function and disease.

NEDD4L, codes for an E3 ubiquitin ligase that regulates autophagy by targeting the proteins ULK1 and ASCT2 for ubiquitination and degradation[42]. Moreover, NEDD4L has been reported to inhibit the activity of mTOR, a key negative regulator of autophagy[43]. Although NEDD4L represents a novel association with endometriosis, a related E3 ligase NEDD4 has been shown to inhibit ferroptosis and promote endometrial lesion stromal cell survival by mediating the ubiquitination and degradation of PTGS2[44].

Notably, both *ATG16L1* and *NEDD4L* have previously been implicated in inflammatory bowel disease (IBD)[45,46]. The SNP identified in *ATG16L1* in our endometriosis study has demonstrated a strong association with Crohn's disease in prior GWAS[45]. This finding is potentially interesting given the reported symptomatic overlap between Crohn's disease and endometriosis[47]. Collectively, this evidence suggests that dysregulation of ATG16L1 and NEDD4L may contribute to impaired autophagic homeostasis in endometriosis, potentially supporting lesion development, inflammation, and disease progression.

Macrophage Biology

Although complex, macrophage dysfunction has been increasingly recognized as a contributor to the development and progression of endometriosis[48]. In patients, peritoneal macrophages produce more VEGF than in healthy individuals, and transcriptomic analyses highlight macrophage-derived cytokines as key drivers of inflammation in endometriosis[49,50]. In mouse models, macrophage depletion reduces lesion size, vascularization, and inflammatory pain, underscoring their central role in disease development[51,52]. During menstruation, macrophages clear endometrial debris via scavenger receptor–mediated phagocytosis[53]. Intriguingly, one of the novel core genes identified in our analysis, *COLEC12*, encodes one of these scavenger receptors involved in host defense and phagocytosis.

We also identified *VSTM1* as a core gene, which encodes an immunoglobulin superfamily protein that enhances IL17A secretion by CD4+ T cells[54]. Although *VSTM1* itself has not been studied in endometriosis, IL17A is known to drive pathogenic M2 macrophage polarization in lesions[55]. Both genes show macrophage-enriched expression (Human Protein Atlas[56]), and while their altered expression in endometriosis remains unconfirmed, dysfunction in either may impair phagocytosis, sustain inflammation, or disrupt tissue remodeling, promoting lesion persistence.

Novel Drug Discovery and Repurposing Targets

Alongside the less well-studied potentially druggable genes that were identified, the analyses found several genes whose protein products are the targets for well-established therapeutics that have not previously been explored in the treatment of endometriosis, i.e., repurposing opportunities.

One example of these is the peptidase, angiotensin-converting enzyme (ACE). ACE plays a major role in the renin-angiotensin-aldosterone system[57], which regulates sodium retention by the kidney and blood pressure by converting the peptide hormone angiotensin I (AT1) to

CONFIDENTIAL © PrecisionLife Ltd 2025 All rights reserved

angiotensin II (AT2), resulting in an increase of the vasoconstrictor activity[58]. Inhibitors of the enzyme have been in clinical use since the 1980s for a variety of indications including coronary artery disease, congestive heart failure, chronic kidney disease and myocardial infarction[59].

The product of ACE activity, AT2, enhances degradation of extracellular matrix proteins in cultures of human endometrial stromal cells, while also improving their viability and increasing proliferation and migration[60]. ACE inhibitors (captopril, ramipril) have been shown to reduce the growth of endometrial implants in rat models[61,62]. The expression of the receptors for AT2 (AGTR1, AGTR2) is also known to be altered in endometrial lesions[63,64].

A link between endometriosis and hypertension has previously been identified[65,66,67], but the opportunity to exploit the existing therapeutic modulators of ACE or the angiotensin II receptor inhibitors has not been investigated.

Limitations of the Analysis

Neither UKB nor AoU provides an ideal dataset on which to be performing this analysis. Both have relatively small case populations, they have low genotype overlap, and both datasets have material levels of phenotypic misclassification in both cases and controls as is common in complex diseases. In highly heterogeneous diseases, we have found that using stricter diagnostic criteria—despite risking the exclusion of some genuine cases—is generally preferable to including a larger number of poorly diagnosed (and potentially true negative) individuals in the case group. We also try to screen out undiagnosed (and potentially true positive) patients who may be considered for the control group due to low historic diagnosis rates. Unfortunately, the option for a larger and better characterized dataset was not available to us, which disadvantaged the study.

