



## OPEN Integrative transcriptomic analysis identifies shared EndMT-related gene signatures in endometriosis and recurrent miscarriage

Yue Liang<sup>1,2,6</sup>, Liangcheng Yu<sup>3,6</sup>, Qingde Zhou<sup>1,4</sup>, Danjie Su<sup>1</sup>, Lu Wang<sup>1</sup>, Cong Li<sup>1</sup>, Jingjing Wang<sup>1</sup>, Haihui Wang<sup>5</sup>, Jie Dong<sup>1</sup>✉, Xifeng Xiao<sup>1</sup>✉ & Xiaohong Wang<sup>1</sup>✉

Endometriosis (EMs) and recurrent miscarriage (RM) represent major reproductive health challenges. This study investigates the involvement of endothelial-mesenchymal transition (EndMT) in these conditions through integrative bioinformatics analysis, focusing on the dysregulation of EndMT-related genes (EndMTRGs). Transcriptomic datasets of EMs and RM were retrieved from the gene expression omnibus (GEO) database. Specifically, GSE120103 includes 18 endometriosis and 18 control samples, and GSE165004 includes 24 recurrent miscarriage and 24 control samples. Differentially expressed gene (DEG) analysis and EndMTRG profiling were performed to identify key pathways and hub genes associated with EndMT. Functional enrichment analyses, including gene ontology (GO), Kyoto encyclopedia of genes and genomes (KEGG), and gene set enrichment analysis (GSEA), were conducted. A protein–protein interaction (PPI) network was constructed via the STRING database, and hub genes were identified using cytoHubba algorithms. Immune cell infiltration patterns were evaluated through single-sample GSEA (ssGSEA). Additionally, support vector machine recursive feature elimination (SVM-RFE) was applied as a supplementary method to assist key gene selection, and nomograms were developed for preliminary risk prediction modeling. A total of 13 EndMTRGs were identified, with key genes such as *FGF2*, *ITGB1*, *VIM*, *NR4A1*, *MAPK1*, *SMAD1*, *TUBB3*, and *CDH11* validated across external datasets. ROC curve analysis demonstrated high diagnostic performance of these genes. Regulatory networks involving RNA-binding proteins and transcription factors were delineated. Immune cell-related gene set enrichment analysis (ssGSEA) revealed distinct immune signatures, notably involving gamma-delta T ( $\gamma\delta$  T) cells and monocytes in EMs, and T follicular helper (Tfh) cells and natural killer (NK) cells in RM. Nomograms based on key genes exhibited reasonable predictive performance. This study elucidates the dysregulated EndMT landscape shared by endometriosis and recurrent miscarriage, identifying common gene signatures and immune features. These findings provide molecular insights into the shared pathogenesis of these conditions and highlight potential targets for future translational research.

**Keywords** Endometriosis, Recurrent miscarriage, Endothelial-mesenchymal transition, Immune infiltration, Biomarker, Machine learning

Endometriosis (EMs) and recurrent miscarriage (RM) are widespread reproductive disorders that significantly affect women globally. Endometriosis, which affects approximately 10% of women of reproductive age, is characterized by the presence of endometrial tissue outside the uterus. This condition often leads to dysmenorrhea, chronic pelvic pain, and infertility<sup>1</sup>. Recurrent miscarriage, defined as two or more consecutive pregnancy losses before 20–24 weeks of gestation, affects approximately 2.5% of couples. It is associated with several factors,

<sup>1</sup>Department of Gynecology and Obstetrics and Reproductive Medicine Center, Tangdu Hospital, Fourth Military Medical University, Xi'an, Shaanxi, People's Republic of China. <sup>2</sup>Department of Gynecology and Obstetrics, 987th Hospital of PLA Joint Logistic Support Force, Baoji, Shaanxi, People's Republic of China. <sup>3</sup>Reproductive Medicine Center and Prenatal Diagnosis Center, 901st Hospital of PLA Joint Logistic Support Force, Hefei, Anhui, People's Republic of China. <sup>4</sup>Center for Women, Children and Reproduction, 924th Hospital of PLA Joint Logistic Support Force, Guilin, Guangxi, People's Republic of China. <sup>5</sup>School of Basic Medical Sciences, Fourth Military Medical University, Xi'an, Shaanxi, People's Republic of China. <sup>6</sup>Yue Liang and Liangcheng Yu contributed equally to this work. ✉email: dongjie2020@fmmu.edu.cn; xxfeng926@163.com; wangxh919@fmmu.edu.cn

including chromosomal abnormalities, uterine defects, and endometrial dysfunction<sup>2</sup>. Both of these conditions present significant diagnostic challenges, as the diagnosis of endometriosis relies primarily on subjective symptom assessments, while recurrent miscarriage is typically diagnosed following post-loss evaluations. Laparoscopy remains the gold standard for confirming endometriosis<sup>3,4</sup>. Current treatments for both conditions focus primarily on alleviating symptoms rather than addressing the underlying disease pathogenesis<sup>5,6</sup>. This therapeutic stagnation can be attributed to the incomplete understanding of the mechanisms underlying these diseases. For instance, there is ongoing debate in the management of recurrent miscarriage, particularly regarding the role of antiphospholipid antibodies and immune modulators<sup>7</sup>, with controversial opinions on their use<sup>8</sup>. The economic burden of these conditions is substantial, with endometriosis alone costing over 69.4 billion dollars annually on a global scale, while recurrent miscarriage imposes a cost of £471 million per year in the UK<sup>9,10</sup>. These figures highlight the urgent need to elucidate the pathogenesis of these conditions and to develop novel diagnostic and therapeutic strategies.

Given this unmet clinical need, recent attention has turned toward novel biological processes that may unify the pathophysiology of both diseases. Endothelial-mesenchymal transition (EndMT) has recently emerged as a critical but underexplored mechanism in both endometriosis and recurrent miscarriage. While both epithelial-mesenchymal transition (EMT) and endothelial-mesenchymal transition (EndMT) are characterized by a shift toward mesenchymal phenotypes, they differ in their cellular origin and biological functions. EMT arises from epithelial cells and has been widely studied in cancer progression and tissue repair, whereas EndMT derives from endothelial cells and plays a central role in vascular remodeling, inflammation, and fibrosis<sup>11–14</sup>. Recent studies have identified EndMT as a critical contributor to gynecological disorders such as endometriosis and recurrent miscarriage, where vascular and immune dysregulation are central features<sup>15</sup>. Therefore, our study specifically focused on EndMT-related gene signatures to elucidate the endothelial-derived molecular mechanisms underlying EMS and RM. Despite the traditional classification of these diseases as epithelial disorders, accumulating evidence suggests that EndMT serves as a mechanistic bridge between vascular dysfunction and immune dysregulation in these conditions<sup>16</sup>. EndMT is characterized by the loss of endothelial markers and the acquisition of mesenchymal features, and it plays a pivotal role in promoting fibrosis in endometriosis through TGF- $\beta$  signaling. These pathological effects highlight the multifaceted roles of EndMT beyond its traditional vascular context. Additionally, it impairs spiral artery remodeling in recurrent miscarriage, which is essential for proper placental development<sup>17,18</sup>. Unlike epithelial-mesenchymal transition (EMT), EndMT is particularly influential in modulating immune interactions, including macrophage polarization and cytokine network regulation<sup>19,20</sup>. However, the precise role of EndMT in the pathogenesis of these diseases remains unclear, especially in terms of the conservation of regulatory pathways, the identification of reliable biomarkers, and the development of targeted therapies. These knowledge gaps impede the development of more effective treatments.

To address these gaps, this study adopts a multi-omics strategy to systematically analyze the common molecular mechanisms of EndMT involved in both endometriosis and recurrent miscarriage. By combining transcriptomic data from publicly available repositories with curated EndMT gene profiles, this study aims to identify key molecular modules involved in the EndMT process, uncover biomarkers that could assist in diagnosis, and explore therapeutic targets grounded in mechanistic insights. This approach will enhance our understanding of how EndMT contributes to the pathogenesis of both diseases, with the potential to propose targeted interventions that could modulate the vascular and immune dysfunctions underlying these conditions.

## Materials and methods

### Data acquisition and sample overview

This study utilized the R package GEOquery (version 2.70.0) to download two datasets from the GEO database (<https://www.ncbi.nlm.nih.gov/geo/>): the Endometriosis dataset (GSE120103) and the Recurrent Miscarriage (RM) dataset (GSE165004)<sup>21,22</sup>. These datasets were selected based on criteria such as sample type consistency, data completeness, and clinical annotation quality. Both datasets consist of human endometrial tissue samples with clearly defined case and control groups. GSE120103 is based on the GPL6480 platform and includes 18 endometriosis and 18 control samples, while GSE165004 uses the GPL16699 platform and includes 24 recurrent miscarriage (RM) and 24 control samples. The detailed information of the GEO microarray chips is summarized in Table 1. Other candidate GEO datasets were initially reviewed but excluded due to inadequate sample size or missing clinical information. The two selected datasets provided the most complete transcriptomic resources for the analysis of endometrial lesions and recurrent pregnancy loss—core focuses of this study.

Raw expression matrices were inspected to determine whether log<sub>2</sub> transformation had been applied; for datasets in non-log<sub>2</sub> scale, log<sub>2</sub> transformation was performed. Between-array quantile normalization was applied using `limma::normalizeBetweenArrays` (v3.58.1). Missing values were imputed using the k-nearest neighbors method implemented in the `impute` R package (default k = 10), and probes with > 20% missing values

	GSE120103	GSE165004
Species	Homo sapiens	Homo sapiens
Platform	GPL6480	GPL16699
Samples in Endometriosis/RM group	18	24
Samples in Control group	18	24
PMID	PMID:30760267	PMID:36369952

**Table 1.** GEO microarray chip information. GEO, gene expression omnibus; AF, atrial fibrillation.

were excluded from analysis. Probe annotation was performed using the official GPL platform files; probes without valid gene symbols were removed, and for genes with multiple probes, the probe with the highest average expression across all samples was retained. Quality control was performed using boxplots before and after normalization for each dataset (Supplementary Fig. S1).

Because the datasets were generated on different microarray platforms, all differential expression analyses were conducted within each dataset separately. Cross-cohort evidence was obtained by identifying genes and pathways showing concordant changes in both datasets. Direct batch-effect correction (e.g., ComBat) and matrix merging were not applied, as platform-specific preprocessing and within-cohort analysis avoided potential cross-platform biases.

