Identification and validation of M2 macrophage-related genes in endometriosis

Hongyan Ding, Hongge Xu, Ting Zhang, Can Shi [*]

Department of Obstetrics and Gynecology, The Affiliated Huai'an No. 1 People's Hospital of Nanjing Medical University, Huai'an, Jiangsu 223300, PR China

A R T I C L E   I N F O

A B S T R A C T

Aims: M2 macrophage is believed to play an important role in the development of endometriosis. This study aimed to identify several key genes related to the M2 macrophage in endometriosis. *Method:* Differential expressed genes between endometriosis and non-endometriosis were identified based on three microarray datasets from the Gene Expression Omnibus database. Gene modules significantly associated with M2 macrophage were identified from the weighted gene co-expression network analysis. Furthermore, by intersecting the differential expressed genes and M2 macrophage-associated module genes, M2 macrophage-related genes in endometriosis were identified. Functional analyses of the Gene Ontology and Kyoto Encyclopedia of Genes and Genomes for these genes were then performed. Following, the least absolute shrinkage and selection operator, random forest, and receiver operating characteristic curves were further conducted to identify the key M2 macrophage-related genes in endometriosis. Finally, the expressions of key genes in endometriosis, as well as their correlations with M2 macrophages were verified in an independent validation cohort.
Results: Totally, 185 M2 macrophage-related genes were identified, and they were mainly enriched in functions associated with the cell cycle, oocyte maturation, and immune response. Following machine learning algorithms, eight key genes were selected in the endometriosis: *PGR, OLFM4, PIP5K1B, CCNA1, BRIP1, CADM1, PRAME,* and *GCNT1*. The eight key genes were confirmed to be negative with M2 macrophage infiltration levels. Furthermore, the expression levels of these genes were significantly lower in the middle secretory stage while relevantly higher in the proliferative stage. The validation analysis also showed similar outcomes with the above results.
Conclusion: Eight M2 macrophage-related genes were identified as potential biomarkers of endometriosis, providing novel understanding of immune cells in the endometriosis.

1. Introduction

Endometriosis is a chronic estrogen-dependent inflammatory disease defined as lesions of endometrial-like tissue outside the uterus [1,2]. It is estimated that endometriosis affects 10% of women of reproductive age [3], equating to 190 million women worldwide [4]. The diagnosis of endometriosis is mainly through laparoscopy and the common treatment method is complete laparoscopic surgical

resection of ectopic endometrial tissues [5]. Combined hormonal contraceptives with or without nonsteroidal anti-inflammatory drugs are first-line options in therapy of endometriosis and have tolerable adverse effect profile [6]. However, it is still unsatisfying that the progression of endometriosis is relatively slow. Usually, it takes 7–10 years before the onset of the endometriosis symptoms, which delays the diagnosis and optimal treatment timing [5]. Moreover, the pathogenesis of endometriosis is poorly understood. Therefore, it is urgent to find the possible diagnostic biomarkers and molecular mechanisms for endometriosis to improve the early diagnosis and treatment of endometriosis.

Recently, the disturbance of the immune microenvironment is demonstrated to play a pivotal role in the pathophysiology and development of endometriosis [7]. Macrophages are ubiquitous immune cells and play important roles in both innate and acquired immunity. The macrophages are abundant throughout female reproductive tissues like ovary, uterus, oviduct and mammary gland [8], and the dysregulation in the endometriosis is common [9]. Increasing of macrophages stimulate cytotoxic T helper cells to release excessive inflammatory cytokines, resulting in promoting the shape of a pro-inflammatory environment in the endometrium and eventually developing an endometriosis [7]. Alternatively activated or M2 macrophage, polarized by Th2 cytokines such as IL-4 and IL-13 and producing anti-inflammatory cytokines such as IL-10 and TGF-β, are anti-inflammatory and immunoregulatory in the diseases [10]. Moreover, aberrant increase and activation of anti-inflammatory M2 macrophage stimulate the abnormal gene expression that is associated with the ectopic endometrium [7]. Moreover, in a deeper investigation, M2 macrophage is elevated in III-IV endometriosis [11]. In addition, specific phenotypes and functions of macrophages also correspond to each phase of the menstrual cycle [12]. While short menstrual cycle is one of the features of high risk of endometriosis [3], and endometrial diseases have a menstrual cycle-dependent epigenetic profile [13]. All these evidences demonstrate that macrophages may play important role in endometriosis progression. Considering the contribution of M2 macrophages in the immunopathogenesis of endometriosis, the identification of M2 macrophage-associated genes may help in the search for potential diagnostic markers of the diseases.

Thanks to the rapid development of bioinformatics, it is convenient to recognize and explore the role of specific genes in diseases. A relevant bioinformatics analysis has reported five M2 macrophage-related genes [14]. Distinguishing from this, the present study aimed to identify key genes related to M2 macrophage through machine learning algorithms. These genes are novel and initially proposed by bioinformatics analyses. In addition to confirming the aberrant expression of these genes in endometriosis and their relationships with M2 macrophages, this study also explored the expression changes of these genes during menstrual cycles in endometriosis. These findings may provide a novel insight for the diagnosis and therapy for patients with endometriosis at the molecular level.

2. Method

2.1. Data source

Based on the Gene Expression Omnibus (GEO) database [15] (http://www.ncbi.nlm.nih.gov/geo/), this study screened microarray datasets with the following criteria: 1) endometriosis is used as keyword; 2) organisms are homo sapiens and samples are endometrium tissues; 3) the sample size is greater than 20. Finally, four datasets (GSE51981, GSE6364, GSE7305 and GSE135485) were obtained.

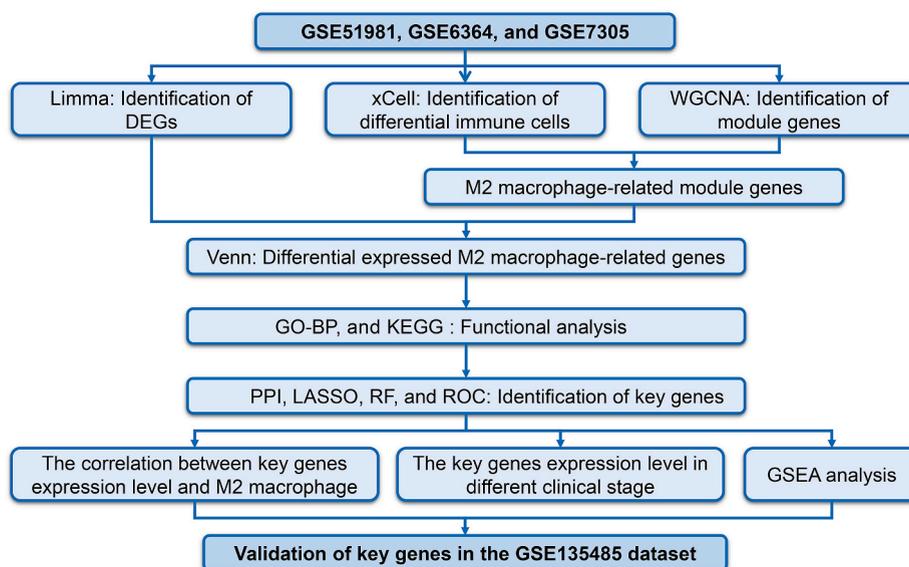**Fig. 1.** The study design. DEGs, differential expressed genes; WGCNA, Weighed Gene Co-expression Network Analysis; GO-BP, Gene Ontology – Biological Process; KEGG, Kyoto Encyclopedia of Genes and Genomes; PPI, Protein-Protein Interaction Networks; RF, Random Forest; LASSO, Least Absolute Shrinkage and Selection Operator; ROC, Receiver Operator Characteristic; GSEA, Gene Set Enrichment Analysis.

