Author/s:
Yang, F;Qi, T;McRae, AF;Rogers, PAW;Montgomery, GW;Mortlock, S

Title:
Regulation of RNA splicing in endometrial tissue and its association with endometriosis

Date:
2025-09-19

Citation:
Yang, F., Qi, T., McRae, A. F., Rogers, P. A. W., Montgomery, G. W. & Mortlock, S. (2025). Regulation of RNA splicing in endometrial tissue and its association with endometriosis. Iscience, 28 (9), pp.113207-. https://doi.org/10.1016/j.isci.2025.113207.

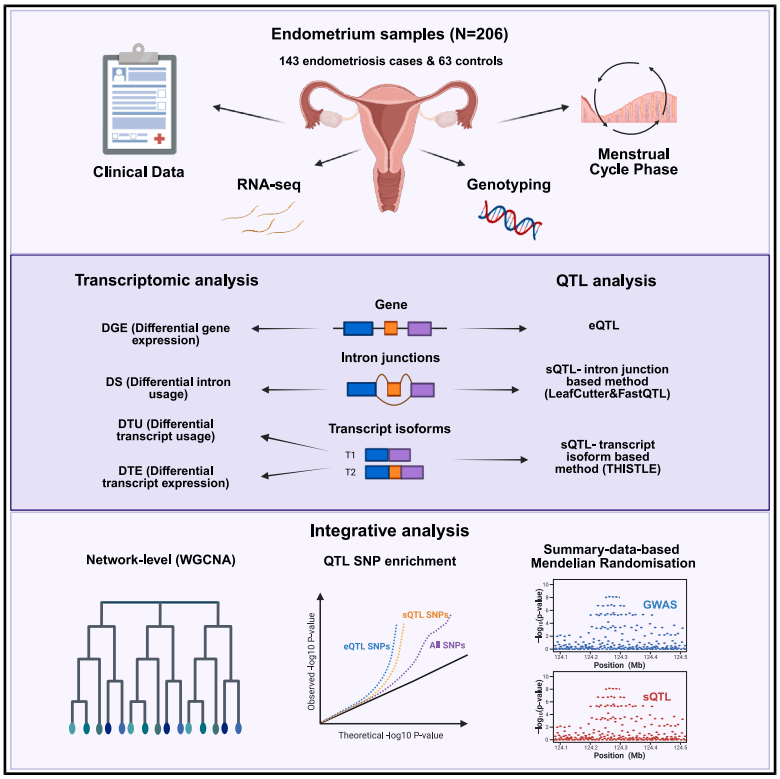
Persistent Link:
https://hdl.handle.net/11343/367552

iScience

Article

Regulation of RNA splicing in endometrial tissue and its association with endometriosis

Graphical abstract



Highlights

- Transcript-level and splicing changes in endometrium vary by menstrual cycle phase
- Mid-secretory phase shows endometriosis-specific splicing differences
- Genetic regulation of splicing identified in over 2,000 endometrial genes
- GREB1 and WASHC3 splicing linked to endometriosis genetic risk

Authors

Fei Yang, Ting Qi, Allan F. McRae, Peter A. W. Rogers, Grant W. Montgomery, Sally Mortlock

Correspondence

s.mortlock@uq.edu.au

In brief

Health sciences; Biological sciences; Molecular biology; Developmental biology; Biology of human development; Complex system biology

Yang et al., 2025, iScience *28*, 113207
September 19, 2025
https://doi.org/10.1016/j.isci.2025.113207

CellPress

iScience

CellPress
OPEN ACCESS

Article

Regulation of RNA splicing in endometrial tissue and its association with endometriosis

Fei Yang,[1] Ting Qi,[1,2,3] Allan F. McRae,[1] Peter A.W. Rogers,[4] Grant W. Montgomery,[1] and Sally Mortlock[1,5,6,*]
[1]The Institute for Molecular Bioscience, The University of Queensland, Brisbane, QLD 4072, Australia
[2]School of Life Sciences, Westlake University, Hangzhou, Zhejiang 310024, China
[3]Westlake Laboratory of Life Sciences and Biomedicine, Hangzhou, Zhejiang 310024, China
[4]University of Melbourne Department of Obstetrics and Gynaecology, and Gynaecology Research Centre, Royal Women's Hospital, Parkville, VIC 3052, Australia
[5]Australian Women and Girls' Health Research Centre, School of Public Health, University of Queensland, Herston, QLD 4006, Australia
[6]Lead contact
*Correspondence: s.mortlock@uq.edu.au
https://doi.org/10.1016/j.isci.2025.113207

SUMMARY

The endometrium, essential for reproduction, undergoes cyclical shedding, remodeling, and regeneration. Using a large endometrial transcriptomic dataset (n = 206), we identified RNA splicing and transcript isoform-level changes across the menstrual cycle and in endometriosis, findings not seen in gene-level analyses. Transcriptomic differences were most pronounced in the mid-secretory (receptive) phase in endometriosis samples. By integrating genotype data, we found evidence of *cis*-genetic effects on splicing in endometrium, identifying 3,296 splicing quantitative trait loci (sQTLs) with the majority of genes with sQTLs (67.5%) not discovered in the gene level eQTLs analysis, indicating the splicing-specific effects. Integrating the sQTLs with the endometriosis genome-wide association study (GWAS) data, we identified two genes (*GREB1* and *WASHC3*) that were significantly associated with endometriosis risk through genetically regulated splicing events. Overall, this study detected transcript isoform-level and RNA splicing level specific changes, providing insights into the dynamic changes in transcriptomic regulation in endometrium and its association with endometriosis.

INTRODUCTION

Endometrium is a specialized tissue that undergoes dynamic changes in cellular composition and function across the menstrual cycle and plays a pivotal role in preparation for implantation and maintenance of pregnancy.[1] Endometrium is thought to be the source of cells initiating many cases of endometriosis, a common gynecological disorder affecting 11% of reproductive aged women.[2,3] Endometriosis is associated with symptoms, such as pelvic pain and infertility, severely impacting the quality of life of those affected and resulting in a heavy economic burden to society.[2,4,5]

Transcriptomic regulation in endometrium and its association with endometriosis are not fully understood, with previous endometrial studies focusing on gene-level expression analysis in both bulk and single-cell resolution RNA sequencing (RNA-seq) methods.[3,6,7] Alternative splicing that can generate multiple mRNAs from a single gene through variation in transcript usage and alternative splicing plays an important role in creating a transcriptome with diverse functions and expanding the proteome complexity.[8] Extensive studies have demonstrated that the coordination of alternative splicing networks contribute to tissue development and cell differentiation in a cell-type- and/or tissue-specific manner.[9–12] Moreover, the associations between human diseases and altered splicing have been widely reported in cancers, and muscular dystrophies, neurological diseases, and myelodysplastic syndromes.[13–15] Compared with gene-level expression analysis, individual transcript-level changes in a single gene and post-transcriptional modifications like alternative RNA splicing, which are more directly related to the downstream functional protein isoforms, have received little attention in reproductive medicine. Recent studies on more than 200 samples[3,16] have found no evidence for differentially expressed genes in endometrium between women with and without endometriosis using both microarray and high-throughput RNA-seq technologies. Therefore, transcript level analyses could add to our understanding of the RNA splicing landscape in human endometrium and may also provide insights into endometriosis pathology.

Genome-wide association studies (GWASs) have identified genetic risk factors associated with endometriosis risk.[17] However, the molecular mechanisms and target genes that link genetic risk factors to this increased risk remain poorly understood.[17] Similar to other complex disorders, the majority of disease associated genetic variants fall within noncoding regions of the

Table 1. Number of patients by menstrual cycle phase and endometriosis case-control status

Menstrual cycle phase	Control	Case	Combined
Menstrual (M)	5	9	14
Early-proliferative (EP)	2	3	5
Mid-proliferative (MP)	16	56	72
Late-proliferative (LP)	10	12	22
Early-secretory (ES)	10	21	31
Mid-secretory (MS)	14	27	41
Late-secretory (LS)	6	15	21
Total number of samples	63	143	206

genome and causal variants may regulate gene expression and RNA splicing of corresponding coding gene targets.[18,19] Despite discovery of a large number of expression quantitative trait loci (eQTLs) where genetic variation affects gene-level expression, few target genes has been linked to genetic variants identified using GWAS,[20] suggesting a role for other fundamental mechanisms where genetic variants regulate RNA splicing (sQTL). Discovery of sQTLs in different tissues provides a valuable resource to identify specific candidate risk genes for diseases using transcriptome-wide association studies.[20–22] The role of sQTLs in endometrium and related diseases such as endometriosis has not been studied.[20]

Here, using genotypes and RNA-seq from endometrial samples from 206 women of European ancestry, we characterized and compared transcript isoform-level and RNA splicing variations with gene-level expression in endometrium across the menstrual cycle and its association with endometriosis. We investigated genetic effects on RNA splicing in endometrium, and prioritized genes with splicing events regulated by variants associated with endometriosis via transcriptome-wide association study analysis. Taken together, this study expands our understanding of transcriptional and genetic regulation in human endometrium and provides insights for pathogenic mechanisms contributing to endometriosis and related infertility.

RESULTS

Transcriptomic changes in endometrium across the menstrual cycle

Dissecting the mechanisms of dynamic transcriptomic changes in endometrium throughout the menstrual cycle in response to fluctuating hormones can provide insights into molecular mechanisms influencing endometrium in health and disease.

Multi-level transcriptomic profiling reveals widespread changes across the cycle

To capture the full complexity of transcriptomic regulation across the menstrual cycle, we examined differential gene-level expression (DGE), transcript-level expression (DTE), transcript usage (DTU), and splicing (DS), between menstrual cycle phases (mid-proliferative [MP], early secretory [ES], mid-secretory [MS], and late secretory [LS]), as each analysis provides complementary insights into distinct layers of gene regulation, including expression magnitude, isoform variation, and alternative splicing events. Table 1 shows detailed information on patient numbers, menstrual cycle phases, and endometriosis status. Following correction for multiple testing, we observed significant differences at all four transcriptomic levels (DGE, DTE, DTU, and DS) across the menstrual cycle (Tables S1, S2, S3, and S4). Fold changes estimated by the DGE analysis were consistent with previously published microarray datasets[16] for the comparison between ES vs. MS (correlation coefficient [R] = 0.98, $p < 2.2e-16$) and MS vs. LS (R = 0.96, $p < 2.2e-16$) (Figure S1).

Phase-specific dynamics in transcript and splicing regulation

Similar to the gene-level results in the previous microarray study,[16] we observed that the biggest transcriptomic changes at transcript- and splicing-levels occurred between MP vs. ES, followed by ES vs. MS and MS vs. LS (Figure 1A). Only a small number of DGE, DTE, DTU, and DS genes changed consistently across all menstrual cycle phase comparisons, the majority of differences were observed between specific phases. Notably, at the DTU level, the cross-phase overlap was greatly attenuated, suggesting that alternative transcript usage may confer some phase specificity. By looking at the transcripts with evidence of DTE across all phases, we observed sets of transcripts whose expression decreases from the proliferative phase to the ES phase and then gradually increases throughout the secretory phase and sets of transcripts whose expression changes in the opposite direction (Figure S2), similar to patterns of gene-level changes reported previously.[16] This demonstrates the dynamic and variable nature of transcriptomic changes at both gene and transcript levels throughout the menstrual cycle.

Transcript isoform and splicing analyses reveal additional layers of regulation

We also compared genes with evidence of DGE, DTE, DTU, and DS to determine whether transcript isoform-level and alternative splicing analyses can detect genes that would have not been found by gene-level expression analysis. By comparing transcriptomic changes between MP and MS phases (Figure 1B), we found 11,912 genes with changes at the DGE level, 11,930 genes at the DTE level, 2,347 genes at DTU level, and 3,205 genes at the DS level (false discovery rate [FDR] <0.05). Further comparisons showed that many of the genes with evidence of transcript-level and splicing changes were not discovered by the DGE analysis, including 576 (24.5%) of DTU genes, 865 (27.0%) of DS genes, and 1536 (12.9%) of DTE genes, indicating transcript isoform-level and splicing-specific changes. The remaining genes were also identified through gene-level expression analysis, confirming that the observed effects are consistent with the overall pattern of gene expression. Additionally, pathway analysis found that genes detected only by transcript isoform-level and splicing analyses (Bonferroni <0.05) were enriched in biologically meaningful pathways, including hormone regulation and cell growth (Figures 1C and S1). Together, the identification of transcript isoform-level and splicing specific changes adds to our understanding of molecular regulation across the menstrual cycle.

Transcript isoform-level and splicing specific changes in endometriosis

Building on our initial analyses across the menstrual cycle, we next explored whether isoform-specific regulatory changes could underlie transcriptomic differences in endometriosis.

