Identification and Reproducibility of Novel Combinatorial Genetic Risk Factors for Endometriosis across 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: Although endometriosis affects approximately 10% of women, timely diagnosis and effective treatments remain out of reach for many patients. Relatively little is known about the genetic drivers of endometriosis even though its heritability is around 50%. A recent large genome wide association study meta-analysis (meta-GWAS) 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 disease signatures (i.e., unique combinations of 2-5 SNPs) that were significantly associated with endometriosis in a White British UK Biobank (UKB) cohort. We assessed the reproducibility of these disease signatures, as well 35 of the 42 previously reported meta-GWAS SNPs, in a multi-ancestry US endometriosis cohort from All of Us (AoU) after controlling for population structure. To ensure that the results do not disfavor historically underserved patient populations, we explicitly evaluated the reproducibility of these signatures across patient groups with different self-reported race/ethnicities. Finally, we characterized drug repurposing opportunities for a subset of the novel candidate genes.

Results: We identified 1,709 disease signatures, comprising 2,957 unique SNPs, that were significantly associated with increased prevalence of endometriosis in the White British UKB cohort. Reproducibility rates in all AoU participants were significantly higher than random and highest for higher frequency disease signatures, ranging from over 68% for signatures with greater than 4% frequency up to 80-88% for signatures with greater than 9% frequency. Reproducibility rates remained broadly unchanged for most frequency bins within self-identified Black/African American and Hispanic/Latino AoU sub-cohorts. Pathways enriched in reproducible disease signatures included cell adhesion, proliferation and migration, cytoskeleton remodeling, and angiogenesis as well as biological processes involved in fibrosis and neuropathic pain.

38% of signatures (644 / 1709) mapped to at least one gene identified in the meta-GWAS, including 20% of signatures (336 / 1709) that contained at least one meta-GWAS SNP. The remaining disease signatures (729 / 1709) were completely novel. Several novel signatures contained SNPs that mapped to genes that were previously associated with endometriosis but not identified by the meta-GWAS. We identified 75 entirely novel candidate genes, of which 25 are tractable drug targets. Of the 25 top SNPs ranked based on frequency in signatures, we identified 9 genes that primarily occurred in entirely novel signatures, all of

CONFIDENTIAL © PrecisionLife Ltd 2025 All rights reserved

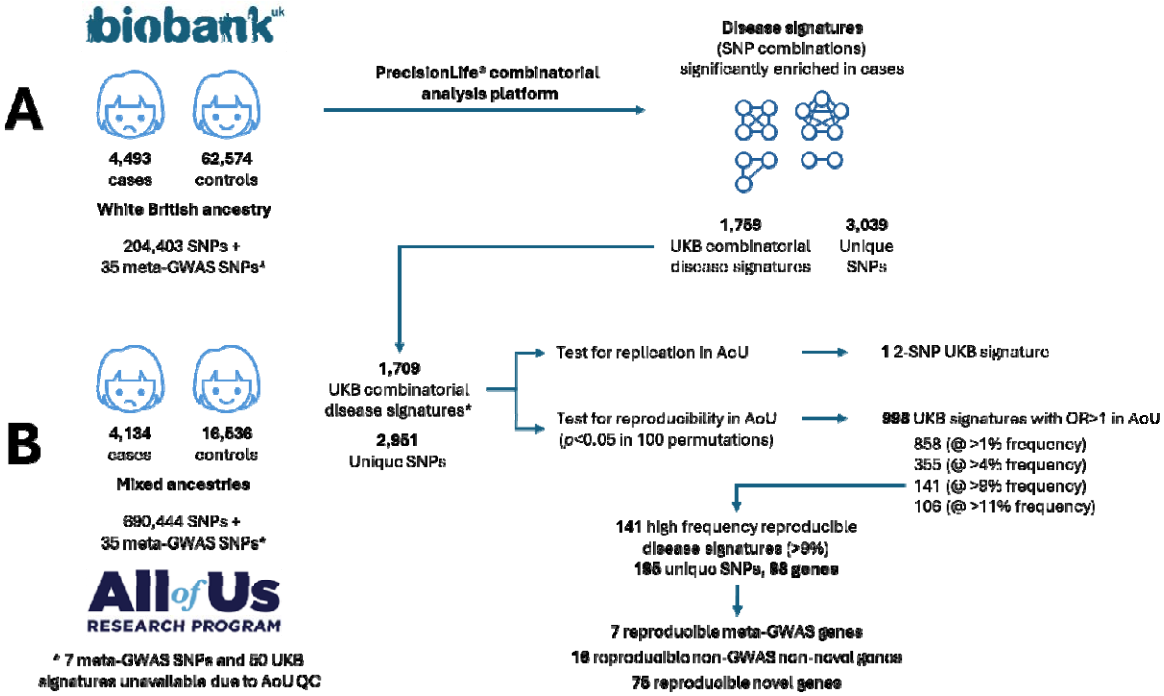
which demonstrated high rates of reproducibility. These 9 genes provide support for hypothesized links between endometriosis and autophagy and macrophage biology.

Conclusions: This study identified a large number of reproducible novel disease signatures, SNPs, and novel candidate genes associated with increased prevalence of endometriosis. It provides support for prior findings from meta-GWAS and other studies linking specific genes to endometriosis. These findings suggest the potential to inform new therapeutic or targeted drug repurposing/repositioning programs aimed at development of new and better treatment options for patients.

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

Lay Summary: Endometriosis affects about one in 10 women, usually between the ages of 15 and 49. It is difficult to diagnose and hard to treat as there are very few effective drugs available. Previous research has found some genes linked to the disease, but we wanted to find more so that better new drugs can be designed that will help more patients. By using a new way to study the disease, we found 75 more genes linked to endometriosis. We believe many of these could be targeted either by new drugs or by existing drugs that were originally developed for another disease. We also showed that these are likely to work in people from different communities, so they could benefit as many patients as possible.

Graphical Abstract



Graphical abstract. A. Discovery of novel combinatorial genetic associations with endometriosis in a White British 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 75 novel, high frequency candidate genes associated with endometriosis in both populations.

Introduction

Endometriosis is a debilitating and, in many cases, progressive chronic disease that impacts approximately 10% of reproductive age women (1,2). Despite its high prevalence and severe pathophysiology, the average time to diagnosis of endometriosis from onset of symptoms is almost nine years in the UK, and treatments such as invasive surgery and hormonal therapies often have limited symptomatic impact (3,4) and significant side effects (5,6).

The long delay between onset of symptoms, diagnosis, and effective treatment occurs in part because the mechanisms underlying the development of endometriosis and its associated symptoms and comorbidities, such as chronic pain and infertility, are poorly understood (7). 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 (8).

A major challenge for research is that endometriosis is a multi-factorial disease, believed to be caused by complex interactions between genetic, hormonal, and environmental factors (9). A recent genome-wide association study (GWAS) meta-analysis based 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 (10). However, these signals only explained at most 5% of the disease variance, indicating that much remains unknown about the biology of this complex disease.