Reliably identifying patients with endometriosis unfortunately currently relies on surgical confirmation, and such data is not available in UKB or AoU. We relied on ICD-10 coding, which is known to be inconsistently and inaccurately applied. This was further complicated by the age distribution of the cases, which in the case of UKB can be decades past reproductive age. This means that older, potentially less reliable diagnostic criteria may have been in place for these individuals. Indeed, although there is no reason to expect that prevalence of endometriosis should be correlated with age, the age distribution of both UKB and AoU endometriosis cases is heavily skewed towards women with lower ages relative to controls (Supplementary Figures 2 & 3). This is consistent with the observation that endometriosis historically has been poorly recognized resulting in high prevalence of mis-/undiagnosed cases. As a result, a subset of the 'controls' in the UKB and AoU cohorts likely represent misphenotyped patients.

Both of these factors work to reduce the number of UKB disease signatures found and the level of their reproducibility reported in AoU. These types of phenotypic misclassifications typically reduce observed effect sizes by artificially increasing the resemblance between cases and controls[68]. Such misclassification poses a challenge for reproducibility analyses, as signal dilution lowers the statistical power of results[69]. For example, it raises the likelihood that a biologically meaningful disease signature could display an odds ratio below 1 simply because of random sampling (as shown in Sardell et al. 2025[28]).

CONFIDENTIAL © PrecisionLife Ltd 2025 All rights reserved

Accordingly, we anticipate that the high rate of phenotypic misclassification in the datasets will have suppressed the overall signature discovery and reproducibility rates we observed. The reproducibility measures reported here therefore likely represent a lower-bound estimate of the true rate. Using more accurate diagnostic protocols would enable more reliable comparisons across study cohorts.

Statistically validating individual disease signatures in an independent dataset is challenging due to the small size of available cohorts, the large number of signatures, the relative rarity of individual signatures, and the complexity of disease biology. This is illustrated by the fact that fewer than 30% of meta-GWAS SNPs significantly replicate in AoU, despite higher frequencies, much less biological complexity, and orders of magnitude fewer features and consequent need for FDR adjustment to account for multiple tests relative to the replication analysis for disease signatures. In contrast, 94% of the meta-GWAS SNP risk alleles are also positively correlated with increased prevalence of endometriosis in AoU, suggesting that the lack of broad statistical replication is likely due to lack of statistical power provided by the relatively small cohort.

Implications of the Study

The results of this study confirm that combinatorial analysis both recapitulates GWAS results and finds novel disease biology, even when working at a considerable disadvantage due to a much smaller, less well-characterized dataset (4,463 cases with low density genotype data vs 60,674 cases using whole genome data).

The most highly reproducible findings, including across multiple ancestries, come from disease signatures that have high frequency (e.g., above 4%). This is encouraging for clinical applications as signatures with higher frequency offer the potential for increased clinical utility across more patients. It shows that the approach can identify biologically and clinically relevant genes that are overlooked by GWAS even in much larger datasets. This is further evidenced by the association of several genes novel to endometriosis that are druggable either as novel therapeutics or potentially the much faster route of targeted clinical trials of drug repurposing/repositioning candidates. Where safe and well-tolerated compounds can be tested directly in humans as a proof-of-concept study, this offers the fastest route to bringing new and effective medicines to patients.

This is important for endometriosis drug development and for patients. In spite of the huge unmet medical need and commercial potential that comes from this, there has been very little progress in the development of novel endometriosis therapeutics. Aside from chronic underinvestment in women's health, there are several technical and scientific reasons for this – the disease is complex (i.e., has multiple etiologies and influences), highly heterogenous, multi-symptomatic, not correlated with a readily measurable biomarker, and there are no good animal models for the disease pathology. These all work to increase uncertainty and risk for proposed drug discovery projects.

Identifying multiple novel genes and mechanisms that reproduce well across different populations offers new routes into the disease biology. Many of these mechanisms can already be modulated using known active compounds. These can be validated readily using existing drug development candidates (repositioning) or on-market/generic compounds (repurposing) in

CONFIDENTIAL © PrecisionLife Ltd 2025 All rights reserved

small proof-of-concept human studies, guided by genetic biomarkers used to recruit patients whose disease is driven by that mechanism. These would provide rapid and cost-effective clinical validation of the disease modifying potential of the novel targets and the ability to select responders.

These same genetic/mechanistic biomarkers can be used during drug development to build a precision regulatory strategy, and to recruit likely responders into clinical trials, accelerating and derisking clinical development. The same tools may become complementary diagnostics to guide therapy selection in the clinic in due course.