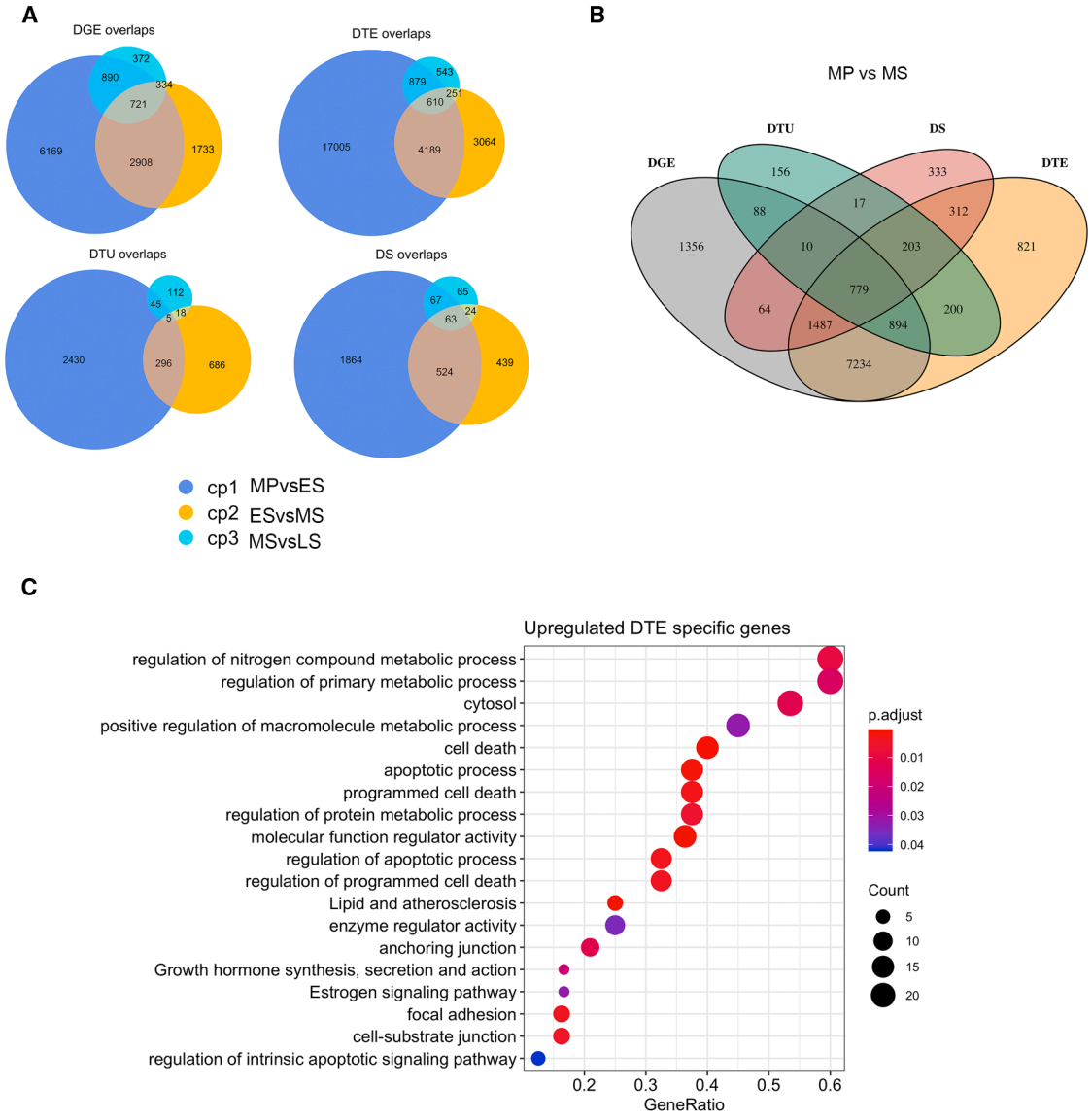


Figure 1. Transcriptomic change across the menstrual cycle

(A) Venn plots show gene-level overlaps of differential gene expression (DGE), differential transcript expression (DTE), differential transcript usage (DTU), and differential splicing (DS) at FDR <0.05 in three consecutive phase comparisons (cp) (cp1: mid-proliferative [MP] vs. early secretory [ES], cp2: ES vs. mid-secretory [MS], cp3: MS vs. late secretory [LS]).

(B) Venn diagram showing gene-level overlaps (FDR <0.05) identified through the four transcriptomic analyses in the comparison of MP vs. MS.

(C) Enriched pathways for up-regulated genes that were only at transcript isoform-level and not detected by DGE analysis. Pathway names are shown on the *y* axis and the enrichment factor on the *x* axis. The circle size indicates the number of genes, and the color bar indicates the adjusted *p* value.

Our previous studies focusing on gene-level expression identified no differentially expressed genes at FDR <0.05 between endometriosis cases and controls.[3,16] In this study, we included the menstrual cycle as a covariate and identified 18 genes (Bonferroni adjusted $p < 0.05$) showing significant evidence of transcript isoform-level and splicing-specific dysregulation associated with endometriosis (Tables 2 and S5). One gene *ZNF217* has been functionally related to hormone regulation and involved in estrogen receptor α (ERα)-mediated signal transduction in breast cancer cells and breast tumor samples.[23] *ZNF217* encodes distinct protein isoforms due to the presence or absence of exon 4.[24] We identified a differentially spliced intron cluster in *ZNF217*, including decreased exon 4-skipping (ΔPSI = −6.4%) in endometriosis (Figure S3).

Co-expression networks captured changes for endometriosis at the MS phase

We also placed transcriptomic changes within a systematic context by performing a weighted gene correlation network analysis (WGCNA),[25] to identify any co-expressed groups of

Table 2. Genes with evidence of transcript-level and splicing dysregulations associated with endometriosis at Bonferroni <0.05

Gene	TranscriptID/IntronCluster	*p* value	adj_pvalue	bf_pvalue	Evidence
ARHGEF12	ENST00000532823.5	6.02E-08	0.004	0.004	DTE
TM7SF3	ENST00000543088.5	4.64E-08	0.001	0.001	DTU
NAPG	ENST00000580224.5	9.48E-08	0.001	0.003	DTU
SLC25A46	ENST00000513807.5	1.94E-07	0.001	0.006	DTU
RRP7BP	ENST00000566851.5	1.52E-07	0.001	0.005	DTU
NEK11	ENST00000508196.5	5.70E-07	0.003	0.018	DTU
CEP44	ENST00000296519.6	7.69E-07	0.003	0.024	DTU
IMPA1	ENST00000518188.5	7.45E-07	0.003	0.023	DTU
DET1	ENST00000444300.1	1.39E-06	0.004	0.044	DTU
KCNJ16	chr17:clu_5500_NA	3.43E-08	0.002	0.002	DS
ZNF217	chr20:clu_59545_NA	1.09E-07	0.002	0.005	DS
KANTR	chrX:clu_13599_NA	1.54E-07	0.002	0.007	DS
PAQR5	chr15:clu_16538_NA	1.92E-07	0.002	0.009	DS
LINC01948	chr5:clu_20866_NA	3.62E-07	0.003	0.017	DS
TSNAXIP1	chr16:clu_48897_NA	6.57E-07	0.004	0.031	DS
NCALD	chr8:clu_24890_NA	6.70E-07	0.004	0.031	DS
PARN	chr16:clu_47322_NA	7.45E-07	0.004	0.035	DS
INTU	chr4:clu_8460_NA	9.89E-07	0.005	0.046	DS

transcripts associated with endometriosis. Results identified 30 co-expressed transcript modules. Only two modules showed significant association with endometriosis; however, this association was not significant after correcting for multiple testing (Figure S4). This result was consistent with our individual transcript level results where only one transcript passed multiple testing at FDR <0.05.

Surprisingly, using the consensus WGCNA in which the same modules were identified in each menstrual cycle phase, we identified that most of those modules showed a different association pattern with endometriosis at each of the four phases (Figure 2). For example, among the top 13 endometriosis associated modules, most were down-regulated in endometriosis at ES phase and up-regulated in MS and LS phases, represented by the different color shading in Figure 2. Specifically, the MS phase captured the greatest number of transcriptomic changes, with significant change for 15 modules following correction for multiple testing (FDR <0.05) (Table S6). Changes for one module remained significant following Bonferroni correction. These results suggest that some gene sets have different roles at different menstrual phases and may only be associated with endometriosis at a specific menstrual cycle phase.

To verify these network results at an individual transcript level, we plotted ten transcripts contributing most to differences for the module most associated with endometriosis at MS phase (MEdarkolivegreen, Beta = −0.55, Bonferroni $p < 0.05$). Compared with the MP phase, all top ten transcripts were significantly down-regulated in endometriosis samples compared with controls at ES, MS, and LS phases (Figures 3A and S4). Pathway analysis of the 181 transcripts (161 genes) in this module revealed that genes were highly enriched for pathways related to fallopian tube and other ciliated cell types and infertility related pathways (Figure 3B). Therefore, this result not only highlighted the important transcriptomic changes at the MS phase but also provided evidence for further investigation of endometriosis related infertility.

Distinct transcriptomic alterations in endometriosis during the mid-secretory phase

Enrichment of isoform- and splicing-level changes in the mid-secretory phase

The identification of transcript-level WGCNA network alterations in endometriosis at the MS phase prompted us to further assess individual levels of transcriptomic dysregulation at this phase. When restricting our transcriptomic level analysis to the MS phase, we identified a larger number of altered transcripts than prior to restriction. This includes DTE for 261 transcripts in 250 genes and DTU for 165 transcripts in 115 genes and 62 alternatively spliced genes (FDR <0.05) between women with and without endometriosis. A proportion also passed the Bonferroni threshold (Bonferroni $p < 0.05$), including 11 DTE in 11 genes, 39 DTU in 29 genes, 11 DS events in 11 genes (Table S7).

Considering the smaller sample size in this analysis when restricting samples to the MS phase ($n = 41$), we evaluated the inflation of all three analyses and found it comparable with analyses using the full sample size ($n = 206$; Figure S5). Additionally, we also explored transcriptomic changes between women with and without endometriosis at the MP phase ($n = 72$) but identified few differences between women with and without endometriosis. These included four DTE, 12 DTU, and three DS events at Bonferroni $p < 0.05$ (Table S8). The results are consistent with the WGCNA network analysis above, suggesting the MS phase captures the most transcriptomic changes.

Predicted functional consequences of dysregulated transcript isoforms and splicing

Using protein domain mapping analysis, we found the altered transcripts usage or splicing were predicted to either affect

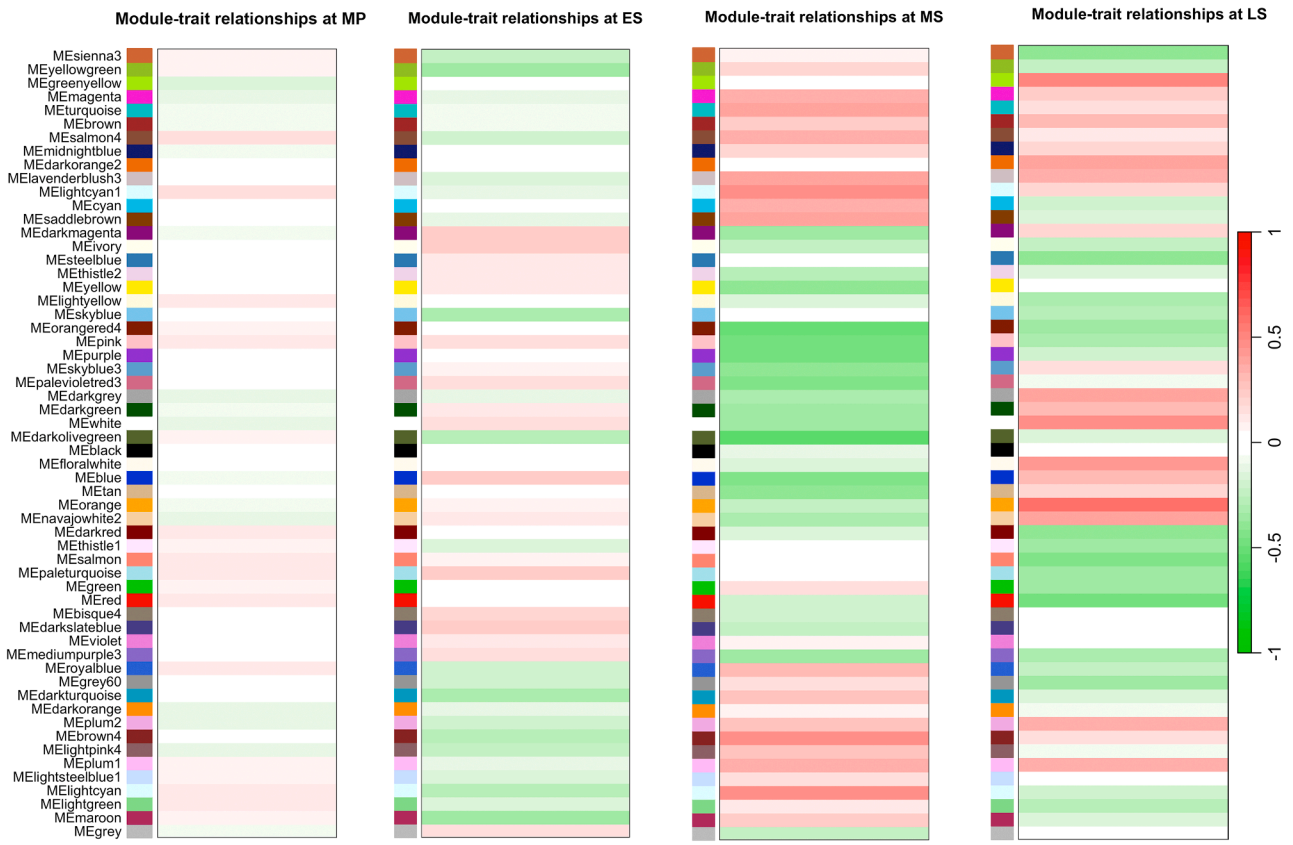


Figure 2. Transcript co-expression network association with endometriosis

Four heatmap plots demonstrating different association patterns between transcript co-expression clusters (*y* axis) and endometriosis (*x* axis) at four different menstrual cycle phases, which from left to right are mid-proliferative (MP), early secretory (ES), mid-secretory (MS), and late secretory (LS) phases. Transcript network clusters on the *y* axis are represented by different color names while their correlation with endometriosis is represented by specific color shading ranging from green (negative correlation) to red (positive correlation).

