Expanding the genetic landscape of endometriosis: Integrative -omics analyses uncover key pathways from a multi-ancestry study of over 900,000 women

Authors and Affiliations

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], Vita Rovite[8], Alan R Sanders[7], Marija Simona Dombrovska[8], Noemie Elhadad[9], Andrew Hill[10], Gail Jarvik[11], James Jaworski[12], Yuan Luo[13], Shinichi Namba[14], Yukinori Okada[14], Yue Shi[15], Yuya Shirai[16], Jonathan Shortt[10], Wei-Qi Wei[17], Chunhua Weng[9], Yuji Yamamoto[16], Sinead Chapman[15], Wei Zhou[15], Digna R. Velez Edwards[18], Shefali Setia-Verma[1]

Corresponding Author: Shefali Setia-Verma

[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 Quantitaive 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, Norway, [5]Center for Neuropsychiatric Research, National Health Research Institutes, Maoli County, Taiwan, [6]Cincinnati Children's Hospital Medical Center, Cincinatti, OH, United States,

[7]Endeavor Health, Evanston, IL, United States, [8]Latviant Biomedical Research and Study Centre, Riga, Riga, Latvia, [9]Biomedical Informatics, Columbia University, New York City, NY, United States, [10]Colorado for Precision Medicine, Aurora, CO, United States, [11]Division of Medical Genetics, University of Washington, Seattle, WA, United States, [12]Division of Epidemiology, Department of Medicine, Vanderbilt University Medical Center, Nashville, Tennessee, United States, [13]Feinberg School of Medicine, Northwestern University, Evanston, IL, United States, [14]Genome Informatics, The University of Tokyo, Tokyo, Japan, [15]Broad Institute, Cambridge, MA, United States, [16]Statistical Genetics, Osaka University, Suita, Osaka, Japan, [17]Biomedical Informatics, Vanderbilt University, Nashville, TN, United States, [18]Division of Quantitative Science, Department of Obstetrics and Gynecology, Vanderbilt University Medical Center, Nashville, Tennessee, United States

Abstract

We report the findings of a genome-wide association study (GWAS) meta-analysis of endometriosis consisting of a large portion (31%) of non-European samples across 14 biobanks worldwide as part of the Global Biobank Meta-Analysis Initiative (GBMI). We identified 45 significant loci using a wide phenotype definition, seven of which are previously unreported and detected first genome-wide significant locus (*POLR2M*) among only African-ancestry. Our narrow phenotypes and surgically confirmed case definitions for endometriosis analyses replicated the known loci near *CDC42*, *SKAP1*, and *GREB1*. Through this large ancestry stratified analyses, we document heritability estimates in range of 10-12% for all ancestral groups. Thirty-eight loci had at least one variant in the credible set after fine-mapping. An imputed

transcriptome-wide association study (TWAS) identified 11 associated genes (two previously unreported), while the proteome-wide association study (PWAS) suggests significant association of R-spondin 3 (RSPO3) with wide endometriosis, which plays a crucial role in modulating the Wnt signaling pathway. Our diverse, comprehensive GWAS, coupled with integrative -omics analysis, identifies critical roles of immunopathogenesis, Wnt signaling, and balance between proliferation, differentiation, and migration of endometrial cells as hallmarks for endometriosis. These interconnected pathways and risk factors underscore a complex, multi-faceted etiology of endometriosis, suggesting multiple targets for precise and effective therapeutic interventions.

Introduction

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

Genome-wide association studies (GWASs) have been an important tool for implicating candidate genes and pathways linked with endometriosis. GWASs for endometriosis have uncovered over 40 loci associated with endometriosis risk. However, like most GWASs historically, the previous endometriosis studies have primarily included European populations with some East-Asian datasets as well[5]. Consequently, they have explained about 7% of

64 phenotypic variance with common variants[6,7] and have fallen short in elucidating the full
65 spectrum of its pathophysiology.

66 The advent of large-scale biobanks has revolutionized genetic research, offering unique
67 opportunities to conduct well-powered studies across diverse populations. As more diverse
68 genomic datasets are developed, like the Penn Medicine Biobank (PMBB)[8], the All of Us
69 Research Program (AOU)[9], and the Million Veterans Program (MVP)[10], that diversity can
70 increase statistical power and enhance the discovery of trait loci[11]. A worldwide consortium of
71 genomic researchers, the global biobank meta-analysis initiative (GBMI), was established to
72 facilitate collaboration in GWAS studies on unprecedented scales. In this study, we have
73 collaborated with 12 biobanks across several countries[9,12–22]. These resources not only enhance
74 statistical power to detect novel associations but also provide rich phenotypic data, enabling
75 more refined analyses. Moreover, the inclusion of diverse ancestries in genetic studies is crucial
76 for improving the generalizability of findings and addressing health disparities. This is
77 particularly pertinent for endometriosis, where significant variations in prevalence and clinical
78 presentation have been observed across different ethnic groups[23].

79 Beyond GWAS, there are many methods which can be employed *in silico* to investigate
80 other elements of the central dogma, providing further validation of GWAS loci and discovery of
81 additional disease-associated genes. Recent advances in multi-omics technologies have further
82 expanded our ability to translate genetic associations into biological insights. Integration of
83 GWAS results with transcriptomic, proteomic, and single-cell data using methods such allows
84 for a more comprehensive understanding of disease mechanisms, potentially identifying novel
85 therapeutic targets and biomarkers. Then, these multi-omic layers of association can be

integrated using tools such as Mergeomics[24], which also identifies gene sets enriched for association with the disease of interest. For endometriosis, where the interplay between genetic predisposition and complex tissue-specific processes is central to its pathogenesis, such integrative approaches are especially valuable. Given these considerations, there is a critical need for a large-scale, ancestrally diverse GWAS of endometriosis that incorporates rigorous phenotyping and leverages multi-omics data to provide a more complete picture of the disease's genetic landscape and underlying biology.

Here, we present the results of large-scale GWAS meta-analyses that significantly advances our understanding of endometriosis pathophysiology. Our study encompasses a diverse cohort from 14 biobanks worldwide, with over 30% non-European participants, enhancing the generalizability of our findings. We employed a comprehensive phenotyping approach, including not only a wide spectrum of endometriosis presentations but also surgically confirmed and procedure-validated (surgery or imaging) phenotypes. This rigorous phenotyping strategy allows for a more nuanced exploration of the genetic architecture underlying various manifestations of endometriosis. Through integrative multi-omics analyses, incorporating transcriptomic, proteomic, and single-cell data from endometrial tissues, we have explored five major hallmarks of endometriosis pathogenesis. These findings provide unprecedented insights into the molecular mechanisms driving this complex disorder, paving the way for improved diagnostics and targeted therapeutic interventions.

Results

EHR-Based Phenotyping of Endometriosis in Genetically Diverse Datasets

We considered six phenotype definitions for our GWASs: wide endometriosis (W), wide endometriosis excluding cases and controls with adenomyosis (Wex), procedure-confirmed narrow endometriosis versus all controls (PCN.v1) and versus procedure-confirmed controls (PCN.v2), and surgically-confirmed narrow endometriosis versus all controls (SCN.v1) and versus surgically-confirmed controls (SCN.v2). The largest phenotype, W, comprised 928,413 women, including 44,125 cases. We leveraged publicly-available FinnGen[25] and MVP[10] summary statistics to increase the sample size. The Wex phenotype, for which several biobanks were omitted, had a total sample size of 434,444 (N_{cases} = 8,122). Sample sizes for the PCN and SCN phenotypes were smaller (Table 1), with summary statistics included from only eight of the 12 contributing GBMI 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, EAS, and EUR. There were not enough EAS studies to conduct meta-analyses for the surgically confirmed narrow (SCN) phenotypes.