Endothelial-to-Mesenchymal Transition (EndMT)-related genes (EndMTRGs) were collected from public databases. Specifically, 142 protein-coding EndMTRGs were obtained from the GeneCards database (<https://www.genecards.org/>)<sup>23</sup> using the keyword “Endothelial-to-Mesenchymal Transition”. A complementary search in PubMed with the same keyword identified 10 additional genes. After merging and deduplication, a total of 150 unique EndMTRGs were retained for subsequent analysis (Table S1).

### Identification of differentially expressed genes (DEGs)

For the GSE120103 dataset, samples were divided into Endometriosis and Control groups. While for the GSE165004 dataset, samples were categorized into Recurrent Miscarriage (RM) and Control groups. Differential gene expression analysis was performed using the Limma R package (version 3.58.1) on both the Endometriosis/Control and RM/Control groups<sup>24,25</sup>. To identify DEGs in the primary analysis, *p*-values were adjusted using the Benjamini–Hochberg (BH) method, and genes with adjusted *p*-value (FDR) < 0.05 were considered statistically significant. As a sensitivity analysis for effect-size stringency, we additionally examined results under absolute log<sub>2</sub> fold change ( $|\log_2FC|$ ) thresholds of 0.30 and 0.58 ( $\approx 1.5$ -fold), showing that the principal cross-cohort conclusions were robust to these thresholds (Supplementary Tables 4). Volcano plots and ranked expression plots of top DEGs were generated with the R package ggplot2 (version 3.4.4) to visually represent the results. To identify EndMT-related DEGs associated with both Endometriosis and Recurrent Miscarriage, we intersected the FDR-controlled DEG lists from both datasets, and a Venn diagram was used for visualization to highlight the core shared EndMT genes. Genes meeting FDR significance in one dataset and nominal significance ( $p < 0.05$ ) with the same direction of change in the other were considered cross-cohort consistent candidates.

### Functional and pathway enrichment analysis

Gene Ontology (GO) analysis, covering Biological Process (BP), Cellular Component (CC), and Molecular Function (MF), as well as Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis, were performed to explore the functional enrichment of the EndMTRDEGs<sup>26–30</sup>. The analyses were conducted using the R package clusterProfiler (v4.10.0), with the genome-wide gene list from each dataset used as the background universe.<sup>31</sup> A significance threshold of  $p < 0.05$  and false discovery rate (FDR, *q*-value) < 0.05 was applied, with *p*-value adjustment performed using the Benjamini–Hochberg (BH) method. Enriched terms were visualized with the R package ggplot2.

### Pathway-level gene set enrichment analysis (GSEA)

Gene Set Enrichment Analysis (GSEA) was performed on the GSE120103 and GSE165004 datasets using the R package clusterProfiler (v4.10.0).<sup>32</sup> Gene sets were obtained from the Molecular Signatures Database (MSigDB, v7.5.1), specifically from the c2 curated gene set collection. Genes were ranked by log<sub>2</sub> fold change (log<sub>2</sub>FC) values, and the genome-wide gene list from each dataset was used as the background universe. Parameters for GSEA included a random seed of 2020, a minimum gene set size of 10, and a maximum of 500 genes per gene set. Statistical significance was determined at  $p < 0.05$  and FDR (*q*-value) < 0.05, with *p*-values adjusted using the Benjamini–Hochberg procedure. Representative enrichment plots and network diagrams were generated to illustrate gene–function associations.

### Protein–protein interaction (PPI) network

Protein–protein interaction (PPI) networks were constructed to analyze the interactions among the identified EndMTRDEGs. PPI data were obtained from the STRING database (version 11.0), with the species restricted to *Homo sapiens* and a medium confidence score threshold set to  $\geq 0.4$ .<sup>33</sup> Interactions included experimental evidence, database-curated associations, predicted associations from gene neighborhood, gene fusions, co-occurrence, and functional associations inferred from co-expression and homology, as provided by STRING. The PPI network was visualized using Cytoscape (version 3.9.1)<sup>34</sup>. Key (hub) genes were identified using the cytoHubba plugin in Cytoscape, applying five ranking algorithms: Degree, maximal clique centrality (MCC), edge percolated component (EPC), density of maximum neighborhood component (DMNC), and maximum neighborhood component (MNC). Among these, the top 10 genes were designated as core EndMTRDEGs<sup>35</sup>. The GeneMANIA website was used to predict functionally similar genes, and an interaction network was constructed for these key genes<sup>36</sup>.

### Regulatory network construction

RNA-binding proteins (RBPs) play a crucial role in gene regulation, including RNA synthesis, splicing, modification, transport, and translation<sup>37</sup>. The StarBase v3.0 database (<https://starbase.sysu.edu.cn/>) was used to predict RBPs interacting with the EndMTRDEGs<sup>38</sup>. Only interactions supported by high-confidence CLIP-seq experimental evidence were retained, and low-confidence or computationally predicted interactions without experimental support were excluded. The mRNA-RBP regulatory network was visualized using Cytoscape software. Additionally, gene expression is regulated by transcription factors (TFs) through the control of the

transcriptional process. The ChIPBase v2.0 database (<http://rna.sysu.edu.cn/chipbase/>) was used to retrieve TFs associated with the EndMTRDEGs. Interactions were filtered to include only those with experimental evidence from ChIP-seq datasets, and the interaction score threshold was set to  $\geq 0.4$  to ensure reliability. The TF–mRNA regulatory network was constructed and visualized using Cytoscape<sup>39</sup>.

### Differential expression and core gene identification

To determine key genes linked to Endometriosis and Recurrent Miscarriage, the Mann–Whitney U test (Wilcoxon rank sum test) was employed to examine the expression discrepancies between the Endometriosis/Control and RM/Control groups within the GSE120103 and GSE165004 datasets. Following differential expression analysis, Receiver Operating Characteristic (ROC) curve analysis was performed to assess the diagnostic potential of the core genes<sup>40</sup>.

### Immune cell-related gene set enrichment analysis (ssGSEA)

Single-sample Gene Set Enrichment Analysis (ssGSEA) was applied to evaluate the relative enrichment of immune cell-related gene sets in the samples from Endometriosis and Recurrent Miscarriage<sup>41</sup>. The immune cell gene sets were obtained from the MSigDB C7 collection and previously published signatures (Bindea et al., *Immunity*, 2013), covering multiple immune cell populations including activated CD8<sup>+</sup> T cells, dendritic cells,  $\gamma\delta$  T cells, natural killer (NK) cells, and regulatory T cells, among others. The ssGSEA method calculates enrichment scores, which represent the relative expression of each predefined gene set within a sample, rather than the absolute proportion of immune cells. Enrichment scores were computed using the GSVA package from Bioconductor and visualized through ggplot2 (version 3.4.4). Differences in enrichment scores between groups were compared, and Spearman correlation analysis was used to examine relationships between immune cell types. R packages including pheatmap (version 1.0.12) and ggplot2 were employed to generate heatmaps and correlation bubble charts.

### SVM-RFE and diagnostic model construction

To further identify core genes associated with Endometriosis and Recurrent Miscarriage, feature selection was conducted using the support vector machine recursive feature Elimination (SVM-RFE) algorithm implemented in the R caret package with a radial basis function (RBF) kernel. The model was evaluated using tenfold cross-validation (method = “cv” in caret), and the optimal number of features was determined automatically by the function based on minimizing the root mean square error (RMSE) from cross-validation results, and visually confirmed at the point of minimal RMSE. The SVM-RFE algorithm, commonly used for pattern recognition and classification, was applied to identify the most relevant genes. Nomograms were created using the R package ‘rms’ based on the expression levels of these pivotal genes. To assess model performance and risk of overfitting, 1000 bootstrap resamples were employed for calibration analysis, and calibration curves were generated to evaluate agreement between predicted and observed probabilities. Receiver operating characteristic (ROC) curves were constructed, and the area under the curve (AUC) with 95% confidence intervals (CIs) was calculated for both individual genes and the composite nomogram scores. The Hosmer–Lemeshow (H–L) test were used to assess model fit.

### Statistical analysis

Data processing and statistical analysis were performed using R software (version 4.2.2). Normality of continuous variables was assessed using the Shapiro–Wilk test. For variables meeting the normality assumption, independent Student’s t-tests were applied to compare continuous variables between the two groups, and one-way ANOVA was used for comparisons among three or more groups. For data that were not normally distributed, the Mann–Whitney U test (Wilcoxon rank-sum test) was applied, and the Kruskal–Wallis test was used for comparing three or more groups.

All p-values were adjusted for multiple comparisons using the Benjamini–Hochberg (BH) method unless otherwise specified. Spearman’s correlation analysis was utilized to determine correlation coefficients between different molecules. For key comparisons, effect sizes (Cohen’s d for t-tests, rank-biserial correlation for Mann–Whitney tests) and 95% confidence intervals (CIs) were calculated to enhance statistical interpretation. All p-values were two-tailed, and statistical significance was set at  $p < 0.05$ .

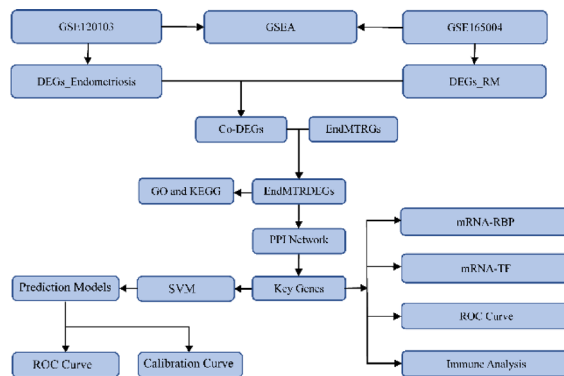
## Results

### Dataset correction

This research mainly using bioinformatics methods to explore the Endometriosis (Endometriosis) and biological characteristics of recurrent miscarriage (RM), flow chart of the overall analysis such as Fig. 1. In data collection GSE120103, samples could be divided into normal sample (Control) group and Endometriosis (Endometriosis), in data collection GSE165004, samples could be divided into normal sample (Control) group and recurrent miscarriage group (RM), The dataset GSE120103 (Fig. S1A–B) and GSE165004 (Fig. S1C–D) standardizing, annotation probe cleaning operations, such as data before and after the standardization and mapped the boxplot of data distribution. Upon normalization, the expression patterns across different samples in the dataset became more consistent.