known protein domain sequences, produce long non-coding RNAs, or generate transcript subject to nonsense mediated mRNA decay (NMD). Examples of predicted functional consequences include *TNC*, *FBLN2*, *PSEN2*, *MALAT1*, *FN1*, and *LEF1* (Figures 4, S6, and S7). Those genes have reported associations with endometriosis and/or infertility through hormone regulation or cellular processes, including cell adhesion, proliferation, invasion, and signal transduction.[26–37] For instance, in *FBLN2* (Figures 4A and 4B), skipping an exon (chr3:13621775-13621915, hg38) which encodes the EGF-like calcium-binding domain (PF07645), results in a shorter transcript isoform (ENST00000295760) with distinct functions compared to other *FBLN2* protein isoforms.[38–40] This gene has also been shown to be regulated in ectopic and eutopic endometrium of a rat model of endometriosis.[34] Other genes have been discussed in the discussion section. Pathway analysis identified that these genes were enriched in pathways related to endometriosis pathogenesis. For example, genes with differential splicing in the MS phase were significantly enriched in cells of mesenchymal origin, endothelial cells, extracellular matrix, and cell migration whereas genes with differential transcript usage were significantly enriched in protein binding pathways attachment (Figure S8). Genes with differential transcript expression were significantly enriched in pathways related to regulation of cell signaling and adhesion (Figure S8).

Genetic regulation of RNA splicing in endometrium

After evaluating the transcriptomic dysregulation in endometrium between women with and without endometriosis, we next sought to identify candidate genes for endometriosis risk with genetically driven effects on RNA splicing. To do this, we first incorporated the corresponding genotype data to map the genetic regulation of RNA splicing (also called splicing quantitative trait loci or sQTL) in endometrium. We then integrated sQTL data with endometriosis GWAS summary statistics to perform a transcriptome-wide association study, prioritizing endometriosis risk genes likely mediated by *cis*-regulatory genetic effects on endometrial RNA splicing.

Mapping genetic effects on RNA splicing

We tested the association between splicing events and 6,230,993 genotyped and imputed SNPs in the same 206 European individuals using two complementary strategies, THISTLE and LeafCutter. Analyzing transcript isoform abundance as the phenotype (THISTLE), we identified 1,433 *cis*-sQTL SNPs (FDR

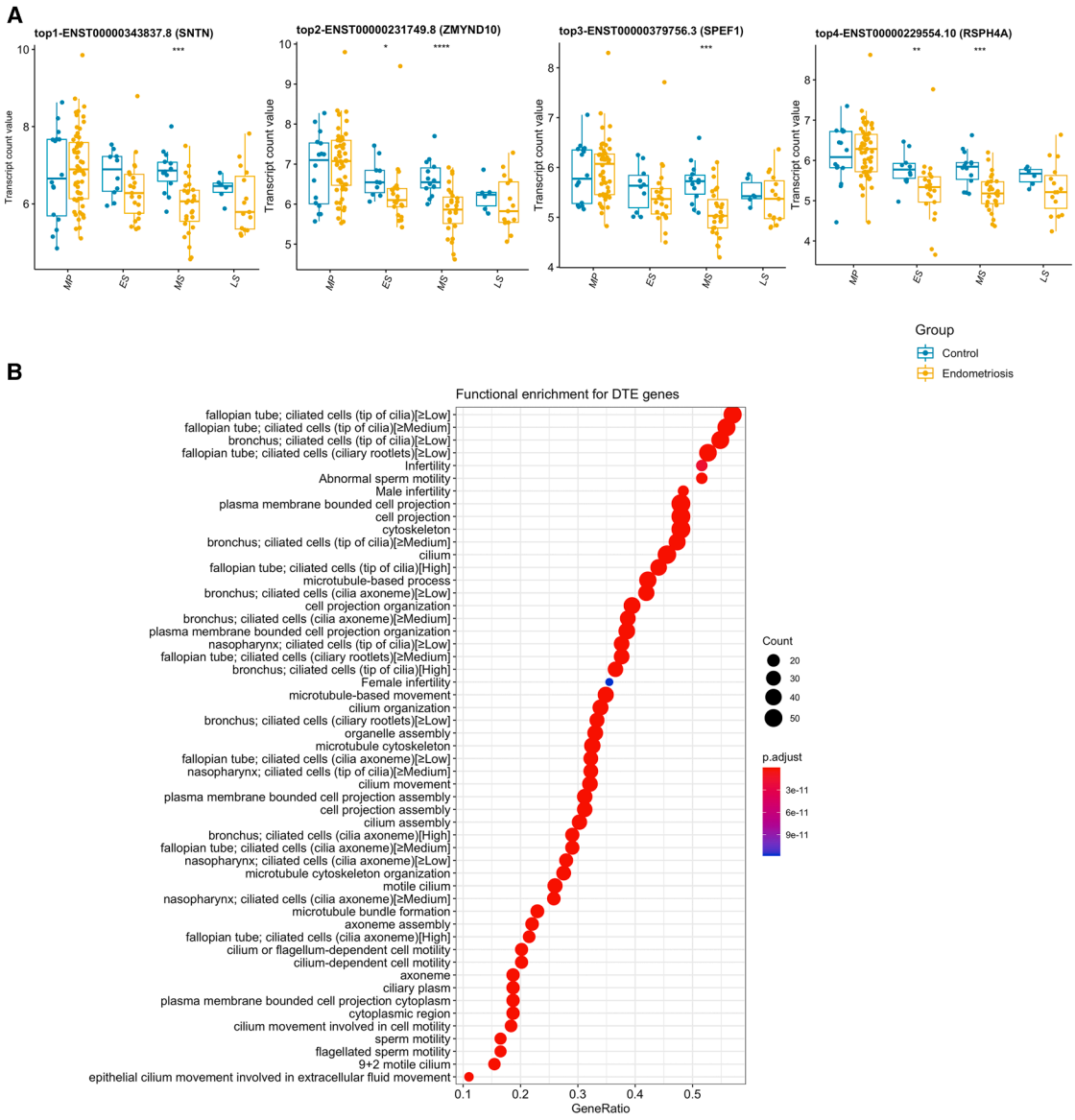


Figure 3. Expression of transcript isoforms and enriched pathways in the endometriosis associated module during the mid-secretory phase

(A) Boxplots show four transcripts with changing expression (*y* axis = transcript count) between women with and without endometriosis at four menstrual cycle phases (*x* axis; mid-proliferative (MP), early secretory (ES), mid-secretory (MS), and late secretory (LS)). These are the top four weighted transcripts in the top associated module (MEdarkolivegreen) with endometriosis at MS phase (Bonferroni $p < 0.05$). *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ****$p < 0.0001$ estimated from differential expression analyses using negative binomial generalized linear models in DESeq2.

(B) Pathway enrichment results for genes involved in the most significantly associated module (MEdarkolivegreen) with endometriosis at MS phase. Pathway names are shown on the *y* axis and the enrichment factor on the *x* axis. The circle size indicates the number of genes, and the color bar indicates the adjusted *p* value.

<0.05) associated with 1,442 genes (sGenes) (Table S9), and when using the intron excision ratio as the phenotype (LeafCutter and QTLtools), we identified 2,072 *cis*-sQTL SNPs (FDR <0.05) associated with 1,727 genes (Table S10). In total, we identified 3,296 *cis*-sQTL SNPs associated with 2,427 genes. Of these, 1,038 *cis*-sQTL SNPs identified through LeafCutter and QTLtools and 715 *cis*-sQTL SNPs identified using THISTLE passed the more stringent Bonferroni threshold of 0.05. Detection of a greater number of sQTLs using LeafCutter and QTLtools compared to THISTLE was in line with the observations in the recent paper from Ting et al. (2022),[20] although with a larger sample size, we expect the number of sQTLs identified by THISTLE would be greater than that by LeafCutter.[20] Approximately, 50% of sGenes identified using THISTLE, were also identified by LeafCutter and QTLtools (Figure S9), supporting the accuracy of both methods and results. Examples of genes identified using

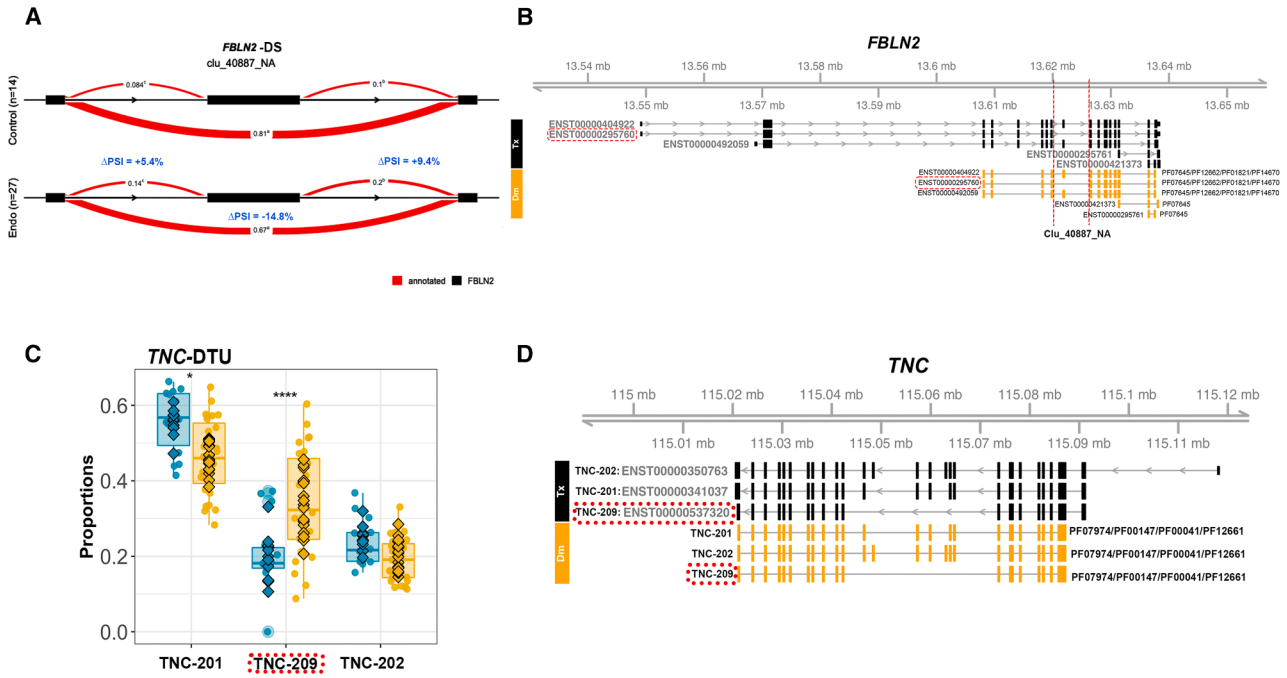


Figure 4. RNA splicing and transcript usage changes between women with and without endometriosis at mid-secretory phase

(A) A significant DS intron cluster in *FBLN2* (clu_40887_NA, Bonferroni $p < 0.05$) shows decreased exon 11 (chr3:13621775-13621915, hg38) skipping (ΔPSI = −14.8%) in endometriosis. Each curved line represents one splicing event and ΔPSI (delta percent spliced in) represents splicing change.

(B) Whole gene view of *FBLN2* with mapped transcript isoforms (Tx) and protein domains (Dm). The dashed line highlights the differentially spliced region in (A), while red rectangle indicates the transcript affected by the exon 11 skipping, along with its associated protein domains.

(C) Boxplot for three transcripts of *TNC* displaying transcript usage change, with * indicating nominally significant $p < 0.05$ while **** indicating Bonferroni significant $p < 0.05$ estimated using Dirichlet-multinomial (DM) models in DRIMSeq for DTU analysis. *y* axis represents the proportional contribution of each transcript to the gene's overall expression.

(D) Whole gene view of *TNC* with mapped Tx and Dm.

both methods include the association between rs16826279 and splicing in *CDC42* (Figure S10), a gene which has been consistently reported to be associated with endometriosis risk.[41] However, the sGenes and sQTLs that do not overlap demonstrated the benefit of using a combination of the two sQTL mapping strategies to capture evidence of RNA splicing.

Splicing-specific genetic effects

To identify potential splicing specific effects sQTLs were compared to gene level eQTLs in endometrium. For this comparison, we considered 2,472 sGenes (FDR <0.05) identified through both sQTL mapping strategies aforementioned (Tables S9 and S10), with the 3,276 endometrial eQTL genes (FDR <0.05) identified from our previous study with the same endometrial RNA-seq data and genotype dataset.[3] We found that of those 2,472 sGenes, 789 (32.5%) genes were also detected in the eQTL analysis, indicating the consistent effects on overall gene expression. However, the majority of sGenes (67.5%) discovered only in the sQTLs analysis suggest splicing-specific effects.