Many affected women schedule ten or more medical consultations before receiving an endometriosis diagnosis (11). The current gold standard for diagnosis is surgical confirmation of the presence of endometrial lesions (12), but these surgeries are invasive, painful and carry a degree of risk (13). While the sensitivity and accuracy of non-invasive imaging methods 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 (14). Thus, there is pressing need for more effective treatment options and more accurate, ideally non-invasive, diagnostic tools for endometriosis.

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. The disease signatures, which are comprised of between 1 to 5 SNP-genotypes, 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. By capturing this additional signal, combinatorial disease risk signatures expand our understanding of complex diseases beyond the single SNP associations identified by GWAS (15).

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) (16), myalgic encephalomyelitis / chronic fatigue syndrome (ME/CFS)(17,18), and long COVID (19). These studies have demonstrated that, by combining additive and non-additive signals from multiple genetic variants, combinatorial analysis can identify significant genetic associations in smaller

CONFIDENTIAL © PrecisionLife Ltd 2025 All rights reserved

datasets than required for GWAS. Validation of the approach has been achieved via studies of reproducibility and replication in independent cohorts, both by the authors (18,20) and independent researchers (21).

Aims of Study

We hypothesized that combinatorial analysis would identify reproducible genetic signal for endometriosis that complements and extends previous findings from GWAS and other studies. We first ran combinatorial analysis on a cohort of endometriosis patients and controls from a UK Biobank (UKB) dataset. This analysis identified 1,709 disease signatures, each comprising between 2 to 5 SNP-genotypes, that were significantly associated with increased risk of endometriosis.

We then sought to demonstrate significant enrichment of reproducing disease associations for these signatures in an independent, diverse ancestry endometriosis cohort from All of Us (AoU) for which we had overlapping genotype data. We wanted to evaluate the relative reproducibility of the disease signatures across all participants and in sub-cohorts of self-reported Black/African American and Hispanic/Latino patients to ensure that the findings of this analysis are unlikely to exacerbate inequalities in medical treatment for historically underserved populations.

To demonstrate the utility of the combinatorial analytics approach, we sought to establish reproducibility of disease signatures that were not linked to the 42 significant meta-GWAS loci and identify unambiguously novel genetic disease associations with endometriosis. We annotated these novel candidate genes from a biological perspective to inform the broader understanding of the biological drivers of endometriosis and used our knowledge of the modulators of these novel candidate genes (with appropriate directionality to identify potential drug repurposing candidates for the associated mechanisms)(22).

Materials and Methods

Generation of UK Biobank Endometriosis Cohort

We used a cohort of endometriosis patients and healthy controls from UKB to identify multi-SNP 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 cohort to include only UKB participants identified as having self-reported ‘White British’ genetic ancestry. We were unable to perform combinatorial analyses on UKB patients with alternative ancestries due to their low representation and small sample sizes in that dataset.

196,188 potential controls were identified that matched the following criteria:

- Females with self-reported White British genetic ancestry

CONFIDENTIAL © PrecisionLife Ltd 2025 All rights reserved

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

We then 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 combinatorial analysis study was hypothesis-free. However, alongside the signatures identified by combinatorial analysis, we also wished to test the relative reproducibility of the lead SNPs identified by a large scale meta-GWAS study of endometriosis (15) (hereafter referred to as 'meta-GWAS SNPs'). Of the 42 meta-GWAS SNPs, 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 35 imputed meta-GWAS SNPs was unlikely to result in compounded error in this study as they comprise a very small fraction (0.02%) of the total dataset. Indeed, we identified only 9 disease signatures (out of 1,709) that contained more than one meta-GWAS SNPs, and no disease signature contained more than two meta-GWAS SNPs, suggesting minimal effects of compounding imputation error.

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)(23). We therefore filtered the UKB dataset to only include SNPs that were also present in CHB, which removed 7 of the 42 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

CONFIDENTIAL © PrecisionLife Ltd 2025 All rights reserved

may represent missed diagnoses due to historic clinical practices. If that is the case, it would result in decreased associations between genetic variants and a diminished ability to identify biologically relevant disease signatures, as well as reducing the reproducibility of the results between datasets (see Limitations of the Analysis).

Statistical / 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 (17,19).

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). 1,000 cycles of fully random permutations are typically 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 AoU Curated Data Repository (CDR) version 8 (accessed on the 24th of July 2025) (24). AoU CDR v8 contains data for 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

CONFIDENTIAL © PrecisionLife Ltd 2025 All rights reserved

SNP genotypes included in the filtered UKB genotype dataset. This resulted in 200,578 SNPs available for signature lookups. This lower SNP genotype coverage further restricts the number of matching reproducible signatures that could be found, and this led to 50 signatures being lost from the reproducibility test.

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 (21) 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 5%. 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).

Statistical Reproducibility of Disease Associations for Combinatorial Signatures in All of Us

We evaluated the degree to which the endometriosis disease signatures identified in the UKB cohort reproduce in AoU, using an approach similar to the one previously used to validate combinatorial disease signatures for long COVID (20). 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.

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 genetic PCs 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 (25) and Benjamini-Hochberg (26) 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 (10) 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. Due to their nature, multi-SNP disease signatures, especially with 4 or 5 SNPs, have much lower frequency than their individual component SNPs, requiring larger sized cohorts to achieve equivalent statistical power. This difference in statistical power is

CONFIDENTIAL © PrecisionLife Ltd 2025 All rights reserved

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. The greater number of features tested 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 enrichment 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. Observing an enrichment of reproducible signatures would suggest 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 the observed reproducibility rate (i.e., fraction of signatures with odds ratio > 1) in the randomized cohort. Repeating this process 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 fraction of total permutations in which 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.

Our previous analysis of long COVID (20) 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 filters for signature frequency in the AoU cohort.

Reproducibility of Disease Signatures in Traditionally Underserved Populations

We also tested whether the disease signatures identified in the White British UKB cohort reproduced among AoU participants who have self-reported 'Black' or 'African American' race as well as patients with self-reported 'Hispanic' or 'Latino/a' ethnicity. These cohorts are not fully mutually exclusive, as they represent responses to different questions within AoU. The aim of this analysis was to test whether use of combinatorial signatures in a clinical setting can avoid exacerbating historical health inequalities driven by biases in representation, not to identify biological differences between human populations. Therefore, we relied on self-reported race/ethnicity which reflects social constructs, rather than genetic ancestry.

CONFIDENTIAL © PrecisionLife Ltd 2025 All rights reserved

We again included genetic PCs as covariates in the reproducibility analyses for each sub-cohort to control for indirect relationships between signature frequency and disease prevalence resulting from population substructure that could potentially manifest as false positives. We calculated 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 and 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 (27). Additional gene assignments for identified SNPs were made using publicly available cis- or trans-eQTL data for ovary and uterine tissues (28,29) with false discovery rate (FDR) < 0.05. Reported genes were limited to protein-coding genes, pseudogenes, and antisense transcripts.

SNPs located outside the upstream and downstream distance thresholds (2 kb and 0.5 kb respectively) of any gene that did not have any data in relevant tissues were not associated with any gene for biological interpretation. This conservative SNP mapping approach is designed to identify the most likely biologically relevant gene(s) for novel target discovery and repurposing analyses. Meta-GWAS SNPs that were assigned to a gene in the meta-GWAS manuscript (10) but could not be mapped to any genes using this approach are listed in Supplementary Table 6.

