Heliyon 9 (2023) e18277

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Identification and validation of risk score model based on gene set activity as a diagnostic biomarker for endometriosis

Yi Zhang [a,*,1], Lulu Wu [b,1], Xiang Wen [c,1], Xiuwei Lv [d]

[a] Department of Gynecology, Second Affiliated Hospital of Hunan University of Traditional Chinese Medicine, Changsha 410005, China
[b] Department of Integrated Traditional Chinese and Western Medicine, The Central Hospital of Wuhan, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430014, China
[c] Department of Pathology, The First People's Hospital of Huizhou City, Huizhou 516000, China
[d] Department of Traditional Chinese Medicine, Rocket Force Medical Center of PLA, Beijing 100088, China

ARTICLE INFO

Keywords:
Endometriosis
Gene set variation analysis
Gene set
Diagnostic biomarker

ABSTRACT

Objective: The enigmatic nature of Endometriosis (EMS) pathogenesis necessitates investigating alterations in signaling pathway activity to enhance our comprehension of the disease's characteristics.

Methods: Three published gene expression profiles (GSE11691, GSE25628, and GSE7305 datasets) were downloaded, and the "combat" algorithm was employed for batch correction, gene expression difference analysis, and pathway enrichment difference analysis. The protein-protein interaction (PPI) network was constructed to identify core genes, and the relative enrichment degree of gene sets was evaluated. The Lasso regression model identified candidate gene sets with diagnostic value, and a risk scoring diagnostic model was constructed for further validation on the GSE86534 and GSE5108 datasets. CIBERSORT was used to assess the composition of immune cells in EMS, and the correlation between EMS diagnostic value gene sets and immune cells was evaluated.

Results: A total of 568 differentially expressed genes were identified between eutopic and ectopic endometrium, with 10 core genes in the PPI network associated with cell cycle regulation. Inflammation-related pathways, including cytokine-receptor signaling and chemokine signaling pathways, were significantly more active in ectopic endometrium compared to eutopic endometrium. Diagnostic gene sets for EMS, such as homologous recombination, base excision repair, DNA replication, P53 signaling pathway, adherens junction, and SNARE interactions in vesicular transport, were identified. The risk score's area under the curve (AUC) was 0.854, as indicated by the receiver operating characteristic (ROC) curve, and the risk score's diagnostic value was validated by the validation cohort. Immune cell infiltration analysis revealed correlations between the risk score and Macrophages M2, Plasma cells, resting NK cells, activated NK cells, and regulatory T cells.

Conclusion: The risk scoring diagnostic model, based on pathway activity, demonstrates high diagnostic value and offers novel insights and strategies for the clinical diagnosis and treatment of Endometriosis.

* Corresponding author.
E-mail address: zhangyi0731@foxmail.com (Y. Zhang).
[1] These authors have contributed equally to this work.

https://doi.org/10.1016/j.heliyon.2023.e18277
Received 18 November 2022; Received in revised form 28 June 2023; Accepted 13 July 2023
Available online 15 July 2023

1. Introduction

Endometriosis (EMS), an estrogen-dependent disease with high prevalence, impacts approximately 190 million women globally, affecting 10–15% of those in reproductive age [1]. This condition primarily presents as growth of endometrial stroma or glandular tissue outside the uterine cavity [2]. EMS patients may suffer from dysmenorrhea, irregular uterine bleeding, chronic pelvic pain, infertility, and other complications [3]. These symptoms considerably decrease patients' quality of life, while the increased utilization of assisted reproductive technology due to infertility raises social costs as well [4]. Therefore, it is imperative to investigate the pathogenesis and treatment of EMS.

Though classified as a benign tumor, EMS exhibits malignancy-like characteristics such as dissemination and invasion due to the ectopic implantation of endometrial tissue [5]. While many studies on the EMS mechanism have primarily focused on the "eutopic endometrium," they have significantly underestimated the influence of the ectopic environment on endometrial tissue. The abnormal proliferation, apoptosis, autophagy, adhesion, invasion, fibrosis, angiogenesis, immune response, and inflammatory response of endometrium at ectopic sites contribute to its growth and development [6]. These pathological processes are not entirely independent of signaling pathway transduction and their upstream and downstream regulatory factors, including Wnt2, which induces β-catenin signaling activity in ectopic endometrial epithelial cells and the expression of growth-related proteins in endometrial epithelial cells, exacerbating the migration and invasion of ectopic endometrial cells [7]. In the ectopic environment, enhanced activation of macrophages, down-regulated phagocytosis, reduced cytotoxicity of NK cells, and variations in the dendritic cell population all contribute to the implantation of EMS ectopic cells [6]. Identifying alterations in endometrial genes and pathways within ectopic environments can aid in understanding the disease's pathogenesis.

With the advancements in high-throughput sequencing technology, genomics and transcriptomics studies, variant functional annotation have elucidated the developmental characteristics of EMS at the gene level, significantly contributing to unveiling the disease's pathogenesis [8]. This study investigated the differences in gene expression, pathway enrichment, and immune microenvironment by integrating gene expression profiles of eutopic and ectopic endometrium across three datasets. The risk scoring model, based on pathway activity, demonstrates high diagnostic value. This research offers novel insights into the pathogenesis and biological characteristics of EMS.

2. Methods

2.1. Data acquisition and processing

The mRNA expression profiles were obtained from the GEO database, encompassing 26 cases of endometriosis and 28 cases of eutopic endometrial samples from the GSE11691, GSE25628, and GSE7305 datasets. These microarray data were collected from the Affymetrix platform and were integrated into a meta-cohort due to the small sample size. Batch correction was implemented using the "combat" algorithm for subsequent analysis. A total of 149 endometriosis and 147 eutopic endometrial samples in GSE141549, and 11 endometriosis and 11 eutopic endometrial samples in GSE5108 were included for diagnostic model validation.

Differential expression analysis was conducted using the limma package, with $P < 0.05$ and $logFC > 1$ as the critical values for differentially expressed genes.

2.2. PPI network

Differentially expressed genes in endometriosis were input into the STRING database, selecting *Homo sapiens* as the organism, and choosing the highest confidence (0.900) for the minimum required interaction score. Disconnected nodes were hidden in the network to generate the PPI network diagram, which was subsequently imported into Cytoscape to obtain the core network by performing topological analysis using the CytoHubba plug-in.

2.3. Pathway enrichment analysis

Alterations in pathway activity among patients were examined through Gene Ontology (GO), KEGG pathway enrichment analysis, Gene Set Enrichment Analysis (GSEA), and Gene Set Variation Analysis (GSVA), calculated using the ClusterProfiler software package. This study primarily employed the GSVA algorithm to evaluate the relative enrichment of a gene set of interest in a sample to observe activity changes in a set of genes (i.e., pathways) corresponding to specific biological conditions.

2.4. Immune cell infiltration level analysis

Infiltrating immune cells in endometrial samples were estimated using the CIBERSORT algorithm, with the default signature matrix permutation number set to 1000. Only samples with $P < 0.05$ were considered suitable for further analysis. The principle of CIBERSORT is to deconvolve the specific marker genes of immune cells and estimate the infiltration proportion of 22 immune cells in a sample.

2.5. Diagnostically valuable signaling pathways identification

Utilizing the limma package, signaling pathways with significantly different activities were identified between ectopic and eutopic samples ($P < 0.05$). LASSO regression analysis was employed to further refine the selection of pathways with the highest diagnostic value, resulting in the establishment of a risk scoring model. For the LASSO regression analysis, the function type was "binomial" and the measurement type was "deviance", with 10-fold cross-validation. The calculation formula is expressed as: risk score = ΣGSVAscorei*βi, where GSVAscorei represents the enrichment score for each pathway, and βi denotes the coefficient for each pathway. A risk score was computed for each patient, and ROC curves were used to assess the diagnostic capability of these risk scores.

