Expanding the genetic landscape of endometriosis: Integrative -omics analyses implicate key genes and pathways in a multi-ancestry study of over one million women

Authors and Affiliations

Corresponding Author: Shefali Setia-Verma

Lindsay A Guare[1], Jagyashila Das[1], Lannawill Caruth[1], Ananya Rajagopalan[2], Alexis T. Akerele[3], Ben M Brumpton[4], Tzu-Ting Chen[5], Leah Kottyan[6], Yen-Feng Lin[5], Elisa Moreno[4], Ashley J Mulford[7], Marija Simona Dombrovska[8], Yuan Luo[9], Vita Rovite[8], Alan R Sanders[7], Craig Teerlink[10], Danielle Candelieri[10], Noemie Elhadad[11], Andrew Hill[12], Gail P. Jarvik[13], James Jaworski[14], Julie Lynch[10], Shinichi Namba[15], Yukinori Okada[15], Yue Shi[16], Yuya Shirai[17], Jonathan Shortt[12], Wei-Qi Wei[18], Chunhua Weng[11], Yuji Yamamoto[17], Penn Medicine Biobank, Regeneron Genetics Center, Global Biobank Meta-analysis Initiative, Sinead Chapman[16], Wei Zhou[16], Todd Edwards[19], Suneeta Senapati[20], Digna R. Velez Edwards[21], Shefali Setia-Verma[1]

[1]Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, PA, United States, [2]Genomics and Computational Biology, University of Pennsylvania, Philadelphia, PA, United States, [3]Division of Quantitative Science, Department of Obstetrics and Gynecology, Department of Microbiology, Immunology, and Physiology, Vanderbilt University Medical Center, and Meharry Medical College, Nashville, TN, United States, [4]HUNT Center for Molecular and Clinical Epidemiology, Norwegian University of Science and Technology, Trondheim,

21 Norway, [5]Center for Neuropsychiatric Research, National Health Research Institutes, Maoli
22 County, Taiwan, [6]Cincinnati Children's Hospital Medical Center, Cincinatti, OH, United
23 States, [7]Endeavor Health, Evanston, IL, United States, [8]Latvian Biomedical Research and Study
24 Centre, Riga, Latvia, [9]Feinberg School of Medicine, Northwestern University, Chicago, IL, United
25 States, [10]VA Informatics and Computing Infrastructure, VA Salt Lake City Healthcare System,
26 Salte Lake City, UT, United States, [11]Biomedical Informatics, Columbia University, New York City,
27 NY, United States, [12]Colorado for Precision Medicine, Aurora, CO, United States, [13]Division of
28 Medical Genetics, University of Washington, Seattle, WA, United States, [14]Division of
29 Epidemiology, Department of Medicine, Vanderbilt University Medical Center, Nashville,
30 Tennessee, United States, [15]Genome Informatics, The University of Tokyo, Tokyo, Japan, [16]Broad
31 Institute, Cambridge, MA, United States, [17]Statistical Genetics, Osaka University, Suita, Osaka,
32 Japan, [18]Biomedical Informatics, Vanderbilt University, Nashville, TN, United States, [19]Division of
33 Epidemiology, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN,
34 United States, [20]Division of Reproductive Endocrinology and Infertility, Department of Obstetrics
35 and Gynecology, Hospital of the University of Pennsylvania, Philadelphia, PA, United
36 States, [21]Division of Quantitative Science, Department of Obstetrics and Gynecology,
37 Department of Microbiology, Immunology, and Physiology, Vanderbilt University Medical
38 Center, Nashville, Tennessee, United States

39

Abstract

We report the findings of a genome-wide association study (GWAS) meta-analysis of endometriosis across 14 biobanks worldwide, including 32% non-European patient participants, as part of the Global Biobank Meta-Analysis Initiative (GBMI). Out of 58 total loci (29 previously unreported), the largest meta-analysis accounted for 46 (20 previously unreported). We detected the first genome-wide significant loci (*2p13.3* and *20q13.2*) uniquely driven by the African-ancestry meta-analysis. Our imaging- and surgery-confirmed phenotypes yielded six additional previously unreported loci. Leveraging our large and diverse study population, we observed SNP heritability estimates of 9-13% for all ancestry groups, and 13 loci had at least one variant in the credible set after fine-mapping. Investigating the complex array of endometriosis comorbidities and risk factors revealed 135 genetically correlated phenotypes and 95 with evidence of vertical pleiotropy, including triglycerides and anxiety disorders. We prioritized 35 disease-relevant cellular contexts from the endometrial cell atlas and found 322 examples of differentially expressed genes in cells from donors with endometriosis. Further high-throughput multi-omic analyses implicated a total of 282 genes in endometriosis pathogenesis. Our diverse, comprehensive GWASs, with downstream analyses spanning molecular to phenotypic scales, provide detailed evidence for aspects of endometriosis including the role of immune cell types, Wnt signaling, and cellular proliferation. These interconnected pathways and risk factors underscore the complex, multi-faceted etiology of endometriosis, suggesting multiple targets for precise and effective therapeutic interventions.

60 # Introduction

61 Endometriosis, a debilitating condition characterized by the growth of endometrial-like tissues
62 outside of the uterus affects approximately 10% of women of reproductive age worldwide[1,2]. The
63 precise mechanisms of these lesions are presently unknown. Endometriosis is incurable, leaving
64 patients to manage severe pain[3] and cope with fertility challenges[4]. Despite its prevalence and
65 impact, the etiology of endometriosis remains poorly understood, hampering efforts to develop
66 effective diagnostic tools and targeted treatments.

67 Genome-wide association studies (GWASs) have been an important tool for suggesting
68 candidate genes and pathways linked with endometriosis[5–7]. GWASs for endometriosis have
69 uncovered over 40 loci associated with endometriosis risk. However, like most GWASs
70 historically, the previous endometriosis studies have primarily included European ancestry
71 populations and at a smaller scale, East-Asian[8]. While the broad sense heritability of
72 endometriosis has been estimated as 47% via a twin study[9], previous endometriosis GWASs have
73 only explained about 7% of phenotypic variance with common variants[5,6] and have fallen short
74 in elucidating the full spectrum of its pathophysiology.

75 The advent of large-scale biobanks has revolutionized genetic research, offering unique
76 opportunities to conduct well-powered studies across diverse populations[10–13]. As more diverse
77 genomic datasets become available like the Penn Medicine Biobank (PMBB)[14], the All of Us
78 Research Program (AOU)[15], and the Million Veterans Program (MVP)[16], they enhance the
79 discovery of trait loci through increased statistical power and variation enriched in non-European
80 populations[10]. A worldwide consortium of genomic researchers, the Global Biobank Meta-
81 Analysis Initiative (GBMI), was established to facilitate collaboration in GWAS studies on

82 unprecedented scales. In this study, we at the PMBB have collaborated with 12 other biobanks
83 across several countries[15,17–27] and incorporated publicly-available summary statistics from
84 FinnGen, totaling 14 global datasets. These resources not only enhance statistical power to
85 detect novel associations but also provide rich phenotypic data, enabling more refined analyses.
86 Moreover, the inclusion of diverse ancestries in genetic studies is crucial to aid in fine-mapping
87 and for improving the generalizability of findings and addressing health disparities. This is
88 particularly pertinent for endometriosis, where significant variations in prevalence and clinical
89 presentation have been observed across different ethnic groups[28].

90 Beyond GWAS, there are many methods which can be employed *in silico* to investigate
91 other elements of the central dogma, providing further validation of GWAS loci and discovery of
92 additional disease-associated genes. Recent advances in multi-omics technologies have further
93 expanded our ability to translate genetic associations into biological insights. Integration of
94 GWAS results with multi-omic QTL and single-cell datasets using methods such as imputed
95 association studies and Mendelian randomization allows for a more comprehensive
96 understanding of disease mechanisms, potentially identifying novel therapeutic targets and
97 biomarkers. For endometriosis, where the interplay between genetic predisposition and complex
98 tissue-specific processes is central, integrative approaches are especially valuable. Given these
99 considerations, there is a critical need for a large-scale, ancestrally diverse GWAS of
100 endometriosis that incorporates rigorous phenotyping and leverages multi-omics data to provide
101 a more complete picture of the disease's genetic landscape and underlying biology.

102 Here, we present the results of a large-scale GWAS meta-analysis that significantly
103 advances our understanding of endometriosis pathophysiology. Our study encompasses a

104 diverse cohort from 14 biobanks worldwide, with over 30% non-European participants,
105 enhancing the generalizability of our findings. We employed a comprehensive phenotyping
106 approach, including not only a wide spectrum of endometriosis presentations but also surgically
107 confirmed and procedure-validated (surgery or imaging) phenotypes. This robust phenotyping
108 strategy enables greater resolution when examining the genetic architecture of endometriosis.
109 Through high-throughput exploration of pleiotropy and integrative multi-omics analyses, we
110 have explored numerous layers of endometriosis pathogenesis. These findings provide
111 unprecedented insights into the risk factors and molecular mechanisms driving this complex
112 disorder, paving the way for improved diagnostics and targeted therapeutic interventions.

113 Results

114 EHR-Based Phenotyping of Endometriosis in Genetically Diverse Datasets

115 Of the six case-control definitions (W - Wide, Wex - Wide excluding adenomyosis, PCN.v2 and
116 PCN.v2 - Procedure-Confirmed Narrow [vs (1) all and (2) confirmed controls], and SCN.v1 and
117 SCN.v2 - Surgically-Confirmed Narrow [vs (1) all and (2) confirmed controls]), the largest
118 phenotype, W, comprised 1,063,657 women (31.9% non-EUR), including 50,265 cases. The Wex
119 phenotype, for which several biobanks were omitted, had a total sample size of 615,050 (N_{cases} =
120 12,262). Sample sizes for the PCN and SCN phenotypes were smaller (Table 1), with summary
121 statistics included from only nine of the 14 biobanks (Supplementary Table 1).

Table 1: All GWAS meta-analysis sample sizes. All six phenotypes had a multi-ancestry meta-analysis as well as single-ancestry meta-analyses for AFR, AMR, CSA+SAS, EAS, and EUR. There were not enough CSA+SAS or EAS studies to conduct meta-analyses for all the narrow phenotypes. Phenotype abbreviations: W = wide, Wex = wide excluding adenomyosis, PCN = procedure-confirmed narrow, SCN = surgically-confirmed narrow, .v1 = versus all controls, .v2 = versus confirmed controls. Ancestry abbreviations: AFR = African, AMR = admixed American, CSA = central/south Asian, SAS = south Asian, EAS = east Asian, EUR = European.

