European Review for Medical and Pharmacological Sciences 2024; 28: 556-570

Identification and verification of feature biomarkers associated with immune cells in recurrent pregnancy loss

M.-Q. JIN, B.-Y. HUANG, D.-Y. LU, J.-Y. HUANG, L. MA

The Reproductive Medicine Center, The Seventh Affiliated Hospital of Sun Yat-Sen University, Shenzhen, China

M-Q. Jin and B-Y. Huang contributed equally to this work

Abstract. – OBJECTIVE: The aim of this study was to investigate the causes, diagnostic markers, and treatment methods for recurrent pregnancy loss (RPL) using bioinformatics approaches.

MATERIALS AND METHODS: Bioinformatics methods were utilized to analyze gene expression databases to identify key genes and modules associated with RPL. Weighted gene co-expression network analysis (WGCNA) was employed to identify gene sets related to maternal-fetal immunity. Gene set variation analysis (GSVA) and protein-protein interaction networks were used to explore signaling pathways and molecular interactions in RPL. Immune cell infiltration was assessed using single-sample gene set enrichment analysis (ssGSEA).

RESULTS: Thirteen genes were identified as potential diagnostic markers, some of which were involved in placental amino acid transport, glucose absorption, and reactive oxygen species production. Several gene sets related to protein transport, steroid synthesis, and glycosaminoglycan degradation were found to be associated with RPL. Immune cell infiltration analysis found that CD56bright NK cells and monocytes showed significantly increased infiltration in RPL and were associated with key hub genes. The validation of hub genes, including *PCSK5, CCND2, SLC5A3, RASAL1, MYZAP, MFAP4*, and *P2RY14*, as potential diagnostic markers, showed promising value.

CONCLUSIONS: This study contributes to a better understanding of the etiology of RPL and potential diagnostic markers. The identified immune-related gene sets, signaling pathways, and immune cell infiltrations provide valuable insights for future research and therapeutic advancements in RPL.

Key Words:

Identification and verification, Feature biomarkers associated, Immune cells, Recurrent pregnancy loss.

Introduction

Recurrent pregnancy loss (RPL), commonly referred to as recurrent miscarriage, is a distressing condition marked by the consecutive loss of two or more pregnancies before the 20th week of gestation. It affects around 1-2% of couples trying to conceive and places a significant emotional and physical burden on them[1,2]. Despite thorough clinical investigation, the multifactorial and complex nature of RPL leaves around 50% of cases unexplained[3]. Genetic, hormonal, anatomical, infectious factors, and immune disruption have been linked to RPL.

Currently, the treatment options for RPL have limited efficacy. Treatment strategies for RPL vary based on the identified causes, resulting in enhanced pregnancy success rates among specific patient populations[4]. Nevertheless, existing treatment interventions exhibit suboptimal effectiveness for RPL cases with unidentified causes. Additionally, diagnosing RPL becomes challenging in patients without a history of consecutive miscarriages. Therefore, the identification and screening of marker genes play a pivotal role in facilitating early RPL diagnosis, discovering new therapeutic targets, and ultimately enhancing pregnancy outcomes.

Unexplained recurrent pregnancy loss (URPL) refers to the consecutive occurrence of two or more spontaneous miscarriages, but the specific cause remains unclear even after common reproductive system issues have been ruled out. Current research in the literature suggests that immune-related factors may be associated with URPL. Some women with URPL have immune system abnormalities, such as decreased natural killer cell activity or T cell subset imbalance. In

Corresponding Authors: Lin Ma, MD; e-mail: malin8@mail.sysu.edu.cn; Jiayu Huang, MD; e-mail: huangjy388@mail.sysu.edu.cn

addition, some studies[5] also indicate the possible existence of abnormal immunological factors such as anti-embryo antibodies or anti-phospholipid antibodies related to URPL.

The maternal immune system needs to maintain a delicate balance between creating an optimal immune environment for fetal development and safeguarding against potential threats like infections and malignancies. Previous research[6] has found that many immune-related diseases may be associated with miscarriage, such as Hashimoto's thyroiditis, systemic lupus erythematosus, antiphospholipid syndrome, asthma, and so on. Successful pregnancy outcomes depend on diverse immune regulatory processes, involving immune cell activation, modulation, cytokine signaling pathways, and immune cell infiltration[7,8]. Abnormalities in the maternal-fetal immune interface, including aberrant immune cell counts, increased cytotoxicity, and imbalances in Th1/Th2/Th17 and Treg cells, are believed to be significant contributors to miscarriage in RPL patients[9,10]. Previous studies[11,12] have demonstrated elevated NK cell density in RPL and repeated implantation failure, along with an increased presence of macrophages in RPL patients. Cytotoxic T cells and dendritic cells (DCs) are also involved in immune regulation during pregnancy[13].

The use of bioinformatic methods has become increasingly popular for analyzing large datasets, such as high-throughput and microarray data. These methods help identify genes that are expressed differently (known as differentially expressed genes or DEGs) and enable various types of analyses[14,15]. This approach has been proven to be highly effective in uncovering the underlying mechanisms of many human diseases. Intending to gain insights into the development of RPL, we conducted a comprehensive genomic analysis using publicly available datasets. Our goal was to identify potential key genes, critical modules, pathways, and infiltrating immune cells associated with the pathogenesis of RPL.

Materials and Methods

Data Download

The data utilized in this study were obtained from the publicly accessible GEO database (Gene Expression Omnibus, https://www.ncbi.nlm.nih.gov/geo/). The whole-genome expression profiles of individuals with recurrent pregnancy loss were obtained by retrieving and downloading the relevant data from the GEO database through the R package "GEOquery" (version 2.62.2). The dataset GSE165004 comprised 24 patients with recurrent pregnancy loss and 24 control samples. GSE180485 included 16 RPL patients and 1 control sample (patients with a history of three or more miscarriages and no prior live births were defined as the disease group, and 16 samples were selected. Patients with no history of miscarriage and having had live births were designated as the control group, and one sample was selected.), while GSE26787 encompassed 5 RPL patients and 5 control samples.

Non-biological technical biases causing batch effects were corrected using the ComBat method from the R package "sva" (version 3.42.0)[16]. The effectiveness of the correction was assessed through principal component analysis (PCA). The data access policies of each respective database were adhered to in this study.

Maternal-fetal immune-related genes were retrieved from Li et al[17] and Alecsandru et al[18] (**Supplementary Table I**).

Differential Analysis of RPL

The R package "limma (version 3.50.0)"[19] was utilized to identify differentially expressed genes (DEGs) between the control group (n=30) and the RPL group (n=45) with the criteria of $|\log_2 \text{Fold Change}| > 0.5$ and $p<0.05$, which were used for subsequent analyses. Heatmaps were generated using the R package "pheatmap" (version 1.0.12), with Euclidean distance and hierarchical clustering methods used for clustering.

Gene Set Variation Analysis (GSVA)

Gene Set Variation Analysis (GSVA) is an unsupervised and non-parametric gene set enrichment method that allows the assessment of associations between gene expression profiles and biological pathways or gene features. To investigate the biological functional differences between the control and RPL groups, the "c2.cp.kegg.v7.5.1.symbols" gene set from the MSigDB database (http://software.broadinstitute.org/gsea/msigdb) was used as the reference gene set, and GSVA analysis was performed using the R package "GSVA (version 1.42.0)". The results were visualized using the R package "pheatmap (version 1.0.12)". Additionally, 50 hallmark gene sets were downloaded from the MSigDB database as reference gene sets, and the GSVA scores of each gene set were calculated in different samples using the ssGSEA function from the GSVA package. Differences in GSVA

scores between the control group and the RPL group for different gene sets were compared using the Limma package.

Weighted Gene Co-expression Network Analysis (WGCNA) and Identification of Significant Modules

The WGCNA algorithm was implemented using the R package WGCNA (version 1.70-3) to construct a co-expression network[20]. Gene expression profile similarity was evaluated by calculating Pearson's correlation coefficients, and the coefficients were weighted using a power function to obtain a scale-free network. The R package "PickSoftThreshold" was used to raise the co-expression similarity to the power of β=14 and to establish a weighted adjacency matrix. Gene modules represented sets of densely connected genes in the co-expression network. Gene modules were identified by WGCNA using hierarchical clustering and were represented with colors. The dynamic tree-cut method was employed to detect modules, whereby the adjacency matrix (a measure of topological similarity) was transformed into a topological overlap matrix (TOM). The modules were identified through clustering analysis. Pearson's correlation analysis was conducted to assess the association between modules and maternal-fetal immunity, by calculating the correlation between module eigengenes (MEs) and maternal-fetal immunity. The modules that showed significant association with maternal-fetal immunity were obtained. Heatmaps were generated to visualize the co-expression module structure, utilizing gene network topological overlap. Hierarchical clustering trees and corresponding eigengene heatmaps were plotted to summarize the relationships between modules. Maternal-fetal immune-related differentially expressed genes (maternal-fetal immune-related DEGs) were obtained from the intersection of DEGs and genes within the maternal-fetal immune-related modules.