2.6. Validation of risk scoring model

To further verify the diagnostic reliability of the risk scoring model, GSVA analysis was conducted on the validation cohort as well. The model pathway was extracted, and the risk score of each sample was calculated based on the model calculation formula. The ROC curve was then utilized to further confirm the diagnostic stability of the model.

2.7. Nomogram construction and verification

The risk score exhibited robust diagnostic value, and patients' disease risks were further assessed. A nomogram was constructed using the "rms" software package to evaluate patients' disease risks, with the predictive value of the nomogram further validated via calibration curve analysis.

2.8. Statistical analysis

Calculations and statistical analyses were performed using R software (https://www.r-project.org/). The Wilcoxon test or the Mann-Whitney *U* test were applied for between-group comparisons of continuous variables related to endometrial distribution. LASSO regression analysis was carried out using the glmnet package. The diagnostic validity of biomarkers was determined through ROC curve analysis. Correlations between infiltrating immune cells and pathway-active biomarkers were analyzed using Spearman's correlation. $P < 0.05$ was considered statistically significant.

3. Results

3.1. Analysis of endometriosis differential genes

Data from 26 cases of ectopic endometrium and 28 cases of eutopic endometrium from GSE11691, GSE25628, and GSE7305 were

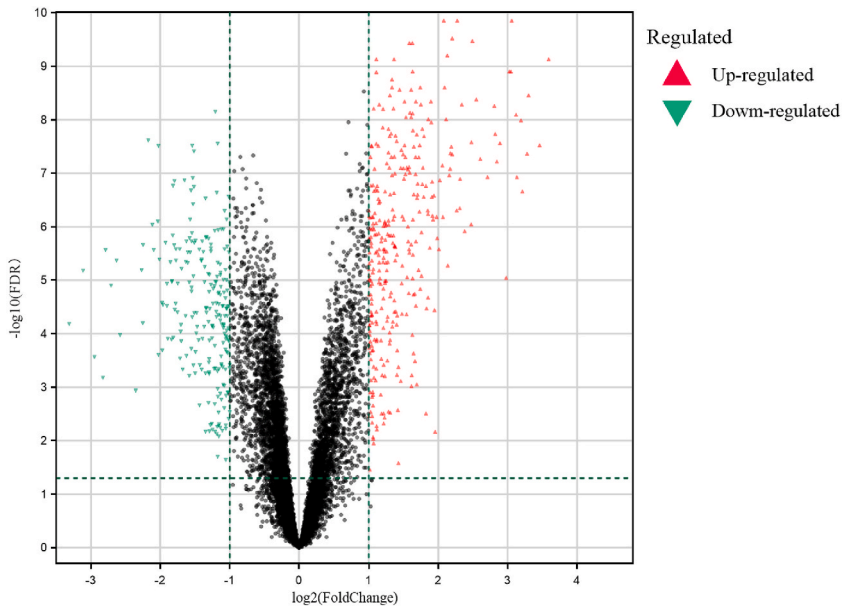

Fig. 1. Volcano plot of DEGs between eutopic and ectopic endometrium in patients with Endometriosis. Each dot represents a gene; red equilateral triangles represent up-regulated genes, blue inverted triangles indicate down-regulated genes, and gray dots signify genes without significant difference in EMS. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

analyzed using the limma package to identify differential genes after batch correction. Ultimately, 568 differential genes were identified, including 338 up-regulated genes and 230 down-regulated genes (Fig. 1).

3.2. Protein-protein interaction (PPI) network construction

To further pinpoint the core genes involved in Endometriosis pathogenesis, the 568 differential genes were entered into the STRING website to obtain the PPI network diagram, which was adjusted using Cytoscape software (Fig. 2A). The core network was acquired based on the connection degree of genes in the network (Fig. 2B), with key genes including CDK1, CCNA2, CDC20, TOP2A, NUSAP1, ASPM, CCNB2, KIF11, KIF20A, and DLGAP5 (Fig. 1). A darker gene color represents a greater number of connections, indicating a core regulatory role for the gene within the network and suggesting a significant relationship with disease onset and progression.

3.3. Pathway enrichment analysis

A functional enrichment analysis was conducted for the 568 differentially expressed genes (DEGs) to better elucidate the pathogenesis of endometriosis (Fig. 3). The biological processes (BP) identified by GO annotation primarily involve extracellular structure organization, epithelial cell proliferation, and extracellular matrix organization. The enriched cellular components (CC) pertain to collagen-containing extracellular matrix, secretory granule lumen, and vesicle lumen. The enriched molecular functions (MF) include extracellular matrix structural constituent, glycosaminoglycan binding, heparin binding, and peptidase inhibitor activity (Fig. 3A). The KEGG pathway analysis indicated that these DEGs participate in various inflammation- and infection-related pathways, such as Complement and coagulation cascades (15 genes, hsa 04610), Cell adhesion molecules (18 genes, hsa 04514), Malaria (9 genes, hsa 05144), *Staphylococcus aureus* infection (12 genes, hsa 05150), Leukocyte transendothelial migration (13 genes, hsa 04670), Focal adhesion (17 genes, hsa 04510), Vascular smooth muscle contraction (13 genes, hsa 04270), and hormone metabolism pathways like steroid hormone biosynthesis (8 genes, hsa 00140) (Fig. 3B).

The GSEA analysis further revealed the primary involvement of enrichment pathways in eutopic endometrium, including cell cycle, DNA replication, oocyte meiosis, proteasome, and spliceosome, signifying more active expression of cytogenetic information in eutopic endometrium (Fig. 3D). The enriched pathways primarily referred to chemokine signaling pathway, complement and coagulation cascades, cytokine-cytokine receptor interaction, and vascular smooth muscle contraction, indicating significant activation of inflammation-related immune signals in the ectopic endometrium (Fig. 3E). The implementation of the GSVA algorithm also disclosed more active homologous recombination, cell cycle, DNA replication, mismatch repair, spliceosome, and other pathways. Additionally, PPAR signaling pathway, Cell adhesion molecules (CAMs), GnRH signaling pathway, and MAPK signaling pathway were found to be significantly active (Fig. 3C). The results from various pathway enrichment analyses demonstrated distinct differences in pathway activity between eutopic and ectopic endometrium, suggesting that immunosuppression of inflammation-related pathways in ectopic endometrium may contribute to disease progression.

3.4. Immune cell infiltration level analysis

The aforementioned results indicated evident crosstalk in the inflammatory immune pathway in ectopic endometrium. To better uncover the microenvironment characteristics of the ectopic endometrium, the differences in the cell composition of the endometrial microenvironment between the two groups of samples were analyzed using the ESTIMATE algorithm. It was found that the stromal

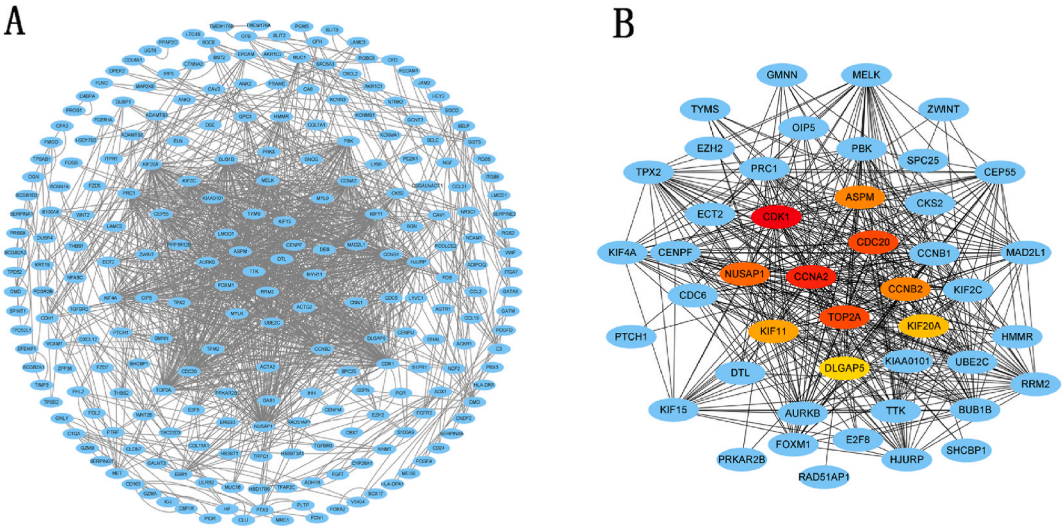

Fig. 2. PPI network. (A) PPI network diagram of endometriosis. (B) Endometriosis core gene network diagram.