Phenotype	Ancestry	N	Cases	Prevalence
W	AFR	96,012	4,289	4.47%
	AMR	67,335	2,136	3.17%
	CSA+SAS	8,164	259	3.17%
	EAS	168,209	7,833	4.66%
	EUR	723,937	35,748	4.94%
	META	1,063,657	50,265	4.73%
Wex	AFR	92,202	1,575	1.71%
	AMR	66,219	1,020	1.54%
	CSA+SAS	5,066	62	1.22%
	EAS	86,667	1,584	1.83%
	EUR	364,896	8,021	2.20%
	META	615,050	12,262	1.99%
PCN.v1	AFR	70,959	1,152	1.62%
	AMR	58,802	750	1.28%
	EAS	78,999	155	0.20%
	EUR	317,481	4,053	1.28%
	META	526,241	6,110	1.16%
PCN.v2	AFR	12,958	1,152	8.89%
	AMR	9,472	750	7.92%
	EAS	1,539	155	10.07%
	EUR	56,134	4,053	7.22%
	META	80,103	6,110	7.63%
SCN.v1	AFR	70,308	501	0.71%
	AMR	58,477	425	0.73%
	EUR	284,932	2,573	0.90%
	META	413,717	3,499	0.85%
SCN.v2	AFR	1,912	501	26.20%
	AMR	2,291	425	18.55%
	EUR	14,147	2,573	18.19%
	META	18,350	3,499	19.07%

Multi-Ancestry Wide Phenotype GWAS Meta-Analysis Results

Across the 11 meta-analyses (six multi-ancestry meta-analyses for the different phenotype definitions plus five single-ancestry W meta-analyses), we identified 3,165 unique genome-wide significant variants ($P < 5 \times 10^{-8}$), corresponding to 5,049 associations (Supplementary Table 2). These associations were aggregated into 249 linkage disequilibrium (LD) clumps (Supplementary Table 3), which further merged into 104 distinct loci across the 11 meta-analyses (Supplementary Table 4). When we consolidate loci shared between meta-analyses, we arrive at 58 total loci. Study-wide loci shown in Figure 1 are labeled with the most significant corresponding gene from the downstream multi-omic analyses (described below) or the cytoband (Supplementary Table 5). The largest analysis (multi-ancestry W phenotype) resulted in 46 loci associated with endometriosis. Besides those, there were three EUR-specific W loci (*GRB14, ZNF536,* and *CHD6*), totaling 49 W loci annotated in Figure 1a. The two AFR and two of the seven EAS W loci were unique among the single-ancestry GWASs but all were overlapping with signals from the multi-ancestry meta-analyses; meanwhile, the AMR and CSA+SAS W GWASs did not yield any genome-wide significant loci. Of the 49 unique W loci, 26 of the lead SNPs were in LD ($R^2 > 0.1$) with the lead SNPs reported in *Rahmioglu et al, 2023*, and an additional three were found within 1 Mb of the previous lead SNPs[6]. The associations between endometriosis and the 20 remaining W loci are reported here for the first time.

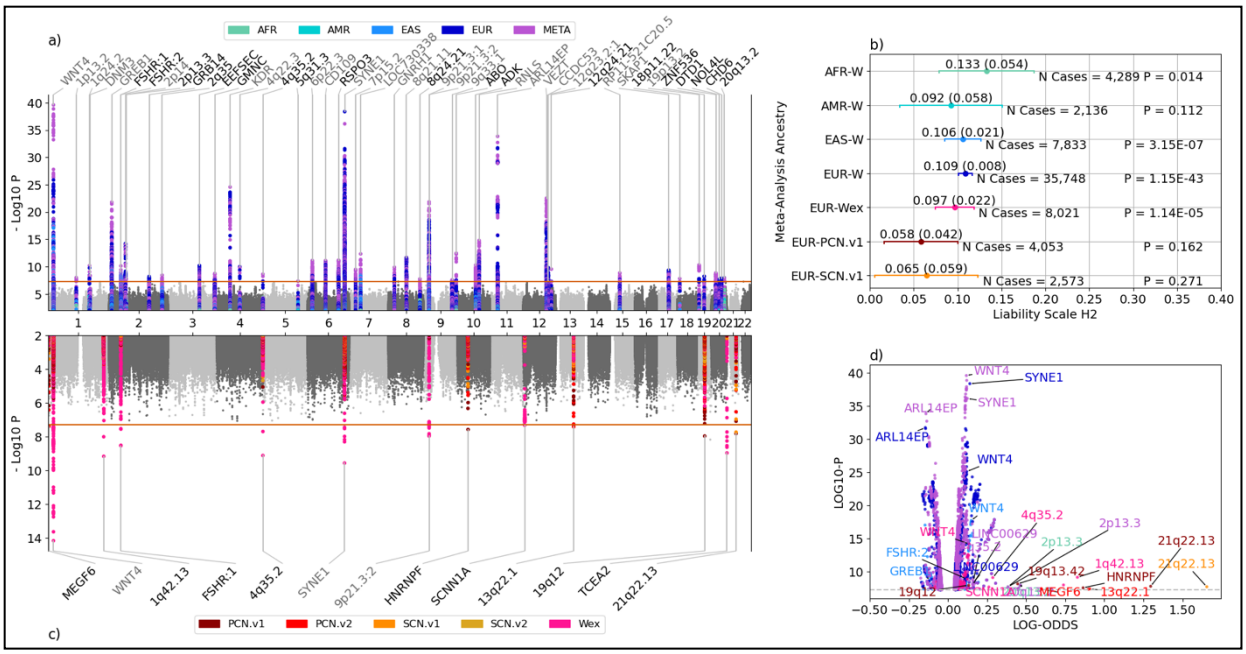

Figure 1: Overview of 11 GWAS meta-analysis results. a) Top left: overlayed Manhattan plots for the five Wide endometriosis phenotype GWASs (multi-ancestry + four single-ancestries). All significant loci are annotated with black text highlighting the previously unreported hits. b) Top right: LD-score regression (using LDSC) heritability estimates for the W GWASs; AMR is excluded because there was not sufficient sample size to use LDSC. c) Bottom left: Overlayed Manhattan plots for the five multi-ancestry GWASs for the other phenotypes. d) Bottom right: volcano plot highlighting variants with an absolute log-odds effect size > 0.1, colored by ancestry group. Phenotype abbreviations: W = wide, Wex = wide excluding adenomyosis, PCN = procedure-confirmed narrow, SCN = surgically-confirmed narrow, .v1 = versus all controls, .v2 = versus confirmed controls. Ancestry abbreviations: AFR = African, AMR = admixed American, EAS = east Asian, EUR = European.

Liability scale genome-wide SNP heritability estimates and standard errors (h^2_{lia}) from the W single-ancestry meta-analyses are shown in Figure 1b. The h^2 estimates and their standard errors were: 0.133 (0.054) for AFR, 0.106 (0.021) for EAS, and 0.108 (0.008) for EUR. Heritability estimates (via the LD-score regression tool, LDSC) for all analyses are provided in Supplementary Table 6. Partitioned LD-score regression (Supplementary Table 7) attributed significant proportions of the phenotypic variance to uterine-specific active enhancers (epigenome map state called "EnhA2") in EAS ($prop_{h2}$ = 33%, enrichment = 26-fold, $P_{prop\ h2} = 1.2 \times 10^{-3}$) and EUR ($prop_{h2}$ = 20%, enrichment = 15-fold, $P_{prop\ h2} = 6.6 \times 10^{-4}$). Two others explained significant portions of the heritability: ncRNA_gene (non-coding RNA genes) in EAS ($prop_{h2}$ = 21%,

167 enrichment = 1.4-fold, $P_{prop\ h2} = 2.6 \times 10^{-3}$) and TssFlnk (regions flanking transcription start sites)
168 in EUR ($prop_{h2}$ = 7%, enrichment = 1.3-fold, $P_{prop\ h2} = 0.031$). Cross-ancestry genetic correlation
169 estimates with Popcorn were under-powered, with none being significantly different from one.
170 The strongest estimate was between EAS and EUR: $\rho_{gi} = 0.76$, P = 0.063 (all genetic impact
171 correlation estimates are available in Supplementary Table 8).

172 Precise Phenotype Analyses Replicate Known Loci and Reveal Additional Loci

173 Across the remaining phenotype meta-analyses, we identified 14 significant loci: eight with Wex,
174 three with PCN.v1, one with both PCN.v1 and SCN.v1, and two with PCN.v2. Three of the eight
175 loci associated with Wex in the multi-ancestry analysis replicate previously reported signals,
176 while the other five are previously unreported: *1q12.13, FSHR:1, 4q35.2, SCNN1A,* and *TCEA2*. All
177 six loci associated with the narrow phenotypes are previously unreported (Figure 1c). The
178 surgically-confirmed phenotype associations displayed the most extreme effect sizes (Figure 1d),
179 including *21q22.13* (rs186666195:T $OR_{SCN.v1}$ = 5.23 and $OR_{PCN.v1}$ = 3.65) and *13q22.1*
180 (rs189908888:T $OR_{PCN.v2}$ = 2.48). The liability scale heritability captured by the EUR Wex GWAS
181 was 0.096 (0.0018). Heritability estimates for the remaining narrow phenotypes were non-
182 significant and are provided in Supplementary Table 6. None of the partitioned heritability
183 estimates for the Wex and narrow phenotypes (EUR-only) reached significance (Supplementary
184 Table 7).

185 Fine-Mapping to Detect Causal Variants at GWAS Loci

186 To characterize shared and population-specific putative causal variants, we conducted statistical
187 fine-mapping (See Methods) on all genome-wide significant regions from multi-population meta-
188 analyses, totaling 63 significant loci in five of the six studies (the SCN.v2 meta-analysis yielded no

189 genome-wide significant results). Among these, 13 loci contained at least one variant with
190 posterior inclusion probability (PIP) greater than 0.5, corresponding to 85 unique causal variants
191 with strong evidence for causality (Supplementary Table 9). The cross-ancestry fine-mapping
192 analyses computed PIPs for each variant across all combinations of ancestry groups included in
193 the overall study. We identify credible sets (CS) for the 13 distinct loci associated with
194 endometriosis as reported in Table 2.

195 Table 2: Credible sets of multi-ancestry GWAS loci. nine are for the W phenotype, two are for Wex, one is for PCN.v1,
196 and one for PCN.v2. An "X" in the ancestry population columns denote whether that ancestry group was used for
197 detecting any SNPs in the credible set (PIP > 0.5). Locus names match Figure 1. Phenotype abbreviations: W = wide,
198 Wex = wide excluding adenomyosis, PCN = procedure-confirmed narrow, .v1 = versus all controls, .v2 = versus
199 confirmed controls. Ancestry abbreviations: AFR = African, AMR = admixed American, EAS = east Asian, EUR =
200 European, SAS = south Asian.

Phenotype	Locus	# CS	AFR	AMR	EAS	EUR	SAS	Credible Set
W	*WNT4*	1			X	X		rs2807365
W	*1p13.2*	4			X	X	X	rs12030576, rs2982742, rs61810746, rs6328
W	*GREB1*	3		X	X	X		rs7578132, rs10929759, rs59129126
W	*FSHR:2*	2		X	X	X	X	rs1504175, rs1504188
W	*CD109*	1	X	X		X		rs77740740
W	*7p15.2*	1	X		X	X		rs1451385
W	*7p12.3*	2		X	X	X		rs10479868, rs11975261
W	*ADK*	2		X	X	X	X	rs4268450, rs11001059
W	*DTD1*	4			X	X	X	rs6081248, rs6081313, rs34478401, rs6045577
Wex	*FSHR:1*	1		X		X		rs2300440
Wex	*9p21.3:2*	3	X		X	X		rs1333054, rs10965274, rs10811669
PCN.v1	*19q12*	7	X			X		rs56124241, rs10411368, rs12981956, rs59841764, rs57255653, rs12973006, rs12972676
PCN.v2	*MEGF6*	1		X		X		rs76419782

201 Of the 13 loci with at least one variant in the credible set, nine were from the W multi-
202 population meta-analysis, two were from Wex, and there was one each from PCN.v1 and PCN.v2.