The mapped genes were annotated using data from over 50 public data sources (see Supplementary Table 7), and their biological relevance analyzed to develop a deeper understanding of the gene 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 (2024). 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.

Pathway Enrichment

Pathway enrichment analysis was performed using gene set enrichment analysis (GSEA)(31) implemented in the gseapy package (32). Enrichment was assessed across three biological process and pathway databases: Gene Ontology (GO) (33), Reactome (34), and WikiPathways (35). Pathways with an adjusted *p*-value threshold of <0.05 were reported as significant.

Reproducibility of Meta-GWAS Disease Associations in All of Us

For comparison, we also assessed the replicability and reproducibility of the 35 meta-GWAS SNPs included in our dataset. 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, which mapped to 1,309 genes. Pathways enriched in the disease signatures included cell adhesion, proliferation and migration (36), epithelial-mesenchymal transition (36,37), cytoskeleton remodeling (38), angiogenesis (39,40), as well as signaling pathways such as TGF-β, PI3K/AKT and MAPK (41), many of which have been extensively linked to endometriosis (Supplementary Figure 4).

We then evaluated the frequency and disease association of these UKB disease signatures in AoU. The most common UKB disease signature occurs in 20% 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.

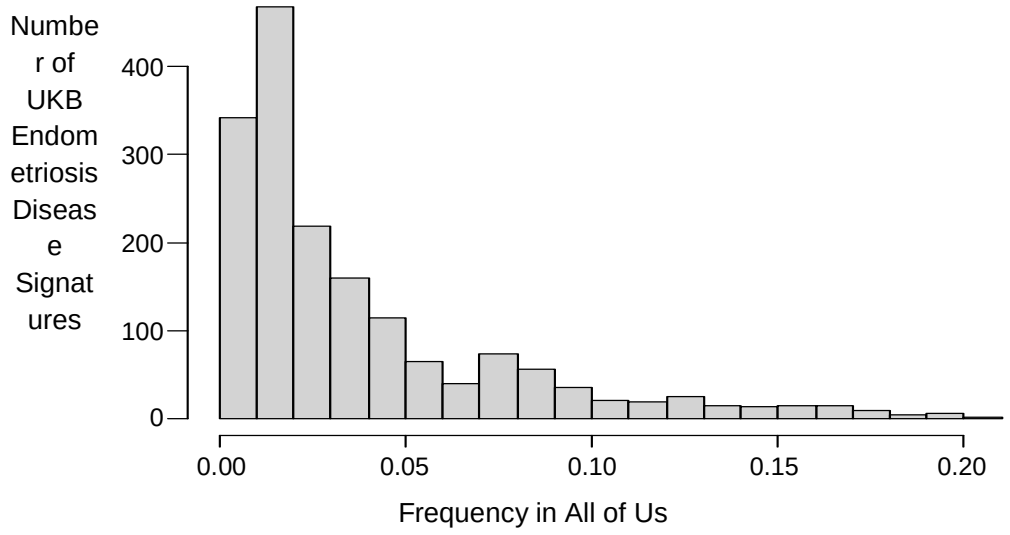


Figure 1. Distribution of UKB endometriosis disease signatures by frequency in AoU

CONFIDENTIAL © PrecisionLife Ltd 2025 All rights reserved

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 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 assessed 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).

CONFIDENTIAL © PrecisionLife Ltd 2025 All rights reserved

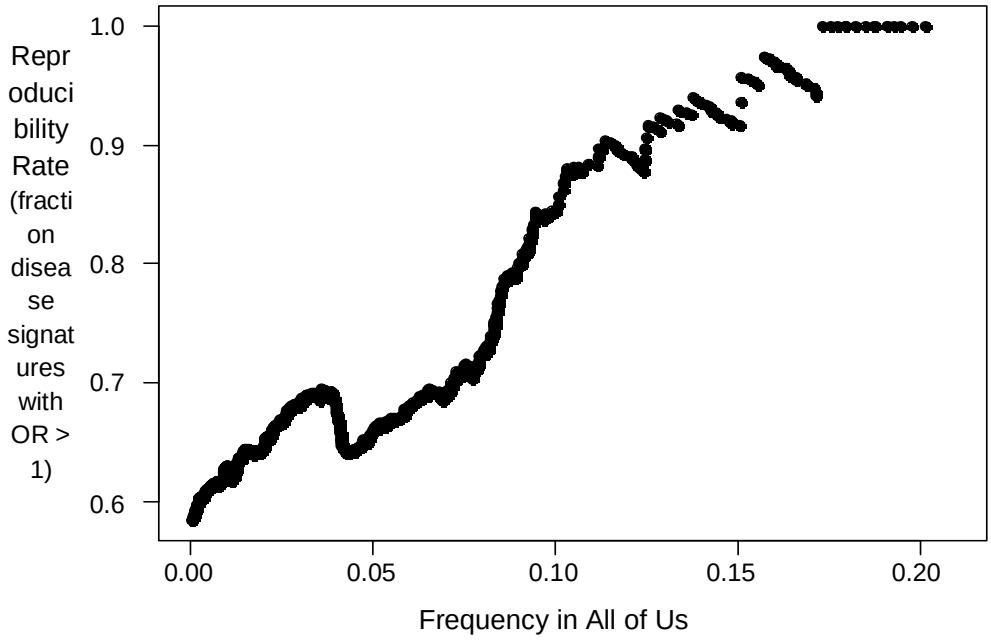


Figure 2. Reproducibility rates for UKB endometriosis disease signatures in AoU under different minimum frequency filters.

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. Including all 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 inherent to 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

CONFIDENTIAL © PrecisionLife Ltd 2025 All rights reserved

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

Reproducibility Across Self-Identified Race/Ethnicity Cohorts

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 / Latino’ AoU cohorts compared to all participants

Minimum signature frequency	# signatures tested	Reproducibility Rate % (#) All Participants	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)

CONFIDENTIAL © PrecisionLife Ltd 2025 All rights reserved

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 6 (17%) of these SNPs in our dataset, 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 8). An additional 4 SNPs, annotated to the genes *CD109*, *ACTL9*, *GDAP1*, and *VEZT* significantly replicate in AoU if we apply a Benjamini-Hochberg FDR correction, corresponding to a cumulative 29% replication.

The GWAS associations for 21 (60%) of the 35 meta-GWAS SNPs are not even nominally significant in AoU ($p > 0.05$), including 9 SNPs (26%) with p-values > 0.5. 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% of the UKB disease signatures (336 / 1709) contain at least one of the lead meta-GWAS SNPs. 17 of the 35 meta-GWAS SNPs are present in at least one disease signature (Table 4). Disease signatures assigned to 11 of these 17 meta-GWAS 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 associated meta-GWAS genes increases by 4 to a total of 21, 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 when considering alternative SNPs 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.

CONFIDENTIAL © PrecisionLife Ltd 2025 All rights reserved

Table 4. Reproducibility of UKB disease signatures linked to meta-GWAS SNPs and their associated genes in the meta-GWAS study.