When fully developed and validated, these types of detailed genetic insights could be used to rapidly and non-invasively differentially triage patients presenting with deep pelvic pain, facilitate rapid referral for definitive diagnosis, evaluate patient prognosis for severity and progression to infertility, endometriomas, and other complications, and help select the most appropriate therapies for an individual patient.

Supporting Health Equity

A key challenge for GWAS results and related risk scores is that they are predominantly derived from European cohorts and often fail to transfer across alternative ancestries[70,71,72,73]. It is important in using genomic analysis to demonstrate that the results support health equity across as many patients as possible, including those with different ancestries. With the predominant focus usually being on achieving better representation in the population datasets, the choice of analytical methods is often overlooked as a critical factor in finding transferrable results that reproduce well in different patient subgroups[74].

Although the initial combinatorial analysis was performed on a white British only cohort, the reproducibility results from this study are broadly consistent across all AoU participants regardless of their self-reported ancestry. This observation is crucial as it implies that these disease signatures can be used to inform precision medicine healthcare without adversely affecting historically underrepresented populations and further increasing health disparities.

Conclusions

The high level of reproducibility of combinatorial disease signatures across different ancestries and both novel and known genes is encouraging for furthering the study of endometriosis and its clinical care. It offers much needed new insights into the disease that will be useful to build new diagnostic and therapeutic options for patients.

By focusing on the sets of novel disease signatures with the highest case frequencies and largest combinatorial networks, we demonstrated broad reproducibility of disease signal among many genes not previously linked to endometriosis. This result highlights the promising potential for using combinatorial analysis to identify novel targets for drug discovery or repurposing and to provide targeted genetic/mechanism-based biomarkers as crucial clinical development tools in a complex heterogenous disease that has resisted more traditional drug development approaches.

Work in endometriosis would clearly benefit from using combinatorial analysis approaches on larger datasets with even wider population diversity, more secure diagnosis, more harmonized

CONFIDENTIAL © PrecisionLife Ltd 2025 All rights reserved

health/symptom surveys and deeper whole genome, longitudinal clinical, proteomic and metabolic data. We hope that such datasets will be made available for deeper study.

Much further work is needed in endometriosis; a rapid and accurate test for the disease would enable patients to access timely care and reduce the potential for years of misdiagnosis, suffering and clinical waste. It would also establish a definitive framework for measuring the public health impact of the disease, informing health policy and helping strategically prioritize much needed women's health research initiatives to make more rapid progress in addressing this massive global challenge and improving patients' lives.

Acknowledgements:

Research described in this article has been conducted using data from the All Of Us Research Program Curated Data Repository (CDR) version 8 and UK Biobank (application number 44288) study. We gratefully acknowledge All of Us and UK Biobank participants for their contributions, without whom this research would not have been possible. We also thank the National Institutes of Health's All of Us Research Program and UK Biobank for making available the participant data examined in this study. Special thanks to Gert Møller and Claus Erik Jensen, who initially developed the combinatorial analytics methodology, Jim Kozubek for prior analyses related to endometriosis, and the rest of the PrecisionLife and Metrodora teams.

Author contributions

SG, AR, RG, MS, JS, SD, CS, and KT contributed to the design of the study. JS designed the reproducibility analyses. MP, KC, JS, and SD performed the analyses described in this manuscript. MP and KC conducted the analyses on the Researcher Workbench. All authors contributed to writing the manuscript and consent to publication.

Funding

The project was funded entirely by PrecisionLife Ltd.

Availability of data and materials

Only data from existing All of Us and UK Biobank study cohorts were analyzed and no new source data were collected for this study. Aggregate-level data for the All of Us cohort is publicly available at https://databrowser.researchallofus.org/ (Public Tier dataset). Individual-level data for the All of Us cohort, available in the Controlled Tier dataset, can be analyzed by approved researchers on the Researcher Workbench. UK Biobank data can be accessed by approved registered users.

Ethics approval and consent to participate

Research described in this article has been conducted using data from All Of Us Research Program and UK Biobank (application number 44288).

UK Biobank has approval from the North West Multi-centre Research Ethics Committee (MREC) as a Research Tissue Bank (RTB) approval, and researchers do not require separate ethical clearance and can operate under this RTB approval.