Phenotype	Ancestry	N	Cases	Prevalence
W	AFR	74,976	2,772	3.70%
	AMR	38,837	1,262	3.25%
	CSA	5,222	208	3.98%
	EAS	165,464	7,814	4.72%
	EUR	643,914	32,069	4.98%
	META	928,413	44,125	4.75%
Wex	AFR	56,314	862	1.53%
	AMR	38,195	620	1.62%
	CSA	5,066	62	1.22%
	EAS	80,375	1,531	1.90%
	EUR	254,494	5,047	1.98%
	META	434,444	8,122	1.87%
PCN.v1	AFR	51,050	823	1.61%
	AMR	36,719	599	1.63%
	EAS	78,999	155	0.20%

	EUR	209,207	3,207	1.53%
	META	375,975	4,784	1.27%
PCN.v2	AFR	9,874	823	8.34%
	AMR	7,113	599	8.42%
	EAS	1,539	155	10.07%
	EUR	35,722	3,207	8.98%
	META	54,248	4,784	8.82%
SCN.v1	AFR	50,592	365	0.72%
	AMR	36,465	345	0.95%
	EUR	208,136	2,137	1.03%
	META	295,193	2,847	0.96%
SCN.v2	AFR	1,436	365	25.42%
	AMR	1,957	345	17.63%
	EUR	12,926	2,137	16.53%
	META	16,319	2,847	17.45%

Multi-Ancestry GWAS Meta-Analysis Results

All 4,983 genome-wide significant variants ($P < 5 \times 10^{-8}$) from all meta-analyses are listed in Supplementary Table 2. The primary meta-analysis performed (multi-ancestry, wide phenotype) yielded 96 regions. From there, physically overlapping clumps were merged, resulting in 42 loci significantly associated with endometriosis. Besides those 42, there were two EUR-specific loci (*LINC01317* and *LINC02456*) and one AFR-specific locus (*POLR2M*), totaling 45 loci annotated in Figure 1a (Supplementary Table 3). All EAS loci overlapped with loci from the multi-ancestry meta-analyses, and the AMR W GWAS did not yield any genome-wide significant loci. Of the 45 unique W loci, 24 of their lead SNPs were in LD ($R^2 > 0.25$) with the lead SNPs reported in *Rahmioglu et al, 2023*, and an additional 14 were found within 50 Mb of the 2023 GWAS lead SNPs. The seven remaining loci were unreported in the 2023 GWAS: *RUVBL1* (rs4058156, P = 3.51E-10), *3q28* (rs9870207, P = 4.17E-09), *5p13.1* (rs55920409, P = 2.66E-08), *VAPA* (rs79380316, P = 3.56E-08), *DTD1* (rs6112068, P = 8.72E-10), *ZHX3* (rs17265513, P = 5.58E-09), *MIRLET7BHG* (rs873492, P = 4.80E-08). The observed SNP heritability, computed with LDSC, was significantly greater than zero ($P < 0.05$) in the multi-ancestry meta-analysis ($h^2_{lia} = 0.070$) as well as the AFR ($h^2_{lia} = 0.129$), EAS ($h^2_{lia} = 0.108$), and EUR ($h^2_{lia} = 0.108$) single-ancestry meta-analyses (Figure1b).

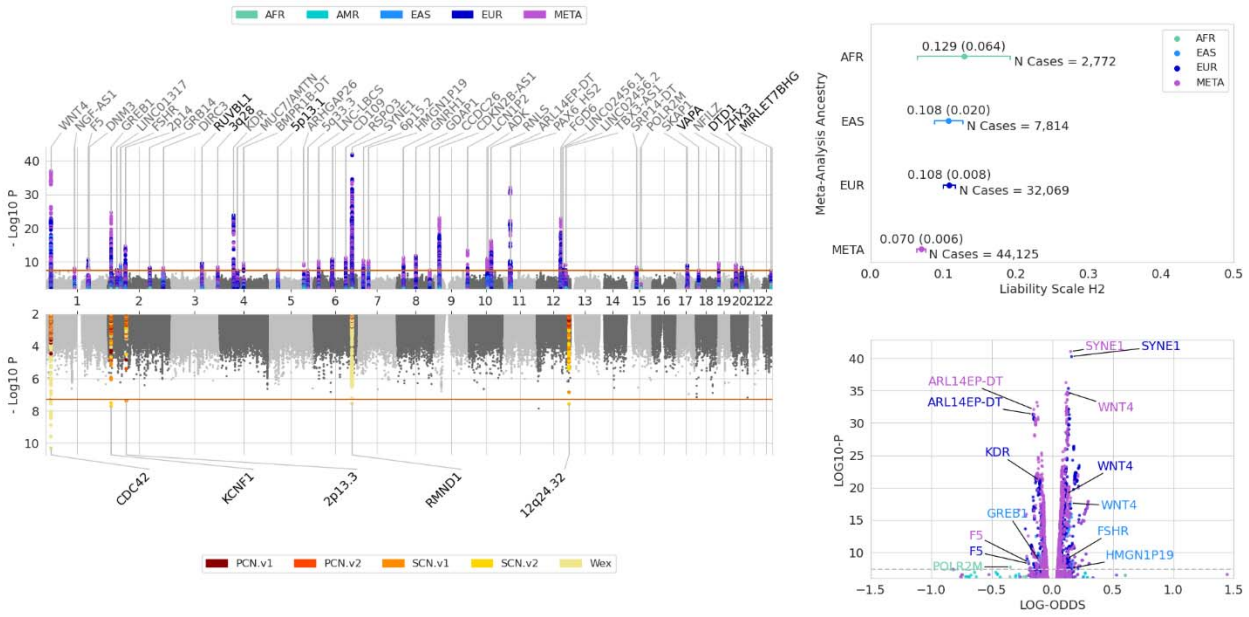

Figure 1: main meta-analysis results. Top left: overlayed Manhattan plots for the five W GWASs (multi-ancestry + four single-ancestries). Significant loci are highlighted, with black text representing the previously unreported hits. Bottom left: Overlayed Manhattan plots for the five multi-ancestry GWASs for the other phenotypes. All loci are previously reported. Top right: LDSC heritability estimates for the W GWASs; AMR is excluded because the heritability estimate was not significant. Bottom right: volcano plot highlighting variants with an absolute log-odds effect size > 0.1, colored by ancestry group.

Precise Phenotype Analyses Replicate Known Loci

For the other phenotype meta-analyses, we identified seven significant loci: two for EUR and multi-ancestry (SCN.v2 and SCN.v1), one for AFR and multi-ancestry (SCN.v1), three just in multi-ancestry (two Wex and one SCN.v1), and one just in AMR (PCN.v1). Both loci associated with Wex in the multi-ancestry analysis replicate known signals. *CDC42* (rs56319427, P = 4.91E-11) is in LD (R2 = 0.91) with the previous *CDC42* locus, and *RMND1* (rs9322319, P = 2.86E-08) is near the previous *SYNE1* locus. The five loci identified with the narrow phenotypes are within 50Mb of known signals, so they are not considered previously unreported. The surgically-confirmed cases vs all controls phenotype (SCN.v1) yielded two loci: 2p13.3 (rs116763937, P = 4.07E-08), also significant in AFR, and *CD163* (rs118078722, P = 1.39E-08), also significant in EUR. When comparing surgically-confirmed cases with only surgically-confirmed controls

(SCN.v2), we identified two loci: *KCNF1* (rs115117372, P = 1.97E-08) and 12q24.32 (rs12581759, P = 2.64E-08), with *KCNF1* also significant in the *EUR* meta-analysis. The final locus, *OTUD7A* (rs17762839, P = 1.72E-08), was associated with the PCN.v1 phenotype in the AMR meta-analysis. These loci are displayed in Figure 1c.

