Journal of Multidisciplinary Healthcare

Dovepress
Taylor & Francis Group

Open Access Full Text Article

ORIGINAL RESEARCH

The Role of Vitamin D Metabolism-Related Genes in Recurrent Pregnancy Loss and Their Immune Microenvironmental Changes

Jiangmei He, Hongmei Liu, Jingru Ji

Department of Eugenics and Genetics, First Hospital of Shanxi Medical University, Taiyuan City, Shanxi Province, People's Republic of China

Correspondence: Jiangmei He, Email acc0351@163.com

Purpose: The importance of vitamin D metabolism has been confirmed in various pregnancy complications. It is unknown, henceforth how vitamin D metabolism contributes to the occurrence of recurrent pregnancy loss (RPL). This study aimed to elucidate its potential mechanisms through bioinformatics analysis.

Methods: Weighted Gene Co-expression Network Analysis (WGCNA) was used to identify module genes linked to vitamin D metabolism after transcriptome datasets were examined to identify differentially expressed genes (DEGs). Machine learning was utilized to refine and identify candidate biomarkers, while Mendelian randomization (MR) assessed their causal relationships with RPL. In addition, the expression was further verified by RT-qPCR and Western blotting. Finally, scRNA-seq uncovered cellular heterogeneity and intercellular communication networks.

Results: We identified 379 DEGs in RPL samples. WGCNA revealed two key modules strongly correlated with vitamin D metabolism. The intersection of DEGs and key module genes yielded 27 candidate genes related to vitamin D metabolism. Machine learning identified DOCK11 and ETV2 as biomarkers, showing consistent expression trends across training and validation sets, both demonstrating AUC values greater than 0.7 in ROC analysis. Functional enrichment analysis indicated that DOCK11 and ETV2 were co-enriched in the pathways of inflammatory responses, interferon gamma response, and TNAF signaling via NFKB. Experimental validation yielded the same results. Single-cell analysis revealed 16 distinct cellular clusters with significant enrichment of DOCK11 and ETV2 in Natural Killer cells, highlighting altered immune interactions in RPL through enhanced signaling from NK cells and cytotoxic CD8+ T cells while reducing signals from macrophages.

Conclusion: This study identified DOCK11 and ETV2 as biomarkers for RPL, revealing the important involvement of NK cells in RPL and providing new directions for the treatment of RPL.

Keywords: recurrent pregnancy loss, vitamin D metabolism, machine learning, single-cell analysis

Introduction

Recurrent pregnancy loss (RPL) refers to describe the occurrence of two or more consecutive pregnancy losses from the moment of conception until the embryo reaches viability or before 24 weeks of gestation.[1] Epidemiological studies indicate that the incidence of RPL is approximately 1% to 5%, and with the full implementation of China's three-child policy and the increasing maternal age, its incidence is rising annually.[2] Currently, RPL has become one of the globally recognized intractable diseases, with complex and diverse etiologies. Genetic factors, uterine anatomical abnormalities, endocrine disorders, viral infections, environmental variables, and immunological factors are the main known causes of RPL.[3,4] RPL patients endure significant physical and psychological stress due to repeated pregnancy failures. Since the exact causes of RPL remain unclear, clinical interventions are relatively limited in their specificity. Helping this population achieve favorable pregnancy outcomes has become a major challenge in the field of reproductive medicine. Therefore, identifying effective biomarkers is of great significance for the early diagnosis of RPL and for exploring its underlying mechanisms.

Received: 20 May 2025
Accepted: 21 October 2025
Published: 21 November 2025

Vitamin D, as a multifunctional steroid hormone, has physiological roles that extend beyond the classical regulation of calcium and phosphorus metabolism. In recent years, its function in maintaining immune homeostasis and suppressing inflammatory responses has gained significant attention.[5] CYP2R1 and the nuclear vitamin D receptor (VDR) are two key genes involved in metabolizing vitamin D into its biologically active form and mediating its corresponding functions. The active metabolite 1,25-dihydroxyvitamin D3 ($1,25(OH)_2D_3$) exerts extensive biological effects by binding to VDR, and experimental studies have demonstrated its potent immunoregulatory effects on various tissues.[6,7] During pregnancy, vitamin D metabolism undergoes dynamic changes. In early pregnancy, the expression of VDR is significantly upregulated in the trophoblast and decidua, along with increased expression of 1α-hydroxylase (CYP27B1), which activates vitamin D. However, compared to normal woman, woman with recurrent miscarriage exhibit lower expression levels of CYP27B1 in the villi and decidua, suggesting that reduced CYP27B1 expression may be associated with recurrent miscarriage.[8–10] Additionally, it has been shown that vitamin D metabolism plays a significant role in a number of pregnancy problems, such as preterm birth, gestational diabetes, and preeclampsia.[11] Since vitamin D plays a significant role in regulating the immune system, a deficiency of vitamin D may increase the susceptibility of pregnant woman to certain pathogens, thereby leading to the occurrence of respiratory diseases, tuberculosis, influenza and COVID-19, and indirectly affecting the stability of pregnancy.[12] Studies have shown that a deficiency of vitamin D can increase the risk of RPL.[13–15] In addition, compared with normal Vit D-RPL and control groups, vitamin D deficiency is associated with increased activity of the NK cells.[14] However, how vitamin D metabolism influences the development of recurrent miscarriage and its underlying molecular mechanisms remain unclear.

An innovative method that provides information at the individual cell level is single-cell RNA sequencing, or scRNA-seq.[16] Given RPL's considerable variability and several stages of evolution, scRNA-seq shows promise as an essential technique for revealing its intricate processes. By integrating multiple biochemical techniques, scRNA-seq enables in-depth analysis of disease-related cells from various perspectives and at multiple levels.[17] Additionally, Mendelian randomization (MR) analysis has gained increasing attention in recent years for leveraging genome-wide association study (GWAS) data to test causal relationships between exposures and disease outcomes. This method not only allows for the measurement of multiple potential biomarkers but also facilitates the evaluation of other traits co-regulated by the exposure [10.1111/rssa.12343]. In our work, we utilized transcriptomic data of RPL from public databases and applied bioinformatics approaches to identify biomarkers associated with vitamin D metabolism in recurrent pregnancy loss. We evaluated the expression levels and diagnostic potential of these biomarkers and employed MR analysis to explore their potential causal relationships with RPL. Finally, we conducted enrichment analysis and immune infiltration analysis to investigate the potential mechanisms by which these biomarkers influence RPL progression. At single-cell level, we further examined the expression of biomarkers, providing novel insights for the clinical prevention and diagnosis of RPL.

Materials and Methods

Data Source

Two transcriptome datasets related to RPL were acquired from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/gds). The training dataset GSE165004 was generated using the GPL16699 platform, and the samples were derived from a total of 48 (control: RPL = 24: 24) endometrial tissue samples of patients with RPL. The validation dataset GSE26787 was based on the GPL570 platform and included 10 (control: RPL = 5: 5) endometrial tissue samples of patients with RPL. In addition, 328 Vitamin D metabolism related genes were collected from the msigdb database (https://www.gsea-msigdb.org/gsea/msigdb). Subsequently, a single-cell dataset was downloaded from the GSA database (https://ngdc.cncb.ac.cn/gsa/browse/CRA002181). The dataset with the accession number CRA002181 contained 3 control samples and 2 samples of patients with RPL.

Differential Expression and Enrichment Analyses

Differentially expressed genes (DEGs) were identified from training set by comparing RPL and control samples, with the criteria set as $|\log_2\text{Fold Change (FC)}| > 0.5$ and FDR < 0.05, applying limma package (v 3.54.0).[18] Visualization of the DEGs

was achieved through a volcano plot and heatmap, created leveraging ggplot2 (v 3.4.1)[19] and ComplexHeatmap (v 2.14.0)[20] packages, respectively.

Weighted Gene Co-Expression Network Analysis (WGCNA)

To further investigate Vitamin D metabolism-related module genes, WGCNA was conducted in training set, using Vitamin D metabolism score as the trait, with WGCNA package (v 1.70.3).[21] Specifically, hierarchical clustering was first performed on the samples to exclude outliers. A soft-thresholding power was then determined by setting the scale-free fit index (R^2) close to 0.8 while maintaining average connectivity near zero, ensuring that the network approximated a scale-free topology. A gene dendrogram was subsequently constructed based on gene similarity, and a correlation matrix between module eigengenes and phenotypic traits was calculated. Lastly, a corresponding heatmap was generated to visualize these correlations. The module with the highest absolute correlation to Vitamin D metabolism was selected as the key module ($|cor| > 0.3$ and $P < 0.05$), and genes with module membership (MM) > 0.7 and gene significance (GS) > 0.65 in the module were selected as key module genes.

Identification and Enrichment Analysis of Candidate Genes

Next, ggvenn package was utilized to identify candidate genes between the DEGs and key module genes. With a view to gain further insight into the signaling pathways and biological mechanisms associated with the candidate genes, GO (adj. $P < 0.05$) and KEGG ($P < 0.05$) enrichment analyses were performed using clusterProfiler package (v 4.6.2).[22]

Recognition of Potential Biomarkers Using Machine Learning Algorithms

After the candidate genes had been identified, they were incorporated into the LASSO regression analysis by means of the glmnet package (v 4.1.8).[23] The optimal model was constructed to acquire the feature genes at the juncture where the model error was minimized and the lambda (λ) value was also minimized. By introducing the λ value, the unimportant feature coefficients can be compressed to zero, thereby reducing the complexity of the model. Subsequently, further analyses were carried out using Boruta package (v 8.0.0),[24] for the purpose of screening out significant genes. Eventually, the feature genes obtained through above two algorithms were intersected to obtain the candidate biomarkers.

Identification of Biomarkers

To further screen the biomarkers, the candidate biomarkers obtained from the machine learning screening were subjected to expression analysis, and the genes with consistent and significant expression trends were selected for subsequent analysis. The diagnostic capability of these potential biomarkers was then assessed. In both the training and validation sets, with RPL status serving as the outcome variable, we drew upon the pROC package (v 1.18.5)[25] to craft individual gene ROC curves and calculate AUC. A biomarker with an AUC exceeding 0.7 signified its competent discrimination between RPL and control samples, thus it was classified as a biomarker of our interest.

MR Analysis

This study followed the Mendelian reporting specifications for randomised studies (STROBE-MR).[26] To further explore the causal relationship between biomarkers and RPL, we regarded these biomarkers as exposure factors and RPL as the outcome. The cis-expression quantitative trait loci (cis-eQTLs) data of these biomarkers were obtained from the eQTLGen Consortium (https://www.eqtlgen.org/) and the IEU OpenGWAS database. Similarly, the GWAS data of RPL (denoted as "finngen_R11_N14_HABITABORT") was sourced from the FREEZE 11 database (https://r11.finngen.fi/), which included a cohort of 120,200 Europeans, with 732 cases of RPL and 119,468 controls.