GSE51981 contained 71 non-endometriosis and 77 endometriosis. GSE6364 contained 16 non-endometriosis and 21 endometriosis. GSE7305 contained 10 non-endometriosis and 10 endometriosis. GSE135485 contained 4 non-endometriosis and 54 endometriosis. Samples in the first three datasets were sequenced by GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array and samples in the last dataset were sequenced with Illumina HiSeq 3000.

The study design was shown in Fig. 1.

2.2. Data processing

The probe expression matrix of GSE51981, GSE6364, and GSE7305 after pretreatment, standardization, and log2 transformation were downloaded. Platform annotation file was also downloaded and was utilized to convert the probe into gene expression matrix. For the gene with numerous probes, the average value of probes was defined as the gene expression level. Data in GSE51981, GSE6364, and GSE7305 were all detected from GPL570, and this study performed a batch effect removing analysis by sva package [16] (version 3.38.0) in R 3.6.1. The combined gene expression data after removing the batch effects was named combat edata, which was utilized for the following analysis.

The GSE135485 dataset sequenced by Illumina HiSeq 3000 was standardized by TMM algorithm in edgeR package (version 3.4) [17,18]. The data were further converted into logCPM value for the following validation analysis.

2.3. Analysis of differential expression

This study performed a Classical Bayesian method using limma package [19] (version 3.10.3) in the combat edata obtained previously. Differential genes expression between endometriosis and non-endometriosis were analysed and relevant p.value and logfold change (FC) were obtained. In addition, multiple testing corrections were carried out to get adjusted p value namely adj.p.value using the Benjamini & Hochberg (BH) method [20]. This study selected differentially expressed thresholds of adj.p.value < 0.05 and |logFC| > 0.5 as the differentially expressed genes (DEGs).

2.4. Analysis of immune cell composition

In order to identify the significantly differential immune cells between disease group and the control group, this study estimated the relative abundance of immune cells and stromal cells in each sample using xCell package [21] in R. Furthermore, the significance p. value was calculated by Wilcox test, and the p value was corrected by BH to obtain p.adjust. Immune cells with significant differences between the disease group and the control group were found for the subsequent analysis with p.adjust <0.05. Then, the CIBERSORT algorithm [22,23] was further used to verify the difference in the proportion of M2 macrophages between two groups.

2.5. Analysis of macrophage M2-related genes

Top 3000 genes with large variation in the combat edata was selected to screen the highly coordinated variation gene modules using weighted correlation network analysis (WGCNA) [24]. In the WGCNA algorithm, the elements in the gene co-expression matrix are the weighted values of gene correlation coefficients, that is, the soft-threshold power. By setting a series of powers, this study calculated the square value of correlation coefficient of connectivity k and p(k) and the average connectivity under each power value, and then selected the appropriate power to allow the connections between genes in the network conform to the scale-free network distribution. Then, genes with similar expression pattern (similarity is greater than 0.3) are combined into a module based on cluster analysis and dynamic pruning method. By calculating the correlation between modules and phenotypes (immune cell infiltration abundance), the modules highly correlated with immune cells were selected as the immune cell-related module and genes in module were selected as immune cell-related genes. Moreover, this study intersected the immune cell-related genes with the DEGs as the differential immune cell-related genes. According to the results, macrophage M2 showed most significant positive and negative correlation with purple and brown modules, respectively. So the purple and brown modules were defined as the key module, and genes in the two modules were defined as a key immune cell (macrophage M2)-related module gene.

2.6. Functional analysis of macrophage M2-related genes and construction of Protein-Protein Interaction (PPI) Networks

Gene Ontology [25] biological process (GO-BP) analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) [26] pathway were conducted in the differential macrophage M2-related genes using clusterProfiler package (version 3.8.1) [27] in R. BH was used for p value correction, after correction, P.adjust <0.05 was considered as a significant enrichment result. Due to the large number of GO-BP results, there was some redundancy. Therefore, simplifyEnrichment (Version: 1.4.0) [28] was further adopted for the further processing. The package used binary cut to divide the obtained GO similarity matrix into several classes. Through the annotation, the corresponding function of each class can be captured.

At the same time, differential macrophage M2-related genes were input into STRING (version 10.0) [29], defined species as homo, PPI parameter as 0.4 (medium confidence) to get the PPI relationship pairs. Cytoscape [30] was utilized to construct a PPI network.

2.7. Selection of key genes

In order to screen which genes in the above macrophage M2-related genes played important role in the disease, two algorithms were carried out. Least Absolute Shrinkage and Selection Operator (LASSO) was used to select the genes using glmnet package (version 4.0-2) [31]. The parameter is set to nfold = 20, that is, 20-fold cross-validation is performed. Random Forest (RF) algorithm [32] were subsequently used to select the Top30 genes. The parameter is set as: ntree = 1000, mtry = 6. The intersecting genes from LASSO and RF algorithms were selected to draw the Receiver Operator Characteristic (ROC) curve and screen genes with areas under the curve (AUC) $\geq$ 0.65 as the final key genes.

2.8. Analysis of expression features for key genes

Pearson correlations were conducted to validate the correlation between key genes and M2 macrophage infiltration level. The relationships between the key genes and clinical features were also analysed. Box plots of the severity of endometriosis (Minimal/Mild and Moderate/Severe) and the expression levels of key genes at different times (endometrial proliferative stage, early secretory stage, and middle secretory stage) were plotted.

2.9. Enrichment pathway analysis of key genes

In order to observe the mechanism of key genes in the process of endometriosis occurrence, Gene Set Enrichment Analysis (GSEA) was conducted for each key gene using GSEA software (version 3.0) [33]. Parameters are set as: Gene sets database: c2.cp.kegg.v7.4. symbols.gmt; Phenopyte labels: Use a gene as the phenotype; Metric for ranking genes: Pearson; Gene list sorting mode: real; Gene list ordering mode: descending; threshold: NOM p-val <0.01.

2.10. Validation of the key genes

The validation data set GSE135485 was used to extract the expression values of key genes in each sample. The differences of the gene expression levels between the disease group and the control group were analysed. xCell algorithm was also used to estimate the macrophage M2 infiltration level of each sample, and Pearson correlation coefficient between key genes and macrophages M2 was calculated.

3. Results

3.1. Identification of DEGs

Data in GSE51981, GSE6364, and GSE7305 were processed to remove batch effects and all samples were generated into combat edata. There was no significant separation between samples after batch effects calibration (Fig. 2A), suggesting the data were able to use for the further analysis. According to the threshold setting in the method, 116 up-regulated and 519 down-regulated DEGs were identified (Fig. 2B).