Similarity of genetic effects on splicing between tissues

Understanding the similarity in genetic regulation of RNA splicing between endometrium and other tissues is important to illustrate endometriosis etiology and the potential for shared transcriptomic profiles between biologically similar tissues. The r_b[42] method was used to estimate the correlation of genetic effects at the top-associated *cis*-sQTLs between endometrium and 48 GTEx tissues. Overall, we observed a positive correlation between tissues, with r_b ranging from 0.86 to 0.90 (Figure S9), which is higher and less variable than an estimated rb ranging from 0.54 to 0.72 by using *cis*-eQTLs in endometrium and GTEx tissues.[3] This result was consistent with previous data in GTEx which indicates that genetic effects on splicing tend to be shared more between tissues when compared with expression.[43] Similar to the correlation estimated between *cis*-eQTLs,[3] reproductive and digestive tissues shared the highest correlation of genetic effects on splicing with endometrium.

Enrichment of eQTLs and sQTLs for endometriosis GWAS signals

To assess whether the large number of identified *cis*-sQTLs could be used to capture molecular effects of genetic risk variants associated with endometriosis, we looked for enrichment of top *cis*-sQTL SNPs (FDR <0.05) among endometriosis GWAS SNPs. We also measured the enrichment of *cis*-eQTLs[3] (FDR <0.05) in the endometriosis GWAS and compared this with *cis*-sQTLs. Both lead *cis*-sQTL and eQTL SNPs were enriched for much stronger associations with endometriosis when compared with genome-wide genetic variants (Figure 5). This suggested the importance of annotating GWAS variants using QTL data source.

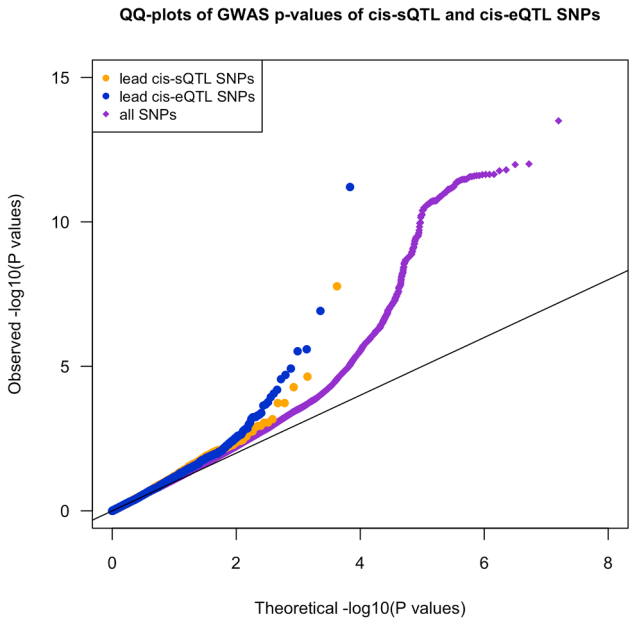


Figure 5. QQ-plots shows enrichment of both lead endometrium *cis*-sQTL and *cis*-eQTL SNPs (FDR <0.05) among endometriosis GWAS signals

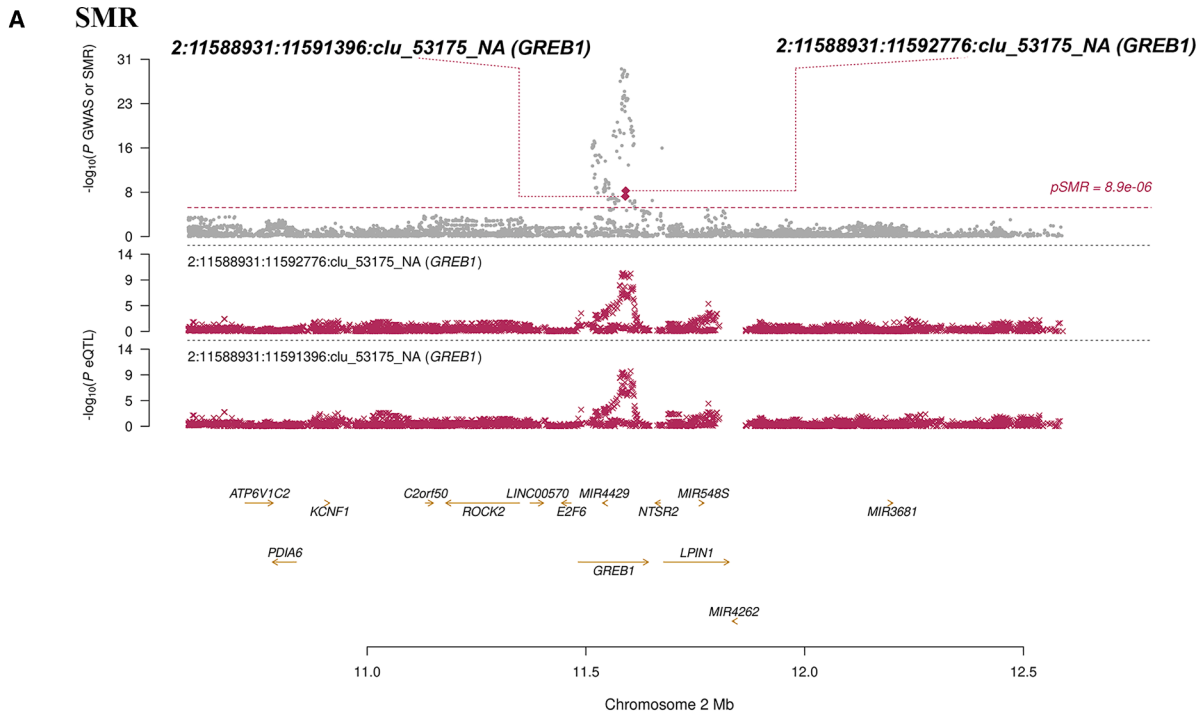
Endometriosis risk genes with genetically driven effects on splicing in endometrium

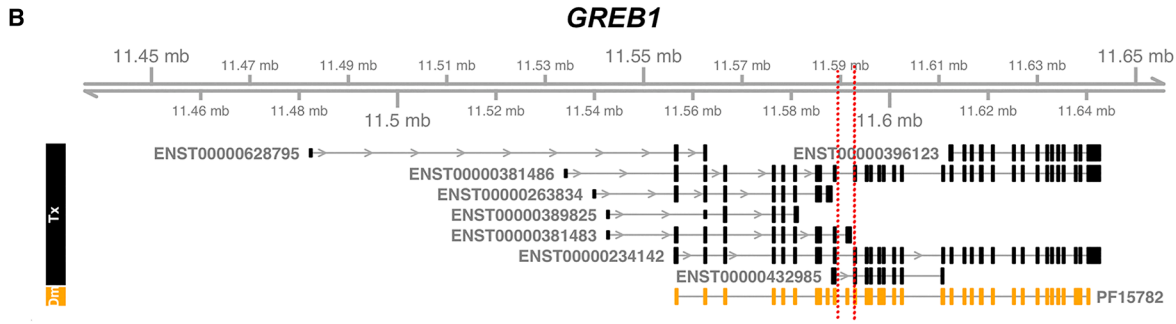
Endometrial *cis*-sQTLs were leveraged to further prioritize risk genes underpinning variant-endometriosis associations from the latest large-scale GWAS analysis on endometriosis[17] using summary data-based Mendelian randomization (SMR).[44] Using the *cis*-sQTLs generated through LeafCutter and QTLtools, SMR analyses identified six genetically regulated splicing events in four genes (*ARL14EP*, *GREB1*, *WASHC3*, and *RP11-624M8.1*) that were significantly associated with endometriosis risk (FDR <0.05), with only two genes *GREB1* and *WASHC3* passing the HEIDI test (p_HEIDI >0.05) (Table S11). We found that two genetically regulated splicing events (a: exon 15 skipping; b: intron excision) in *GREB1* were significantly associated with endometriosis risk (Figure 6A). By further annotating the splicing events (Figures 6B–6D), we found that the alternative exon 15 encodes a stop codon resulting in a truncated transcript isoform (ENST00000381483), which is different from the full-length transcript (ENST00000381486). *WASHC3* was associated with endometriosis through the genetic regulation of exon 6 skipping in WASHC3 (Figure S11). We further found that genetic regulated exon 6 skipping in WASHC3 might affect the production of a specific transcript isoform (ENST00000538761) subject to NMD. The *GREB1* associated variant, rs12470971, was an intronic variant within *GREB1*, and the *WASHC3* associated variant, rs143511013, was located in a non-coding transcript exon of *LINC02456*, both positioned more than 10 bp from canonical splice sites and without predicted direct splicing effects based on SpliceAI analysis. Furthermore, rs12470971 is over 5 kb from the exon 15 skipping event in *GREB1*, and rs143511013 is over 16 kb from the exon 7 skipping event in *WASHC3*. Additionally, no signal passed FDR tests in the SMR when using *cis*-sQTLs from THISTLE when using the corresponding genotype and transcript usage ratio matrix.

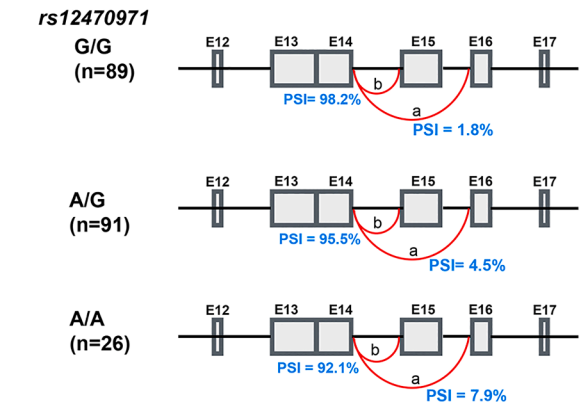
DISCUSSION

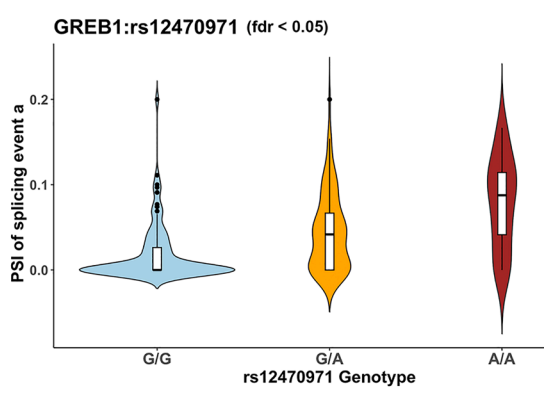
In this study, we conducted transcriptome analysis and genome-wide sQTL mapping within human endometrium using endometrial RNA-seq and genotype data. By employing diverse analytical approaches, including DTE, DTU, and DS, we investigated multiple layers of transcriptomic regulation in both healthy and diseased endometrium. Notably, 30% of genes regulated at transcript level across the menstrual cycle were undetected by standard gene-level expression analyses and were enriched in hormonally responsive and cell growth pathways. Our phase-specific analysis of the MS phase, a critical window of endometrial receptivity, revealed unique, context-specific splicing differences between women with and without endometriosis. Similarly, sQTL analysis revealed distinct genotype-dependent splicing events that are often overlooked by traditional approaches that average across genotypes. Several of these splicing events were linked to known endometriosis risk loci, providing insights into the genetic regulation of splicing and its potential role in endometriosis pathogenesis. Together, our findings emphasize the complementary nature of multi-layered transcriptomic approaches and highlight transcript-level and splicing regulation as key contributors to endometrial function and disease, offering potential target genes for endometriosis treatment and research.

More than half of women with infertility also have endometriosis.[45] It has been suggested that the association between endometriosis and infertility is due to decreased endometrial receptivity and defective implantation at the MS phase, but the causal effects remain unclear.[46,47] The identification of transcriptomic changes associated with endometriosis in the MS phase suggests cellular dysregulation at the MS phase might be related to endometriosis pathogenesis and infertility. For example, there was a transcript switch in *PSEN2* (Figure S6), a member of the Notch pathway, involved in the cellular regulation and transition during the window of uterine receptivity.[48] *PSEN2* has been identified as a potential receptor of protopanaxadiol that could improve endometriosis associated infertility and miscarriage.[49] We also found one *TNC* transcript that encodes a truncated protein isoform that was significantly up-regulated at transcript-usage level in endometriosis at the mid-secretory phase (Figure 4). Previous studies have demonstrated functional differences of *TNC* protein isoforms in tumors compared with normal tissue.[50,51] Interestingly the *TNC* gene encoded protein, tenascin C, classified as an adhesion-modulating protein, is also elevated in eutopic endometrial tissue in endometriosis when compared with controls.[52] Tenascin C protein concentrations have been suggested as a potential serum biomarker to distinguish endometriosis from abnormal pain at the secretory phase.[52] The function of TNC protein isoforms have not been investigated in endometriosis. Taken together, genes with evidence of RNA splicing and transcript alternations in endometriosis may serve as potential biomarkers and therapeutic targets, especially considering the promising translational value of RNA splicing in clinical applications.[53,54]









(legend on next page)

Most splicing variants associated with endometriosis were not uniquely expressed in the MS phase but showed case control differences only during this phase, suggesting a potential phase-specific disease effect. These splicing changes may be driven by dynamic hormonal, molecular, and epigenetic factors that fluctuate across the menstrual cycle. Steroid hormones have been shown to influence alternative splicing in a context-specific manner through multiple mechanisms, including regulating splicing factor expression, recruiting co-regulatory proteins, altering splicing of key regulators, and modulating transcriptional elongation.[55] In addition, epigenetic modifications such as DNA methylation and histone marks can influence transcription and splice site selection by affecting chromatin accessibility and transcriptional kinetics, contributing to differential splicing across menstrual phases.[56,57] These effects may also be modulated by changes in cell type composition across the cycle.