Meta-GWAS gene(s)	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%

Reproducible 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 exhibited the strongest evidence for a consistent link to endometriosis in the AoU replication analysis. A total of 195 unique SNPs mapping to 98 genes (see Extended Table 1) were identified among the 141 disease signatures that have frequency >9% and that also replicate in AoU (odds ratio > 1). Of these, 7 genes were previously identified in the endometriosis meta-GWAS study. 16 additional genes have been previously associated with endometriosis according to Open Targets (42) or Pharos (43), providing independent confirmation of a potential role in disease, but were not identified by the meta-GWAS (see Supplementary Table 9). The remaining 75 candidate genes represent novel disease-associated genes. 25 of these are tractable drug targets (see Extended Table 2).

CONFIDENTIAL © PrecisionLife Ltd 2025 All rights reserved

Of the 98 candidate genes mapping to high frequency disease signatures, more than 50% are expressed in endometrium and ovary (nTPM > 1; see Extended Table 3). Pathways enriched in these candidate genes included cell-adhesion and proliferation mechanisms, cytoskeletal organization along with neurological and PI3K/ERK1/2 signaling processes (Supplementary Figure 5, Extended Table 4). Many of these novel genes occur in meta-GWAS signatures which exhibit strong reproducibility rates overall, but it is challenging to quantify the contribution of meta-GWAS genes vs. novel genes to this reproducibility.

To validate the ability of combinatorial analytics to identify novel reproducible candidate genes associated with endometriosis, we focused on the set of 25 SNPs that were found in 25 or more disease signatures (termed 'top SNPs'). This criterion reflects the network geometry-based approach used 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 top 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 candidate genes are included in Table 4 and Supplementary Table 8).

Of the remaining 14 annotated top 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 'top novel SNPs'. Importantly, all 9 top novel SNPs exhibit a strong enrichment of disease signatures that reproduce in AoU (Table 5). Reproducibility rates for the top novel 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.

The other 5 top 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, 3 show high rates of reproducibility, while 2 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 ('top SNPs').

Gene	# signatures	Reproducibility rate of all signatures in AoU % (#)	# non-meta-GWAS signatures mapped to gene	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. These UKB endometriosis disease signatures exhibited a significant enrichment of reproducible disease signal in AoU. That is, we observed significantly more UKB disease signatures that are also positively correlated with increased prevalence of endometriosis in AoU than expected by chance. These results include many new genes and mechanisms, and the broader biological context captured by these combinatorial disease signatures provides potential to improve our understanding of disease biology and build clinical development tools to support novel therapeutics programs.

Reproducibility rates are especially high for higher-frequency signatures, which is encouraging for precision medicine applications as signatures with higher frequency offer potential clinical utility across more patients. The increased reproducibility of higher-frequency signatures is consistent with observations from a reproducibility study of long COVID disease signatures (20) and likely reflects the lower sampling variance associated with increased sample sizes when estimating the signatures' odds ratios. That is, the observed odds ratio of biologically relevant low-frequency signatures is more likely to be less than 1 due to random sampling effects.

Despite relying on a much smaller dataset, combinatorial analysis identified many of the same gene-disease associations previously identified by a large scale meta-GWAS (10) as well as several novel genetic associations with endometriosis. The latter include many genes that have previously been linked to endometriosis in non-GWAS studies, providing the first genetic support for their role in disease.

We identified 75 novel gene associations that have high frequency and reproduce in the independent, ancestrally diverse AoU dataset, including 9 candidate genes associated with the highly reproducible top novel SNPs. Although none of the 9 top candidate genes have been previously linked to endometriosis, several play critical roles in pathways and biological processes that have been hypothesized to be important to disease biology.

First Genetic Evidence in Endometriosis for Many Genes

Identification of genes previously linked to endometriosis via non-genetic studies also provides support for the biological validity of the combinatorial analysis results. For example, we identified disease signatures linked to serotransferrin (TF), an iron binding protein which transports Fe^{3+} ions from sites of absorption and heme degradation to where it is utilized or stored.

CONFIDENTIAL © PrecisionLife Ltd 2025 All rights reserved

Dysregulation of iron homeostasis is believed to be involved in endometriosis (44). Elevated levels of iron occur in endometriotic tissues (44–46) and these have been associated with inflammation, oxidative stress, cellular damage, and ferroptosis (47). TF levels are elevated in peritoneal fluid from women with endometriosis (48,49) and TF saturation is correlated with disease (50). Populations of stromal cells and macrophages from endometrial tissues express elevated levels of the transferrin receptor 1, which is responsible for cellular uptake of TF and iron, resulting in increased iron loading (45,51).

Serotransferrin and antioxidants provide protection for mouse oocytes against the oxidative damage and dysmaturity induced by ovarian follicular fluid from endometriosis patients (52,53). Our observations lend support to the hypothesized importance of iron management in endometriosis and its potential utility for therapeutic manipulation.

Additional examples of genes with previous non-genetic associations are listed in Extended Table 2.

Novel Insights into the Genetics of Endometriosis

We identified 9 novel candidate endometriosis genes associated with the set of top SNPs 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 candidate 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 (54), has previously been implicated in the pathophysiology of endometriosis. Autophagic activity is suppressed in ectopic endometrial tissue, facilitating the survival, immune evasion, and proliferation of ectopic cells (55–57). Among the top novel candidate genes identified in our study, *ATG16L1* and *NEDD4L* both play 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 (58). While *ATG16L1* has not previously been associated with endometriosis, knockdown of interacting autophagy genes *ATG5* and *ATG7* has been shown to impair decidualization in human endometrial stromal cells (59), suggesting a potential 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 (60). Moreover, NEDD4L has been reported to inhibit the activity of mTOR, a key negative regulator of autophagy (61). 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 (62).

Notably, both *ATG16L1* and *NEDD4L* have been implicated in inflammatory bowel disease (IBD) (63,64). The *ATG16L1* SNP identified in this study was also strongly associated with Crohn's disease in a published GWAS (63). This finding is consistent with the reported symptomatic overlap and high rates of co-prevalence between Crohn's disease and endometriosis (65). Collectively, this evidence suggests that dysregulation of *ATG16L1* and

CONFIDENTIAL © PrecisionLife Ltd 2025 All rights reserved

NEDD4L may contribute to impaired autophagic homeostasis in endometriosis, potentially supporting lesion development, inflammation, and disease progression.

Macrophage Biology

Macrophage dysfunction has been hypothesized to contribute to the development and progression of endometriosis (66). Peritoneal macrophages produce more *VEGF* in endometriosis patients relative to healthy individuals, and transcriptomic analyses highlight macrophage-derived cytokines as key drivers of inflammation in endometriosis (67,68). In mouse models, macrophage depletion reduces lesion size, vascularization, and inflammatory pain, underscoring their central role in disease development (69,70). During menstruation, macrophages clear endometrial debris via scavenger receptor–mediated phagocytosis (71). Intriguingly, one of the top novel candidate genes identified in our analysis, *COLEC12*, encodes one of these scavenger receptors involved in host defense and phagocytosis.