The MR analysis was based on the following three fundamental premises: (1) There is a significant and robust correlation between instrumental variables (IVs) and biomarkers; (2) IVs are independent of confounding factors; (3) IVs affect RPL only through the biomarker pathway. Before commencing the analysis, we used the extract_instrument function in the TwoSampleMR package (v 0.6.3)[27] to select the appropriate IVs. Under the strict threshold ($P < 1\times10^{-5}$), we selected single nucleotide polymorphisms (SNPs) that had a significant association with the exposure factor. To ensure the robustness and relevance, we excluded SNPs that had a significant association with the exposure factor less

than three times during the selection process. We also set parameters such as "r2 = 0.1 and kb = 100" to ensure that SNPs in linkage disequilibrium (LD) were excluded, thereby maintaining the independence of SNPs. In addition, the F-statistic was calculated to measure the robustness of these SNPs, and an F-value higher than 10 indicated that the set of instrumental variables was reliable. Then, we used the harmonise_data function to align the effect alleles with their respective effect sizes, laying the foundation for subsequent analysis and ensuring the integrity and consistency of the data set. The MR analysis was conducted by using the mr function combined with five algorithms: MR-Egger, weighted median, inverse variance weighted (IVW), simple mode, and weighted mode. In this study, the IVW method could effectively elucidate the causal relationship and found a statistically significant association between biomarkers and RPL ($P < 0.05$). Finally, the robustness of the MR study results was evaluated by sensitivity analysis, including tests of heterogeneity, horizontal pleiotropy, and leave-one-out (LOO) method.

Immune Infiltration Analysis

The composition of immune cells within the immune microenvironment of patients with RPL was evaluated by computing the infiltration of 22 types of immune cells in both RPL samples and control samples within the training set, utilizing the CIBERSORT algorithm along with the LM22 gene set.[28] The disparities in the proportions of immune cells between the groups were analyzed, and the correlation between the immune cells and biomarkers was explored through Spearman correlation analysis.

Functional and Annotation Analyses

Following this, the analysis focused on delineating the pathways affected by biomarkers in the progression of RPL. The "c2.cp.kegg_legacy.v2024.1.Hs.symbols.gmt and h.all.v2023.2.Hs.symbols.gmt" were obtained from MSigDB (https://www.gsea-msigdb.org/gsea/msigdb) to serve as the gene set. First, "Spearman" correlation analysis was performed between each biomarker and all genes, followed by Gene Set Enrichment Analysis (GSEA) pathway enrichment analysis using the clusterProfiler package (|NES|>1, adj.P < 0.05).

Development of Regulatory Network and Drug Prediction Studies

Predictions for drugs targeting biomarkers were sourced from DSigDB database (https://dsigdb.tanlab.org/). The highest scoring drug for each biomarker underwent molecular docking. Active compounds' 3D structures were initially downloaded in SDF format from PubChem database (https://pubchem.ncbi.nlm.nih.gov/), converted to PDB format via Babel GUI, and further processed. The corresponding protein structures were acquired from PDB (https://www.rcsb.org/). Using PyMOL software (v 3.1.1),[29] water molecules and small ligands were removed from these protein structures. AutoDock software (v 1.5.7)[30] facilitated the molecular docking process, with docking results visualized in PyMOL.

ScRNA-Seq Analysis

After the transcriptomic analysis, to further explore the mechanisms of RPL at the cellular level and understand the heterogeneity of biomarkers at this scale, a series of single-cell analyses were conducted within the CRA002181 dataset. Initially, cells in the dataset underwent quality control (QC), clustering, and annotation with the support of the Seurat package (v 5.1.0).[31] Metrics such as gene count, cell count, and the percentage of mitochondrial genes were computed. Cells with less than 200 genes and less than 3 cells covered with genes were filtered out, while cells with more than 10% mitochondrial genes were filtered out, cells with ≤ 200 and ≥ 3000 genes in the cell were removed, and genes with count numbers ≤ 200 and ≥ 20000 were removed. Changes in nFeature_RNA, nCount_RNA, and percent_mt before and after QC were displayed, with cells meeting the criteria included in downstream analysis. Next, data were normalized, and the top 2000 highly variable genes were identified via vst method. PCA was performed to rank the variance percentage of each principal component and to generate a PCA elbow plot. Significant PCs enriched with genes having the lowest *P*-values were retained by JackStrawPlot function, and the appropriate PCs were chosen for further analysis. Unsupervised clustering was then conducted through FindNeighbors and FindClusters functions, with cell clusters visualized using UMAP method at a resolution of 0.2. Following clustering, cell types were annotated based on marker

genes from CellMarker database, and a bubble plot was created to display the specificity of marker genes across these cell types.

Identification and Cell Communication Analysis of Key Cells

We further analyzed the expression profiles of the biomarkers at the single-cell level to investigate their expression in annotated cells of RPL and normal samples. Subsequently, we constructed cellular communication analyses to detect the expression and pairing of receptors and ligands between cell types in order to infer interactions between different cells. The distribution of annotated cells relative to other cell types in RPL and normal samples was shown in the form of proportions and numbers. Subsequently, communication analyses were performed using the CellChat software package (v 1.6.1),[32] where the aggregation of cell-cell communication networks was calculated and signals were visualised for each cell population.

Reverse Transcription-Quantitative Polymerase Chain Reaction (RT-qPCR)

Ten samples (5 RPL samples, 5 normal samples) were procured from the First Hospital of Shanxi Medical University, with all patients providing informed consent and approval granted by the Ethics Committee of the First Hospital of Shanxi Medical University. The guidelines outlined in the Declaration of Helsinki were followed. RNA was extracted using the FastPure Complex Tissue/Cell Total RNA Isolation Kit (Vazyme, NJ). RNA purity was assessed using a Nano-500 microspectrophotometer. cDNA synthesis was performed using ABScript III RT Master Mix with gDNA Remover (RK20429, ABclonal, Wuhan). RT-qPCR analysis was carried out with the Genious 2X SYBR Green Fast RT-qPCR Mix (RK21205, ABclonal, Wuhan), and GAPDH served as the internal reference gene. Primer sequences were listed in Table 1. To guarantee the precision of the experimental findings, each experimental group was subjected to three separate replicates. Relative gene expression levels were calculated using the $2^{-\Delta\Delta Ct}$ method.

Western Blotting

The protein expression levels of biomarkers were analyzed for six samples (3 RPL samples, 3 normal samples). Protein extraction was performed tissues from RPL and control using RIPA lysis buffer (Beyotime, Shanghai, China), and protein concentrations were measured via a BCA assay kit (Beyotime, Shanghai). Proteins were mixed with 5× loading buffer (Servicebio, Beijing) at a 4:1 ratio, denatured at 95°C for 10 minutes, and stored at −20°C or −80°C. SDS-PAGE was conducted using separating and stacking gels tailored to protein molecular weights. Stacking gel electrophoresis was run at 80 V for 30–40 minutes, followed by separating gel electrophoresis at 120 V until the marker reached the bottom. Proteins were transferred onto PVDF membranes at 200 mA for 1 hour in an ice bath. Membranes were blocked with 5% skimmed milk for 30 minutes after TBST rinsing. Primary antibodies were applied overnight at 4°C, followed by incubation with secondary antibodies (1:5000 dilution) for 30 minutes at room temperature. Finally, membranes were treated with ECL luminescent solution and visualized using a chemiluminescence imaging system. The experiment was conducted with biological replicates (N = 3).

Table 1 The Primer Sequence of Biomarkers

Primer	Sequences 5'-3'
H-GAPDH	F:5'-GGAGTCCACTGGCGTCTTCA −3'
	R:5'-GTCATGAGTCCTTCCACGATACC −3'
DOCK11	F:5'-CAGTGACGGTAGCCCAAAGG-3'
	R:5'-GCACAGTGTGTAATGTTTCCCTG-3'
ETV2	F:5'-CTGGAAAGGTACAAGCTCATCC-3'
	R:5'-AACTTCTGGGTGCAGTAACGC-3'
SOAT1	F:5'-GGTGCGCTCTCACAACCTTT-3'
	R:5'-GAGGTGCTCTCAAATCCTTCG-3'

https://doi.org/10.2147/JMDH.S541670

Statistical Analysis

All analyses were executed in R software (v 4.2.2). Differences between groups were analyzed by Wilcoxon test. $P < 0.05$ was considered statistically significant.

Results

Identification of DEGs and Key Modules Associated with Vitamin D Metabolism in RPL Samples

This analysis of the RPL and control samples within training set revealed 379 DEGs: 167 up-regulated and 212 down-regulated (Figure 1A and B). WGCNA was performed with vitamin D metabolism as the trait to screen key module genes for vitamin D metabolism. The analysis revealed no outlier samples in the training set and confirmed a scale-free network topology when optimal soft-thresholding power was set to 12 (Figure 1C). A total of 11 modules were identified through