203 Twelve of the credible sets included variants detectable in part due to the incorporation of
204 additional genetic ancestry groups (AMR, AFR, and SAS) previously absent from analyses. In
205 particular, two loci (*FSHR:2* and *ADK.)* derived their credible sets from incorporating four of the
206 five ancestry groups. All PIPs from MESuSiE are available in Supplementary Table 9.

207 GWAS Variant-Set Gene Enrichment Analyses

208 From 11 meta-analyses, we identified 78 associations with 46 unique gene regions that were
209 enriched for significant GWAS variants. Eleven gene regions from MAGMA did not overlap with
210 significant GWAS loci, of which 9 were detected by W analyses. Two of the 9 were significant in
211 both the META and EUR wide phenotype GWASs: *CCHCR1* and *KCTD13*. The other two non-GWAS
212 significant MAGMA hits were for the Wex (*GLB1L3*, $P = 3.66 \times 10^{-7}$) and PCN.v1 (*LRRC2*, $P = 3.40$
213 $\times 10^{-7}$) phenotypes. All significant MAGMA results are reported in Supplementary Table 10.

214 Exploring Pleiotropy via Genetic Correlation and Mendelian Randomization

215 Out of 3,065 genetic correlation tests spanning complex trait GWASs from the UK Biobank
216 (UKBB) phenome, the All of Us (AOU) phenome, microbiome, and partial proteome datasets, 890
217 traits were nominated for bidirectional Mendelian randomization analysis. A total of 135
218 phenotypes showed significant genetic correlation with endometriosis at a Bonferroni-corrected
219 threshold of $P < 1.63 \times 10^{-5}$, including age at first live birth (rg = -0.24), sleeplessness/insomnia (rg
220 = 0.19), headache (rg = 0.30), and antidepressants prescription (rg = 0.39). Evidence for vertical
221 pleiotropy was observed in 95 exposure-outcome pairs, including associations suggestive of
222 endometriosis contributing to risk of hyperlipidemia, prescription of antacids, and reduced levels
223 of Apolipoprotein F and *Bifidobacteriaceae*. Nominated genetic correlation traits and their MR

results are shown in Figure 2. Detailed results of the analyses are provided in Supplementary Tables 11 and 12 for genetic correlation and MR, respectively.

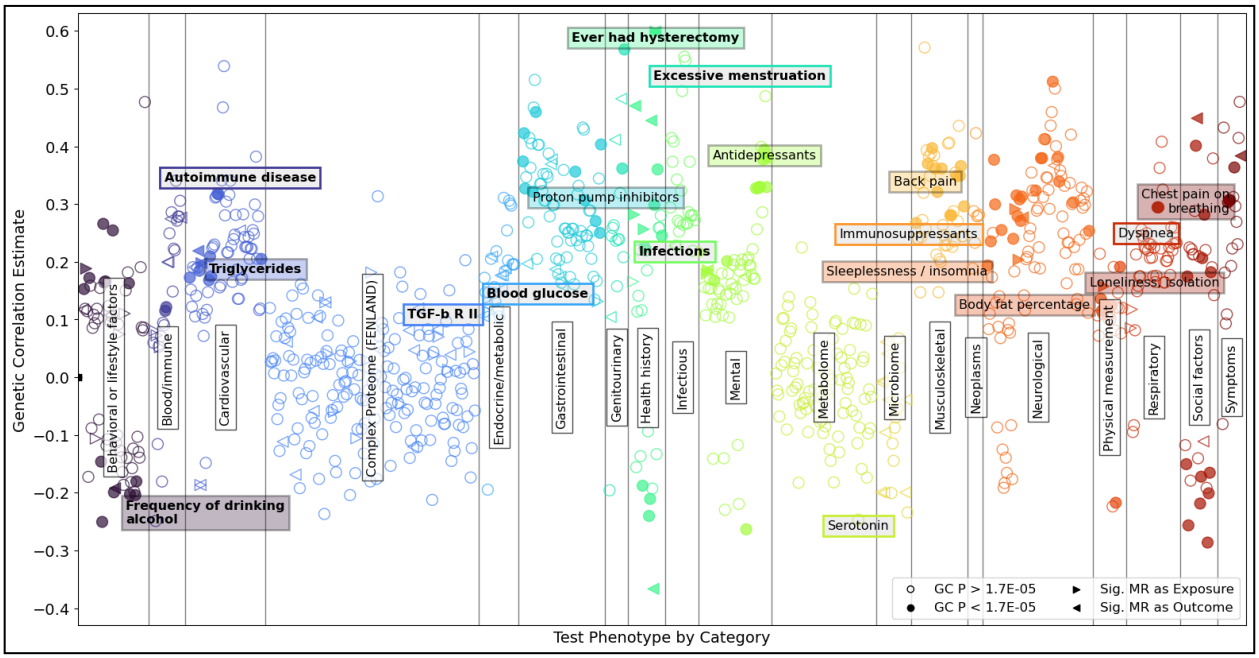

Figure 2: Genetic correlation (GC) and Mendelian randomization (MR) analyses. Each scatter point, colored by category, represents a GC estimate for which the p-value was less than 0.05. Filled shapes indicate a Bonferroni-significant GC value. Triangles indicate significant causal relationships detected by MR. Selected results are highlighted with boxes. A color fill corresponds to a significant GC estimate whereas a colored outline did not reach multiple testing GC significance. Bold text in the annotation boxes indicates a significant MR pair.

Identifying Disease-Relevant Cell Types and Crucial Genes

We prioritized 34 cellular contexts (cell type – menstrual stage pairs) with significant regression coefficients ($P < 2.08 \times 10^{-4}$) for comparison of gene expression among the top MAGMA-identified genes (Figure 3). Regression results of endometriosis status on single cell disease risk score (scDRS) value across all tested cell type-menstrual stage pairs are provided in Supplementary Table 13.

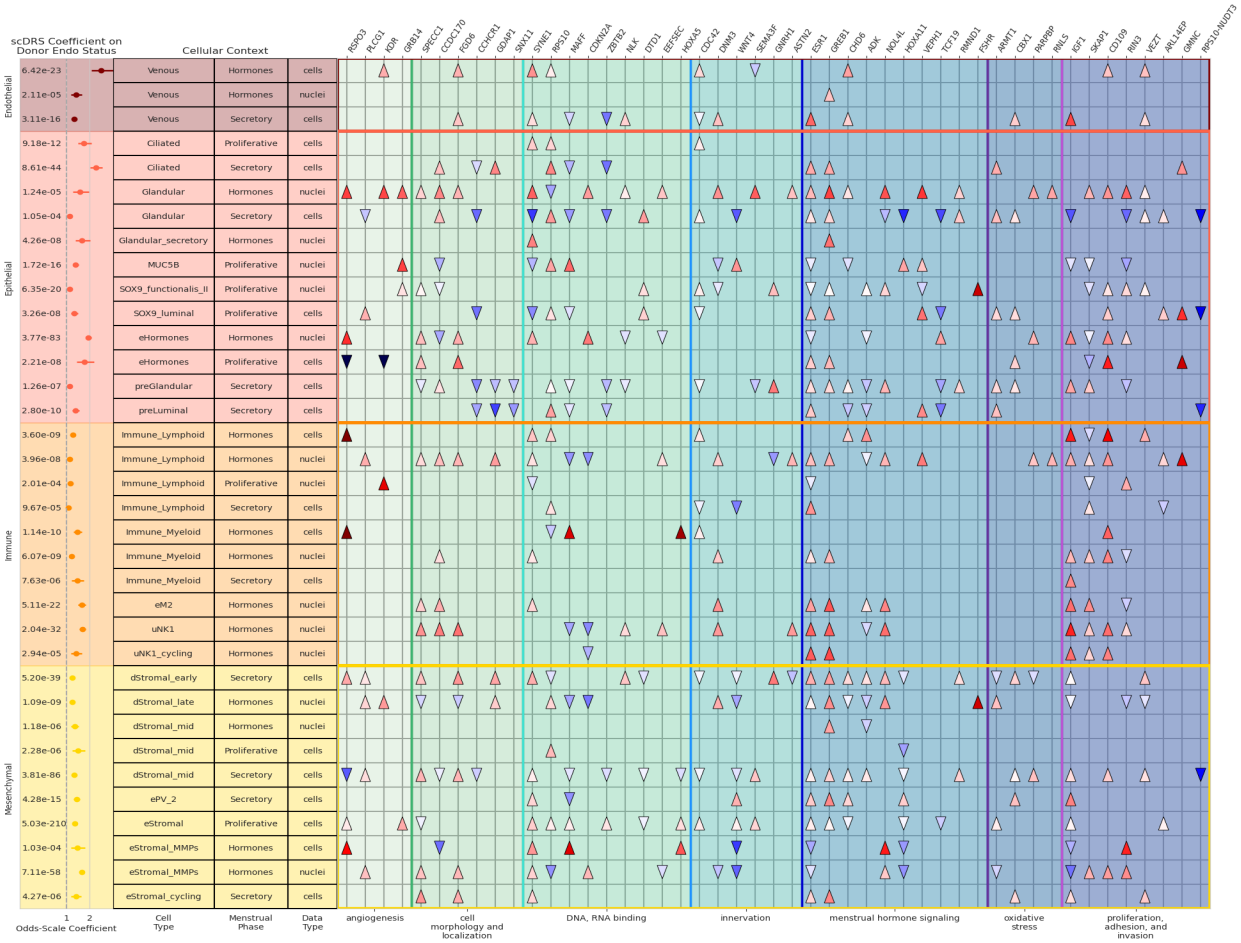

Figure 3: Single cell analyses – disease relevance scores (scDRSs) and differential gene expression. The left warm-colored panels show the odd-scale regression coefficients when comparing cells from donors with endometriosis to donors without endometriosis. Each row is annotated with the menstrual stage, cell type, and data type. The rows are ordered by cell type lineage. The right cool-colored panels show the results of two-sided t-tests comparing the gene expression of the top MAGMA genes, with significant tests represented by upward tringles for overexpression and downward triangles for underexpression.

Of the 34 cell contexts with significant regression effect estimates, 14 were derived from donors in the “hormones” menstrual phase, five from the proliferative phase, and 15 from the secretory phase. These contexts spanned all four cellular lineage groups: three endothelial, eight epithelial, ten immune, and 13 mesenchymal. The largest regression coefficient was observed in venous cells from donors on hormones (odds-scale coefficient = 2.34, P = 1.02 x 10^{-18}). The most statistically robust signal emerged in endometrial stromal cells sampled during the proliferative phase (odds-scale coefficient = 1.40, P = 1.58 x 10^{-231}). Among the genes driving the performance

252 of each regression model, *ESR1* was most frequently differentially expressed, reaching
253 significance in 24 of the cellular contexts, with 19 of those being overexpression and 5
254 demonstrating depletion. *GREB1* was the most consistently upregulated gene (23 contexts: 10
255 hormones, 10 secretory, and 3 proliferative), whereas *ADK* and *ZBTB2* were the most consistently
256 downregulated (7 contexts each). Decidualized stromal cells from the mid-secretory phase
257 exhibited the largest number of differentially expressed genes overall (21 genes: 13 upregulated
258 and 8 downregulated).