3.2. Identification of differential expressed M2 macrophage-related genes

The immune infiltration levels of 64 stromal and immune cells were calculated using xCell package in R. There are significant differential infiltration levels between the non-endometriosis and endometriosis groups in four cells: hematopoietic stem cell (HSC),
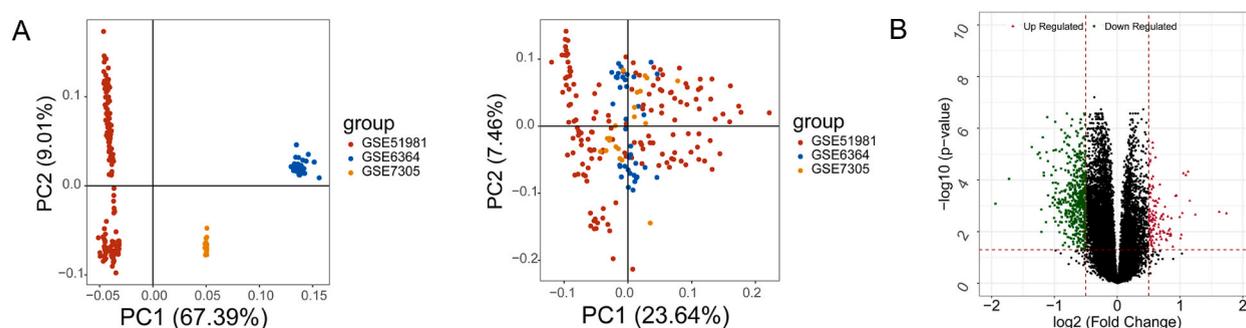

Fig. 2. Identification of DEGs. (A) PCA shows gene distributions before (left) and after (right) batch effect removal. (B) The volcano plot of DEGs. The red triangle represents the up-regulated genes, the green square represents the down-regulated genes, and the black indicates that the gene difference is not significant. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

M2 macrophage, plasma cells, and activated dendritic cell (aDC) ($p < 0.05$). The infiltration level of HSC, M2 macrophage, plasma cells, and aDC in the endometriosis group were significantly higher than in the Non-endometriosis group (Fig. 3A). The infiltration level of macrophage M2 in xCell was provided in Table S1. To further verity the infiltration difference of M2 macrophage between two groups, the CIBERSORT algorithm was implemented. The results also suggested an increased level of M2 macrophage infiltration in the endometriosis group than that in the non-endometriosis group (Fig. 3B), which is consistent with the above findings.

The Top3000 genes in the combat edata were selected to conduct a WGCNA analysis. Finally, the genes were integrated into nine
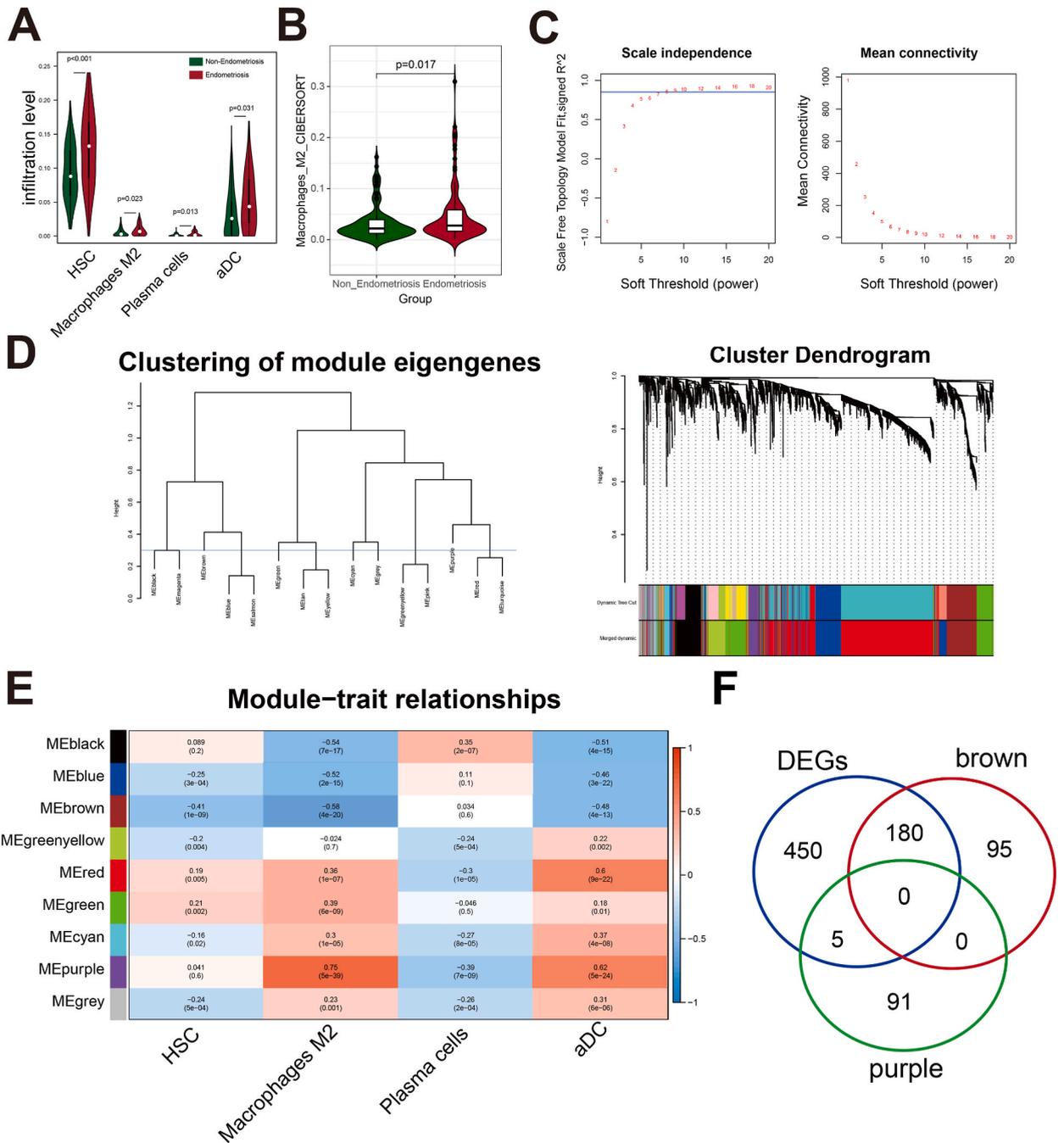

Fig. 3. Identification of differential expressed M2 macrophage-related genes. (A) Screening of four immune cells with significant infiltration differences between endometriosis and non-endometriosis groups based on the xCell algorithm. (B) The CIBERSORT algorithm was applied to confirm the infiltration difference of M2 macrophages between two groups. (C) Selection of the soft-threshold powers in WGCNA analysis. (D) Dendrogram of all genes clustered based on the measurement of dissimilarity. (E) Heat plot of correlation between different modules and immune cells. (F) Venn plot of intersecting of DEGs and M2 macrophage-related module genes.

modules (Fig. 3C and D). Following, the correlations between the gene modules and the immune infiltration level were calculated (Fig. 3E). The brown (275 genes, r = −0.58, $p < 0.001$) and the purple (96 genes, r = 0.75, $p < 0.001$) modules were most significantly negative and positive with the M2 macrophage (Fig. 3E). Therefore, the brown and the purple modules were selected to be the key M2-related gene modules for the further analysis. Intersecting the key M2-related gene modules with the DEGs, 180 differential M2-related brown module genes and 5 differential M2-related purple module genes were obtained (Fig. 3F).