GeneMANIA

GeneMANIA website (http://genemania.org) is a valuable resource that can predict functional relationships between genes and central genes, including protein-protein interactions, protein-DNA interactions, pathways, physiological and biochemical reactions, co-expression, and co-localization[21]. The GeneMANIA website was used to construct key genes' protein-protein interaction (PPI) network.

Receiver Operating Characteristic (ROC) Curve

The Receiver Operating Characteristic (ROC) curve is widely used as an effective method for evaluating the performance of diagnostic tests. The ROC curve is a comprehensive index reflecting the sensitivity and specificity of continuous variables, and it represents the relationship between sensitivity and specificity graphically. The most commonly used measure is the Area Under the Curve (AUC), derived from the receiver's operating characteristic curve. The R package "pROC" (version 1.18.0) was utilized to generate ROC curves and calculate the Area Under the Curve (AUC) for the selection of feature genes and evaluation of their diagnostic value[22]. The area under the ROC curve (AUC) generally ranges between 0.5 and 1, where a value closer to 1 indicates better diagnostic performance.

Immune Infiltration Analysis

Single-sample GSEA (ssGSEA) is a variation of the Gene Set Enrichment Analysis (GSEA) algorithm that provides a score for each sample and gene set pair instead of calculating enrichment scores for groups of samples (i.e., Control *vs.* Disease) and sets of genes (i.e., pathways). It calculates separate enrichment scores for each pairing of a sample and gene set. Each ssGSEA enrichment score indicates the extent to which the genes within a specific gene set are collectively up or downregulated in a sample.

Based on the 28 types of immune cells downloaded from the TISIDB (Tumor and Immune System Interactions Database) (http://cis.hku.hk/TISIDB/index.php), including Activated CD8 T cell, Central memory CD8 T cell, effector memory CD8 T cell, activated CD4 T cell, central memory CD4 T cell, effector memory CD4 T cell, T follicular helper cell, Gamma delta T cell, type 1 T helper cell, Type 17 T helper cell, Type 2 T helper cell, regulatory T cell, activated B cell, Immature B cell, memory B cell, natural killer cell, CD56bright natural killer cell, CD56dim natural killer cell, myeloid-derived suppressor cell, natural killer T cell, activated dendritic cell, plasmacytoid dendritic cell, immature dendritic cell, macrophage, eosinophil, mast cell, monocyte and neutrophil[23], every immunocyte's relative enrichment score was quantified from each sample's gene expression profile. Variations of the immune cell infiltration level among samples in RPL and control groups were illustrated through the R package ggplot2 (version 3.3.6)[24].

RBP-mRNA Network Construction

The widely used open-source platform StarBase (https://starbase.sysu.edu.cn/tutorialAPI.php#RBPTarget) was utilized for analysis of ncRNA interactions and to study the associations between mRNA and RNA-binding proteins (RBPs) expression using CLIP-seq, degradome-seq, and RNA-RNA interaction data. To identify key mRNA-RBP pairs in RPL, we defined the cutoff criteria as $p<0.05$, clusterNum≥5, and clipExpNum≥5. Subsequently, the RBP-mRNA network was built with the software Cytoscape (version 3.9.1).

Statistical Analysis

The statistical analyses were performed using R software version 4.1.2 (Vienna, Austria). The Spearman's correlation test was used to infer the correlation between two variables. Differences between two groups were assessed using the Wilcoxon test, while differences among three or more groups were evaluated using the Kruskal-Wallis test. We considered *p*-values lower than 0.05 in both tails to be indicative of statistical significance.

Results

DEGs Identification

Through the comparison of recurrent pregnancy loss samples and a healthy control group, a total of 299 differentially expressed genes (DEGs) were identified, and these genes showed statistically significant differences between the two groups ($p<0.05$, $|\text{Log}_2$ fold change$|>0.5$). Among the recurrent miscarriage samples, 154 genes were upregulated, and 145 genes were downregulated. All DEGs were visualized through a volcano plot (Figure 1A). In addition, a heat map was used to display the top 5 upregulated genes (*TSPAN14*, *PCGF1*, *EMD*, *ZDHHC9*, *FAM166B*) and the top 5 downregulated genes (*ZNF90*, *PSIP1*, *SNRPE*, *C2orf69*, *CCNYL1*) (Figure 1B). Rank-sum tests also revealed significant differences in the expression of the top ten genes between recurrent miscarriage samples and control samples ($p<0.05$, Figure 1C).

GSVA

To explore the functional annotations of recurrent pregnancy loss, we performed GSVA analysis to assess the relative expression differences of pathways between the two groups. The GSVA analysis enriched numerous pathways with distinct expressions visualized through a heatmap. Compared to the control group, the RPL group exhibited significantly lower expression in the KEGG_PROTEIN_EXPORT and KEGG_RENIN_ANGIOTENSIN_SYSTEM pathways, while the expression of pathways related to KEGG_TERPENOID_BACKBONE_BIOSYNTHESIS and KEGG_GLYCOSAMINOGLYCAN_DEGRADATION was significantly higher (Figure 2).

Construction of Weighted Gene Co-expression Network and Module Identification

WGCNA was conducted to study gene sets related to maternal-fetal immunity. Scale independence and average connectivity analysis showed that when the minimum soft threshold (β) was set to 14 (Figure 3A), the average connectivity was close to 0 and the scale independence was above 0.85. Five co-expression modules were identified, with irrelevant genes assigned to the grey module, which was disregarded in subsequent analysis (Figure 3B). To understand the relationships between modules and determine their relevance, module eigengenes (MEs) were correlated. A feature gene network was visualized through a dendrogram and a heatmap (Figure 3C). To comprehend the physiological significance of genes within modules, the 5 MEs were associated with maternal-fetal immunity, and the most significant associations were sought. Based on the heatmap of module-trait relationships (Figure 3D), genes clustered in the blue module (n=925) exhibited the strongest negative correlation with maternal-fetal immunity ($r=-0.64$, $p<0.05$). Consequently, the blue module was primarily considered as it may more accurately indicate maternal-fetal immunity.

A total of 13 immune-related DEGs were obtained by taking the intersection of DEGs and the genes in the maternal-fetal immune-related modules, which were regarded as hub genes (**Supplementary Table II**).

Hub Gene Interaction Analysis

We used the GeneMANIA database to construct a PPI network (Figure 4) for a total of 33 genes, including 13 hub genes with known interactions and 20 genes associated with the hub genes.