CONFIDENTIAL © PrecisionLife Ltd 2025 All rights reserved

Institutional Reviewing Board (IRB) approval was obtained prior to enrollment of patients in the All of Us Research Program. Informed consent for all participants is conducted in person or through an eConsent platform that includes primary consent, HIPAA Authorization for Research use of EHRs and other external health data, and Consent for Return of Genomic Results. The protocol was reviewed by the Institutional Review Board (IRB) of the All of Us Research Program (IRB Approval Date: Dec 03, 2021). The All of Us IRB follows the regulations and guidance of the NIH Office for Human Research Protections for all studies, ensuring that the rights and welfare of research participants are overseen and protected uniformly. The All of Us Research Program is supported by the National Institutes of Health, Office of the Director: Regional Medical Centers (OT2 OD026549; OT2 OD026554; OT2 OD026557; OT2 OD026556; OT2 OD026550; OT2 OD 026552; OT2 OD026553; OT2 OD026548; OT2 OD026551; OT2 OD026555); Inter agency agreement AOD 16037; Federally Qualified Health Centers HHSN 263201600085U; Data and Research Center: U2C OD023196; Genome Centers (OT2 OD002748; OT2 OD002750; OT2 OD002751); Biobank: U24 OD023121; The Participant Center: U24 OD023176; Participant Technology Systems Center: U24 OD023163; Communications and Engagement: OT2 OD023205; OT2 OD023206; and Community Partners (OT2 OD025277; OT2 OD025315; OT2 OD025337; OT2 OD025276). Results reported are in compliance with the All of Us Data and Statistics Dissemination Policy disallowing disclosure of group counts under 20 to protect participant privacy.

Declaration of Competing Interest

AR is an employee of the Complex Disorders Alliance, SG is co-chair of the Complex Disorder Alliance's Scientific Advisory Board. JS, SD, KT, KC, MP, CS, AM and SG are employees of PrecisionLife Ltd. S.G. is a shareholder of PrecisionLife, Ltd.

CONFIDENTIAL © PrecisionLife Ltd 2025 All rights reserved

References

[1] Ochoa D, Hercules A, Carmona M, Suveges D, *et al*. The next-generation open targets platform: reimagined, redesigned, rebuilt. *Nucleic Acids Res*. 2023;51:D1353–9. doi: 10.1093/nar/gkac1046.

[2] World Health Organization (24 March 2023) Endometriosis Fact Sheet https://www.who.int/news-room/fact-sheets/detail/endometriosis accessed 3 August 2025.

[3] Rogers PA, D'Hooghe TM, Fazleabas A, et al. Priorities for endometriosis research: recommendations from an international consensus workshop. *Reprod Sci* 2009;16(4):335-46. doi: 10.1177/1933719108330568.

[4] Meuleman C, Vandenabeele B, Fieuws S, Spiessens C, Timmerman D, D'Hooghe T. High prevalence of endometriosis in infertile women with normal ovulation and normospermic partners. *Fertil Steril* 2009;92(1):68-74. doi: 10.1016/j.fertnstert.2009.11.027.

[5] "Dismissed, ignored and belittled" The long road to endometriosis diagnosis in the UK, Endometriosis UK Diagnosis Report, March 2024 https://www.endometriosis-uk.org/sites/default/files/2024-03/Endometriosis%20UK%20diagnosis%20survey%202023%20report%20March.pdf, accessed 3 August 2025.

[6] Smolarz B, Szyłło K, Romanowicz H. Endometriosis: Epidemiology, classification, pathogenesis, treatment and genetics (review of literature). *Int J Mol Sci*. 2021;22(19):10554. doi: 10.3390/ijms221910554.

[7] Zondervan KT, Becker CM, Missmer SA. Endometriosis. *N Engl J Med*. 2020;382(13):1244-1256. doi: 10.1056/NEJMra1810764.

[8] NICE 2017 Endometriosis: diagnosis and management https://www.nice.org.uk/guidance/ng73/chapter/recommendations, accessed 3 August 2025

[9] Byrne D, Curnow T, Smith P, Cutner A, Saridogan E, Clark TJ; BSGE Endometriosis Centres. Laparoscopic excision of deep rectovaginal endometriosis in BSGE endometriosis centres: a multicentre prospective cohort study. *BMJ Open*. 2018;8(4):e018924. doi: 10.1136/bmjopen-2017-018924.

[10] Gordts S, Puttemans P, Segaert I, Valkenburg M, Schutyser V, Campo R, Gordts S. Diagnosis and treatment of early-stage endometriosis by Transvaginal Hydro laparoscopy. *Facts Views Vis Obgyn*. 2023;15(1):45-52. doi: 10.52054/FVVO.15.1.057.