The top novel candidate gene *VSTM1* encodes an immunoglobulin superfamily protein that enhances IL17A secretion by CD4+ T cells (72). Although *VSTM1* itself has not been studied in endometriosis, *IL17A* is known to drive pathogenic M2 macrophage polarization in lesions (73). Both genes show macrophage-enriched expression (74), 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 less well-studied but potentially druggable novel candidate genes, the reproducing disease signatures for endometriosis include several genes whose protein products are targets for well-established therapeutics that have not previously been explored in the treatment of endometriosis, i.e., drug repurposing opportunities. We ranked these drug repurposing candidates based on several criteria, including a mechanistic link to the disease, target selectivity, appropriate directionality of modulation, and a modality and safety profile suitable for endometriosis patients (22).

One example of a high-ranked candidate drug repurposing target included in the endometriosis disease signatures is the peptidase angiotensin-converting enzyme (ACE). ACE plays a major role in the renin-angiotensin-aldosterone system (75), which regulates sodium retention by the kidney and blood pressure by converting the peptide hormone angiotensin I (AT1) to angiotensin II (AT2), resulting in an increase of the vasoconstrictor activity (76). Although *ACE* was not identified in the endometriosis meta-GWAS, 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 (77). ACE inhibitors (captopril, ramipril) reduce the growth of endometrial implants in rat models (78,79) and the expression of the receptors for AT2 (AGTR1, AGTR2) is altered in endometrial lesions (80,81).

ACE inhibitors have been in widespread clinical use since the 1980s for a variety of indications including coronary artery disease, congestive heart failure, chronic kidney disease and myocardial infarction (82). A link between endometriosis and hypertension has

CONFIDENTIAL © PrecisionLife Ltd 2025 All rights reserved

previously been identified (83–85) but the opportunity to exploit existing therapeutic modulators of ACE or the angiotensin II receptor inhibitors has not been investigated.

Many other shortlisted repurposing candidates identified offer clear mechanistic hypotheses for their roles in lesion development, inflammation, fibrosis and – in a few cases – endometriosis-associated infertility. Future work based on these results will prioritize the drug repurposing candidates (see Extended Table 2) mapped to high frequency disease signatures as described above. Candidates will only be selected if they correspond to a clear hypothesis for how they might benefit endometriosis patients safely. Additional research will investigate the potential for using the disease signatures identified in this study as biomarkers to stratify patients by mechanistic subtype and match patients with drugs that target their unique disease mechanisms.

Limitations of the Analysis

We observed a significant enrichment of reproducible disease signatures in the output of the combinatorial analysis, illustrating the potential suitability for applying these disease signatures in aggregate to provide insights into the biology of endometriosis. However, it was only possible to statistically validate replication of a single individual disease signature, preventing us from confidently asserting which of the reproducible disease signatures are most important for disease biology.

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 the meta-GWAS SNPs replicate significantly in AoU, despite higher frequencies, much less biological complexity, and orders of magnitude fewer features and correspondingly reduced need for FDR adjustment to account for multiple tests relative to the replication analysis required 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 for both sets of results is likely due to lack of statistical power provided by the relatively small cohort.

Reliance on patient data provided by UKB and AoU also introduces several potential limitations for this study. Both have relatively small populations of diagnosed endometriosis cases, they have low genotype coverage and overlap, and both datasets have material levels of phenotypic misclassification in both cases and controls as is common in complex diseases. Each of these factors are expected to reduce the number of UKB disease signatures found during the initial combinatorial analysis and the level of their reproducibility reported in AoU.

A key limitation is the necessary reliance on ICD-10 coding, which is often inconsistently and inaccurately applied, to identify cases and controls. Surgical confirmation is needed to reliably identify patients with endometriosis, and such data is not consistently available in UKB or AoU. As a result, some participants with ICD-10 codes for endometriosis may not actually have the disease. The observed negative correlation between age and endometriosis in both datasets is likely to be an artifact of high prevalence of mis-/undiagnosed cases amongst older women due to lack of recognition of the disease and use of older, less reliable diagnostic criteria when they were younger. This pattern suggests that

CONFIDENTIAL © PrecisionLife Ltd 2025 All rights reserved

a subset of the ‘controls’ in the UKB and AoU study cohorts likely represent misphenotyped patients.

Phenotypic misclassifications typically reduce observed effect sizes by artificially increasing the resemblance between cases and controls and this is equally problematic for GWAS (86). Such misclassification also poses a challenge for reproducibility analyses, as signal dilution lowers the statistical power of results (87). 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 (21). Accordingly, we anticipate that phenotypic misclassification in UKB will have suppressed the overall number of statistically significant signatures found during the combinatorial analysis. Likewise, the reproducibility measures reported here likely represent a lower-bound estimate of the reliability of the combinatorial analytics approach due to phenotypic misclassification within AoU.

Although more robustly phenotyped genomic datasets of endometriosis patients are available, they do not include sufficient controls. Many genetic association studies merge case data with data for controls that were genotyped as part of different studies. However, differences in sample collection, genotyping, and SNP calling between studies can introduce technological ‘batch effects’, i.e., significant differences in allele frequencies between cases and controls related to dataset preparation rather than the disease status of the individuals. When this occurs, signatures that are indirectly associated with disease due to batch effects in the data can be very difficult to separate from true biological associations (18,88). Any benefits to this study that could be gained from using better phenotyped case data would therefore likely be outweighed by the reduced confidence in the observed disease associations.

Wider Implications of the Findings

The study’s findings include several candidate genes and mechanisms novel to endometriosis. This is important to provide new avenues of research to find new clinical options for a disease that is currently poorly diagnosed and whose main therapies are surgery and hormone treatment, neither of which is particularly effective or appropriate for many women (89).

Many of these genes are associated with targets or mechanisms that are plausible for the disease given evidence from other studies, and which can be modulated using known safe and well-tolerated active compounds. These can therefore be validated relatively readily using existing drug development candidates (repositioning) or on-market/generic compounds (repurposing) in small proof-of-concept human studies, ideally 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 complementary diagnostic tools to select patients most likely to respond to these drugs in a clinical setting.

Given the high degree of reproducibility, these same genetic/mechanistic biomarkers can also potentially be used during novel drug development to build a precision regulatory strategy, and to recruit likely responders into clinical trials, accelerating and derisking their clinical development.

CONFIDENTIAL © PrecisionLife Ltd 2025 All rights reserved

When fully developed and validated, such detailed genetic insights could also potentially be used in a clinical setting for rapid and non-invasive differential triage of patients presenting with deep pelvic pain to facilitate quicker referral of women to achieve a definitive diagnosis, and to evaluate patient prognosis for severity and progression to infertility, endometriomas, and other complications.

Supporting Health Equity

It is important that the results of genomic analyses support health equity across as many patients as possible, including those with different ancestries. GWAS results and related risk scores are often derived predominantly from European cohorts and may fail to transfer well across populations (90–93). Better population coverage in the datasets used for genetic analyses is widely recognized as important for ensuring that study findings and derived tools do not disadvantage populations of patients, but the choice of analytical methods is often overlooked as a critical factor in finding transferable results that reproduce well across different patient subgroups (94).

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 race/ethnicity. This observation is crucial as it implies that these disease signatures can be used to inform precision medicine healthcare without adversely affecting historically underserved populations and further increasing health disparities.

Conclusions

The high level of reproducibility of combinatorial disease signatures across different ancestries, based on combinations of relatively common variants, and the finding of a large number of both novel and known genes is encouraging for furthering the study of endometriosis and its clinical care.