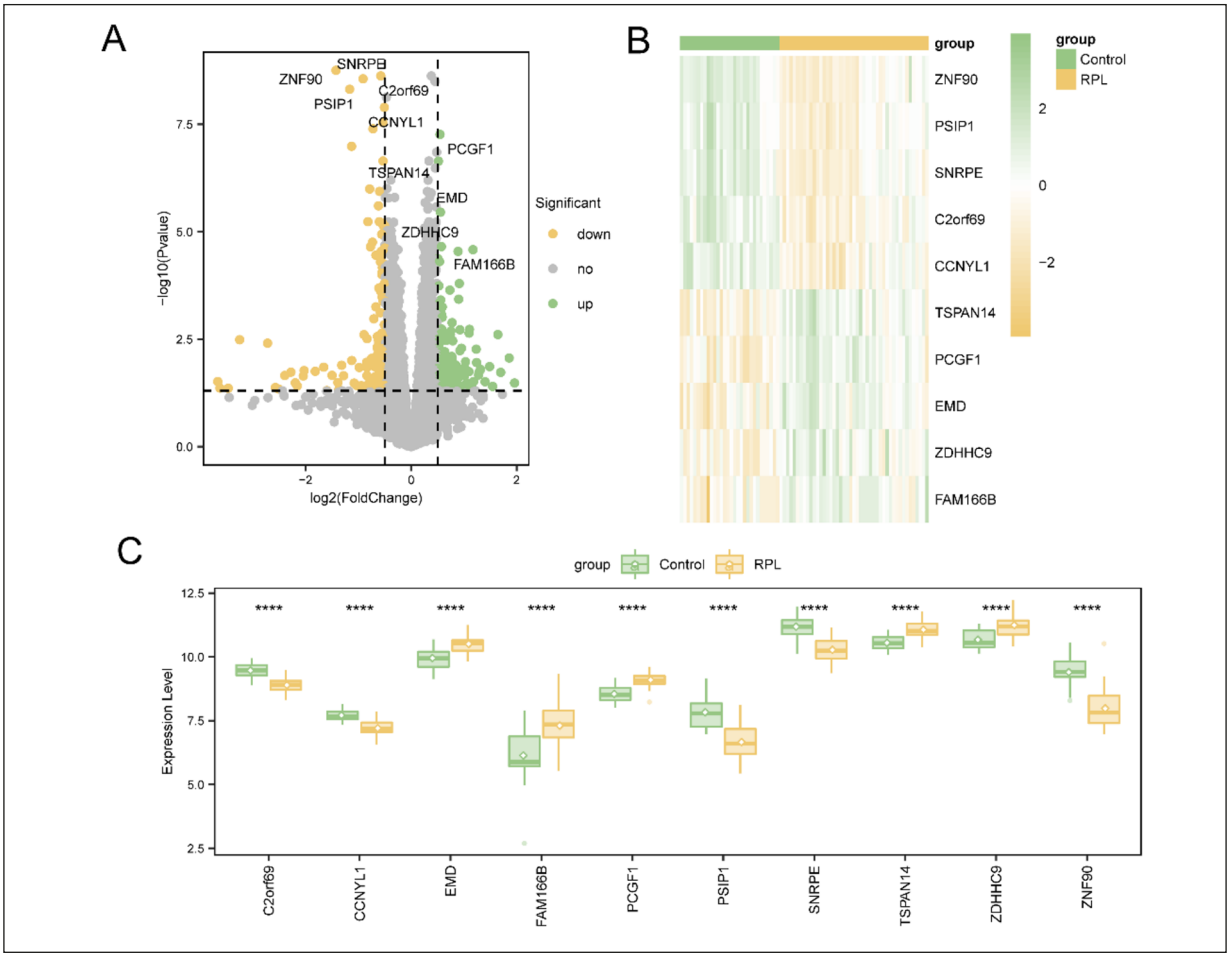

Figure 1. DEGs related to recurrent pregnancy loss (RPL). **A**, The volcano plot illustrates the distribution of DEGs between RPL and control group samples. Gray dots indicate genes with no significant expression correlation. **B**, The heatmap shows the top 5 upregulated and downregulated DEGs. **C**, The boxplot displays the expression level differences of genes between RPL and control group samples, with significance determined by the rank-sum test. The asterisks indicate the *p*-values (****p<0.0001, ***p<0.001, **p<0.01, *p<0.05).

Diagnostic Value of Hub Genes

To further validate the diagnostic value of hub genes, we used ROC curves to assess the hub genes. We found that *DEK* (AUC=0.813), *CCND2* (AUC=0.742), *UCP2* (AUC=0.736), *TRIM6* (AUC=0.715), *RASAL1* (AUC=0.707), *SLC38A5* (AUC=0.688), *CD109* (AUC=0.687), *SLC5A3* (AUC=0.686), *MYZAP* (AUC=0.677), *MFAP4* (AUC=0.675), *CCDC64B* (AUC=0.669), *PCSK5* (AUC=0.668), *P2RY14* (AUC=0.66) all had an area under the ROC curve (AUC) greater than 0.6 (Figure 5A-5L, (**Supplementary Figure 1**). This indicates that hub genes have discriminatory capacity as potential biomarkers for RPL.

Immune Infiltration

Immune cell infiltration may play an important role in the pathogenesis of recurrent pregnancy loss. Therefore, we investigated the association between RPL/control samples and infiltrating immune cells. Among the 28 types of immune cells, there were significant differences in the immune cell infiltration abundance between the two groups for 5 immune cell types (p<0.05) (Figure 6A). Among them, two immune cell types (CD56bright natural killer cell, Monocyte) showed significantly higher levels of infiltration in the RPL group compared to the control group (Figure 6A). The overall level of immune cell infiltration between the RPL and control groups is shown in Figure 6B.

We also examined the significant correlations between each hub gene and the corresponding immune cells. *CCDC64B* demonstrated a significant correlation with regulatory T cells (R=-0.593, p<0.001) (Figure 6C); *UCP2* exhib-

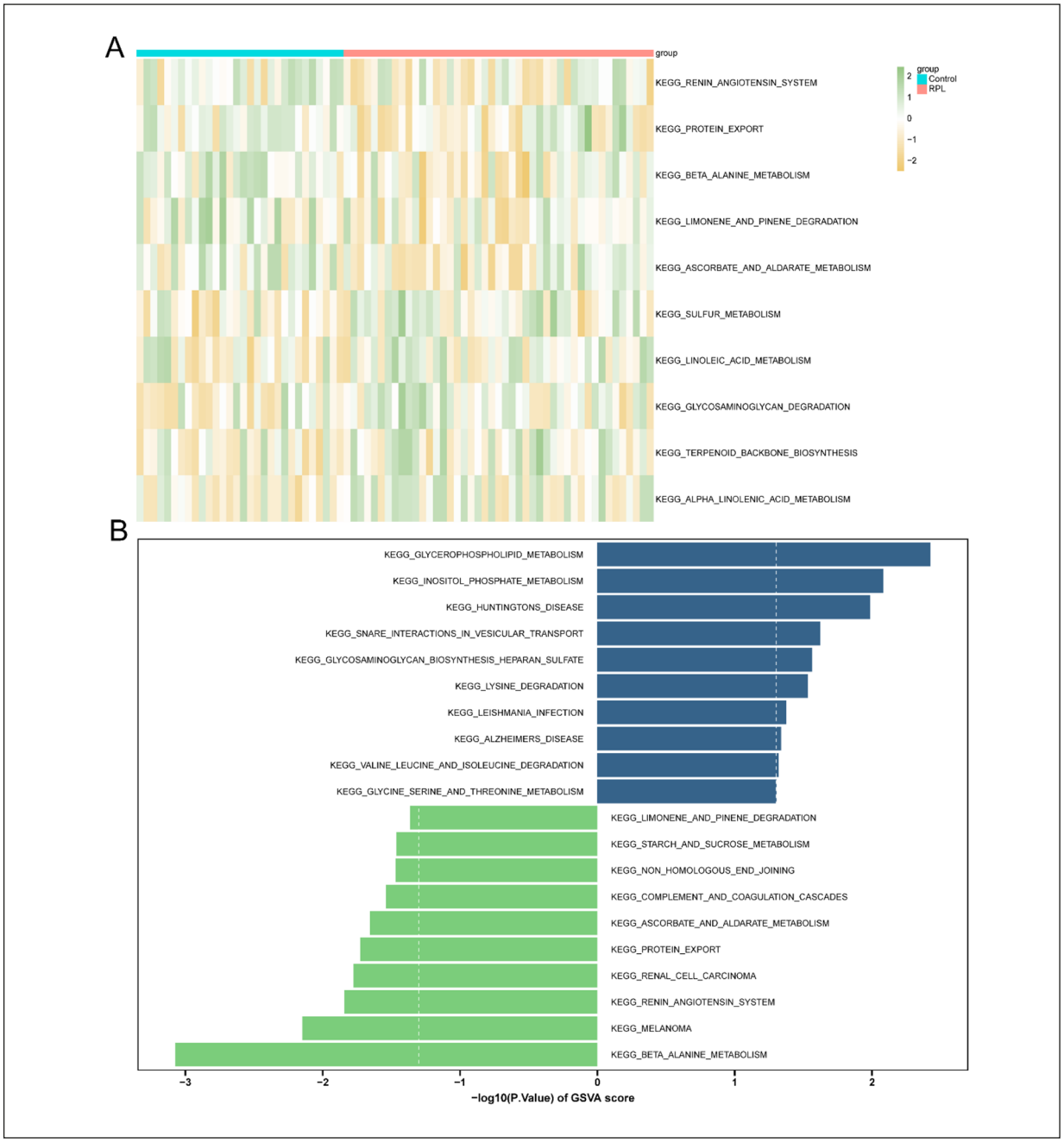

Figure 2. Differentially enriched pathways between RPL and control. **A**, Heatmap for significantly enriched pathways. **B**, Differences in pathway activities scored by GSVA.

ited a significant correlation with monocytes (R=0.621, p<0.001) (Figure 6D).

Additionally, correlations among immune cells were investigated, and there was a general positive correlation among immune cells (Figure 6E).