### Endometriosis and recurrent miscarriage related endothelial-mesenchymal transition related differentially expressed genes

Significant transcriptional alterations were observed in both the GSE120103 and GSE165004 datasets. Differential expression analysis was conducted using the limma R package. To identify Endothelial–Mesenchymal Transition–Related Differentially Expressed Genes (EndMTRDEGs), the overlapping genes between the DEGs



**Fig. 1.** Technology roadmap. This figure illustrates the overall study design, detailing the datasets used, the analytical workflow, and the key bioinformatics methods applied.

from both datasets and a predefined set of EndMT-related genes were determined. Under the primary criterion of  $FDR < 0.05$ , 24 EndMTRDEGs were identified in GSE120103 and 15 in GSE165004. Thirteen genes were present in the intersection: SOX4, TUBB3, ITGB1, NR4A1, PARP1, VIM, SOX7, KITLG, SMAD1, CDH11, MAPK1, FGF2, and ROBO4 (Fig. 2E).

Several of these genes, including SOX4, VIM, ITGB1, and FGF2, showed consistent direction of change across datasets, with VIM and ITGB1 subsequently emerging as top-ranked predictive features in the machine learning analysis. Sensitivity analyses applying  $|\log_2FC|$  thresholds of 0.30 and 0.58 confirmed that the overall cross-cohort overlap pattern was robust to moderate but not strict effect-size cut-offs.

### Gene ontology (GO) and path (KEGG) enrichment analysis

GO and KEGG enrichment analyses were conducted using the R package clusterProfiler (v4.10.0) to investigate the functional roles of the 13 EndMTRDEGs, with the genome-wide gene list from each dataset serving as the background universe. The GO analysis included BP, CC, and MF categories, while the KEGG analysis focused on canonical pathways. Significance was defined as  $p < 0.05$  and FDR (BH-adjusted)  $< 0.05$ . The EndMTRDEGs were enriched in angiogenesis- and tissue-remodeling-related processes, involving classical signaling pathways such as PI3K–Akt and MAPK (Fig. 3A; Table 2). A network diagram was constructed to visualize the associations between EndMTRDEGs and enriched terms, with node size reflecting the number of genes and edges representing gene–term relationships (Fig. 3B–E).

### Gene set enrichment analysis (GSEA)

GSEA was performed on the GSE120103 and GSE165004 datasets using gene sets from the MSigDB c2 curated collection, with genes ranked by  $\log_2FC$  values and the genome-wide gene list from each dataset used as the background universe. Parameters were set with  $seed = 2020$ ,  $minGSSize = 10$ , and  $maxGSSize = 500$ . Significance was determined at  $p < 0.05$  and FDR (BH-adjusted)  $< 0.05$ . In both datasets, GSEA revealed significant enrichment of inflammatory, immune-related, and endothelial–mesenchymal transition (EndMT)–related processes, consistent with the biological context of endometriosis and recurrent miscarriage. Representative enrichment plots are shown in Fig. 4A, E, and the enriched gene sets are listed in Tables 3 and 4. Network diagrams illustrate the associations between EndMTRDEGs and enriched functional pathways (Fig. 4B–E, G–J).

### Proteins-protein interaction network (PPI network)

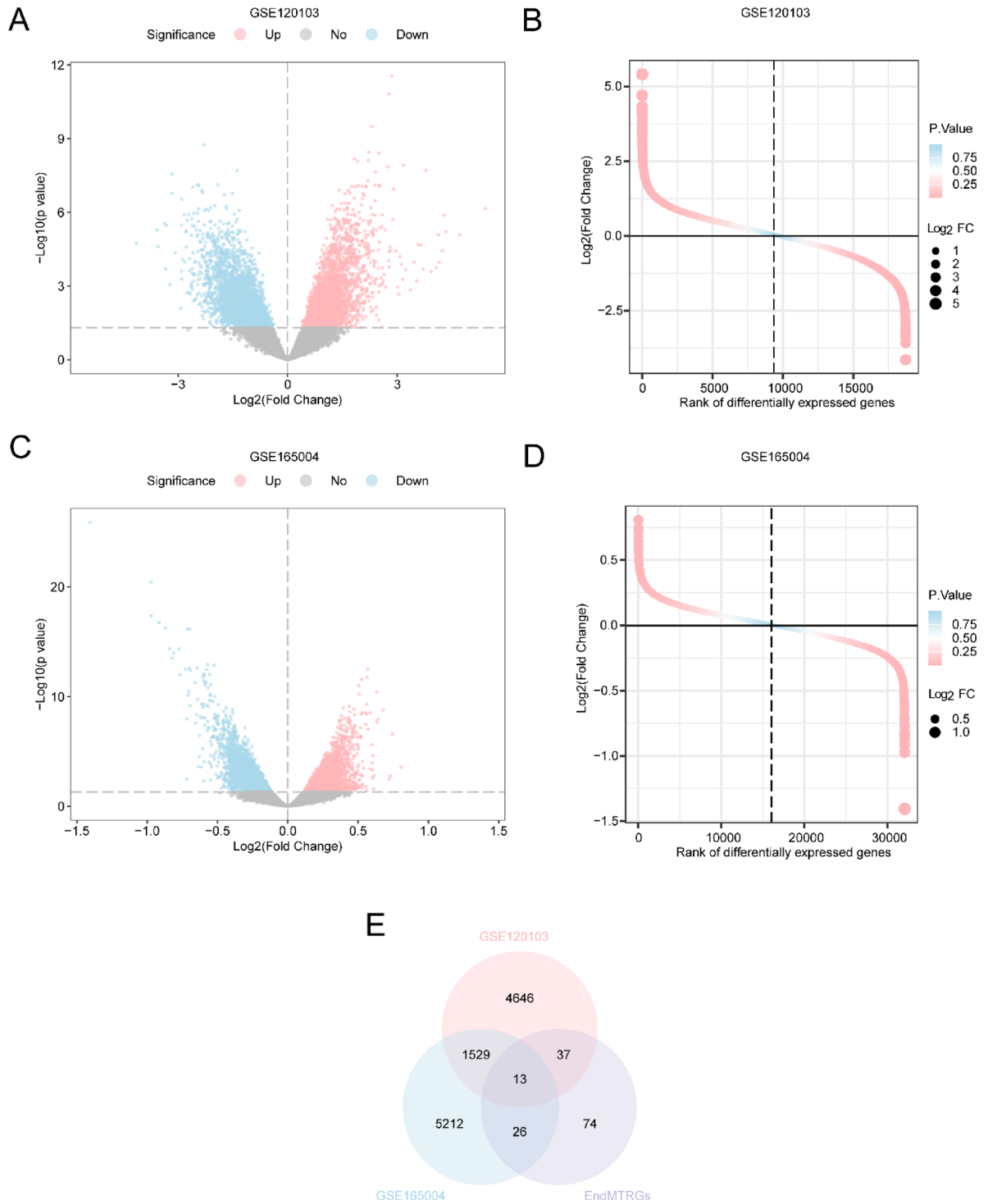
A protein–protein interaction (PPI) network of the 13 EndMT-related differentially expressed genes was generated using the STRING database (v11.0) with the species set to Homo sapiens and a medium confidence threshold (interaction score  $\geq 0.4$ ). Interactions included experimental validation, curated database information, and computational predictions from co-expression and homology. The network was visualized using Cytoscape (Fig. 5A).

Hub genes were identified using the cytoHubba plugin with five ranking algorithms (MCC, MNC, EPC, Degree, DMNC). Genes ranked in the top 10 by each algorithm were collected, and those appearing in at least three ranking lists were defined as hub genes. These hub genes were located at critical network nodes and may represent potential biomarkers or therapeutic targets related to EndMT processes (Fig. 5B–C).

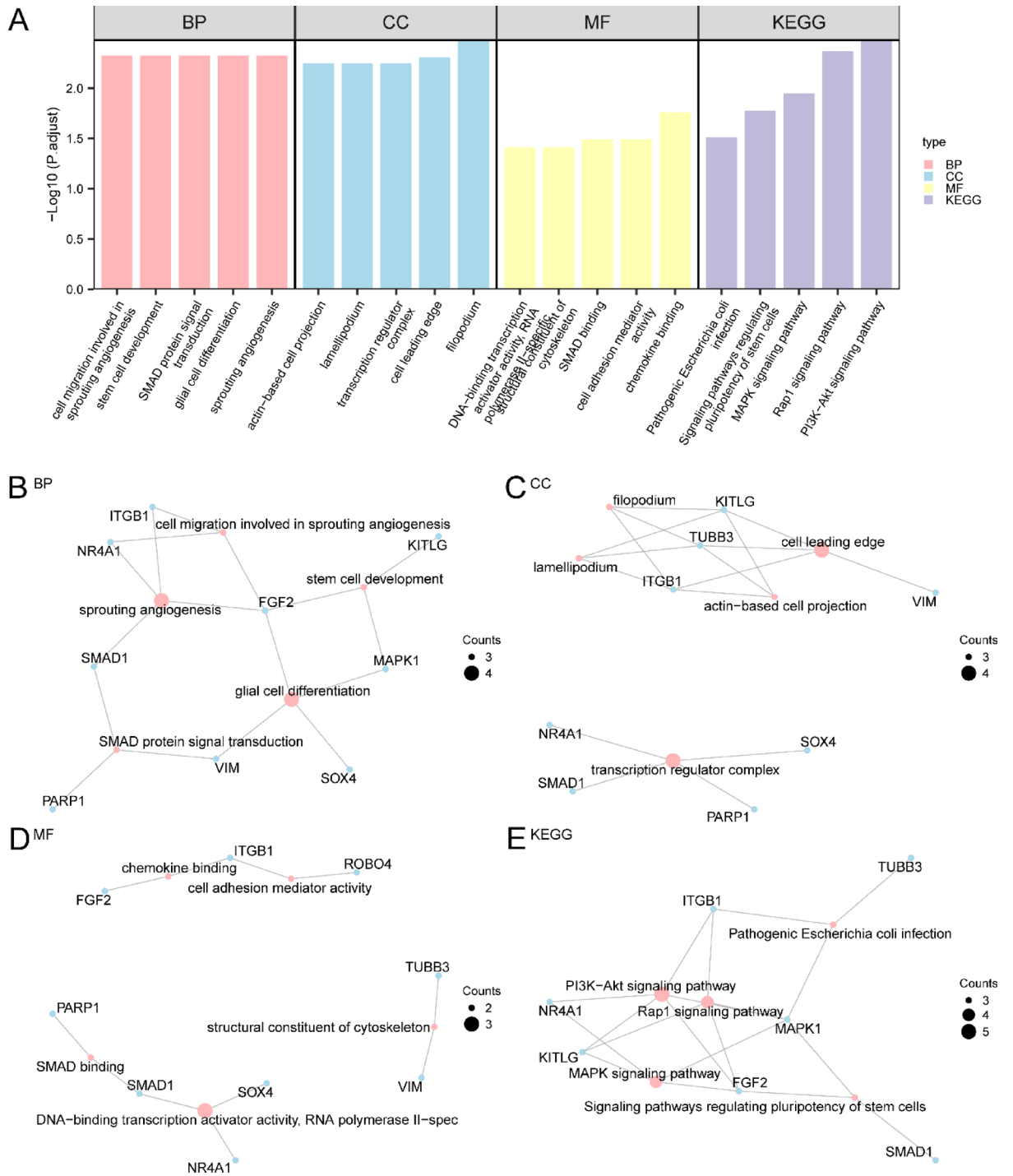
We further used GeneMANIA to predict genes functionally similar to the identified hub genes and to construct an extended interaction network, highlighting shared domains, co-expression patterns, and other functional associations (Fig. 5D).