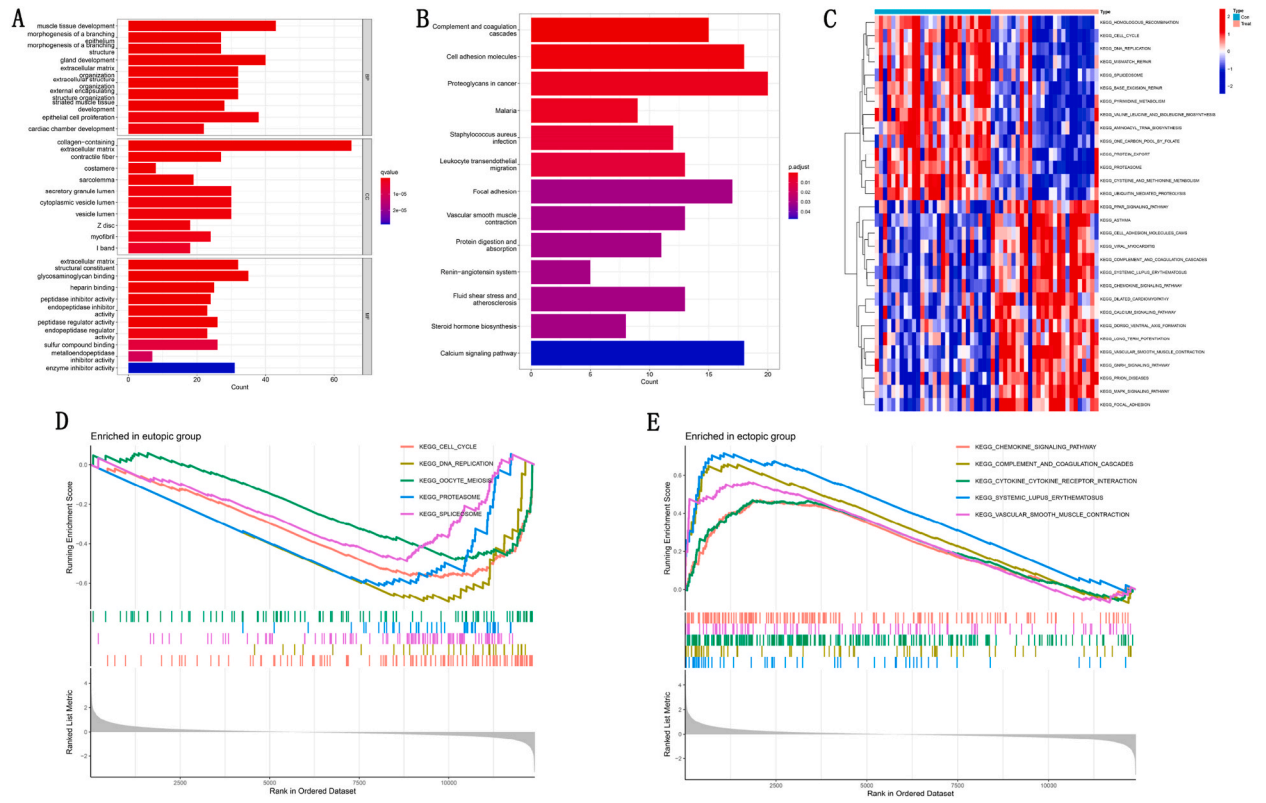

Fig. 3. Differential gene functional enrichment analysis. (A) GO enrichment analysis. (B) KEGG enrichment analysis. (C) GSVA enrichment analysis. (D) Eutopic endometrium GSEA enrichment analysis. (E) Ectopic endometrium GSEA enrichment analysis.

score and immune score in ectopic endometrium samples were significantly higher compared to eutopic endometrium samples, with statistically significant differences ($P < 0.001$) (Fig. 4A). This suggests a higher content of immune cells in ectopic endometrial tissue, as well as more stromal cells and other components related to the endometrial microenvironment.

The CIBERSORT algorithm was further employed to evaluate immune cell infiltration of samples in the two groups. The activated proportions of plasma cells, Macrophages M2, and Mast cells in ectopic endometrial cells were markedly higher than those in eutopic endometrium, accompanied by a higher infiltration tendency of monocytes and neutrophils. In contrast, the eutopic endometrium exhibited significantly higher proportions of T cells follicular helper, T cells regulatory, NK cells resting, NK cells activated, and Dendritic cells compared to the other group (Fig. 4B). Immune checkpoint expression analysis revealed higher expression of LAG3 and PDCD1LG2 in ectopic intima and up-regulated expression of IDO1 in eutopic endometrium (Fig. 4C). These results suggested a significant decrease in the infiltration of innate and adaptive immune cells in ectopic endometrium, representing a state of immunosuppression. Meanwhile, the enrichment of inflammatory cells such as monocytes and neutrophils exacerbated the progression of chronic inflammation, resulting in immune escape.

3.5. Identification of signaling pathways with diagnostic value and nomogram construction

To enhance the accuracy of diagnosing eutopic and ectopic endometrium, differential pathways of samples from the two groups were calculated utilizing the limma package and GSVA algorithm. LASSO regression determined six gene sets with the highest diagnostic value: KEGG_HOMOLOGOUS_RECOMBINATION, KEGG_BASE_EXCISION_REPAIR, KEGG_DNA_REPLICATION, KEGG_P53_SIGNALING_PATHWAY, KEGG_ADHERENS_JUNCTION, and KEGG_SNARE_INTERACTIONS_IN_VESICULAR_TRANSPORT (Fig. 5A and B). Subsequently, a risk score model was constructed, where risk score = ΣGSVAscorei *βi. (Table 1). In the training group, the risk score for the ectopic endometrium group was significantly higher than that of the eutopic endometrium group (Fig. 5E), with KEGG_HOMOLOGOUS_RECOMBINATION, KEGG_BASE_EXCISION_REPAIR, KEGG_DNA_REPLICATION, and KEGG_P53_SIGNALING_PATHWAY exhibiting lower expression in ectopic endometrium, while KEGG_ADHERENS_JUNCTION and KEGG_SNARE_INTERACTIONS_IN_VESICULAR_TRANSPORT displayed higher expression. The ROC curve of the risk scoring model had an AUC of 0.854, surpassing 0.75, which validated the diagnostic value of the risk scoring model (Fig. 5C). Based on the six gene sets of the risk model, a nomogram was constructed to quantify the risk of patients developing endometriosis (Fig. 5F). A calibration curve suggested that the nomogram demonstrated high prediction accuracy (Fig. 5G).