Fine-Mapping to Detect Causal Variants at GWAS Loci

To elucidate shared and population-specific putative causal variants, we conducted statistical fine-mapping on all hits from multi-population meta-analyses, totaling 48 loci in five (PCN.v2 had no significant hits) studies (See Methods). Of those 48 loci, 38 had at least one variant with posterior probability of being causal (PIP) greater than 0.5, totaling 132 causal variants (Supplementary Table 4). The cross-ancestry fine-mapping analyses computed PIPs for each variant across all combinations of ancestry groups included in the overall study. We identified credible sets (CS) for the 38 distinct loci associated with endometriosis as reported in Table 2.

Table 2: credible sets of multi-ancestry GWAS loci. 34 are for the W phenotype, 2 are for Wex, and 2 are for SCN.v2 (SCN cases versus surgically-confirmed controls). An "X" in the AFR-EUR columns denote whether that ancestry group was used for detecting any SNPs in the credible set (PIP > 0.5). Locus names match the GWAS.

Phenotype	Locus	Locus P	# CS	AFR	AMR	EAS	EUR	Credible Set
W	*WNT4*	1.12E-37	3			X	X	rs2473331, rs61768001, rs16826658
W	*MIR4418*	1.85E-09	3	X		X	X	rs1934478, rs10917216, rs12097230
W	*NGF-AS1*	1.11E-08	4			X	X	rs12030576, rs11102915, rs6656381, rs12075799
W	*DNM3*	2.38E-08	2		X	X	X	rs479960, rs569179
W	*GREB1*	2.45E-25	2		X	X	X	rs10929759, rs10165819
W	*FSHR*	8.59E-10	6	X	X	X	X	rs12614817, rs6716567, rs2134811, rs13032266, rs4420736, rs6722885
W	*2p14*	2.23E-15	2			X	X	rs2683680, rs2860517
W	*DIRC3*	6.11E-09	3			X	X	rs13018792, rs67436597, rs72613753
W	*EEFSEC*	1.00E-09	2		X		X	rs2955102, rs2811529
W	*3q28*	4.17E-09	7			X	X	rs9870207, rs6778588, rs9290954, rs9837216, rs9872656, rs9834978, rs9830307

W	*KDR*	6.46E-24	2	X	X	X	X	rs7680198, rs17081840
W	*MUC7/AMTN*	4.27E-08	4	X			X	rs13105826, rs4694318, rs28751942, rs28503247
W	*BMPR1B-DT*	3.88E-10	3			X	X	rs11938840, rs6532503, rs7680214
W	*5p13.1*	2.66E-08	2			X	X	rs13165498, rs2220801
W	*ARHGAP26*	5.06E-10	2			X	X	rs7728894, rs1799293
W	*5q33.3*	3.49E-08	6	X		X	X	rs1316379, rs13170063, rs2963462, rs2901003, rs11135035, rs17055653
W	*SYNE1*	9.67E-43	7		X	X	X	rs9371219, rs7740302, rs9371528, rs6927162, rs58415480, rs9371246, rs6557210
W	*6p15.2*	4.40E-11	5			X	X	rs1812673, rs13231733, rs11976790, rs1451383, rs1451385
W	*HMGN1P19*	5.37E-11	2			X	X	rs10951860, rs11975261
W	*GNRH1*	6.06E-12	2			X	X	rs7819740, rs13277090
W	*CDKN2B-AS1*	9.16E-24	9			X	X	rs7028213, rs10122243, rs10757288, rs1333054, rs1537377, rs10965274, rs2779747, rs10738612, rs1095899
W	*LCN1P2*	4.67E-14	1			X	X	rs633862
W	*RNLS*	5.69E-17	6			X	X	rs792208, rs792212, rs111958828, rs10788601, rs7922551, rs1935578
W	*ARL14EP-DT*	8.96E-33	5	X		X	X	rs606651, rs606642, rs476669, rs601681, rs76242056
W	*PAX6_HS2*	1.86E-09	2			X	X	rs6484567, rs2207550
W	*FGD6*	1.41E-23	2	X		X	X	rs10777675, rs6538622
W	*LINC02456*	7.72E-10	5	X		X	X	rs7306496, rs12319554, rs10437892, rs7296495, rs11111353
W	*TBX3-AS1*	9.23E-10	3			X	X	rs11067288, rs11067298, rs117549885
W	*SRP14-DT*	4.16E-09	5	X	X	X	X	rs4924409, rs2412474, rs78053299, rs12441483, rs7162269
W	*VAPA*	3.56E-08	2			X	X	rs604079, rs486270
W	*NFILZ*	2.25E-10	1			X	X	rs2967676
W	*DTD1*	8.72E-10	4			X	X	rs6075379, rs6081270, rs34478401, rs6045578
W	*ZHX3*	1.29E-08	3			X	X	rs6102464, rs6102467, rs4610114
W	*MIRLET7BHG*	4.80E-08	3	X	X		X	rs11702897, rs9626905, rs3859891
Wex	*CDC42*	4.91E-11	2			X	X	rs3754496, rs7529389
Wex	*RMND1*	2.86E-08	3	X	X	X	X	rs3778608, rs6913515, rs6904364
SCNv2	*KCNF1*	1.97E-08	5		X		X	rs13387419, rs77606953, rs10194777, rs2091817, rs1317430
SCNv2	*12q24.32*	2.64E-08	2		X			rs10773384, rs10847311

Of the 38 loci with at least one variant in the credible set, 34 were from the W multi-population meta-analysis, 2 were from Wex, and 2 were from SCN.v2. Fourteen of the credible

sets included variants detectable in part due to new genetic ancestry groups incorporated in the analyses (AMR and AFR). In particular, three loci leveraged the incorporation of all four ancestries to compute the credible set: *FSHR*, *KDR*, and *SRP14-DT*.

GWAS Variant Gene Enrichment Analyses

From the wide multi-population meta-analysis, we identified 47 unique gene regions that were enriched for significant GWAS variants. Eight genes were specific to the multi-ancestry W meta-analysis: *CDKN2A, CCHCR1, AP3M1, SPECC1, IGF1, DTD1, DNM3, and TCF19.* Eleven were specific to the EUR meta-analysis: *CBX1, GRB14, HOXA5, RPS10, SNX11, CHD6, VIP, LOC105375205, PARPBP, NOL4L, NR2C1,* and one (*CEBPE*) was specific to the AMR meta-analysis. Two additional genes were identified for the other phenotypes: *TCEA2* with Wex in the multi-ancestry meta-analysis, and *HAVCR2* in the PCN.v1 EAS meta-analysis. Full MAGMA results are given in Supplementary Table 5.

Post-GWAS Multi-Omic Analyses Fuel Pathway Identification

To identify associations between endometriosis and both the transcriptome and proteome, we performed TWAS using uterine tissue eQTL weights from GTEx and PWAS on whole blood pQTLs, inputting the multi-ancestry W GWAS summary statistics (See Methods). The TWAS tested 9,642 genes, and the PWAS tested 300 proteins. There were 11 genes achieving genome-wide significance (Bonferroni P value < 0.05) in the TWAS and 1 in the PWAS, shown in Table 3.