### Construction of control network

RNA-binding proteins (RBPs) associated with the endothelial–mesenchymal transition-related differentially expressed genes (EndMTRDEGs) were identified using the StarBase database. Cytoscape software was used to construct and visualize the mRNA–RBP regulatory network (Fig. 6 A). This network includes 8 EndMTRDEGs and 117 RBPs. Further details are presented in Table S2.



**Fig. 2.** Differential gene expression analysis. **(A)** Volcano plot showing differentially expressed genes (DEGs) between Endometriosis and Control groups in the GSE120103 dataset. **(B)** Ranked expression plot of DEGs in GSE120103. **(C)** Volcano plot showing DEGs between Recurrent Miscarriage (RM) and Control groups in the GSE165004 dataset. **(D)** Ranked expression plot of DEGs in GSE165004. **(E)** Venn diagram depicting the overlap between DEGs and endothelial-to-mesenchymal transition-related genes (EndMTRGs) in both datasets.



**Fig. 3.** GO and KEGG enrichment analysis of EndMTRDEGs. **(A)** Bar plot showing Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis. **(B)** Network diagram of Biological Process (BP) terms. **(C)** Network diagram of Cellular Component (CC) terms. **(D)** Network diagram of Molecular Function (MF) terms. **(E)** Network diagram of KEGG pathways. Node sizes indicate the number of associated genes.

Transcription factors (TFs) associated with the EndMTRDEGs were extracted from the ChIPBase database. The resulting mRNA-TF regulatory network was subsequently built and visualized using Cytoscape software (Fig. 6B). This network includes 8 EndMTRDEGs and 81 TFs, with detailed information available in Table S3.

**Expression patterns and group separation analysis of hub genes**

We analyzed the expression levels of eight hub genes (FGF2, ITGB1, VIM, NR4A1, MAPK1, SMAD1, TUBB3, CDH11) in the GSE120103 dataset. Violin plots (Fig. 7A) revealed significant differential expression between

Ontology	ID	Description	Gene ratio	Bg ratio	p value	p.adjust
BP	GO:0002040	Sprouting angiogenesis	4/13	186/18,800	6.18594E-06	0.004753578
BP	GO:0010001	Glial cell differentiation	4/13	218/18,800	1.15853E-05	0.004753578
BP	GO:0060395	SMAD protein signal transduction	3/13	83/18,800	2.29861E-05	0.004753578
BP	GO:0048864	Stem cell development	3/13	88/18,800	2.73981E-05	0.004753578
BP	GO:0002042	Cell migration involved in sprouting angiogenesis	3/13	94/18,800	3.3387E-05	0.004753578
CC	GO:0030175	Filopodium	3/13	105/19,594	4.1129E-05	0.003290321
CC	GO:0031252	Cell leading edge	4/13	416/19,594	0.000122998	0.004919913
CC	GO:0005667	Transcription regulator complex	4/13	483/19,594	0.000218455	0.005647961
CC	GO:0030027	Lamellipodium	3/13	202/19,594	0.000286088	0.005647961
CC	GO:0098858	Actin-based cell projection	3/13	217/19,594	0.000352998	0.005647961
MF	GO:0019956	Chemokine binding	2/13	33/18,410	0.000240055	0.017283962
MF	GO:0098631	Cell adhesion mediator activity	2/13	64/18,410	0.000905326	0.032185493
MF	GO:0046332	SMAD binding	2/13	78/18,410	0.001341062	0.032185493
MF	GO:0005200	Structural constituent of cytoskeleton	2/13	104/18,410	0.002367221	0.038686773
MF	GO:0001228	DNA-binding transcription activator activity, RNA polymerase II-specific	3/13	462/18,410	0.003722609	0.038686773
KEGG	hsa04151	PI3K-Akt signaling pathway	5/10	359/8779	2.36519E-05	0.00338222
KEGG	hsa04015	Rap1 signaling pathway	4/10	210/8779	5.96788E-05	0.004267035
KEGG	hsa04010	MAPK signaling pathway	4/10	299/8779	0.000235409	0.011221161
KEGG	hsa04550	Signaling pathways regulating pluripotency of stem cells	3/10	143/8779	0.000467009	0.016695585
KEGG	hsa05130	Pathogenic Escherichia coli infection	3/10	198/8779	0.001206299	0.030901775

**Table 2.** Results of GO and KEGG enrichment analysis for EndMTRDEGs. GO, gene ontology; BP, biological process; CC, cellular component; MF, molecular function; KEGG, Kyoto encyclopedia of genes and genomes; EndMTRDEGs, endothelial-to-mesenchymal transition-related differentially expressed genes.

the Endometriosis and Control groups for several of these genes. To evaluate diagnostic performance, receiver operating characteristic (ROC) curve analysis was conducted (Fig. 7B–C). These hub genes exhibited moderate to high diagnostic potential, with FGF2 and ITGB1 showing particularly strong performance.

In the GSE165004 dataset, the expression patterns of these genes were also compared between Recurrent Miscarriage (RM) and Control groups (Fig. 7D), and significant differences were observed for most of them. ROC analysis (Fig. 7E–F) further confirmed that FGF2 maintained high diagnostic accuracy, while other hub genes demonstrated moderate predictive performance.

These results suggest that the identified hub genes may serve as potential diagnostic markers for both Endometriosis and RM. Across both datasets, group differences corresponded to moderate-to-large effect sizes (Cohen's *d* range: 0.65–1.25), with 95% confidence intervals excluding zero for all significant comparisons, supporting the robustness and clinical relevance of these findings.

### Immune cell-related gene set enrichment analysis (ssGSEA)

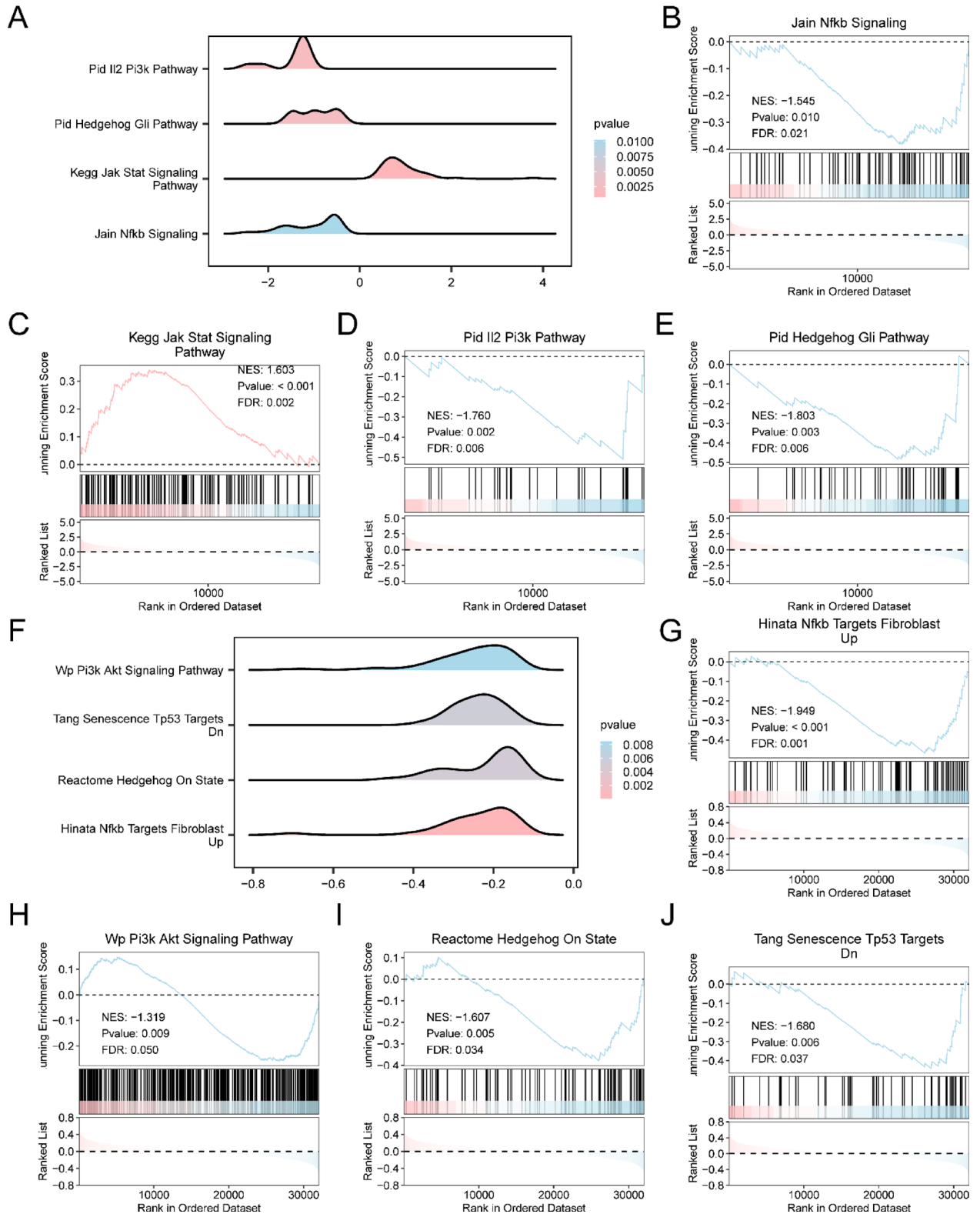
The single-sample Gene Set Enrichment Analysis (ssGSEA) algorithm was used to calculate enrichment scores for 28 immune cell-related gene sets in both the GSE120103 and GSE165004 datasets. Group differences in enrichment scores, as well as correlations among immune cell types and between key gene expression and immune cell-related enrichment scores, were systematically assessed.