Signaling Pathways Related to Hub Genes

Further analysis using Gene Set Variation Analysis (GSVA) was conducted to investigate the differences between RPL patients and the control group in 50 Hallmark signaling pathways. In

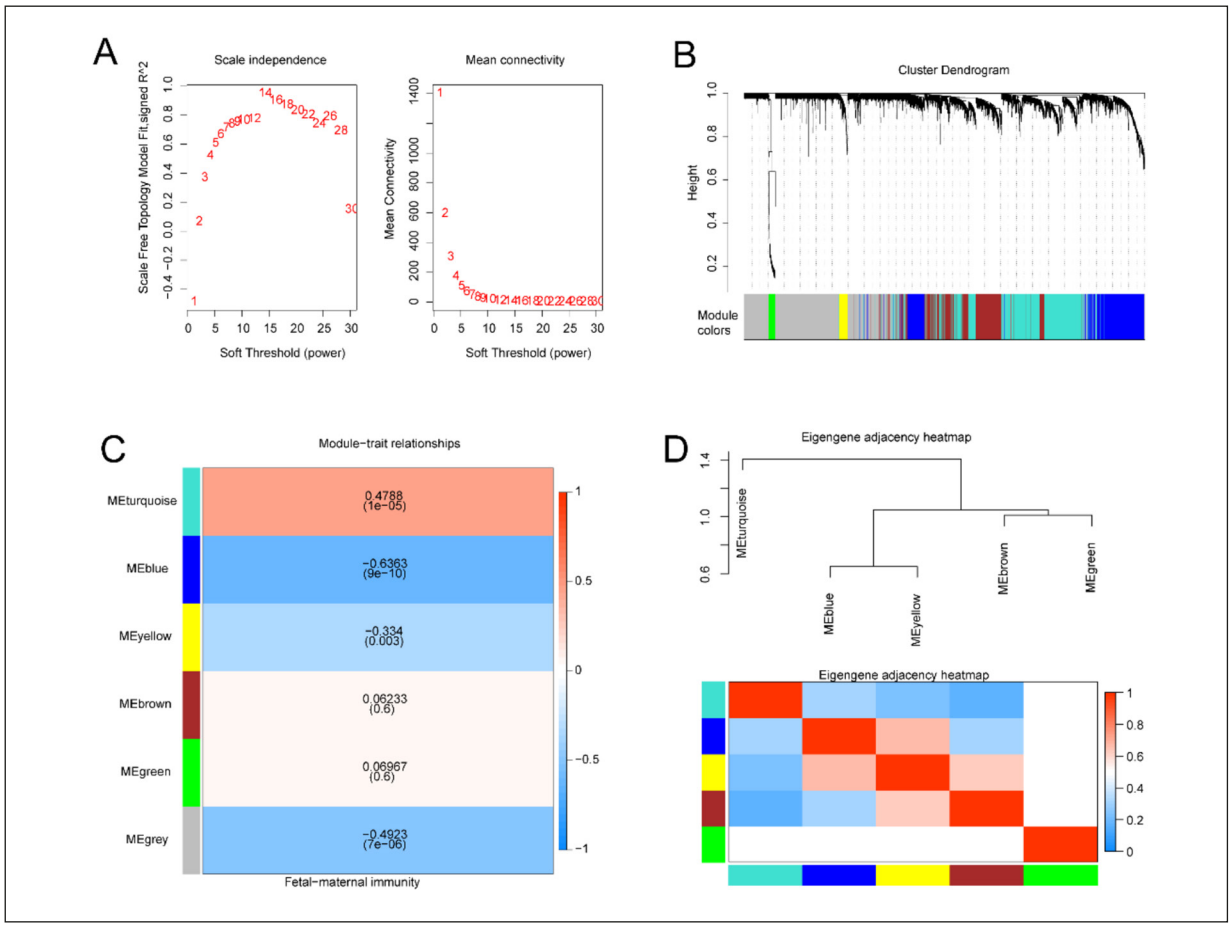

Figure 3. Construction of WGCNA Co-expression Network. **A**, Soft threshold β=14, scale-free topology fitting index (R2). **B**, Analysis of gene expression networks in RPL identified different modules of co-expression data. **C**, Relationships between modules. Top: Hierarchical clustering of module eigengenes summarizing the modules identified in the clustering analysis. Branches (meta-modules) in the dendrogram group positively correlated feature genes together. Bottom: Heatmap of correlations in the feature gene network. Each row and column in the heatmap corresponds to the feature genes of a module (color-coded). In the heatmap, red indicates high adjacency, and blue indicates low adjacency. Red squares along the diagonal represent meta-modules. **D**, Associations between consensus module eigengenes and maternal-fetal immunity. Each row in the table corresponds to a consensus module, and each column to a sample or trait. Numbers in the table report the correlations of the corresponding module eigengenes and traits, with the *p*-values printed below the correlations in parentheses. The color-coded legend indicates the color-coded representation of the correlations.

RPL patients, three Hallmark signaling pathways were significantly upregulated, which are HALLMARK_BILE_ACID_METABOLISM, HALLMARK_KRAS_SIGNALING_DN, HALLMARK_PEROXISOME. Nine pathways were significantly downregulated, including HALLMARK_ANGIOGENESIS, HALLMARK_APICAL_SURFACE, HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION, HALLMARK_ESTROGEN_RESPONSE_EARLY, HALLMARK_IL2_STAT5_SIGNALING, HALLMARK_IL6_JAK_STAT3_SIGNALING, HALLMARK_INTERFERON_ALPHA_RESPONSE, HALLMARK_KRAS_SIGNALING_UP, HALLMARK_UV_RESPONSE_DN (Figure 7A).

We also analyzed the correlation between the five most significantly differentially expressed hub genes and 50 Hallmark signaling pathways (Figure 7B).

Construction and Functional Annotation of the Crosstalk Between the Hub mRNAs and RBPs

To investigate the interaction between RNA binding proteins (RBPs) and mRNA, we utilized

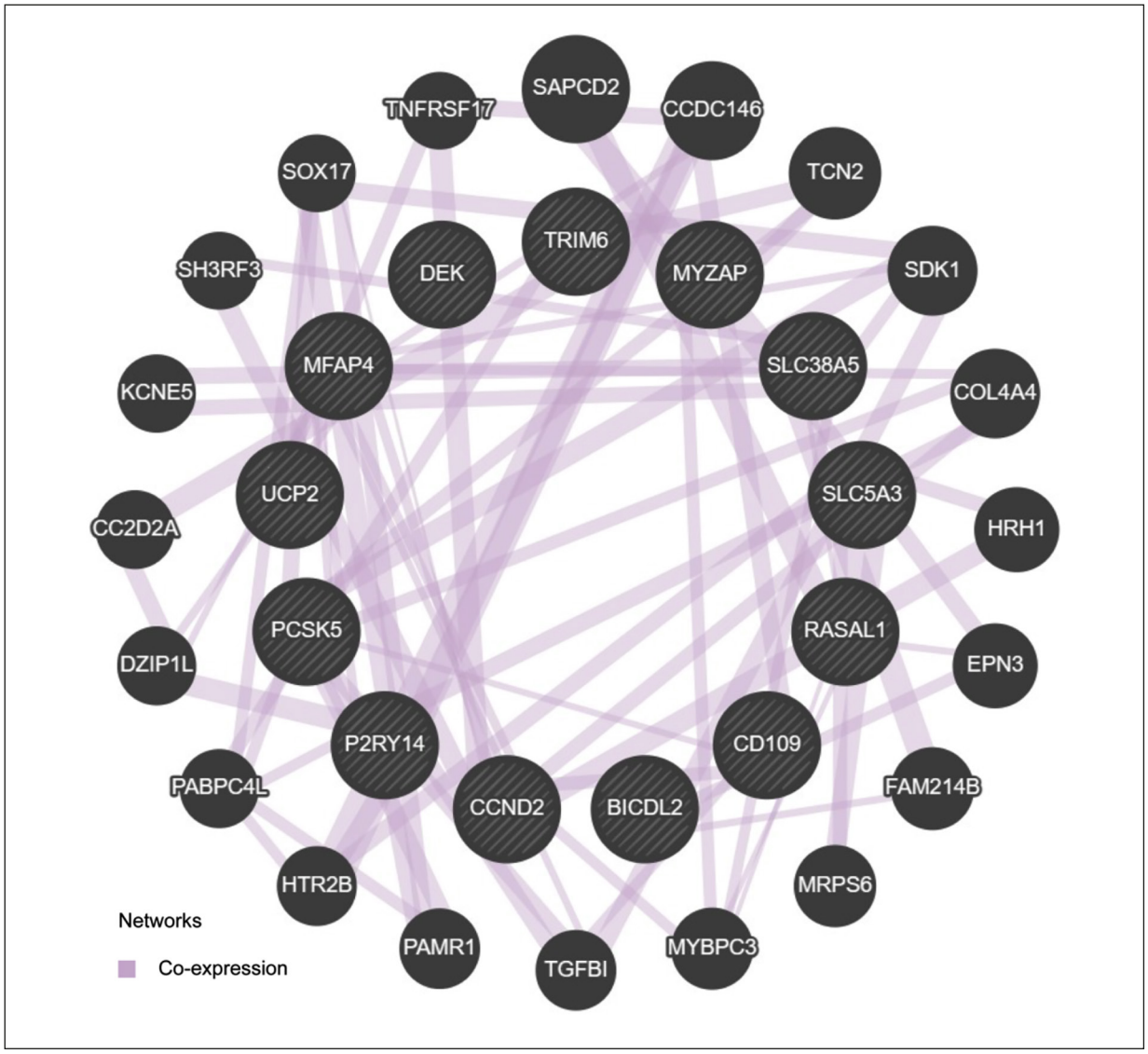

Figure 4. Interaction analysis of hub genes characterized gene co-expression network.

the StarBase online database to search for 13 hub mRNAs and retrieved the corresponding mRNA/RBP pairs for 11 hub mRNA. Using the provided relationships between target genes from the online dataset, we constructed an RBP-mRNA network comprising 76 nodes, 65 RBPs, 11 mRNAs, and 250 edges. Figure 8 illustrates the network.