Table 3: 11 significant TWAS genes and one significant PWAS protein. The p-value thresholds were adjusted for the numbers of tests (TWAS P < 0.05 / 9,642, and PWAS P < 0.05 / 300).

Gene Symbol	Z-score	Beta	P-value
RMND1	8.69	4.92	3.7E-18
ARL14EP	-8.35	-0.16	6.8E-17

ZBTB2	-6.86	-1.77	6.7E-12
VEZT	6.62	0.44	3.5E-11
KDR	-6.17	-0.54	6.7E-10
ARMT1	6.13	0.77	8.9E-10
RP11-521C20.2	-5.11	-0.27	3.3E-07
RP11-521C20.5	-5.11	-0.13	3.3E-07
CCDC88B	-4.79	-0.20	1.7E-06
DTD1	4.77	0.20	1.9E-06
NOL4L	4.57	0.77	4.8E-06
RSPO3	6.08	0.14	1.2E-09

Of 11 significant genes identified from TWAS, nine of them were proximal to our significant multi-ancestry W GWAS loci. Three of them are the nearest genes to GWAS loci: *ARL14EP, KDR,* and *DTD1*. Three TWAS genes (*RMND1, ZBTB2,* and *ARMT1*) are related to the chromosome six cluster of loci which also includes *SYNE1* and *ESR1*. In the GWAS results, *RMD1* was more prominent in the Wex phenotype. *VEZT* (TWAS P = 3.5E-11) was most significant in the EAS GWAS results and was overlapping with the multi-ancestry GWAS locus *FGD6*. Two LincRNA genes identified in TWAS, *RP11-521C20.2, RP11-521C20.5*, are close to the GWAS locus *SRP14-DT*. The remaining two TWAS genes, *NOL4L* and *CCDC88B,* are not proximal to any GWAS loci we identified (at least 20 Mb away). PWAS identified RSPO3 (P = 1.2E-09) as the only significantly-associated protein, which was also a significant locus from our GWAS meta-analysis.

Identifying Disease-Relevant Cell Types and Crucial Genes

We prioritized 18 cell type – menstrual stage pairs with AUROC > 0.6 in which to compare gene expression for the 35 top MAGMA genes in the multi-ancestry W meta-analysis.

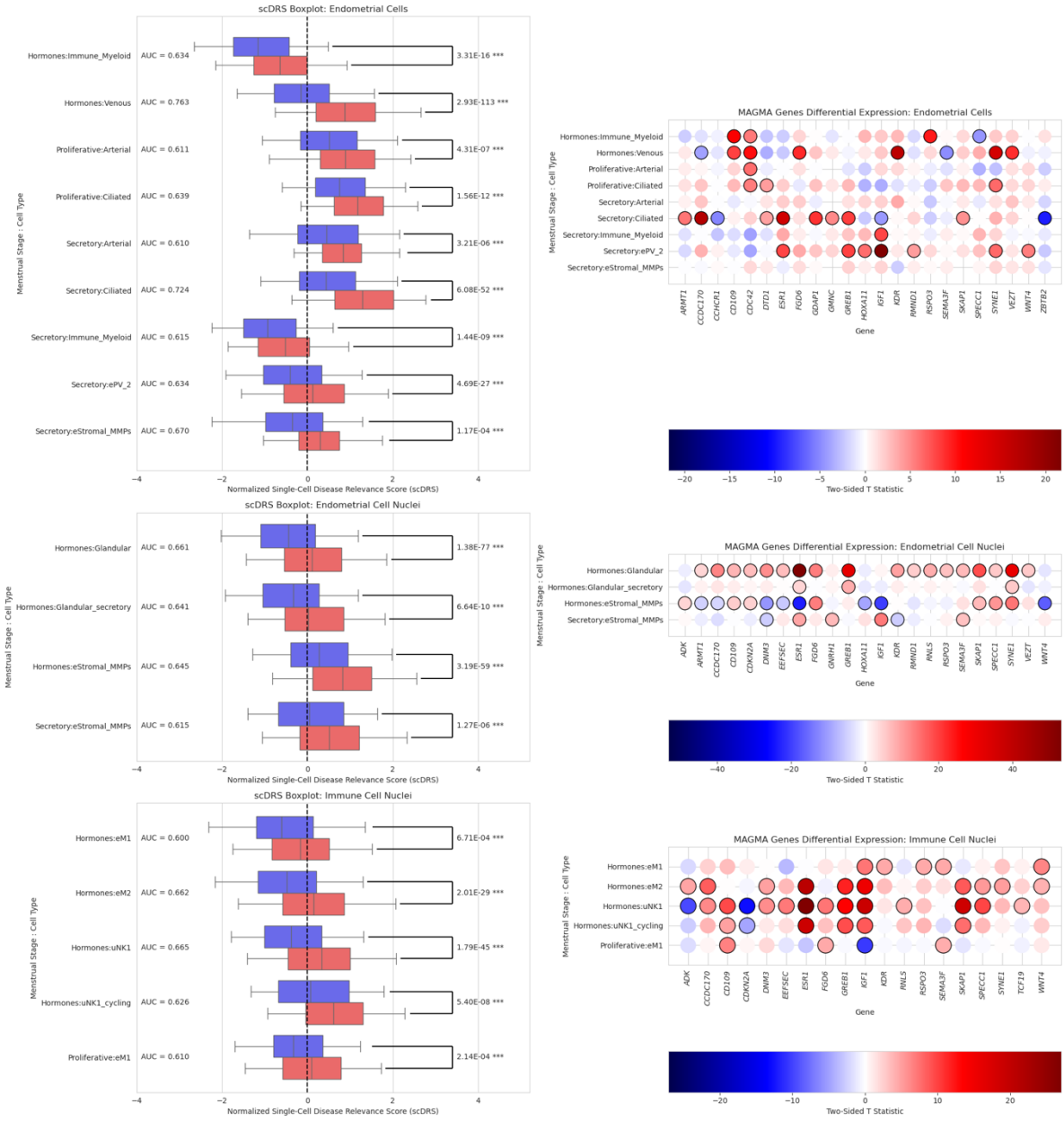

Figure 3: single cell analyses – disease relevance scores (scDRSs) and differential gene expression. The left panels show the distributions of scDRS comparing cells from donors with endometriosis (red) to donors without endometriosis (blue). Each set of boxes is annotated on the y-axis with the menstrual stage and cell type, an AUROC for the scDRS, and a P-value for a one-sided t-test of independent means. The right panels show the results of two-sided t-tests comparing the gene expression of the top MAGMA genes, with significant tests circled with a black border.

Of the 18 cell type – menstrual stage pairs with significantly different mean scDRS and AUROC > 0.6, nine came from donors in the “Hormones” menstrual phase (two endometrial cell types, three endometrial nucleus types, and four immune nucleus types). Of those nine, the

gene *CD109* was more highly expressed in six: immune myeloid cells, venous cells, glandular nuclei, stromal MMP (matrix metalloproteinase – expressing) nuclei, uterine natural killer nuclei, and cycling uterine natural killer nuclei.

The expression of *IGF1* was significantly different between cells from endometriosis cases and controls in ten cell type – menstrual stage pairs (three endometrial cell types, two endometrial nucleus types, and all five immune nucleus types). It did not always have the same direction; for seven out of those ten groups, expression of *IGF1* was greater, while for the remaining three (secretory phase ciliated cells, stromal MMP nuclei from donors on hormones, and proliferative phase uterine macrophage nuclei), it was downregulated.