3.3. The function of M2-related genes and construction of the PPI network

GO-BP and KEGG pathway enrichment analyses were conducted to explore the function of the M2-related genes. As there were too
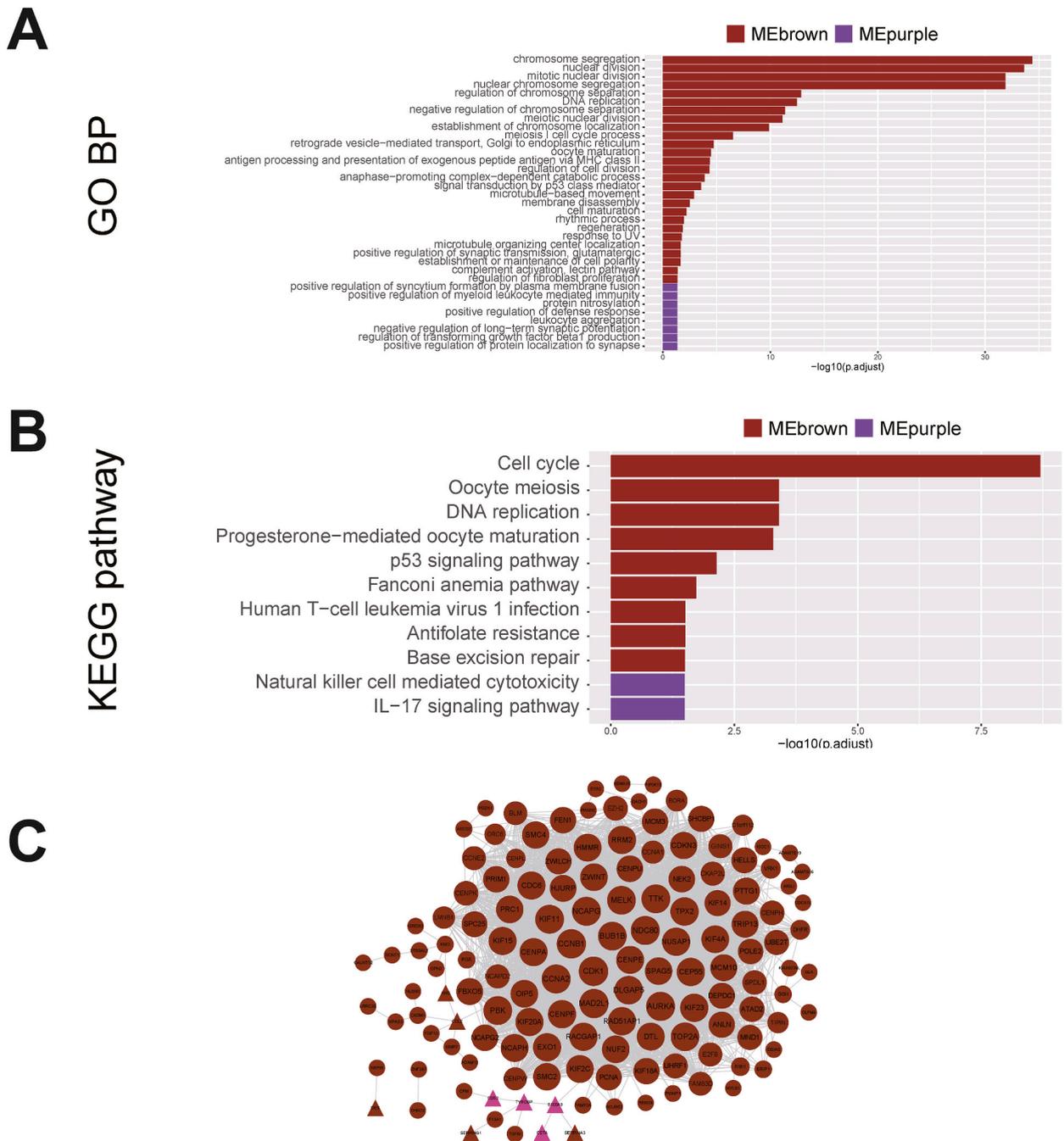

Fig. 4. The function analysis of the differential expressed M2 macrophage-related. (A) The most significant item in each GO BP clusters. (B) The KEGG pathway significantly enriched in differential module genes. (C) The PPI network constructed using differential expressed module genes.
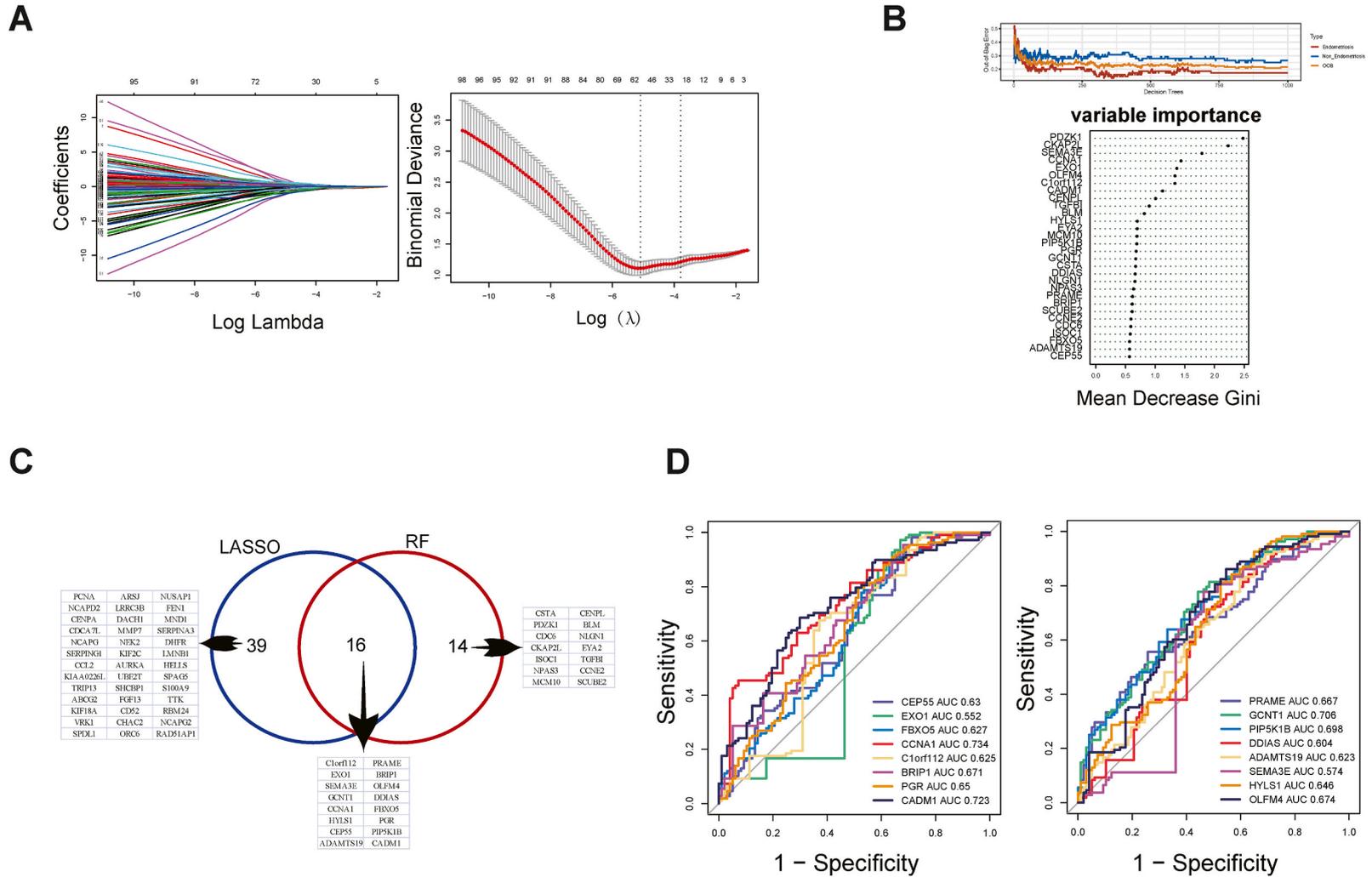
Fig. 5. Screening of the key genes. (A) LASSO coefficient distribution (left) and the likelihood bias of LASSO coefficient distribution (right). (B) The Top30 genes, ranked by importance, were selected using the RF algorithm. (C) Intersecting Venn plot of genes from LASSO and RF. (D) ROC curves of 16 intersecting genes, and eight genes with AUC >0.65 were screened as the key genes.

