

Integration of Single Cell and Bulk RNA-Sequencing Reveals Key Genes and Immune Cell Infiltration to Construct a Predictive Model and Identify Drug Targets in Endometriosis

Hanke Zhang, Yuqing Fang, Dan Luo, Yan-Hui Li

Department of Obstetrics and Gynecology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, 430022, People's Republic of China

Correspondence: Yan-Hui Li, Department of Obstetrics and Gynecology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, 430022, People's Republic of China, Tel +86-13407159190, Email liyanhui251@163.com

Purpose: Endometriosis is a common chronic neuroinflammatory disease with a poorly understood pathogenesis. Molecular changes and specific immune cell infiltration in the eutopic endometrium are critical to disease progression. This study aims to explore immune mechanisms and molecular differences in the proliferative eutopic endometrium of endometriosis by integrating bulk RNA-seq and single-cell RNA sequencing (scRNA-seq) data, and to develop diagnostic and predictive models for the disease.

Methods: Gene expression profiles from the proliferative endometrium of endometriosis patients and healthy controls were obtained from the Gene Expression Omnibus. Single-cell RNA-seq data were processed using R packages, and cell clusters' contributions to endometriosis were calculated. Differentially expressed genes (DEGs) from bulk RNA-seq were intersected with significant mesenchymal cell genes from scRNA-seq, and a predictive model was constructed using LASSO analysis. Key gene mechanisms were explored through Gene Set Enrichment and Variation Analyses. miRNA networks and transcriptional regulation analyses were conducted, and potential drugs were predicted using the Connectivity Map database. RT-qPCR validated key gene expression.

Results: Mesenchymal cells in the proliferative eutopic endometrium were identified as major contributors to endometriosis pathogenesis. LASSO regression identified eight key genes: SYNE2, TXN, NUPR1, CTSK, GSN, MGP, IER2, and CXCL12. The predictive model based on these genes achieved AUC values of 1.00 and 0.8125 in training and validation cohorts. Immune infiltration analysis showed increased CD8⁺ T cells and monocytes in the eutopic endometrium of endometriosis patients. Drug target prediction indicated that drugs like Retinol, Orantinib, Piperacillin, and NECA were negatively correlated with the expression profiles of endometriosis. RT-qPCR validated gene expression in patients aligned with bioinformatics analysis.

Conclusion: Significant transcriptomic changes and altered immune cell infiltration in the proliferative eutopic endometrium potentially contribute to endometriosis pathogenesis. Our predictive model based on the key genes demonstrates high diagnostic accuracy, offering insights for diagnosis and potential treatment strategies.

Keywords: endometriosis, eutopic endometrium, transcriptomics, single-cell sequencing, predictive model, immune cell infiltration

Introduction

Endometriosis is a chronic neuroinflammatory disease strongly associated with chronic pelvic pain in women. It is estimated to affect 6–10% of women of reproductive age,¹ with its incidence rising to 50% among infertile women.² Additionally, epidemiological research indicates that women with endometriosis have a higher likelihood of developing ovarian cancer,³ breast cancer, melanoma, asthma, rheumatoid arthritis, and cardiovascular diseases.⁴

The pathogenesis of endometriosis remains incompletely understood. The primary current treatments include surgical excision of lesions and pharmacotherapy to suppress ovarian hormone production.⁵ However, more than half of the patients who undergo surgery require additional procedures within 5.5 years.⁶ Moreover, endometrioma cystectomy may

have long-term and detrimental effects on ovarian reserve.⁷ Regarding pharmacological treatment, current guidelines recommend hormonal therapies, such as combined oral contraceptives, progestins, and GnRH agonists and antagonists.⁸ However, it is important to note that these treatments may cause potential side effects, including bone loss and low estrogen symptoms induced by GnRH analogs.^{9,10} Therefore, further elucidation of the pathogenesis of endometriosis and the development of more effective therapeutic strategies that alleviate symptoms without compromising fertility are critical research priorities.^{5,11}

Various hypotheses, including retrograde menstruation, intracavitary dissemination, and vascular or lymphatic metastasis, have been proposed to explain the mechanisms of endometriosis. However, these theories do not fully capture the diverse pathological subtypes of the disease.¹² Emerging evidence suggests that the cellular and molecular characteristics of the eutopic endometrium in endometriosis patients are pivotal in the disease's pathogenesis. Studies have documented significant alterations in gene and protein expression within this tissue,^{13,14} yet analyses focusing on single markers often fail to capture the full spectrum of regulatory changes. To address this limitation, high-throughput techniques such as microarray and RNA sequencing (RNA-seq) have been employed to provide a comprehensive analysis of transcriptional changes and to identify critical genes and pathways involved in endometriosis.^{15,16} Nonetheless, inconsistencies in identifying differentially expressed genes across RNA-seq studies may arise from variations in sample collection, including differences in menstrual cycle phase, endometriosis subtype, and hormonal treatment backgrounds.^{17,18} Therefore, it is crucial to standardize these variables in RNA-seq analyses to ensure consistent and reliable results.

Additionally, the endometrium is a dynamic, multicellular tissue that responds to steroid hormones. Its epithelial cells are supported by a stromal layer, which is rich in vasculature and populated by a diverse and fluctuating array of immune cells. During menstruation, the luminal layer undergoes breakdown and is subsequently shed.⁵ The shed endometrial cells include epithelial cells, stromal fibroblasts, vascular cells, and various immune cells, particularly neutrophils, monocytes/macrophages, and uterine natural killer (uNK) cells.¹⁹ While all these cell types may potentially contribute to peritoneal lesions, contingent on their ability to survive in the peritoneal cavity and evade immune surveillance, there remains significant debate regarding which specific cell types play a pivotal role in the pathogenesis of endometriosis. Changes in immune cells within the eutopic endometrium (eg, macrophages, NK cells, T cells, B cells, etc.) have garnered significant attention. Although findings regarding immune cell alterations in the eutopic endometrium in endometriosis are not yet fully consistent across studies,^{5,20} existing evidence supports the involvement of immune cell dysfunction in the pathophysiological mechanisms of endometriosis. This dysfunction contributes to the establishment of an unfavorable environment for embryo implantation, the formation of ectopic lesions, fibrosis, and pain.²⁰

Since bulk RNA-seq is typically performed at the tissue or multi-cellular level, the genetic data obtained represents an average from multiple cells or cell types, which overlooks intercellular heterogeneity.²¹ In contrast, single-cell RNA sequencing (scRNA-seq), which has emerged over the past decade, enables the acquisition of genetic information at the single-cell level, revealing the gene structure and expression state of each individual cell. This approach captures the heterogeneity and genetic characteristics of rare cells that traditional methods fail to detect. Additionally, scRNA-seq facilitates the more precise discovery and validation of disease-related biomarkers, particularly those that may be diluted or obscured in conventional techniques.²¹ The advent of scRNA-seq technologies has also provided unprecedented insights into the unique roles of different cell populations in the development of endometriosis.²² By employing scRNA-seq, we can conduct an in-depth analysis of the transcriptomic profiles of individual cell types, thereby enhancing our understanding of their interactions and their collective contribution to the progression of endometriosis.^{23,24}

In this study, we systematically downloaded and analyzed Bulk RNA-seq and scRNA-seq data from the Gene Expression Omnibus (GEO) database, focusing on proliferative phase endometrial samples from endometriosis patients and healthy controls. We developed a predictive model to assess the risk of endometriosis onset, which was validated using external cohorts to confirm its predictive accuracy. The model integrates a combination of genetic biomarkers capable of predicting the development of endometriosis, with the goal of supporting early diagnosis and guiding personalized treatment and care. Additionally, we characterized the immune cell infiltration landscape within the eutopic endometrium of endometriosis, identified novel regulatory genes strongly associated with endometriosis pathogenesis, and validated these findings through in vitro experiments. Furthermore, we conducted drug target prediction analyses for

potential therapeutic interventions. Collectively, our findings offer novel insights that may enhance the clinical management of endometriosis.

In this study, we systematically downloaded and analyzed Bulk RNA-seq and single-cell RNA-seq data from the Gene Expression Omnibus (GEO) database, focusing on proliferative phase endometrial samples from endometriosis patients and healthy controls. First, through in-depth analysis of scRNA-seq data and the calculation of contributions from various cell subtypes, we identified the key role of mesenchymal cells in the pathogenesis of endometriosis. Next, we developed a predictive model that, for the first time, integrated a combination of eight genes (SYNE2, TXN, NUPR1, CTSK, GSN, MGP, IER2, and CXCL12) to predict the onset of endometriosis, and validated its predictive accuracy using external cohorts. The model integrates key genes capable of predicting the development of endometriosis, aiming to support early diagnosis and guide personalized treatment and care. In addition, we characterized the immune cell infiltration landscape in the eutopic endometrium of endometriosis patients, identified novel regulatory genes closely associated with the pathogenesis of endometriosis, and validated these findings through in vitro experiments. Furthermore, we conducted drug target prediction analyses for potential therapeutic interventions. Overall, this study provides novel insights that could enhance the clinical management of endometriosis and holds significant potential for practical applications.

Materials and Methods

Dataset Source

Single-cell RNA sequencing (scRNA-seq) datasets and bulk RNA-seq datasets were downloaded from the Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/info/datasets.html>). The selection criteria for the datasets were as follows: (1) completeness and quality of sequencing data; (2) comprehensive patient information, with clear documentation of the menstrual cycle phase at the time of sample collection, and exclusion of patients who had used oral contraceptives or GnRH analogs prior to sampling; (3) availability of corresponding in vitro validation for the sequencing data; (4) inclusion of a healthy control group; (5) sufficiently large sample sizes. For scRNA-seq, we selected GSE179640 and GSE213216, which are currently the largest available scRNA datasets for endometriosis, with total sample sizes of 59 and 51, respectively. For bulk RNA-seq, we selected GSE25628 and GSE153739 due to the completeness of their data, comprehensive sample collection information, extensive in vitro validation, and frequent citation in other studies.^{25,26} As previously discussed, dynamic changes in gene expression and immune cell infiltration occur in the eutopic endometrium across different phases of the menstrual cycle.^{5,20} However, many sequencing studies have not accounted for the menstrual cycle phase in their analyses,^{22,27} and there has been limited focus on immune infiltration and gene expression alterations in the proliferative phase of the eutopic endometrium. In this study, we specifically selected sequencing data from the proliferative phase of the eutopic endometrium in the aforementioned datasets to better explore the gene expression profiles and immune cell infiltration during this phase.

The GSE179640 dataset includes scRNA-seq data from the endometrial tissue of 3 healthy controls and the eutopic endometrial tissue of 3 endometriosis patients, while the GSE213216 dataset includes scRNA-seq data from the endometrial tissue of 3 healthy controls and the eutopic endometrial tissue of 7 endometriosis patients, all of which were incorporated into the subsequent analysis. Additionally, the bulk RNA-seq data from the endometrial tissue of 6 healthy controls and the eutopic endometrial tissue of 8 endometriosis patients in the GSE25628 dataset (based on the GPL571 platform), as well as from the endometrial tissue of 3 healthy controls and the eutopic endometrial tissue of 4 endometriosis patients in the GSE153739 dataset (based on the GPL18573), were included in this study for further analysis. A review of the relevant literature confirmed that all eutopic endometrial samples from the above-mentioned patients were in the proliferative phase. Detailed information for the aforementioned datasets was provided in [Table S1](#).

ScRNA-Seq Data Processing

To handle ScRNA-seq data, samples were integrated using the anchors technique in the R package “Seurat”.²⁸ Core cells were obtained by filtering the scRNA-seq data. Cells with significant mitochondrial content (>20%), as well as those with fewer than 200 or more than 6000 genes (filtering out ruptured cells and potential non-singlet cells, respectively), were

removed from each sample. The gene expression of core cells was normalized using a linear regression model, and the top 2000 genes with highly variable properties were identified using analysis of ANOVA. Principal component analysis (PCA) was performed on single-cell samples, with the top 20 principal components (PC) chosen for further analysis, and the Harmony algorithm²⁹ was used for batch effect correction. The top 20 principal components were subjected to overall dimensionality reduction analysis using the uniform manifold approximation and projection (UMAP) algorithm,³⁰ which determined the positional relationships between clusters. Subsequently, The R package “singleR”³¹ was initiated to automatically annotate the clusters using reference data from comprehensive databases such as the Human Primary Cell Atlas, PanglaoDB, and ImmuneCellExpressionData. These databases offer detailed information on human cell types, including those relevant to the endometrium. Following this, manual annotation was performed by cross-referencing the marker genes with the CellMarker database³² and recent literature. Any inconsistencies between the automated and manual annotations were resolved by reviewing the marker gene profiles and adjusting the annotations to ensure biological accuracy. This combined approach enhanced the precision and reliability of the cell type identification in our study.