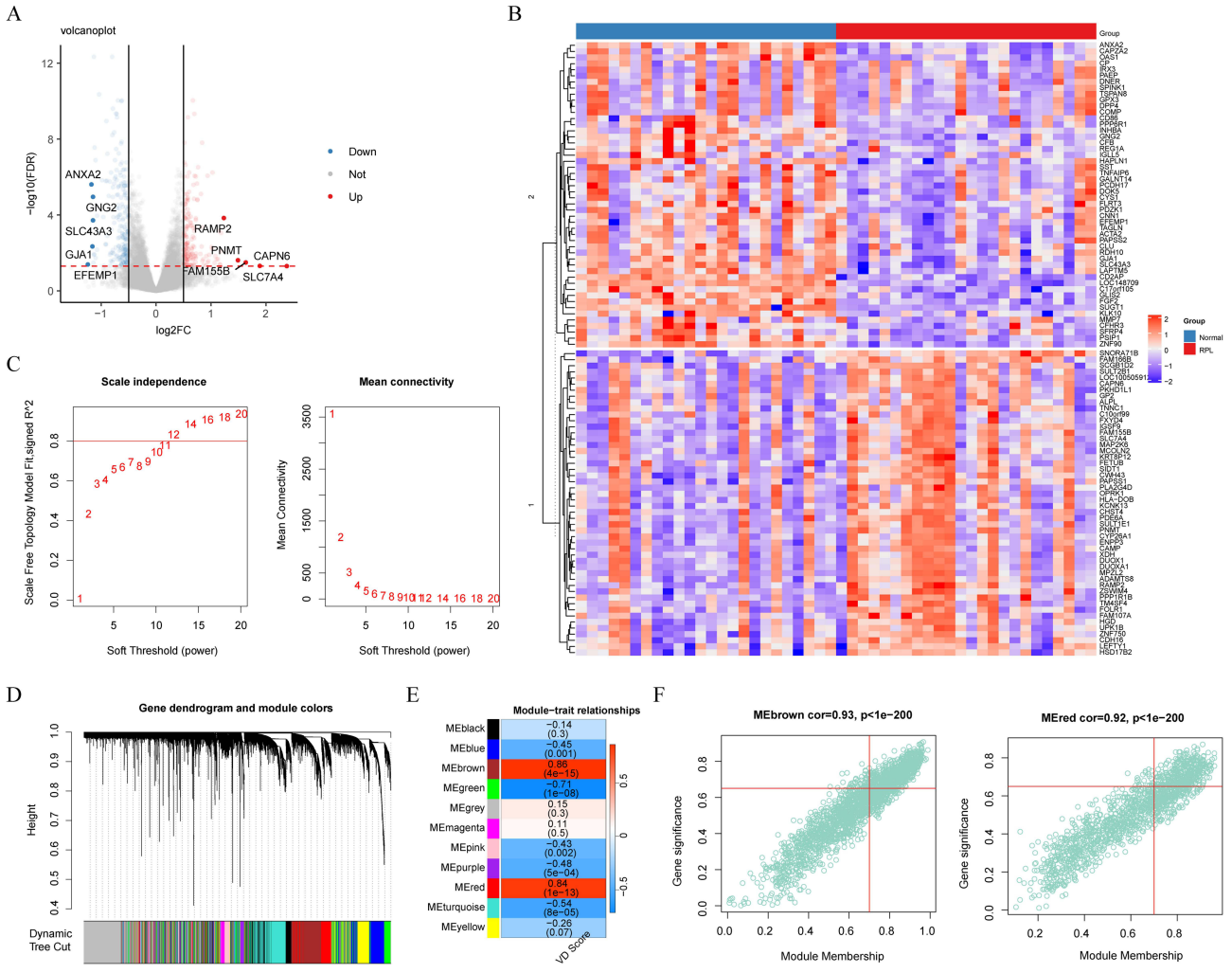

Figure 1 Identification of differentially expressed genes (DEGs) and key modules associated with vitamin d metabolism in recurrent pregnancy loss (RPL) samples (**A** and **B**) Volcano plot and heatmap illustrating DEGs between RPL and control subjects. The red dotted line in Figure A represents the value of -log10(0.05), and this line was used to determine whether the p value is significant. (**C**) Topology analysis depicting soft-thresholding power in network construction. The horizontal axis of the figure represents the power value of the weight parameter, and the vertical axis of the figure on the left represents the square of the fitting coefficients of log(k) and log(p(k)) in the corresponding network. The vertical axis on the right represents the mean of all gene adjacency functions in the corresponding gene module. The red line was the soft threshold selection judgment line, with R^2=0.8. The first soft threshold higher than this line was selected. (**D**) Hierarchical clustering dendrogram categorizing genes into modules based on expression patterns. Different colors represent different modules, where grey defaults to genes that cannot be classified to any module. (**E**) Heatmap of module-trait correlations. (**F**) Correlation analysis between key modules and gene importance in WGCNA. The vertical axis represents the correlation between genes and diseases, and the horizontal axis represents the correlation between genes and modules. Select the genes in the upper right quadrant, that is, those with a disease correlation greater than 0.65 and a module correlation greater than 0.7 (with the red line as the dividing line), as the module genes.

systematic clustering tree analysis among genes (Figure 1D). Among them, the MEbrown module (cor = 0.86, P = 4×10^{-15}) and the MEred module (cor = 0.84, P = 1×10^{-13}) showed a relatively strong positive correlation with vitamin D metabolism (Figure 1E). Subsequently, the genes with module similarity greater than 0.7 and trait correlation greater than 0.65 in each module were selected as the key genes of the corresponding modules. In total, there were 970 key module genes screened out in this way (Figure 1F).

Identification and Functional Enrichment of Candidate Genes Associated with Vitamin D Metabolism in RPL

After screening the DEGs and the key module genes, in order to identify the genes related to vitamin D metabolism in the context of RPL, we took the intersection of DEGs and the key module genes. It was found that there were 27 genes related to vitamin D metabolism among the DEGs, and these genes were named candidate genes (Figure 2A).

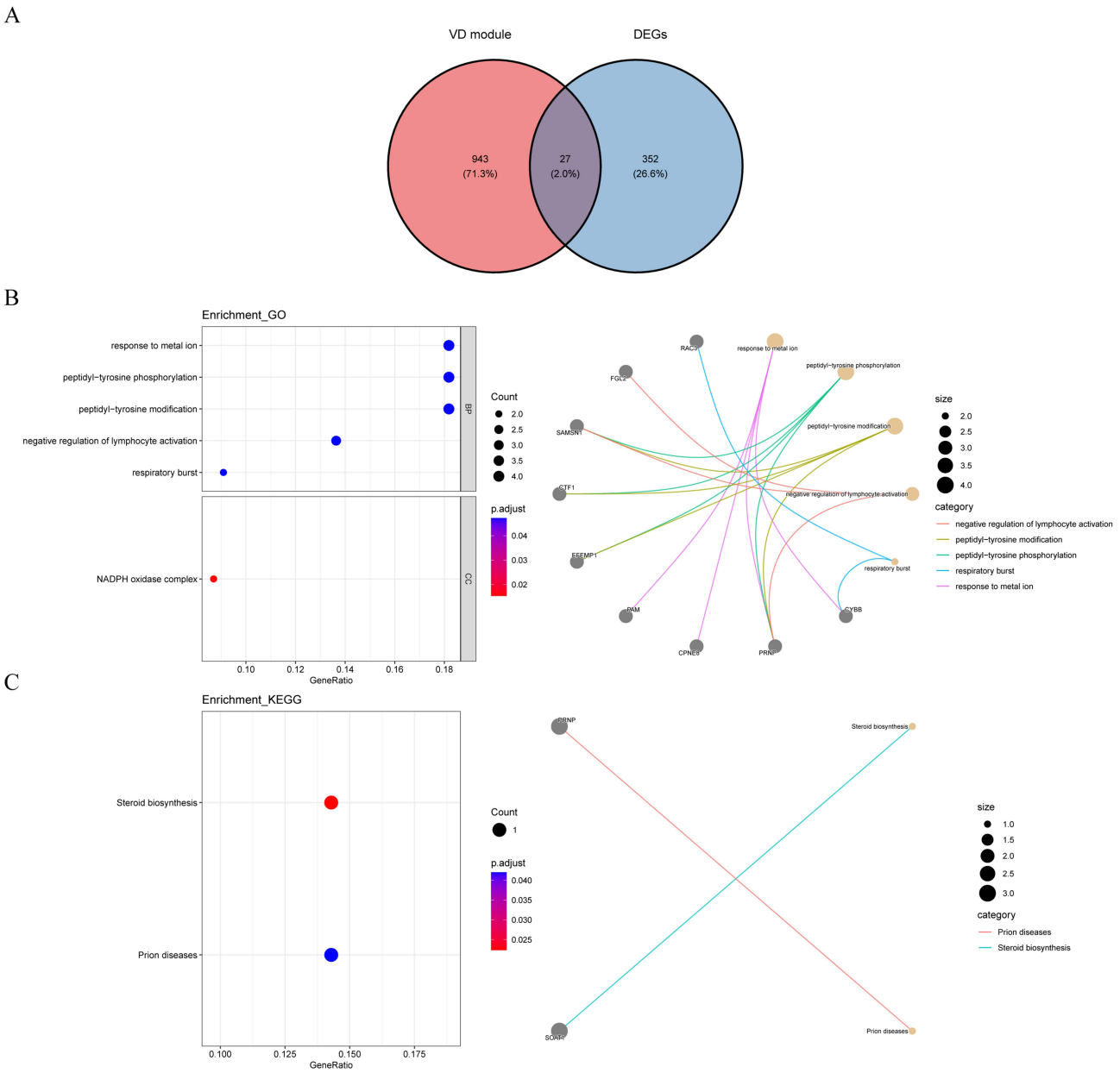

Figure 2 Identification and analysis of candidate genes. (**A**) Venn diagram identifying candidate genes by overlapping DEGs and key module genes. (**B** and **C**) Functional enrichment analysis of candidate genes through Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways.

https://doi.org/10.2147/JMDH.S541670

Subsequently, to gain a deeper understanding of the potential biological functions and the involved pathways of these candidate genes, we carried out enrichment analysis. Regarding the GO analysis, the candidate genes were enriched in a total of 17 GO terms, including 16 BP terms and 1 CC term (Figure 2B), specifically such as "response to metal ion", "peptidyl-tyrosine phosphorylation", and "NADPH oxidase complex", this complex participated in redox reactions within the cell. In terms of the KEGG analysis, only two pathways were enriched, namely the "Steroid biosynthesis" pathway and "Prion diseases" pathway (Figure 2C).

DOCK11 and ETV2 Were Determined as Biomarkers

Subsequently, further screening of biomarkers was carried out by means of machine learning. The results of the LASSO analysis showed that when Lambda.min was set to 0.05, the optimal number of genes was 5, specifically SOAT1, DOCK11, CYBB, RAC3, and ETV2 (Figure 3A). Further, using the Boruta algorithm to rank according to importance, 8 feature genes were obtained (Figure 3B). By taking the intersection of the feature genes obtained from the above two algorithms, a total of 3 candidate biomarkers were obtained, specifically DOCK11, SOAT1, and ETV2 (Figure 3C).

DOCK11 and ETV2 were determined as biomarkers through the verification of gene expression levels and ROC analysis. The differential expression analysis revealed consistent expression patterns for both genes across the training and validation sets, with statistically significant differences observed. Specifically, the expression of DOCK11 was downregulated and the expression of ETV2 was upregulated in RPL samples (Figure 4A). RT-qPCR analysis revealed an upregulation of ETV2 expression in RPL (Figure 4B) (Table S1). Similarly, Western blot results demonstrated a downregulation of DOCK11 and an upregulation of ETV2, aligning with the findings reported by bioinformatics (Figure 4C) (Table S2).