[11] Simko S, Wright KN The future of diagnostic laparoscopy. *Cons Reprod Fertil.*, 2022;*3*(2):R91–R95. doi: 10.1530/RAF-22-0007.

[12] Quesada J, Härmä K, Reid S, Rao T, Lo G, Yang N, Karia S, Lee E, Borok N. Endometriosis: A multimodal imaging review. Eur J Radiol. 2023;158:110610. doi: 10.1016/j.ejrad.2022.110610.

[13] Abbott JA, Hawe J, Clayton RD, Garry R. The effects and effectiveness of laparoscopic excision of endometriosis: a prospective study with 2-5 year follow-up. *Hum Reprod*. 2003;18(9):1922-7. doi: 10.1093/humrep/deg275.

[14] Brown J, Crawford TJ, Datta S, & Prentice A. Oral contraceptives for pain associated with endometriosis. *Cochrane Database Syst. Rev.*, 2018;*5*(5):CD001019. doi: 10.1002/14651858.CD001019.pub3.

[15] Vannuccini S, Clemenza S, Rossi M, Petraglia F. Hormonal treatments for endometriosis: The endocrine background. *Rev Endocr Metab Disord*. 2022;23(3):333-355. doi: 10.1007/s11154-021-09666-w.

[16] Somigliana E, Vigano' P, Invernici D, Fornelli G, Merli CEM, Vercellini P. Risk of endometriosis progression in infertile women trying to conceive naturally or using IVF. *Hum Reprod*. 2025;40(7):1249-1256. doi: 10.1093/humrep/deaf090.

[17] Malvezzi H, Marengo EB, Podgaec S, et al. Endometriosis: current challenges in modeling a multifactorial disease of unknown etiology. *J Transl Med* 2020;18:311. doi: 10.1186/s12967-020-02471-0.

CONFIDENTIAL © PrecisionLife Ltd 2025 All rights reserved

[18] Burney RO, Giudice LC. Pathogenesis and pathophysiology of endometriosis. *Fertil Steril*. 2012;98(3):511-9. doi: 10.1016/j.fertnstert.2012.06.029.

[19] Karp BI, Stratton P. Endometriosis-associated chronic pelvic pain. *Med*. 2023;4(3):143-146. doi: 10.1016/j.medj.2023.02.006.

[20] Eisenberg VH, Decter DH, Chodick G, Shalev V, Weil C. Burden of endometriosis: Infertility, comorbidities, and healthcare resource utilization. *J Clin Med*. 2022;11(4):1133. doi: 10.3390/jcm11041133.

[21] Rahmioglu N, Mortlock S, Ghiasi M, et al. The genetic basis of endometriosis and comorbidity with other pain and inflammatory conditions. *Nat Genet*. 2023;55(3):423-436. doi: 10.1038/s41588-023-01323-z.

[22] Gardner S. Combinatorial analytics: an essential tool for the delivery of precision medicine and precision agriculture. *Artif Intell Life Sci*. 2021;1: 100003. doi: 10.1016/j.ailsci.2021.100003.

[23] Das S, Taylor K, Beaulah S, Gardner S. Systematic indication extension for drugs using patient stratification insights generated by combinatorial analytics. *Patterns*. 2022;3(6):100456. doi: 10.1016/j.patter.2022.100496.

[24] Sardell J, Das S, Stubberfield C, Malinowski A, Strivens M, Gardner S. Actively protective combinatorial analysis *Artif Intell Life Sci.*, 2025;7:100125. doi: 10.1016/j.ailsci.2025.100125.

[25] Das S, Taylor K, Kozubek J, Sardell J, Gardner S. Genetic risk factors for ME/CFS identified using combinatorial analysis. *J Transl Med*. 2022;20(1):598. doi: 10.1186/s12967-022-03815-8.

[26] Taylor K, Pearson M, Das S, Sardell J, Chocian K, Gardner S. Genetic risk factors for severe and fatigue dominant long COVID and commonalities with ME/CFS identified by combinatorial analysis. *J Transl Med*. 2023;21:775. doi: 10.1186/s12967-023-04588-4.

[27] The All of Us Research Program Genomics Investigators. Genomic data in the All of Us research program. *Nature*. 2024;627:340–6. doi: 10.1038/s41586-023-06957-x.