WNT4 and *ESR1* are well-known endometriosis-associated genes. *WNT4's* expression was significantly different between endometriosis cases and controls in four cell type – menstrual stage pairs. *WNT4* expression was higher in both macrophage nucleus populations (eM1 and eM2) in donors taking exogenous hormones and secretory phase perivascular (ePV_2) nuclei. In contrast, *WNT4* expression was lower in stromal MMP nuclei from donors on hormones. *ESR1* expression was significantly different in nine cell type – menstrual stage pairs (three secretory, six hormones). All but one association were positive – the one negative T statistic was found in stromal MMP nuclei from donors on hormones. The three secretory groups with *ESR1* upregulated in donors with endometriosis were ciliated cells, perivascular (ePV_2) cells, and stromal MMP nuclei.

Mergeomics identifies significant biological processes from multi-omic integration

Using marker set enrichment analysis (MSEA) on KEGG database gene sets[26], we identified one significantly enriched pathway that provides insight into the molecular mechanisms underlying endometriosis. The WNT signaling pathway (hsa04310) was the most significantly enriched biological process, driven by markers including *RUVBL1, WNT4,* and *CTBP2* (Supplementary Figure XX). Other significant pathways (FDR $P < 0.05$) were Focal Adhesion (hsa04510), B cell receptor signaling pathway (hsa04662), Notch signaling pathway (hsa04330), and chronic myeloid leukemia (hsa05220). along with pathways in cancer and small cell lung cancer. The detailed module enrichment result is shown in Supplementary Table 6.

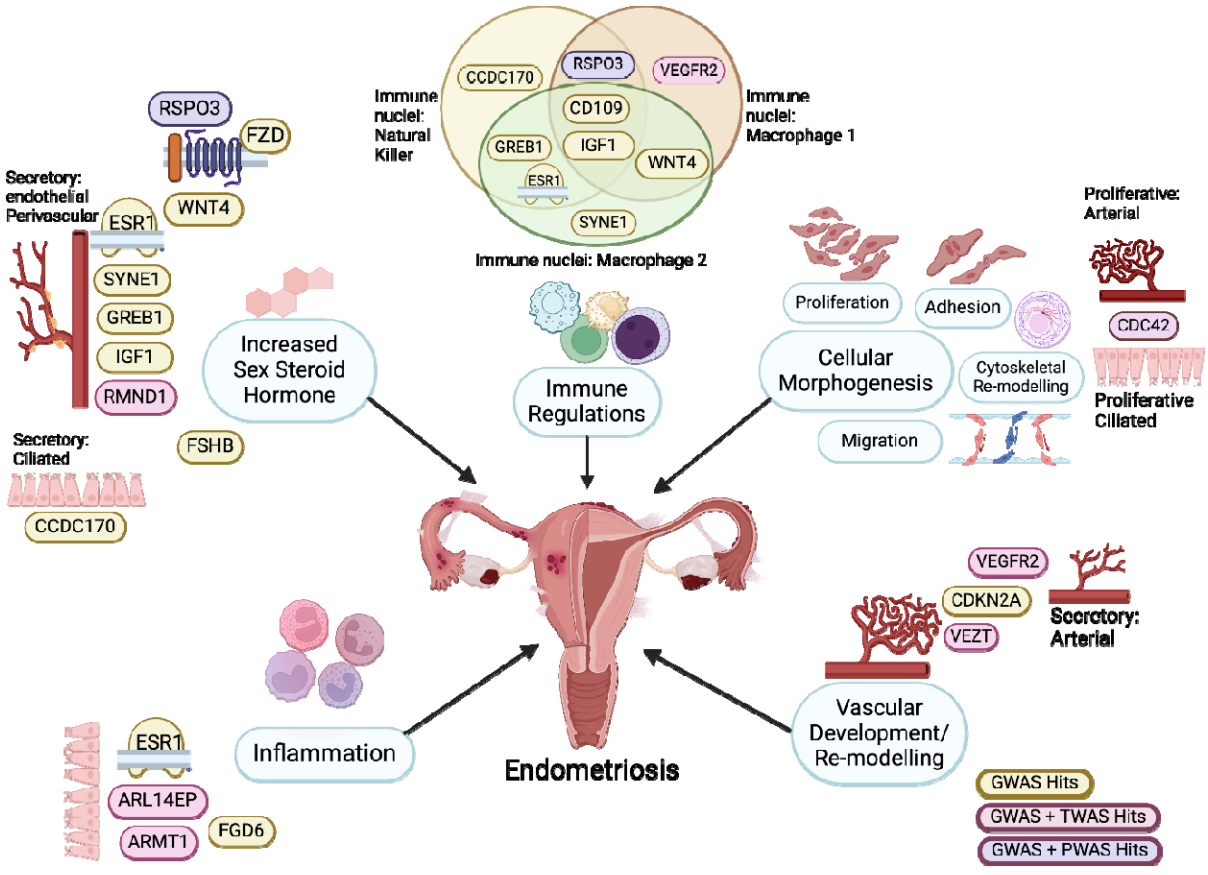

Figure 4: Summary of genes identified from multi-omics analysis and their potential role in each of the hallmarks of endometriosis.

Discussion

We report the largest multi-population GWAS for endometriosis to date incorporating nearly a million participants from 14 diverse biobanks across the world and increasing the total number of GWAS loci. Although multiple significant loci obtained from our study have been previously implicated in endometriosis pathophysiology, we have identified seven previously unreported GWAS loci and two previously unreported TWAS genes associated with endometriosis, which provides important insights into the underlying genetic architecture of endometriosis across population. We have summarized the major associations based on their involvement in specific biological functions (increased sex steroid hormones, immune regulation, inflammation, cell morphogenesis and vascular development and remodeling) in Figure 4, these could also be defined as five hallmarks of endometriosis.

Our ancestry stratified meta-analyses revealed a consistent observed SNP heritability of 10-12% for endometriosis for three studied genetic ancestry populations (EUR, EAS, and AFR), which is higher than previously observed SNP heritability of 7% based solely on European populations. The higher and more consistent heritability we observed can be attributed to the unprecedented scale and diversity of our study cohort. The previous GWASs have not included any African-ancestry or Admixed-American-ancestry (AMR) populations, which were essential in the computation of 14 / 38 (37%) credible sets in our fine-mapping analysis. We detected one African-only GWAS locus, *POLR2M*, and one AMR-only MAGMA gene, *CEBPE*. By

268 incorporating biobanks from various regions and ancestry backgrounds worldwide, we were
269 able to capture a more comprehensive genetic landscape of endometriosis.

270 Given that diagnosis codes from the EHR can be unreliable for accurately phenotyping
271 endometriosis, we employed narrow phenotyping algorithms which incorporated procedure
272 codes (surgical and imaging). Despite the diminished sample sizes, there were still genome-
273 wide significant loci: one for PCN.v1, two for SCN.v1, and two for SCN.v2. Both GWAS loci for
274 SCN.v2 (*KCNF1* and *12q24.32*) were successfully fine-mapped, with five and two putative causal
275 variants in their credible sets, respectively. The EAS-only PCN.v1 meta-analysis yielded a
276 significantly enriched gene, *HAVCR2*, an immune receptor gene near the *5q33.3* GWAS locus,
277 which was not present in the other meta-analyses. Leveraging multi-modal EHR data for more
278 precise phenotyping significantly enhanced our ability to refine and validate genetic
279 associations and uncover deeper biological insights.