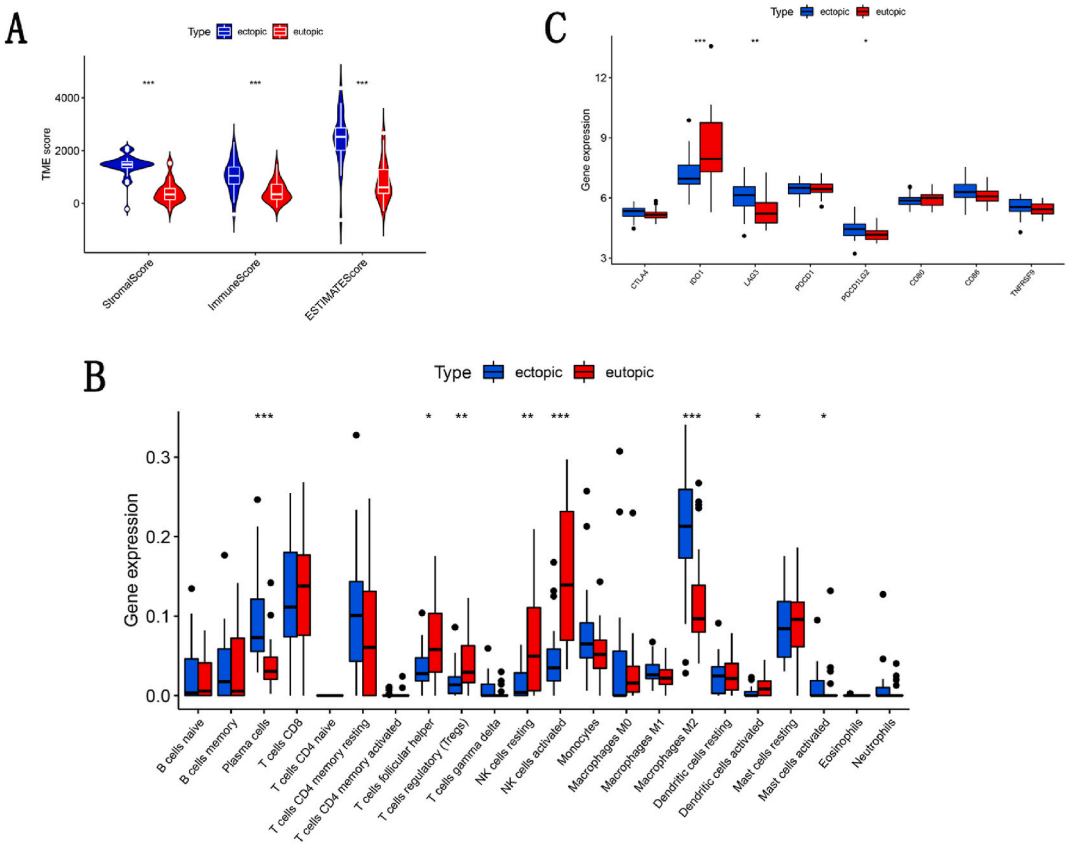

Fig. 4. Analysis of immune cell infiltration. (A) Tumor microenvironment score. (B) Analysis of immune cell infiltration. (C) Analysis of differential expression of immune checkpoints.

3.6. Risk scoring model validation

The risk model validation was conducted on 149 endometriosis and 147 eutopic endometrial samples in GSE141549 of the two groups from GSE141549 and 11 cases of the two groups from GSE5108, respectively. The samples were processed using the GSVE algorithm, in which the risk score of each sample was calculated based on the model calculation formula. In comparison with eutopic endometrium, the channel risk scores of ectopic endometrium were significantly increased (Fig. 6B and D). The ROC curve dedicated that the AUC values of the verification group were all above 0.75 (Fig. 6A and C), demonstrating the diagnostic stability of the risk model once more.

3.7. Risk score and immune correlation analysis

The immune correlation analysis of risk score-related pathways revealed a negative correlation of KEGG_ADHERENS_JUNCTION with T cells regulatory and T cells CD8, and a positive correlation of KEGG_BASE_EXCISION_REPAIR with T cells regulatory, T cells follicular helper, NK cells resting, NK cells activated, and Dendritic cells activated, while a negative correlation with Plasma cells and Macrophages M2. KEGG_DNA_REPLICATION was positively correlated with T cells regulatory and NK cells activated, and negatively correlated with Plasma cells and Macrophages M2. KEGG_HOMOLOGOUS_RECOMBINATION was positively correlated with T cells regulatory, NK cells resting, and NK cells activated, and negatively correlated with Plasma cells and Macrophages M2. KEGG_P53_SIGNALING_PATHWAY displayed a positive correlation with NK cells resting, NK cells activated, and Dendritic cells activated, and a negative correlation with Plasma cells and Macrophages M2. KEGG_SNARE_INTERACTIONS_IN_VESICULAR_TRANSPORT was positively correlated with Macrophages M2, and negatively correlated with NK cells activated and T cells CD8, with statistically significant differences (Fig. 7A). The risk score was analyzed for immune correlation, revealing a positive correlation of the Endometriosis risk score with Macrophages M2 and Plasma cells, and a negative correlation with NK cells resting, NK cells activated, and T cells regulatory (Fig. 7B–G).