many GO-BP items and it was redundancy, semantic similarity enrichment analysis was further conducted for GO-BP. Among them, 27 GO-BP clusters and eight GO-BP clusters were obtained from the brown and purple module genes, respectively. The most significant items in each cluster were visualized in Fig. 4A. Genes in the module brown were more likely to involve in chromosome segregation and nuclear division, while the purple module genes were mainly enriched in the regulation of syncytium formation and immunity. The KEGG pathway of the M2-related genes were enriched in the cell cycle, oocyte meiosis, DNA replication, progesterone-mediated oocyte maturation, p53 signaling pathway, human T-cell leukemia virus 1 infection, antifolate resistance, base excision repair, natural killer cell mediated cytotoxicity, and IL-17 signaling pathway (Fig. 4B).

Furthermore, a PPI network was constructed in the 185 differential M2-related module genes (Fig. 4C). A total of 3030 interaction pairs composed of 140 gene proteins were obtained. The list of these 140 genes and their connections was provided in Table S2. Of these, *CDK1*, *CCNB1*, and *MAD2L1* may contribute more to the progression of endometriosis, since they have more degrees of connections in this network.

3.4. Screening of key genes

LASSO regression analysis identified 55 genes from 140 PPI genes (Fig. 5A). RF algorithm was used to select the Top30 genes (Fig. 5B). Following, intersecting the 55 genes from LASSO and 30 genes from RF, 16 genes were finally obtained (Fig. 5C). Finally, ROC analysis was performed in the intersecting genes and eight genes with AUC ≥0.65 were selected as the key genes (Fig. 5D). The eight key genes were *PGR*, *OLFM4*, *PIP5K1B*, *CCNA1*, *BRIP1*, *CADM1*, *PRAME*, and *GCNT1*.
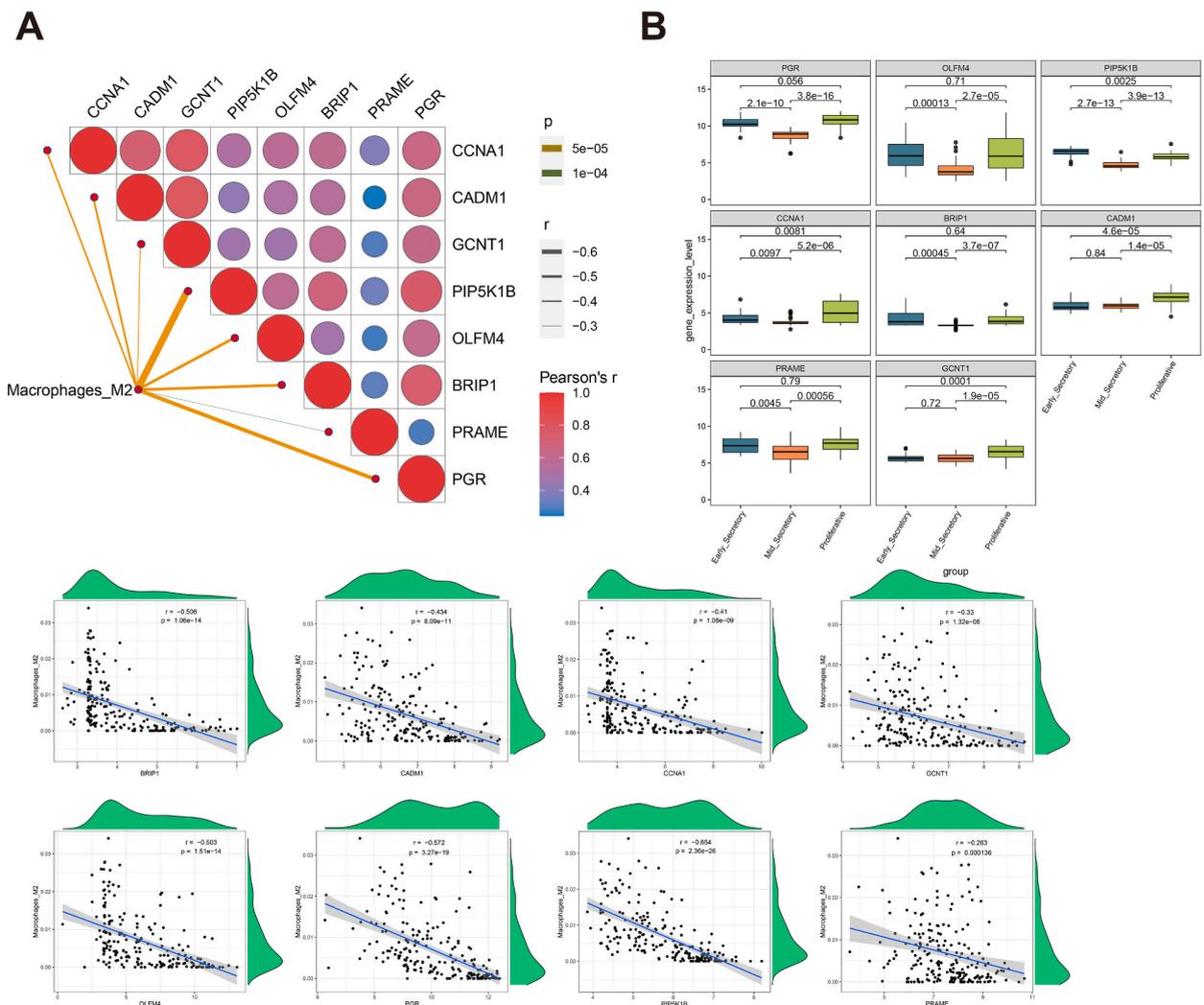

Fig. 6. The analysis for key genes. (A) The correlation of each key gene expression and the M2 macrophage infiltration level. (B) Box plots of the gene expression level in different stage.

3.5. The analysis for key genes

The correlations of each key gene and the M2 macrophage were evaluated, and the expressions of eight key genes were all significantly negative with the infiltration levels of M2 macrophage (Fig. 6A). In addition, the box plots of the gene expression level in different stage were plotted (Fig. 6B). The expression levels of the eight key genes were significantly lower in the middle secretory stage ($p < 0.05$) while relevantly higher in the proliferative stage. This study also compared the expression difference of the eight genes between different severity degree of endometriosis group in proliferative, early, and mid secretory phases of menstrual cycle,

Fig. 7. The GSEA enrichment pathways of eight key genes.
The enrichment pathways for *BRIP1*, *CADM1*, *CCNA1*, *GCNT1*, *PIP5K1B*, *PRAME*, *PGR*, and *OLFM4* are shown in panel A–H, respectively.

respectively (Fig. S1). The results showed that only *PRAME* in moderate/severe group were significantly lower than in the minimal/mild group in the early secretory stage. However, the expression difference of these genes between different severity degree of endometriosis group was not found in proliferative and mid secretory phases.