We also investigated genetic regulation of RNA splicing in endometrial tissue by mapping sQTLs using two complementary methods (THISTLE and LeafCutter & QTLtools). Most genes identified in the sQTL mapping analysis (67.5%) (FDR <0.05) were not detected in the eQTL analysis, also suggesting specific genetic regulation of splicing in the endometrium. As shown in this study, sQTLs in endometrium were enriched in endometriosis GWAS statistics, indicating the important contribution of sQTLs in mediating genetic effects. Using the endometrium sQTL data generated in this study along with the publicly available sQTL data in GTEx tissues, we highlight the shared genetic effects on RNA splicing between endometrium and reproductive and digestive tissues, such as ovary, colon, and esophagus. This supports observations of the genetic correlation and shared genetic risk between endometriosis and gastrointestinal disorders and ovarian cancer.[58–61] Endometrial sQTLs are a valuable resource to further investigate other molecular mechanisms and inform diagnostic and therapeutic strategies for reproductive diseases.

Lastly, by integrating endometrial sQTL data and endometriosis GWAS summary statistics,[41] we prioritized two genes (*GREB1* and *WASHC3*) associated with endometriosis risk through genetically driven effects on RNA splicing. This contrasts with previous eQTL studies in endometrium[3] that did not detect genetic effects on expression of both genes, highlighting the added resolution offered by transcript level analysis. *GREB1*, an estrogen-responsive gene that mediates interactions between estrogen receptors and corresponding proteins, has variants significantly associated with endometriosis risk in women from the GWAS studies.[17,41] Our results show that rs12470971 is associated with exon 15 skipping in *GREB1*, resulting in a truncated transcript isoform (ENST00000381483), and previous work suggests that different *GREB1* isoforms can differentially regulate cellular proliferation in hormone-responsive tissues such as breast cancer cells.[62] This raises the possibility that genetic regulation of *GREB1* splicing could contribute to endometriosis pathogenesis via altered protein isoforms in the endometrium. Notably, while rs12470971 lies within an intron of *GREB1*, it was not predicted to directly impact splicing using the SpliceAI tool and was located more than 5 kb from the splicing event and 10 bp from canonical splice sites, suggesting that the splicing effect may be mediated through more complex regulatory mechanisms. Interestingly, rs12470971 is in high LD with rs1865573 (r^2 = 0.98) a variant located in a predicted enhancer in uterus and reported to be associated with endometriosis risk through its *cis*-genetic regulation of *GREB1* methylation in endometrium and an sQTL in ovary[57] Both variants are in high LD with rs11674184 (r^2 = 0.76) previously associated with risk for both endometriosis and clear cell ovarian cancer.[55] This finding further suggests a potential interaction between genetically regulated splicing and methylation at *GREB1* locus, contributing to both risk of endometriosis and ovarian cancer.

The *WASHC3* locus has not been previously linked to endometriosis risk. In our analysis, we identified that rs143511013 was associated with exon 6 skipping in *WASHC3*, and this SNP is in high LD with the index SNP (rs10860864, r^2 = 0.95) located in the *IGF1* locus downstream of *WASHC3* on chromosome 12 and associated with endometriosis risk from the latest endometriosis GWAS.[17] Although rs143511013 resides in a non-coding exon of *LINC02456* and is not predicted to directly affect splicing, protein mapping analysis demonstrated that exon 3 skipping in *WASHC3* is associated with the production of a specific transcript isoform (ENST00000538761), which functions in NMD, a translation-dependent mRNA surveillance mechanism. Recent Mendelian randomization analysis has also reported associations between *WASHC3* plasma protein levels and endometriosis risk.[63] Together, these results suggest *WASHC3* may play a role in endometriosis pathogenesis and illustrate how non-coding variants can modulate splicing indirectly, via mechanisms like RNA-binding protein interactions or long-range chromatin effects, typical of sQTLs.[43]

The splicing changes observed in endometriosis, particularly during the mid-secretory phase, may reflect functionally relevant disruptions in endometrial biology. Genes such as *GREB1*, *WASHC3*, *PSEN2*, and *TNC* showed differential splicing or isoform usage, implicating pathways involved in cell signaling, adhesion, and hormonal response. While direct links between genetic risk variants and splicing remain complex, evidence,

Figure 6. Genetic control of RNA splicing in human endometrium and transcriptome-wide association

(A) SMR results show the association of *GREB1* with endometriosis through the *cis*-effect of rs12470971 on splicing events in *GREB1*. The top track shows −log10(*p*-values) of SNPs from the endometriosis GWAS. The middle track shows −log10(*p*-values) from the *cis*-sQTLs analysis through Leafcutter and QTLtools. The bottom track shows the gene annotation in this region.

(B) Whole view of *GREB1* with mapped transcripts (Tx) and protein domains (Dm). The region highlighted in the red rectangle indicating the spliced region which contains two splicing events associated with different genotypes of rs12470971 (FDR <0.05) in (C).

(C) Association between sQTL rs12470971 genotypes and splicing events a (exon 15 skipping) and b (intron excision). The blue number represents the average PSI (percent spliced in) value for the splicing event in each genotype of rs12470971.

(D) Violin plot shows the association between different rs12470971 genotypes (*x* axis) and PSI value (*y* axis) of splicing event a.

including our identification of splicing alterations in *GREB1*, a gene near known endometriosis and ovarian cancer risk loci, supports this as a plausible mechanism. Recent endometriosis GWAS studies have also shown that SNPs in noncoding regions like lncRNAs can influence splicing and RNA metabolism.[64] These findings highlight the intersection of genetic, epigenetic, and transcriptomic regulation in a tissue- and phase-specific context, underscoring the importance of splicing analyses in understanding endometriosis pathogenesis.

Limitations of the study

There are several limitations to these results. Firstly, the accuracy of transcript-level analysis and RNA splicing measurement using a transcript usage ratio strategy is limited by the short-read RNA-seq which cannot precisely characterize full-length transcript isoforms. This could be improved by the emerging long-read sequencing technology, which can sequence single reads of 30 kb or longer covering the full length of most transcripts.[65] Secondly, the power of the sQTL analysis and the following transcriptome-wide association analysis are dependent on the sample size of the endometrial dataset. Despite of the sample being large relative to existing transcriptomic datasets in human endometrium, more powerful datasets will be required to capture and validate QTLs with smaller effects. Thirdly, endometriosis is a heterogeneous disease with variation in clinical presentation and response to treatment that suggests the presence of subphenotypes of disease. To better understand the potential role of transcriptomic regulation in the endometrium in endometriosis and related infertility, larger sample sizes with detailed subphenotype information, including fertility information, will be required. Finally, our study identified candidate genes with transcript-level and splicing changes in endometriosis with strong statistical support and many of them were also predicted to affect protein isoforms. Functional validation studies of these effects on transcript regulation will be required elucidated the molecular mechanism and their role disease.

In conclusion, this comprehensive transcriptome-wide analysis and genome-wide sQTL analysis add important biological knowledge and evidence to our understanding of dynamic changes in human endometrium throughout the menstrual cycle, as well as associations between transcriptomic changes and endometriosis. The endometrial sQTL dataset generated is a valuable resource to pinpoint the precise mechanism by which genetic variants are associated with other reproductive diseases.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by Dr Sally Mortlock (s.mortlock@imb.uq.edu.au).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Data: The RNA-seq data on which this analysis was based are deposited in GEO: GSE234354. Endometriosis GWAS summary statistics from the International Endometriosis Genetics Consortium is not publicly available due to data restrictions. GWAS summary statistics containing data from 23andMe, Inc. were made available under a data use agreement that protects participant privacy. Please contact dataset-request@23andme.com or visit research.23andMe.com/collaborate for more information and to apply to access the data.
- The V8 release of the GTEx sQTL summary data in SMR binary format was downloaded from Yang Lab (https://yanglab.westlake.edu.cn/software/smr/#DataResource) and was formatted into text files containing all variant-intron associations based on nominal mode.
- Code: The study used standard analysis tools described in the STAR Methods and the key resources table with parameter settings applied in these analyses.
- Any additional information required to analyze the data reported in this work paper is available from the lead contact Dr Sally Mortlock upon request.

ACKNOWLEDGMENTS

We would like to thank the surgeons, pathologists, theater and booking staff at The Royal Women's Hospital, Melbourne and laboratory staff at the Institute for Molecular Bioscience, The University of Queensland, and most of all the patients who contribute to our research. Research reported in this publication was supported by research grants (GNT1026033, GNT1046880, GNT1049472, GNT1050208, GNT1105321, GNT1147846, GNT1078399, and GNT1177194) from the Australian National Health and Medical Research Council (NHMRC).

AUTHOR CONTRIBUTIONS

F.Y., S.M., and G.W.M. designed the study with input from the other authors. G.W.M., S.M., and P.A.W.R. coordinated data collection, quality control of data, data management, and analysis of the original datasets. F.Y. and S.M. ran additional quality control and filtering of datasets. Data analysis was performed by F.Y. which was interpreted by all authors. T. Q. and A.F.M. helped with RNA splicing and transcript-level expression analyses. F.Y., S.M., and G.W.M. drafted the report with input from all other authors. The final manuscript has been critically revised and approved by all authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Participant information
- METHOD DETAILS
  - Sample information
  - RNA sequencing
  - Genotyping
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - RNA quantification
  - Differential transcript/gene analysis
  - Differential transcript usage analysis (DTU)
  - Differential splicing (DS) analysis
  - Protein domain mapping
  - Transcript co-expression network analysis
  - Functional enrichment of genes and modules
  - Processing of genotype data
  - Splicing quantitative trait loci (sQTL)
  - Similarity of sQTLs between tissues
  - Enrichment of sQTLs and eQTLs in GWAS
  - Transcriptome-wide association analysis

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2025.113207.

Received: October 11, 2024
Revised: March 19, 2025
Accepted: July 22, 2025
Published: July 24, 2025