259 Eight differentially expressed genes overlap with seven of GWAS loci that had not been
260 previously associated with endometriosis. Gene expression results across all cell type–menstrual
261 stage contexts are provided in Supplementary Table 14. *FSHR* demonstrated strong
262 overexpression in two nucleus types: proliferative-phase SOX9 functionalis II (11.8-fold, P = 1.79
263 x 10^{-7}) and late decidualized stromal from donors on hormones (11.0-fold, P = 1.36 x 10^{-22}).
264 Menstrual-stage-dependent association directions were observed for *EEFSEC*; in particular,
265 uterine natural killer nuclei had a higher expression of *EEFSEC* gene in donors taking exogenous
266 hormones (1.52-fold, P = 3.5 x 10^{-12}), but in the same type of nuclei in the secretory phase, it was
267 reduced (0.702-fold, P = 1.50 x 10^{-9}). Two immune nucleus types from donors taking hormones
268 showed high enrichment of *GMNC*, lymphoid (7.99-fold, P = 4.1 x 10^{-8}) and myeloid (6.02-fold, P
269 = 2.45 x 10^{-5}). Relative expression of *RSPO3* within stromal cells peaked during the early secretory
270 phase (1.78-fold, P = 1.65 x 10^{-14}) relative to the proliferative stage (1.17-fold, P = 1.79 x 10^{-8}) and
271 the mid secretory phase (0.325-fold, P = 5.00 x 10^{-133}). Two genes from the *ADK* locus, *AP3M1*
272 and *ADK* shared associations with two cellular contexts, late-secretory decidualized stromal cells
273 and pre-luminal secretory cells. *DTD1* was differentially expressed in five stromal cell contexts,

274 four positively (early, mid, late decidualized and proliferative nondecidualized) and one
275 negatively (nondecidualized secretory-stage stromal cells: 0.91-fold, $P = 3.84 \times 10^{-13}$). Six of the
276 ten *NOL4L*-associated contexts involved nucleus types from donors exposed to exogenous
277 hormones, all demonstrating positive association (1.46- to 2.72-fold, $P < 1.36 \times 10^{-5}$).

278 ## Post-GWAS Multi-Omic Analyses Fuel Pathway Identification

279 To identify associations between endometriosis and additional layers of the central dogma and
280 beyond (p - proteome, s - splice-ome, e - transcriptome, m - methylome, and ed - RNA edit-ome),
281 we performed imputed association studies (xWAS) and QTL-wide summary-based Mendelian
282 randomization (SMR), inputting the eleven GWASs' summary statistics (See Methods). In total,
283 xWAS implicated 54 unique molecular phenotypes, including 2 proteins, 73 introns, and 35
284 transcripts (Supplementary Table 15). Causal relationships involving 235 unique genes were
285 identified by SMR, represented by 1 pQTL, 57 sQTLs, 34 eQTLs, 353 mQTLs, and 6 edQTLs (Figure
286 4; Supplementary Table 16).

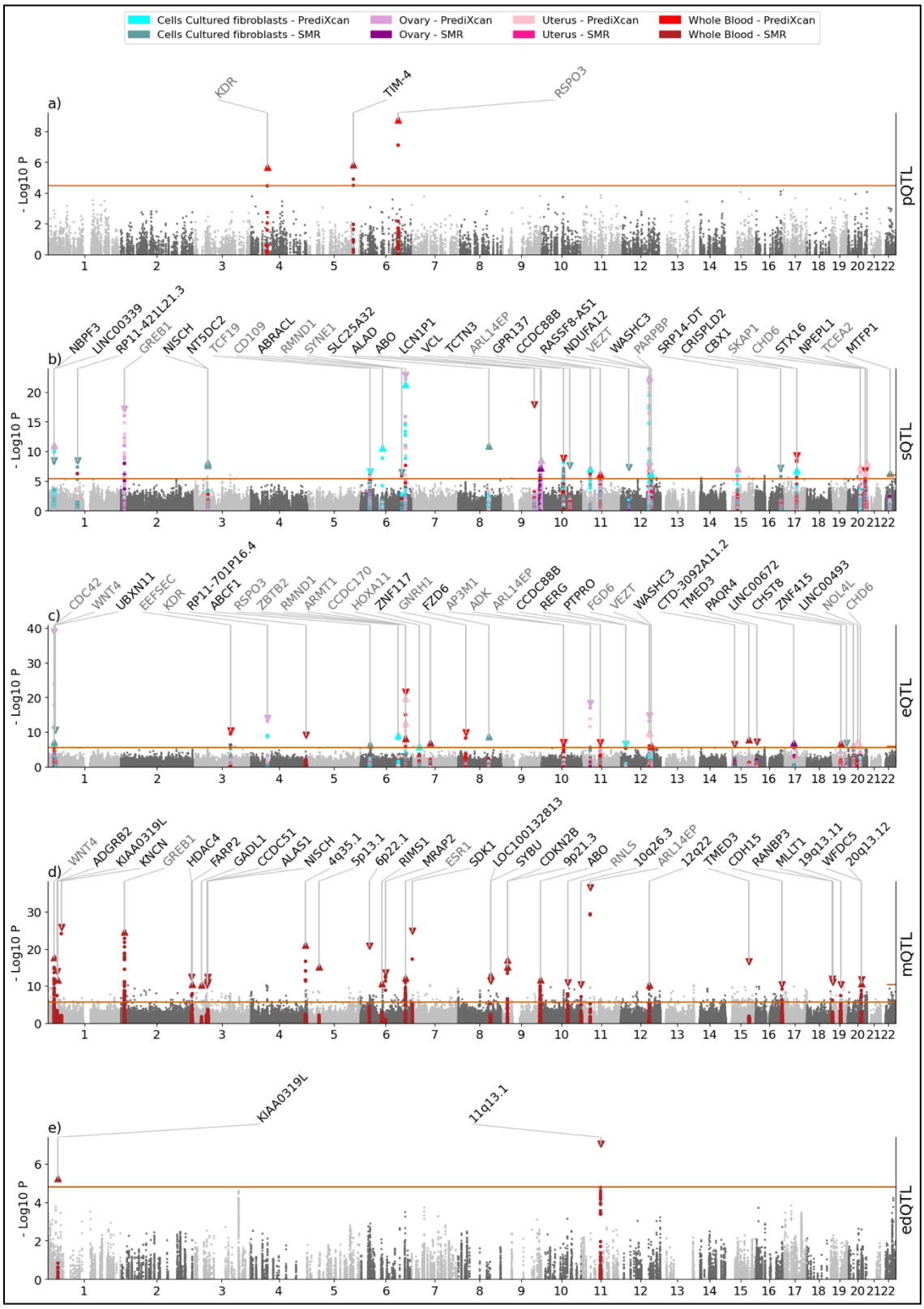

Figure 4: xQTL summary-based Mendelian randomization (xSMR) and PrediXcan imputed "x"ome-wide association study (xWAS) results across four tissue types and five -omic layers: a) proteome, b) splice-ome, c) transcriptome, d)

methylome, and e) RNA edit-ome. The p-value thresholds were adjusted for the numbers of tests per analysis (count of sites per combination of GWAS – QTL dataset). The number of genes displayed on each panel is limited to the 35 most significant. Genes annotated in gray overlap with MAGMA genes whereas genes annotated in black do not. Triangles point upward for a positive effect and down for a negative effect.

Of the 266 genes implicated by xQTL summary-based Mendelian randomization (xSMR) and PrediXcan imputed "x"ome-wide association studies (xWAS), 32 overlapped with genome-level associations identified by MAGMA. Above and beyond the GWAS-related genes, pSMR and PWAS detected one protein, sSMR and SWAS identified introns at 33 genes, eSMR and TWAS revealed 44 genes, mSMR identified methylation sites at 164 genes, and edSMR implicated RNA editing sites at three genes. Several genes were supported by multiple layers of evidence, notably *KDR* and *RSPO3* which exhibited associations across MAGMA, the proteome, and the transcriptome. In contrast, we detected associations with *CCDC88B* and *WASHC3 in* both the splice-ome and transcriptome, but no corresponding associations were detected at the genome-level by MAGMA.

The only pSMR signal without significant heterogeneity was the protein TIM-4, with a positive association across three GWASs: AMR W (B = 0.594, $P_{SMR} = 3.26 \times 10^{-5}$), META PCN.v1 (B = 0.402, $P_{SMR} = 1.48 \times 10^{-6}$), and META PCN.v2 (B = 0.384, $P_{SMR} = 1.27 \times 10^{-5}$). *VCL* was the only gene in the intersection of the sSMR with SWAS results, represented by three distinct splicing events with associations across all four tissues, chr10:74109156-74111909 (B = -0.0803, $P_{SMR} = 2.81 \times 10^{-7}$, $P_{SWAS} = 1.32 \times 10^{-9}$), chr10:74109156-74114184 (B = 0.0692, $P_{SMR} = 8.37 \times 10^{-8}$, $P_{SWAS} = 2.93 \times 10^{-9}$), and chr10:74112112-74114184 (B = -0.0735, $P_{SMR} = 1.51 \times 10^{-6}$, $P_{SWAS} = 4.52 \times 10^{-8}$). The most-replicated transcriptome result was the association of endometriosis with reduced *EEFSEC* expression in both whole blood and cultured fibroblast cells (min $P_{SMR} = 5.51 \times 10^{-7}$ and min $P_{TWAS} = 4.31 \times 10^{-11}$). The gene with the highest number of significant CpG sites (12) from

314 mSMR was *ESR1*. Methylation sites from intron 1 had positive SMR coefficients (highest B = 0.197,
315 P_{SMR} = 2.16 x 10^{-7}; probe cg04211581) whereas methylation sites from intron 2 had negative SMR
316 coefficients (lowest B = -0.283, P_{SMR} = 3.02 x 10^{-7}; probe cg24764793). Of the six significant RNA-
317 editing sites, the most robust site (chr3:156540497) showed six significant edSMR associations
318 across three tissues. This site in the *SSR3* gene demonstrated stronger effect sizes with the SCN.v1
319 (surgically-confirmed) GWAS than with the wide phenotype (B = -0.0769, P_{SMR} = 2.61 x 10^{-5} for
320 fibroblast cells and B = -0.0731, P_{SMR} = 3.06 x 10^{-5} for uterus).