Contribution of Different Cell Subpopulations to Endometriosis

To assess the contribution of each cell subpopulation to endometriosis, this study evaluated changes in both cell abundance and gene expression. Differential gene expression analysis was first conducted to identify the characteristic genes, focusing on the top 100 genes most highly expressed in the control versus endometriosis groups. The differential expression levels (Fold Change, FC) and percentage proportion (PctProp) of these genes in each cell subpopulation were then quantified. The product of FC and PctProp for each gene within each subpopulation was calculated. To normalize the contribution, the square root of $FC * PctProp$ was computed, considering both gene expression intensity and cell representation. This methodology, commonly used in single-cell transcriptomics studies, allows for quantification of the gene contributions from different cell subpopulations.^{33,34}

Identification and Functional Enrichment Analysis of Differentially Expressed Genes in Eutopic Endometrium of Endometriosis

RNA-seq data from the eutopic endometrium of 6 healthy controls and 8 endometriosis patients in the GSE25628 dataset were processed using the limma tool to compute differentially expressed genes (DEGs) across the groups. The DEGs selection criteria were $P\text{-Value} < 0.05$ and $|\log_2FC| > 1$. The DEGs volcano plots and heatmaps were generated using the ggplot2³⁵ and pheatmap³⁶ programs, respectively.

Construction and Validation of Disease Risk Model

To create a novel disease risk prediction model, the machine learning algorithm Least Absolute Shrinkage and Selection Operator (Lasso)³⁷ was used to identify genes indicative of disease risk. The analysis steps are summarized as follows. The intersection of DEGs from the GSE25628 dataset and mesenchymal cell cluster genes obtained from the above scRNA-seq data analysis was defined as candidate genes. The GSE25628 dataset served as the training set (13 samples), while the GSE153739 dataset was used as the validation set (7 samples). Univariate Cox proportional hazards regression analysis was performed on candidate genes in the training set to identify characteristic genes associated with the risk of developing endometriosis. Variables with p-values < 0.05 were included in the least absolute shrinkage and selection operator (LASSO) regression analysis, performed using the R package “glmnet”³⁸ to reduce the number of genes in the final risk model. The disease risk model was constructed using the formula: $\text{risk score} = \text{gene expression value}_1 \times \beta_1 + \text{gene expression value}_2 \times \beta_2 + \dots + \text{gene expression value}_n \times \beta_n$ (where gene expression value represents the gene expression level, and β represents the corresponding LASSO regression coefficient). The “survROC” package³⁹ was used to plot receiver operating characteristic (ROC) curves to evaluate the performance of the risk score in predicting the occurrence of endometriosis. Furthermore, the validity of the disease risk model was verified using the internal validation set and external datasets GSE25628 and GSE153739, respectively.

Immuno-Infiltration Analysis

The CIBERSORT algorithm⁴⁰ utilizes deconvolution analysis on immune cell subtype expression matrices through support vector regression to differentiate 22 immune cell types based on 547 biomarkers, including T cells, B cells, plasma cells, and myeloid cell subtypes. To evaluate the immune cell infiltration profiles in the endometrium of endometriosis patients, we uploaded the standardized GSE153739 gene expression matrix to the CIBERSORT website (<https://cibersort.stanford.edu/>) and applied the LM22 signature matrix to estimate the relative proportions of 22 immune cell types, retaining only data with a p-value < 0.05. Additionally, we applied immune cell signature gene sets derived from previous studies^{41,42} and performed single sample gene set enrichment analysis (ssGSEA) to calculate the absolute immune cell infiltration scores for each sample in the GSE153739 dataset. The results were presented as bar plots, and immune cell correlations were calculated using the Psych package.

Functional Pathway Analysis in Endometriosis Using Gene Set Enrichment Analysis and Gene Set Variation Analysis

Using the clusterProfiler package (Version 4.12.0), we conducted GSEA enrichment analysis on target gene sets in endometrium samples from both the endometriosis and healthy control groups to investigate the functional pathway differences associated with endometriosis development. Target marker pathway genes related to immune response, cell signaling, and other relevant processes were obtained from the Molecular Signature Database (MSigDB) (<http://www.gsea-msigdb.org/gsea/index.jsp>). Subsequently, Gene set variation analysis (GSVA) enrichment analysis, a nonparametric and unsupervised algorithm, was performed on these target gene sets in the endometrial samples from both groups, and the differences in GSVA enrichment scores between the two groups were analyzed using the R package limma.

MicroRNAs Network Construction

MicroRNAs (miRNA) are small non-coding RNAs that have been shown to regulate gene expression by promoting the degradation of mRNAs or inhibiting the translation of mRNAs.⁴³ We further analyzed the key genes identified in the above analysis to determine if there are significant miRNAs involved in their transcription or degradation. Using the miRcode database (<http://www.mircode.org>), we identified miRNAs related to these key genes (Table S2) and visualized the gene-miRNA network using Cytoscape software (version 3.10.2).

Regulatory Network Analysis of Key Genes

This study used the R package “RcisTarget (version 1.23.1)”⁴⁴ to predict transcription factors for the selected DEGs set identified in the above analysis. All computations were based on motifs, and the normalized enrichment score (NES) for each motif depends on the total number of motifs in the database. In addition to motifs annotated by the source data, further annotations were inferred based on motif similarity and gene sequences. The process of estimating the over-expression of each motif in the gene set consists of two steps. First, the area under the curve (AUC) for each motif-gene set pair is calculated based on the recovery curve generated by ranking the gene set according to the motif. Next, the NES for each motif is determined using the AUC distribution of all motifs within the gene set. We used the RcisTarget.hg19.motifDBs.cisbpOnly.500bp as the motif-gene rankings database.

Connectivity Map Drug Prediction

The Connectivity Map (CMap) is a tool used to analyze the relationship between gene expression profiles and drug responses. It aims to explore the connection between gene expression and drug actions, helping to identify potential therapeutic candidates or elucidate drug mechanisms.^{45,46} In this study, we uploaded 150 upregulated and 150 down-regulated differentially expressed genes (DEGs) from ectopic endometrial tissues of endometriosis patients and healthy controls to CMap (version 1.1.1.43, <https://clue.io>).^{45,46} CMap compares these gene expression profiles with those of over 1.5 million human cell line samples treated with more than 5,000 bioactive compounds, generating connectivity scores. These scores, ranging from -1 to 1, assess the degree of similarity between the query signature and CMap

compounds: positive scores indicate stimulation of the query signature by the compound, while negative scores suggest inhibitory effects. As a result, small molecules with potential therapeutic effects were identified.

Sample Collection and Quantitative Polymerase Chain Reaction (qPCR)

A total of four patients with endometrioma treated by laparoscopic surgery were recruited from the Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, between October 2023 and March 2024. All patients had a pathological diagnosis confirming endometriosis with clinical stages III–IV according to the revised American Fertility Society (rAFS) classification,⁴⁷ with an average age of 30.50 years (age range 25–36 years). Additionally, four non-endometriotic patients who underwent surgery for uterine fibroids, benign ovarian cysts, etc. were included as controls. These control patients had an average age of 32.25 years (age range 28–37 years). All participants had regular menstrual cycles (21–35 days) and had not received hormone therapy for three months prior to surgery. Specimen collection was performed during the proliferative phase of the menstrual cycle. The medical ethics committee of Union Hospital, Tongji Medical College, Huazhong University of Science and Technology (approval no. 2023–0349), and all patients provided written informed consent before surgery.

The collected tissue specimens were used to extract total RNA using TRIzol reagent (Takara, Japan). The concentration and purity of the extracted total RNA were measured using a Nanodrop 2000 spectrophotometer (Thermo Scientific). A total of 1 µg of RNA was used to synthesize the first strand of cDNA in a 20 µL reaction system using the PrimeScript™ RT kit (Takara, Japan). Real-time quantitative PCR was performed using the SYBR Green Master Mix (Vazyme, China) kit in a 25 µL reaction system. Primer sequences are listed in [Table S3](#), with GAPDH used as an internal reference. The relative gene expression levels were calculated using the $2^{-\Delta\Delta CT}$ method.

Statistical Analysis

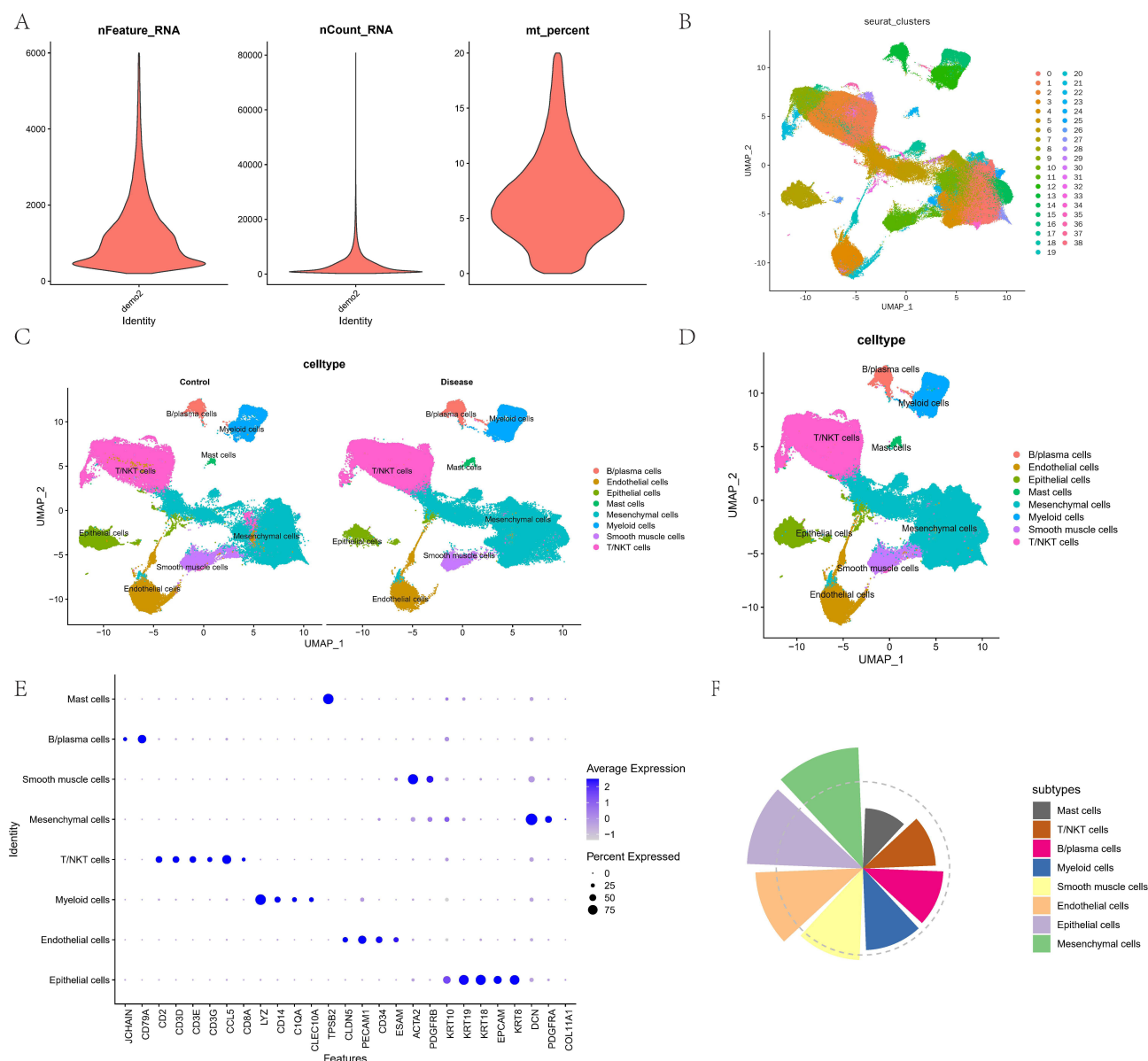
All statistical analyses and presentations were carried out with open-source R (version 4.3.1). The cell proportions between two groups were compared using an unpaired, two-tailed Student's *t*-test. To determine statistical significance in gene expression or gene signatures between two groups, an unpaired, two-tailed Wilcoxon rank-sum test with Bonferroni correction was used. A P-value of less than 0.05 indicated statistical significance.

The Wilcoxon rank-sum test was used to compare gene expression between groups due to the non-normal distribution of the data. This non-parametric test is appropriate for comparing independent groups with skewed data or outliers, ensuring robustness. To control for type I errors from multiple comparisons, Bonferroni correction was applied. However, this stringent approach may reduce statistical power and increase the risk of overlooking biologically or clinically relevant findings. Future research could explore alternative methods, such as false discovery rate (FDR) control, to improve sensitivity while maintaining appropriate error control.