[28] Sardell J, Pearson M, Chocian K, Das S, Taylor K, Strivens M, Gupta R, Rochlin A, Gardner S. Reproducibility of genetic risk factors identified for long COVID using combinatorial analysis across US and UK patient cohorts with diverse ancestries. *J Transl Med*. 2025;23(1):516. doi: 10.1186/s12967-025-06535-x.

[29] Sørensen E, Christiansen L, Wilkowski B, et al. Data Resource Profile: The Copenhagen Hospital Biobank (CHB). *Int J Epidemiol*. 2021;50(3):719-720e. doi: 10.1093/ije/dyaa157. PMID: 33169150.

[30] Neyman J, Pearson ES. On the use and interpretation of certain test criteria for purposes of statistical inference. *Biometrika*. 1928;20A:175–240. doi: 10.3238/arztebl.m2024.0113.

[31] Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple hypothesis testing. *J R Stat Soc B*. 1995;57:289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x.

[32] Fung JN, Girling JE, Lukowski SW, et al. The genetic regulation of transcription in human endometrial tissue. *Hum Reprod*. 2017;32:893–904. doi: 10.1093/humrep/dex006. PMID: 28177073.

[33] GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. Nat Genet. 2013;45(6):580-5. doi: 10.1038/ng.2653.

[34] GlobalData Pharma Market Data and Insights available from https://www.globaldata.com/industries-we-cover/pharmaceutical/ Accessed 6 Oct 2024.

[35] Ochoa D, Hercules A, Carmona M, Suveges D, *et al*. The next-generation open targets platform: reimagined, redesigned, rebuilt. *Nucleic Acids Res*. 2023;51:D1353–9. doi: 10.1093/nar/gkac1046.

[36] Klionsky DJ, Petroni G, Amaravadi RK, et al. Autophagy in major human diseases. *EMBO J*. 2021;40(19):e108863. doi: 10.15252/embj.2021108863.

CONFIDENTIAL © PrecisionLife Ltd 2025 All rights reserved

[37] Kobayashi H, Imanaka S, Yoshimoto C, Matsubara S, Shigetomi H. Molecular mechanism of autophagy and apoptosis in endometriosis: Current understanding and future research directions. *Reprod Med Biol*. 2024;23(1):e12577. doi: 10.1002/rmb2.12577.

[38] Yang HL, Mei J, Chang KK, Zhou WJ, Huang LQ, Li MQ. Autophagy in endometriosis. *Am J Transl Res*. 2017;9(11):4707-4725.

[39] Zhan L, Li J, Wei B. Autophagy in endometriosis: Friend or foe? *Biochem Biophys Res Commun*. 2018;495(1):60-63. doi: 10.1016/j.bbrc.2017.10.145.

[40] Mestre Citrinovitz AC, Strowitzki T, Germeyer A. Decreased autophagy impairs decidualization of human endometrial stromal cells: a role for ATG Proteins in endometrial physiology. *Int J Mol Sci*. 2019;20(12):3066. doi: 10.3390/ijms20123066.

[41] Gammoh N. The multifaceted functions of ATG16L1 in autophagy and related processes. *J Cell Sci*. 2020;133(20):jcs249227. doi: 10.1242/jcs.249227.

[42] Lee DE, Yoo JE, Kim J, Kim S, Kim S, Lee H, Cheong H. NEDD4L downregulates autophagy and cell growth by modulating ULK1 and a glutamine transporter. *Cell Death Dis*. 2020;11(1):38. doi: 10.1038/s41419-020-2242-5.

[43] Li G, Song Z, Wu C, Li X, Zhao L, Tong B, Guo Z, Sun M, Zhao J, Zhang H, Jia L, Li S, Wang L. Downregulation of NEDD4L by EGFR signaling promotes the development of lung adenocarcinoma. *J Transl Med*. 2022;20(1):47. doi: 10.1186/s12967-022-03247-4.

[44] Lao J, Hu P, Wan Y, Shu M, Chen J, Yang H. *NEDD4* knockdown suppresses human endometrial stromal cell growth and invasion by regulating PTGS2-mediated ferroptosis in endometriosis. *Curr Mol Med*. 2024. doi: 10.2174/0115665240311438241011052341.

[45] de Lange KM, Moutsianas L, Lee JC, Lamb CA, Luo Y, Kennedy N, et al. Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nat Genet*. 2017;49(2):256-261. doi: 10.1038/ng.3760.