3.6. The GSEA enrichment pathway of the key genes

The GSEA enrichment of KEGG pathway analysis for eight genes was performed. The results showed that these genes were significantly enriched in pathways related to genomic instability, such as base excision repair, DNA replication, mismatch repair, nucleotide excision repair, RNA degradation spliceosome etc. (Fig. 7A–H).

3.7. Validation of the key genes

GSE135485 was utilized as a validation dataset. The eight gene expression levels in the samples of non-endometriosis and endometriosis groups were analysed (Fig. 8A). All the eight genes showed downregulated trend in the endometriosis groups. Except for the *OLFM4* and *CADM1*, the other 6 genes were all significant differences between the non-endometriosis and endometriosis groups ($p < 0.05$). The ROC curves showed five genes (*PGR* AUC = 0.796, *PIP5K1B* AUC = 0.759, *CCNA1* AUC = 0.912, *BRIP1* AUC = 0.773, *CADM1* AUC = 0.694, and *GCNT1* AUC = 0.829) with AUCs >0.65 (Fig. 8B). These results further demonstrated that the eight key genes were significantly associated with the endometriosis. Moreover, the Pearson coefficients were calculated to evaluate the correlation between the key genes and the M2 macrophage (Fig. 8C). The seven key genes were significantly negatively related to the M2 macrophage.
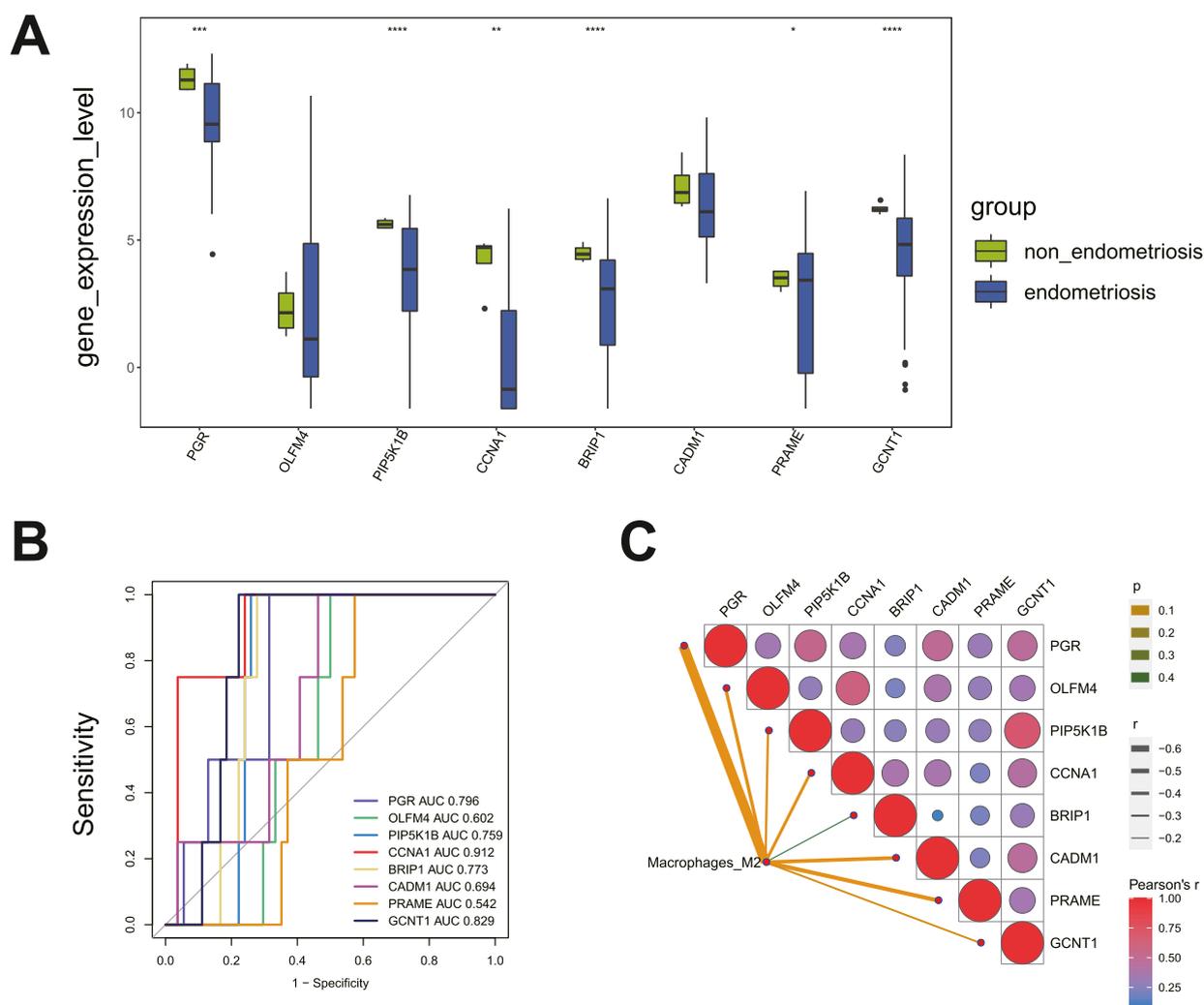

Fig. 8. Validation of the key genes in the GSE135485. (A) The eight gene expression level in the non-endometriosis and endometriosis groups. (B) The ROC curves for the eight key genes. (C) The correlation between the key genes and the M2 macrophage.

4. Discussion

Endometriosis, affects around 10 % of reproductive women, with growth or deposition of endometrial tissue at extra uterine sites. M2 macrophages play important roles in the endometriosis which may stem from the regulation of M2 macrophage-related genes [7]. In the present study, by analyzing the data downloaded from the public GEO database, this study identified 185 M2 macrophage-related genes through a series bioinformatics method. These M2 macrophage-related genes were mainly enriched in the oocyte-related function, cell cycle, and immune cell cytokines-related function (Fig. 4A–B). The PPI network suggested that these genes had close interaction relationships and might play an important role in endometriosis progression. To obtain the optimal genes related to M2-macrophage, further LASSO, RF, and ROC analysis were conducted to screen the optimal key genes in the endometriosis. Finally, eight key genes were obtained: *PGR, OLFM4, PIP5K1B, CCNA1, BRIP1, CADM1, PRAME,* and *GCNT1.*