Results

Cell Subpopulation Annotation and Contribution Analysis to Endometriosis in scRNA-Seq Data

The overall schematic outline of the current study is illustrated in [Figure S1](#). We downloaded scRNA-seq data files for GSE179640 and GSE213216 from the GEO database. GSE179640 included 6 samples: proliferative endometrium from healthy controls (n=3) and endometriosis patients (n=3). GSE213216 included 10 samples: proliferative endometrium from healthy controls (n=3) and endometriosis patients (n=7). The combined data from both datasets were processed through standardization, normalization, PCA, and Harmony analysis. Initial filtering was based on nFeature_RNA (>6000) and percent.mt (<20) ([Figure 1A](#)). Ultimately, in the UMAP analysis, 39 clusters were identified based on the unlabeled gene expression profiles ([Figure 1B](#)). To link these clusters to known cell types or functional states, we further annotated these subpopulations using the “SingleR”³¹ package, the CellMarker database,³² and previously reported literature. After both automatic and manual annotation, a total of 39 clusters were classified into eight distinct cell types: B cells/plasma cells, endothelial cells, epithelial cells, mast cells, mesenchymal cells, myeloid cells, smooth muscle cells,



and T/NKT cells (Figure 1C and D). The expression patterns of key marker genes across these cell types were visualized using bubble plots (Figure 1E).

We then analyzed scRNA-seq data from endometrial tissue samples of both healthy controls and endometriosis patients. From these data, we identified the top 100 genes with the highest expression levels in eutopic endometrial tissue (see Table S4 for the detailed list of top 100 highest expression genes in eutopic endometrium vs healthy endometrium). Subsequently, we calculated the differential expression levels (FC) and PctProp of these genes within each cell subtype (refer to Table S4 for the detailed results). The contribution of each cell type to endometriosis was then assessed by calculating the square root of $FC \times PctProp$. Our results indicate that mesenchymal cells have the highest contribution to the pathology of endometriosis (Figure 1F).

Identification of DEGs in Proliferative Phase Endometrial Bulk RNA-Seq Data

Additionally, we obtained the GSE25628 dataset from the GEO public database, which includes bulk RNA-seq data from 13 endometrial tissue samples: 6 from healthy controls and 7 from patients with endometriosis. Differential gene expression analysis between the two groups was conducted using the limma package. Genes were considered differentially expressed if they met the criteria of $P\text{-Value} < 0.05$ and $|\log_2\text{FC}| > 1$. This analysis identified a total of 2,084 differentially expressed genes, with 810 genes upregulated and 1,274 genes downregulated. (Figure 2A, B and Table S5).

Construction and Validation of a Characteristic Gene-Based Predictive Model for Endometriosis

We intersected the DEGs from the bulk RNA-seq dataset GSE25628 with the genes of the mesenchymal cells cluster identified in the scRNA-seq dataset, resulting in 67 intersecting genes (Figure 3A and Box S1). We used the GSE25628 dataset as the training set and the GSE153739 dataset as the validation set, and performed feature selection using Lasso regression. LASSO is a regression technique that allows for many covariates in a model. It uniquely regularizes the absolute values of regression coefficients, potentially shrinking some to zero, thereby enabling variable selection.⁴⁸ The Lasso regression identified 8 genes as characteristic genes for endometriosis, which we selected as candidate genes for further study and used to construct a predictive model (Figure 3B–D). The model formula is as follows: $\text{RiskScore} = \text{SYNE2} \times (-0.119478802623019) + \text{TXN} \times (-0.0938631926715449) + \text{NUPR1} \times (-0.0241295692262499) + \text{CTSK} \times 0.0227825216179987 + \text{GSN} \times 0.0244531595971548 + \text{MGP} \times 0.0595150231025177 + \text{IER2} \times 0.134667729309355 + \text{CXCL12} \times 0.18988373081461$. The results showed that the predictive model constructed with these 8 genes demonstrated good diagnostic performance, with an AUC of 1 (Figure 3E). We further validated the predictive model using the GSE153739 dataset as the validation set, and the results indicated that the model had strong stability, with an AUC of 0.8125 (Figure 3F). Information on the performance metrics of the predictive model is detailed in Table S6.

In this study, we further analyzed endometrial tissue from healthy controls and eutopic endometrium from endometriosis patients to assess the expression levels of the 8 key genes identified by our machine learning model. RT-PCR results showed significant upregulation of CXCL12, IER2, MGP, and GSN and downregulation of NUPR1 and TXN in endometriosis. Although the expression levels of CTSK and SYNE2 did not show significant differences, their trends were consistent with

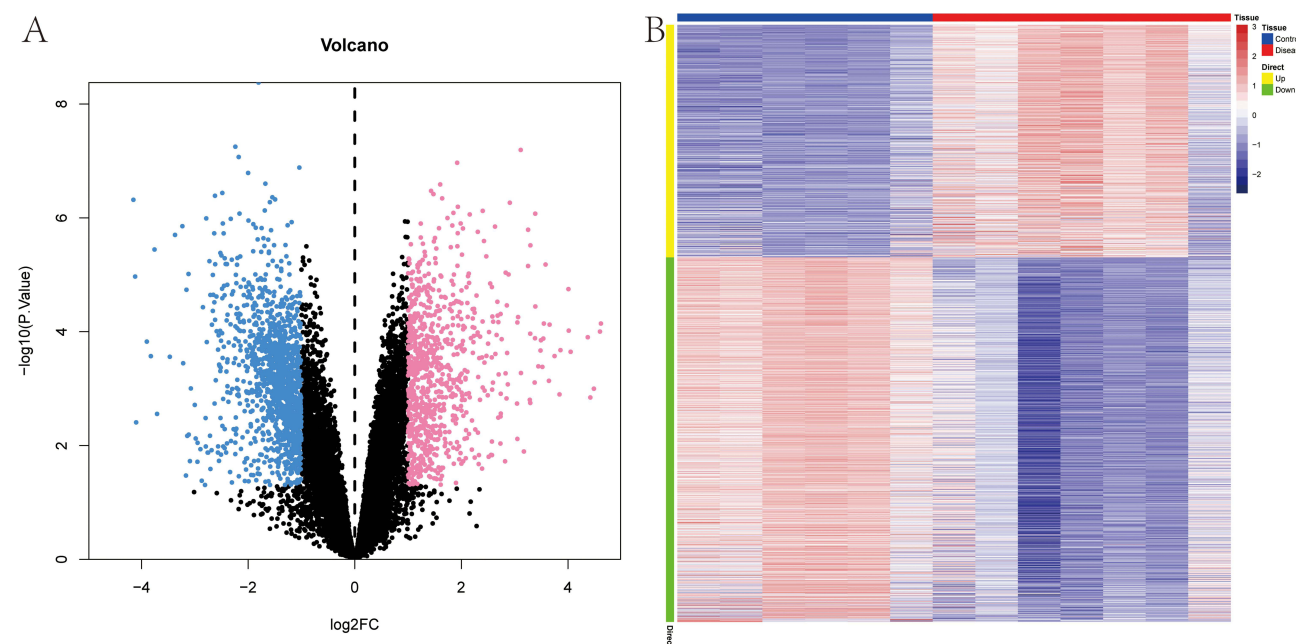


Figure 2 The identification of differentially expressed genes (DEGs) between the eutopic endometrium of healthy controls and endometriosis patients. (A). Volcano plot of DEGs between the endometrium of healthy controls and the eutopic endometrium of endometriosis patients in the GSE25628 dataset. $P < 0.05$ and $|\log_2\text{FoldChange}| > 1$ were considered significant DEGs. Red dots indicate upregulated genes, while blue dots indicate downregulated genes. (B). Heatmap of DEGs.

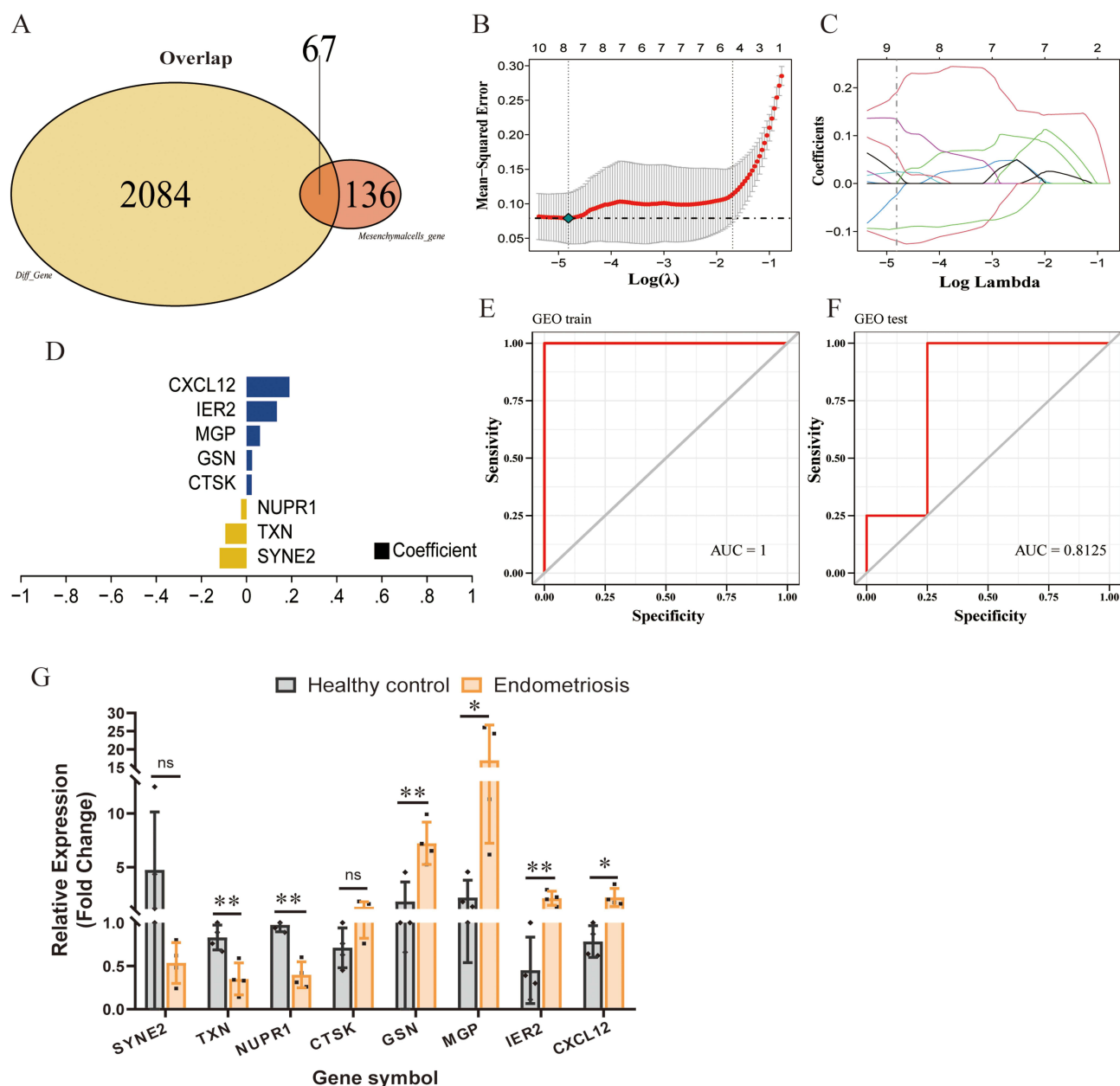


Figure 3 Construction and validation of a characteristic gene-based predictive model for endometriosis. (A). The DEGs from the Bulk RNA-seq dataset GSE25628 were intersected with the genes of the mesenchymal cell cluster identified in the scRNA-seq dataset, resulting in 67 intersecting genes. (B). Tenfold cross-validation for tuning parameter selection in the LASSO model. (C). Distribution of LASSO coefficients for the DEGs. Positive coefficients (blue) indicate a positive association, while negative coefficients (yellow) indicate a negative association. Bar width represents the magnitude of each gene's contribution to the predictive model. (D). Histogram of LASSO coefficients for the eight key genes. (E and F). Predictive performance of the training and validation sets. (G). The RT-PCR analysis results of key gene expression levels in endometrial samples from healthy controls and patients with endometriosis. * $p < 0.05$, ** $p < 0.01$, ns: not significant.

model predictions (Figure 3G and Table S7). These RT-PCR results further confirm the reliability of our predictive model. These 8 genes were identified as key candidate genes for our research. Subsequently, we presented the expression of these 8 key genes in 8 different cell types: Mast cells, B/plasma cells, Smooth muscle cells, Mesenchymal cells, T/NKT cells, Myeloid cells, Endothelial cells, and Epithelial cells (Figure 4A and B).