321 Including genomic signals from MAGMA, we identified 282 genes across six -omics types,
322 each supported by respective biological evidence of endometriosis association. We performed
323 chi-squared tests for this gene list with gene sets curated by MSIGDB (Supplementary Table 17)
324 and identified several significant groups. From the Biocarta category of gene sets, there was only
325 one gene set identified, the CREM pathway (P = 1.66 x 10^{-7}). Top regulatory gene sets include
326 target genes of *MIR128_1_5P* (P = 9.68 x 10^{-11}), *ZNF92* (P = 7.70 x 10^{-10}), and *MIR124_5P* (P = 1.78
327 x 10^{-7}). The most significant biological processes from the GO (gene ontology) were sinoatrial
328 node cell development (P = 3.44 x 10^{-8}), mitochondrial fission (P = 6.99 x 10^{-8}), and regulation of
329 DNA ligation (P = 6.08 x 10^{-7}). The strongest hit from the GO molecular function category was
330 Wnt protein binding (P = 5.89 x 10^{-10}).

331 Discussion

332 We report the most diverse GWAS of over a million women for endometriosis to date
333 incorporating participants from 14 biobanks across the world. Out of 58 loci from 11 meta-
334 analyses, 29 are previously unreported (20 W, 3 Wex, 6 PCN and SCN) while 26 are in LD with (R^2
335 > 0.1) and three more are proximal to (1 Mb) known signals. We have specified biological

336 evidence of the genes or proteins relevant to 15 of the previously unreported loci. In addition to
337 specifying 85 putative causal variants, we also detected shared genetic architecture between
338 endometriosis and 135 complex polygenic phenotypes. Then, with the integration of five
339 additional -omics layers, we have implicated a total of 282 genes associated with endometriosis.

340 Our ancestry-stratified meta-analyses revealed a consistent observed SNP heritability of
341 9-13% for endometriosis for four studied genetic ancestry populations with sufficient size (AFR,
342 AMR, EAS, and EUR), which is higher than previously observed SNP heritability of 7% based on a
343 majority-European population[6]. The higher and more consistent heritability we observed can be
344 attributed to the unprecedented scale and diversity of our study cohort. The previous GWASs
345 have not included any African-ancestry (AFR), Admixed-American-ancestry (AMR), or
346 Central/South Asian (CSA+SAS) populations, which were essential in the computation of 12 / 13
347 (92%) credible sets in our fine-mapping analysis. We detected two multi-ancestry loci whose
348 associations were driven by the AFR cohorts (they were only also significant in the AFR meta-
349 analysis): *2p13.3* and *20q13.2*. By incorporating biobanks from various regions and ancestry
350 backgrounds worldwide, we were able to capture a more comprehensive genetic landscape of
351 endometriosis.

352 Given that diagnosis codes from the EHR can lack sensitivity for accurately phenotyping
353 endometriosis[29], we employed narrow phenotyping algorithms which incorporated procedure
354 codes (surgical and imaging). Despite the diminished sample sizes, there were still 13 genome-
355 wide significant loci: eight for Wex, four for PCN.v1, two for PCN.v2, and one for SCN.v1 (*21q22.13*
356 is shared between PCN.v1 and SCN.v1). Four of the loci from the more precise phenotypes were
357 successfully fine-mapped, including *FSHR:1* for Wex and *19q12* for PCN.v1, with one and seven

358 putative causal variants in their credible sets, respectively. MAGMA detected regional
359 associations with three genes (*LRRC2* with PCN.v1 and *GLB1L3* and *TCEA2* with Wex) that were
360 not detected in the largest GWAS. Leveraging multi-modal EHR data for more precise
361 phenotyping enhanced our ability to refine and validate genetic associations and uncover deeper
362 biological insights.

363 Endometriosis has multifaceted risk factors and complex clinical presentations with a
364 wide range of comorbidities and concomitant conditions. We explored genetic correlation and
365 pleiotropy with endometriosis on an unprecedented scale (3,065 tests), leveraging phenome-
366 wide summary statistics from All of Us and the UK Biobank along with metabolite, microbiome,
367 and complex proteome GWASs. With the addition of Mendelian randomization, we have yielded
368 invaluable hypothesis-generating results for further validation, even if genetic variants do not
369 always meet the requisite instrumental variable assumptions for detecting causality. Five
370 metabolites showed nominally significant ($p < 0.05$) genetic correlation with endometriosis,
371 including serotonin which is known to be related to endometrial stromal cell decidualization[30]
372 but is also impacted by SSRIs like citalopram which may improve endometriosis symptoms[31]. We
373 have observed vertical pleiotropy supporting a potential causal relationship between
374 endometriosis and the reduction of the gut microbiome family *bifidobacteriaciae* (order
375 *bifidobacteriales*, class *actinobacteria*), which plays a role in estrogen regulation[32,33].

376 Results from across our multi-omic analyses underscore the potential role of R-spondin 3
377 (*RSPO3*) in endometriosis pathophysiology as a previously-characterized promotor of fibrotic
378 proliferation[34]. *RSPO3*, known to modulate risk of endometriosis[35], interacts with *WNT4* via
379 Frizzled (*FZD*) receptors, influencing the WNT/Ca2+ and WNT/β-catenin pathways[36]. Wnt protein

380 binding (GOMF) was a significant pathway in our gene set analyses, with five genes overlapping
381 with our biologically-implicated endometriosis gene set: *CTHRC1* (mSMR), *FZD6* (fibroblast-
382 eSMR), *PTPRO* (fibroblast-, ovary-, and whole blood-TWAS), *RECK* (fibroblast-sSMR), and
383 *TRABD2A* (mSMR). Specifically, within the Wnt protein family, evidence for *WNT4's* inverse
384 association with endometriosis is found in six endometrial cellular contexts, mSMR (positive
385 coefficient) and uterus- and ovary-TWAS (negative coefficient); methylation suppresses gene
386 expression. *WNT4* interacts with *ESR1* (Estrogen Receptor Alpha), key regulator of endometrial
387 cell proliferation and survival throughout the menstrual cycle[37]. *ESR1* was positively associated
388 with endometriosis in 19 endometrial cellular contexts, and mSMR returned causal CpG site
389 associations with intron-dependent effect directions. There were three stromal cellular contexts
390 in which the expressions of *RSPO3*, *WNT4*, and *ESR1* were significantly different depending on
391 endometriosis status. During the proliferative phase, all three were upregulated in
392 nondecidualized stromal cells from donors with endometriosis; in the early secretory phase,
393 *WNT4* switched to being downregulated, and in the mid secretory phase, both *WNT4* and *RSPO3*
394 were diminished. Uncovering the interplay between these key genes throughout the menstrual
395 cycle will be essential for understanding endometriosis pathogenesis.

396 One hallmark of endometriotic lesions is cellular adhesion, the process by which cells
397 migrate and interact with the extracellular matrix[38,39]. An essential cytoskeletal protein which
398 plays a role in regulating adhesion and motility is vinculin (Vcn)[40]. We detected three
399 endometriosis-associated splicing events via sSMR and SWAS which support that expression of
400 an isoform of Vcn, metavinculin (MVcn, +exon 19), may reduce endometriosis risk. The AA
401 genotype at rs7896966 is associated with increased inclusion of exon 19[41], and our GWAS

402 identified that the AA genotype is also associated with decreased rates of endometriosis. It has
403 been demonstrated that MVcn expression contributes to fewer but larger focal adhesions per
404 cell[40] and that the inhibition of focal adhesion via FAK decreases the formation of endometriotic
405 lesions in mice[39]. While the GO gene set for focal adhesion assembly was not statistically
406 significant, there were four genes that overlapped with that pathway: *VCL* (sSMR), *WNT4* (mSMR,
407 uterus- and ovary-TWAS), *RHOD* (fibroblast-, uterus-, and ovary-eSMR), and *KDR* (six cellular
408 contexts, mSMR, PWAS, and fibroblast-, uterus-, and ovary-TWAS). *RHOD* expression (codes for
409 the GTPase RhoD) had a negative causal relationship with endometriosis, which aligns with
410 previously published results that show that silencing RhoD leads to less efficient cell migration[42].
411 *KDR* exhibited mixed effect directions. Plasma PWAS along with endothelial cells and
412 decidualized stromal cells supported positive associations whereas TWAS and stromal matrix
413 metalloproteinase cells from the secretory phase indicated a negative association with *KDR*
414 expression. When KDR (a.k.a. vascular endothelial growth factor receptor 2; *VEGFR2*) is
415 stimulated by vascular endothelial growth factor (VEGF), it stimulates DNA synthesis and
416 promotes focal-adhesion-related cell migration[43]. Although more investigation is required to
417 precisely elucidate how the proteins in this pathway interact in the context of endometriosis, our
418 data clearly support the hypothesis that the regulation of focal adhesion may be a strong
419 candidate for therapeutic targets[39].

420 Given the strong interplay between endometriosis and regulation of the menstrual cycle,
421 it is important to take note of genetic and molecular signatures which may indicate hormonal
422 dysregulation. *GREB1* (codes for Growth Regulating Estrogen Receptor Binding 1) is a well-
423 characterized endometriosis-associated gene. We identified 23 cellular contexts in which *GREB1*

424 expression is elevated in cells from donors with endometriosis, mostly from the cells and nuclei
425 sampled during the secretory phase and from donors exposed to exogenous hormones. All five
426 associations between elevated *GREB1* and endometriosis under the immune lineage were in
427 nucleus types from donors on hormones. Since hormones are a common first-line treatment of
428 endometriosis symptoms[44], it is important to investigate how these medications interact with
429 immune cell types that contribute to an inflammatory response. We have also replicated
430 associations with known methylation and splice sites at the *GREB1* locus[45], while also expanding
431 upon the multi-omic molecular signature. *GREB1* has a truncated isoform, GREB1c, and we
432 demonstrate that splicing events in ovarian and uterine tissue associated with this isoform
433 increase risk of endometriosis while splicing events associated with the canonical isoform
434 (GREB1a) are associated with decreased disease risk. Importantly, *GREB1* was not associated with
435 endometriosis on the eQTL or pQTL levels, emphasizing that in order to promote mechanistic
436 hypotheses of endometriosis, we must continue to incorporate comprehensive layers of multi-
437 omics.

438 Despite the large sample size, heritability observed with GWAS (10-13%) still fails to
439 measure up to the broad sense heritability estimation from a twin study of 47%[9]. Rare variants,
440 structural variants, nonlinear effects, or gene-environment interactions might contribute to
441 endometriosis risk but remain undetected in our analyses. As we leveraged the GWAS results to
442 study other -omics from the central dogma and beyond, one limitation was that we lacked
443 individual-level data. The results captured by xWAS and xSMR relied on previously-trained *cis*-
444 xQTL models that are limited to capturing only genotype-driven molecular trait variability. This
445 may leave out complex genetic architecture (*i.e. trans*-xQTLs) or environmentally-driven

446 molecular signatures potentially relevant to our phenotype of interest[46]. While our innovative
447 and comprehensive *in silico* multi-omic approach yielded many testable hypotheses, we did not
448 perform any experimental functional validation. Furthermore, although our GWAS constitutes
449 the most diverse genetic study of endometriosis performed so far (32% non-European), we still
450 lacked the power to draw many meaningful conclusions in understudied populations. We want
451 to emphasize that the field should continue prioritizing the inclusion of diverse cohorts in genetic
452 studies.