280 Our meta-GWAS and PWAS analyses underscore the potential role of RSPO3 in
281 endometriosis pathophysiology. *RSPO3*, previously known to modulate risk of endometriosis,
282 interacts with *WNT4* via Frizzled (*FZD*) receptors, influencing the WNT/Ca2+ and WNT/β-catenin
283 pathways[27,28]. *WNT4* interacts with *ESR1* (Estrogen Receptor Alpha), key regulator of
284 endometrial cell proliferation and survival throughout the menstrual cycle[29]. Additionally,
285 scDRS analyses identified *ESR1, WNT4, SYNE1, GREB1, IGF1*, and *FSHB* as relevant in
286 endometrial perivascular cells (ePV) from donors in the secretory phase, while *RMND1*
287 (Required for Meiotic Nuclear Division 1 homolog), *ARMT1*, and *CCDC170* were linked to
288 secretory phase ciliated cells. Collectively, these genes likely promote increased sex steroid
289 hormone production and signaling. This hormonal milieu may, in turn, fuel inflammatory

responses and dysregulate immune function, potentially driving the progression of endometriosis[6,30,31].

Our meta-GWAS and TWAS analyses identified *VEZT* (Vezatin) *and ARL14EP* as significant. *VEZT* exhibits menstrual cycle-dependent expression in endometrial glands during the secretory phase and is also relevant in venous cells of individuals on exogenous hormones[32]. Together with *FGD6*, *VEZT* plays a crucial role in cytoskeletal remodeling. Dysregulation of this process can accelerate cell proliferation, migration, and invasion into the endometrium, promoting the development of ectopic lesions. This dysregulation may also lead to uncontrolled epithelial-mesenchymal transition (EMT), potentially driving endometriosis progression[32,33]. *ARL14EP* has previously been linked to depression which is a known endometriosis comorbidity[34]. It was also prominent in ciliated cells from the secretory phase in scDRS analysis. *ARL14EP* is involved in MHCII movement via the TGF-β pathway, suggesting a role in immune modulation, a hallmark of endometriosis[35,36]. Last, we also identified *VEGFR2 (KDR)*, significant in arterial cells from the secretory phase, and *CDKN2A,* both known to regulate angiogenesis and endothelial cell adhesion[37,38]. *CDC42*, another significant hit, plays a role in vascular remodeling along with *VEGFR2*[39–41]. Together, these findings highlight pathways in vascular and immune modulation, guiding future functional studies on endometriosis.

Despite the large sample size, heritability observed with GWAS (10-12%) still fails to measure up to the broad sense heritability estimation from a twin study of 47%[42]. Rare variants, structural variants, nonlinear effects, or gene-environment interactions might contribute to endometriosis risk but remain undetected in our analyses. As we leveraged the GWAS results to study other levels of the central dogma, one limitation was that we lacked

individual-level transcriptomic/proteomic data. The TWAS and PWAS performance relied on previously-trained *cis*-eQTL expression models, limiting the power, since the models do not capture all the gene expression variability and complex genetic architecture potentially relevant to our phenotype of interest[43]. Mergeomics assumes that overlapping signals between GWAS, TWAS, and PWAS are biologically relevant. The tool primarily considers only the association p-value, but not the directionality of association, which may limit the validity of detected biological signals. Additionally, Mergeomics may not fully capture complex interactions between genes or the functional effects of non-coding variants. While this tool does not allow us to infer causality, it does generate testable biological hypotheses for further functional assays.

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

Methods

Phenotyping

For the EHR-linked biobanks, we used structured data to define the phenotypes. Cases for wide endometriosis (W) were any women with a history of ICD-9, ICD-10, or SNOMED codes for endometriosis. For the wide phenotype excluding adenomyosis (Wex), women with a history of ICD-9, ICD-10, or SNOMED codes for uterine endometriosis were excluded from both cases and controls. Our two narrow case definitions were procedure-confirmed and surgically-confirmed (PCN and SCN). These narrow phenotypes were designed to capture confirmed cases and controls of endometriosis based on procedure history including hysterectomies, laparoscopies and ultrasounds. The list of procedure codes (CPT-4 or OPCS) for PCN included all the surgery codes in SCN but added non-obstetric ultrasound codes to account for potential imaging diagnoses. Each of the two narrow phenotypes was tested in two versions: cases versus all controls (PCN.v1 and SCN.v1) and cases versus confirmed controls (PCN.v2 and SCN.v2) where confirmed controls were those who had history of the corresponding procedure codes for each phenotype with no history of endometriosis diagnosis. For additional details on the phenotyping algorithms, see Supplemental Appendix 1.

Genotyping by Biobank

Most contributing biobanks acquired 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 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 Supplemental Appendix 2.

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[44] (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 Supplemental Appendix 3.

Contributing biobanks used linear mixed models to estimate the effect of each variant on the six phenotypes. Tools implemented for these tests include SAIGE[45] and Regenie[46]. 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 testing (SAIGE or Regenie step 2), variants were required to have a minor allele count of at least 20, and an imputation quality of at least 0.6. Individuals were included if their overall genotyping rate was at least 95%. A summary of each biobank's exact covariates and QC procedures is in Supplemental Appendix 4.

After biobank analyses were completed and collected, summary statistics were cleaned and lifted into genome build hg38 as necessary. Then, inverse variance-weighted, fixed-effect meta-analyses were performed using GWAMA[47]. GWAMA adjusts each study based on its overall genomic inflation factor. Prior to meta-analysis, input summary statistics were restricted to variants with a minor allele frequency of at least 0.005 to ensure stability in the estimates. We started by applying the clumping function of Plink 1.9[48] with the lead SNP p-value threshold set to $5x10^{-8}$, the secondary SNP threshold set to $1x10^{-3}$, minimum R^2 of 0.1, and a window of 1000kb. 1KG was used as a linkage disequilibrium (LD) reference since the meta-analyses were multi-ancestry. After the variant clumps were identified, any clumps that were physically overlapping with one another were merged into loci. Then, we tested if any loci were in LD with ($R^2 > 0.25$) or proximal to tag SNPs of the 39 autosomal lead SNPs from the 2023 GWAS[49] to determine whether our hits were known or previously unreported[10].

Estimation of Observed SNP Heritability

We estimated heritability with the LD-Score Regression tool, LDSC[50], for four ancestry-stratified meta-analyses and the multi-ancestry meta-analysis. LD scores were computed from the full 1KG for the multi-ancestry summary statistics, and from the respective super-population subgroups from 1KG for each ancestry-stratified meta-analysis.

Variant-Set Enrichment for Identifying Enriched Gene Regions

MAGMA[51] was used to assess gene regions within the GWAS results for enrichment. MAGMA uses a model based on multiple regression to test the association of a phenotype with groups of variants. Testing groups of variants increases statistical power over the single-variant

GWAS tests performed[51]. We utilized gene region definitions from NCBI for human genome build 38, testing about 18,000 genes for each GWAS.

Statistical Fine-Mapping for Identification of Putative Causal Variants

MESuSiE[52] was used to identify causal variants in each significant GWAS locus. We extracted the summary statistics from each locus's region in addition to computing LD matrices for that region in all ancestry groups comprising the overall study. For single-ancestry meta-analyses, only one LD matrix was needed, but the loci from the multi-ancestry meta-analyses were fine-mapped using LD matrices from AFR, AMR, EAS, and EUR. MESuSiE analyzes whether causal variants are shared between ancestry groups by computing a posterior probability of causality (PIP) for each ancestry alone and for each combination of two or more ancestries. SNPs with any PIP value greater than 0.5 are part of the credible set for that locus.