Further combinatorial analysis approaches in endometriosis would benefit from access to larger datasets with even wider population diversity, more secure diagnosis, more harmonized 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 in the future.

Much more 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, but it also requires there to be a range of good precision therapeutic options once those women are diagnosed. Progress on both fronts will help strategically prioritize much needed women's health research initiatives to make more rapid progress in addressing this massive global challenge and improving patients' lives.

Abbreviations

***AoU*:** All of Us

***FDR*:** false discovery rate

***GWAS*:** genome-wide association study

***IBD*:** inflammatory bowel disease

CONFIDENTIAL © PrecisionLife Ltd 2025 All rights reserved

***PCA*:** principal component analysis

***SNP*:** single nucleotide polymorphism

***UKB*:** UK Biobank

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.

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.

Availability of Data and Materials

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

Declaration of Competing Interest

AR is an employee of the Complex Disorders Alliance, SG is an SAB member of Our Future Health. JS, GM, SD, KT, KC, MS, CS, AM and SG are employees of PrecisionLife Ltd. SG and GM are shareholders of and have assigned patent rights to PrecisionLife Ltd.

CONFIDENTIAL © PrecisionLife Ltd 2025 All rights reserved

Funding

The project was funded by PrecisionLife Ltd with support in part from the EU EIC accelerator project Transformative Non-invasive Causal Mechanostics Platform to Effectively Triage and Treat Endometriosis (TRANSCEND/101217760) and from the EU Horizon 2020 funded project Finding Endometriosis using Machine Learning (FEMaLe/101017562).

Author Contributions

SG, JS, GM, SD, CS, and KT contributed to the design of the study. JS designed the reproducibility analyses. GM, SD, KC, MS, and JS performed the analyses described in this manuscript. KC, SD and MS conducted the analyses on the Researcher Workbench. CS, AM, KT and SD analyzed the genes identified and SG, KT, SD and AR helped source the datasets used. All authors contributed to writing the manuscript and consent to publication.

Acknowledgements

We gratefully acknowledge All of Us and UK Biobank participants for their contributions, without whom this research would not have been possible. Special thanks to Gert Møller and Claus Erik Jensen, who initially developed the combinatorial analytics methodology, Matthew Pearson and James Kozubek for prior analyses related to endometriosis, and the rest of the PrecisionLife and Complex Disorders Alliance teams for their contributions to supporting the platform and study. 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 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.

CONFIDENTIAL © PrecisionLife Ltd 2025 All rights reserved

References

1. World Health Organization. https://www.who.int/news-room/fact-sheets/detail/endometriosis. 2023. Endometriosis Fact Sheet.

2. Rogers PAW, D'Hooghe TM, Fazleabas A, Gargett CE, Giudice LC, Montgomery GW, et al. Priorities for endometriosis research: Recommendations from an international consensus workshop. Reprod Sci. 2009 Apr 30;16(4):335–46.

3. 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 Sep 1;18(9):1922–7.

4. Brown J, Crawford TJ, Datta S, Prentice A. Oral contraceptives for pain associated with endometriosis. Cochrane Database of Systematic Reviews. 2018 May 22;2018(5).

5. Vannuccini S, Clemenza S, Rossi M, Petraglia F. Hormonal treatments for endometriosis: The endocrine background. Rev Endocr Metab Disord. 2022 Jun 17;23(3):333–55.

6. 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 Jul 1;40(7):1249–56.

7. Smolarz B, Szyłło K, Romanowicz H. Endometriosis: Epidemiology, classification, pathogenesis, treatment and genetics (review of literature). Int J Mol Sci. 2021 Sep 29;22(19):10554.

8. Burney RO, Giudice LC. Pathogenesis and pathophysiology of endometriosis. Fertil Steril. 2012 Sep;98(3):511–9.

9. Malvezzi H, Marengo EB, Podgaec S, Piccinato C de A. Endometriosis: current challenges in modeling a multifactorial disease of unknown etiology. J Transl Med. 2020 Dec 12;18(1):311.

10. Rahmioglu N, Mortlock S, Ghiasi M, Møller PL, Stefansdottir L, Galarneau G, et al. The genetic basis of endometriosis and comorbidity with other pain and inflammatory conditions. Nat Genet. 2023 Mar 13;55(3):423–36.

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

12. NICE. https://www.nice.org.uk/guidance/ng73/chapter/recommendations. 2017. Endometriosis: diagnosis and management.

CONFIDENTIAL © PrecisionLife Ltd 2025 All rights reserved

13. Byrne D, Curnow T, Smith P, Cutner A, Saridogan E, Clark TJ. Laparoscopic excision of deep rectovaginal endometriosis in BSGE endometriosis centres: a multicentre prospective cohort study. BMJ Open. 2018 Apr;8(4):e018924.

14. Quesada J, Härmä K, Reid S, Rao T, Lo G, Yang N, et al. Endometriosis: A multimodal imaging review. Eur J Radiol. 2023 Jan;158:110610.

15. Gardner S. Combinatorial analytics: An essential tool for the delivery of precision medicine and precision agriculture. Artificial Intelligence in the Life Sciences. 2021 Dec;1:100003.

16. Sardell J, Das S, Taylor K, Stubberfield C, Malinowski A, Strivens M, et al. Actively protective combinatorial analysis: A scalable novel method for detecting variants that contribute to reduced disease prevalence in high-risk individuals. Artificial Intelligence in the Life Sciences. 2025 Jun;7:100125.

17. Das S, Taylor K, Kozubek J, Sardell J, Gardner S. Genetic risk factors for ME/CFS identified using combinatorial analysis. J Transl Med. 2022 Dec 14;20(1):598.

18. Sardell J, Das S, Pearson M, Kolobkov D, Malinowski A, Fullwood L, et al. Identification of novel reproducible combinatorial genetic risk factors for myalgic encephalomyelitis in the DecodeME patient cohort and commonalities with long COVID. medRxiv. 2025 Dec 2;

19. 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 Nov 1;21(1):775.

20. Sardell J, Pearson M, Chocian K, Das S, Taylor K, Strivens M, et al. 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 May 8;23(1):516.

21. Cheng Z. Integrative genome-wide association studies of COVID-19 susceptibility and hospitalization reveal risk loci for long COVID. medRxiv. 2025 Feb 14;

22. Das S, Taylor K, Beaulah S, Gardner S. Systematic indication extension for drugs using patient stratification insights generated by combinatorial analytics. Patterns. 2022 Jun;3(6):100496.

23. Sørensen E, Christiansen L, Wilkowski B, Larsen MH, Burgdorf KS, Thørner LW, et al. Data Resource Profile: The Copenhagen Hospital Biobank (CHB). Int J Epidemiol. 2021 Jul 9;50(3):719–720e.

24. Bick AG, Metcalf GA, Mayo KR, Lichtenstein L, Rura S, Carroll RJ, et al. Genomic data in the All of Us research program. Nature. 2024 Mar 14;627(8003):340–6.

CONFIDENTIAL © PrecisionLife Ltd 2025 All rights reserved

25. Neyman J, Pearson ES. On the use and interpretation of certain test criteria for purposes of statistical inference: Part I. Biometrika. 1928 Jul;20A(1/2):175.

26. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B Stat Methodol. 1995 Jan 1;57(1):289–300.

27. Brodie A, Azaria JR, Ofran Y. How far from the SNP may the causative genes be? Nucleic Acids Res. 2016 Jul 27;44(13):6046–54.

28. Fung JN, Girling JE, Lukowski SW, Sapkota Y, Wallace L, Holdsworth-Carson SJ, et al. The genetic regulation of transcription in human endometrial tissue. Hum Reprod. 2017 Apr 1;32(4):893–904.

29. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, et al. The Genotype-Tissue Expression (GTEx) project. Nat Genet. 2013 Jun 29;45(6):580–5.

30. GlobalData. https://www.globaldata.com/industries-we-cover/pharmaceutical/. 2024. GlobalData Pharma Market Data and Insights .

31. Reimand J, Isserlin R, Voisin V, Kucera M, Tannus-Lopes C, Rostamianfar A, et al. Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap. Nat Protoc. 2019 Feb 21;14(2):482–517.

32. Fang Z, Liu X, Peltz G. GSEApy: a comprehensive package for performing gene set enrichment analysis in Python. Bioinformatics. 2023 Jan 1;39(1).

33. Huntley RP, Sawford T, Mutowo-Meullenet P, Shypitsyna A, Bonilla C, Martin MJ, et al. The GOA database: Gene Ontology annotation updates for 2015. Nucleic Acids Res. 2015 Jan 28;43(D1):D1057–63.

34. Gillespie M, Jassal B, Stephan R, Milacic M, Rothfels K, Senff-Ribeiro A, et al. The reactome pathway knowledgebase 2022. Nucleic Acids Res. 2022 Jan 7;50(D1):D687–92.

35. Agrawal A, Balcı H, Hanspers K, Coort SL, Martens M, Slenter DN, et al. WikiPathways 2024: next generation pathway database. Nucleic Acids Res. 2024 Jan 5;52(D1):D679–89.

36. Zondervan KT, Becker CM, Missmer SA. Endometriosis. New England Journal of Medicine. 2020 Mar 26;382(13):1244–56.

37. Vincent-Mistiaen ZI. Epithelial-mesenchymal transition links inflammation and fibrosis in the pathogenesis of endometriosis: a narrative review. F S Rev. 2025 Jun;6(1):100089.

CONFIDENTIAL © PrecisionLife Ltd 2025 All rights reserved

38. Arendt W, Kleszczyński K, Gagat M, Izdebska M. Endometriosis and cytoskeletal remodeling: The functional role of actin-binding proteins. Cells. 2025 Feb 28;14(5):360.

39. Samimi M, Pourhanifeh MH, Mehdizadehkashi A, Eftekhar T, Asemi Z. The role of inflammation, oxidative stress, angiogenesis, and apoptosis in the pathophysiology of endometriosis: Basic science and new insights based on gene expression. J Cell Physiol. 2019 Nov 19;234(11):19384–92.

40. Tan Y, Flynn WF, Sivajothi S, Luo D, Bozal SB, Davé M, et al. Single-cell analysis of endometriosis reveals a coordinated transcriptional programme driving immunotolerance and angiogenesis across eutopic and ectopic tissues. Nat Cell Biol. 2022 Aug 21;24(8):1306–18.

41. Zhang M, Xu T, Tong D, Li S, Yu X, Liu B, et al. Research advances in endometriosis-related signaling pathways: A review. Biomedicine & Pharmacotherapy. 2023 Aug;164:114909.

42. Ochoa D, Hercules A, Carmona M, Suveges D, Baker J, Malangone C, et al. The next-generation Open Targets Platform: reimagined, redesigned, rebuilt. Nucleic Acids Res. 2023 Jan 6;51(D1):D1353–9.

43. Kelleher KJ, Sheils TK, Mathias SL, Yang JJ, Metzger VT, Siramshetty VB, et al. Pharos 2023: an integrated resource for the understudied human proteome. Nucleic Acids Res. 2023 Jan 6;51(D1):D1405–16.

44. Wyatt J, Fernando SM, Powell SG, Hill CJ, Arshad I, Probert C, et al. The role of iron in the pathogenesis of endometriosis: a systematic review. Hum Reprod Open. 2023 Jun 1;2023(3).

45. Lousse JC, Defrère S, Van Langendonckt A, Gras J, González-Ramos R, Colette S, et al. Iron storage is significantly increased in peritoneal macrophages of endometriosis patients and correlates with iron overload in peritoneal fluid. Fertil Steril. 2009 May;91(5):1668–75.

46. Ng SW, Norwitz SG, Taylor HS, Norwitz ER. Endometriosis: the role of iron overload and ferroptosis. Reprod Sci. 2020 Jul 19;27(7):1383–90.

47. Mielke Cabello LA, Meresman G, Darici D, Carnovale N, Heitkötter B, Schulte M, et al. Assessment of the ferroptosis regulators: glutathione peroxidase 4, acyl-coenzyme A synthetase long-chain family member 4, and transferrin receptor 1 in patient-derived endometriosis tissue. Biomolecules. 2024 Jul 21;14(7):876.

48. Mathur SP. Autoimmunity in endometriosis: Relevance to infertility. American Journal of Reproductive Immunology. 2000 Aug 28;44(2):89–95.

CONFIDENTIAL © PrecisionLife Ltd 2025 All rights reserved

49. Ferrero S, Gillott DJ, Remorgida V, Anserini P, Ragni N, Grudzinskas JG. Proteomic analysis of peritoneal fluid in fertile and infertile women with endometriosis. J Reprod Med. 2009 Jan;54(1):32–40.

50. Xu G, Chen L, Li Q. Association of iron metabolism markers, socioeconomic and lifestyle factors with endometriosis: A cross-sectional study. Journal of Trace Elements in Medicine and Biology. 2023 Jul;78:127175.

51. Mori M, Ito F, Shi L, Wang Y, Ishida C, Hattori Y, et al. Ovarian endometriosis-associated stromal cells reveal persistently high affinity for iron. Redox Biol. 2015 Dec;6:578–86.

52. Ni Z, Li Y, Song D, Ding J, Mei S, Sun S, et al. Iron-overloaded follicular fluid increases the risk of endometriosis-related infertility by triggering granulosa cell ferroptosis and oocyte dysmaturity. Cell Death Dis. 2022 Jul 4;13(7):579.

53. Ren Z, Huang J, Zhou C, Jia L, Li M, Liang X, et al. Transferrin and antioxidants partly prevented mouse oocyte oxidative damage induced by exposure of cumulus-oocyte complexes to endometrioma fluid. J Ovarian Res. 2020 Dec 26;13(1):139.

54. Klionsky DJ, Petroni G, Amaravadi RK, Baehrecke EH, Ballabio A, Boya P, et al. Autophagy in major human diseases. EMBO J. 2021 Oct 30;40(19).

55. 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 Jan 20;23(1).