In addition, the results of the ROC analysis showed that in both the training set and the validation set, the AUC values of these two genes were greater than 0.7, indicating that they had good diagnostic value (Figure 4D). The MR analysis did not reveal a causal relationship between the biomarkers and RPL, suggesting that these biomarkers were not direct contributors to the onset of the disease. Instead, it was likely that the occurrence of the disease leads to changes in gene

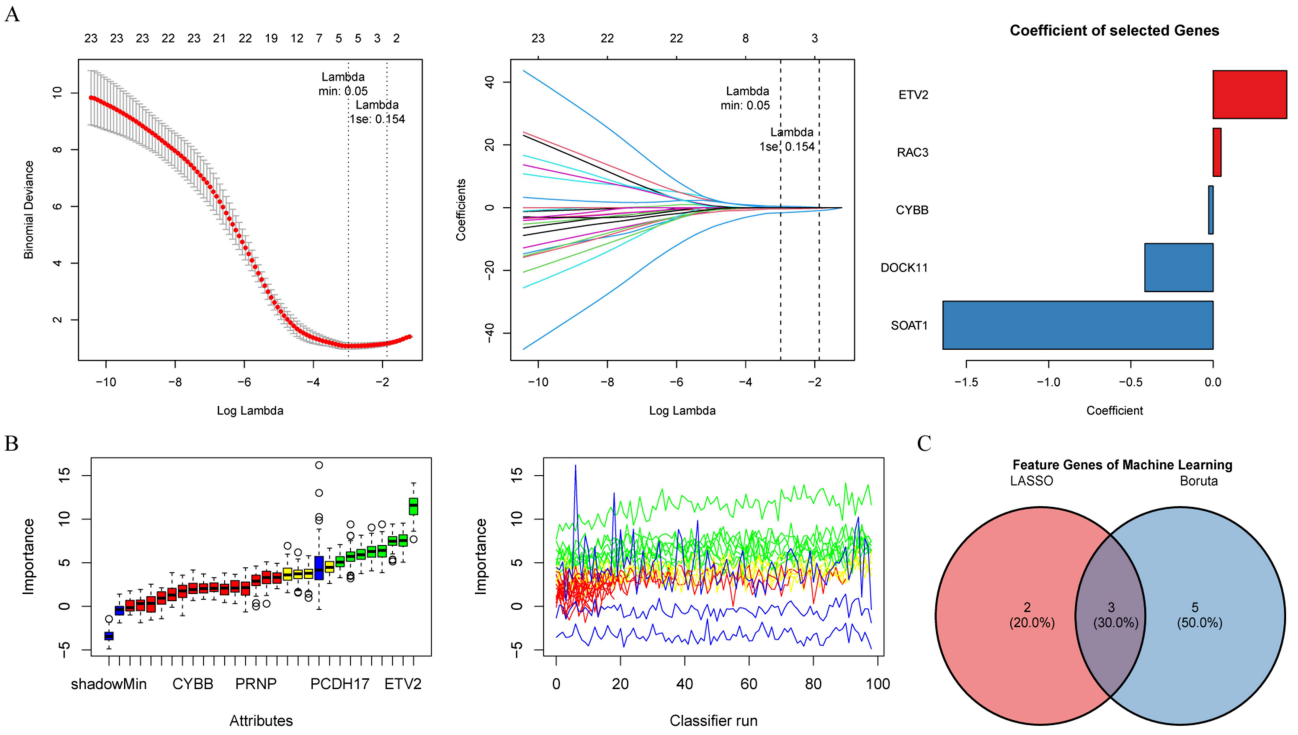

Figure 3 Identification of biomarkers through machine learning. (**A**) Least absolute shrinkage and selection operator (LASSO) regression analysis: gene selection performed through 10-fold cross-validation. (**B**) Variable selection using the Boruta algorithm. (**C**) Intersection of the two algorithms to identify candidate biomarkers.

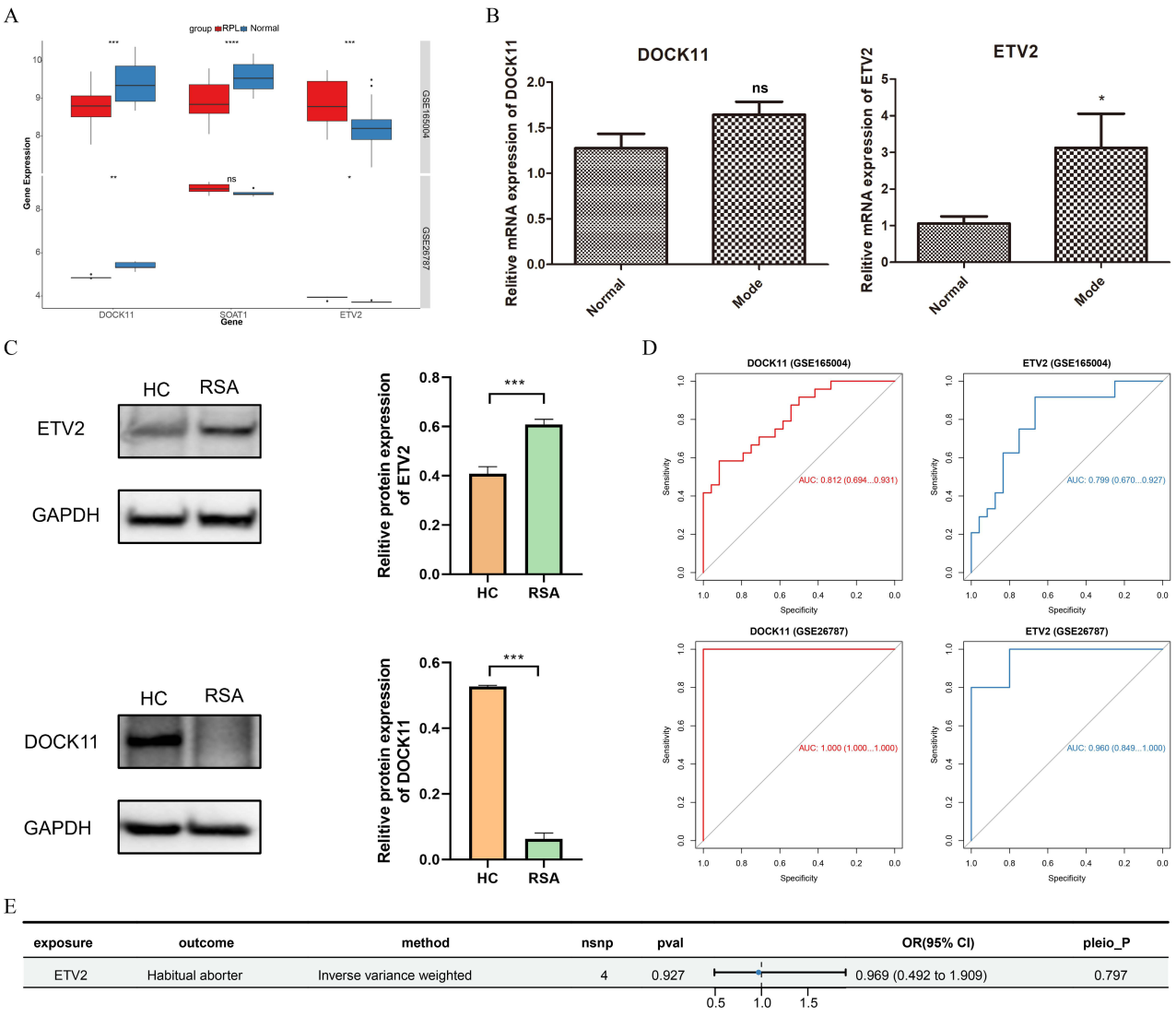

Figure 4 Identification of biomarkers. (**A**) Expression analysis of biomarkers between endometriosis and controls in training and validation sets. (**B** and **C**) Reverse transcription-quantitative polymerase chain reaction (RT-qPCR) and Western blotting to verify biomarkers expression. (**D**) Receiver operating characteristic (ROC) curve of the biomarkers. (**E**) Mendelian randomisation analysis revealed causal relationship between biomarkers and RPL. *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001, ns: P > 0.05.

expression associated with these biomarkers (Figure 4E). They may act as downstream effector molecules in the disease process, participating in the reprogramming of the immune microenvironment and the regulation of placental function, rather than being upstream factors driving the occurrence of diseases. This discovery further emphasizes the complexity and multi-factorial nature of RPL, suggesting that in future research, we should pay more attention to the dynamic changes of these genes during disease progression and their interactions with other risk factors.

Functional Characterization of Biomarkers

The GSEA further supported the influence of biomarkers on the progression of RPL. The enrichment analysis conducted on the basis of the KEGG gene set disclosed that biomarkers were consistently enriched within pathways related to chemokinetic signaling, cellular dynamics receptor interactions, and natural cell toxicity (Figure 5A). Moreover, the outcomes derived from the enrichment analysis using the Hallmark gene set demonstrated that both of these biomarkers were jointly enriched in pathways associated with inflammatory responses, interferon gamma response, and TNAF signaling via NFKB (Figure 5B).

https://doi.org/10.2147/JMDH.S541670

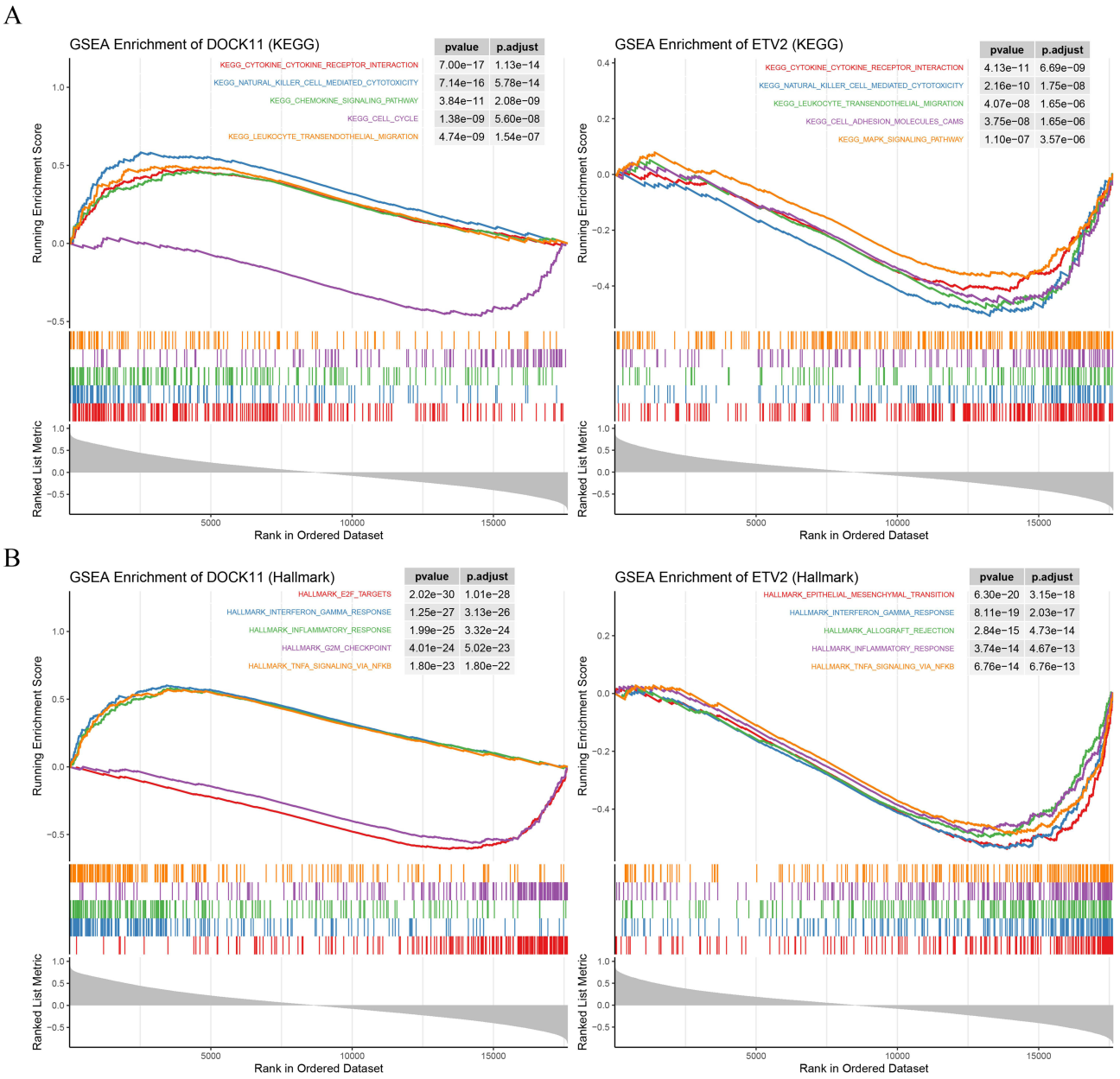

Figure 5 Functional characterization of biomarkers. (**A**) Gene Set Enrichment Analysis (GSEA) results for biomarkers base on KEGG gene set. (**B**) GSEA results for biomarkers base on Hallmark gene set.