Analysis of Immuno-Infiltration in Eutopic Endometria of Endometriosis

The microenvironment, which was primarily composed of immune cells, extracellular matrix, various growth factors, inflammatory factors, and specific physicochemical characteristics, played a significant role in influencing disease diagnosis and clinical treatment sensitivity. Immunologic dysfunction played a crucial role in the implantation of endometriotic tissue. In

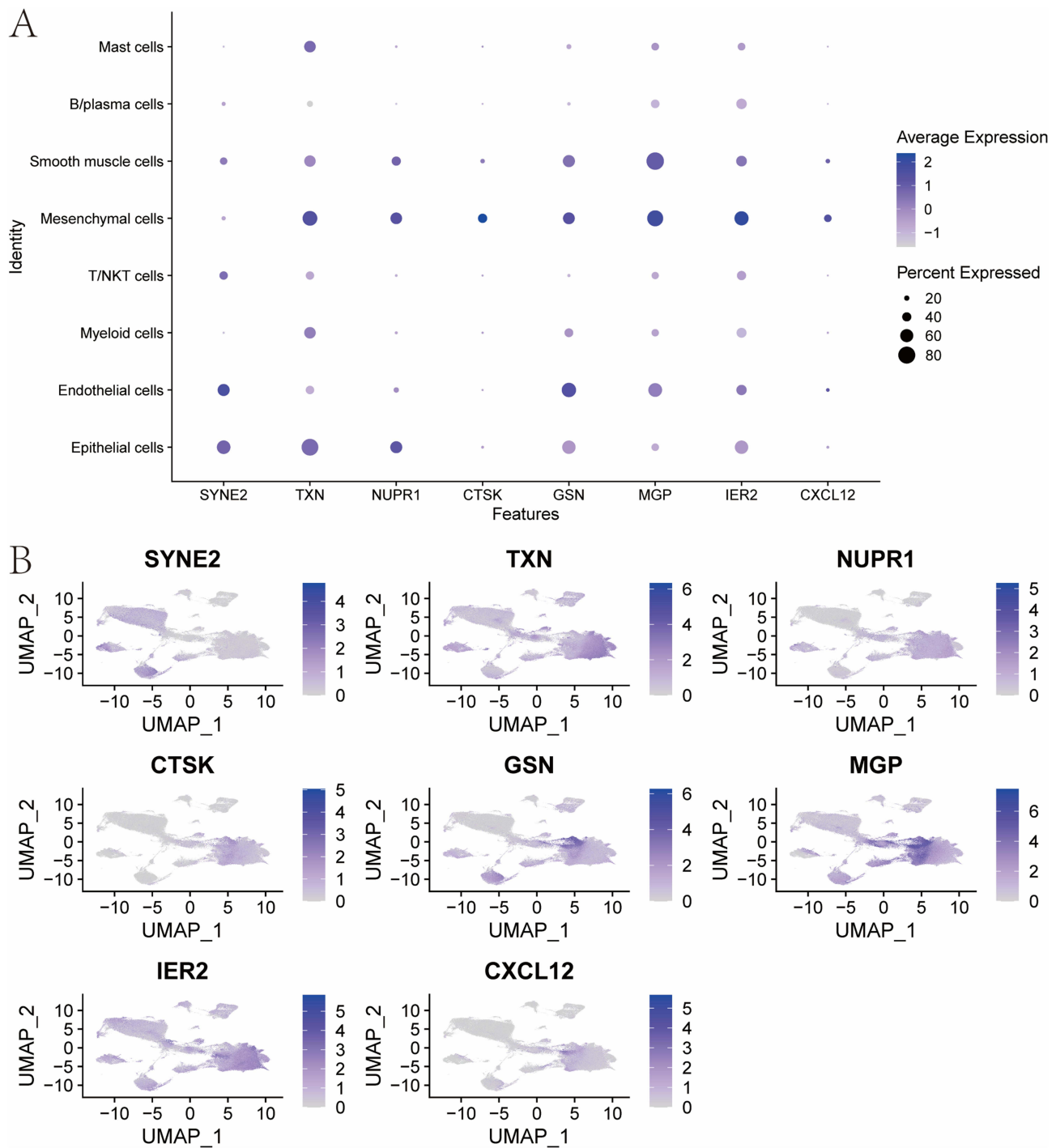


Figure 4 Expression of the 8 key genes across cell clusters. **(A)** Bubble plots visualizing the expression levels of the 8 key genes across each cell cluster. **(B)** UMAP plots showing the distribution and expression levels of the 8 key genes (SYNE2, TXN, NUPR1, CTSK, GSN, MGP, IER2, CXCL12) in the scRNA-seq dataset. Colors range from gray to blue, with deeper blue indicating higher expression levels.

recent years, several studies suggested that immune dysregulation was also present in the microenvironment of the eutopic endometrium in patients with endometriosis.^{49,50} Our study delved into the potential molecular mechanisms of endometriosis by analyzing the relationship between key genes and immune infiltration. In this study, we used both CIBERSORT and ssGSEA algorithms to analyze the distribution of immune cells in each eutopic endometrial sample and the correlations between different immune cell types (Figures 5A, B, and S2A, B). CIBERSORT analysis revealed that, compared to the

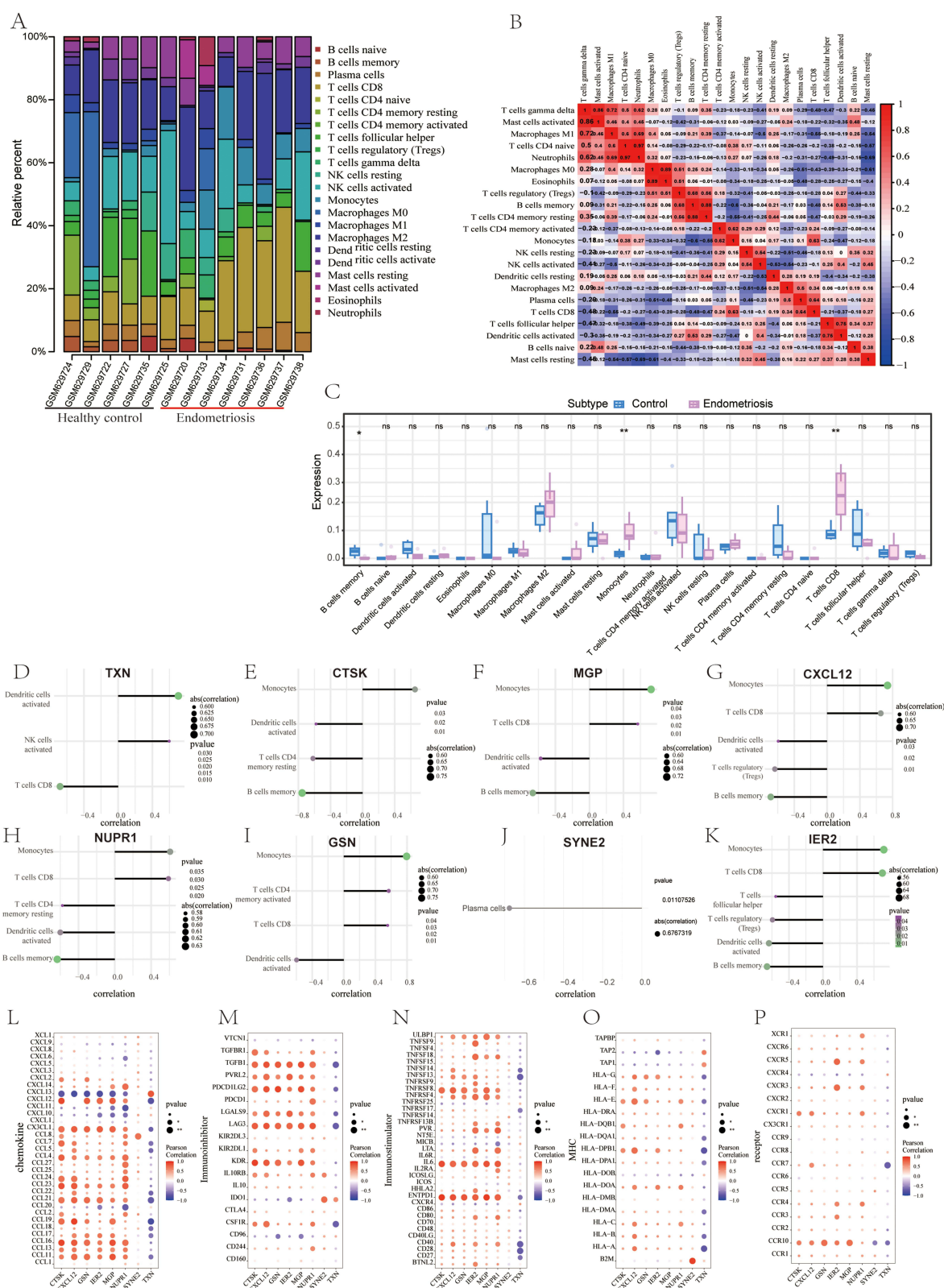


Figure 5 Analysis of Immuno-infiltration in Eutopic Endometria of Endometriosis. **(A)** The stacked bar chart shows the relative proportions of different immune cell types in the endometria of healthy controls and endometriosis patients. **(B)** The correlation heatmap displays the correlations between different immune cell types, with colors ranging from blue (negative correlation) to red (positive correlation). **(C)** The box plot compares the expression levels of different immune cell types in the endometria of healthy controls and endometriosis patients. **(D-K)** The correlation dot plot shows the correlation analysis between the eight key genes (TXN, CTSK, MGP, CXCL12, NUPR1, GSN, SYNE2, and IER2 respectively) and the infiltrating immune cell types in the endometrial tissues. **(L-P)** Correlation analysis of the eight key genes with various chemokines, immunoinhibitors, immunostimulators, MHC, and receptors. Dot color indicates the Pearson correlation coefficient (red for positive, blue for negative), and dot size reflects correlation significance, with larger dots indicating smaller p-values. * $p < 0.05$, ** $p < 0.01$, ns: not significant.

healthy control group, the levels of CD8 T cells and monocytes were significantly increased in endometriosis patients, while the levels of memory B cells were significantly decreased (Figure 5C). The study also highlighted a strong correlation between key genes and immune cells (Figure 5D–K), indicating that NUPR1, CTSK, GSN, MGP, IER2, and CXCL12 were significantly positively correlated with monocytes, while NUPR1, CTSK, MGP, IER2, and CXCL12 were significantly negatively correlated with memory B cells. Furthermore, the enrichment of 28 immune cell gene sets by ssGSEA further validated the immune cell infiltration distribution in the eutopic endometrium of endometriosis (Figure S2C–K). The study further explored the correlations between these 8 key genes and various immune factors, including immunomodulators, chemokines, and receptors, using data from the TISIDB (Tumor-Immune System Interactions Database)⁵¹ (Figure 5L–P). These findings suggested that these key genes were closely linked to immune cell infiltration and played critical roles within the immune microenvironment.

GSEA Analysis of Specific Signaling Pathways Involved in Key Genes

In this study, we further analyzed the specific signaling pathways associated with the 8 key genes to explore the potential molecular mechanisms by which these genes may influence disease progression. The GSEA results revealed that CTSK is enriched in pathways such as the Calcium signaling pathway and the MAPK signaling pathway (Figure 6A); CXCL12 is enriched in pathways including the Calcium signaling pathway and the cGMP-PKG signaling pathway (Figure 6B); GSN is enriched in pathways such as the cGMP-PKG signaling pathway and the Relaxin signaling pathway (Figure 6C); IER2 is enriched in the cGMP-PKG signaling pathway and the Relaxin signaling pathway (Figure 6D); MGP is enriched in pathways such as the cGMP-PKG signaling pathway and the Relaxin signaling pathway (Figure 6E); NUPR1 is enriched in the Apelin signaling pathway and the cGMP-PKG signaling pathway (Figure 6F); SYNE2 is enriched in pathways such as the Chemokine signaling pathway and the IL-17 signaling pathway (Figure 6G); and TXN is enriched in the Fanconi anemia pathway and the mRNA surveillance pathway (Figure 6H). Detailed information on the GSEA analysis of signaling pathways related to key genes can be found in Table S8.

GSVA and miRNA Interaction Analysis of Eight Key Genes

GSVA results show that high CTSK expression is associated with enrichment in IL6_JAK_STAT3_SIGNALING and IL2_STAT5_SIGNALING pathways (Figure 7A). High CXCL12 expression is linked to WNT_BETA_CATENIN_SIGNALING and HEDGEHOG_SIGNALING pathways (Figure 7B). High GSN expression enriches IL6_JAK_STAT3_SIGNALING and TGF_BETA_SIGNALING pathways (Figure 7C). High IER2 expression is associated with WNT_BETA_CATENIN_SIGNALING and IL6_JAK_STAT3_SIGNALING pathways (Figure 7D). High MGP expression enriches HEDGEHOG_SIGNALING and WNT_BETA_CATENIN_SIGNALING pathways (Figure 7E). High NUPR1 expression is linked to IL6_JAK_STAT3_SIGNALING and IL2_STAT5_SIGNALING pathways (Figure 7F). High SYNE2 expression enriches IL2_STAT5_SIGNALING and TGF_BETA_SIGNALING pathways (Figure 7G). High TXN expression is associated with MTORC1_SIGNALING and DNA_REPAIR pathways (Figure 7H).