56. Yang HL, Mei J, Chang KK, Zhou WJ, Huang LQ, Li MQ. Autophagy in endometriosis. Am J Transl Res. 2017;9(11):4707–25.

57. Zhan L, Li J, Wei B. Autophagy in endometriosis: Friend or foe? Biochem Biophys Res Commun. 2018 Jan;495(1):60–3.

58. 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 Jun 23;20(12):3066.

59. Gammoh N. The multifaceted functions of ATG16L1 in autophagy and related processes. J Cell Sci. 2020 Oct 15;133(20).

60. Lee DE, Yoo JE, Kim J, Kim S, Kim S, Lee H, et al. NEDD4L downregulates autophagy and cell growth by modulating ULK1 and a glutamine transporter. Cell Death Dis. 2020 Jan 20;11(1):38.

61. Li G, Song Z, Wu C, Li X, Zhao L, Tong B, et al. Downregulation of NEDD4L by EGFR signaling promotes the development of lung adenocarcinoma. J Transl Med. 2022 Dec 28;20(1):47.

CONFIDENTIAL © PrecisionLife Ltd 2025 All rights reserved

62. 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. 2025 Sep;25(9):1154–66.

63. de Lange KM, Moutsianas L, Lee JC, Lamb CA, Luo Y, Kennedy NA, et al. Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. Nat Genet. 2017 Feb 9;49(2):256–61.

64. 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. Journal of Clinical Investigation. 2025 Feb 3;135(3).

65. Chiaffarino F, Cipriani S, Ricci E, Roncella E, Mauri PA, Parazzini F, et al. Endometriosis and inflammatory bowel disease: A systematic review of the literature. European Journal of Obstetrics & Gynecology and Reproductive Biology. 2020 Sep;252:246–51.

66. Hogg C, Panir K, Dhami P, Rosser M, Mack M, Soong D, et al. Macrophages inhibit and enhance endometriosis depending on their origin. Proceedings of the National Academy of Sciences. 2021 Feb 9;118(6).

67. McLaren J, Prentice A, Charnock-Jones DS, Millican SA, Müller KH, Sharkey AM, et al. Vascular endothelial growth factor is produced by peritoneal fluid macrophages in endometriosis and is regulated by ovarian steroids. Journal of Clinical Investigation. 1996 Jul 15;98(2):482–9.

68. Beste MT, Pfäffle-Doyle N, Prentice EA, Morris SN, Lauffenburger DA, Isaacson KB, et al. Molecular network analysis of endometriosis reveals a role for c-Jun–regulated macrophage activation. Sci Transl Med. 2014 Feb 5;6(222).

69. 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 Aug;175(2):547–56.

70. Forster R, Sarginson A, Velichkova A, Hogg C, Dorning A, Horne AW, et al. Macrophage-derived insulin-like growth factor-1 is a key neurotrophic and nerve-sensitizing factor in pain associated with endometriosis. The FASEB Journal. 2019 Oct 10;33(10):11210–22.

71. Park M, Kim YS, Song H. Macrophages: a double-edged sword in female reproduction and disorders. Exp Mol Med. 2025 Feb 3;57(2):285–97.

72. Guo X, Zhang Y, Wang P, Li T, Fu W, Mo X, et al. VSTM1-v2, a novel soluble glycoprotein, promotes the differentiation and activation of Th17 cells. Cell Immunol. 2012 Jul;278(1–2):136–42.

CONFIDENTIAL © PrecisionLife Ltd 2025 All rights reserved

73. Miller JE, Ahn SH, Marks RM, Monsanto SP, Fazleabas AT, Koti M, et al. IL-17A modulates peritoneal macrophage recruitment and M2 polarization in endometriosis. Front Immunol. 2020 Feb 14;11.

74. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Tissue-based map of the human proteome. Science (1979). 2015 Jan 23;347(6220).

75. Wu CH, Mohammadmoradi S, Chen JZ, Sawada H, Daugherty A, Lu HS. Renin-angiotensin system and cardiovascular functions. Arterioscler Thromb Vasc Biol. 2018 Jul;38(7).

76. Wong MKS. Angiotensin converting enzymes. In: Handbook of Hormones. Elsevier; 2016. p. 263-e29D-4.

77. 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 Jul;225:116280.

78. Oktem M, Ozcan P, Erdem O, Karakaya C, Cenksoy C, Guner H, et al. The effect of captopril on endometriotic implants in a rat model. European Journal of Obstetrics & Gynecology and Reproductive Biology. 2014 Sep;180:120–5.

79. Mazumdar P, Biswas SS. Ramipril ameliorates endometriosis by inducing oxidative stress-mediated apoptosis in the wistar rat. J Mol Histol. 2025 Apr 24;56(2):117.

80. 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.

81. 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 Jul 29;

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

83. Mu F, Rich-Edwards J, Rimm EB, Spiegelman D, Forman JP, Missmer SA. Association between endometriosis and hypercholesterolemia or hypertension. Hypertension. 2017 Jul;70(1):59–65.

84. 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 Jan 21;25(1):245.

CONFIDENTIAL © PrecisionLife Ltd 2025 All rights reserved

85. 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 Jun 26;21(1):91.

86. Beesley LJ, Mukherjee B. Statistical inference for association studies using electronic health records: handling both selection bias and outcome misclassification. Biometrics. 2022 Mar 3;78(1):214–26.

87. Burstein D, Hoffman G, Mathur D, Venkatesh S, Therrien K, Fanous AH, et al. Detecting and adjusting for hidden biases due to phenotype misclassification in genome-wide association studies. medRxiv. 2023.

88. Lee SH, Nyholt DR, Macgregor S, Henders AK, Zondervan KT, Montgomery GW, et al. A simple and fast two-locus quality control test to detect false positives due to batch effects in genome-wide association studies. Genet Epidemiol. 2010 Dec 22;34(8):854–62.

89. Becker CM, Gattrell WT, Gude K, Singh SS. Reevaluating response and failure of medical treatment of endometriosis: a systematic review. Fertil Steril. 2017 Jul;108(1):125–36.

90. Tam V, Patel N, Turcotte M, Bossé Y, Paré G, Meyre D. Benefits and limitations of genome-wide association studies. Nat Rev Genet. 2019 Aug 8;20(8):467–84.

91. Moreno-Grau S, Vernekar M, Lopez-Pineda A, Mas-Montserrat D, Barrabés M, Quinto-Cortés CD, et al. Polygenic risk score portability for common diseases across genetically diverse populations. Hum Genomics. 2024 Sep 2;18(1):93.

92. Peterson RE, Kuchenbaecker K, Walters RK, Chen CY, Popejoy AB, Periyasamy S, et al. Genome-wide association studies in ancestrally diverse populations: opportunities, methods, pitfalls, and recommendations. Cell. 2019 Oct;179(3):589–603.

93. Kachuri L, Chatterjee N, Hirbo J, Schaid DJ, Martin I, Kullo IJ, et al. Principles and methods for transferring polygenic risk scores across global populations. Nat Rev Genet. 2024 Jan 24;25(1):8–25.

94. Lehmann B, Bräuninger L, Cho Y, Falck F, Jayadeva S, Katell M, et al. Methodological opportunities in genomic data analysis to advance health equity. Nat Rev Genet. 2025 Sep 15;26(9):635–49.

CONFIDENTIAL © PrecisionLife Ltd 2025 All rights reserved