The analysis revealed altered enrichment score patterns across both datasets, which may indicate potential involvement of immune dysregulation in the pathogenesis of Endometriosis and Recurrent Miscarriage (RM). Notably, several immune cell-related gene sets demonstrated distinct enrichment profiles and correlation patterns with hub genes, indicating a possible link between immune microenvironment remodeling and disease-related transcriptional changes.

These findings were visualized through heatmaps and correlation plots, highlighting distinct immune landscapes and their associations with key gene expression (Fig. 8A–F).

### Identification of hub genes and construction and evaluation of predictive models

To identify hub genes associated with endometriosis and recurrent miscarriage, SVM-RFE (Support Vector Machine-Recursive Feature Elimination) analysis was conducted on datasets GSE120103 and GSE165004 to screen for key gene signatures using the R caret package with an RBF kernel and tenfold cross-validation, with optimal features selected based on minimal RMSE from cross-validation results. Note that the set of differentially expressed genes identified from the initial analysis is not identical to the final set of key genes selected by machine learning. This discrepancy arises from differences in sample distribution, feature selection criteria, and algorithm-specific characteristics. This analysis identified 4 gene signatures from each dataset: *ITGB1*, *VIM*, *TUBB3*, and *NR4A1* from GSE120103, and *FGF2*, *SMAD1*, *VIM*, and *ITGB1* from GSE165004. Subsequently, the intersection of the gene signatures from the two datasets was analyzed, revealing *ITGB1* and *VIM* as hub genes (Fig. S2).



**Fig. 4.** Gene Set Enrichment Analysis (GSEA) results. **(A)** GSEA enrichment score plot for GSE120103 dataset. **(B–E)** GSEA enrichment plots for Jain NF-κB signaling, JAK-STAT pathway, PI3K/AKT pathway, and Hedgehog signaling pathway in GSE120103. **(F)** GSEA enrichment score plot for GSE165004 dataset. **(G–J)** GSEA enrichment plots for Hinata NF-κB fibroblast targets, WP PI3K/AKT pathway, Hedgehog pathway, and TP53 senescence targets in GSE165004.

ID	Set size	Enrichment score	NES	p value	p.adjust	q value
JAIN_NFKB_SIGNALING	68	-0.382343356	-1.544609704	0.010251295	0.033618802	0.020763553
KEGG_JAK_STAT_SIGNALING_PATHWAY	140	0.340767515	1.602678976	0.000780256	0.003818163	0.002358164
PID_IL2_PI3K_PATHWAY	33	-0.509324969	-1.759753451	0.002197524	0.009268324	0.005724277
PID_HEDGEHOG_GLI_PATHWAY	45	-0.483128673	-1.802659076	0.002506912	0.010381525	0.006411809

**Table 3.** Results of GSEA for datasets GSE120103. GSEA: gene set enrichment analysis.

ID	Set size	Enrichment score	NES	p value	p.adjust	q value
HINATA_NFKB_TARGETS_FIBROBLAST_UP	79	-0.468088924	-1.94867302	7.23E-05	0.001719703	0.001367636
WP_PI3K_AKT_SIGNALING_PATHWAY	332	-0.257958607	-1.319358604	0.008512333	0.062312775	0.049555777
REACTOME_HEDGEHOG_ON_STATE	84	-0.378733728	-1.607380387	0.00499314	0.042773855	0.034016967
TANG_SENESCENCE_TP53_TARGETS_DN	54	-0.440380814	-1.679957431	0.005519669	0.046291007	0.036814069

**Table 4.** Results of GSEA for datasets GSE165004. GSEA, gene set enrichment analysis.

These genes are not only associated with both endometriosis and recurrent miscarriage but also serve as critical link genes between the two conditions, underscoring their significant biological importance. Subsequently, we constructed predictive nomograms based on the key genes (Fig. 9A,D) and evaluated their calibration and discrimination. Calibration curves, generated from 1,000 bootstrap resamples, demonstrated excellent agreement between predicted and observed probabilities, suggesting minimal overfitting risk. ROC curves (Fig. 9 B–C, E–F) showed high discriminative ability, with the AUC for the nomogram predicting endometriosis being 0.925 (95% CI: 0.89–1.00) and for recurrent miscarriage being 0.951 (95% CI: 0.85–0.99). Both nomograms exhibited higher AUC values compared to those of ITGB1 or VIM alone, indicating promising predictive performance and potential clinical utility. The Hosmer–Lemeshow test yielded p-values greater than 0.05 for both models, indicating good model fit. The expression differences and ROC analyses for the key genes in GSE165004 are shown in Fig. 9.

## Discussion

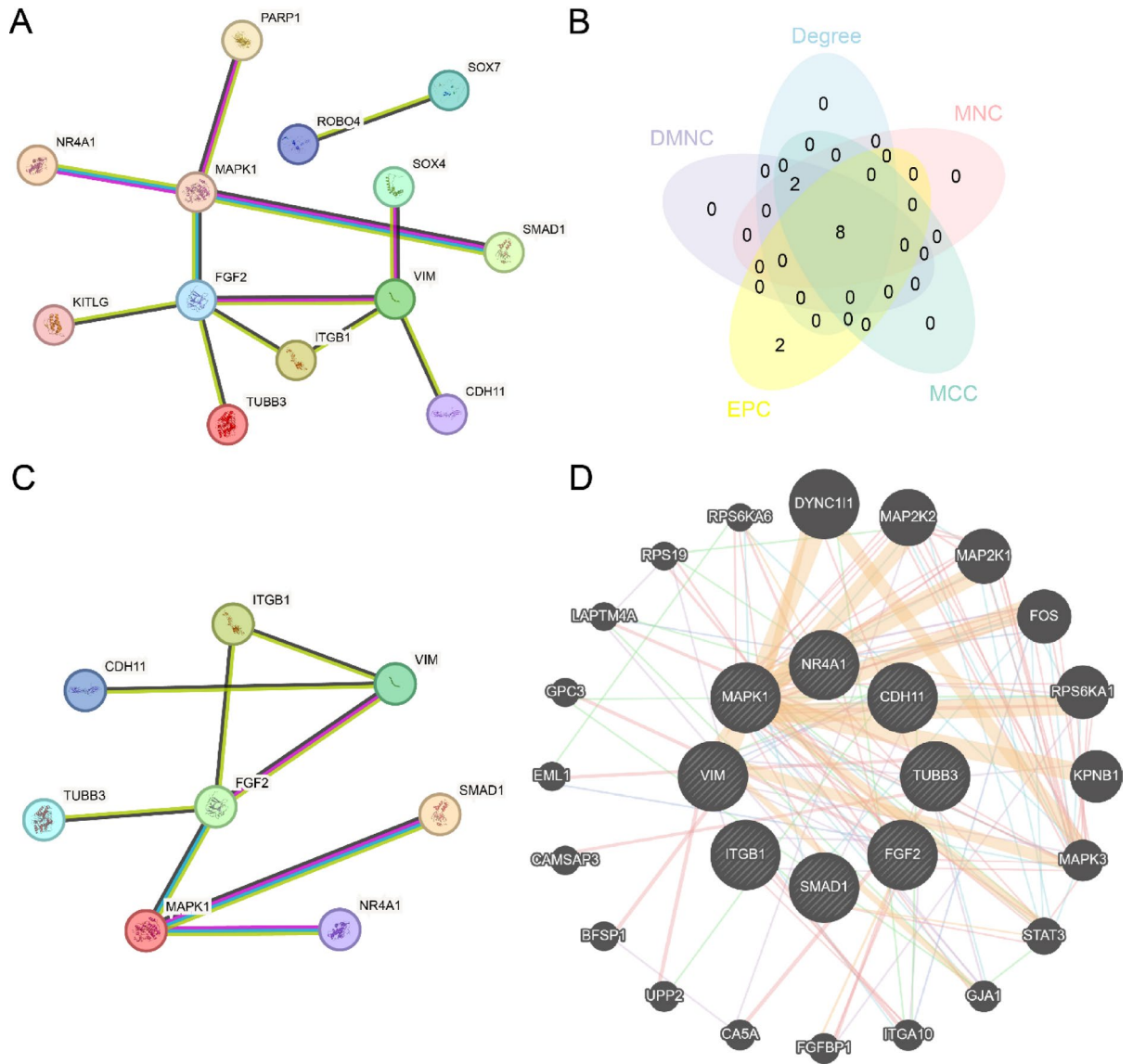
Endometriosis (EMS) and recurrent miscarriage (RM) are two prevalent gynecological disorders that significantly impact women's reproductive health, often resulting in infertility or recurrent pregnancy loss. EMS is characterized by the ectopic growth of endometrial tissue outside the uterine cavity, accompanied by chronic inflammation, cyclical bleeding, and fibrosis<sup>42</sup>. RM typically occurs during early pregnancy and is associated with genetic, hormonal, immunological, and anatomical abnormalities, ultimately impairing women's fertility<sup>43</sup>. Importantly, both conditions involve pathological processes such as aberrant angiogenesis and immune dysregulation, mechanisms closely linked to endothelial-mesenchymal transition (EndMT).