In addition, using the miRcode database (<http://www.mircode.org>), we performed reverse prediction on the eight key genes, identifying 83 miRNAs and a total of 253 mRNA-miRNA interactions. These interactions were visualized using Cytoscape (Figure 7I).

Transcription Factor Enrichment and Motif Analysis of Key Genes

In this study, we analyzed eight key genes and found that they are regulated by multiple transcription factors. Using GSEA, we evaluated the enrichment of these transcription factors. Motif-TF annotation and key gene selection results indicated that the motif with the highest normalized enrichment score (NES) was cisbp_M5609, with an NES value of 6.24 (Figure 8A and B). We also identified the motifs enriched in these key genes and their corresponding transcription factors, as detailed in Figure 8C, which illustrates the relationships between these genes and the associated motifs.

Correlation Analysis Between Key Genes and Endometriosis-Related Disease Genes

This study obtained endometriosis-related disease genes through the GeneCards database (<https://www.genecards.org/>). The Relevance Score in the GeneCards database is used to measure the degree of association between a gene and

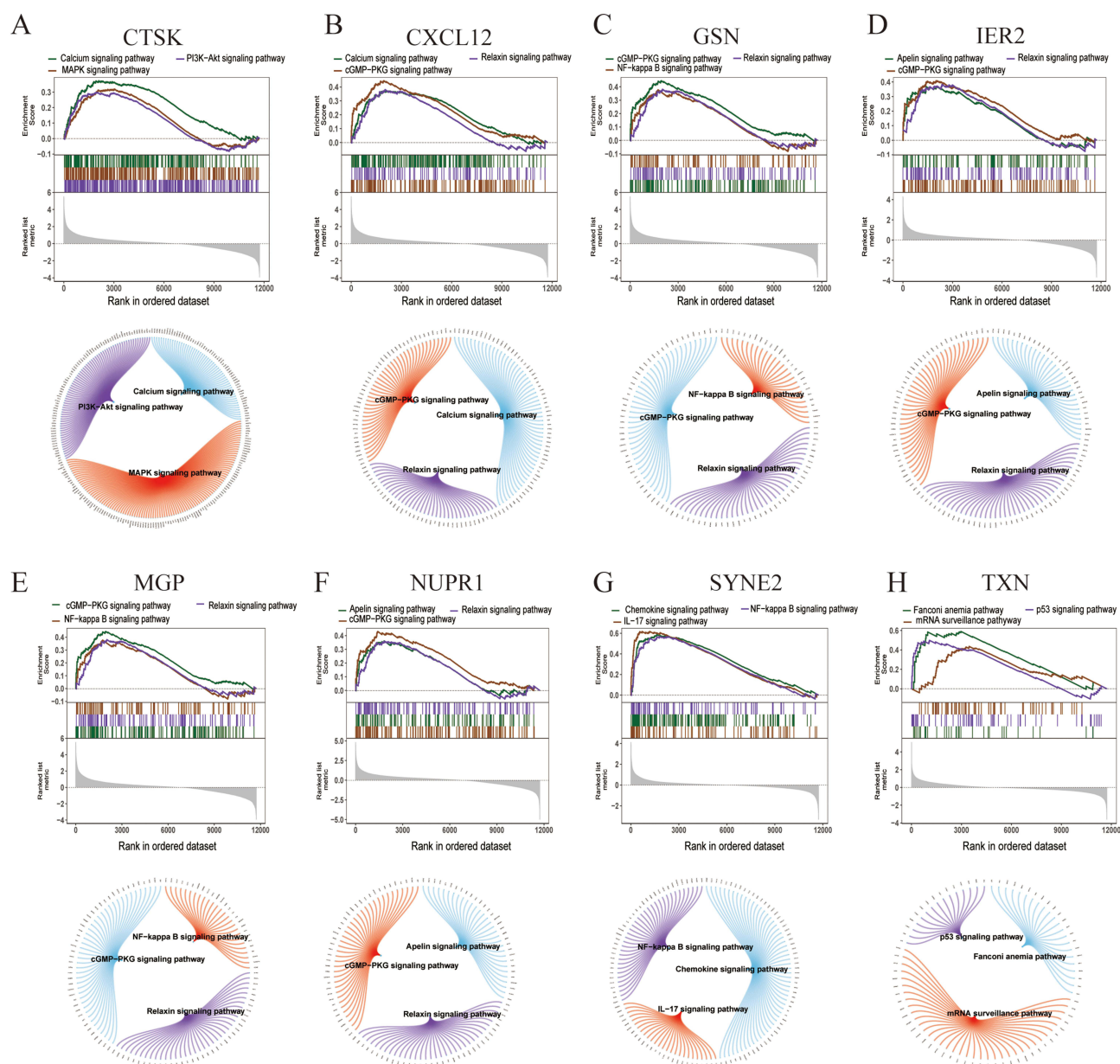


Figure 6 GSEA and Pathway Analysis of Key Genes. (A-H). Gene Set Enrichment Analysis (GSEA) and pathway analysis for eight key genes (CTSK, CXCL12, GSN, IER2, MGP, NUPR1, SYNE2, TXN). The upper panels show enrichment curves for top associated pathways, highlighting gene enrichment as they rank higher in the dataset. The lower panels display circular diagrams of key pathways, with segment width and color intensity indicating the significance of each pathway in the gene's function.

a specific disease; the higher the score, the stronger the association with the target disease. We selected the top 20 genes with the highest Relevance Scores (Top 20) and analyzed their expression levels in endometriotic and healthy control endometrial tissues, exploring the differences in expression between the two groups (Figure 9A). The results showed that the expression of genes such as CYP17A1, ESR1, ESR2, and GNRH1 differed significantly between the two groups. Additionally, we found that the expression levels of eight key genes were significantly correlated with the expression levels of the Top 20 regulatory genes in endometriosis. Among them, CXCL12 and MMP2 were significantly positively correlated ($r=0.963$), while CTSK and ESR1 were significantly negatively correlated ($r=-0.885$) (Figure 9B).

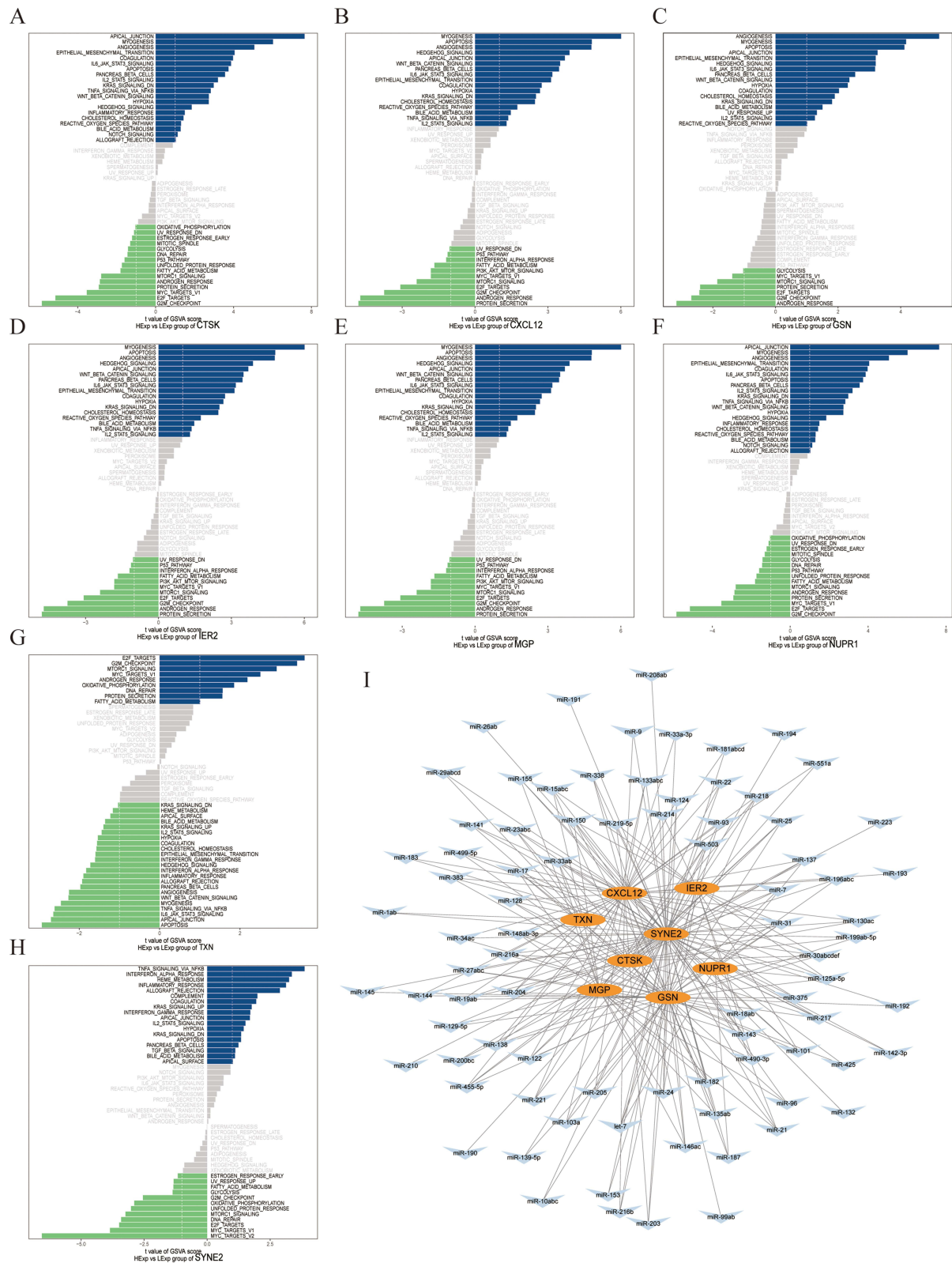


Figure 7 Pathway Activity Comparison via GSVA and miRNA-mRNA Interaction Network for Key Genes. (A-H). Gene Set Variation Analysis (GSVA) comparing pathway activity between high-expression (HExp) and low-expression (LExp) groups for eight key genes (CTSK, CXCL12, GSN, IER2, MGP, NUPR1, SYNE2, TXN). Positive t-values (blue bars) indicate pathways more active in the HExp group, while negative t-values (green bars) indicate pathways more active in the LExp group. (I). miRNA-mRNA interaction network for key genes. The network illustrates the interaction between eight key genes (CTSK, CXCL12, GSN, IER2, MGP, NUPR1, SYNE2, TXN) and their predicted targeting miRNAs. The Orange nodes represent the key genes, while the blue nodes represent the miRNAs.