REFERENCES

1. Critchley, H.O.D., Maybin, J.A., Armstrong, G.M., and Williams, A.R.W. (2020). Physiology of the endometrium and regulation of menstruation. Physiol. Rev. *100*, 1149–1179.
2. Rowlands, I.J., Abbott, J.A., Montgomery, G.W., Hockey, R., Rogers, P., and Mishra, G.D. (2021). Prevalence and incidence of endometriosis in Australian women: a data linkage cohort study. BJOG *128*, 657–665.
3. Mortlock, S., Kendarsari, R.I., Fung, J.N., Gibson, G., Yang, F., Restuadi, R., Girling, J.E., Holdsworth-Carson, S.J., Teh, W.T., Lukowski, S.W., et al. (2020). Tissue specific regulation of transcription in endometrium and association with disease. Hum. Reprod. *35*, 377–393.
4. Chapron, C., Marcellin, L., Borghese, B., and Santulli, P. (2019). Rethinking mechanisms, diagnosis and management of endometriosis. Nat. Rev. Endocrinol. *15*, 666–682.
5. Zondervan, K.T., Becker, C.M., and Missmer, S.A. (2020). Endometriosis. N. Engl. J. Med. *382*, 1244–1256.
6. Garcia-Alonso, L., Handfield, L.F., Roberts, K., Nikolakopoulou, K., Fernando, R.C., Gardner, L., Woodhams, B., Arutyunyan, A., Polanski, K., Hoo, R., et al. (2021). Mapping the temporal and spatial dynamics of the human endometrium *in vivo* and *in vitro*. Nat. Genet. *53*, 1698–1711.
7. Wang, W., Vilella, F., Alama, P., Moreno, I., Mignardi, M., Isakova, A., Pan, W., Simon, C., and Quake, S.R. (2020). Single-cell transcriptomic atlas of the human endometrium during the menstrual cycle. Nat. Med. *26*, 1644–1653.
8. Matlin, A.J., Clark, F., and Smith, C.W.J. (2005). Understanding alternative splicing: towards a cellular code. Nat. Rev. Mol. Cell Biol. *6*, 386–398.
9. Baralle, F.E., and Giudice, J. (2017). Alternative splicing as a regulator of development and tissue identity. Nat. Rev. Mol. Cell Biol. *18*, 437–451.
10. Dillman, A.A., Hauser, D.N., Gibbs, J.R., Nalls, M.A., McCoy, M.K., Rudenko, I.N., Galter, D., and Cookson, M.R. (2013). mRNA expression, splicing and editing in the embryonic and adult mouse cerebral cortex. Nat. Neurosci. *16*, 499–506.
11. Giudice, J., Xia, Z., Wang, E.T., Scavuzzo, M.A., Ward, A.J., Kalsotra, A., Wang, W., Wehrens, X.H.T., Burge, C.B., Li, W., and Cooper, T.A. (2014). Alternative splicing regulates vesicular trafficking genes in cardiomyocytes during postnatal heart development. Nat. Commun. *5*, 3603.
12. Bhate, A., Parker, D.J., Bebee, T.W., Ahn, J., Arif, W., Rashan, E.H., Chorghade, S., Chau, A., Lee, J.H., Anakk, S., et al. (2015). ESRP2 controls an adult splicing programme in hepatocytes to support postnatal liver maturation. Nat. Commun. *6*, 8768.
13. Chabot, B., and Shkreta, L. (2016). Defective control of pre-messenger RNA splicing in human disease. J. Cell Biol. *212*, 13–27.
14. Scotti, M.M., and Swanson, M.S. (2016). RNA mis-splicing in disease. Nat. Rev. Genet. *17*, 19–32.
15. Panousis, N.I., Bertsias, G.K., Ongen, H., Gergianaki, I., Tektonidou, M.G., Trachana, M., Romano-Palumbo, L., Bielser, D., Howald, C., Pamfil, C., et al. (2019). Combined genetic and transcriptome analysis of patients with SLE: distinct, targetable signatures for susceptibility and severity. Ann. Rheum. Dis. *78*, 1079–1089.
16. Fung, J.N., Mortlock, S., Girling, J.E., Holdsworth-Carson, S.J., Teh, W.T., Zhu, Z., Lukowski, S.W., McKinnon, B.D., McRae, A., Yang, J., et al. (2018). Genetic regulation of disease risk and endometrial gene expression highlights potential target genes for endometriosis and polycystic ovarian syndrome. Sci. Rep. *8*, 11424.
17. Rahmioglu, N., Mortlock, S., Ghiasi, M., Møller, P.L., Stefansdottir, L., Galarneau, G., Turman, C., Danning, R., Law, M.H., Sapkota, Y., et al. (2023). The genetic basis of endometriosis and comorbidity with other pain and inflammatory conditions. Nat. Genet. *55*, 423–436.
18. Wang, D., Liu, S., Warrell, J., Won, H., Shi, X., Navarro, F.C.P., Clarke, D., Gu, M., Emani, P., Yang, Y.T., et al. (2018). Comprehensive functional genomic resource and integrative model for the human brain. Science *362*, eaat8464.
19. Gandal, M.J., Zhang, P., Hadjimichael, E., Walker, R.L., Chen, C., Liu, S., Won, H., van Bakel, H., Varghese, M., Wang, Y., et al. (2018). Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. Science *362*, eaat8127.
20. Qi, T., Wu, Y., Fang, H., Zhang, F., Liu, S., Zeng, J., and Yang, J. (2022). Genetic control of RNA splicing and its distinct role in complex trait variation. Nat. Genet. *54*, 1355–1363.
21. Li, Y.I., Wong, G., Humphrey, J., and Raj, T. (2019). Prioritizing Parkinson's disease genes using population-scale transcriptomic data. Nat. Commun. *10*, 994.
22. Li, Y.I., Knowles, D.A., Humphrey, J., Barbeira, A.N., Dickinson, S.P., Im, H.K., and Pritchard, J.K. (2018). Annotation-free quantification of RNA splicing using LeafCutter. Nat. Genet. *50*, 151–158.
23. Nguyen, N.T., Vendrell, J.A., Poulard, C., Győrffy, B., Goddard-Léon, S., Bièche, I., Corbo, L., Le Romancer, M., Bachelot, T., Treilleux, I., and Cohen, P.A. (2014). A functional interplay between ZNF217 and estrogen receptor alpha exists in luminal breast cancers. Mol. Oncol. *8*, 1441–1457.
24. Bellanger, A., Le, D.T., Vendrell, J., Wierinckx, A., Pongor, L.S., Solassol, J., Lachuer, J., Clezardin, P., Győrffy, B., and Cohen, P.A. (2021). Exploring the significance of the exon 4-skipping isoform of the ZNF217 oncogene in breast cancer. Front. Oncol. *11*, 647269.
25. Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinf. *9*, 559.
26. Liang, Z., Wu, Q., Wang, H., Tan, J., Wang, H., Gou, Y., Cao, Y., Li, Z., and Zhang, Z. (2022). Silencing of lncRNA MALAT1 facilitates erastin-induced ferroptosis in endometriosis through miR-145-5p/MUC1 signaling. Cell Death Discov. *8*, 190.
27. Yu, J., Chen, L.H., Zhang, B., and Zheng, Q.M. (2019). The modulation of endometriosis by lncRNA MALAT1 via NF-κB/iNOS. Eur. Rev. Med. Pharmacol. Sci. *23*, 4073–4080.
28. Albayrak, İ.G., Azhari, F., Çolak, E.N., Balcı, B.K., Ülgen, E., Sezerman, U., Baştu, E., and Günel, T. (2021). Endometrial gene expression profiling of recurrent implantation failure after *in vitro* fertilization. Mol. Biol. Rep. *48*, 5075–5082.
29. Tamm, K., Rõõm, M., Salumets, A., and Metsis, M. (2009). Genes targeted by the estrogen and progesterone receptors in the human endometrial cell lines HEC1A and RL95-2. Reprod. Biol. Endocrinol. *7*, 150.
30. Flores, I., Rivera, E., Ruiz, L.A., Santiago, O.I., Vernon, M.W., and Appleyard, C.B. (2007). Molecular profiling of experimental endometriosis identified gene expression patterns in common with human disease. Fertil. Steril. *87*, 1180–1199.
31. Shukla, V., Popli, P., Kaushal, J.B., Gupta, K., and Dwivedi, A. (2018). Uterine TPPP3 plays important role in embryo implantation via modulation of β-catenin. Biol. Reprod. *99*, 982–999.
32. Kobayashi, H., Yamashita, Y., Iwase, A., Yoshikawa, Y., Yasui, H., Kawai, Y., Uchida, K., Uno, N., Akatsuka, S., Takahashi, T., et al. (2012). The ferroimmunomodulatory role of ectopic endometriotic stromal cells in ovarian endometriosis. Fertil. Steril. *98*, 415–422.e12.
33. Aston, K.I., Uren, P.J., Jenkins, T.G., Horsager, A., Cairns, B.R., Smith, A.D., and Carrell, D.T. (2015). Aberrant sperm DNA methylation predicts male fertility status and embryo quality. Fertil. Steril. *104*, 1388–1397.e1-5.
34. Matalliotaki, C., Matalliotakis, M., Rahmioglu, N., Mavromatidis, G., Matalliotakis, I., Koumantakis, G., Zondervan, K., Spandidos, D.A., Goulielmos,

G.N., and Zervou, M.I. (2019). Role of FN1 and GREB1 gene polymorphisms in endometriosis. Mol. Med. Rep. *20*, 111–116.

35. Cardoso, J.V., Perini, J.A., Machado, D.E., Pinto, R., and Medeiros, R. (2020). Systematic review of genome-wide association studies on susceptibility to endometriosis. Eur. J. Obstet. Gynecol. Reprod. Biol. *255*, 74–82.
36. Hiskey, M., Chinn, Z., Cohen, C., and Hanley, K. (2016). LEF1 overexpression is associated with preneoplastic and neoplastic glands of the endometrium. Am. J. Clin. Pathol. *146*.
37. Koler, M., Achache, H., Tsafrir, A., Smith, Y., Revel, A., and Reich, R. (2009). Disrupted gene pattern in patients with repeated *in vitro* fertilization (IVF) failure. Hum. Reprod. *24*, 2541–2548.
38. Zhang, H., Hui, D., and Fu, X. (2020). Roles of fibulin-2 in carcinogenesis. Med. Sci. Monit. *26*, e918099.
39. Ma, Y., Nenkov, M., Schröder, D.C., Abubrig, M., Gassler, N., and Chen, Y. (2021). Fibulin 2 is hypermethylated and suppresses tumor cell proliferation through inhibition of cell adhesion and extracellular matrix genes in non-small cell lung cancer. Int. J. Mol. Sci. *22*, 11834.
40. Law, E.W.L., Cheung, A.K.L., Kashuba, V.I., Pavlova, T.V., Zabarovsky, E. R., Lung, H.L., Cheng, Y., Chua, D., Lai-Wan Kwong, D., Tsao, S.W., et al. (2012). Anti-angiogenic and tumor-suppressive roles of candidate tumor-suppressor gene, Fibulin-2, in nasopharyngeal carcinoma. Oncogene *31*, 728–738.
41. Sapkota, Y., Steinthorsdottir, V., Morris, A.P., Fassbender, A., Rahmioglu, N., De Vivo, I., Buring, J.E., Zhang, F., Edwards, T.L., Jones, S., et al. (2017). Meta-analysis identifies five novel loci associated with endometriosis highlighting key genes involved in hormone metabolism. Nat. Commun. *8*, 15539.
42. Qi, T., Wu, Y., Zeng, J., Zhang, F., Xue, A., Jiang, L., Zhu, Z., Kemper, K., Yengo, L., Zheng, Z., et al. (2018). Identifying gene targets for brain-related traits using transcriptomic and methylomic data from blood. Nat. Commun. *9*, 2282.
43. The GTEx Consortium (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science *369*, 1318–1330.
44. Zhu, Z., Zhang, F., Hu, H., Bakshi, A., Robinson, M.R., Powell, J.E., Montgomery, G.W., Goddard, M.E., Wray, N.R., Visscher, P.M., and Yang, J. (2016). Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nat. Genet. *48*, 481–487.
45. Macer, M.L., and Taylor, H.S. (2012). Endometriosis and infertility: a review of the pathogenesis and treatment of endometriosis-associated infertility. Obstet. Gynecol. Clin. North Am. *39*, 535–549.
46. Vargas, E., García-Moreno, E., Aghajanova, L., Salumets, A., Horcajadas, J.A., Esteban, F.J., and Altmäe, S. (2022). The mid-secretory endometrial transcriptomic landscape in endometriosis: a meta-analysis. Hum. Reprod. Open *2022*, hoac016.
47. Horton, J., Sterrenburg, M., Lane, S., Maheshwari, A., Li, T.C., and Cheong, Y. (2019). Reproductive, obstetric, and perinatal outcomes of women with adenomyosis and endometriosis: a systematic review and meta-analysis. Hum. Reprod. Update *25*, 592–632.
48. Afshar, Y., Miele, L., and Fazleabas, A.T. (2012). Notch1 is regulated by chorionic gonadotropin and progesterone in endometrial stromal cells and modulates decidualization in primates. Endocrinology *153*, 2884–2896.
49. Lai, Z.Z., Yang, H.L., Shi, J.W., Shen, H.H., Wang, Y., Chang, K.K., Zhang, T., Ye, J.F., Sun, J.S., Qiu, X.M., and Li, M.Q. (2021). Protopanaxadiol improves endometriosis associated infertility and miscarriage in sex hormones receptors-dependent and independent manners. Int. J. Biol. Sci. *17*, 1878–1894.
50. Guttery, D.S., Hancox, R.A., Mulligan, K.T., Hughes, S., Lambe, S.M., Pringle, J.H., Walker, R.A., Jones, J.L., and Shaw, J.A. (2010). Association of invasion-promoting tenascin-C additional domains with breast cancers in young women. Breast Cancer Res. *12*, R57.
51. Hancox, R.A., Allen, M.D., Holliday, D.L., Edwards, D.R., Pennington, C.J., Guttery, D.S., Shaw, J.A., Walker, R.A., Pringle, J.H., and Jones, J.L. (2009). Tumour-associated tenascin-C isoforms promote breast cancer cell invasion and growth by matrix metalloproteinase-dependent and independent mechanisms. Breast Cancer Res. *11*, R24.
52. Irungu, S., Mavrelos, D., Worthington, J., Blyuss, O., Saridogan, E., and Timms, J.F. (2019). Discovery of non-invasive biomarkers for the diagnosis of endometriosis. Clin. Proteomics *16*, 14.
53. Di, C., Syafrizayanti, Zhang, Q., Chen, Y., Wang, Y., Zhang, X., Liu, Y., Sun, C., Zhang, H., and Hoheisel, J.D. (2019). Function, clinical application, and strategies of Pre-mRNA splicing in cancer. Cell Death Differ. *26*, 1181–1194.
54. El Marabti, E., and Abdel-Wahab, O. (2021). Therapeutic modulation of RNA splicing in malignant and non-malignant disease. Trends Mol. Med. *27*, 643–659.
55. Le Billan, F., Umogbai, G., and Cummins, C.L. (2023). Regulation of alternative splicing by steroid hormones. Endocrinology *164*, bqad081.
56. Retis-Resendiz, A.M., González-García, I.N., León-Juárez, M., Camacho-Arroyo, I., Cerbón, M., and Vázquez-Martínez, E.R. (2021). The role of epigenetic mechanisms in the regulation of gene expression in the cyclical endometrium. Clin. Epigenetics *13*, 116.
57. Mortlock, S., Houshdaran, S., Kosti, I., Rahmioglu, N., Nezhat, C., Vitonis, A.F., Andrews, S.V., Grosjean, P., Paranjpe, M., Horne, A.W., et al. (2023). Global endometrial DNA methylation analysis reveals insights into mQTL regulation and associated endometriosis disease risk and endometrial function. Commun. Biol. *6*, 780.
58. Mortlock, S., Corona, R.I., Kho, P.F., Pharoah, P., Seo, J.H., Freedman, M. L., Gayther, S.A., Siedhoff, M.T., Rogers, P.A.W., Leuchter, R., et al. (2022). A multi-level investigation of the genetic relationship between endometriosis and ovarian cancer histotypes. Cell Rep. Med. *3*, 100542.
59. Adewuyi, E.O., Mehta, D., Sapkota, Y., International Endogene Consortium, 23andMe Research Team, Auta, A., Yoshihara, K., Nyegaard, M., Griffiths, L.R., Montgomery, G.W., et al. (2021). Genetic analysis of endometriosis and depression identifies shared loci and implicates causal links with gastric mucosa abnormality. Hum. Genet. *140*, 529–552.
60. Nabi, M.Y., Nauhria, S., Reel, M., Londono, S., Vasireddi, A., Elmiry, M., and Ramdass, P.V.A.K. (2022). Endometriosis and irritable bowel syndrome: a systematic review and meta-analyses. Front. Med. *9*, 914356.
61. Yang, F., Wu, Y., Hockey, R., Doust, J., Mishra, G.D., Montgomery, G.W., and Mortlock, S.; the International Endometriosis Genetics Consortium (2022). Evidence of shared genetic factors in the aetiology of gastrointestinal disorders and endometriosis and clinical implications for disease management. Preprint at medRxiv. https://doi.org/10.1101/2022.10.20.22281201.
62. Haines, C.N., Braunreiter, K.M., Mo, X.M., and Burd, C.J. (2018). GREB1 isoforms regulate proliferation independent of ERα co-regulator activities in breast cancer. Endocr. Relat. Cancer *25*, 735–746.
63. Tao, T., Mo, X., and Zhao, L. (2024). Identifying novel potential drug targets for endometriosis via plasma proteome screening. Front. Endocrinol. *15*, 1416978.
64. Sheu, J.J.C., Lin, W.Y., Liu, T.Y., Chang, C.Y.Y., Cheng, J., Li, Y.H., Chen, C.M., Tseng, C.C., Ding, W.Y., Chung, C., et al. (2024). Ethnic-specific genetic susceptibility loci for endometriosis in Taiwanese-Han population: a genome-wide association study. J. Hum. Genet. *69*, 573–583.
65. Castaldi, P.J., Abood, A., Farber, C.R., and Sheynkman, G.M. (2022). Bridging the splicing gap in human genetics with long-read RNA sequencing: finding the protein isoform drivers of disease. Hum. Mol. Genet. *31*, R123–R136.
66. Teh, W.T., Chung, J., Holdsworth-Carson, S.J., Donoghue, J.F., Healey, M., Rees, H.C., Bittinger, S., Obers, V., Sloggett, C., Kendarsari, R., et al. (2023). A molecular staging model for accurately dating the endometrial biopsy. Nat. Commun. *14*, 6222.
67. Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. Nat. Methods *14*, 417–419.