Verification of Hub Genes

To further validate the robustness of hub genes as potential biomarkers, we conducted ROC analysis on the independent GEO dataset GSE161969 to validate the diagnostic performance of hub genes in RPL. In the dataset GSE161969 (**Supplementary Figure 2**), *PCSK5* (AUC=1), *CCND2* (AUC=0.917), *SLC5A3* (AUC=0.917), *RASAL1* (AUC=0.833), *MYZAP* (AUC=0.833), *MFAP4* (AUC=0.833), *P2RY14* (AUC=0.75) the validation results showed that the hub genes as potential biomarkers of RPL has strong robustness.

Discussion

Recurrent pregnancy loss (RPL) is a condition that significantly impacts women's reproductive health. The incidence of miscarriage is challenging to quantify, with most studies[1] indicating that natural miscarriage transpires in 15-25% of pregnancies. Recurrent pregnancy loss, as

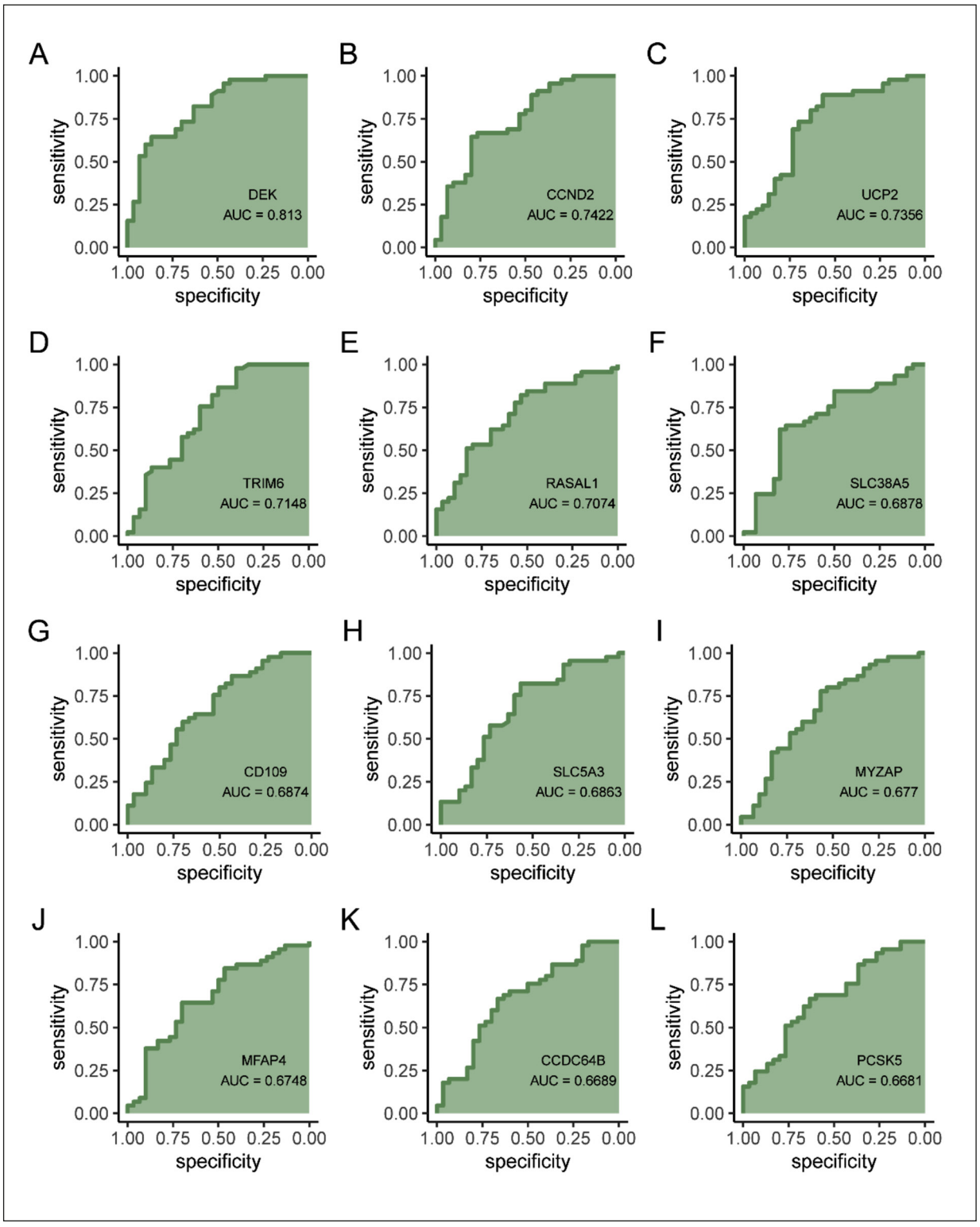

Figure 5. ROC curves of hub genes. **A**, *DEK*. **B**, *CCND2*. **C**, *UCP2*. **D**, *TRIM6*. **E**, *RASAL1*. **F**, *SLC38A5*. **G**, *CD109*. **H**, *SLC5A3*. **I**, *MYZAP*. **J**, *MFAP4*. **K**, *CCDC64B*. **L**, *PCSK5*.