Tissue-Specific Transcriptome-Wide Association Study

We used GWAS summary statistics from our multi-population W meta-analysis to estimate the association between the variants on expression of genes and the complex trait, endometriosis. We used precomputed eQTL weights from the PredictDB data repository (http://predictdb.org/) which represented *cis*-level associations. The database files comprise of a list of genetic variants (cis-SNPs) used to predict gene expression, located within 1 Mb of a gene, eQTL weights, and covariance information. Those models along with the GWAS summary statistics were input into the summary-based PrediXcan (S-PrediXcan) pipeline for the transcriptome-wide association study (TWAS), scanning for genes whose expression levels were associated with endometriosis[53]. We used Bonferroni $P < 0.05$ (corrected for 9,642 tests) to identify statistically significant TWAS genes.

Proteome-Wide Association Study

We performed a proteome-wide association study (PWAS) to investigate the association between genetically predicted protein levels and endometriosis. We used summary-level GWAS data from the multi-ancestry W meta-analysis, and pQTL models developed by *Schubert et al*[54]. The models were derived from SOMAscan assay of plasma proteins measured in multi-ethnic Trans-omics for Precision Medicine (TOPMed) Multi-omics pilot study, capturing 1mb cis-regulatory window. We used baseline models only, essentially those models were trained under standard conditions without any additional adjustments for other covariates or interaction terms, typically a reference model for general prediction of gene expression based solely on genetic variation. Only models with a predictive correlation (rho) of at least 0.1 and a z-score p-value threshold of 0.05, meaning that the association between the SNPs and gene expression is statistically significant, were included. The PWAS was thereby performed on 309 most robust, and reliable models for the analysis. We employed the S-PrediXcan pipeline to perform the PWAS. We used Bonferroni-correct P value threshold ($P < 0.05 / 309$) for identifying significant PWAS genes.

Integration with Single-Cell Data

Single cell data are rich resources for understanding the biology of diseases. A recently-published endometrial single cell atlas includes gene expression data for cells, nuclei, immune cells, and immune nuclei, covering over 40 different cell types, three menstrual phases, and four different endometrial pathologies[55]. We leveraged these data to compute single-cell disease-relevance scores (scDRSs) based on the associated genes from the GWAS, as computed by MAGMA. Disease risk scores were computed and normalized using the scDRS software[56]. We

441 removed any cell types labeled as "hormones," as the authors did in the atlas release paper
442 when they performed fGWAS.

443 For each cell type and menstrual phase, we compared the scDRSs of cells from donors
444 with and without endometriosis. We performed one-sided t-tests to compare the distributions,
445 hypothesizing that the scDRSs of cells from endometriosis cases should be higher than that of
446 cells from endometriosis controls. We also computed the area under the receiver-operating
447 curve (AUROC) to see how well the scDRS distinguished between endometriosis case and
448 control donors. From there, we prioritized cell type – menstrual phase combinations with
449 significant t-tests and AUROC > 0.6 for further examination. We compared the expression of the
450 top MAGMA genes between case and control donors using t-tests to identify which genes were
451 driving the performance of the scDRS.

452 ## Merge-Omics

453 Using the GWAS, TWAS, and PWAS results, we conducted a marker set enrichment analysis
454 (MSEA) with Mergeomics[24]. MSEA identifies key marker sets (genes or proteins) significantly
455 enriched for associations across multi-omics data. The markers were mapped to Kyoto
456 Encyclopedia of Genes and Genomes (KEGG) pathways using gene annotations. We used GWAS,
457 TWAS and PWAS markers (rsIDs for GWAS and gene names for TWAS and PWAS), along with
458 their corresponding –log10 transformed p-values as input for the tool. We enriched for KEGG
459 pathways with FDR < 0.05.

Acknowledgments

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

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

This phase of the eMERGE Network was initiated and funded by the NHGRI through the following grants: U01HG008657 (Group Health Cooperative/University of Washington); U01HG008685 (Brigham and Women's Hospital); U01HG008672 (Vanderbilt University Medical Center); U01HG008666 (Cincinnati Children's Hospital Medical Center); U01HG006379 (Mayo Clinic); U01HG008679 (Geisinger Clinic); U01HG008680 (Columbia University Health Sciences); U01HG008684 (Children's Hospital of Philadelphia); U01HG008673 (Northwestern University); U01HG008701 (Vanderbilt University Medical Center serving as the Coordinating Center); U01HG008676 (Partners Healthcare/Broad Institute); and U01HG008664 (Baylor College of Medicine).

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

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

The Trøndelag Health Study (The HUNT Study) is a collaboration between HUNT Research Centre (Faculty of Medicine and Health Sciences, NTNU, Norwegian University of Science and Technology), Trøndelag County Council, Central Norway Regional Health Authority, and the Norwegian Institute of Public Health. The genotyping in HUNT was financed by the National Institutes of Health; University of Michigan; the Research Council of Norway; the Liaison Committee for Education, Research and Innovation in Central Norway; and the Joint Research Committee between St Olavs hospital and the Faculty of Medicine and Health Sciences, NTNU. The genetic investigations of the HUNT Study are a collaboration between researchers from the HUNT Center for Molecular and Clinical Epidemiology (formerly known as the K.G. Jebsen

Center for Genetic Epidemiology as of August 1st 2023), NTNU, and the University of Michigan Medical School and the University of Michigan School of Public Health. We thank HUNT participants for donating their time, samples, and information to help others; clinicians and other employees at Nord-Trøndelag Hospital Trust for their support and for contributing to data collection.

References

1. Christ, J. P. *et al.* Incidence, prevalence, and trends in endometriosis diagnosis: a United States population-based study from 2006 to 2015. *American Journal of Obstetrics and Gynecology* **225**, 500.e1-500.e9 (2021).
2. Agarwal, S. K. *et al.* Clinical diagnosis of endometriosis: a call to action. *American Journal of Obstetrics and Gynecology* **220**, 354.e1-354.e12 (2019).
3. Fauconnier, A. & Chapron, C. Endometriosis and pelvic pain: epidemiological evidence of the relationship and implications. *Human Reproduction Update* **11**, 595–606 (2005).
4. Coccia, M. E., Nardone, L. & Rizzello, F. Endometriosis and Infertility: A Long-Life Approach to Preserve Reproductive Integrity. *International Journal of Environmental Research and Public Health* **19**, 6162 (2022).
5. Fatumo, S. *et al.* A roadmap to increase diversity in genomic studies. *Nat Med* **28**, 243–250 (2022).
6. Sapkota, Y. *et al.* Meta-analysis identifies five novel loci associated with endometriosis highlighting key genes involved in hormone metabolism. *Nat Commun* **8**, 15539 (2017).

7. Rahmioglu, N. *et al.* The genetic basis of endometriosis and comorbidity with other pain and inflammatory conditions. *Nat Genet* **55**, 423–436 (2023).
8. Verma, A. *et al.* The Penn Medicine BioBank: Towards a Genomics-Enabled Learning Healthcare System to Accelerate Precision Medicine in a Diverse Population. *Journal of Personalized Medicine* **12**, 1974 (2022).
9. Genomic data in the All of Us Research Program. *Nature* **627**, 340–346 (2024).
10. Verma, A. *et al.* Diversity and Scale: Genetic Architecture of 2,068 Traits in the VA Million Veteran Program. 2023.06.28.23291975 Preprint at https://doi.org/10.1101/2023.06.28.23291975 (2023).
11. Zhou, W. *et al.* Global Biobank Meta-analysis Initiative: Powering genetic discovery across human disease. *Cell Genomics* **2**, 100192 (2022).
12. Nagai, A. *et al.* Overview of the BioBank Japan Project: Study design and profile. *Journal of Epidemiology* **27**, S2–S8 (2017).
13. Pulley, J., Clayton, E., Bernard, G. R., Roden, D. M. & Masys, D. R. Principles of human subjects protections applied in an opt-out, de-identified biobank. *Clin Transl Sci* **3**, 42–48 (2010).
14. Roden, D. *et al.* Development of a Large-Scale De-Identified DNA Biobank to Enable Personalized Medicine. *Clinical Pharmacology & Therapeutics* **84**, 362–369 (2008).
15. Wiley, L. K. *et al.* Building a vertically integrated genomic learning health system: The biobank at the Colorado Center for Personalized Medicine. *The American Journal of Human Genetics* **111**, 11–23 (2024).