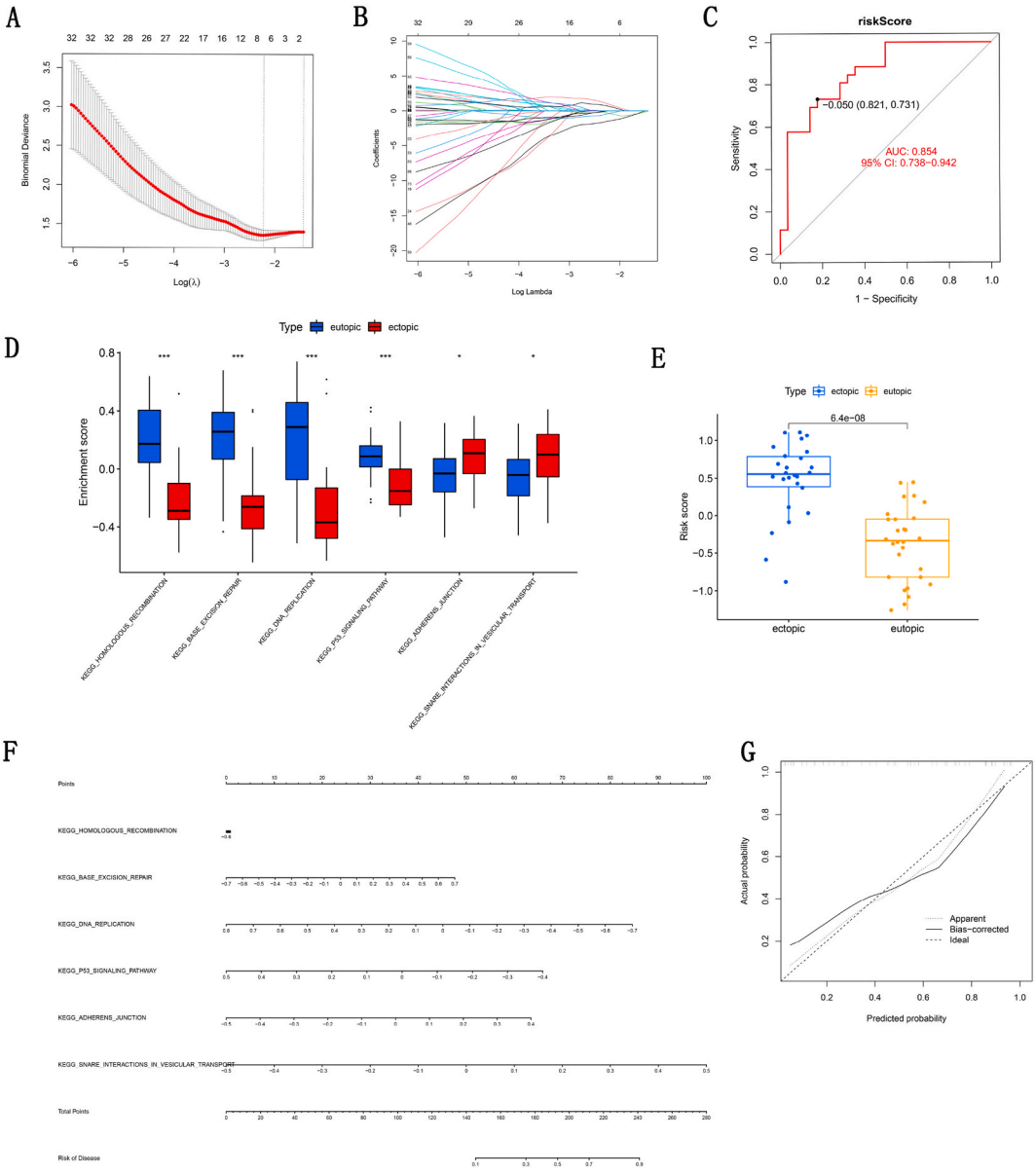

Fig. 5. The process of screening diagnostic signaling pathways for Endometriosis. (A) Determination of the optimal penalty coefficient λ for LASSO regression. (B) Determination of the gene coefficient of the model. (C) ROC analysis of the risk score of the signaling pathway diagnostic model. (D) The scoring boxplot of six gene sets with the most diagnostic values in eutopic endometrium and ectopic endometrium. (E) The risk score difference analysis between eutopic endometrium and ectopic endometrium.(F) Nomogram construction. (G) Calibration curve to evaluate the risk prediction capability of the nomogram.

Table 1
Risk score sheet.

Channels	βi
KEGG_HOMOLOGOUS_RECOMBINATION	−0.0233436380671805
KEGG_BASE_EXCISION_REPAIR	−0.350182582932821
KEGG_DNA_REPLICATION	−0.382228926089417
KEGG_P53_SIGNALING_PATHWAY	−1.45995041399291
KEGG_ADHERENS_JUNCTION	0.664788627626051
KEGG_SNARE_INTERACTIONS_IN_VESICULAR_TRANSPORT	0.965206706652808

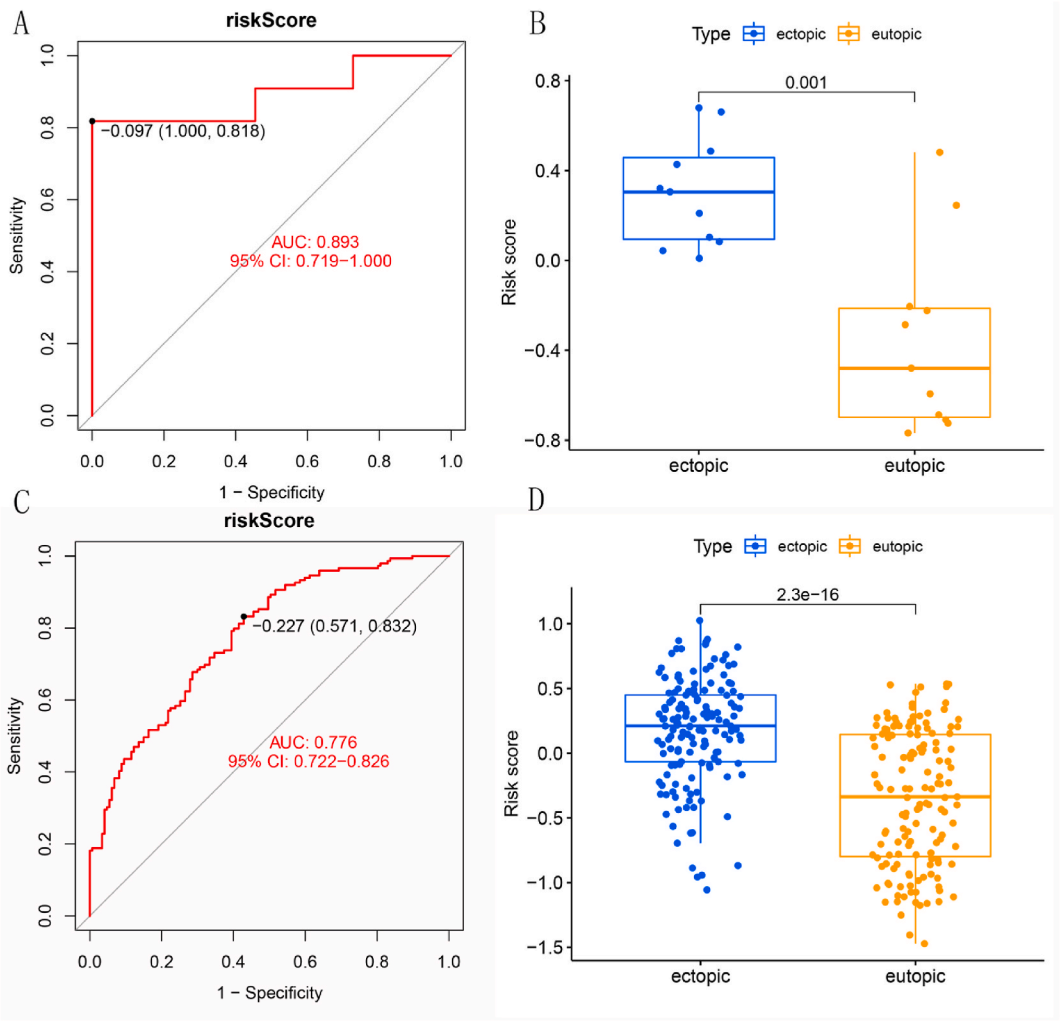

Fig. 6. Risk score model validation. (A) Validation group GSE5108 ROC validation. (B) Risk score difference analysis of eutopic endometrium and ectopic endometrium in validation group GSE5108. (C) Validation group GSE141549 ROC validation. (D) Risk score difference analysis of eutopic endometrium and ectopic endometrium in validation group GSE141549.

4. Discussion

Endometriosis (EMS) is considered a critical disease affecting fertile women, and a permanent treatment plan remains elusive due to its unclear pathogenesis. Current treatment objectives mainly focus on pain relief, symptom alleviation, and improving the quality of life. Therefore, an in-depth exploration of EMS pathogenesis could provide effective guidance for disease treatment. In this study, based on the differential analysis of gene expression profiles of eutopic and ectopic endometrium, 568 differentially expressed genes were identified, suggesting a significant difference in gene expression between the two groups (Fig. 1). KEGG enrichment analysis demonstrated that these genes were involved in inflammation-related signaling pathways, primarily exhibiting a pro-inflammatory effect in EMS with abnormal expression (Fig. 3B). A PPI network analysis was further conducted on the differential genes, revealing the central location of CDK1, CCNA2, CDC20, TOP2A, NUSAP1, ASPM, CCNB2, KIF11, KIF20A, and DLGAP5 within the network (Fig. 2A and B). These genes are mainly associated with cell cycle regulation and likely play a crucial role in cell proliferation and survival, thereby influencing the occurrence and development of EMS genes. Studies have demonstrated the up-regulated cyclin-dependent kinase 1 (CDK1) in ectopic endometrial tissue, which induces proliferation of endometriosis cells [9]; The number of lesions in rectosigmoid tissue is proportional to the expression of Topoisomerase II alpha (TOP2A) [10]; inhibition of Kinesin family member 20A (KIF20A) reduces endometriosis lesions in an endometriosis mouse model [11]; while no related studies or reports on other core genes in the mechanism of EMS are available, they all play roles in tumor growth regulation: Cyclin A2 (CCNA2) promotes the invasion and migration of non-small cell lung cancer cells [12]; Cell division cycle 20 (CDC20), as a subunit of the mitotic checkpoint complex (MCC), plays a critical role in anaphase and the exit process from mitosis, the activation of which facilitates tumor progression and stemness in gastric cancer [13]; NUSAP1 (Nucleolar spindle associated protein 1) contributes to the proliferation, invasion, and