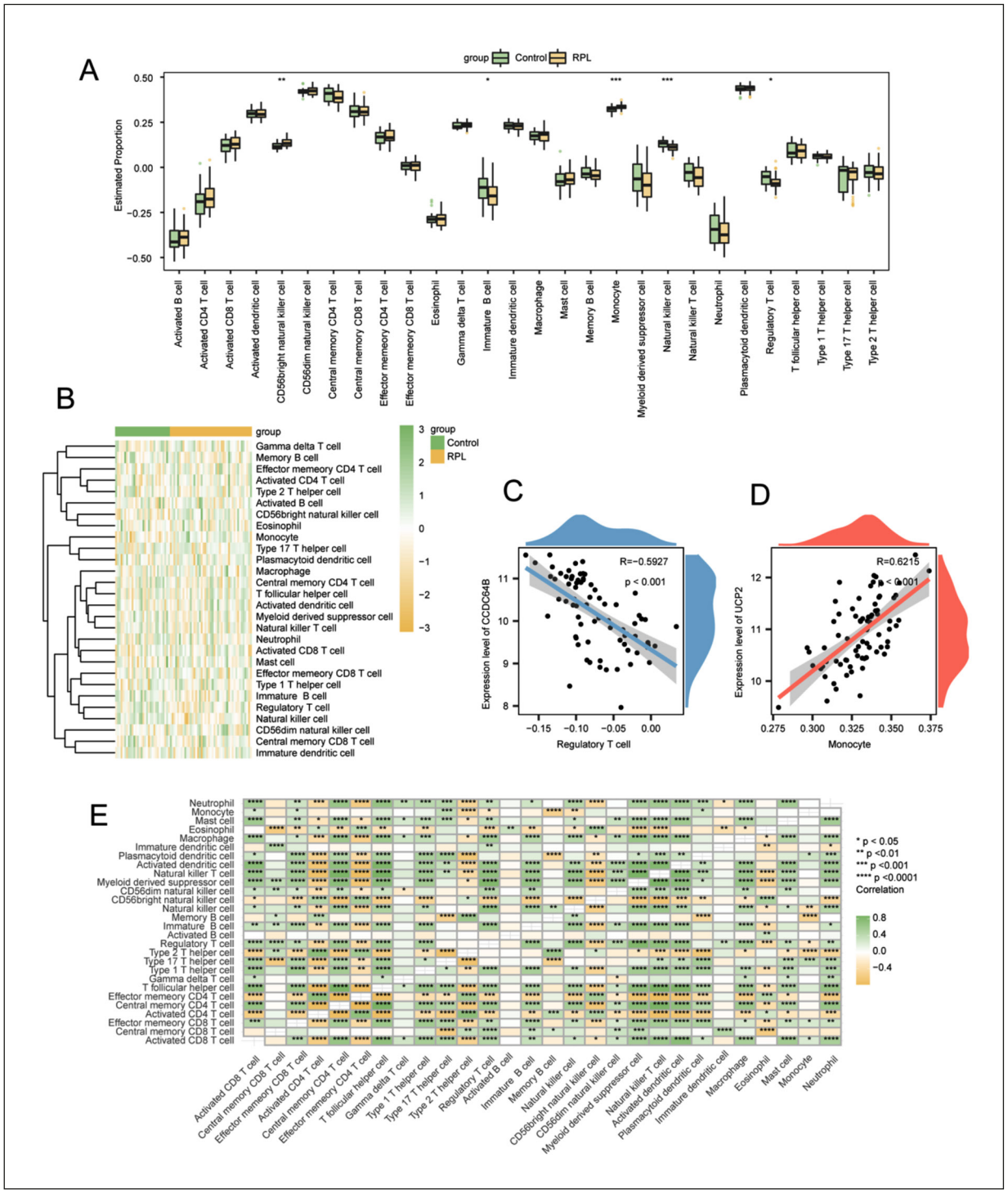

Figure 6. Differences in immune infiltration between the RPL and control groups. **A**, Bar plot showing the abundance of immune infiltration. **B**, Heatmap displaying the degree of immune infiltration. **C**, Scatter plot depicting the correlation between *UCP2* and monocytes. **D**, Scatter plot depicting the correlation between *CCDC64B* and regulatory T cells. **E**, Correlation among immune cells. Asterisks indicate *p*-values: ****$p<0.0001$, ***$p<0.001$, **$p<0.01$, *$p<0.05$.

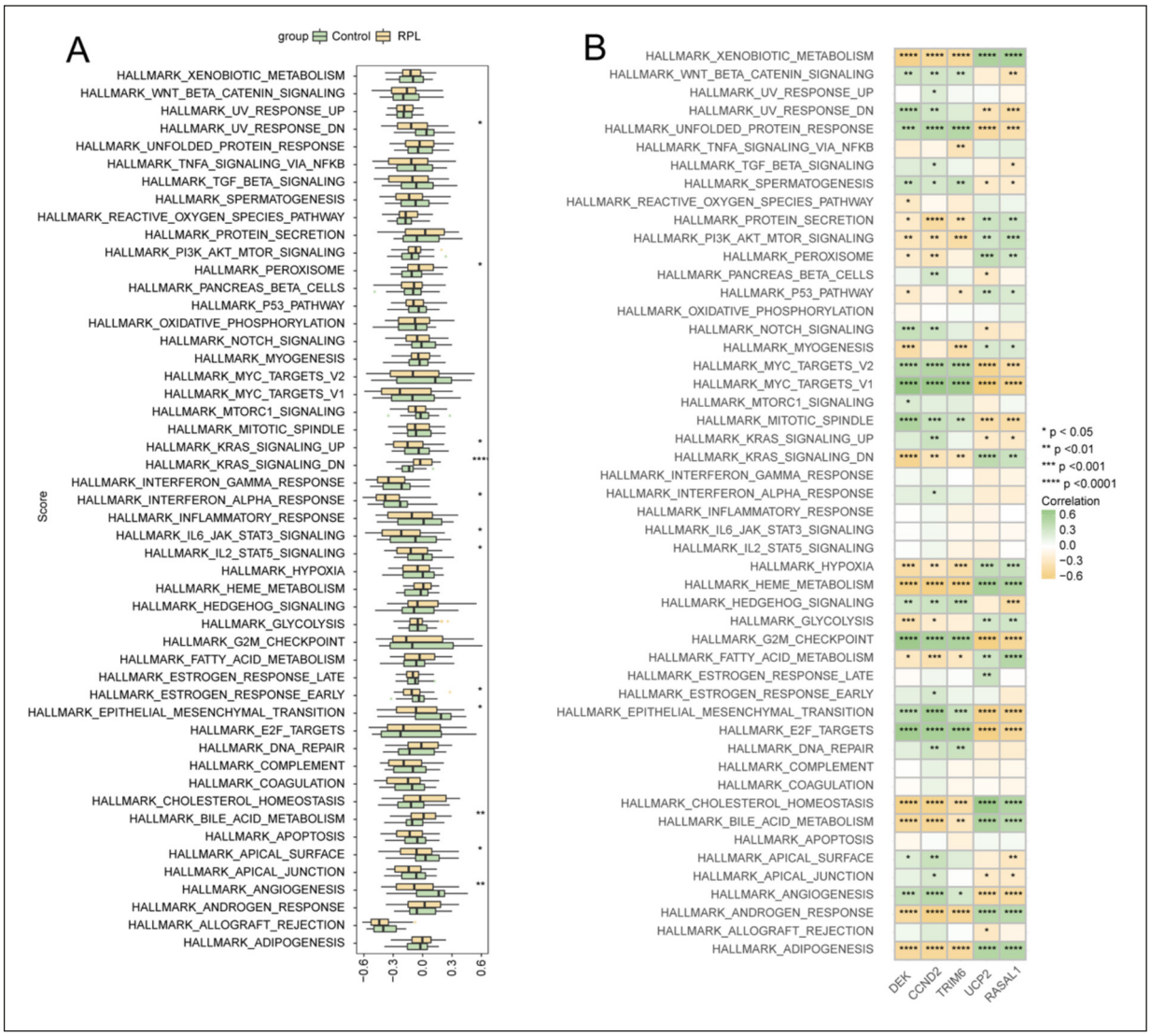

Figure 7. Correlation between hub genes and 50 HALLMARK signaling pathways. **A**, Comparison of 50 HALLMARK signaling pathways between the RPL group and the control group. **B**, Correlation between hub genes and 50 HALLMARK signaling pathways. ****p<0.0001, ***p<0.001, **p<0.01, *p<0.05.

defined by ESHRE[2,25], encompasses the consecutive loss of two or more pregnancies before 28 weeks in the same partner, and is observed in 1-2% of women. RPL imposes significant physical and psychological distress on women and places a substantial emotional burden on families. The etiology of RPL is intricate and characterized by substantial heterogeneity. Established causes encompass genetic, anatomical, endocrine, coagulation, infectious, immune, and psychological factors[1]. Notably, immune factors have emerged as a prominent and demanding area of investigation in reproductive research. Clinically, almost half of recurrent pregnancy loss instances are categorized as unexplained, yielding substantial challenges in clinical identification and early intervention. Furthermore, diagnosing recurrent pregnancy loss relies on the presence of prior miscarriages, posing a challenge in identifying miscarriage risks and implementing early interventions for women without a history of miscarriage. Therefore, the discovery of precise diagnostic markers and analysis of immune cell infiltration patterns in RPL holds immense significance in enhancing pregnancy outcomes for affected patients. The rapid advancements in science and technology have led to bioinformatics emerging as a potent