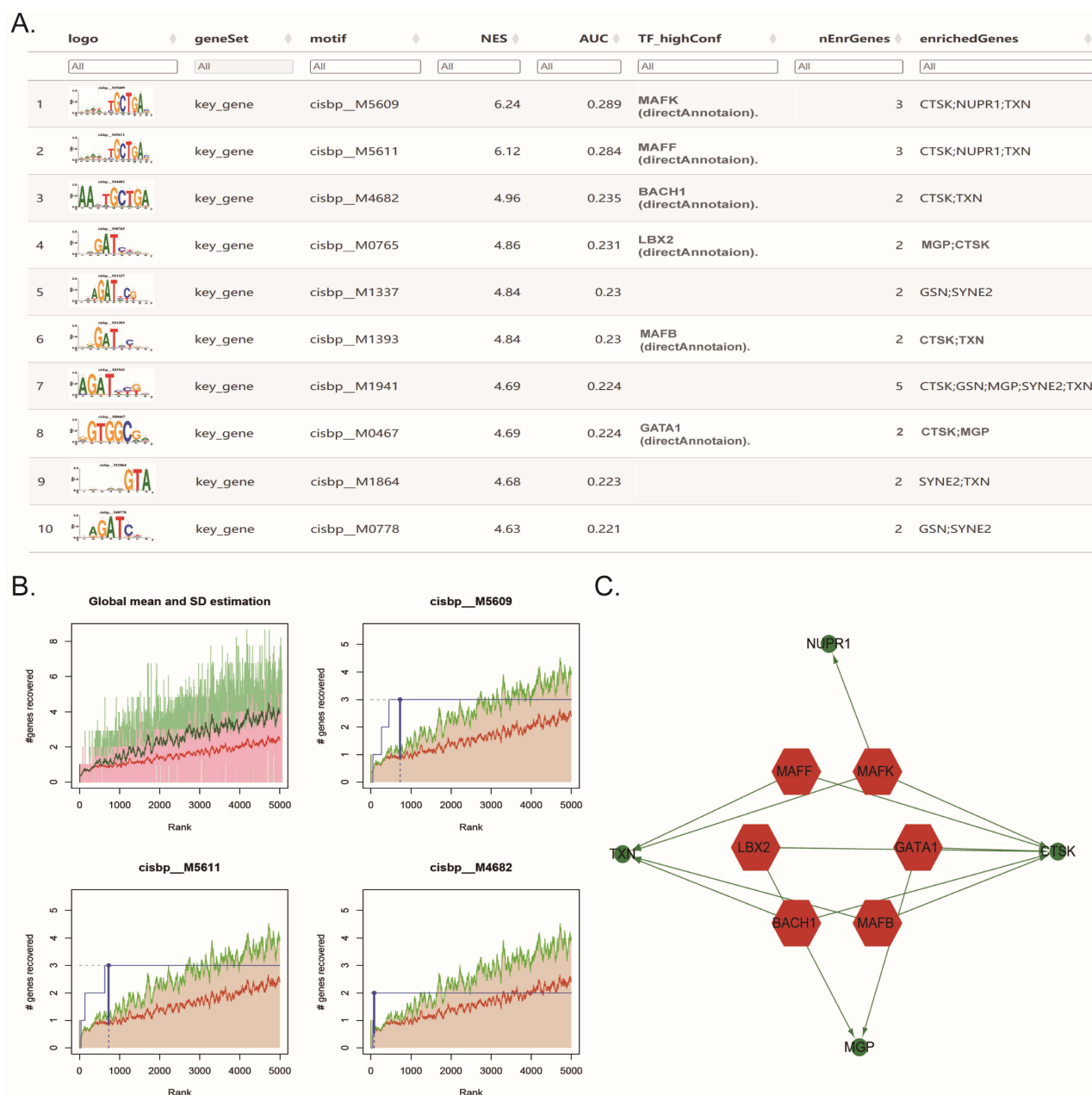


Figure 8 Transcription factor (TF) motif enrichment analysis for key genes. **(A).** Summary table of top enriched TF motifs associated with the key genes, including the normalized enrichment score (NES), area under the curve (AUC), and transcription factors annotated to motifs (TF_highConf). **(B).** Global mean and standard deviation (SD) estimation plots for the top TF motifs, illustrating the rank distribution of associated genes. **(C)** Interaction network showing the relationship between the enriched TFs (red hexagons) and the key genes (green circles), highlighting potential regulatory interactions.

Drug-Target Predictions Based on Common DEGs

By analyzing the Bulk RNA-seq datasets GSE25628 and GSE153739, we identified common DEGs between the two datasets and selected the top 150 upregulated and downregulated genes based on differential expression. Drug-target predictions for these differential genes were performed using the CMap.⁴⁶ CMap, a transcriptomics-based drug repurposing tool, has been used to identify potential therapeutic candidates for various conditions, including endometriosis. Several studies have demonstrated its capability to identify compounds with therapeutic potential for this disease.^{52,53} In our study, the results indicated that the expression profiles perturbed by drugs such as Retinol (Figure 9C), Orantinib (Figure 9D), Piperacillin

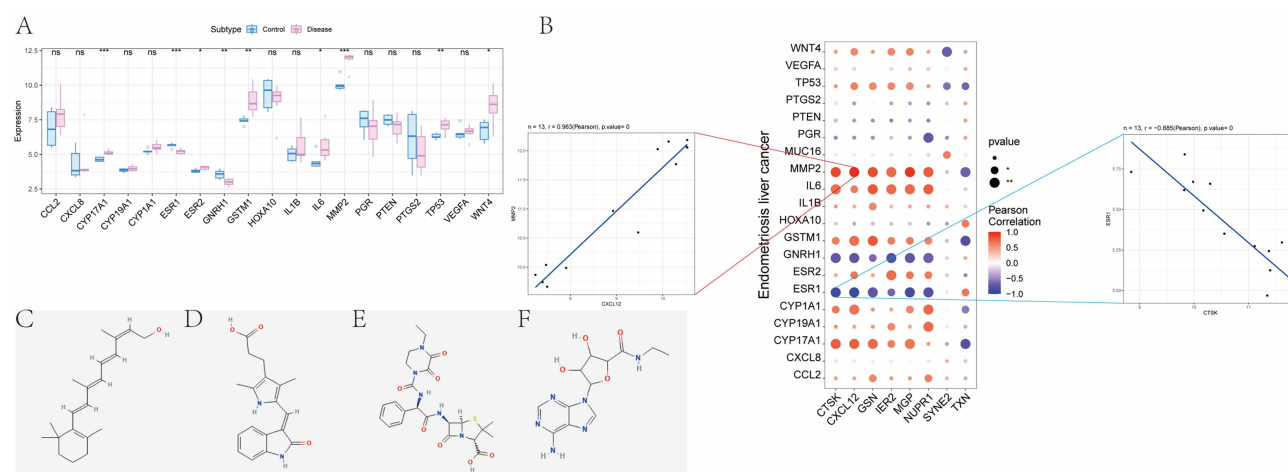


Figure 9 Drug-Target Predictions for Endometriosis. **(A)** Box plots showing gene expression levels in control (blue) and endometriosis (pink) groups. Significance: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. **(B)** Pearson correlation analysis of key genes with endometriosis-associated genes, with blue indicating negative and red indicating positive correlations. **(C-F)** The chemical structures of retinol, orantinib, piperacillin, and NECA, respectively. These four drugs were identified as potential treatments for endometriosis through the Connectivity Map (CMap) database.

(Figure 9E), and NECA (Figure 9F) were significantly negatively correlated with the expression profiles perturbed by endometriosis, suggesting that these drugs may have the potential to alleviate or even reverse the disease state. (Table S9).

Discussion

As a prevalent chronic condition, endometriosis can significantly impact the well-being of women and their families.⁵⁴ Despite the lack of comprehensive understanding regarding its pathogenesis, it is postulated that multifactorial interactions contribute to the development of endometriosis, while current knowledge in this field remains limited, particularly regarding immune dysregulation.⁵⁵ Therefore, further research is urgently warranted to comprehensively elucidate the immunopathological mechanisms underlying endometriosis and to investigate potential diagnostic biomarkers and optimal immunotherapy strategies for this condition.

Immune Cell Profile Dysregulation in Eutopic Endometrium of Endometriosis Immune Dysfunction

Immune dysfunction has been identified as a key factor in the implantation of endometriotic tissue. Recent research indicates that immune dysregulation also occurs within the microenvironment of the eutopic endometrium in endometriosis patients.^{49,50} Huang X et al⁵⁰ utilized the 10x Genomics platform for scRNA-seq to analyze endometrial cells from endometriosis patients and controls. They identified differences in immune cell proportions and IL-10 secretion during the secretory phase, with elevated proinflammatory cytokines in endometriosis. These findings suggest that immune microenvironment alterations in the eutopic endometrium may impair endometrial receptivity. In our current study, according to CIBERSORT and ssGSEA analysis, we observed a higher abundance of CD8⁺ T cells and monocytes and an increasing trend in M2 macrophages in patients with endometriosis compared to the control group, which aligns with previous research findings.^{56,57} In 2021, Ma J et al performed scRNA-seq on endometrial tissue from three healthy controls and three endometriosis patients.⁵⁷ The study identified significant differences in cell subtypes and gene expression between the eutopic endometrium of endometriosis patients and healthy controls, with monocytes notably more prevalent in the eutopic endometrium of endometriosis. The mononuclear macrophage system is considered a pivotal factor in the promotion of endometriosis, wherein mononuclear cells are recruited to the lesion site and induced to differentiate into macrophages, which serve as the principal immune cells responsible for pro-inflammatory chemokine production.^{58,59} Previous studies have reported a higher number of CD8⁺ T cells in the eutopic endometrium of endometriosis patients compared to normal controls.⁵⁶ Furthermore, increased levels of CD8⁺ T cells are associated with an elevated risk of infertility among individuals with endometriosis. In patients with endometriosis, CD8 MAIT cells represent the predominant subset in both peritoneal fluid and peripheral blood. These cells have the potential to induce a pro-inflammatory

state and create an immunosuppressive microenvironment by secreting IL-17 and initiating a Th17-biased immune response.⁶⁰ Unfortunately, despite compelling evidence of dysregulation in key regulators of CD8+ T cell function such as PD-1, PD-L1, and CD4+ T regulatory cells in endometriosis, functional analyses of CD8+ T cells have not yet been conducted to date.^{20,61,62}

Identifying Key Genes and Models for Predicting Endometriosis

Historically, the lack of specificity and sensitivity in a single biomarker or group of biomolecules has hindered their utilization as diagnostic tests for endometriosis.⁶³ Using integrated scRNA-Seq and bulk RNA-Seq datasets with machine learning algorithms, we developed a predictive model for endometriosis and identified eight key genes (SYNE2, TXN, NUPR1, CTSK, GSN, MGP, IER2, and CXCL12). These findings were validated through qRT-PCR using clinical specimens. The identified genes were closely associated with immune cell infiltration.

Previous studies have shown that CXCL12 is significantly upregulated in endometriosis lesions compared to healthy endometrial tissue,⁶⁴ consistent with our findings. CXCL12 binds to CXCR4/7, activating multiple signaling pathways,^{65,66} inhibiting autophagy in stromal cells, and promoting the proliferation, migration, and invasion of endometriotic cells. GSEA analysis revealed that CXCL12 is closely linked to monocyte infiltration, suggesting it recruits immune cells, causing immune dysregulation in the endometrial microenvironment and accelerating disease progression.

CTSK, a lysosomal cysteine protease, activates the TLR7/9 pathway in antigen-presenting cells, promoting Th17 differentiation.^{67,68} CTSK has been proposed as a diagnostic marker for endometriosis and correlates with immune scores.⁶⁹ Our study showed that CTSK expression is elevated in endometriosis lesions, with GSEA analysis indicating its enrichment in the MAPK pathway, which is crucial for implantation, growth, invasion, proliferation, and apoptosis in endometriotic tissue.⁷⁰

We also found that SYNE2 expression is significantly reduced in the eutopic endometrium of endometriosis, which aligns with prior studies.⁷¹ GSEA analysis indicated that SYNE2 is enriched in the chemokine and IL-17 signaling pathways, both key in immune regulation.^{72–74}

TXN (Thioredoxin), a gene critical for cellular redox balance, plays a central role in immune regulation and oxidative stress.⁷⁵ Despite ongoing debate about its expression in endometriosis,^{76,77} TXN may contribute to disease pathogenesis by modulating oxidative stress and immune pathways like NF- κ B and MAPK.

MGP is low in normal endometrium but significantly elevated in endometriosis lesions, especially in glandular and endothelial cells.⁷⁸ While no studies have reported on GSN,⁷⁹ IER2, and NUPR1 expression in endometriosis, GSEA analysis shows these genes, along with MGP, are enriched in the cGMP-PKG and Relaxin pathways. The Relaxin pathway may regulate endometriosis by inhibiting IL-8 expression in stromal cells.⁸⁰

Collectively, these genes hold potential as novel diagnostic biomarkers for endometriosis; however, further investigation is warranted to elucidate their precise role in the pathogenesis of this condition. Moreover, our study primarily focused on endometriosis lesions, and additional clinical studies are necessary to determine whether these genes exhibit differential expression in peripheral blood and can serve as non-invasive biomarkers for diagnosis.