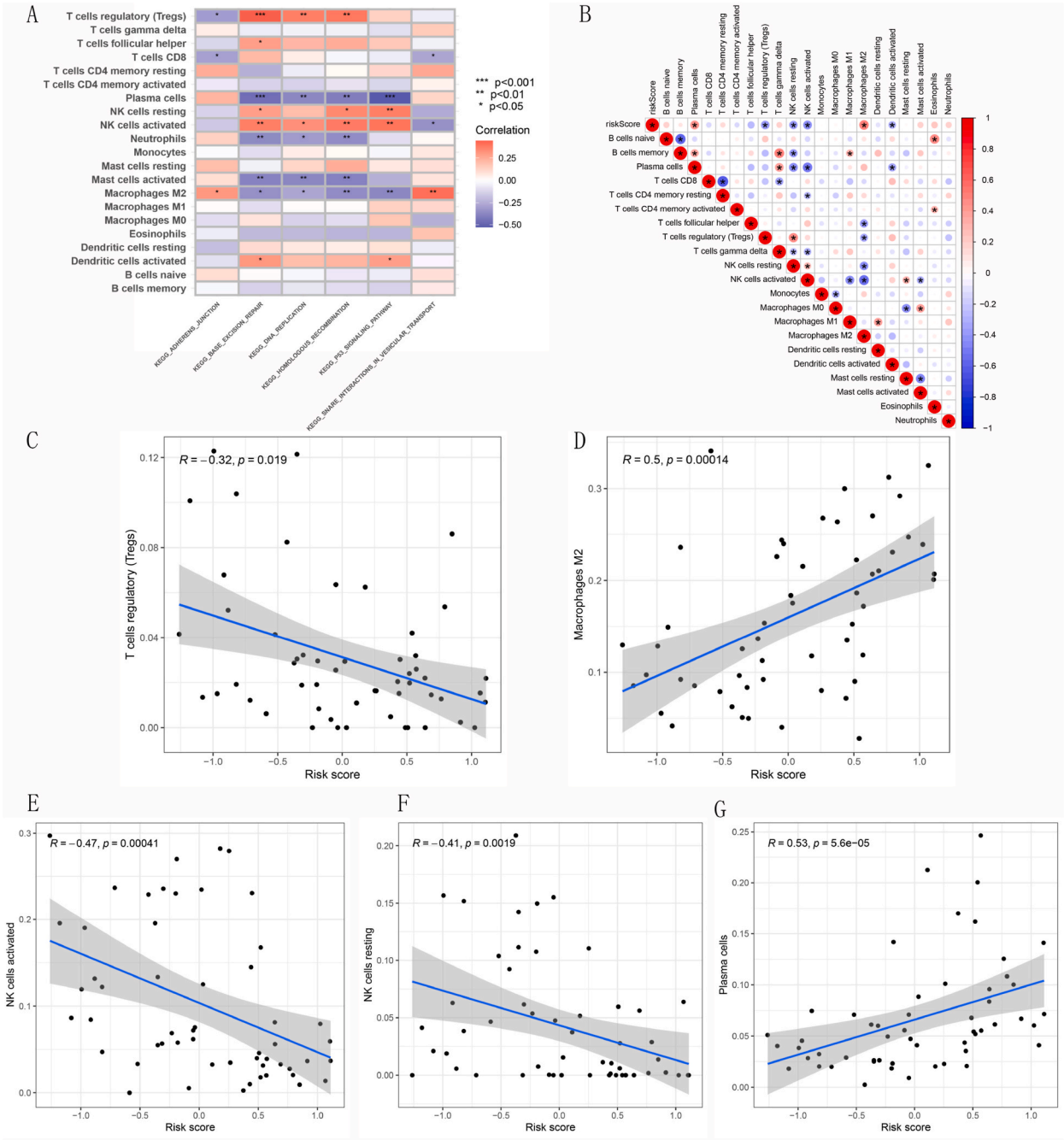

Fig. 7. Risk score and immune correlation analysis. (A) Risk-related pathway and immune correlation analysis. (B) Risk score and immune Spearman's analysis. (C) EMS risk score and Spearman's analysis of T cells regulatory. (D) EMS risk score and Spearman's analysis of Macrophages M2. (E) EMS risk score and Spearman's analysis of NK cells activated. (F) EMS risk score and Spearman's analysis of NK cells resting. G EMS risk score and Spearman's analysis of Plasma cells.

migration of nephroblastoma through the PI3K/Akt signaling pathway [14]; the overexpression of spindle microtubules (ASPM) increases tumor invasiveness [15]; upregulation of Cyclin B2 (CCNB2) [16] exacerbates lymphovascular invasion in breast cancer; these findings suggest the potential target therapy value of these core genes that can serve as diagnostic biomarkers for EMS. The abnormal activation of signaling pathways typically induces phenotypic alterations in normal tissues and progression toward malignant features. To further explore the disease characteristics of EMS, differences in pathway activity between ectopic and eutopic endometrium were detected using GSEA and GSVA algorithms (Fig. 3C–E). It was observed that inflammation-related pathways, such as cytokine-cell receptor signaling pathway and chemokine signaling pathway, were significantly more active in ectopic endometrium compared to

normal endometrium. Multiple studies have demonstrated the association of endometriosis occurrence with chemokine/cytokine interactions, for example, estrogen-regulated chemokine elevation promotes the migration of bone marrow-derived stem cells and mesenchymal stem cells, playing key roles in ectopic endometrium [17,18]; cytokines regulate mitosis and angiogenesis, accelerating the implantation and growth of ectopic endometrium on the peritoneal surface [19]. Meanwhile, this process triggers inflammatory cascades, inducing the secretion of angiogenic factors, tumor necrosis factors, and monocyte chemokine, which contribute to the adhesion of ectopic endometrium [20]. Activation of the PPAR pathway inhibits the expression of adhesion factors during inflammation, diminishing the adhesion of endometrioid tissue to ectopic sites [21], while reducing inflammatory secretion of endometrial basal cells [22]. The inhibited VEGF expression results in decreased microvessel density [23], helping to prevent endometriosis. The aberrant activation of the MAPK signaling pathway participates in the inflammatory response, cell proliferation, cell differentiation, and apoptosis in endometriosis [24]. These findings suggest that inflammation-related signaling pathways play a promoting role in EMS, accelerating ectopic intimal cell proliferation, angiogenesis, and adhesion and fixation. Chronic inflammation progression may potentially create an immunosuppressive environment to resist immune cell attacks.

Pathway data are important for understanding the relationship between genes, proteins and many other molecules in livng organisms [25]. Significant variations in signaling pathway activity were observed between eutopic endometrium and ectopic endometrium, based on which an attempt was made to diagnose both types. A total of 106 pathways exhibiting differences between the two groups were identified using the limma package and the GSVA algorithm. Six gene sets with the highest diagnostic value were determined through LASSO regression analysis: homologous recombination, base excision repair, DNA replication, P53 signaling pathway, adherens junction, and snare interactions in vesicular transport. These gene sets primarily involve DNA damage repair-related pathways and intercellular communication pathways (Table 1). DNA damage occurs approximately 10,000 times per cell each day [26] due to misincorporation, deamination, depurination, methylation, and oxidation of DNA bases. The presence of DNA damage or DNA replication stress activates the DNA damage response (DDR) pathway [27], which regulates genome stability by repairing damaged DNA or removing defective cells through programmed cell death [28]. The tumor suppressor protein P53 recognizes the presence of DNA double-strand breaks (DSBs) and activates signaling pathways that regulate tumor progression and promote apoptosis. Mutations in the P53 gene affect DNA damage repair and promote tumor progression [29]. Homologous recombination (HRR) can repair DSB damage, and studies have indicated that Oxabicycloheptene sulfonate (OBHS), acting as an estrogen receptor alpha (ERα) ligand, impairs RNA polymerase II (Pol II) loading at the promoters of estrogen-responsive HRR genes. The downregulation of homologous recombination results in increased DNA damage, apoptosis, and cell cycle arrest, effectively preventing and treating in vivo estrogen-dependent endometriosis [30]. Base excision repair (BER), a vital DNA repair approach, can specifically excise N-β-glycosidic bonds on damaged nucleotides, remove small damaged DNA fragments, and synthesize new fragments by DNA polymerase I. These new fragments are eventually connected by DNA ligase into a newly repaired DNA strand [26]. The activity of these pathways is significantly down-regulated in ectopic endometrium, indicating a defect or inhibition of the DNA damage repair pathway in ectopic endometrium. This results in a loss of control of ectopic endometrial cells during genetic information expression, leading to cell proliferation. Adherens junctions, mediated by the calcium-dependent transmembrane protein family of cadherins, form homodimers in the intercellular space, serving to keep adjacent cells together and maintain cell and tissue polarity.