453 In addition to replicating findings from previous studies, we uncovered new insights that
454 advanced our understanding of the genetic underpinnings of endometriosis. Our integrative
455 multi-omics approach, combining genomic data with single-cell analyses along with proteomic,
456 splice-omic, transcriptomic, methylomic, and RNA edit-omic tests, has biologically implicated 282
457 genes, providing unprecedented insights into the molecular mechanisms driving this chronic
458 condition. These findings lay a robust foundation for future functional studies to elucidate the
459 precise roles of identified genes and pathways in endometriosis pathogenesis. Moreover, our
460 results have important clinical implications, potentially informing the development of more
461 accurate diagnostic tools, personalized risk prediction models, and targeted therapeutic
462 interventions. As we move forward, this work emphasizes the critical importance of large-scale,
463 diverse genetic studies in unraveling the complexities of multifactorial diseases like
464 endometriosis, paving the way for improved patient outcomes and a deeper understanding of
465 women's health issues globally.

466 # Methods

467 ## Phenotyping

468 Endometriosis diagnosis requires direct observation of lesions by laparoscopic surgery or
469 sometimes imaging (MRI/ultrasound). We can extract diagnoses from medical records via
470 diagnosis billing codes in the form of ontologies like the international classification of diseases
471 (ICD) or the systematized nomenclature of medicine (SNOMED). For the EHR-linked biobanks, we
472 used structured data to define the phenotypes. Cases for wide endometriosis (W) were any
473 women with a history of ICD-9 (617.*), ICD-10 (N80.*), or SNOMED codes for endometriosis. For
474 the wide phenotype excluding adenomyosis (Wex), women with a history of ICD-9 (617.0), ICD-
475 10 (N80.0), or SNOMED codes for uterine endometriosis were excluded from both cases and
476 controls. Our two narrow case definitions were procedure-confirmed and surgically-confirmed
477 (PCN and SCN). These narrow phenotypes were designed to capture confirmed cases and controls
478 of endometriosis based on procedure history including hysterectomies, laparoscopies and
479 ultrasounds. The list of procedure codes (CPT-4 or OPCS) for PCN included all the surgery codes
480 in SCN but added non-obstetric ultrasound codes to account for potential imaging diagnoses.
481 Each of the two narrow phenotypes was tested in two versions: cases versus all controls (PCN.v1
482 and SCN.v1) and cases versus confirmed controls (PCN.v2 and SCN.v2) where confirmed controls
483 were those who had history of the corresponding procedure codes for each phenotype with no
484 history of endometriosis diagnosis. For additional details on the phenotyping algorithms, see
485 Supplemental Appendix 1.

Genotyping by Biobank

Most contributing biobanks genotyped several hundred thousand variants. After those data are collected, then the rest of the genomic variants can be imputed probabilistically based on a reference panel such as TOPMED. Some biobanks such as AOU used short read whole genome sequencing (WGS) to gather their genomic data. The sequences from the short reads are aligned and compared to the reference to produce variant calls. References and platform information for each biobank's genotyping method and QC can be found in Supplementary Table 18.

Genetic Association Testing

Each contributing biobank performed ancestry-stratified association testing with up to six phenotypes (W, Wex, PCN.v1, PCN.v2, SCN.v1, and SCN.v2), depending on whether they had at least 50 cases for that phenotype and ancestry combination. Genetically-informed ancestry (GIA) is defined based on genetic distance from subpopulations of a reference group such as the 1000 Genomes Project dataset[47] (1KG). Genetic distance is measured using a dimensionality reduction technique such as principal component analysis (PCA). Ancestry is then assigned using a classifier such as a mixture model or k-nearest neighbors. For references and methods used for assigning GIA within each biobank, see Supplementary Table 19.

Contributing biobanks used linear mixed models to estimate the effect of each variant on the six phenotypes. Tools implemented for these tests include SAIGE[48] and Regenie[49]. Association tests were adjusted for principal components, age at EHR data collection, and any biobank-specific batch variables (for example, collection site within eMERGE). For estimation of the null models (SAIGE or Regenie step 1), variants were required to have a call rate of at least 95%, a minor allele frequency of 0.01, and a Hardy-Weinberg p-value of at least $1x10^{-6}$. For association

508 testing (SAIGE or Regenie step 2), variants were required to have a minor allele count of at least
509 20, and an imputation quality of at least 0.6. Individuals were included if their overall genotyping
510 rate was at least 95%. A summary of each biobank's exact covariates and QC procedures is in
511 Supplementary Table 20. We leveraged publicly-available FinnGen[50] summary statistics to
512 increase the sample size.

513 After biobank analyses were completed and collected, summary statistics were cleaned
514 and lifted into genome build hg38 as necessary. Then, inverse variance-weighted, fixed-effect
515 meta-analyses were performed using GWAMA (software for genome-wide association meta-
516 analysis)[51]. GWAMA adjusts each study based on its overall genomic inflation factor. Prior to
517 meta-analysis, input summary statistics were restricted to variants with a minor allele frequency
518 of at least 0.005 to ensure stability in the estimates.

519 To define significantly-associated loci, we started by applying the clumping function of
520 Plink 1.9[52] with the lead SNP p-value threshold set to $5x10^{-8}$, the secondary SNP threshold set to
521 1, minimum R^2 of 0.1, and a window of 1000kb. 1KG was used as a linkage disequilibrium (LD)
522 reference, matching the whole dataset to the multi-ancestry meta-analyses and single-ancestry
523 subsets of the dataset to their respective meta-analyses. After the variant clumps were identified,
524 any clumps that were physically overlapping with one another were merged into loci. Then, we
525 tested if any loci were in LD with ($R^2 > 0.25$) or proximal to tag SNPs of the 39 autosomal lead
526 SNPs from the 2023 GWAS[6] to determine whether our hits were known or previously
527 unreported[16]. Each locus is labeled with its cytoband, but if the corresponding gene region was
528 statistically significant in the MAGMA, xSMR, or xWAS tests (described below), or if the GWAS

529 locus is in contact with a gene based on endometrial cell chromatin looping data[53], that gene
530 symbol was added to the locus label.

531 Estimation of Observed SNP Heritability and Genetic Correlation Amongst Endometriosis 532 GWASs

533 We estimated heritability with the LD-Score Regression tool, LDSC[54], for all single-ancestry
534 meta-analyses representing at least two thousand cases. LD scores were computed for HapMap
535 variants[55] from the respective super-population subgroups from 1KG for each ancestry-stratified
536 meta-analysis. We also estimated partitioned heritability[56] with variant annotation groups
537 coming from ENSEMBL regulatory features GRCh38.105[57] and uterus-specific regulatory features
538 from EpiMap[58]. We estimated cross-ancestry genetic correlation of our meta-analyses using
539 Popcorn[59]. Popcorn cross-population LD scores were derived from the same set of variants and
540 samples from 1KG as in LD score computation for LD score regression.

541 Statistical Fine-Mapping for Identification of Putative Causal Variants

542 The multi-ancestry sum of the single effects model (MESuSiE)[60] was used to identify
543 putative causal variants in each significant GWAS locus from the multi-ancestry meta-analyses.
544 For each locus, we extracted the single-ancestry summary statistics from that region in addition
545 to computing ancestry-specific LD matrices for AFR, AMR, EAS, EUR, and SAS. MESuSiE analyzes
546 whether causal variants are shared between ancestry groups by computing a posterior inclusion
547 probability (PIP) for each ancestry alone and for each combination of two or more ancestries.
548 SNPs with any PIP value greater than 0.5 are part of the credible set for that locus.

549 ## Variant-Set Enrichment for Identifying Enriched Gene Regions

550 MAGMA[61] was used to assess gene regions within the GWAS results for enrichment.
551 MAGMA uses a model based on multiple regression to test the association of a phenotype with
552 groups of variants. Testing groups of variants increases statistical power over the single-variant
553 GWAS tests performed[61]. We utilized gene region definitions from NCBI for human genome build
554 38, testing between 18,691-18,846 genes for each GWAS meta-analysis (MAGMA test counts and
555 p-value thresholds are available in Supplementary Table 21).

556 ## Phenome-Wide and Multi-Omic Genetic Correlation and Mendelian Randomization Tests

557 The AFR, EAS, and EUR wide phenotype meta-analyses were utilized for LDSC[54] genetic
558 correlation (GC) tests with phenome-wide GWAS summary statistics from AOU[62] and UKB (EUR-
559 only). In AOU, we tested for correlation with binary traits that had at least 2,000 cases
560 (medications and phecodeX phenotypes) and quantitative traits with at least 2,000 samples
561 overall (labs and biometric measurements), totaling 1,728 tests. From UKB, we utilized summary
562 statistics for diagnoses and survey variables, both binary and quantitative, with heritability
563 estimates for which the Z-score was at least 4 (527 studies)[63]. Summary statistics were also
564 obtained for 169 metabolites[64], 165 gut microbiome taxa[65], and 487 proteins[66] which we were
565 unable to map to one genomic location (mappable proteins were utilized in SMR described
566 below). Reference LD scores were computed from the corresponding 1KG superpopulation
567 groups across Hapmap variants[55]. For the larger-sample tests (AOU and UKBB phenome), genetic
568 correlation results with a p-value of < 0.05 were candidates for mendelian randomization, and
569 for the smaller tests (proteome and microbiome), traits with an h^2 p-value or rg p-value < 0.05
570 were nominated. Effect estimates were not available for the metabolites, so we could not

571 perform MR on those. Significant GC estimates were those that reached a p-value < 1.63×10^{-5}
572 (correcting for 3,065 total tests).

573 To distinguish horizontal and vertical pleiotropy, we performed secondary mendelian
574 randomization (MR) tests on the 890 nominated traits using the MendelianRandomization R
575 package[67]. Several MR methods were utilized as a sensitivity analysis for the estimates: inverse-
576 variance weighted (IVW), debiased IVW, lasso-penalized IVW, weighted median, maximum
577 likelihood, and MR-Egger (random-effects). We tested in both directions: endometriosis causing
578 the test trait and vice-versa). Instrumental variables (IVs) were the genome-wide significant
579 variants from the HapMap-filtered[55] exposure summary statistics, and we used correlation
580 matrices calculated with PLINK 2.0[68] to adjust for the covariance between IVs. Significant causal
581 relationships were those for which the maximum p-value across methods was < 0.05, the
582 minimum p-value was < 2.84×10^{-5} (1,760 total bidirectional tests), all estimates were the same
583 direction, and that direction matched the direction of the genetic correlation estimate.

584 ## Integration with Single-Cell Data

585 Single cell data are rich resources for understanding the biology of diseases. A recently
586 published endometrial single cell atlas includes gene expression data for cells, nuclei, immune
587 cells, and immune nuclei, covering over 40 different cell types, three menstrual phases, and four
588 different endometrial pathologies[69]. We leveraged these data to compute single-cell disease-
589 relevance scores (scDRSs) based on the associated genes from GWAS, as computed by MAGMA.
590 Disease risk scores were computed and normalized using the scDRS software[70].

591 For each cell type and menstrual phase ("cellular context"), we computed the logistic
592 regression coefficient of the scDRS comparing cells from donors with and without endometriosis.