To provide a comprehensive overview of shared molecular mechanisms in EMS and RM, we integrated differential expression analysis, functional enrichment, network construction, and immune profiling. We identified a core set of EndMT-related DEGs involved in angiogenesis, immune modulation, and cell migration, with functional enrichment analyses highlighting pathways such as PI3K-Akt, MAPK, and Rap1 signaling. PPI network analysis revealed hub genes like FGF2, ITGB1, and VIM, which were associated with diagnostic potential in our bioinformatics analysis.

Among these, FGF2 has been shown to promote EndMT through the activation of the TGF- $\beta$ /SMAD signaling pathway, thereby contributing to endothelial plasticity and fibrotic transformation in multiple disease contexts<sup>44–46</sup>. VIM (vimentin), a canonical mesenchymal cytoskeletal protein, serves as a robust marker of EndMT and is actively involved in cytoskeletal remodeling and cellular migration during transition processes<sup>47,48</sup>. ITGB1 (Integrin  $\beta$ 1) mediates endothelial adhesion to the extracellular matrix and activates signaling cascades such as FAK and AKT, facilitating EndMT and matrix remodeling, especially in fibrotic and inflammatory environments<sup>49,50</sup>. Together, these genes may represent core regulatory nodes that bridge vascular remodeling and immune dysregulation, reinforcing their pathophysiological relevance in both endometriosis and recurrent miscarriage.

These findings, together with distinct immune-related enrichment score profiles, are consistent with the hypothesis that EndMT-associated dysregulation of vascular and immune pathways may act as a convergent mechanism contributing to the pathogenesis of both diseases. This integrated perspective provides a foundation for future mechanistic studies and potential therapeutic targeting.

Immune dysregulation emerged as a convergent mechanism linking EMS and RM<sup>51,52</sup>. Significant alterations were observed in enrichment scores of key immune cell-related gene sets, including  $\gamma\delta$  T cells, monocytes, natural killer (NK) cells, regulatory T cells (Tregs), follicular helper T cells (Tfh), B cell subsets, dendritic cells, eosinophils, mast cells, and Th2 cells. A positive correlation between  $\gamma\delta$  T cells and monocytes suggests coordinated pro-inflammatory activities, whereas an inverse relationship between MAPK1 expression and mast cell-related enrichment scores may indicate feedback regulation. These immune alterations collectively suggest a systemic imbalance, characterized by enhanced humoral responses, impaired maternal–fetal tolerance, and



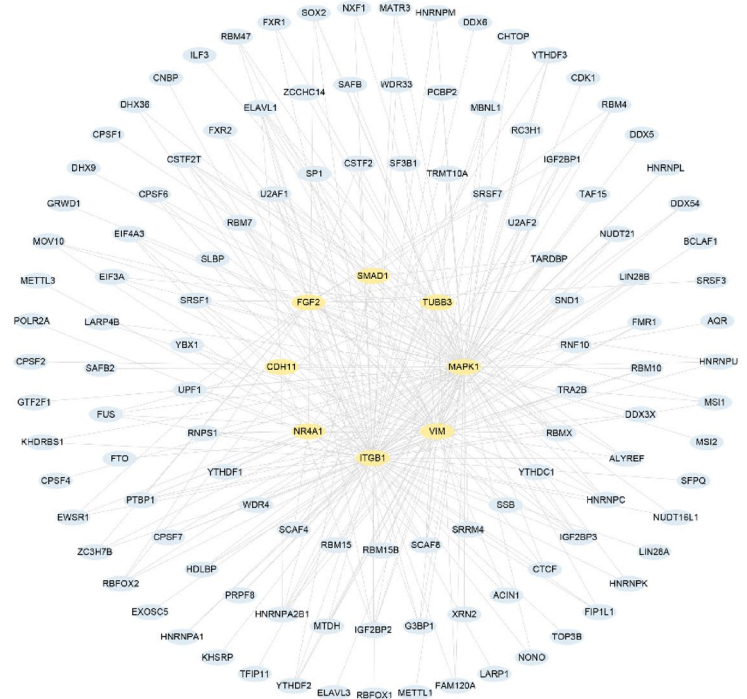
**Fig. 5.** Protein–protein interaction (PPI) network of EndMTRDEGs and key gene identification. **(A)** PPI network of 13 EndMTRDEGs constructed using the STRING database. **(B)** Venn diagram showing overlap of top hub genes identified by cytoHubba algorithms. **(C)** Refined PPI network for eight selected key genes. **(D)** GeneMANIA-predicted network of functionally similar genes.

chronic inflammation<sup>53–58</sup>. Such immune disturbances likely exacerbate ectopic lesion formation in EMS and disrupt implantation processes in RM, contributing to shared pathogenic pathways.

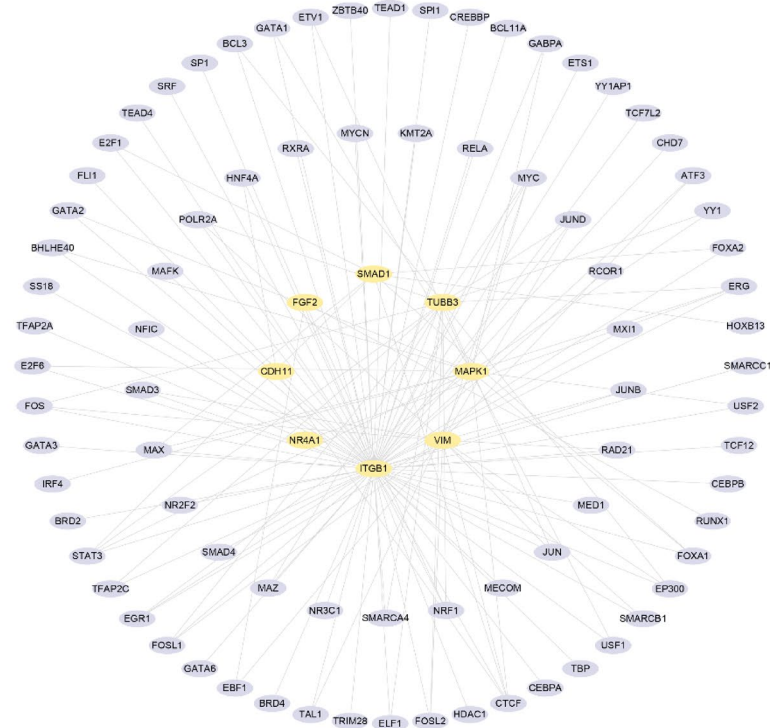
While this study provides novel insights into the molecular and immunological landscapes of EMS and RM, certain limitations should be acknowledged. The relatively small sample sizes and the retrospective, cross-sectional design may limit the generalizability of findings. Although we conducted internal validation through systematic bioinformatics analyses, the lack of external dataset validation and experimental evidence remains a limitation. The functional roles of the identified hub genes and pathways in the pathogenesis of EMS and RM require further investigation. Future studies should incorporate independent external cohorts, in vitro and in vivo experiments, and advanced techniques such as single-cell sequencing and spatial transcriptomics to enhance the reliability and biological interpretability of the findings.

In conclusion, this study identified a set of shared EndMT-related gene signatures through integrative transcriptomic analysis, highlighting common molecular mechanisms underlying endometriosis and recurrent miscarriage. These findings provide a foundational framework for understanding the overlapping pathogenic processes of angiogenesis, immune dysregulation, and EndMT, offering candidate targets for future translational research and precision therapies.