The eight key genes were all demonstrated to be negatively related to M2 macrophage, further confirming their closely correlations with the M2 macrophage. *PGR*, progesterone receptor, has been demonstrated to be important in regulating female reproduction [34]. A study has confirmed that the increased methylation of *PGR* isoforms may be associated with the reduced gene expression level, thereby impairing endometrial receptivity in patients with endometriosis [35]. The loss of PGR nuclear positivity in the proliferative endometrium of patients with endometriosis also suggests a decrease in the expression level of *PGR* [36]. These findings are consistent with our study, suggesting that *PGR* expression levels are regulated by methylation of PGR isoforms, which may further affect the endometrial receptivity. In addition, this study also found the elevated level of M2 macrophages and their significant negative correlations with *PGR* levels in patients with endometriosis. Jeong et al. revealed that a large number of CD163+ (M2) macrophages was associated with PGR negativity in tumor [37]. These evidences demonstrate that patients with endometriosis may be in a state of immunosuppression, leading to the increased infiltration of M2 macrophages, which can cause damage to the endometrium. *OLFM4,* olfactomedin-4, an extracellular matrix protein, is one of the top downregulated genes in patients with endometriosis compared with non-endometriosis controls [38]. However, it is highly expressed in human endometrium and has the highest expression level in proliferative-phase endometrium [39]. These conclusions are consistent with the present study, as this study found that *OLFM4* expression in endometriosis is menstrual cycle-dependent, with decreased expression in the midsecretory phase and relatively high expression in the proliferative phase (Fig. 6B). In ulcerative colitis, the expression level of *OLFM4* is also significantly negatively correlated with the infiltration of M2 macrophages [40]. Therefore, this study further hypothesized that *OLFM4* may play an anti-inflammatory and endometrial stabilizing role by negatively regulating M2 macrophages. *PIP5K1B, CCNA1, BRIP1, CADM1, PRAME,* and *GCNT1* were demonstrated to involve in various diseases and cancers, but their roles in endometriosis and their relationships with M2 macrophages remain unreported. Therefore, subsequent functional experiments are needed to investigate their immunoregulatory mechanisms.

In a menstrual cycle, endometrium undergoes transition from estrogen-dominant proliferative (follicular) phases, progesterone-dominant secretory (luteal) phases (early-secretory, mid-secretory, and late-secretory), to menstrual phase. In each endometrial cycle phase, the transcriptional profile exhibited distinct differences [41]. The transcriptomic profiles of uterine linings in patients with endometriosis will alter [38], demonstrating that the molecular feature of endometriosis may be a marker of onset of endometriosis. In the present study, the results of testing set and validation set indicated that the expressions of eight key genes were decreased in endometriosis. Meanwhile, in patients with endometriosis, the expression of these eight key genes shows a similar pattern, that is, the expression is decreased in the mid-secretion stage, but significantly up-regulated in the proliferative stage. Therefore, the expression of these eight key genes is considered to be menstrual cycle-dependent, and their aberrant down-regulation in endometriosis may be associated with the reduced expression during the mid-secretory phase. It is known that the short menstrual cycle is one of the risk factors for endometriosis [3], so increasing the expression levels of these genes in the mid-secretory phase may be beneficial in reducing the risk of endometriosis. However, these speculations still need to be validated by the clinical sample-based sequencing analyses.

There are still several limitations in the present study. Firstly, the down-regulation of these eight genes in endometriosis needs to be verified by *in vitro* experiments. Secondly, the relationships between these eight genes and M2 macrophages, as well as their regulatory mechanisms in endometriosis still need to be excavated by conducting *in vivo* and *in vitro* experiments. Finally, the lack of clinical information such as regional and ethnic limit the generalisability of this study. Therefore, how the eight genes affect the disease progression of endometriosis through the menstrual cycle still needs to be confirmed by experiments based on a large number of clinical samples.

5. Conclusions

Using machine learning algorithms, this study identified eight M2 macrophage-related genes (*PGR, OLFM4, PIP5K1B, CCNA1, BRIP1, CADM1, PRAME,* and *GCNT1*) that may be potential biomarkers of endometriosis. These genes were confirmed to be significantly under-expressed in endometriosis and were significantly negatively correlated with the infiltration of M2 macrophages. Furthermore, the expression pattern of these genes may be menstrual cycle-dependent. However, these conclusions need to be further explored by *in vitro* and *in vivo* experiments, as well as clinical sample-based analyses.

Funding

Data Availability statement

All data generated or analysed during this study are included in this article.

CRediT authorship contribution statement

Hongyan Ding: Writing – original draft, Conceptualization. **Hongge Xu:** Formal analysis, Data curation. **Ting Zhang:** Validation, Methodology, Investigation. **Can Shi:** Writing – review & editing, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Not applicable.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2023.e22258.