593 These regression models were adjusted for age; one source dataset only had an age range (18-
594 36), so we assigned age with a random uniform distribution to those cells. These regression
595 coefficients were used to prioritize cellular contexts (cell type – menstrual stage pairs). Significant
596 contexts were those which had a coefficient one-sided p-value of < 0.05 / 240 tests. For
597 prioritized contexts, we compared the expression of the top MAGMA genes (34 genes) between
598 case and control donors using t-tests to identify which genes were driving the distinguishing
599 effect of the scDRS. Significant t-tests were those with a p-value < 0.05 / 1,190 (34 genes x 35
600 significant contexts).

601 ## Multi-Omic Imputed Association Studies and Mendelian Randomization

602 We used GWAS summary statistics from the 11 meta-analyses to estimate the association
603 between endometriosis, transcript expression (TWAS), protein expression (PWAS), and intron
604 inclusion (SWAS). For TWAS and SWAS, we used precomputed eQTL and sQTL weights,
605 respectively, from the PredictDB data repository (http://predictdb.org/) which represented *cis*-
606 level associations in uterine tissue, ovarian tissue, whole blood, and cultured fibroblast cells. For
607 PWAS, we used whole blood pQTL models from *Schubert et al*[71] which are based on the
608 SOMAscan assay data measured in multi-ethnic Trans-omics for Precision Medicine (TOPMed)
609 Multi-omics pilot study. We used baseline models only (no covariates) which had a predictive
610 correlation (rho) of at least 0.1 and a z-score p-value threshold of 0.05, ensuring that the
611 association between the SNPs and protein expression was statistically significant. The xQTL
612 models along with the GWAS summary statistics were the input for summary-based PrediXcan
613 (S-PrediXcan)[72]. We used Bonferroni-correct p-value thresholds to identify statistically significant
614 associations for each tissue study. The alpha value was 0.05 and the numbers of tests for each

615 GWAS-tissue-QTL combination can be found in Supplementary Table S22. For labeling sQTLs,
616 GTEx provided a mapping from introns to transcripts.

617 Summary-Based Mendelian Randomization (SMR) is a publicly-available softare tool
618 which leverages uploaded GWAS summary statistics to detect pleiotropic associations between
619 molecular phenotypes as the exposure and a trait of interest as the outcome[73]. We tested all
620 combinations between the 11 GWASs and the following QTL datasets: GTEx eQTLs, sQTLs, and
621 edQTLs for ovary, uterus, whole blood, and fibroblast cells, plasma pQTLs from the FENLAND
622 study[66], McRae whole blood mQTLs[74], and Hatton blood mQTLs derived from EAS and EUR
623 populations[75]. The QTL was the exposure with endometriosis being the outcome. Significant
624 causal associations were identified as those with a p-value less than 0.05 / the number of tests
625 for each GWAS-tissue-dataset combination. Then, significantly heterogeneous results (HEIDI P <
626 0.05 / the number of significant tests) were discarded. The test counts and the corrected p-value
627 thresholds for the xSMR analyses are available in Supplemental Table S23.

628 Once we aggregated all genes resulting from MAGMA, xWAS, and xSMR, we performed basic
629 gene set testing using chi-squared statistics comparing our gene list to each pathway's list of
630 genes. The gene sets were from MSIGDB and included hallmark pathways, cell type signatures,
631 co-regulation pathways, and GO ontology pathways. The numbers of gene set tests corrected for
632 are available in Supplementary Table S24.

633 Data and Code Availability

634 All data required to recreate figures are available in supplemental tables S1-S24. GWAS summary
635 statistics will be uploaded to the GWAS catalog upon manuscript acceptance.

636 The availability of individual-level data varies by biobank. AOU, eMERGE, and UKBB are
637 available to researchers by application. The other biobanks (BBJ, BioVU, CCPM, FinnGen, GHI,
638 HUNT, Latvian National Biobank, MGBB, MVP, PMBB, and TWB) only allow access to institutional
639 researchers via their respective IRB protocols.

640 Code written and run by the coordinating center (PMBB) is available in this github
641 repository: https://github.com/Setia-Verma-Lab/guare_et_al_endometriosis_gwas_2025.

642 Acknowledgments

643 Research reported in this publication was supported by the Eunice Kennedy Shriver National
644 Institute of Child Health and Human Development of the National Institutes of Health under
645 award number R01HD110567.

646 We acknowledge the Penn Medicine BioBank (PMBB) for providing data and thank the
647 patient-participants of Penn Medicine who consented to participate in this research program.
648 We would also like to thank the Penn Medicine BioBank team and Regeneron Genetics Center
649 for providing genetic variant data for analysis. The PMBB is approved under IRB protocol# 813913
650 and supported by Perelman School of Medicine at University of Pennsylvania, a gift from the
651 Smilow family, and the National Center for Advancing Translational Sciences of the National
652 Institutes of Health under CTSA award number UL1TR001878.

653 This phase of the eMERGE Network was initiated and funded by the NHGRI through the
654 following grants: U01HG008657 (Group Health Cooperative/University of Washington);
655 U01HG008685 (Brigham and Women's Hospital); U01HG008672 (Vanderbilt University Medical
656 Center); U01HG008666 (Cincinnati Children's Hospital Medical Center); U01HG006379 (Mayo

657 Clinic); U01HG008679 (Geisinger Clinic); U01HG008680 (Columbia University Health Sciences);
658 U01HG008684 (Children's Hospital of Philadelphia); U01HG008673 (Northwestern University);
659 U01HG008701 (Vanderbilt University Medical Center serving as the Coordinating Center);
660 U01HG008676 (Partners Healthcare/Broad Institute); and U01HG008664 (Baylor College of
661 Medicine).

662 We gratefully acknowledge All of Us participants for their contributions, without whom
663 this research would not have been possible. We also thank the National Institutes of Health's All
664 of Us Research Program for making available the participant data examined in this study.

665 This research has been conducted using the Taiwan Biobank Resource. We thank all the
666 participants and investigators of the Taiwan Biobank. We thank the National Center for Genome
667 Medicine, Academia Sinica, Taiwan, for the technical support in the genotyping. We thank
668 National Core Facility for Biopharmaceuticals (NCFB, 112-2740-B-492-001) and National Center
669 for High-performance Computing (NCHC) of National Applied Research Laboratories (NARLabs)
670 of Taiwan for providing computational resources and storage resources. This study was
671 supported by the National Health Research Institutes (NP-109, 110, 111, 112, 113-PP-09),
672 Ministry of Science and Technology (MOST 109-2314-B-400-017, 110-2314-B-400-028-MY3), and
673 National Science and Technology Council (NSCT 113-2628-B-400-002), Taiwan.

674 The Trøndelag Health Study (The HUNT Study) is a collaboration between HUNT Research
675 Centre (Faculty of Medicine and Health Sciences, NTNU, Norwegian University of Science and
676 Technology), Trøndelag County Council, Central Norway Regional Health Authority, and the
677 Norwegian Institute of Public Health. The genotyping in HUNT was financed by the National
678 Institutes of Health; University of Michigan; the Research Council of Norway; the Liaison

679 Committee for Education, Research and Innovation in Central Norway; and the Joint Research
680 Committee between St Olavs hospital and the Faculty of Medicine and Health Sciences, NTNU.
681 The genetic investigations of the HUNT Study are a collaboration between researchers from the
682 HUNT Center for Molecular and Clinical Epidemiology (formerly known as the K.G. Jebsen Center
683 for Genetic Epidemiology as of August 1st 2023), NTNU, and the University of Michigan Medical
684 School and the University of Michigan School of Public Health. We thank HUNT participants for
685 donating their time, samples, and information to help others; clinicians and other employees at
686 Nord-Trøndelag Hospital Trust for their support and for contributing to data collection.

687 The authors thank Million Veteran Program (MVP) staff, researchers, and volunteers, who
688 have contributed to MVP, and especially participants who previously served their country in the
689 military and now generously agreed to enroll in the study.

690 Support for title page creation and format was provided by AuthorArranger, a tool
691 developed at the National Cancer Institute.

692 References

693 1. Christ, J. P. *et al.* Incidence, prevalence, and trends in endometriosis diagnosis: a United
694 States population-based study from 2006 to 2015. *American Journal of Obstetrics and*
695 *Gynecology* **225**, 500.e1-500.e9 (2021).
696 2. Agarwal, S. K. *et al.* Clinical diagnosis of endometriosis: a call to action. *American Journal of*
697 *Obstetrics and Gynecology* **220**, 354.e1-354.e12 (2019).
698 3. Fauconnier, A. & Chapron, C. Endometriosis and pelvic pain: epidemiological evidence of the
699 relationship and implications. *Human Reproduction Update* **11**, 595–606 (2005).

700 4. Coccia, M. E., Nardone, L. & Rizzello, F. Endometriosis and Infertility: A Long-Life Approach to
701 Preserve Reproductive Integrity. *International Journal of Environmental Research and Public*
702 *Health* **19**, 6162 (2022).
703 5. Sapkota, Y. *et al.* Meta-analysis identifies five novel loci associated with endometriosis
704 highlighting key genes involved in hormone metabolism. *Nat Commun* **8**, 15539 (2017).
705 6. Rahmioglu, N. *et al.* The genetic basis of endometriosis and comorbidity with other pain and
706 inflammatory conditions. *Nat Genet* **55**, 423–436 (2023).
707 7. Koller, D. *et al.* Multi-ancestry genome-wide association study of endometriosis and its
708 clinical manifestations in ∼1.4 million women: translating gene discovery into pathogenic
709 mechanisms and therapeutic targets. 2025.09.03.25335012 Preprint at
710 https://doi.org/10.1101/2025.09.03.25335012 (2025).
711 8. Fatumo, S. *et al.* A roadmap to increase diversity in genomic studies. *Nat Med* **28**, 243–250
712 (2022).
713 9. Saha, R. *et al.* Heritability of endometriosis. *Fertility and Sterility* **104**, 947–952 (2015).
714 10. Zhou, W. *et al.* Global Biobank Meta-analysis Initiative: Powering genetic discovery across
715 human disease. *Cell Genomics* **2**, 100192 (2022).
716 11. Zhao, H. *et al.* Proteome-wide Mendelian randomization in global biobank meta-analysis
717 reveals multi-ancestry drug targets for common diseases. *Cell Genomics* **2**, (2022).
718 12. Tsuo, K. *et al.* Multi-ancestry meta-analysis of asthma identifies novel associations and
719 highlights the value of increased power and diversity. *Cell Genomics* **2**, (2022).
720 13. Faro, V. L. *et al.* Novel ancestry-specific primary open-angle glaucoma loci and shared biology
721 with vascular mechanisms and cell proliferation. *CR Med* **5**, (2024).