In comparison to eutopic endometrium, E-cadherin-negative epithelial cells increased in peritoneal and ovarian ectopic endometrial tissues [31], while expressions of T-cadherin and E-cadherin decreased [32]. The reduction in adhesion junctions promoted epithelial-mesenchymal transition, increasing the invasiveness of ectopic endometrial cells and contributing to the occurrence of EMS. Extracellular vesicles, small membrane vesicles ranging from 30 to 1000 nm in size, promote disease progression by transferring cargo molecules and serve as key mediators of intercellular communication in the microenvironment [33]. Relevant studies have indicated that endometriosis cyst stromal cell (ECSC)-derived extracellular vesicles facilitate remodeling of the ectopic environment for further implantation and metastasis [34]. The active expressions of these two signaling pathways in ectopic endometrium likely promote the malignant development of EMS by accelerating adhesion and colonization of endometrium. The varied activities of these critical pathways in EMS highlight potential therapeutic targets. Risk score diagnostic model was constructed based on these six pathways (Fig. 5A and B). ROC curve analysis showed an AUC value of 0.854 in this model (Fig. 5C), indicating high diagnostic accuracy. Two validation cohorts confirmed the diagnostic reliability of the model, providing a certain clinical reference value.

To explore the connection between EMS and immune factors, further analysis was conducted on the characteristics of the immune microenvironment in EMS, including immune microenvironment exploration and immune cell infiltration analysis (Fig. 4A–C). Higher scores were observed for stromal cells and immune cells in ectopic endometrial cells, with increased infiltration of plasma cells, M2 macrophages, and activated mast cells, and decreased infiltration of immune killer cells such as follicular helper T cells, regulatory T cells, resting NK cells, activated NK cells, and dendritic cells. This indicates significant immunosuppression in ectopic endometrium. Additionally, the EMS risk score was directly proportional to the infiltration degree of plasma cells and M2 macrophages, while inversely proportional to the infiltration degree of regulatory T cells, resting NK cells, and activated NK cells (Fig. 7A–G). The infiltration level of corresponding immune cells can be estimated based on the calculation of the EMS risk score. Studies suggest that EMS occurrence may be related to decreased phagocytosis and cytotoxicity of immune cells, along with increased pro-inflammatory and chemotactic effects [35]. Chemokines in the ectopic environment specifically activate and induce differentiation from B cells to plasma cells [36], as well as activate related mast cells to promote chemokine pathways and focal fibrosis. Due to the inflammatory nature of the disease, immune response suppression occurs by inhibiting the activation of effector T cells [37], leading to inhibited cytotoxic activity of NK cells [35]. This promotes the escape of endometrial tissue from immune surveillance by NK cells, fostering the development of endometriosis [38]. M2 macrophages facilitate angiogenesis in ectopic endometrial lesions [39], contributing to the survival of endometriosis foci. These findings align with the presented conclusions.

In addition, our study has certain limitations. Firstly, we did not conduct model validation and experimental validation in more

realistic cohorts. Secondly, the diagnostic performance of the diagnostic model may be improved by adjusting the parameters and analyzing more samples. In summary, to address these shortcomings, we will further verify the accuracy of the model by expanding the sample size in follow-up studies, and explore the biological functions of these pathways in endometriosis through more basic experiments. We will also continue to optimize the model to promote the identification of more accurate diagnostic markers for endometriosis.

5. Conclusion

Our study reveals the characteristics of gene expression and immune cell infiltration in endometriosis, and the pathway risk diagnostic model, comprising homologous recombination, base excision repair, DNA replication, P53 signaling pathway, adherens junction, and snare interactions in vesicular transport, demonstrates high diagnostic value. This research offers novel insights and alternative strategies for the clinical diagnosis and treatment of EMS.

Author contribution statement

Yi Zhang: Conceived and designed the experiments; Analyzed and interpreted the data; Wrote the paper.

Lulu Wu: Conceived and designed the experiments; Performed the experiments; Contributed reagents, materials, analysis tools or data.

Xiang Wen, Xiuwei Lv: Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data.

Data availability statement

Data associated with this study has been deposited at GEO database under the accession number (GSE11691, GSE25628, GSE7305, GSE5108 and GSE141549).

Declaration of competing interest

The authors declare that no potential conflict of interest that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank Qi Liu (master student at English major of foreign language college, Huanggang Normal University) for helping with translation.

References

[1] A.L. Shafrir, L.V. Farland, D.K. Shah, H.R. Harris, M. Kvaskoff, K. Zondervan, et al., Risk for and consequences of endometriosis: a critical epidemiologic review, Best Pract. Res. Clin. Obstet. Gynaecol. 51 (2018) 1–15.
[2] J. Feng, S. Zhang, J. Chen, J. Yang, J. Zhu, Long-term trends in the incidence of endometriosis in China from 1990 to 2019: a joinpoint and age-period-cohort analysis, Gynecol. Endocrinol. 37 (2021) 1041–1045.
[3] S.E. Bulun, B.D. Yilmaz, C. Sison, K. Miyazaki, L. Bernardi, S. Liu, et al., Endometriosis, Endocr. Rev. 40 (2019) 1048–1079.
[4] C. Xie, C. Lu, Y. Liu, Z. Liu, Diagnostic gene biomarkers for predicting immune infiltration in endometriosis, BMC Wom. Health 22 (2022).
[5] F. de Almeida Asencio, H.A. Ribeiro, P. Ayrosa Ribeiro, M. Malzoni, L. Adamyan, A. Ussia, et al., Symptomatic endometriosis developing several years after menopause in the absence of increased circulating estrogen concentrations: a systematic review and seven case reports, Gynecol. Surg. 16 (2019).
[6] S.W. Hung, R. Zhang, Z. Tan, J.P.W. Chung, T. Zhang, C. Wang, Pharmaceuticals targeting signaling pathways of endometriosis as potential new medical treatment: a review, Med. Res. Rev. 41 (2021) 2489–2564.
[7] H. Zhang, J. Xue, M. Li, X. Zhao, D. Wei, C. Li, Metformin regulates stromal-epithelial cells communication via Wnt2/beta-catenin signaling in endometriosis, Mol. Cell. Endocrinol. 413 (2015) 61–65.
[8] H. Zhou, T. Arapoglou, X. Li, Z. Li, X. Zheng, J. Moore, A. Asok, S. Kumar, E.E. Blue, S. Buyske, et al., FAVOR: functional annotation of variants online resource and annotator for variation across the human genome, Nucleic Acids Res. 51 (D1) (2023) D1300–D1311.
[9] P. Chen, R. Mamillapalli, S. Habata, H.S. Taylor, Endometriosis cell proliferation induced by bone marrow mesenchymal stem cells, Reprod. Sci. 28 (2021) 426–434.
[10] M.A. Bassi, V. Arias, F.N. D'Amico, B.Y. Gueuvoghlanian-Silva, M.S. Abrao, S. Podgaec, Deep invasive endometriosis lesions of the rectosigmoid may Be related to alterations in cell kinetics, Reprod. Sci. 22 (2015) 1122–1128.
[11] H. Ferrero, A. Corachan, A. Quinonero, C. Bougeret, P. Pouletty, A. Pellicer, et al., Inhibition of KIF20A by BKS0349 reduces endometriotic lesions in a xenograft mouse model, Mol. Hum. Reprod. 25 (2019) 562–571.
[12] J.S. Ruan, H. Zhou, L. Yang, L. Wang, Z.S. Jiang, S.M. Wang, CCNA2 facilitates epithelial-to-mesenchymal transition via the integrin alphavbeta3 signaling in NSCLC, Int. J. Clin. Exp. Pathol. 10 (2017) 8324–8333.
[13] W. Li, H. Chen, Z. Wang, J. Liu, X. Lei, W. Chen, Chromobox 4 (CBX4) promotes tumor progression and stemness via activating CDC20 in gastric cancer, J. Gastrointest. Oncol. 13 (2022) 1058–1072.
[14] B. Luo, S. Feng, T. Li, J. Wang, Z. Qi, Y. Zhao, et al., Transcription factor HOXB2 upregulates NUSAP1 to promote the proliferation, invasion and migration of nephroblastoma cells via the PI3K/Akt signaling pathway, Mol. Med. Rep. 25 (2022).
[15] Z. Feng, J. Zhang, Y. Zheng, J. Liu, T. Duan, T. Tian, Overexpression of abnormal spindle-like microcephaly-associated (ASPM) increases tumor aggressiveness and predicts poor outcome in patients with lung adenocarcinoma, Transl. Cancer Res. 10 (2021) 983–997.
[16] A.I. Aljohani, M.S. Toss, K.A. El-Sharawy, S. Mirza, G.R. Ball, A.R. Green, et al., Upregulation of Cyclin B2 (CCNB2) in breast cancer contributes to the development of lymphovascular invasion, Am. J. Cancer Res. 12 (2022) 469–489.
[17] W. Zhang, X. Li, H. Li, X. Lu, J. Chen, L. Li, et al., 17 beta-estradiol promotes bone marrow mesenchymal stem cell migration mediated by chemokine upregulation, Biochem. Biophys. Res. Commun. 530 (2020) 381–388.