Deciphering Differences in Immune Cell Infiltration Between RPL Patients and Healthy Individuals

We explored the differences in the immune microenvironment between patients with RPL and healthy individuals. We used the CIBERSORT algorithm and the LM22 gene set to calculate the proportions of 22 immune cells in the training set (Table S3). The stacked bar chart displayed the infiltration abundances of 22 types of immune cells, while the box plot showed the significant differences in four types of immune cells between RPL patients and the healthy control group ($P < 0.05$) (Figure 6A). Specifically, except that the infiltration of CD8 T cells was higher in RPL patients ($P < 0.05$), the infiltration levels of the other three types (memory resting CD4 T cells, M2 Macrophages and M1 Macrophages) of immune cells were lower in RPL patients ($P < 0.05$) (Figure 6B) (Table S4). These findings indicated that there were differences in the immune microenvironment between RPL patients and the healthy control group. Overall, there was

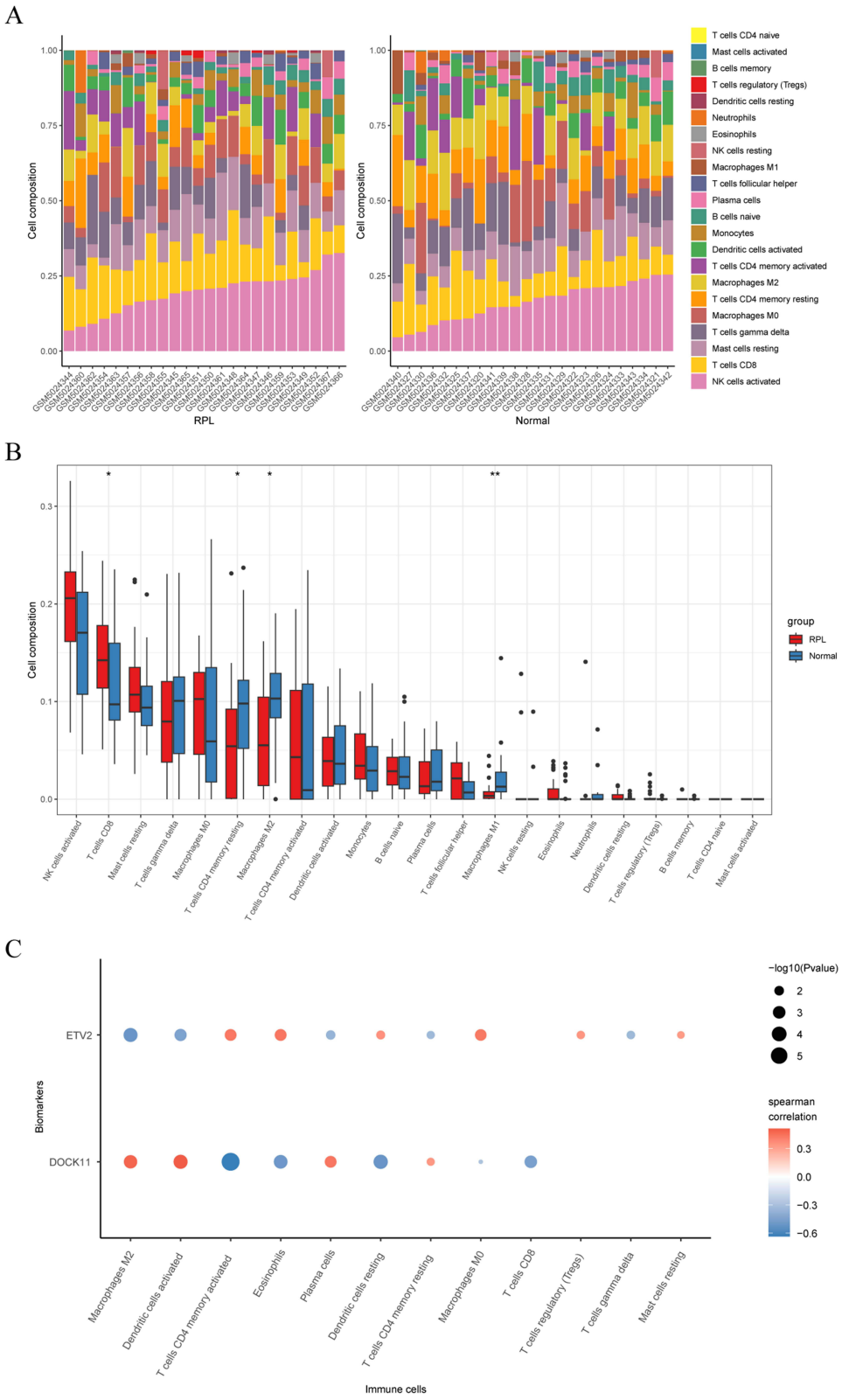

Figure 6 Immune correlation analysis for biomarkers and immune cell. (**A**) Stacked bar chart visualizing the infiltration abundance of 22 immune cell types. The horizontal axis represents the RPL samples and Normal samples in the training set, and the vertical axis represents the immune cell proportions. (**B**) Box plot illustrating the differences in infiltration abundance of these immune cells between RPL patients and controls. The horizontal axis represents cell types, and the vertical axis represents immune cell proportions. **P* < 0.05, ***P* < 0.01. (**C**) Heatmap of the correlation between immune cells and biomarkers. The horizontal axis represents immune cell types, and the vertical axis represents biomarkers. The size of the dots represents the significance of the correlation, while the color and depth of the dots indicate the direction and magnitude of the correlation.

https://doi.org/10.2147/JMDH.S541670

a significant correlation between immune cells and biomarkers. Among them, the DOCK11 had the highest correlation with activated memory CD4 T cells (cor= −0.635, P < 0.001) (Figure 6C) (Table S5).

Targeted Drug Research for Biomarkers

Using public databases, a drug-biomarker network was then assembled with 2 nodes and 17 edges, identifying 4 drugs predicted to target ETV2 and 7 targeting DOCK11 (Figure 7A) (Table S6). We continued the molecular docking analysis by selecting the highest scoring drug for each biomarker. The binding energies of ETV2 were −5.1 kcal/mol with tolazoline and −6 kcal/mol with Primaquine, suggesting stable interactions. Similarly, DOCK11 had a binding energy of −7.6 kcal/mol with 8-Bromo-cAMP and −8.7 kcal/mol with calcitriol (Figure 7B–E).

Single-Cell Analysis Revealed Distinct Cellular Clusters in RPL

After data processing, Figure S1A displays the results for nFeature RNA, nCount RNA, and percent_m. Figure S1B highlighted 2,000 highly variable genes. The top 30 principal components were selected using PCA (Figure S1C). The t-SNE dimensionality reduction revealed 16 distinct cellular clusters (Figure 8A), which were further classified into 11 cell types based on the expression of marker genes: Cytotoxic CD8+ T cells, Natural Killer cells, Macrophages,