722 14. Verma, A. *et al.* The Penn Medicine BioBank: Towards a Genomics-Enabled Learning
723 Healthcare System to Accelerate Precision Medicine in a Diverse Population. *Journal of*
724 *Personalized Medicine* **12**, 1974 (2022).
725 15. Genomic data in the All of Us Research Program. *Nature* **627**, 340–346 (2024).
726 16. Verma, A. *et al.* Diversity and scale: Genetic architecture of 2068 traits in the VA Million
727 Veteran Program. *Science* **385**, eadj1182 (2024).
728 17. Nagai, A. *et al.* Overview of the BioBank Japan Project: Study design and profile. *Journal of*
729 *Epidemiology* **27**, S2–S8 (2017).
730 18. Pulley, J., Clayton, E., Bernard, G. R., Roden, D. M. & Masys, D. R. Principles of human subjects
731 protections applied in an opt-out, de-identified biobank. *Clin Transl Sci* **3**, 42–48 (2010).
732 19. Roden, D. *et al.* Development of a Large-Scale De-Identified DNA Biobank to Enable
733 Personalized Medicine. *Clinical Pharmacology & Therapeutics* **84**, 362–369 (2008).
734 20. Wiley, L. K. *et al.* Building a vertically integrated genomic learning health system: The biobank
735 at the Colorado Center for Personalized Medicine. *The American Journal of Human Genetics*
736 **111**, 11–23 (2024).
737 21. Gottesman, O. *et al.* The Electronic Medical Records and Genomics (eMERGE) Network: past,
738 present, and future. *Genet Med* **15**, 761–771 (2013).
739 22. David, S. P. *et al.* Personalized medicine in a community health system: the NorthShore
740 experience. *Front. Genet.* **14**, (2023).
741 23. Brumpton, B. M. *et al.* The HUNT study: A population-based cohort for genetic research. *Cell*
742 *Genomics* **2**, (2022).

743 24. Rovite, V. *et al.* Genome Database of the Latvian Population (LGDB): Design, Goals, and
744 Primary Results. *Journal of Epidemiology* **28**, 353–360 (2018).
745 25. Boutin, N. T. *et al.* The Evolution of a Large Biobank at Mass General Brigham. *Journal of*
746 *Personalized Medicine* **12**, 1323 (2022).
747 26. Fan, C.-T., Lin, J.-C. & Lee, C.-H. Taiwan Biobank: A Project Aiming to Aid Taiwan's Transition
748 into a Biomedical Island. *Pharmacogenomics* **9**, 235–246 (2008).
749 27. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature*
750 **562**, 203–209 (2018).
751 28. Katon, J. G., Plowden, T. C. & Marsh, E. E. Racial disparities in uterine fibroids and
752 endometriosis: a systematic review and application of social, structural, and political context.
753 *Fertility and Sterility* **119**, 355–363 (2023).
754 29. McKillop, M. Phenotyping Endometriosis from Observational Health Data. (Columbia
755 University, 2019). doi:10.7916/d8-1est-dh56.
756 30. Tian, J. *et al.* Dysregulation of endometrial stromal serotonin homeostasis leading to
757 abnormal phosphatidylcholine metabolism impairs decidualization in patients with recurrent
758 implantation failure. *Hum Reprod Open* **2024**, hoae042 (2024).
759 31. Hosseini, R., Mirzaei, S., Asgari, Z., Hajiloo, N. & Garrosi, L. Efficacy of adjunct therapy with
760 citalopram to improve health-related quality of life and associated symptoms in patients with
761 endometriosis: A randomized clinical trial. *Onkologia i Radioterapia* **18**, 1 (2024).
762 32. Uzuner, C., Mak, J., El-Assaad, F. & Condous, G. The bidirectional relationship between
763 endometriosis and microbiome. *Front. Endocrinol.* **14**, (2023).

764 33. Salliss, M. E., Farland, L. V., Mahnert, N. D. & Herbst-Kralovetz, M. M. The role of gut and
765 genital microbiota and the estrobolome in endometriosis, infertility and chronic pelvic pain.
766 *Hum Reprod Update* **28**, 92–131 (2022).
767 34. Zhang, M. *et al.* Targeting the Wnt signaling pathway through R-spondin 3 identifies an anti-
768 fibrosis treatment strategy for multiple organs. *PLoS One* **15**, e0229445 (2020).
769 35. Tao, T., Mo, X. & Zhao, L. Identifying novel potential drug targets for endometriosis via plasma
770 proteome screening. *Front Endocrinol (Lausanne)* **15**, 1416978 (2024).
771 36. Scholz, B. *et al.* Endothelial RSPO3 Controls Vascular Stability and Pruning through Non-
772 canonical WNT/Ca(2+)/NFAT Signaling. *Dev Cell* **36**, 79–93 (2016).
773 37. Gallagher, C. S. *et al.* Genome-wide association and epidemiological analyses reveal common
774 genetic origins between uterine leiomyomata and endometriosis. *Nat Commun* **10**, 4857
775 (2019).
776 38. Psilopatis, I. *et al.* The Hallmarks of Endometriosis. *Geburtshilfe und Frauenheilkunde* **84**,
777 555–563 (2024).
778 39. Nagai, T. *et al.* Focal Adhesion Kinase-Mediated Sequences, Including Cell Adhesion,
779 Inflammatory Response, and Fibrosis, as a Therapeutic Target in Endometriosis. *Reprod. Sci.*
780 **27**, 1400–1410 (2020).
781 40. Lee, H. T. *et al.* Vinculin and metavinculin exhibit distinct effects on focal adhesion properties,
782 cell migration, and mechanotransduction. *PLoS One* **14**, e0221962 (2019).
783 41. Garrido-Martín, D., Borsari, B., Calvo, M., Reverter, F. & Guigó, R. Identification and analysis
784 of splicing quantitative trait loci across multiple tissues in the human genome. *Nat Commun*
785 **12**, 727 (2021).

786 42. Blom, M., Reis, K., Heldin, J., Kreuger, J. & Aspenström, P. The atypical Rho GTPase RhoD is a
787 regulator of actin cytoskeleton dynamics and directed cell migration. *Experimental Cell*
788 *Research* **352**, 255–264 (2017).
789 43. Kanno, S. *et al.* Roles of two VEGF receptors, Flt-1 and KDR, in the signal transduction of VEGF
790 effects in human vascular endothelial cells. *Oncogene* **19**, 2138–2146 (2000).
791 44. Brown, J. & Farquhar, C. An Overview of Treatments for Endometriosis. *JAMA* **313**, 296–297
792 (2015).
793 45. Mortlock, S. *et al.* Global endometrial DNA methylation analysis reveals insights into mQTL
794 regulation and associated endometriosis disease risk and endometrial function. *Commun Biol*
795 **6**, 780 (2023).
796 46. Mai, J., Lu, M., Gao, Q., Zeng, J. & Xiao, J. Transcriptome-wide association studies: recent
797 advances in methods, applications and available databases. *Commun Biol* **6**, 1–10 (2023).
798 47. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65
799 (2012).
800 48. Zhou, W. *et al.* Efficiently controlling for case-control imbalance and sample relatedness in
801 large-scale genetic association studies. *Nat Genet* **50**, 1335–1341 (2018).
802 49. Mbatchou, J. *et al.* Computationally efficient whole-genome regression for quantitative and
803 binary traits. *Nat Genet* **53**, 1097–1103 (2021).
804 50. Kurki, M. I. *et al.* FinnGen provides genetic insights from a well-phenotyped isolated
805 population. *Nature* **613**, 508–518 (2023).
806 51. Mägi, R. & Morris, A. P. GWAMA: software for genome-wide association meta-analysis. *BMC*
807 *Bioinformatics* **11**, 288 (2010).

808 52. Purcell, S. *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based
809 Linkage Analyses. *Am J Hum Genet* **81**, 559–575 (2007).
810 53. O'Mara, T. A., Spurdle, A. B. & Glubb, D. M. Analysis of Promoter-Associated Chromatin
811 Interactions Reveals Biologically Relevant Candidate Target Genes at Endometrial Cancer Risk
812 Loci. *Cancers (Basel)* **11**, 1440 (2019).
813 54. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in
814 genome-wide association studies. *Nat Genet* **47**, 291–295 (2015).
815 55. Manolio, T. A. & Collins, F. S. The HapMap and Genome-Wide Association Studies in Diagnosis
816 and Therapy*. *Annual Review of Medicine* **60**, 443–456 (2009).
817 56. Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide
818 association summary statistics. *Nat Genet* **47**, 1228–1235 (2015).
819 57. Harrison, P. W. *et al.* Ensembl 2024. *Nucleic Acids Research* **52**, D891–D899 (2024).
820 58. Boix, C. A., James, B. T., Park, Y. P., Meuleman, W. & Kellis, M. Regulatory genomic circuitry
821 of human disease loci by integrative epigenomics. *Nature* **590**, 300–307 (2021).
822 59. Brown, B. C., Ye, C. J., Price, A. L. & Zaitlen, N. Transethnic Genetic-Correlation Estimates from
823 Summary Statistics. *The American Journal of Human Genetics* **99**, 76–88 (2016).
824 60. Gao, B. & Zhou, X. MESuSiE enables scalable and powerful multi-ancestry fine-mapping of
825 causal variants in genome-wide association studies. *Nat Genet* **56**, 170–179 (2024).
826 61. Leeuw, C. A. de, Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: Generalized Gene-Set
827 Analysis of GWAS Data. *PLOS Computational Biology* **11**, e1004219 (2015).

828 62. Overview of the All by All tables available on the All of Us Researcher Workbench. *User*
829 *Support* https://support.researchallofus.org/hc/en-us/articles/27049847988884-Overview-
830 of-the-All-by-All-tables-available-on-the-All-of-Us-Researcher-Workbench (2025).
831 63. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat*
832 *Genet* **47**, 1236–1241 (2015).
833 64. Hysi, P. G. *et al.* Metabolome Genome-Wide Association Study Identifies 74 Novel Genomic
834 Regions Influencing Plasma Metabolites Levels. *Metabolites* **12**, 61 (2022).
835 65. Kurilshikov, A. *et al.* Large-scale association analyses identify host factors influencing human
836 gut microbiome composition. *Nat Genet* **53**, 156–165 (2021).
837 66. Pietzner, M. *et al.* Mapping the proteo-genomic convergence of human diseases. *Science* **374**,
838 eabj1541 (2021).
839 67. Patel, A. *et al.* MendelianRandomization v0.9.0: updates to an R package for performing
840 Mendelian randomization analyses using summarized data. *Wellcome Open Res* **8**, 449
841 (2023).
842 68. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer
843 datasets. *GigaScience* **4**, s13742-015-0047–8 (2015).
844 69. Marečková, M. *et al.* An integrated single-cell reference atlas of the human endometrium.
845 *Nat Genet* **56**, 1925–1937 (2024).
846 70. Zhang, M. J. *et al.* Polygenic enrichment distinguishes disease associations of individual cells
847 in single-cell RNA-seq data. *Nat Genet* **54**, 1572–1580 (2022).
848 71. Schubert, R. *et al.* Protein prediction for trait mapping in diverse populations. *PLOS ONE* **17**,
849 e0264341 (2022).

850 72. Gamazon, E. R. *et al.* A gene-based association method for mapping traits using reference
851 transcriptome data. *Nat Genet* **47**, 1091–1098 (2015).
852 73. Zhu, Z. *et al.* Integration of summary data from GWAS and eQTL studies predicts complex
853 trait gene targets. *Nat Genet* **48**, 481–487 (2016).
854 74. McRae, A. F. *et al.* Identification of 55,000 Replicated DNA Methylation QTL. *Sci Rep* **8**, 17605
855 (2018).
856 75. Hatton, A. A. *et al.* Genetic control of DNA methylation is largely shared across European and
857 East Asian populations. *Nat Commun* **15**, 2713 (2024).
858
859
860