Drug Target Prediction for Endometriosis Based on Eutopic Endometrium Sequencing Data

Through drug-target prediction analysis, we identified Retinol, Orantinib, Piperacillin, and NECA as compounds with expression profiles significantly inversely correlated with those associated with endometriosis. These findings suggest their potential therapeutic value in mitigating or reversing the disease state. Retinol and its derivatives, collectively known as retinoids, play a crucial role in endometrial development, maintenance and stromal decidualization, and blastocyst implantation. Dysregulated retinoid metabolism may disrupt mitochondrial function and facilitate the implantation and growth of ectopic cells, potentially serving as a pivotal factor in the pathogenesis of endometriosis.^{81–84} Retinoids have found widespread use in the treatment of various skin diseases, malignant tumors, and immune disorders. However, their poor solubility in aqueous solutions, increased catabolism upon intravenous injection, and potential for systemic side effects have prompted researchers to develop novel retinoid delivery agents with improved tolerability.⁸⁵ In the context of endometriosis, it is worth investigating strategies for targeted delivery of retinoids to the lesion site and specific immune cell populations. Orantinib (SU6668) exhibits multipotent inhibition of VEGF, FGF, and PDGF receptor

tyrosine kinase activity, effectively suppressing the angiogenesis and vascularization processes in ectopic endometrial tissue.⁸⁶ However, further investigation is required to elucidate its specific regulatory mechanism. The primary adverse events associated with orantinib treatment were edema, ascites, and elevations in levels of aspartate aminotransferase and alanine aminotransferase.⁸⁷

Limitations

Although this study explored the correlation between immune cell infiltration in the endometrium and endometriosis occurrence, and developed a predictive model using bioinformatics based on identified key genes, there are still some limitations. First, this study analyzed sequencing data only from the proliferative phase of the endometrium, without investigating dynamic changes in the immune microenvironment and transcriptomic expression across the entire menstrual cycle. This limitation may restrict the ability to fully capture the molecular characteristics of the eutopic endometrium in endometriosis. Longitudinal studies are warranted to explore these dynamic changes and provide a more comprehensive understanding. Second, the relatively small sample size, particularly in the validation set ($n = 7$), and the lack of detailed clinical follow-up data may limit the generalizability and predictive accuracy of the model. Furthermore, the influence of comorbidities or other health conditions was not fully considered, which could impact biomarker identification and model stability. Future studies should include larger sample sizes and more comprehensive clinical datasets to improve model reliability and applicability. Third, the reliance on *in silico* analyses introduces methodological constraints, including potential biases in publicly available sequencing datasets and the inherent limitations of computational predictions. While *in vitro* validation using qPCR provided initial support for the findings, the small number of clinical samples analyzed may reduce the generalizability of these results. Expanding the validation cohort to include larger clinical sample sizes will be crucial for confirming these findings and enhancing their broader applicability. Lastly, drug-target predictions in this study identified potential therapeutic agents, such as Retinol and Orantinib, with expression profiles negatively correlated with those associated with endometriosis. However, their efficacy and underlying mechanisms remain to be validated *in vivo*. Rigorous experimental studies utilizing endometriosis models are essential to confirm their therapeutic potential and elucidate their mechanisms of action. Addressing these limitations in future research will help strengthen the study's findings, advance understanding of endometriosis pathophysiology, and support the clinical translation of these insights.

Conclusion

This study elucidates the role of immune dysregulation in the pathogenesis of endometriosis and identifies potential diagnostic biomarkers (SYNE2, TXN, NUPR1, CTSK, GSN, MGP, IER2, CXCL12) as well as candidate drugs for this condition. These findings offer valuable insights for the development of early detection methods and targeted therapeutic strategies for endometriosis.

Abbreviations

AUC, Area Under the Curve; CMap, Connectivity Map; DEGs, Differentially Expressed Genes; GEO, Gene Expression Omnibus; GSEA, Gene Set Enrichment Analysis; GSVA, Gene Set Variation Analysis; LASSO, Least Absolute Shrinkage and Selection Operator; NES, Normalized Enrichment Score; PCA, Principal Component Analysis; qPCR, Quantitative Polymerase Chain Reaction; RT-qPCR, Reverse Transcription Quantitative Polymerase Chain Reaction; scRNA-seq, Single-cell RNA sequencing; UMAP, Uniform Manifold Approximation and Projection.

Data Sharing Statement

The dataset supporting the conclusions of this article are included within the article and [Supplementary Tables](#).

Ethics Approval and Consent to Participate

Our experimental protocol was approved by the Medical Ethics Committee of Union Hospital, Tongji Medical College, Huazhong University of Science and Technology (Ethics No. 2024-0404) and was conducted in accordance with the principles of the Declaration of Helsinki. All research participants provided written informed consent.

Acknowledgments

We extend our gratitude to all contributors who uploaded usable data to the NCBI GEO, TISIDB, GeneCards, miRcode and CMap databases, as well as the platforms that provide publicly available data. We also thank our et.al. at the Reproductive Medicine Center, Union Hospital, Tongji Medical College, and Huazhong University of Science and Technology for their assistance with sample collection.

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

This work was supported by the Hubei Provincial Natural Science Foundation of China (No. 2024AFB639) and the National Natural Science Foundation of China (No. 82301848).

Disclosure

The authors report no conflicts of interest in this work.

References

1. Zondervan KT, Becker CM, Missmer SA. Endometriosis. *N Engl J Med*. 2020;382(13):1244–1256. doi:10.1056/NEJMr1810764
2. Mishra B VV, Agarwal P, Aggarwal R, Aggarwal R. Prevalence, clinical and laparoscopic features of endometriosis among infertile women. *J Obstet Gynaecol India*. 2017;67:208–212. doi:10.1007/s13224-016-0931-x
3. Barnard ME, Farland LV, Yan B, et al. Endometriosis typology and ovarian cancer risk. *JAMA*. 2024;332(6):482–489. doi:10.1001/jama.2024.9210
4. Kvaskoff M, Mu F, Terry KL, et al. Endometriosis: a high-risk population for major chronic diseases? *Hum Reprod Update*. 2015;21(4):500–516. doi:10.1093/humupd/dmv013
5. Saunders PTK, Horne AW. Endometriosis: etiology, pathobiology, and therapeutic prospects. *Cell*. 2021;184(11):2807–2824. doi:10.1016/j.cell.2021.04.041
6. Saraswat L, Ayansina D, Cooper KG, et al. Impact of endometriosis on risk of further gynaecological surgery and cancer: a national cohort study. *BJOG*. 2018;125(1):64–72. doi:10.1111/1471-0528.14793
7. Younis JS, Shapso N, Fleming R, et al. Impact of unilateral versus bilateral ovarian endometriotic cystectomy on ovarian reserve: a systematic review and meta-analysis. *Hum Reprod Update*. 2019;25(3):375–391. doi:10.1093/humupd/dmy049
8. Becker CM, Bokor A, Heikinheimo O, et al. ESHRE guideline: endometriosis. *Hum Reprod Open*. 2022;2022(2):hoac009. doi:10.1093/hropen/hoac009
9. Wu D, Hu M, Hong L, et al. Clinical efficacy of add-back therapy in treatment of endometriosis: a meta-analysis. *Arch Gynecol Obstet*. 2014;290(3):513–523. doi:10.1007/s00404-014-3230-8
10. Sauerbrun-Cutler MT, Alvero R. Short- and long-term impact of gonadotropin-releasing hormone analogue treatment on bone loss and fracture. *Fertil Steril*. 2019;112(5):799–803. doi:10.1016/j.fertnstert.2019.09.037
11. Rogers PA, Adamson GD, Al-Jefout M, et al. Research priorities for endometriosis. *Reprod Sci*. 2017;24(2):202–226. doi:10.1177/1933719116654991
12. Vercellini P, Viganò P, Somigliana E, et al. Endometriosis: pathogenesis and treatment. *Nat Rev Endocrinol*. 2014;10(5):261–275. doi:10.1038/nrendo.2013.255
13. Focarelli R, Luddi A, De Leo V, et al. Dysregulation of GdA expression in endometrium of women with endometriosis: implication for endometrial receptivity. *Reprod Sci*. 2018;25(4):579–586. doi:10.1177/1933719117718276
14. Wang M, Hao C, Huang X, et al. Aberrant expression of lncRNA (HOXA11-AS1) and homeobox A (HOXA9, HOXA10, HOXA11, and HOXA13) genes in infertile women with endometriosis. *Reprod Sci*. 2018;25(5):654–661. doi:10.1177/1933719117734320
15. Cui D, Ma J, Liu Y, et al. Analysis of long non-coding RNA expression profiles using RNA sequencing in ovarian endometriosis. *Gene*. 2018;673:140–148. doi:10.1016/j.gene.2018.06.046
16. Sherwin JRA, Sharkey AM, Mihalyi A, et al. Global gene analysis of late secretory phase, eutopic endometrium does not provide the basis for a minimally invasive test of endometriosis. *Hum Reprod*. 2008;23(5):1063–1068. doi:10.1093/humrep/den078
17. Hu S, Yao G, Wang Y, et al. Transcriptomic changes during the pre-receptive to receptive transition in human endometrium detected by RNA-Seq. *J Clin Endocrinol Metab*. 2014;99(12):E2744–53. doi:10.1210/jc.2014-2155
18. Fassbender A, Rahmiloglu N, Vitonis AF, et al. World endometriosis research foundation endometriosis phenome and biobanking harmonisation project: IV. Tissue collection, processing, and storage in endometriosis research. *Fertil Steril*. 2014;102(5):1244–1253. doi:10.1016/j.fertnstert.2014.07.1209
19. Armstrong GM, Maybin JA, Murray AA, et al. Endometrial apoptosis and neutrophil infiltration during menstruation exhibits spatial and temporal dynamics that are recapitulated in a mouse model. *Sci Rep*. 2017;7(1):17416. doi:10.1038/s41598-017-17565-x
20. Vallvé-Juanico J, Houshdaran S, Giudice LC. The endometrial immune environment of women with endometriosis. *Hum Reprod Update*. 2019;25(5):564–591. doi:10.1093/humupd/dmz018