[46] Liang J, Wang N, Yao Y, Wang Y, An X, Wang H, et al. NEDD4L mediates intestinal epithelial cell ferroptosis to restrict inflammatory bowel diseases and colorectal tumorigenesis. *J Clin Invest*. 2024;135(3):e173994. doi: 10.1172/JCI173994.

[47] Chiaffarino F, Cipriani S, Ricci E, Roncella E, Mauri PA, Parazzini F, Vercellini P. Endometriosis and inflammatory bowel disease: A systematic review of the literature. *Eur J Obstet Gynecol Reprod Biol*. 2020;252:246-251. doi: 10.1016/j.ejogrb.2020.06.051.

[48] Hogg C, Panir K, Dhami P, Rosser M, Mack M, Soong D, Pollard JW, Jenkins SJ, Horne AW, Greaves E. Macrophages inhibit and enhance endometriosis depending on their origin. *Proc Natl Acad Sci USA*. 2021;118(6):e2013776118. doi: 10.1073/pnas.2013776118.

[49] McLaren J, Prentice A, Charnock-Jones DS, Millican SA, Müller KH, Sharkey AM, Smith SK. Vascular endothelial growth factor is produced by peritoneal fluid macrophages in endometriosis and is regulated by ovarian steroids. *J Clin Invest*. 1996;98(2):482-9. doi: 10.1172/JCI118815.

[50] Beste MT, Pfäffle-Doyle N, Prentice EA, Morris SN, Lauffenburger DA, Isaacson KB, Griffith LG. Molecular network analysis of endometriosis reveals a role for c-Jun-regulated macrophage activation. *Sci Transl Med*. 2014;6(222):222ra16. doi: 10.1126/scitranslmed.3007988.

[51] Bacci M, Capobianco A, Monno A, Cottone L, Di Puppo F, Camisa B, et al. Macrophages are alternatively activated in patients with endometriosis and required for growth and vascularization of lesions in a mouse model of disease. Am J Pathol. 2009;175(2):547-56. doi: 10.2353/ajpath.2009.081011.

[52] Forster R, Sarginson A, Velichkova A, Hogg C, Dorning A, Horne AW, Saunders PTK, Greaves E. Macrophage-derived insulin-like growth factor-1 is a key neurotrophic and nerve-sensitizing factor in pain associated with endometriosis. *FASEB J*. 2019;33(10):11210-11222. doi: 10.1096/fj.201900797R.

CONFIDENTIAL © PrecisionLife Ltd 2025 All rights reserved

[53] Park M, Kim YS, Song H. Macrophages: a double-edged sword in female reproduction and disorders. *Exp Mol Med*. 2025;57(2):285-297. doi: 10.1038/s12276-025-01392-6.

[54] Guo X, Zhang Y, Wang P, Li T, Fu W, Mo X, Shi T, Zhang Z, Chen Y, Ma D, Han W. VSTM1-v2, a novel soluble glycoprotein, promotes the differentiation and activation of Th17 cells. Cell Immunol. 2012;278(1-2):136-42. doi: 10.1016/j.cellimm.2012.07.009.

[55] Miller JE, Ahn SH, Marks RM, Monsanto SP, Fazleabas AT, Koti M, Tayade C. IL-17A modulates peritoneal macrophage recruitment and M2 polarization in endometriosis. *Front Immunol*. 2020;11:108. doi: 10.3389/fimmu.2020.00108.

[56] Uhlén M, Fagerberg L, Hallström BM, et al. Proteomics. Tissue-based map of the human proteome. *Science*. 2015;347(6220):1260419. doi: 10.1126/science.1260419.

[57] Wu CH, Mohammadmoradi S, Chen JZ, Sawada H, Daugherty A, Lu HS. Renin-Angiotensin System and Cardiovascular Functions. *Arterioscler Thromb Vasc Biol*. 2018;38(7):e108-e116. doi: 10.1161/ATVBAHA.118.311282.

[58] Wong MKS. Angiotensin Converting Enzymes. *Handbook of Hormones*. 2016:263–e29D-4. doi: 10.1016/B978-0-12-801028-0.00254-3.

[59] Singh B, Cusick AS, Goyal A, Patel P. ACE Inhibitors. 2025 May 4. In: *StatPearls [Internet]. Treasure Island (FL)*: StatPearls Publishing; 2025

[60] Zanon P, Terraciano PB, Quandt L, Palma Kuhl C, Pandolfi Passos E, Berger M. Angiotensin II - AT1 receptor signalling regulates the plasminogen-plasmin system in human stromal endometrial cells increasing extracellular matrix degradation, cell migration and inducing a proinflammatory profile. *Biochem Pharmacol*. 2024;225:116280. doi: 10.1016/j.bcp.2024.116280.