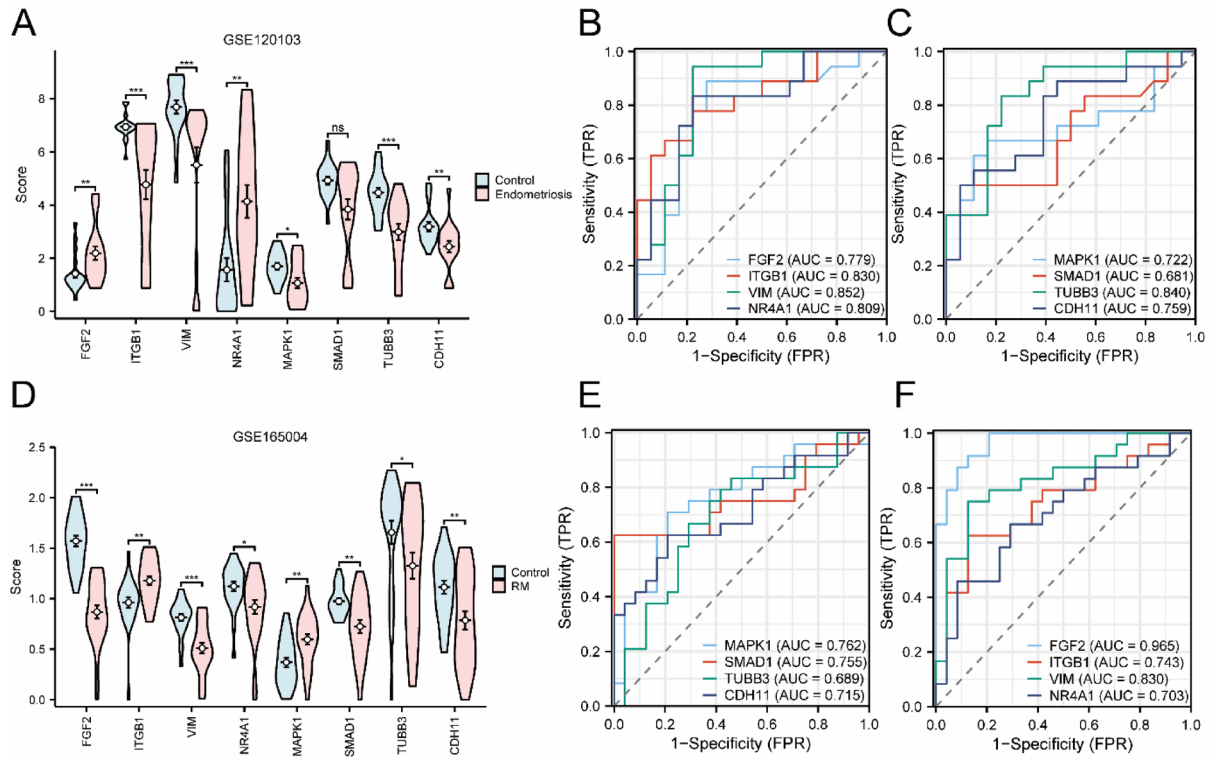
A



B

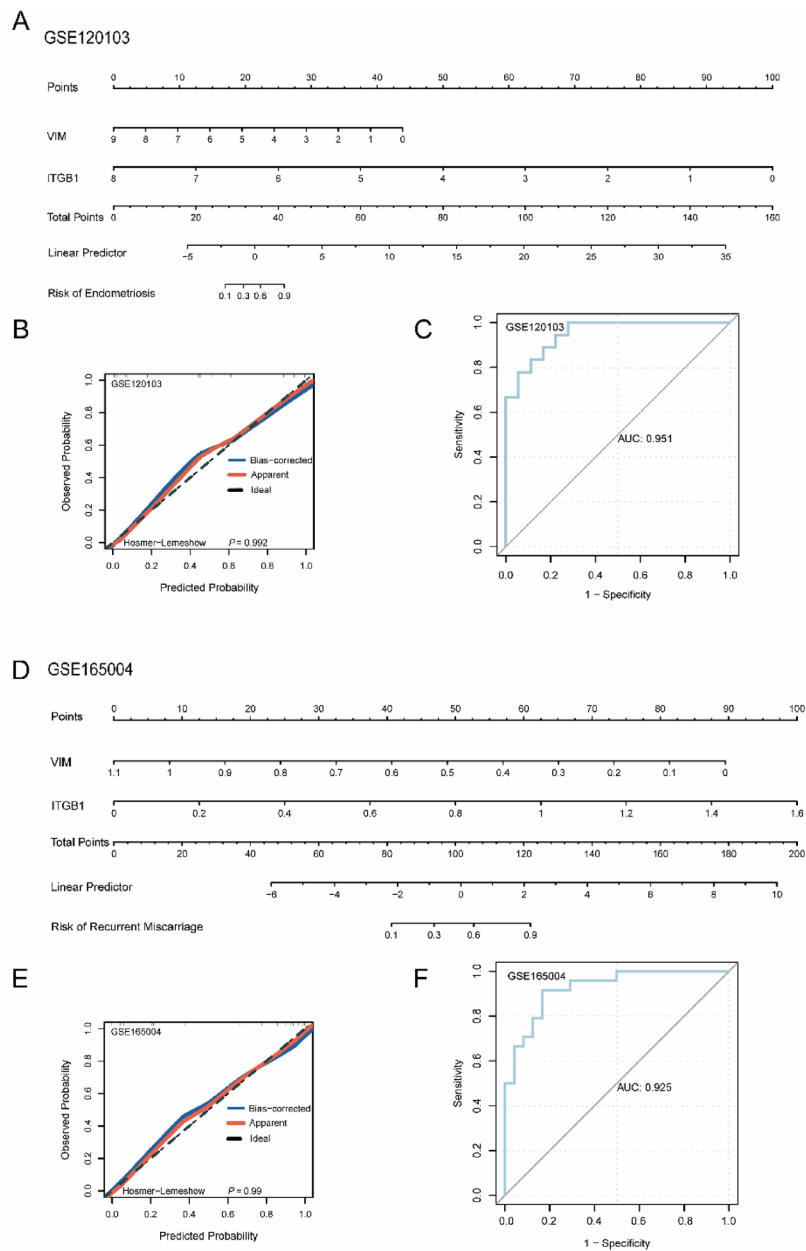


**Fig. 6.** Regulatory networks of key genes. (A) mRNA–RNA-binding protein (RBP) interaction network based on StarBase predictions. (B) mRNA–transcription factor (TF) regulatory network based on ChIPBase data.



**Fig. 7.** Validation of key gene expression levels and diagnostic evaluation. (A, D) Violin plots comparing expression levels of key genes between disease and control groups in GSE120103 and GSE165004. (B–C, E–F) Receiver operating characteristic (ROC) curves demonstrating diagnostic value, with area under the curve (AUC) values indicated.





**Fig. 9.** Nomogram models for diagnosis of EMs and RM. (**A, D**) Nomogram models constructed using key genes to predict risk scores for EMs and RM. (**B, E**) Calibration plots assessing model prediction accuracy. (**C, F**) ROC curves evaluating the diagnostic performance of nomogram models.

### Data availability

The transcriptomic datasets analyzed during this study are publicly available in the Gene Expression Omnibus (GEO) repository: GSE120103 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE120103>) GSE165004 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE165004>) Selected source code and processed data files used in this study are provided as supplementary material (Supplementary\_Code\_and\_Data.zip) for transparency and reproducibility. This includes scripts for data cleaning, differential gene expression, enrichment analyses (GO/KEGG/GSEA), PPI construction and immune infiltration profiling. Additional processed data and scripts are available upon reasonable request from the corresponding author.

Received: 5 May 2025; Accepted: 15 September 2025

Published online: 21 October 2025

### References

- Zondervan, K. T., Becker, C. M. & Missmer, S. A. Endometriosis. *N. Engl. J. Med.* **382**(13), 1244–1256. <https://doi.org/10.1056/NEJMra1810764> (2020).

2. Dimitriadis, E., Menkhorst, E., Saito, S., Kutteh, W. H. & Brosens, J. J. Recurrent pregnancy loss. *Nat. Rev. Dis. Primers* **6**(1), 98. <https://doi.org/10.1038/s41572-020-00228-z> (2020).
3. Allaire, C., Bedaiwy, M. A. & Yong, P. J. Diagnosis and management of endometriosis. *CMAJ* **195**(10), E363–E371. <https://doi.org/10.1503/cmaj.220637> (2023).
4. Bender Atik, R. et al. ESHRE guideline: Recurrent pregnancy loss: An update in 2022. *Hum. Reprod. Open* **2023**(1), 2. <https://doi.org/10.1093/hropen/hoad002> (2023).
5. Marquardt, R. M., Tran, D. N., Lessey, B. A., Rahman, M. S. & Jeong, J. W. Epigenetic dysregulation in endometriosis: Implications for pathophysiology and therapeutics. *Endocr. Rev.* **44**(6), 1074–1095. <https://doi.org/10.1210/edrv/bnad020> (2023).
6. Deng, T., Liao, X. & Zhu, S. Recent advances in treatment of recurrent spontaneous abortion. *Obstet. Gynecol. Surv.* **77**(6), 355–366. <https://doi.org/10.1097/ogx.0000000000001033> (2022).
7. Cafun, R. G. et al. The low prevalence and high cost of routine antiphospholipid antibody testing in patients with recurrent pregnancy loss. *Fertil. Steril.* **119**(6), 1078–1080. <https://doi.org/10.1016/j.fertnstert.2023.03.026> (2023).
8. Banjar, S. et al. Intravenous immunoglobulin use in patients with unexplained recurrent pregnancy loss. *Am. J. Reprod. Immunol.* **90**(2), e13737. <https://doi.org/10.1111/aji.13737> (2023).
9. Della Corte, L. et al. The burden of endometriosis on women's lifespan: A narrative overview on quality of life and psychosocial wellbeing. *Int. J. Environ. Res. Public Health* <https://doi.org/10.3390/ijerph17134683> (2020).
10. Quenby, S. et al. Miscarriage matters: The epidemiological, physical, psychological, and economic costs of early pregnancy loss. *Lancet* **397**(10285), 1658–1667. [https://doi.org/10.1016/s0140-6736\(21\)00682-6](https://doi.org/10.1016/s0140-6736(21)00682-6) (2021).
11. Kalluri, R. & Weinberg, R. A. The basics of epithelial-mesenchymal transition. *J. Clin. Invest.* **119**(6), 1420–1428. <https://doi.org/10.1172/jci39104> (2009).
12. Yang, J. & Weinberg, R. A. Epithelial-mesenchymal transition: At the crossroads of development and tumor metastasis. *Dev. Cell* **14**(6), 818–829. <https://doi.org/10.1016/j.devcel.2008.05.009> (2008).
13. Kovacic, J. C. et al. Endothelial to mesenchymal transition in cardiovascular disease: JACC state-of-the-art review. *J. Am. Coll. Cardiol.* **73**(2), 190–209. <https://doi.org/10.1016/j.jacc.2018.09.089> (2019).
14. Piera-Velazquez, S. & Jimenez, S. A. Endothelial to mesenchymal transition: Role in physiology and in the pathogenesis of human diseases. *Physiol. Rev.* **99**(2), 1281–1324. <https://doi.org/10.1152/physrev.00021.2018> (2019).
15. Zhang, Z., Fang, Z., Ge, J. & Li, H. Endothelial-to-mesenchymal transition in cardiovascular diseases. *Trends Mol. Med.* <https://doi.org/10.1016/j.molmed.2025.05.002> (2025).
16. Sun, X., Nkenkor, B., Mastikhina, O., Soon, K. & Nunes, S. S. Endothelium-mediated contributions to fibrosis. *Semin. Cell Dev. Biol.* **101**, 78–86. <https://doi.org/10.1016/j.semcdb.2019.10.015> (2020).
17. Liang, G. et al. Tenascin-X mediates flow-induced suppression of EndMT and atherosclerosis. *Circ. Res.* **130**(11), 1647–1659. <https://doi.org/10.1161/circresaha.121.320694> (2022).
18. Xu, C. et al. EndMT: New findings on the origin of myofibroblasts in endometrial fibrosis of intrauterine adhesions. *Reprod. Biol. Endocrinol.* **20**(1), 9. <https://doi.org/10.1186/s12958-022-00887-5> (2022).
19. Gui, Z. et al. Macrophage polarization induces endothelium-to-myofibroblast transition in chronic allograft dysfunction. *Ren. Fail.* **45**(1), 2220418. <https://doi.org/10.1080/0886022x.2023.2220418> (2023).
20. Shu, D. Y., Butcher, E. & Saint-Geniez, M. EMT and EndMT: Emerging roles in age-related macular degeneration. *Int. J. Mol. Sci.* <https://doi.org/10.3390/ijms21124271> (2020).
21. Davis, S. & Meltzer, P. S. GEOquery: A bridge between the gene expression omnibus (GEO) and bioconductor. *Bioinformatics* **23**(14), 1846–1847. <https://doi.org/10.1093/bioinformatics/btm254> (2007).
22. Clough, E. et al. NCBI GEO: Archive for gene expression and epigenomics data sets: 23-year update. *Nucleic Acids Res.* **52**(D1), D138–D144. <https://doi.org/10.1093/nar/gkad965> (2024).
23. Fishilevich, S. et al. GeneHancer: Genome-wide integration of enhancers and target genes in GeneCards. *Database (Oxford)* <https://doi.org/10.1093/database/bax028> (2017).
24. Bhat, M. A., Sharma, J. B., Roy, K. K., Sengupta, J. & Ghosh, D. Genomic evidence of Y chromosome microchimerism in the endometrium during endometriosis and in cases of infertility. *Reprod. Biol. Endocrinol.* **17**(1), 22. <https://doi.org/10.1186/s12958-019-0465-z> (2019).
25. Keleş, I. D. et al. Gene pathway analysis of the endometrium at the start of the window of implantation in women with unexplained infertility and unexplained recurrent pregnancy loss: is unexplained recurrent pregnancy loss a subset of unexplained infertility?. *Hum. Fertil. (Camb)* **26**(5), 1129–1141. <https://doi.org/10.1080/14647273.2022.2143299> (2023).
26. Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M. & Tanabe, M. KEGG: Integrating viruses and cellular organisms. *Nucleic Acids Res.* **49**(D1), D545–D551. <https://doi.org/10.1093/nar/gkaa970> (2021).
27. Mi, H. et al. PANTHER version 16: A revised family classification, tree-based classification tool, enhancer regions and extensive API. *Nucleic Acids Res.* **49**(D1), D394–d403. <https://doi.org/10.1093/nar/gkaa1106> (2021).
28. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**(1), 27–30. <https://doi.org/10.1093/nar/28.1.27> (2000).
29. Kanehisa, M. Toward understanding the origin and evolution of cellular organisms. *Protein Sci.* **28**(11), 1947–1951. <https://doi.org/10.1002/pro.3715> (2019).
30. Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M. & Ishiguro-Watanabe, M. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res.* **51**(D1), D587–D592. <https://doi.org/10.1093/nar/gkac963> (2023).
31. Yu, G. Thirteen years of cluster profiler. *Innovation (Camb)* **5**(6), 100722. <https://doi.org/10.1016/j.xinn.2024.100722> (2024).
32. de Jong, A., Kuipers, O. P. & Kok, J. FUNAGE-Pro: Comprehensive web server for gene set enrichment analysis of prokaryotes. *Nucleic Acids Res.* **50**(W1), W330–W336. <https://doi.org/10.1093/nar/gkac441> (2022).
33. Szklarczyk, D. et al. The STRING database in 2023: Protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res.* **51**(D1), D638–D646. <https://doi.org/10.1093/nar/gkac1000> (2023).
34. Majeed, A. & Mukhtar, S. Protein–protein interaction network exploration using cytoscape. *Methods Mol. Biol.* **2690**, 419–427. [https://doi.org/10.1007/978-1-0716-3327-4\\_32](https://doi.org/10.1007/978-1-0716-3327-4_32) (2023).
35. Chin, C. H. et al. cytoHubba: Identifying hub objects and sub-networks from complex interactome. *BMC Syst. Biol.* **8**(4), S11. <https://doi.org/10.1186/1752-0509-8-4-s11> (2014).
36. Franz, M. et al. GeneMANIA update 2018. *Nucleic Acids Res.* **46**(W1), W60–W64. <https://doi.org/10.1093/nar/gky311> (2018).
37. Singh, A. RNA-binding protein kinetics. *Nat. Methods* **18**(4), 335. <https://doi.org/10.1038/s41592-021-01122-6> (2021).
38. Li, J. H., Liu, S., Zhou, H., Qu, L. H. & Yang, J. H. starBase v2.0: Decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res.* **42**, D92–D97. <https://doi.org/10.1093/nar/gkt1248> (2014).
39. Huang, J. et al. ChIPBase v3.0: The encyclopedia of transcriptional regulations of non-coding RNAs and protein-coding genes. *Nucleic Acids Res.* **51**(D1), D46–d56. <https://doi.org/10.1093/nar/gkac1067> (2023).
40. Nahm, F. S. Receiver operating characteristic curve: Overview and practical use for clinicians. *Korean J. Anesthesiol.* **75**(1), 25–36. <https://doi.org/10.4097/kja.21209> (2022).
41. Xiao, B. et al. Identification and verification of immune-related gene prognostic signature based on ssGSEA for osteosarcoma. *Front. Oncol.* **10**, 607622. <https://doi.org/10.3389/fonc.2020.607622> (2020).
42. Kvaskoff, M. et al. Endometriosis and cancer: A systematic review and meta-analysis. *Hum. Reprod.* **27**(2), 393–420. <https://doi.org/10.1093/humupd/dmaa045> (2021).