References

[1] C. Chapron, L. Marcellin, B. Borghese, P. Santulli, Rethinking mechanisms, diagnosis and management of endometriosis, Nat. Rev. Endocrinol. 15 (2019) 666–682.
[2] A. Mińko, A. Turoń-Skrzypińska, A. Ryl, P. Bargiel, Z. Hilicka, K. Michalczyk, P. Łukowska, I. Rotter, A. Cymbaluk-Płoska, Endometriosis-A multifaceted problem of a modern woman, Int. J. Environ. Res. Publ. Health 18 (2021).
[3] A.L. Shafrir, L.V. Farland, D.K. Shah, H.R. Harris, M. Kvaskoff, K. Zondervan, S.A. Missmer, Risk for and consequences of endometriosis: a critical epidemiologic review, Best Pract. Res. Clin. Obstet. Gynaecol. 51 (2018) 1–15.
[4] M. Kvaskoff, Y. Mahamat-Saleh, L.V. Farland, N. Shigesi, K.L. Terry, H.R. Harris, H. Roman, C.M. Becker, S. As-Sanie, K.T. Zondervan, A.W. Horne, S.A. Missmer, Endometriosis and cancer: a systematic review and meta-analysis, Hum. Reprod. Update 27 (2021) 393–420.
[5] T. Falcone, R. Flyckt, Clinical management of endometriosis, Obstet. Gynecol. 131 (2018) 557–571.
[6] R. Edi, T. Cheng, Endometriosis: evaluation and treatment, Am. Fam. Physician 106 (2022) 397–404.
[7] J. Vallvé-Juanico, S. Houshdaran, L.C. Giudice, The endometrial immune environment of women with endometriosis, Hum. Reprod. Update 25 (2019) 564–591.
[8] R. Wu, K.H. Van der Hoek, N.K. Ryan, R.J. Norman, R.L. Robker, Macrophage contributions to ovarian function, Hum. Reprod. Update 10 (2004) 119–133.
[9] J. Wu, H. Xie, S. Yao, Y. Liang, Macrophage and nerve interaction in endometriosis, J. Neuroinflammation 14 (2017) 53.
[10] A. Shapouri-Moghaddam, S. Mohammadian, H. Vazini, M. Taghadosi, S.A. Esmaeili, F. Mardani, B. Seifi, A. Mohammadi, J.T. Afshari, A. Sahebkar, Macrophage plasticity, polarization, and function in health and disease, J. Cell. Physiol. 233 (2018) 6425–6440.
[11] O.B. Poli-Neto, J. Meola, E.S.J.C. Rosa, D. Tiezzi, Transcriptome meta-analysis reveals differences of immune profile between eutopic endometrium from stage I-II and III-IV endometriosis independently of hormonal milieu, Sci. Rep. 10 (2020) 313.
[12] L.L. Wang, Z.H. Li, H. Wang, J. Kwak-Kim, A.H. Liao, Cutting edge: the regulatory mechanisms of macrophage polarization and function during pregnancy, J. Reprod. Immunol. 151 (2022), 103627.
[13] M. Saare, V. Modhukur, M. Suhorutshenko, B. Rajashekar, K. Rekker, D. Sõritsa, H. Karro, P. Soplepmann, A. Sõritsa, C.M. Lindgren, N. Rahmioglu, A. Drong, C. M. Becker, K.T. Zondervan, A. Salumets, M. Peters, The influence of menstrual cycle and endometriosis on endometrial methylome, Clin. Epigenet. 8 (2016) 2.
[14] Z. Cui, R. Bhandari, Q. Lei, M. Lu, L. Zhang, M. Zhang, F. Sun, L. Feng, S. Zhao, Identification and exploration of novel macrophage M2-related biomarkers and potential therapeutic agents in endometriosis, Front. Mol. Biosci. 8 (2021), 656145.
[15] T. Barrett, T.O. Suzek, D.B. Troup, S.E. Wilhite, W.-C. Ngau, P. Ledoux, D. Rudnev, A.E. Lash, W. Fujibuchi, R. Edgar, NCBI GEO: mining millions of expression profiles—database and tools, Nucleic Acids Res. 33 (2005) D562–D566.
[16] J.T. Leek, W.E. Johnson, H.S. Parker, A.E. Jaffe, J.D. Storey, The sva package for removing batch effects and other unwanted variation in high-throughput experiments, Bioinformatics 28 (2012) 882–883.
[17] O. Nikolayeva, M.D. Robinson, edgeR for differential RNA-seq and ChIP-seq analysis: an application to stem cell biology, Methods Mol. Biol. 1150 (2014) 45–79.
[18] M.D. Robinson, D.J. McCarthy, G.K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data, Bioinformatics 26 (2010) 139–140.
[19] G.K. Smyth, Limma: linear models for microarray data, in: R. Gentleman, V.J. Carey, W. Huber, R.A. Irizarry, S. Dudoit (Eds.), Bioinformatics and Computational Biology Solutions Using R and Bioconductor. Vol, Springer New York, New York, NY, 2005, pp. 397–420.
[20] D. Ghosh, Wavelet-based Benjamini-Hochberg procedures for multiple testing under dependence, Math. Biosci. Eng. 17 (2019) 56–72.
[21] D. Aran, Z. Hu, A.J. Butte, xCell: digitally portraying the tissue cellular heterogeneity landscape, Genome Biol. 18 (2017) 220.
[22] Z. Wu, K. Lei, H. Li, J. He, E. Shi, Transcriptome-based network analysis related to M2-like tumor-associated macrophage infiltration identified VARS1 as a potential target for improving melanoma immunotherapy efficacy, J. Transl. Med. 20 (2022) 489.
[23] K. Xiao, S. Zhao, J. Yuan, Y. Pan, Y. Song, L. Tang, Construction of molecular subtypes and related prognostic and immune response models based on M2 macrophages in glioblastoma, Int. J. Gen. Med. 15 (2022) 913–926.
[24] P. Langfelder, S. Horvath, WGCNA: an R package for weighted correlation network analysis, BMC Bioinf. 9 (2008) 559.
[25] M. Ashburner, C.A. Ball, J.A. Blake, D. Botstein, H. Butler, J.M. Cherry, A.P. Davis, K. Dolinski, S.S. Dwight, J.T. Eppig, M.A. Harris, D.P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J.C. Matese, J.E. Richardson, M. Ringwald, G.M. Rubin, G. Sherlock, Gene Ontology: tool for the unification of biology, Nat. Genet. 25 (2000) 25–29.
[26] M. Kanehisa, S. Goto, KEGG: Kyoto Encyclopedia of genes and Genomes, Nucleic Acids Res. 28 (2000) 27–30.
[27] G. Yu, L.G. Wang, Y. Han, Q.Y. He, clusterProfiler: an R package for comparing biological themes among gene clusters, OMICS 16 (2012) 284–287.

[28] Z. Gu, D. Hübschmann, simplifyEnrichment: an R/bioconductor package for clustering and visualizing functional enrichment results, bioRxiv. 21 (2023) 190–202.

[29] D. Szklarczyk, A. Franceschini, S. Wyder, K. Forslund, D. Heller, J. Huerta-Cepas, M. Simonovic, A. Roth, A. Santos, K.P. Tsafou, STRING v10: protein–protein interaction networks, integrated over the tree of life, Nucl. Acids Res. 43 (2015) D447–52.

[30] P. Shannon, A. Markiel, O. Ozier, N.S. Baliga, J.T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker, Cytoscape: a software environment for integrated models of biomolecular interaction networks, Genome Res. 13 (2003) 2498–2504.

[31] J. Friedman, T. Hastie, R. Tibshirani, Glmnet: lasso and elastic-net regularized generalized linear models, R Package Version 1 (2009).

[32] A. Liaw, M. Wiener, Package 'randomForest': breiman and Cutler's random forests for classification and regression, R Development Core Team 4 (2014) 6–10.

[33] A. Subramanian, P. Tamayo, V.K. Mootha, S. Mukherjee, B.L. Ebert, M.A. Gillette, A. Paulovich, S.L. Pomeroy, T.R. Golub, E.S. Lander, J.P. Mesirov, Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles, Proc Natl Acad Sci U S A 102 (2005) 15545–15550.

[34] L. Chen, B.Y. Zhang, G.D. Feng, W. Xiang, Y.X. Ma, H. Chen, M.X. Chu, P.Q. Wang, The mechanism of miRNA-mediated PGR signaling pathway in regulating female reproduction, Yi Chuan 38 (2016) 40–51.

[35] C.V. Rocha-Junior, M.G. Da Broi, C.L. Miranda-Furtado, P.A. Navarro, R.A. Ferriani, J. Meola, Progesterone receptor B (PGR-B) is partially methylated in eutopic endometrium from infertile women with endometriosis, Reprod. Sci. 26 (2019) 1568–1574.

[36] M. Colón-Caraballo, M. García, A. Mendoza, I. Flores, Human endometriosis tissue microarray reveals site-specific expression of estrogen receptors, progesterone receptor, and Ki67, Appl. Immunohistochem. Mol. Morphol. 27 (2019) 491–500.

[37] H. Jeong, I. Hwang, S.H. Kang, H.C. Shin, S.Y. Kwon, Tumor-associated macrophages as potential prognostic biomarkers of invasive breast cancer, J Breast Cancer 22 (2019) 38–51.

[38] Z. Mirza, U.A. Abdel-Dayem, Uncovering potential roles of differentially expressed genes, upstream regulators, and canonical pathways in endometriosis using an in silico genomics approach, Diagnostics 10 (2020).

[39] H. Dassen, C. Punyadeera, B. Delvoux, I. Schulkens, C. Marchetti, R. Kamps, J. Klomp, F. Dijcks, A. de Goeij, T. D'Hooghe, C. Kyama, A. Ederveen, G. Dunselman, P. Groothuis, A. Romano, Olfactomedin-4 regulation by estrogen in the human endometrium requires epidermal growth factor signaling, Am. J. Pathol. 177 (2010) 2495–2508.

[40] T. He, K. Wang, P. Zhao, G. Zhu, X. Yin, Y. Zhang, Z. Zhang, K. Zhao, Z. Wang, K. Wang, Integrative computational approach identifies immune-relevant biomarkers in ulcerative colitis, FEBS Open Bio 12 (2022) 500–515.

[41] A.P. Ponnampalam, G.C. Weston, A.C. Trajstman, B. Susil, P.A. Rogers, Molecular classification of human endometrial cycle stages by transcriptional profiling, Mol. Hum. Reprod. 10 (2004) 879–893.