21. Jovic D, Liang X, Zeng H, et al. Single-cell RNA sequencing technologies and applications: a brief overview. *Clin Transl Med.* **2022**;12(3):e694. doi:10.1002/ctm2.694
22. Fonseca MA, Haro M, Wright KN, et al. Single-cell transcriptomic analysis of endometriosis. *Nat Genet.* **2023**;55(2):255–267. doi:10.1038/s41588-022-01254-1
23. Li Q, Shi J, Yi D, et al. The pathogenesis of endometriosis and adenomyosis: insights from single-cell RNA sequencing. *Biol Reprod.* **2024**;110(5):854–865. doi:10.1093/biolre/iaoe032
24. Potter SS. Single-cell RNA sequencing for the study of development, physiology and disease. *Nat Rev Nephrol.* **2018**;14(8):479–492. doi:10.1038/s41581-018-0021-7
25. Chen S, Chai X, Wu X. Bioinformatical analysis of the key differentially expressed genes and associations with immune cell infiltration in development of endometriosis. *BMC Genom Data.* **2022**;23(1):20. doi:10.1186/s12863-022-01036-y
26. Bane K, Desouza J, Shetty D, et al. Endometrial DNA damage response is modulated in endometriosis. *Hum Reprod.* **2021**;36(1):160–174. doi:10.1093/humrep/deaa255
27. Tan Y, Flynn WF, Sivajothi S, et al. Single-cell analysis of endometriosis reveals a coordinated transcriptional programme driving immunotolerance and angiogenesis across eutopic and ectopic tissues. *Nat Cell Biol.* **2022**;24(8):1306–1318. doi:10.1038/s41556-022-00961-5
28. Stuart T, Butler A, Hoffman P, et al. Comprehensive integration of single-cell data. *Cell.* **2019**;177(7):1888–1902. doi:10.1016/j.cell.2019.05.031
29. Korsunsky I, Millard N, Fan J, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods.* **2019**;16(12):1289–1296. doi:10.1038/s41592-019-0619-0
30. Becht E, McInnes L, Healy J, et al. Dimensionality reduction for visualizing single-cell data using UMAP. *Nat Biotechnol.* **2019**;37(1):38–44. doi:10.1038/nbt.4314
31. Aran D, Looney AP, Liu L, et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat Immunol.* **2019**;20(2):163–172. doi:10.1038/s41590-018-0276-y
32. Zhang X, Lan Y, Xu J, et al. CellMarker: a manually curated resource of cell markers in human and mouse. *Nucleic Acids Res.* **2019**;47:D721–D728. doi:10.1093/nar/gky900
33. Hu Y, Wang Y, Zhi L, et al. SDC4 protein action and related key genes in nonhealing diabetic foot ulcers based on bioinformatics analysis and machine learning. *Int J Biol Macromol.* **2024**;283(Pt 2):137789. doi:10.1016/j.ijbiomac.2024.137789
34. Wu Y, Jiang T, Hua J, et al. Integrated bioinformatics-based analysis of hub genes and the mechanism of immune infiltration associated with acute myocardial infarction. *Front Cardiovasc Med.* **2022**;9:831605. doi:10.3389/fcvm.2022.831605
35. Gustavsson EK, Zhang D, Reynolds RH, et al. ggtranscript: an R package for the visualization and interpretation of transcript isoforms using ggplot2. *Bioinformatics.* **2022**;38(15):3844–3846. doi:10.1093/bioinformatics/btac409
36. Li G-M, Zhang C-L, Rui R-P, et al. Bioinformatics analysis of common differential genes of coronary artery disease and ischemic cardiomyopathy. *Eur Rev Med Pharmacol Sci.* **2018**;22(11):3553–3569. doi:10.26355/eurrev_201806_15182
37. Liu Z, Liu L, Weng S, et al. Machine learning-based integration develops an immune-derived lncRNA signature for improving outcomes in colorectal cancer. *Nat Commun.* **2022**;13(1):816. doi:10.1038/s41467-022-28421-6
38. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw.* **2010**;33(1):1–22. doi:10.18637/jss.v033.i01
39. Hebert PD, Cywinska A, Ball SL, DeWaard JR. Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London Series B.* **2003**;270(1512):313–321. doi:10.1098/rspb.2002.2218
40. Newman AM, Liu CL, Green MR, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods.* **2015**;12(5):453–457. doi:10.1038/nmeth.3337
41. Bindea G, Mlecnik B, Tosolini M, et al. Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. *Immunity.* **2013**;39(4):782–795. doi:10.1016/j.immuni.2013.10.003
42. Charoentong P, Finotello F, Angelova M, et al. Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade. *Cell Rep.* **2017**;18(1):248–262. doi:10.1016/j.celrep.2016.12.019
43. Ha M, Kim VN. Regulation of microRNA biogenesis. *Nat Rev Mol Cell Biol.* **2014**;15(8):509–524. doi:10.1038/nrm3838
44. Aibar S, González-Blas CB, Moerman T, et al. SCENIC: single-cell regulatory network inference and clustering. *Nat Methods.* **2017**;14(11):1083–1086. doi:10.1038/nmeth.4463
45. Musa A, Ghorais LS, Zhang S-D, et al. A review of connectivity map and computational approaches in pharmacogenomics. *Brief Bioinform.* **2018**;19(3):506–523. doi:10.1093/bib/bbx023
46. Subramanian A, Narayan R, Corsello SM, et al. A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell.* **2017**;171(6):1437–1452. doi:10.1016/j.cell.2017.10.049
47. Canis M, Donnez JG, Guzick DS, et al. Revised American society for reproductive medicine classification of endometriosis: 1996. *Fertil Steril.* **1997**;67(5):817–821. doi:10.1016/s0015-0282(97)81391-x
48. Tibshirani R. The lasso method for variable selection in the Cox model. *Stat Med.* **1997**;16(4):385–395. doi:10.1002/(sici)1097-0258(19970228)16:4
49. Jørgensen H, Fedorcsak P, Isaacson K, et al. Endometrial cytokines in patients with and without endometriosis evaluated for infertility. *Fertil Steril.* **2022**;117(3):629–640. doi:10.1016/j.fertnstert.2021.11.024
50. Huang X, Wu L, Pei T, et al. Single-cell transcriptome analysis reveals endometrial immune microenvironment in minimal/mild endometriosis. *Clin Exp Immunol.* **2023**;212(3):285–295. doi:10.1093/cei/uxad029
51. Ru B, Wong CN, Tong Y, et al. TISIDB: an integrated repository portal for tumor-immune system interactions. *Bioinformatics.* **2019**;35(20):4200–4202. doi:10.1093/bioinformatics/btz210
52. Oskotsky TT, Bhoja A, Bunis D, et al. Identifying therapeutic candidates for endometriosis through a transcriptomics-based drug repositioning approach. *iScience.* **2024**;27(4):109388. doi:10.1016/j.isci.2024.109388
53. Zheng W, Wu J, Gu J, et al. Modular Characteristics and Mechanism of Action of Herbs for Endometriosis Treatment in Chinese Medicine: a Data Mining and Network Pharmacology-Based Identification. *Front Pharmacol.* **2020**;11:147. doi:10.3389/fphar.2020.00147
54. Zondervan KT, Becker CM, Koga K, Missmer SA, Taylor RN, Viganò P. Endometriosis. *Nat Rev Dis Primers.* **2018**;4(1):9. doi:10.1038/s41572-018-0008-5
55. Abramiuk M, Grywalska E, Małkowska P, Sierawska O, Hryniewicz R, Niedźwiedzka-Rystwej P. The role of the immune system in the development of endometriosis. *Cells.* **2022**;11(13):2028. doi:10.3390/cells11132028

56. Wu XG, Chen JJ, Zhou HL, et al. Identification and validation of the signatures of infiltrating immune cells in the eutopic endometrium endometria of women with endometriosis. *Front Immunol.* **2021**;12:671201. doi:10.3389/fimmu.2021.671201
57. Ma J, Zhang L, Zhan H, et al. Single-cell transcriptomic analysis of endometriosis provides insights into fibroblast fates and immune cell heterogeneity. *Cell Biosci.* **2021**;11:1–19. doi:10.1186/s13578-021-00637-x
58. Zhu T, Du Y, Jin B, Zhang F, Guan Y. Identifying immune cell infiltration and hub genes related to M2 macrophages in endometriosis by bioinformatics analysis. *Reprod Sci.* **2023**;30(11):3388–3399. doi:10.1007/s43032-023-01227-7
59. Jeljeli M, Riccio LGC, Chouzenoux S, et al. Macrophage immune memory controls endometriosis in mice and humans. *Cell Rep.* **2020**;33(5):108325. doi:10.1016/j.celrep.2020.108325
60. Li C, Lu Z, Bi K, et al. CD4(+)/CD8(+) mucosa-associated invariant T cells foster the development of endometriosis: a pilot study. *Reprod Biol Endocrinol.* **2019**;17(1):78. doi:10.1186/s12958-019-0524-5
61. Kisovar A, Becker CM, Granne I, Southcombe JH. The role of CD8+ T cells in endometriosis: a systematic review. *Front Immunol.* **2023**;14:1225639. doi:10.3389/fimmu.2023.1225639
62. Wu L, Lv C, Su Y, et al. Expression of programmed death-1 (PD-1) and its ligand PD-L1 is upregulated in endometriosis and promoted by 17beta-estradiol. *Gynecol Endocrinol.* **2019**;35(3):251–256. doi:10.1080/09513590.2018.1519787
63. May KE, Conduit-Hulbert SA, Villar J, Kirtley S, Kennedy SH, Becker CM. Peripheral biomarkers of endometriosis: a systematic review. *Hum Reprod Update.* **2010**;16(6):651–674. doi:10.1093/humupd/dmq009
64. Zhou C, Feng M, Chen Y, et al. Unraveling immunotherapeutic targets for endometriosis: a transcriptomic and single-cell analysis. *Front Immunol.* **2023**;14:1288263. doi:10.3389/fimmu.2023.1288263
65. Krikun G. The CXCL12/CXCR4/CXCR7 axis in female reproductive tract disease: review. *Am J Reprod Immunol.* **2018**;80(5):e13028. doi:10.1111/aji.13028
66. Mei J, Zhu XY, Jin LP, Duan ZL, Li DJ, Li MQ. Estrogen promotes the survival of human secretory phase endometrial stromal cells via CXCL12/CXCR4 up-regulation-mediated autophagy inhibition. *Hum Reprod.* **2015**;30(7):1677–1689. doi:10.1093/humrep/dev100
67. Hirai T, Kanda T, Sato K, et al. Cathepsin K is involved in development of psoriasis-like skin lesions through TLR-dependent Th17 activation. *J Immunol.* **2013**;190(9):4805–4811. doi:10.4049/jimmunol.1200901
68. Asagiri M, Hirai T, Kunigami T, et al. Cathepsin K-dependent toll-like receptor 9 signaling revealed in experimental arthritis. *Science.* **2008**;319:624–627. doi:10.1126/science.1150110
69. Zhu R, Liu Y, Zhou J, Lv Z, Shi K, Xiong J. Development and validation of the diagnostic model of 7 gene in endometriosis. *Curr Med Chem.* **2024**;32. doi:10.2174/0109298673283426231220100011
70. Bora G, Yaba A. The role of mitogen-activated protein kinase signaling pathway in endometriosis. *J Obstet Gynaecol Res.* **2021**;47(5):1610–1623. doi:10.1111/jog.14710
71. Santin A, Spedicati B, Morgan A, et al. Puzzling out the genetic architecture of endometriosis: whole-exome sequencing and novel candidate gene identification in a deeply clinically characterised cohort. *Biomedicines.* **2023**;11(8). doi:10.3390/biomedicines11082122
72. Miller JE, Ahn SH, Marks RM, et al. IL-17A modulates peritoneal macrophage recruitment and M2 polarization in endometriosis. *Front Immunol.* **2020**;11:108. doi:10.3389/fimmu.2020.00108
73. Shi JL, Zheng ZM, Chen M, Shen HH, Li MQ, Shao J. IL-17: an important pathogenic factor in endometriosis. *Int J Med Sci.* **2022**;19(4):769–778. doi:10.7150/ijms.71972
74. Li Y, Zhou Z, Liang X, et al. Gut microbiota disorder contributes to the production of IL-17A that exerts chemotaxis via binding to IL-17RA in endometriosis. *J Inflamm Res.* **2024**;17:4199–4217. doi:10.2147/jir.S458928
75. Muri J, Kopf M. Redox regulation of immunometabolism. *Nat Rev Immunol.* **2021**;21(6):363–381. doi:10.1038/s41577-020-00478-8
76. Huang -Y-Y, Wu C-H, Liu C-H, et al. Association between the genetic variants of glutathione peroxidase 4 and severity of endometriosis. *Int J Environ Res Public Health.* **2020**;17(14):5089. doi:10.3390/ijerph17145089
77. Seo SK, Yang HI, Lee KE, et al. The roles of thioredoxin and thioredoxin-binding protein-2 in endometriosis. *Hum Rprod.* **2010**;25(5):1251–1258. doi:10.1093/humrep/deq027
78. Van Langendonck A, Punyadeera C, Kamps R, et al. Identification of novel antigens in blood vessels in rectovaginal endometriosis. *Mol Hum Reprod.* **2007**;13(12):875–886. doi:10.1093/molehr/gam073
79. Wang Y, Bi X, Luo Z, Wang H, Ismtula D, Guo C. Gelsolin: a comprehensive pan-cancer analysis of potential prognosis, diagnostic, and immune biomarkers. *Front Genet.* **2023**;14:1093163. doi:10.3389/fgene.2023.1093163
80. Yoshino O, Ono Y, Honda M, et al. Relaxin-2 may suppress endometriosis by reducing fibrosis, scar formation, and inflammation. *Biomedicines.* **2020**;8(11):467. doi:10.3390/biomedicines8110467
81. Harris HR, Eke AC, Chavarro JE, Missmer SA. Fruit and vegetable consumption and risk of endometriosis. *Hum Reprod.* **2018**;33(4):715–727. doi:10.1093/humrep/dey014
82. Jiang Y, Chen L, Taylor RN, Li C, Zhou X. Physiological and pathological implications of retinoid action in the endometrium. *J Endocrinol.* **2018**;236(3):R169–r188. doi:10.1530/joe-17-0544
83. Anderson G. Endometriosis pathoetiology and pathophysiology: roles of Vitamin A, estrogen, immunity, adipocytes, gut microbiome and melatonergic pathway on mitochondria regulation. *Biomol Concepts.* **2019**;10(1):133–149. doi:10.1515/bmc-2019-0017
84. Pierzchalski K, Taylor RN, Nezhat C, et al. Retinoic acid biosynthesis is impaired in human and murine endometriosis. *Biol Reprod.* **2014**;91(4):84. doi:10.1095/biolreprod.114.119677
85. Ferreira R, Napoli J, Enver T, Bernardino L, Ferreira L. Advances and challenges in retinoid delivery systems in regenerative and therapeutic medicine. *Nat Commun.* **2020**;11(1):4265. doi:10.1038/s41467-020-18042-2
86. Laschke MW, Elitzsch A, Vollmar B, Vajkoczy P, Menger MD. Combined inhibition of vascular endothelial growth factor (VEGF), fibroblast growth factor and platelet-derived growth factor, but not inhibition of VEGF alone, effectively suppresses angiogenesis and vessel maturation in endometriotic lesions. *Hum Reprod.* **2006**;21(1):262–268. doi:10.1093/humrep/dei308
87. Kudo M, Cheng AL, Park JW, et al. Orantinib versus placebo combined with transcatheter arterial chemoembolisation in patients with unresectable hepatocellular carcinoma (ORIENTAL): a randomised, double-blind, placebo-controlled, multicentre, Phase 3 study. *Lancet Gastroenterol Hepatol.* **2018**;3(1):37–46. doi:10.1016/s2468-1253(17)30290-x

Journal of Inflammation Research**Dovepress**

Taylor & Francis Group

Publish your work in this journal

The Journal of Inflammation Research is an international, peer-reviewed open-access journal that welcomes laboratory and clinical findings on the molecular basis, cell biology and pharmacology of inflammation including original research, reviews, symposium reports, hypothesis formation and commentaries on: acute/chronic inflammation; mediators of inflammation; cellular processes; molecular mechanisms; pharmacology and novel anti-inflammatory drugs; clinical conditions involving inflammation. The manuscript management system is completely online and includes a very quick and fair peer-review system. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/journal-of-inflammation-research-journal>