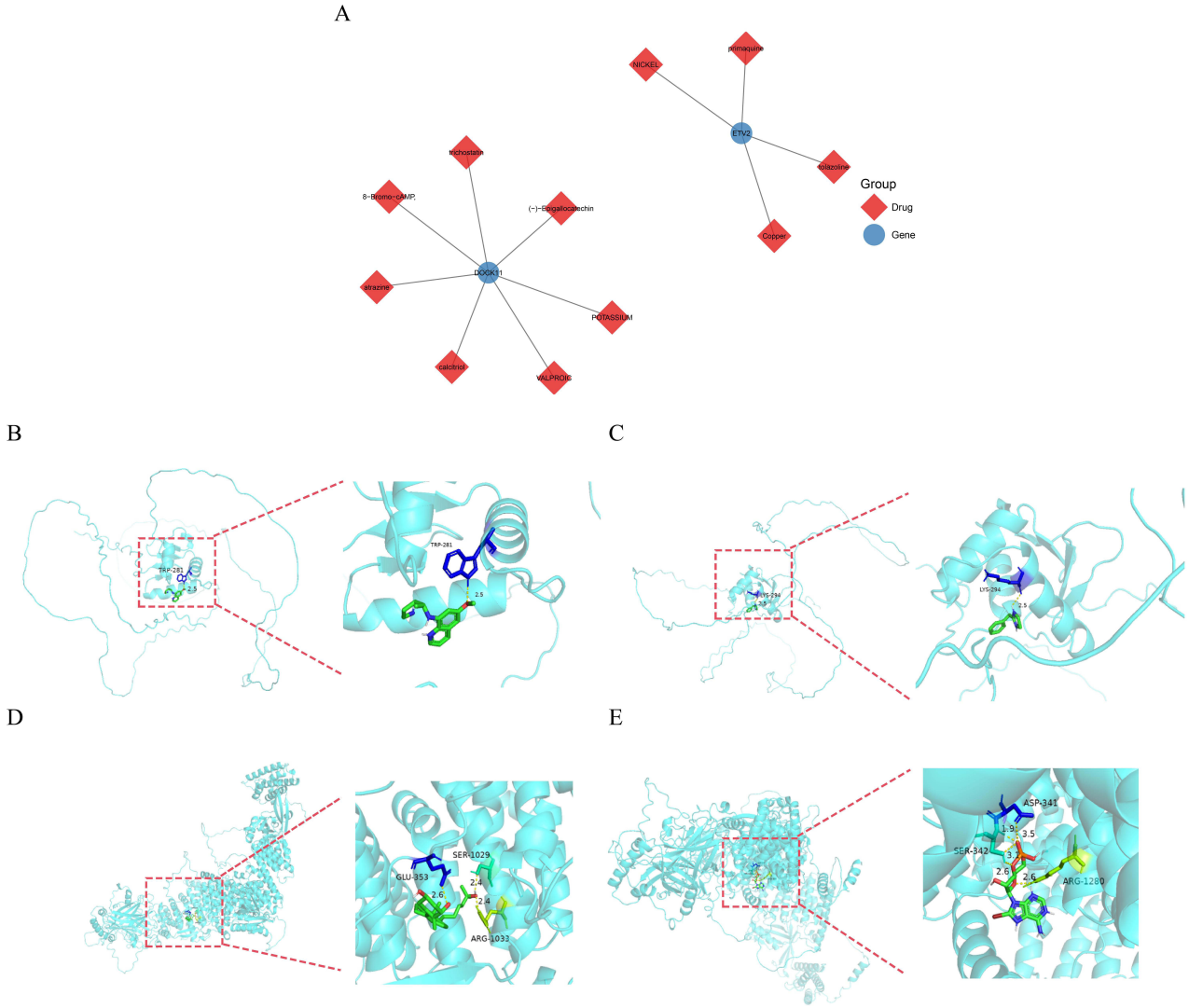

Figure 7 Potential regulatory mechanisms of biomarkers. (**A**) Construction of Drug-biomarkers regulatory network. The red rhombus represents drugs and the blue circle represents genes. (**B**–**E**) Molecular docking binding site prediction of ETV2 with Primaquine (**B**) and tolazoline (**C**), DOCK11 with calcitriol (**D**) and 8-Bromo-cAMP (**E**).

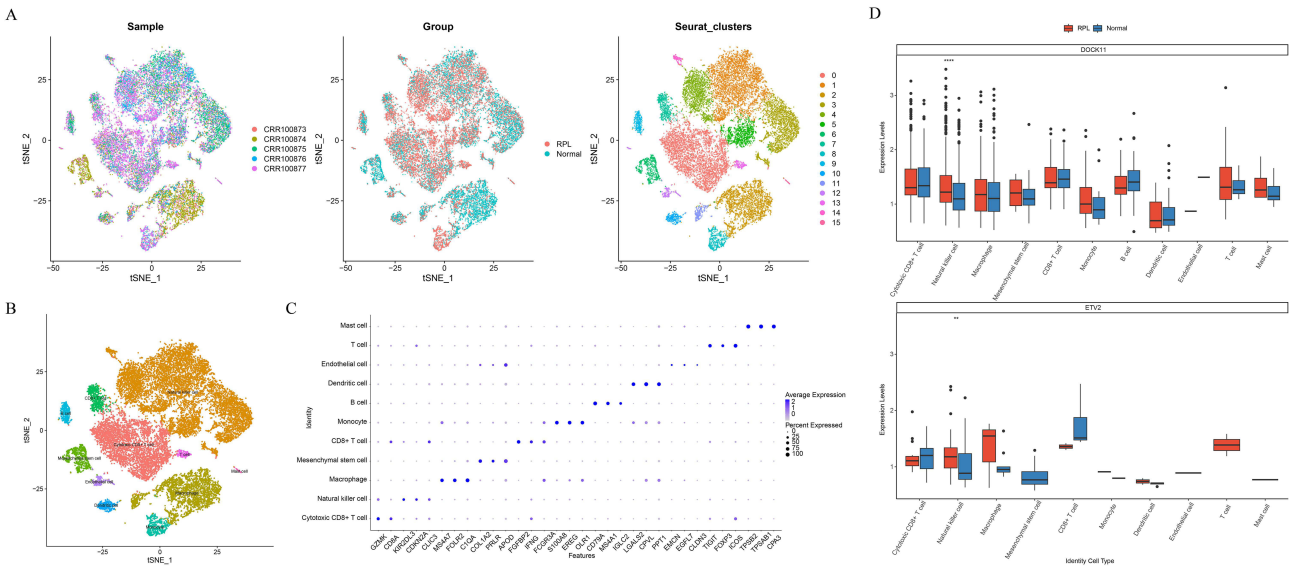

Figure 8 Single-cell RNA sequencing (scRNA-seq) analysis. (**A**) Clustering analysis identified 16 distinct cell clusters. Each point in the figure represents a single cell. (**B**) Annotation revealed 11 cell types: Cytotoxic CD8+ T cells, Natural Killer cells, Macrophages, Mesenchymal Stem Cells, CD8+ T cells, Monocytes, B cells, Dendritic Cells, Endothelial Cells, T cells, and Mast Cells. (**C**) Bubble plot illustrating the marker genes used for annotation. (**D**) Violin plots visualize the expression distribution of biomarkers across 11 cell types. $^{**}P < 0.01$, $^{****}P < 0.0001$.

Mesenchymal Stem Cells, CD8+ T cells, Monocytes, B cells, Dendritic Cells, Endothelial Cells, T cells, and Mast Cells (Figure 8B and C). We then analyzed the distribution of biomarkers across different cell types and found that both biomarkers were significantly enriched in Natural Killer cells (Figure 8D).

We determined the maximum number of interactions and interaction weights for each cell in order to more accurately infer the network's node sizes and edge weights across various groups. The results indicated that, compared to the normal group, the RPL group exhibited a decrease in the number of interactions but a significant increase in interaction strength (Figure 9A and B). In the RPL group, input signals from NK cells, cytotoxic CD8+ T cells, and T cells were enhanced, while input signals from macrophages were relatively diminished (Figure 9C). This shift might reflect a reprogramming of the immune microenvironment in disease states. Subsequently, we illustrated how various signaling pathways from different cellular inputs to NK cells changed their communication probabilities between diseased and normal tissues. Notably affected signaling pathways included HLA-E-CD94:NKG2E, HLA-E-CD94:NKG2C, and HLA-E-KLRC2 (Figure 9D). These alterations could provide new insights into understanding the immune evasion mechanisms associated with RPL.

Discussion

RPL is a complex clinical issue, and its pathogenesis involves multiple factors, including genetic factors, immune system disorders, hormonal imbalances, and infections, etc.[33] Vitamin D regulates immune function by modulating the production of certain Th1 and Th2 cytokines. This regulation alters the immune response, promoting maternal immune tolerance, which is essential for maintaining a normal pregnancy.[34–36] However, abnormal vitamin D levels can have detrimental effects on immune function and pregnancy outcomes. Low vitamin D levels have been found to be substantially linked to an increased risk of RPL,[37,38] further suggesting the potential critical role of vitamin D in the pathogenesis of RPL. In this study, differential expression analysis and WGCNA identified 27 genes related to vitamin D metabolism. Two key biomarkers, DOCK11 and ETV2, associated with vitamin D metabolism in RPL were selected using two machine learning algorithms (LASSO and Boruta). ROC curve analysis, qRT-PCR, and Western blotting were employed to validate the diagnostic efficacy of these markers. Additionally, single-cell data was used to explore the correlation and expression patterns of these biomarkers with cell subpopulations. Furthermore, GSEA analysis, immune infiltration assessment, drug prediction, and molecular docking analyses were conducted to further uncover the potential molecular mechanisms underlying the development of RPL.

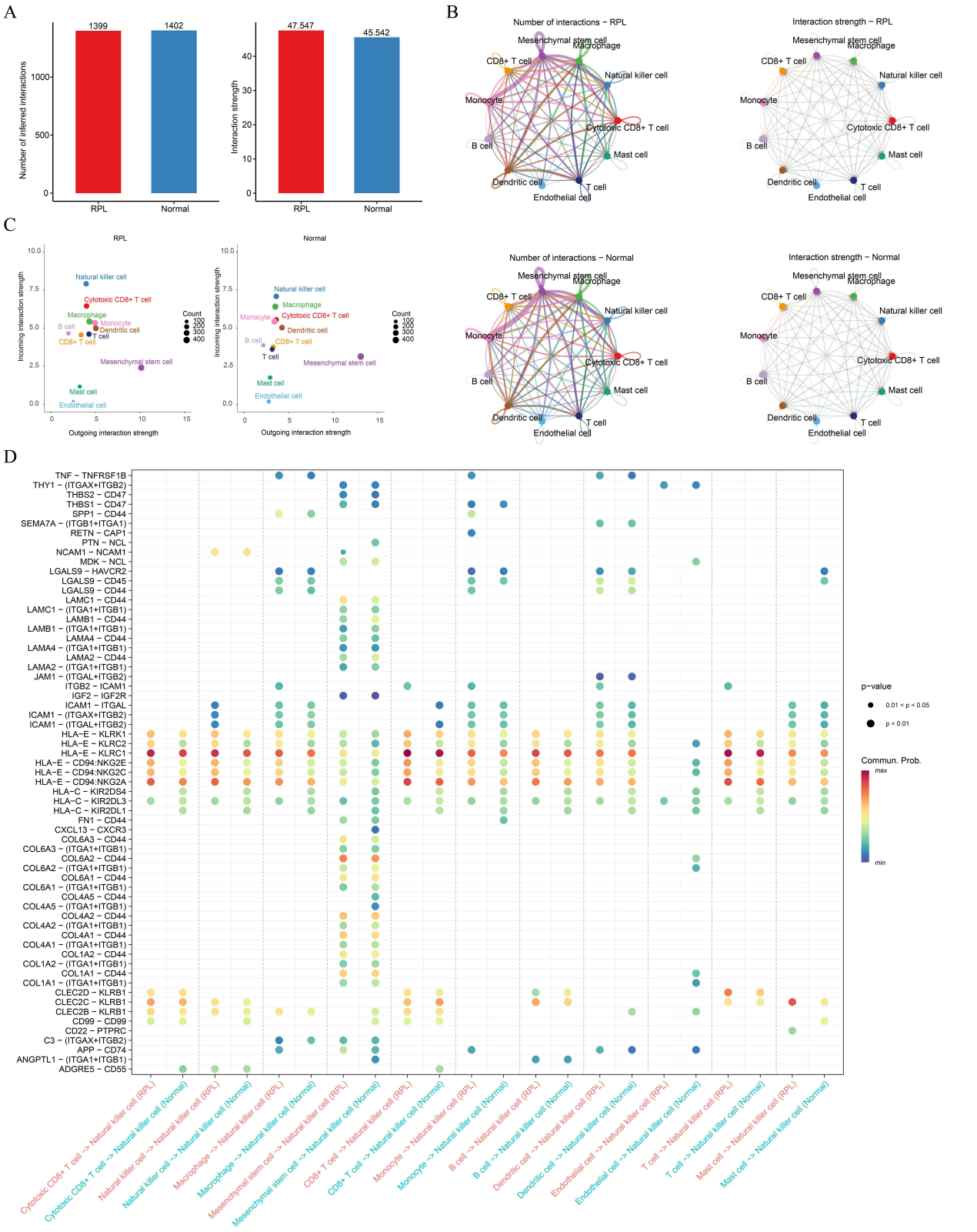

Figure 9 Cell communication analysis. (**A** and **B**) Annotated cellular communication network illustrating the number and strength of interactions in RPL and control. (**C**) Changes in the strength of cellular communication interactions. The horizontal axis represents the output intensity of different cells, while the vertical axis denotes the input intensity of different cells. (**D**) NK cell input signal interaction pathway communication probability. The horizontal axis represents different input cell types, and the vertical axis represents communication-related pathways.