16. Gottesman, O. *et al.* The Electronic Medical Records and Genomics (eMERGE) Network: past, present, and future. *Genet Med* **15**, 761–771 (2013).

17. David, S. P. *et al.* Personalized medicine in a community health system: the NorthShore experience. *Front. Genet.* **14**, (2023).

18. Brumpton, B. M. *et al.* The HUNT study: A population-based cohort for genetic research. *Cell Genomics* **2**, (2022).

19. Rovite, V. *et al.* Genome Database of the Latvian Population (LGDB): Design, Goals, and Primary Results. *Journal of Epidemiology* **28**, 353–360 (2018).

20. Boutin, N. T. *et al.* The Evolution of a Large Biobank at Mass General Brigham. *Journal of Personalized Medicine* **12**, 1323 (2022).

21. Fan, C.-T., Lin, J.-C. & Lee, C.-H. Taiwan Biobank: A Project Aiming to Aid Taiwan's Transition into a Biomedical Island. *Pharmacogenomics* **9**, 235–246 (2008).

22. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).

23. Katon, J. G., Plowden, T. C. & Marsh, E. E. Racial disparities in uterine fibroids and endometriosis: a systematic review and application of social, structural, and political context. *Fertility and Sterility* **119**, 355–363 (2023).

24. Ding, J. *et al.* Mergeomics 2.0: a web server for multi-omics data integration to elucidate disease networks and predict therapeutics. *Nucleic Acids Research* **49**, W375–W387 (2021).

25. Kurki, M. I. *et al.* FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature* **613**, 508–518 (2023).

564 26. Kanehisa, M. & Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids*
565 *Research* **28**, 27–30 (2000).
566 27. Tao, T., Mo, X. & Zhao, L. Identifying novel potential drug targets for endometriosis via
567 plasma proteome screening. *Front. Endocrinol.* **15**, (2024).
568 28. Scholz, B. *et al.* Endothelial RSPO3 Controls Vascular Stability and Pruning through Non-
569 canonical WNT/Ca2+/NFAT Signaling. *Developmental Cell* **36**, 79–93 (2016).
570 29. Gallagher, C. S. *et al.* Genome-wide association and epidemiological analyses reveal
571 common genetic origins between uterine leiomyomata and endometriosis. *Nat Commun*
572 **10**, 4857 (2019).
573 30. Marla, S. *et al.* Genetic risk factors for endometriosis near estrogen receptor 1 and
574 coexpression of genes in this region in endometrium. *Molecular Human Reproduction* **27**,
575 gaaa082 (2021).
576 31. García-Gómez, E. *et al.* Regulation of Inflammation Pathways and Inflammasome by Sex
577 Steroid Hormones in Endometriosis. *Front. Endocrinol.* **10**, (2020).
578 32. Holdsworth-Carson, S. J. *et al.* Endometrial vezatin and its association with endometriosis
579 risk. *Human Reproduction* **31**, 999–1013 (2016).
580 33. Szymański, M. *et al.* Role of Cyclins and Cytoskeletal Proteins in Endometriosis: Insights into
581 Pathophysiology. *Cancers* **16**, 836 (2024).
582 34. Adewuyi, E. O. *et al.* Genetic analysis of endometriosis and depression identifies shared loci
583 and implicates causal links with gastric mucosa abnormality. *Hum Genet* **140**, 529–552
584 (2021).

35. Paul, P. *et al.* A Genome-wide Multidimensional RNAi Screen Reveals Pathways Controlling MHC Class II Antigen Presentation. *Cell* **145**, 268–283 (2011).

36. Hogg, C. *et al.* Macrophages inhibit and enhance endometriosis depending on their origin. *Proceedings of the National Academy of Sciences* **118**, e2013776118 (2021).

37. Jin, Y. *et al.* Deletion of Cdc42 Enhances ADAM17-Mediated Vascular Endothelial Growth Factor Receptor 2 Shedding and Impairs Vascular Endothelial Cell Survival and Vasculogenesis. *Molecular and Cellular Biology* **33**, 4181–4197 (2013).

38. Barry, D. M. *et al.* Cdc42 is required for cytoskeletal support of endothelial cell adhesion during blood vessel formation in mice. *Development* **142**, 3058–3070 (2015).

39. Sugino, N., Kashida, S., Karube-Harada, A., Takiguchi, S. & Kato, H. Expression of vascular endothelial growth factor (VEGF) and its receptors in human endometrium throughout the menstrual cycle and in early pregnancy. (2002) doi:10.1530/rep.0.1230379.

40. Cerione, R. A. Cdc42: new roads to travel. *Trends in Cell Biology* **14**, 127–132 (2004).

41. Slack, M. A. & Gordon, S. M. Protease Activity in Vascular Disease. *Arterioscler Thromb Vasc Biol* **39**, e210–e218 (2019).

42. Saha, R. *et al.* Heritability of endometriosis. *Fertility and Sterility* **104**, 947–952 (2015).

43. Mai, J., Lu, M., Gao, Q., Zeng, J. & Xiao, J. Transcriptome-wide association studies: recent advances in methods, applications and available databases. *Commun Biol* **6**, 1–10 (2023).

44. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).

45. Zhou, W. *et al.* Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat Genet* **50**, 1335–1341 (2018).

607 46. Mbatchou, J. *et al.* Computationally efficient whole-genome regression for quantitative and
608 binary traits. *Nat Genet* **53**, 1097–1103 (2021).
609 47. Mägi, R. & Morris, A. P. GWAMA: software for genome-wide association meta-analysis.
610 *BMC Bioinformatics* **11**, 288 (2010).
611 48. Purcell, S. *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based
612 Linkage Analyses. *Am J Hum Genet* **81**, 559–575 (2007).
613 49. Rahmioglu, N. *et al.* The genetic basis of endometriosis and comorbidity with other pain and
614 inflammatory conditions. *Nat Genet* **55**, 423–436 (2023).
615 50. LD Score regression distinguishes confounding from polygenicity in genome-wide
616 association studies | Nature Genetics. https://www.nature.com/articles/ng.3211.
617 51. Leeuw, C. A. de, Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: Generalized Gene-Set
618 Analysis of GWAS Data. *PLOS Computational Biology* **11**, e1004219 (2015).
619 52. Gao, B. & Zhou, X. MESuSiE enables scalable and powerful multi-ancestry fine-mapping of
620 causal variants in genome-wide association studies. *Nat Genet* **56**, 170–179 (2024).
621 53. Gamazon, E. R. *et al.* A gene-based association method for mapping traits using reference
622 transcriptome data. *Nat Genet* **47**, 1091–1098 (2015).
623 54. Schubert, R. *et al.* Protein prediction for trait mapping in diverse populations. *PLOS ONE* **17**,
624 e0264341 (2022).
625 55. Marečková, M. *et al.* An integrated single-cell reference atlas of the human endometrium.
626 *Nat Genet* **56**, 1925–1937 (2024).
627 56. Zhang, M. J. *et al.* Polygenic enrichment distinguishes disease associations of individual cells
628 in single-cell RNA-seq data. *Nat Genet* **54**, 1572–1580 (2022).

629