68. Soneson, C., Love, M.I., and Robinson, M.D. (2015). Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000Res. *4*, 1521.

69. Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

70. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

71. Nowicka, M., and Robinson, M.D. (2016). DRIMSeq: a Dirichlet-multinomial framework for multivariate count outcomes in genomics. F1000Res. *5*, 1356.

72. Hahne, F., and Ivanek, R. (2016). Visualizing genomic data using Gviz and bioconductor. In Statistical Genomics: Methods and Protocols, E. Mathé and S. Davis, eds. (New York, NY: Springer New York), pp. 335–351.

73. Rainer, J., Gatto, L., and Weichenberger, C.X. (2019). ensembldb: an R package to create and use Ensembl-based annotation resources. Bioinformatics *35*, 3151–3153.

74. Kolberg, L., Raudvere, U., Kuzmin, I., Vilo, J., and Peterson, H. (2020). gprofiler2 – an R package for gene list functional enrichment analysis and namespace conversion toolset g:Profiler. F1000Res. *9*, ELIXIR-709.

75. Delaneau, O., Ongen, H., Brown, A.A., Fort, A., Panousis, N.I., and Dermitzakis, E.T. (2017). A complete tool set for molecular QTL discovery and analysis. Nat. Commun. *8*, 15452.

76. van de Geijn, B., McVicker, G., Gilad, Y., and Pritchard, J.K. (2015). WASP: allele-specific software for robust molecular quantitative trait locus discovery. Nat. Methods *12*, 1061–1063.

77. Jaganathan, K., Kyriazopoulou Panagiotopoulou, S., McRae, J.F., Darbandi, S.F., Knowles, D., Li, Y.I., Kosmicki, J.A., Arbelaez, J., Cui, W., Schwartz, G.B., et al. (2019). Predicting splicing from primary sequence with deep learning. Cell *176*, 535–548.e24.

78. Love, M.I., Soneson, C., and Patro, R. (2018). Swimming downstream: statistical analysis of differential transcript usage following Salmon quantification. F1000Res. *7*, 952.

79. Letunic, I., Khedkar, S., and Bork, P. (2021). SMART: recent updates, new developments and status in 2020. Nucleic Acids Res. *49*, D458–D460.

80. Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G.A., Sonnhammer, E.L.L., Tosatto, S.C.E., Paladin, L., Raj, S., Richardson, L.J., et al. (2021). Pfam: The protein families database in 2021. Nucleic Acids Res. *49*, D412–D419.

81. Pandurangan, A.P., Stahlhacke, J., Oates, M.E., Smithers, B., and Gough, J. (2018). The SUPERFAMILY 2.0 database: a significant proteome update and a new webserver. Nucleic Acids Res. *47*, D490–D494.

82. Zhang, B., and Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. Stat. Appl. Genet. Mol. Biol. *4*, Article17.

83. Li, Y.I., van de Geijn, B., Raj, A., Knowles, D.A., Petti, A.A., Golan, D., Gilad, Y., and Pritchard, J.K. (2016). RNA splicing is a primary link between genetic variation and disease. Science *352*, 600–604.

84. Wu, Y., Zeng, J., Zhang, F., Zhu, Z., Qi, T., Zheng, Z., Lloyd-Jones, L.R., Marioni, R.E., Martin, N.G., Montgomery, G.W., et al. (2018). Integrative analysis of omics summary data reveals putative mechanisms underlying complex traits. Nat. Commun. *9*, 918.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
RNA-seq raw data for endometrial tissue	Teh et al.[66]	GSE234354
Endometrial sQTL data	This study	Tables S9 and S10
Endometrial eQTL data	Mortlock et al.[3]	http://reproductivegenomics.com.au/shiny/eeqtl2/
GTEx sQTL V8 release in SMR binary format	GTEx consortium[43] and Yanglab	https://yanglab.westlake.edu.cn/software/smr/#DataResource
Endometriosis GWAS summary statistic	International Endometriosis Genetics Consortium Rahmioglu et al.[17]	GCST90205183
Endometriosis GWAS summary statistic	International Endometriosis Genetics Consortium Sapkota et al.[41]	N/A (Not publicly available due to 23andMe data restrictions, but can contact author for request process)
Software and algorithms		
Salmon	Patro et al.[67]	https://combine-lab.github.io/salmon/getting_started/
tximport (v1.18)	Soneson et al.[68]	https://bioconductor.org/packages/release/bioc/vignettes/tximport/inst/doc/tximport.html
edgeR (v3.32.1)	Robinson et al.[69]	https://bioconductor.org/packages/release/bioc/html/edgeR.html
DESeq2 (v1.30.1)	Love et al.[70]	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
DRIMSeq	Nowicka et al.[71]	https://bioconductor.org/packages/release/bioc/html/DRIMSeq.html
LeafCutter	Yang et al.[22]	http://davidaknowles.github.io/leafcutter/
Gviz (v1.34.1)	Hahne et al.[72]	https://bioconductor.org/packages/devel/bioc/vignettes/Gviz/inst/doc/Gviz.html
ensembldb (v2.14.1)	Rainer et al.[73]	https://bioconductor.org/packages/release/bioc/html/ensembldb.html
WGCNA	Langfelder et al.[25]	https://horvath.genetics.ucla.edu/html/CoexpressionNetwork/Rpackages/WGCNA/Tutorials/
gprofiler2 (v0.6.4)	Kolberg et al.[74]	https://cran.r-project.org/web/packages/gprofiler2/vignettes/gprofiler2.html
THISTLE	Qi et al.[20]	https://yanglab.westlake.edu.cn/software/osca/#THISTLE
QTLtools	Delaneau et al.[75]	https://qtltools.github.io/qtltools/
WASP	Van de Geijn et al.[76]	https://github.com/bmvdgeijn/WASP/
SMR	Zhu et al.[44]	https://yanglab.westlake.edu.cn/software/smr/#DataResource
SpliceAI	Jaganathan et al.[77]	https://spliceailookup.broadinstitute.org/

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Participant information

Information on sample collection has been previously published by Mortlock et al. 2020.[3] Briefly, informed consent was obtained from 206 women of European ancestry aged 18-49 years, including 143 with endometriosis (cases) and 63 without endometriosis (controls) recruited at the Royal Women's Hospital (RWH) (n=184) and Melbourne IVF Clinic (IVF) (n=22).[3] Inclusion criteria for individuals in RWH was based on the result of laparoscopic surgery in response to suspected endometriosis with pelvic pain as the primary symptom, while selected individuals in IVF were undergoing treatment for subfertility. The study received approval from the

Human Research Ethics Committees of the Royal Women's Hospital (Projects 11–24 and 16–43), Melbourne IVF (Project 05–11), and the University of Queensland (2020/HE002852). All participants provided informed consent.

METHOD DETAILS

Sample information

Endometrial tissue samples were collected from all 206 women by curettage during investigative laparoscopy surgery, followed by the surgical diagnosis of endometriosis status for RWH women or self-reported endometriosis status for women undergoing treatment for subfertility from the IVF clinic. Menstrual cycle phase of each endometrial biopsy was assessed by an experienced pathologist and categorised into seven different phases described previously.[3] The endometrial tissue was collected into RNALater (Life Technologies, USA) and subsequently frozen at -80°C before later being used for RNA extraction and next generation sequencing. A whole blood sample was collected from each individual prior to the laparoscopic surgery and prepared for DNA extraction and genotype analysis.

RNA sequencing

RNA extraction was performed using the Allprep DNA/RNA Mini Kit (Qiagen, CA). RNA quality and concentration was assessed, and libraries were generated using the Illumina TruSeq Stranded Total RNA Gold protocol. Libraries were sequenced at a length of 75 bp pair-ended reads to a mean depth of 37,490,673 reads on the Illumina HiSeq4000 platform for 178 RWH samples, and 120 bp pair-ended reads to a mean depth of 30,818,062 reads on the Illumina HiSeq2000 platform for 28 IVF samples.[3]

Genotyping

DNA was extracted from blood samples and genotyping was performed using HumanCoreExome or Infinium PsychArray chips (Illumina, USA) for all 206 individuals as described previously.[3]

QUANTIFICATION AND STATISTICAL ANALYSIS

RNA quantification

The quality of reads was assessed, and trimming was performed as described previously.[3] Subsequently, the mapping-based mode in Salmon[67] was used for an accurate quantification of transcript abundance. Briefly, the salmon index for the endometrium transcriptome was built using both the GRCh38 reference transcriptome from GENCODE release 38 and the latest GRCh38 genome reference from NCBI. Transcript abundance was then quantified using the index with parameters of –gcBias for fragment-level GC bias correction and -l -A for automatic library type detection.

Transcript-level counts, abundance and effective transcript lengths were generated from Salmon transcript-level abundances in the R package tximport v1.18.0[68] for the subsequent differential transcript expression (DTE) and differential transcript usage (DTU) analysis. The scaling option *dtuScaledTPM*, which was designed for DTU analyses by scaling transcript-per-million (TPM) abundance, was used to generate counts from abundance. Alternatively, by setting *countsFromAbundance* = *"scaledTPM"*, we also generated adjusted counts for each transcript for the following DTE analysis by accounting the potential differential transcript usage within the gene, which is also the outstanding feature of tximport to improve the gene-level analysis.[68] For this reason, the same parameter was also applied for the gene-level counts for the following differential gene expression (DGE) analysis between menstrual cycle phases.

Differential transcript/gene analysis

To assess the differential transcript and gene expression (DTE/DGE), the *filterByExpr* function implemented in edgeR (v3.32.1) was used to filter low-expressed transcripts or genes. The filtered tximport output which contained a count matrix, length matrix and covariate matrix, were then imported into the *DESeqDataSetFromTximport* function implemented in the DESeq2 R package (v1.30.1) to generate the DESeq2 data frame for the differential expression analyses using negative binomial generalized linear models. Differences between women with and without endometriosis were tested as well as between menstrual cycle phases. Batch effects derived from flow-cell and lane were included as covariates in all models and the menstrual cycle phase and endometriosis status were included as covariates when comparing between women with and without endometriosis and menstrual phases, respectively, to correct for the known confounders. Only transcripts passing a multiple-testing threshold of Benjamini-Hochberg FDR < 0.05 were considered as significantly differentially expressed. Meanwhile, a more stringent Bonferroni test was also performed in this study. We focused our differential expression analysis on four pairwise comparisons among menstrual cycle phases: mid-proliferative vs. early secretory (MP vs. ES), early secretory vs. mid-secretory (ES vs. MS), mid-secretory vs. late secretory (MS vs. LS), and mid-proliferative vs. mid-secretory (MP vs. MS). These comparisons were selected based on two considerations: (1) each phase included a sufficient sample size (more than 20 samples), and (2) our previous microarray study[16] indicated minimal transcriptomic differences between the proliferative phases. To control for multiple testing, a stringent Bonferroni correction was applied, using a significance threshold of $P < 0.05$ divided by the maximum number of transcripts/genes tested across the four comparisons, and then multiplied by 4 to account for the number of pairwise tests.