DOCK1 belongs to the DOCK family of proteins and is one of the key components of the integrin signaling pathway. Studies have shown that DOCK1 can promote cell proliferation and inhibit apoptosis by activating downstream proteins such as ERK and AKT.[39] Previous research has reported that the DOCK1 inhibitor TBOPP induces miscarriage in mice by inactivating the DUSP4/ERK pathway, and the expression of DOCK1 in the placental villi of women with RPL is significantly reduced.[40] This finding aligns with our research results, where we further confirmed the low expression of DOCK1 through Western blotting. However, due to tissue specificity, the PCR results did not show significant differences. The function of DOCK1 has also been widely studied in other pregnancy-related diseases and malignant conditions. For example, in endometrial cancer, Xie et al found that DOCK1 regulates malignant biological behaviors through the c-Raf/ERK pathway.[41] Additionally, in preeclampsia, DOCK1 deficiency has been shown to lead to placental dysfunction by coordinating inflammatory responses and oxidative stress.[42] These studies further support the important role of DOCK1 in pregnancy-related diseases, suggesting that it may participate in the regulation of placental function and the maintenance of pregnancy outcomes through various signaling pathways. Research on DOCK1 has also been conducted in other types of diseases. For instance, in breast cancer, MiR-486-5p suppresses breast cancer cell epithelial-mesenchymal transition induced by IL-22 by inhibiting DOCK1.[43]

During early mouse embryo development, several key transcription factors (such as SCL, ETV2, Runx1, GATA1, and GATA2) are essential for the differentiation of mesodermal cells into the hematopoietic lineage. Among them, ETV2 is considered a critical transcription factor in hematopoietic differentiation. Studies have shown that ETV2 is expressed during a brief window in embryonic development, beginning at E7.0, with expression levels starting to decrease after E8.5.[44] At E7.75, ETV2 is involved in the development of hematopoietic endothelial cells in the yolk sac. If ETV2 is mutated during early embryo development, the embryo cannot survive due to the loss of hematopoietic stem cells (HSCs) and endothelial cells. Moreover, overexpression of Etv2 can induce the formation of Flk1+ mesodermal cells in mouse embryos and, under the suppression of BMP, Notch, and Wnt signaling, promotes the differentiation of Flk1+ mesodermal cells into HSCs and endothelial cells by alleviating the inhibitory effects.[45] In Zhao et al's study, the interaction between ETV2, VEGF, and FLK1 was shown to induce and regulate vascular regeneration. After ischemic injury to the hind limb of adult mice, Etv2 expression in endothelial cells was reactivated, promoting vascular regeneration through the regulation of the VEGF/FLK1 signaling pathway.[46] Xu et al found that after bone marrow transplantation or HSC injury, the Etv2 gene in hematopoietic stem/progenitor cells is activated by reactive oxygen species (ROS), which promotes autocrine, migration, and regeneration of these cells.[39] Interestingly, the expression of c-Kit can alleviate the proliferation defects of hematopoietic stem/progenitor cells and the short-term bone marrow transplant failure caused by ETV2 deletion in this study. In our study, we found that ETV2 is underexpressed in patients with RPL, which is consistent with its critical role in embryonic development and angiogenesis. The low expression of ETV2 may lead to impaired placental angiogenesis, thus affecting normal embryonic development. This discovery implies that ETV2 may be a significant factor in occurrence of RPL by regulating angiogenesis and placental function. Combined with existing research, the low expression of ETV2 may exacerbate placental dysfunction and abnormal embryonic development by affecting the VEGF/FLK1 signaling pathway or other downstream signaling molecules such as c-Kit. Our research provides new evidence for the role of ETV2 in RPL and lays the foundation for future exploration of its potential as a therapeutic target.

Through single-cell analysis, we discovered differential expression of two biomarkers in NK cells, further suggesting that NK cells may play a key role in the pathogenesis of RPL. The immune system of the female endometrium is crucial for the success of pregnancy. Early in pregnancy, about 40% of decidual cells are immune cells resident in the endometrium, with uterine natural killer (uNK) cells playing a significant role. Regulated by sex hormones, uNK cells (a subpopulation derived from peripheral blood NK cells) are recruited to the uterus.[47] By releasing cytokines and growth factors including interleukin-15 (IL-15) and VEGF, NK cells dramatically enhance their infiltration into the endometrium during the first trimester of pregnancy, controlling endometrial receptivity and angiogenesis.[48] Studies have shown that NK cells play a central role in the success or failure of pregnancy. In patients with RPL, a significant imbalance in NK cells is observed in peripheral blood and decidual tissue.[1,49] Braun et al[50] further confirmed that NK cells in the endometrium are closely related to RPL, highlighting how crucial particular immune cell subsets are to the endometrial microenvironment's ability to support pregnancy. Our findings further support the critical role of NK cells in RPL.

In our research, through drug prediction and molecular docking analyses, we identified some drugs that might target DOCK11 and ETV2. These drugs include tolazoline, Primaquine, 8-Bromo-cAMP and calcitriol. Tolazoline is a non-selective α-adrenergic antagonist and vasodilator. Tolazoline was used in a full-term newborn with neonatal encephalopathy and pulmonary hemorrhage. After intravenous infusion, the newborn survived and there was no evidence of long-term disability.[51] Primaquine is mainly used to treat malaria.[52] Primaquine is usually used in combination with other therapies, such as chloroquine or artemisinin drugs, which target the reproductive active forms of parasites. 8-Bromo-cAMP is a cell-permeable CAMP analogue that can act as a CAMP-dependent protein kinase activator.[53] Calcitriol is the active hormonal form of vitamin D and a key regulator of mineral and bone metabolism.[54] Calcitriol exerts a variety of biological functions, including controlling growth and cell differentiation, regulating hormone secretion and regulating reproductive function. Moreover, calcitriol can stimulate the synthesis of estradiol and progesterone in the human placenta.[55] Furthermore, as calcitriol is an immunosuppressant that can regulate the synthesis of various cytokines, including decidual prolactin, it may contribute to the establishment of fetal placental units.[55] Therefore, among these drugs, calcitriol has the potential for clinical translational research due to its clear clinical safety. While other compounds have shown good binding stability in molecular docking analysis, these drugs still need to be verified through preclinical studies and clinical trials, which provides a direction for the future development of new therapeutic strategies.

This study found that DOCK11 and ETV2 were abnormally expressed in the endometrial tissues of RPL patients and had good diagnostic value, suggesting that they might serve as potential biomarkers for clinical auxiliary diagnosis. In the future, by detecting the mRNA or protein expression levels of DOCK11 and ETV2 in endometrial tissue, a more accurate RPL risk stratification model can be established to identify high-risk populations and implement early intervention.

However, this study has some limitations. First, due to small sample size of RPL patients and the limited availability of public datasets, the validation dataset GSE26787 only has 5 samples of RPL patients, which may limit the universality and applicability of the research results. In addition, although the public dataset used in this study has minimized interference through a strict data standardization process, the inherent heterogeneity in sample collection and processing of the public dataset may still affect the stability of the research results. To increase the results' robustness and wide applicability, future studies should concentrate on increasing the sample size, adding more public datasets, and incorporating multi-center and multi-ethnic independent cohorts. Second, as this study mainly relies on the GWAS data of the European population for MR analysis, it is necessary to verify its universality in the GWAS data of the East Asian population in the future. Finally, additional in vitro and in vivo confirmation is required because the molecular indicators found in this study are novel and their precise roles in RPL have not yet been thoroughly investigated. Additionally, further exploration of the molecular mechanisms and validation of their potential as diagnostic or therapeutic targets are needed to advance basic and clinical research on RPL.

Conclusion

This study revealed 379 DEGs related to vitamin D metabolism, with 167 genes upregulated and 212 genes downregulated. Through WGCNA, key modules significantly associated with vitamin D metabolism were identified, and 27 candidate genes were selected. Machine learning methods further identified DOCK11 and ETV2 as potential biomarkers, with consistent expression trends in RPL samples, demonstrating good diagnostic value. Additionally, immune cell infiltration analysis showed changes in the immune microenvironment of RPL patients, with enhanced signals in NK cells and CD8 T cells, while macrophage signaling was weakened. Single-cell analysis indicated that these two biomarkers were significantly enriched in natural killer cells. These results offer fresh perspectives on the pathogenesis and immune evasion mechanisms of RPL.

Abbreviations

RPL, recurrent pregnancy loss; WGCNA, Weighted Gene Co-expression Network Analysis; DEGs, differentially expressed genes; MR, Mendelian randomization; VDR, vitamin D receptor; $1,25(OH)_2D_3$, 1,25-dihydroxyvitamin D3; GWAS, genome-wide association study; GEO, Gene Expression Omnibus; IVW, inverse variance weighted; GSEA, Gene Set Enrichment Analysis; QC, quality control.

Data Sharing Statement

The datasets generated during and/or analysed during the current study are available in the GEO database (https://www.ncbi.nlm.nih.gov/geo/), accession number: [GSE165004 and GSE26787]. The scRNA-Seq data was downloaded from the GSA database (https://ngdc.cncb.ac.cn/gsa/browse/CRA002181), accession number is CRA002181.

Ethics Approval and Consent to Participate

The guidelines outlined in the Declaration of Helsinki were followed. The study was approved by the Ethics Review Committee for Scientific Research, First Hospital of Shanxi Medical University (No:KYYJ-2025-038). All experiments were performed in accordance with relevant named guidelines and regulations and consent was obtained from all participants and/or their legal guardians.

Acknowledgments

We would like to sincerely thank the authors for their scientific contribution.

Author Contributions

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

Funding

There is no funding to report.

Disclosure

The authors report no conflicts of interest in this work.