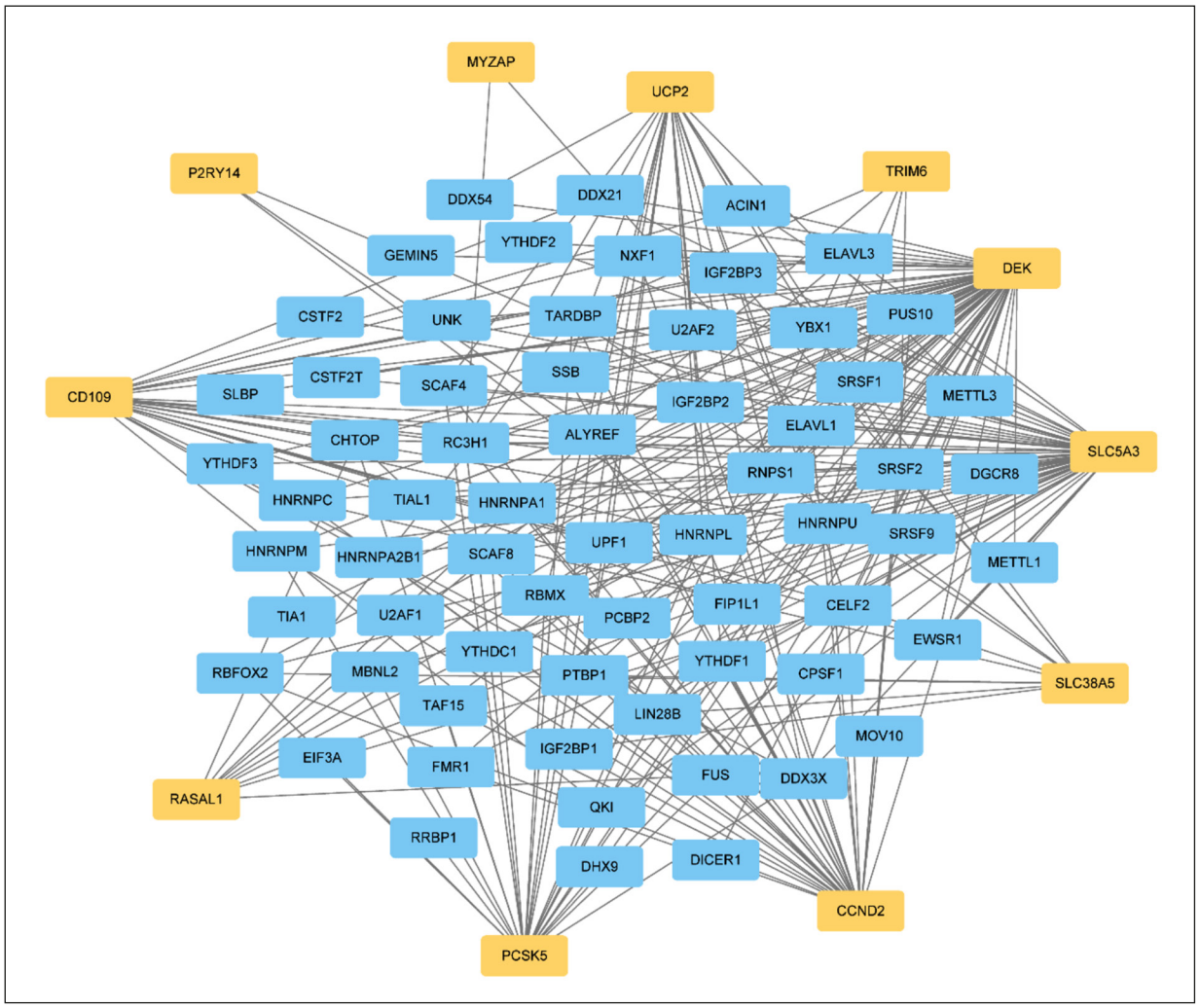

Figure 8. RBP-mRNA regulatory network. Blue represents RNA binding proteins (RBPs), and yellow represents mRNAs.

approach for molecular marker screening. In this study, we attempted to identify diagnostic markers for RPL and further explore the role of immune cell infiltration in RPL.

During the process of pregnancy, there are two major maternal-fetal immune interfaces. The first interface involves the interaction between maternal immune cells in the decidua and fetal trophoblast cells, which disappear along with the regression of invasive trophoblast cells and the degradation of related decidual lymphocytes in the third month of pregnancy. The second interface involves the interaction between maternal immune cells in the circulatory system and syncytiotrophoblast cells on the surface of the placental villi, which appear with the initiation of uteroplacental circulation at 8-9 weeks[26]. In the human placenta, the syncytiotrophoblast surface lacks MHC expression, which means it is incapable of triggering antigenic stimulation in maternal T cells. In this regard, it is widely accepted that immune interactions between the mother and fetus in humans may primarily be mediated by NK cells rather than T cells, like the local immune environment in the uterus[27].

Among the 13 hub genes identified in this study, namely *DEK*, *CCND2*, *UCP2*, *TRIM6*, *RASAL1*, *SLC38A5*, *CD109*, *SLC5A3*, *MYZAP*, *MFAP4*, *CCDC64B*, *PCSK5*, and *P2RY14*, *SLC38A5* and *SLC5A3* belong to the solute carrier family 38 (SLC38) and solute carrier family 5 (SLC5), respectively. The SLC38 functions as a transporter for amino acids, facilitating their movement from the extracellular environment into the cells. This transport of amino acids across the placenta, facilitated by amino acid transporters, plays a

crucial role in providing energy and nutrients for the developing fetus. Aberrations in placental amino acid transporters have been implicated in abnormalities in fetal growth and development[28]. Members of the solute carrier family 5 (SLC5) primarily mediate the absorption and transport of monosaccharides and polysaccharides[29]. Uncoupling protein 2 (*UCP2*) is located in the inner mitochondrial membrane and downregulates the production of reactive oxygen species (ROS). In models[30] of acute infection, *UCP2* exhibits regulatory effects on innate immunity by modulating ROS production, cytokine, and chemokine generation, as well as the recruitment of phagocytes. *CD109* is a surface molecule antigen belonging to the cluster of differentiation (CD) family. Within the transforming growth factor-beta (TGF-β)/Smads signaling pathway, *CD109* serves as a novel accessory receptor that negatively regulates TGF-β in human keratinocytes. It plays a significant role in the initiation and progression of specific tumors[31].

The AUC value of DEK was determined to be 0.813 based on the ROC curve analysis, suggesting its excellent diagnostic value. This implies that DEK exhibits a robust capability to distinguish RPL from the control group. Nonetheless, future validation of DEK as a biomarker for RPL will necessitate increasing the sample size. We further validated the potential biomarkers of recurrent pregnancy loss (RPL), and the results show that hub genes have strong robustness. ROC analysis conducted on the independent GEO dataset GSE161969 demonstrated the excellent diagnostic performance of hub genes as potential biomarkers. These results strongly support the robustness and accuracy of hub genes as potential biomarkers for recurrent pregnancy loss. These findings provide strong evidence for the potential application of hub genes in clinical diagnosis, offering new perspectives and methods for early screening and diagnosis of RPL.

GSEA provides valuable insights into genes with relatively smaller fold changes on a large scale. Through GSEA analysis of gene profiles from datasets, we discovered numerous gene sets that were highly enriched in the RPL group. One of these gene sets, 'RENIN ANGIOTENSIN SYSTEM', refers to the Renin-Angiotensin System (RAS) pathway, which plays a crucial role in regulating blood pressure, electrolyte balance, and fluid homeostasis. Another gene set, 'PROTEIN EXPORT', involves the general biological process of protein export, which entails the transportation of proteins from the cytoplasm to specific cellular compartments or extracellular spaces. Additionally, 'TERPENOID BACKBONE BIOSYNTHESIS' corresponds to the terpenoid backbone biosynthesis pathway in the KEGG database. Terpenoids, also known as isoprenoids, are a diverse class of natural products found in various organisms, including plants, fungi, and bacteria. They have important roles in cellular processes and demonstrate a wide range of biological activities. Lastly, 'KEGG GLYCOSAMINOGLYCAN DEGRADATION' represents the glycosaminoglycan (GAG) degradation pathway. Glycosaminoglycans are complex polysaccharides present in the extracellular matrix and cell surfaces of various organs and tissues. They play essential roles in cellular processes, tissue development, and structural integrity.

To investigate the impact of immune cell infiltration in RPL further, we conducted a thorough evaluation of RPL immune infiltration using ssGSEA. Our findings suggest a potential association between increased infiltration of CD56bright natural killer cells and monocytes and the occurrence and development of RPL. Upon analyzing the correlation between *UCP2* and immune cells, a significant positive correlation with monocytes was observed. Monocytes serve as precursor cells for macrophages in the bloodstream. Within the decidual immune cell population, decidual macrophages comprise 10%. Previous studies[32] have substantiated the pivotal role of decidual macrophages in establishing and maintaining immune tolerance at the maternal-fetal interface, promoting trophoblast invasion, facilitating angiogenesis, remodeling spiral arteries, and engulfing apoptotic cells. Moreover, preliminary research[33] has indicated the involvement of *UCP2* in regulating inflammation and metabolic processes within macrophages. Drawing upon these findings, we hypothesize that *UCP2* may impact the functionality of monocytes and macrophages, thus potentially contributing to the occurrence of RPL.

Conclusions

Our study identified the DEGs, WGCNA modules, hub genes, enriched pathways, and infiltrated immune cells that likely contribute to the development of RPL. These findings enhance our understanding of the underlying mechanisms of RPL and have the potential to influence future advancements in disease theranostics.

Conflict of Interest
The authors declare that they have no conflict of interests.

Acknowledgements
We express our gratitude to the Gene Expression Omnibus (GEO) database for providing valuable data free of charge for scientific research purposes.

Authors' Contribution
MJ and BH conducted the statistical analysis and wrote the original draft. DL and BH contributed to picture processing and article reviewing. LM contributed to the conception and design of the study. LM and JH provided valuable scientific suggestions, and supervision, and made final revisions to the manuscript.