Differential transcript usage analysis (DTU)

We extended the differential analysis to investigate RNA splicing. Currently, there are two main strategies to quantify alternative RNA splicing including DRIMSeq, focusing on transcript-level quantification, and LeafCutter, focusing on intron-excision event. DRIMSeq[71] was designed specifically for DTU analysis. By assuming a Dirichlet-multinomial distribution for each gene, DRIMSeq calculated a gene-wise precision parameter. To address the complexity of the human transcriptome such as the diversity and number of transcripts per gene which may lead to a higher FDR compared to simpler organisms,[71] we used the function *dmFilter* implemented in DRIMSeq R package (v1.18.0) to filter out lowly expressed transcripts prior to the DTU analysis. Transcripts with <10 counts in at least the total number of samples in the smaller group, or for which their expression proportion for a single gene was smaller than ten percent, were removed. Additionally, genes with less than 10 counts in 90% of samples were also excluded. The remaining transcripts were used as an input for DRIMSeq for the DTU analysis. As recommended in the DTU analysis workflow presented by Michael I. Love et al.,[78] the stageR R package (v1.12.0) which allows detection of both the genes containing evidence of DTU and the included transcripts participating in the DTU, and *Post-hoc* which helps to avoid the excess of FDR bounds because of the sensitivity of DRIMSeq to detect DTU, were both applied following DRIMSeq analysis. The same comparisons and covariates were applied as in the DTE analysis. Transcripts with a final adjusted p-value less than 0.05 were considered as significantly differentially used between women with and without endometriosis. Meanwhile, the same Bonferroni correction strategy was also applied as in the DTE analysis above.

Differential splicing (DS) analysis

As mentioned above, differential RNA splicing analysis was also conducted by Leafcutter,[22] which focuses on uncovering alternative intron excision events by using short RNA-sequencing reads that span exon-exon junctions, without estimating transcript abundance and without requiring transcript annotations. Leafcutter was used to identify clusters of alternative excised introns across samples and then to compare intron usage in terms of percentage splicing intron (PSI) changes across an entire cluster to detect differential splicing between women with and without endometriosis and between menstrual cycle phases using the Dirichlet-Multinomial generalized linear model. To reduce the number of tests, the rarely used introns present in less than five samples were removed, and small clusters supported by less than 20 reads in at least four individuals in each group were also filtered out. The same comparisons, covariates, and Bonferroni correction strategy were applied as in the DTE and DTU analyses.

Protein domain mapping

Genes with transcriptomic changes like DTU or DS were mapped onto transcripts using Gviz (v1.34.1)[72] and ensembldb (v2.14.1)[73] bioconductor R packages. Protein domain datasets from Smart[79] and Pfam[80] and Superfamily[81] were used to map protein coordinates onto transcript coordinates using the *proteinToGenome* function of ensembldb.

Transcript co-expression network analysis

To place the transcriptomic changes within a system-level network architecture and further investigate the molecular mechanisms of endometriosis we applied the weighted gene correlation network analysis (WGCNA)[25] to transcript level quantifications in this study. WGCNA analysis, which takes the relationship between measured transcripts into consideration, has been widely accepted to explore the system-level functionality of expression profiles.[82] The filtered transcript matrix generated during the DTE analysis using the DESeq2 R package was used as input for the WGCNA analysis. As suggested, the variance stabilizing transformation (vst) function implemented in DESeq2 was performed to generate the normalized transcript-level count matrix. To assess the effects of covariates PCA plots for each covariate were generated using *plotPCA* function and then all covariates including batch effects derived from flow-cell and lane and the menstrual cycle phase were corrected using the *empiricalBayesLM* implemented in WGCNA R package (v1.70.3). The lowest soft-threshold power β which achieves the approximate scale-free topology of $R^2 > 0.8$ was selected to calculate adjacency. We then used the implemented function *blockwiseModules* to construct signed networks and detect modules or correlated transcripts. Specifically, the hierarchical clustering tree (dendrogram) of transcripts was constructed using hierarchical clustering based on dissimilarity, 1-Topological Overlap Matrix (TOM). Modules were defined using the dynamic cut tree with a deepSplit of 2 and robust correlation method biweight midcorrelation (bicor), minimum module size of 30 and mergeCutHeight of 0.25. The association between each module and endometriosis status was assessed using a linear regression model. Only an association with FDR < 0.05 was considered significant.

Given that endometrium is a complex tissue and undergoes different menstrual cycle phases which has been reported to strongly influence gene expression,[16] we proposed that there might be different expression patterns between women with and without endometriosis in various menstrual cycle phases. Expression differences that occur only in a particular menstrual cycle phase may not be detected when combining endometrium samples at different phases together. To address this hypothesis, the consensus WGCNA analysis was conducted at four menstrual cycle phases with sample sizes >20 to build transcript-modules separately and identify their associations with endometriosis at each phase. This includes the mid-proliferative (MP (n=72)), early secretory (ES (n=31), mid-secretory (MS (n=41)) and late secretory (LS (n=21)) phases. Following the same filtering and normalization process as the standard WGCNA analysis above, a total number of 45,027 common transcripts expressed at four phases remained and 58 transcript

modules were subsequently identified. Similarly, the association between each module and endometriosis status at each phase was assessed using linear regression models. The Benjamini-Hochberg test was used for the correction of multiple testing (FDR < 0.05).

Functional enrichment of genes and modules

The *gprofiler2* R v0.6.4 package was adopted in this study for the pathway analysis of genes with differential spliced cluster or transcript usages under different conditions. Background genes are selected from all categories. All nine types of background gene sets from *gprofiler2* were applied, including commonly used Gene Ontology and KEGG pathways. The Benjamini-Hochberg (FDR) multiple testing correction was applied. A pathway was considered significantly enriched at an adjusted-p value < 0.05.

Processing of genotype data

Raw genotype data underwent quality control (QC) based on the outlined protocol.[16] The 282,625 SNPs (hg19) remaining following QC were phased and imputed on the Michigan Imputation Server. Imputed SNPs were filtered out if they had a low imputation quality of $R^2 < 0.8$, missing rate of more than 5%, minor allele frequency of less than 0.01% and Hardy–Weinberg equilibrium $< 1 \times 10^{-6}$. CrossMap (v.0.2.8) was used to lift over SNP positions to Ensembl genome build 38 (GRCh38) and those that failed to remap were manually mapped according to the dbSNP151 GRCh38 release 7. A total number of 6,230,993 SNPs remained for subsequent analysis.

Splicing quantitative trait loci (sQTL)

After evaluating the transcriptomic dysregulation in endometrium between women with and without endometriosis, we next sought to identify any genetic drivers for such changes by investigating the genetic regulation of RNA splicing (also called splicing quantitative trait loci or sQTL). As mentioned previously, there are two main strategies to quantify alternative splicing, (1) DRIMSeq focusing on transcript-level quantification and (2) Leafcutter focusing on intron-excision event-level. The transcript-level strategy favours detecting sQTLs arising from complex splicing events by using an annotated transcript expression ratio as the phenotype. However, the intron-excision event-level strategy has the advantage of identifying different sQTLs as it does not rely on transcript annotations and can be more sensitive to subtle splicing events. Given that each strategy has its own advantages, we used these two complementary strategies for splicing quantification and subsequent sQTL mapping, in order to capture the overall landscape of genetic regulation on splicing.

Briefly, for the transcript-level strategy, the transcript expression matrix used for the DTE analysis above was used as the phenotype input while the genetic information of the same 206 individuals was used as genotype input for THISTLE (testing for heterogeneity between transcript-eQTL effects),[20] a recently published powerful transcript-level sQTL method. THISTLE tests the association of genetic variants with abundance difference between transcripts, which is equivalent to testing the heterogeneity in transcript-eQTL effect between transcripts. SNPs within a 1 Megabase (MB) window of the transcript (*cis*-sQTLs) in either direction were tested. RNA-sequencing lane and flow cell were included as technical covariates in the model along with menstrual cycle phase and endometriosis status. Both nominal and permutation-based procedures were used separately to map sQTLs. Both Benjamini-Hochberg and a more stringent Bonferroni test on permutation p-value were used to correct for multiple testing. Any sQTLs at FDR < 5% were considered significant. Bonferroni significant sQTLs were also reported.

For the intron-excision event level strategy, the intron excision ratio matrix generated by leafcutter was used as the phenotype for QTLtools.[75] Leafcutter only uses reads that cross the junctions, and if a variant is covered by reads that also span injunctions, it may lead to spurious associations between genetic variant and intron excision rate. Therefore, read-mapping bias caused by allele-specific reads was removed by applying WASP[76] for HISAT2 v2.0.5 alignments. Then, similar to the first two steps of the differential splicing analysis, the clusters of alternatively excised introns were identified using alignment results as an input to Leafcutter.[22] Leafcutter was then used to quantify and normalize the intron excision ratios, which are the proportion of reads supporting each alternatively excised intron to total number of reads within its cluster. Introns used in more than 60% of individuals were retained and used as the phenotype matrix for the sQTL mapping analysis. The corresponding genotype matrix for the same 206 individuals used for the THISTLE analysis was also used for QTLtools to test the associations between genetic variants within 1Mb of introns and corresponding intron excision rates. Both nominal pass and adaptive permutations mode (–permute 1000 10000) implemented in QTLtools were applied. The permutation p-value from QTLtools was first controlled for the number of introns tested per cluster by Bonferroni as a more conservative approach to correct for multiple testing. We then performed both Benjamini-Hochberg and Bonferroni testing on the corrected p-value. The number of sQTL was defined as the number of introns clusters with a significant intron (FDR < 0.05), and thus we take the minimum p-value of all introns for each cluster. Bonferroni significant sQTLs were also reported.

Similarity of sQTLs between tissues

To measure the similarity of genetic effects on cis-splicing QTLs in endometrium and other tissues, we adopted a summary-data-based approach r_b[42] using sQTLs identified in our 206 endometrial tissue samples and another 48 GTEx tissues, and further compared these with the previously reported similarity of genetic effects on cis-expression between those tissues.[3] The V8 release of the GTEx sQTL summary data in SMR binary format was downloaded from Yanglab (https://yanglab.westlake.edu.cn/software/smr/#DataResource) and was formatted into text files containing all variant-intron associations based on nominal mode. To avoid ascertainment bias, we selected the top cis-sQTLs (nominal p < 5e-8) at gene level in Cells_Cultured_fibroblasts from GTEx as

the reference tissue, then extracted those common variant-intron association pairs in endometrium and each of the GTEx tissues, and finally calculated the correlation of cis-effects between two tissues. During the analysis, the errors in the estimated sQTL effects due to small sample size and sample overlap were corrected by r_b, and the direction of sQTL effects was also considered.

Enrichment of sQTLs and eQTLs in GWAS

Previous evidence strongly suggested that RNA splicing serves as a primary link between genetic variation and diseases.[83] Therefore, to verify whether endometrium sQTLs can help explain the molecular mechanisms of endometriosis-related variants, we investigated the enrichment of association p-values among the top endometrium sQTLs (FDR < 0.05) identified in this study. The European ancestry-based endometriosis GWAS data published in 2017[41] was used for the enrichment analysis. Specifically, we extracted top sQTL SNPs (FDR < 0.05) which were also tested in this GWAS study and then plotted the distribution of -log10(p-value) of their association with endometriosis. Additionally, to compare with genetic variation explained by gene expression, we also plotted the enrichment of endometriosis association p-values among top eQTL SNPs (FDR < 0.05) published previously.[3]

Transcriptome-wide association analysis

To verify whether genetic variants had a causal effect on endometriosis mediated by splicing or if genetic variants had pleiotropic effects on both splicing and endometriosis risk, we applied summary data-based Mendelian randomisation (SMR).[44] The SMR was conducted by integrating sQTL summary statistics with summary statistics data from European ancestry GWAS meta-analysis on endometriosis that was published in 2023.[17] HEIDI (heterogeneity in dependent instruments) test, which is implemented in SMR and used to distinguish pleiotropy from linkage, was also carried out. We define associations that passed both the SMR test at genome-wide significance level ($P_{smr} < P_{smr}$ /number of tested transcripts) and HEIDI test at $P_{HEIDI} \geq 0.05$ as significant. Given the limited power of our sQTL summary data generated from 206 RNA-seq samples, and the presence of secondary signals in multiple endometriosis risk loci, a multi-SNP-based SMR was also carried out by leveraging information from all significant SNPs in each region, followed by the HEIDI test.[84] Variants significant in the SMR analysis were further annotated using SpliceAI,[77] an online deep learning tool developed by Illumina that predicts the impact of genetic variants on RNA splicing.