[61] Oktem M, Ozcan P, Erdem O, Karakaya C, Cenksoy C, Guner H, Karabacak O, Dursun P. The effect of captopril on endometriotic implants in a rat model. *Eur J Obstet Gynecol Reprod Biol*. 2014;180:120-5. doi: 10.1016/j.ejogrb.2014.06.028.

[62] Mazumdar P, Biswas SS. Ramipril ameliorates endometriosis by inducing oxidative stress-mediated apoptosis in the wistar rat. *J Mol Histol*. 2025;56(2):117. doi: 10.1007/s10735-025-10397-4.

[63] Nakao T, Chishima F, Sugitani M, Tsujimura R, Hayashi C, Yamamoto T. Expression of Angiotensin II Types 1 and 2 Receptors in Endometriotic Lesions. *Gynecol Obstet Invest*. 2017;82(3):294-302. doi: 10.1159/000447591.

[64] Zhang Z, Yuan Y, He L, Yao X, Chen J. Involvement of angiotensin II receptor type 1/NF-κB signaling in the development of endometriosis. *Exp Ther Med*. 2020;20(4):3269-3277. doi: 10.3892/etm.2020.9071.

[65] Mu F, Rich-Edwards J, Rimm EB, Spiegelman D, Forman JP, Missmer SA. Association between endometriosis and hypercholesterolemia or hypertension. *Hypertension*. 2017;70(1):59-65. doi: 10.1161/HYPERTENSIONAHA.117.09056.

[66] Parsa S, Noroozpoor R, Dehghanbanadaki H, Khateri S, Moradi Y. Endometriosis and risk of cardiovascular disease: a systematic review and meta-analysis. *BMC Public Health*. 2025;25(1):245. doi: 10.1186/s12889-025-21486-0.

[67] Sharifipour F, Mohaghegh Z, Javanbakht Z, Siahkal SF, Azizi F. The relationship between hypertensive disorders in pregnancy and endometriosis: a systematic review and meta-analysis. *Reprod Health*. 2024;21(1):91. doi: 10.1186/s12978-024-01833-x.

[68] Beesley LJ, Mukherjee B. Statistical inference for association studies using electronic health records: handling both selection bias and outcome misclassification. *Biometrics*. 2022;78(1):214–26. https://doi.org/10.1111/biom.13400.

[69] Burstein D, Hoffman G, Mathur D, Venkatesh S, Therrien K, Fanous AH, Bigdeli TB, Harvey PD, Roussos P, Voloudakis G. Detecting and adjusting for hidden biases due to phenotype misclassification in genome-wide association studies. medRxiv. 2023. https://doi.org/10.1101/2023.01.17.23284670.

CONFIDENTIAL © PrecisionLife Ltd 2025 All rights reserved

[70] Tam V, Patel N, Turcotte M, Bossé Y, Paré G, Meyre D. Benefits and limitations of genome-wide association studies. *Nat Rev Genet*. 2019;20(8):467–84. https://doi.org/10.1038/s41576-019-0127-1.

[71] Moreno-Grau S, Vernekar M, Lopez-Pineda A, Mas-Montserrat D, et al. Polygenic risk score portability for common diseases across genetically diverse populations. *Hum Genomics*. 2024;18(1):93. doi: 10.1186/s40246-024-00664-y.

[72] Peterson RE, Kuchenbaecker K, Walters RK, Chen CY, *et al*. Genome-wide association studies in ancestrally diverse populations: opportunities, methods, pitfalls, and recommendations. *Cell*. 2019;179:589-603. doi: 10.1016/j.cell.2019.08.051.

[73] Kachuri L, Chatterjee N, Hirbo J, Schaid DJ, Martin I, *et al*. Principles and methods for transferring polygenic risk scores across global populations. *Nat Rev Genet*. 2024;25:8-25. doi: 10.1038/s41576-023-00637-2.

[74] Lehmann B, Bräuninger L, Cho Y, Falck F, Jayadeva S, Katell M, Nguyen T, Perini A, et al. Methodological opportunities in genomic data analysis to advance health equity. *Nat Rev Genet*. 2025. doi: 10.1038/s41576-025-00839-w. doi: 10.1038/s41576-023-00637-2.

CONFIDENTIAL © PrecisionLife Ltd 2025 All rights reserved