Funding
This research received no external funding.

Ethics Approval
This study is conducted based on the GEO database and is not involved in animal or human experiments. The patients involved in the database have obtained ethical approval. Users are able to download pertinent data free of charge for research purposes and may publish relevant articles based on the data acquired. Our study is based on open-source data, so there are no ethical issues in this study.

Data Availability
Publicly available datasets were analyzed in this study. The data can be accessed at the GEO website: GSE165004, GSE180485, GSE26787, GSE161969.

Informed Consent
No informed consent was required.

ORCID ID
Lin Ma: 0000-0001-8545-3380

References

1) Practice Committee of the American Society for Reproductive Medicine. Evaluation and treatment of recurrent pregnancy loss: a committee opinion. Fertil Steril 2012; 98: 1103-1111.
2) The ESHRE Guideline Group on RPL, Bender Atik R, Christiansen OB, Elson J, Kolte AM, Lewis S, Middeldorp S, Nelen W, Peramo B, Quenby S, Vermeulen N, Goddijn M. ESHRE guideline: recurrent pregnancy loss. Hum Reprod Open 2018; 2018: hoy004.
3) Alijotas-Reig J, Garrido-Gimenez C. Current Concepts and New Trends in the Diagnosis and Management of Recurrent Miscarriage. Obstet Gynecol Surv 2013; 68: 445-466.
4) Coomarasamy A, Dhillon-Smith RK, Papadopoulou A, Al-Memar M, Brewin J, Abrahams VM, Maheshwari A, Christiansen OB, Stephenson MD, Goddijn M, Oladapo OT, Wijeyaratne CN, Bick D, Shehata H, Small R, Bennett PR, Regan L, Rai R, Bourne T, Kaur R, Pickering O, Brosens JJ, Devall AJ, Gallos ID, Quenby S. Recurrent miscarriage: evidence to accelerate action. Lancet 2021; 397: 1675-1682.
5) Moffett A, Shreeve N. Local immune recognition of trophoblast in early human pregnancy: controversies and questions. Nat Rev Immunol 2023; 23: 222-235.
6) Xu YS, Liao RY, Huang D, Wang D, Zhang L, Li YZ. Evidence from Mendelian randomization: increased risk of miscarriage in patients with asthma. Eur Rev Med Pharmacol Sci 2023; 27: 11587-11596.
7) Xie M, Li Y, Meng YZ, Xu P, Yang YG, Dong S, He J, Hu Z. Uterine Natural Killer Cells: A Rising Star in Human Pregnancy Regulation. Front Immunol 2022; 13: 918550.
8) Wang F, Qualls AE, Marques-Fernandez L, Colucci F. Biology and pathology of the uterine microenvironment and its natural killer cells. Cell Mol Immunol 2021; 18: 2101-2113.
9) Saito S, Nakashima A, Shima T, Ito M. Th1/Th2/Th17 and Regulatory T-Cell Paradigm in Pregnancy: Th1/Th2/Th17 AND TREG CELLS IN PREGNANCY. Am J Reprod Immunol 2010; 63: 601-610.
10) Fu B, Li X, Sun R, Tong X, Ling B, Tian Z, Wei H. Natural killer cells promote immune tolerance by regulating inflammatory T H 17 cells at the human maternal–fetal interface. Proc Natl Acad Sci 2013; 110: E231-E240.
11) Laird SM. A review of immune cells and molecules in women with recurrent miscarriage. Hum Reprod Update 2003; 9: 163-174.
12) Chen X, Mariee N, Jiang L, Liu Y, Wang CC, Li TC, Laird S. Measurement of uterine natural killer cell percentage in the periimplantation endometrium from fertile women and women with recurrent reproductive failure: establishment of a reference range. Am J Obstet Gynecol 2017; 217: 680.e1-680.e6.
13) Sci-Hub | Dynamic Function and Composition Changes of Immune Cells During Normal and Pathological Pregnancy at the Maternal-Fetal Interface. Frontiers in Immunology, 10 | 10.3389/fimmu.2019.02317 [Internet]. [cited 2023 Aug 15]. Available from: https://sci-hub.ru/10.3389/fimmu.2019.02317.
14) Tajti F, Kuppe C, Antoranz A, Ibrahim MM, Kim H, Ceccarelli F, Holland CH, Olauson H, Floege J, Alexopoulos LG, Kramann R, Saez-Rodriguez J. A Functional Landscape of CKD Entities From Public Transcriptomic Data. Kidney Int Rep 2020; 5: 211-224.

15) Malone JH, Oliver B. Microarrays, deep sequencing and the true measure of the transcriptome. BMC Biol 2011; 9: 34.

16) Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics 2012; 28: 882-883.

17) Li Q, Chen S, Dong X, Fu S, Zhang T, Zheng W, Tian Y, Huang D. The Progress of Research on Genetic Factors of Recurrent Pregnancy Loss. Genet Res 2023; 2023: 1-16.

18) Alecsandru D, Klimczak AM, Garcia Velasco JA, Pirtea P, Franasiak JM. Immunologic causes and thrombophilia in recurrent pregnancy loss. Fertil Steril 2021; 115: 561-566.

19) Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 2015; 43: e47-e47.

20) Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 2008; 9: 559.

21) Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, Franz M, Grouios C, Kazi F, Lopes CT, Maitland A, Mostafavi S, Montojo J, Shao Q, Wright G, Bader GD, Morris Q. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. Nucleic Acids Res 2010; 38: W214-W220.

22) Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, Müller M. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics 2011; 12: 77.

23) Ru B, Wong CN, Tong Y, Zhong JY, Zhong SSW, Wu WC, Chu KC, Wong CY, Lau CY, Chen I, Chan NW, Zhang J. TISIDB: an integrated repository portal for tumor–immune system interactions. Bioinformatics 2019; 35: 4200-4202.

24) Ito K, Murphy D. Application of ggplot2 to Pharmacometric Graphics. CPT Pharmacomet Syst Pharmacol 2013; 2: 79.

25) The ESHRE Guideline Group on RPL, Bender Atik R, Christiansen OB, Elson J, Kolte AM, Lewis S, Middeldorp S, Mcheik S, Peramo B, Quenby S, Nielsen HS, Van Der Hoorn M-L, Vermeulen N, Goddijn M. ESHRE guideline: recurrent pregnancy loss: an update in 2022. Hum Reprod Open 2022; 2023: hoad002.

26) Sargent IL, Borzychowski AM, Redman CWG. NK cells and human pregnancy – an inflammatory view. Trends Immunol 2006; 27: 399-404.

27) Hiby SE, Walker JJ, O'Shaughnessy KM, Redman CWG, Carrington M, Trowsdale J, Moffett A. Combinations of Maternal KIR and Fetal HLA-C Genes Influence the Risk of Preeclampsia and Reproductive Success. J Exp Med 2004; 200: 957-965.

28) Farley DM, Choi J, Dudley DJ, Li C, Jenkins SL, Myatt L, Nathanielsz PW. Placental Amino Acid Transport and Placental Leptin Resistance in Pregnancies Complicated by Maternal Obesity. Placenta 2010; 31: 718-724.

29) Matoba S, Nakamuta S, Miura K, Hirose M, Shiura H, Kohda T, Nakamuta N, Ogura A. Paternal knockout of Slc38a4 /SNAT4 causes placental hypoplasia associated with intrauterine growth restriction in mice. Proc Natl Acad Sci 2019; 116: 21047-21053.

30) Rousset S, Emre Y, Joinlambert O, Hurtaud C, Ricquier D, Cassarddoulcier A. The uncoupling protein 2 modulates the cytokine balance in innate immunity. Cytokine 2006; 35: 135-142.

31) Tanabe M, Hosokawa K, Nguyen MAT, Nakagawa N, Maruyama K, Tsuji N, Urushihara R, Espinoza L, Elbadry MI, Mohiuddin M, Katagiri T, Ono M, Fujiwara H, Chonabayashi K, Yoshida Y, Yamazaki H, Hirao A, Nakao S. The GPI-anchored protein CD109 protects hematopoietic progenitor cells from undergoing erythroid differentiation induced by TGF-β. Leukemia 2022; 36: 847-855.

32) Parasar P, Guru N, Nayak NR. Contribution of macrophages to fetomaternal immunological tolerance. Hum Immunol 2021; 82: 325-331.

33) Van Dierendonck XAMH, Sancerni T, Alves-Guerra M-C, Stienstra R. The role of uncoupling protein 2 in macrophages and its impact on obesity-induced adipose tissue inflammation and insulin resistance. J Biol Chem 2020; 295: 17535-17548.