[18] M. Irene, M. Ramanaiah, C. Emine, S.E. Gulcin, S.T. Hugh, Bone marrow stem cell chemotactic activity is induced by elevated CXCl12 in endometriosis, Reprod. Sci. 24 (2017).
[19] N.C. Llarena, E.G. Richards, A. Priyadarshini, D. Fletcher, T. Bonfield, R.L. Flyckt, Characterizing the endometrial fluid cytokine profile in women with endometriosis, J. Assist. Reprod. Genet. 37 (2020).
[20] D.I. Lebovic, F. Bentzien, V.A. Chao, E.N. Garrett, Y.G. Meng, R.N. Taylor, Induction of an angiogenic phenotype in endometriotic stromal cell cultures by interleukin-1 beta, Mol. Hum. Reprod. 6 (2000) 269–275.
[21] A. Vallee, J.N. Vallee, A. Le Blanche, Y. Lecarpentier, PPARgamma agonists: emergent therapy in endometriosis, Pharmaceuticals 14 (2021).
[22] A.L. Rocha, F.M. Reis, F. Petraglia, New trends for the medical treatment of endometriosis, Expet Opin. Invest. Drugs 21 (2012) 905–919.
[23] A. Nenicu, C. Korbel, Y. Gu, M.D. Menger, M.W. Laschke, Combined blockade of angiotensin II type 1 receptor and activation of peroxisome proliferator-activated receptor-gamma by telmisartan effectively inhibits vascularization and growth of murine endometriosis-like lesions, Hum. Reprod. 29 (2014) 1011–1024.
[24] G. Bora, A. Yaba, The role of mitogen-activated protein kinase signaling pathway in endometriosis, J. Obstet. Gynaecol. Res. 47 (2021) 1610–1623.
[25] H. Zhou, J. Jin, H. Zhang, B. Yi, M. Wozniak, L. Wong, IntPath–an integrated pathway gene relationship database for model organisms and important pathogens, BMC Syst. Biol. 6 (Suppl 2) (2012) S2.
[26] J.F. Alhmoud, J.F. Woolley, M.A. Al, M.I. Malki, DNA damage/repair management in cancers, Cancers 12 (2020).
[27] A. Marechal, L. Zou, DNA damage sensing by the ATM and ATR kinases, Cold Spring Harbor Perspect. Biol. 5 (2013).
[28] D. Hanahan, R.A. Weinberg, Hallmarks of cancer: the next generation, Cell 144 (2011) 646–674.
[29] T.D. Halazonetis, V.G. Gorgoulis, J. Bartek, An oncogene-induced DNA damage model for cancer development, Science 319 (2008) 1352–1355.
[30] J. Wu, J. Yan, P. Fang, H.B. Zhou, K. Liang, J. Huang, Three-dimensional oxabicycloheptene sulfonate targets the homologous recombination and repair programmes through estrogen receptor alpha antagonism, Cancer Lett. 469 (2020) 78–88.
[31] R. Gaetje, S. Kotzian, G. Herrmann, R. Baumann, A. Starzinski-Powitz, Nonmalignant epithelial cells, potentially invasive in human endometriosis, lack the tumor suppressor molecule E-cadherin, Am. J. Pathol. 150 (1997) 461–467.
[32] I. Biyik, U. Kalkan, S. Simsek, The deep infiltrating endometriosis tissue has lower T-cadherin, E-cadherin, progesterone receptor and oestrogen receptor than endometrioma tissue, Taiwan J. Obstet. Gynecol. 60 (2021) 1059–1065.
[33] Y.L. Hsu, J.Y. Hung, W.A. Chang, Y.S. Lin, Y.C. Pan, P.H. Tsai, et al., Hypoxic lung cancer-secreted exosomal miR-23a increased angiogenesis and vascular permeability by targeting prolyl hydroxylase and tight junction protein ZO-1, Oncogene 36 (2017) 4929–4942.
[34] J. Qiu, Y. Lin, X.Y. Tang, Y. Ding, X.F. Yi, K.Q. Hua, Extracellular vesicle-mediated transfer of the lncRNA-TC0101441 promotes endometriosis migration/invasion, Exp. Cell Res. 388 (2020), 111815.
[35] G. Zou, J. Wang, X. Xu, P. Xu, L. Zhu, Q. Yu, et al., Cell subtypes and immune dysfunction in peritoneal fluid of endometriosis revealed by single-cell RNA-sequencing, Cell Biosci. 11 (2021) 98.
[36] A. Hever, R.B. Roth, P. Hevezi, M.E. Marin, J.A. Acosta, H. Acosta, et al., Human endometriosis is associated with plasma cells and overexpression of B lymphocyte stimulator, Proc. Natl. Acad. Sci. U. S. A. 104 (2007) 12451–12456.
[37] A. Delbandi, M. Mahmoudi, A. Shervin, Z. Moradi, T. Arablou, A.H. Zarnani, Higher frequency of circulating, but not tissue regulatory T cells in patients with endometriosis, J. Reprod. Immunol. 139 (2020), 103119.
[38] A. Fukui, C. Mai, S. Saeki, M. Yamamoto, R. Takeyama, T. Kato, et al., Pelvic endometriosis and natural killer cell immunity, Am. J. Reprod. Immunol. 85 (2021), e13342.
[39] M. Bacci, A. Capobianco, A. Monno, L. Cottone, F. Di Puppo, B. Camisa, et al., Macrophages are alternatively activated in patients with endometriosis and required for growth and vascularization of lesions in a mouse model of disease, Am. J. Pathol. 175 (2009) 547–556.