43. de Ziegler, D. & Frydman, R. F. Recurrent pregnancy losses, a lasting cause of infertility. *Fertil. Steril.* **115**(3), 531–532. <https://doi.org/10.1016/j.fertnstert.2020.12.004> (2021).
44. Obora, K. et al. Inflammation-induced miRNA-155 inhibits self-renewal of neural stem cells via suppression of CCAAT/enhancer binding protein  $\beta$  (C/EBP $\beta$ ) expression. *Sci. Rep.* **7**, 43604. <https://doi.org/10.1038/srep43604> (2017).
45. de Araújo, P. A., Maciel-Honda, P. O., de Oliveira, C.-F., Dos Santos, G. G. & Martins, M. L. Efficacy of chlorine, sodium chloride and trichlorfon baths against monogenean *Dawestrema cycloancistrum* parasite of pirarucu *Arapaima gigas*. *J. Fish Dis.* **46**(2), 113–126. <https://doi.org/10.1111/jfd.13725> (2023).
46. Lee, C. et al. VEGF-B prevents excessive angiogenesis by inhibiting FGF2/FGFR1 pathway. *Signal Transduct. Target. Ther.* **8**(1), 305. <https://doi.org/10.1038/s41392-023-01539-9> (2023).
47. Kizawa, A., Kawahara, A., Takimura, Y., Nishiyama, Y. & Hihara, Y. RNA-seq profiling reveals novel target genes of LexA in the cyanobacterium *Synechocystis* sp. PCC 6803. *Front Microbiol.* **7**, 193. <https://doi.org/10.3389/fmicb.2016.00193> (2016).
48. Zhang, X. et al. Vimentin protein in situ expression predicts less tumor metastasis and overall better survival of endometrial carcinoma. *Dis. Mark.* **2022**, 5240046. <https://doi.org/10.1155/2022/5240046> (2022).
49. Shi, D. et al. Pharmacokinetic analysis of 6-O-[(18)F]FEE for PET imaging of EGFR mutation. *Bioorg. Med. Chem. Lett.* **85**, 129217. <https://doi.org/10.1016/j.bmcl.2023.129217> (2023).
50. Zhilan, T. et al. Salidroside promotes pro-angiogenesis and repair of blood brain barrier via Notch/ITGB1 signal path in CSVD Model. *J. Adv. Res.* **68**, 429–444. <https://doi.org/10.1016/j.jare.2024.02.019> (2025).
51. Shen, H. H. et al. Ovarian hormones-autophagy-immunity axis in menstruation and endometriosis. *Theranostics* **11**(7), 3512–3526. <https://doi.org/10.7150/thno.55241> (2021).
52. Alecsandru, D., Klimczak, A. M., Garcia Velasco, J. A., Pirtea, P. & Franasiak, J. M. Immunologic causes and thrombophilia in recurrent pregnancy loss. *Fertil. Steril.* **115**(3), 561–566. <https://doi.org/10.1016/j.fertnstert.2021.01.017> (2021).
53. Li, L., Liao, Z., Ye, M. & Jiang, J. Recombinant human IL-37 inhibited endometriosis development in a mouse model through increasing Th1/Th2 ratio by inducing the maturation of dendritic cells. *Reprod. Biol. Endocrinol.* **19**(1), 128. <https://doi.org/10.1186/s12958-021-00811-3> (2021).
54. Yang, X., Tian, Y., Zheng, L., Luu, T. & Kwak-Kim, J. The update immune-regulatory role of pro- and anti-inflammatory cytokines in recurrent pregnancy losses. *Int. J. Mol. Sci.* <https://doi.org/10.3390/ijms24010132> (2022).
55. Li, Z. et al. The role of thymus- and extrathymus-derived regulatory T cells in maternal-fetal tolerance. *Front Immunol.* **14**, 1109352. <https://doi.org/10.3389/fimmu.2023.1109352> (2023).
56. Wei, P. et al. Identification and validation of a signature based on macrophage cell marker genes to predict recurrent miscarriage by integrated analysis of single-cell and bulk RNA-sequencing. *Front Immunol.* **13**, 1053819. <https://doi.org/10.3389/fimmu.2022.1053819> (2022).
57. Wang, W., Sung, N., Gilman-Sachs, A. & Kwak-Kim, J. T helper (Th) cell profiles in pregnancy and recurrent pregnancy losses: Th1/Th2/Th9/Th17/Th22/Tfh cells. *Front Immunol.* **11**, 2025. <https://doi.org/10.3389/fimmu.2020.02025> (2020).
58. Abramiuk, M. et al. The role of the immune system in the development of endometriosis. *Cells* <https://doi.org/10.3390/cells11132028> (2022).

### Author contributions

Yue Liang and Liangcheng Yu contributed equally to this work and share first authorship. Yue Liang and Liangcheng Yu conceived study design, data acquisition, comprehensive bioinformatics analysis, and manuscript drafting. Danjie Su and Lu Wang provided technical supervision and guidance on data analysis strategies. Qingde Zhou contributed to manuscript editing, language refinement, and formatting. Haihui Wang participated in data retrieval and preprocessing from public databases. Cong Li and Jingjing Wang assisted in data processing and visualization tasks. Jie Dong, Xifeng Xiao, and Xiaohong Wang supervised the study and critically revised the manuscript. Xiaohong Wang is the primary corresponding author. Jie Dong and Xifeng Xiao are additional corresponding authors. All authors read and approved the final manuscript.

### Funding

This work was supported by the National Natural Science Foundation of China (Grant Nos. 82271734 and 82101794).

### Declarations

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-20302-4>.

**Correspondence** and requests for materials should be addressed to J.D., X.X. or X.W.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025