References

1. Guo C, Cai P, Jin L, et al. Single-cell profiling of the human decidual immune microenvironment in patients with recurrent pregnancy loss. *Cell Discovery*. 2021;7(1):1. doi:10.1038/s41421-020-00236-z
2. Wang XH, Xu S, Zhou XY, et al. Low chorionic villous succinate accumulation associates with recurrent spontaneous abortion risk. *Nat Commun*. 2021;12(1):3428. doi:10.1038/s41467-021-23827-0
3. Bortoletto P, Lucas ES, Melo P, et al. Miscarriage syndrome: linking early pregnancy loss to obstetric and age-related disorders. *EBioMedicine*. 2022;81:104134. doi:10.1016/j.ebiom.2022.104134
4. Li J, Wang L, Ding J, et al. Multiomics studies investigating recurrent pregnancy loss: an effective tool for mechanism exploration. *Front Immunol*. 2022;13. 826198. doi:10.3389/fimmu.2022.826198
5. Ma Y. Identification and characterization of noncalcemic, tissue-selective, nonsecosteroidal vitamin D receptor modulators. *J Clin Invest*. 2006;116 (4):892–904. doi:10.1172/jci25901
6. Cheng JB, Levine MA, Bell NH, Mangelsdorf DJ, Russell DW. Genetic evidence that the human CYP2R1 enzyme is a key vitamin D 25-hydroxylase. *Proc Natl Acad Sci*. 2004;101(20):7711–7715. doi:10.1073/pnas.0402490101
7. Barry EL, Rees JR, Peacock JL, et al. Genetic variants in CYP2R1, CYP24A1, and VDR modify the efficacy of vitamin D3 supplementation for increasing serum 25-hydroxyvitamin D levels in a randomized controlled trial. *J Clin Endocrinol Metab*. 2014;99(10):E2133–E2137. doi:10.1210/jc.2014-1389
8. Wagner MM, Jukema JW, Hermes W, et al. Assessment of novel cardiovascular biomarkers in women with a history of recurrent miscarriage. *Pregnancy Hypertens*. 2018;11:129–135. doi:10.1016/j.preghy.2017.10.012
9. Zhou JC, Zhu Y, Gong C, et al. The GC2 haplotype of the vitamin D binding protein is a risk factor for a low plasma 25-hydroxyvitamin D concentration in a Han Chinese population. *Nutr Metab*. 2019;16(1):5. doi:10.1186/s12986-019-0332-0
10. Lafi ZM, Irshaid YM, El-Khateeb M, Ajlouni KM, Hyassat D. Association of rs7041 and rs4588 polymorphisms of the vitamin D binding protein and the rs10741657 polymorphism of CYP2R1 with vitamin D status among jordanian patients. *Genet Test Mol Bioma*. 2015;19(11):629–636. doi:10.1089/gtmb.2015.0058
11. Aguilar-Cordero MJ, Lasserrot-Cuadrado A, Mur-Villar N, León-Ríos XA, Rivero-Blanco T, Pérez-Castillo IM. Vitamin D, preeclampsia and prematurity: a systematic review and meta-analysis of observational and interventional studies. *Midwifery*. 2020;87:102707. doi:10.1016/j.midw.2020.102707
12. Papagni R, Pellegrino C, Di Gennaro F, et al. Impact of vitamin D in prophylaxis and treatment in tuberculosis patients. *Int J Mol Sci*. 2022;23 (7):3860. doi:10.3390/ijms23073860

13. Tamblyn JA, Pilarski NSP, Markland AD, et al. Vitamin D and miscarriage: a systematic review and meta-analysis. *Fertil Steril*. 2022;118 (1):111–122. doi:10.1016/j.fertnstert.2022.04.017
14. Al Balawi AN, Alblwi NAN, Soliman R, et al. Impact of vitamin D deficiency on immunological and metabolic responses in women with recurrent pregnancy loss: focus on VDBP/HLA-G1/CTLA-4/ENTPD1/adenosine-fetal-maternal conflict crosstalk. *BMC Pregnancy Childbirth*. 2024;24 (1):709. doi:10.1186/s12884-024-06914-0
15. Farazmand T, Rahbarian R, Jalali M, Ghahremani A, Razi A, Namdar Ahmadabad H. Vitamin D levels in non-pregnant women with a history of recurrent pregnancy loss with and without autoantibodies. *Casp J Intern Med*. 2024;15(2):266–272. doi:10.22088/cjim.15.2.266
16. Stuart T, Satija R. Integrative single-cell analysis. *Nat Rev Genet*. 2019;20(5):257–272. doi:10.1038/s41576-019-0093-7
17. Wu AR, Wang J, Streets AM, Huang Y. Single-cell transcriptional analysis. *Annu Rev Anal Chem*. 2017;10(1):439–462. doi:10.1146/annurev-anchem-061516-045228
18. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47. doi:10.1093/nar/gkv007
19. Gustavsson EK, Zhang D, Reynolds RH, Garcia-Ruiz S, Ryten M. ggtranscript: an R package for the visualization and interpretation of transcript isoforms using ggplot2. *Bioinformatics*. 2022;38(15):3844–3846. doi:10.1093/bioinformatics/btac409
20. Gu Z, Hübschmann D. Make Interactive Complex Heatmaps in R. *Bioinformatics*. 2022;38(5):1460–1462. doi:10.1093/bioinformatics/btab806
21. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-Seq data. *BMC Bioinformatics*. 2013;14(1):7. doi:10.1186/1471-2105-14-7
22. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics*. 2012;16(5):284–287. doi:10.1089/omi.2011.0118
23. Li Y, Lu F, Yin Y. Applying logistic LASSO regression for the diagnosis of atypical Crohn's disease. *Sci Rep*. 2022;12(1):11340. doi:10.1038/s41598-022-15609-5
24. Yue S, Li S, Huang X, et al. Machine learning for the prediction of acute kidney injury in patients with sepsis. *J Transl Med*. 2022;20(1):215. doi:10.1186/s12967-022-03364-0
25. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*. 2011;12(1):77. doi:10.1186/1471-2105-12-77
26. Skrivankova VW, Richmond RC, Woolf BAR, et al. Strengthening the reporting of observational studies in epidemiology using mendelian randomization: the STROBE-MR statement. *JAMA*. 2021;326(16):1614. doi:10.1001/jama.2021.18236
27. Hemani G, Zheng J, Elsworth B, et al. The MR-base platform supports systematic causal inference across the human phenome. *ELife*. 2018;7: e34408. doi:10.7554/eLife.34408
28. Newman AM, Liu CL, Green MR, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods*. 2015;12(5):453–457. doi:10.1038/nmeth.3337
29. Seeliger D, De Groot BL. Ligand docking and binding site analysis with PyMOL and AutoDock/Vina. *J Comput-Aided Mol Des*. 2010;24 (5):417–422. doi:10.1007/s10822-010-9352-6
30. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem*. 2010;31(2):455–461. doi:10.1002/jcc.21334
31. Satija R, Farrell JA, Gennert D, Schier AF, Regev A. Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol*. 2015;33 (5):495–502. doi:10.1038/nbt.3192
32. Jin S, Guerrero-Juarez CF, Zhang L, et al. Inference and analysis of cell-cell communication using CellChat. *Nat Commun*. 2021;12(1):1088. doi:10.1038/s41467-021-21246-9
33. Di Gennaro F, Guido G, Frallonardo L, et al. Chronic endometritis and antimicrobial resistance: towards a multidrug-resistant endometritis? An expert opinion. *Microorganisms*. 2025;13(1):197. doi:10.3390/microorganisms13010197
34. Warner JO, Warner JA. The foetal origins of allergy and potential nutritional interventions to prevent disease. *Nutrients*. 2022;14(8):1590. doi:10.3390/nu14081590
35. Ota K, Dambaeva S, Kim MW, et al. 1,25-Dihydroxy-vitamin D3 regulates NK-cell cytotoxicity, cytokine secretion, and degranulation in women with recurrent pregnancy losses. *Eur J Immunol*. 2015;45(11):3188–3199. doi:10.1002/eji.201545541
36. Ota K, Dambaeva S, Han AR, Beaman K, Gilman-Sachs A, Kwak-Kim J. Vitamin D deficiency may be a risk factor for recurrent pregnancy losses by increasing cellular immunity and autoimmunity. *Hum Reprod*. 2014;29(2):208–219. doi:10.1093/humrep/det424
37. Kwak-Kim J, Skariah A, Wu L, Salazar D, Sung N, Ota K. Humoral and cellular autoimmunity in women with recurrent pregnancy losses and repeated implantation failures: a possible role of vitamin D. *Autoimmun Rev*. 2016;15(10):943–947. doi:10.1016/j.autrev.2016.07.015
38. Li N, Wu HM, Hang F, Zhang YS, Li MJ. Women with recurrent spontaneous abortion have decreased 25(OH) vitamin D and VDR at the fetal-maternal interface. *Braz J Med Biol Res*. 2017;50(11):e6527. doi:10.1590/1414-431x20176527
39. Xu CX, Lee TJ, Sakurai N, et al. ETV2/ER71 regulates hematopoietic regeneration by promoting hematopoietic stem cell proliferation. *J Exp Med*. 2017;214(6):1643–1653. doi:10.1084/jem.20160923
40. Xu Y, Liu X, Zeng W, et al. DOCK1 insufficiency disrupts trophoblast function and pregnancy outcomes via DUSP4-ERK pathway. *Life Sci Alliance*. 2024;7(2):e202302247. doi:10.26508/lsa.202302247
41. Xie S, Jin Y, Wang J, Li J, Peng M, Zhu X. DOCK1 regulates the malignant biological behavior of endometrial cancer through c-Raf/ERK pathway. *BMC Cancer*. 2024;24(1):296. doi:10.1186/s12885-024-12030-1
42. Xu Y, Qin X, Zeng W, et al. DOCK1 deficiency drives placental trophoblast cell dysfunction by influencing inflammation and oxidative stress, hallmarks of preeclampsia. *Hypertens Res*. 2024;47(12):3434–3446. doi:10.1038/s41440-024-01920-3
43. Li H, Mou Q, Li P, et al. MiR-486-5p inhibits IL-22-induced epithelial-mesenchymal transition of breast cancer cell by repressing Dock1. *J Cancer*. 2019;10(19):4695–4706. doi:10.7150/jca.30596
44. Das S, Gupta V, Bjorge J, et al. ETV2 and VEZF1 interaction and regulation of the hematoendothelial lineage during embryogenesis. *Front Cell Dev Biol*. 2023;11:1109648. doi:10.3389/fcell.2023.1109648
45. Kim JY, Lee DH, Kim JK, et al. ETV2/ER71 regulates the generation of FLK1(+) cells from mouse embryonic stem cells through miR-126-MAPK signaling. *Stem Cell Res Ther*. 2019;10(1):328. doi:10.1186/s13287-019-1466-8

46. Zhao H, Choi K. A CRISPR screen identifies genes controlling Etv2 threshold expression in murine hemangiogenic fate commitment. *Nat Commun*. 2017;8(1):541. doi:10.1038/s41467-017-00667-5
47. Dosiou C, Giudice LC. Natural killer cells in pregnancy and recurrent pregnancy loss: endocrine and immunologic perspectives. *Endocr Rev*. 2005;26(1):44–62. doi:10.1210/er.2003-0021
48. Strunz B, Bister J, Jönsson H, et al. Continuous human uterine NK cell differentiation in response to endometrial regeneration and pregnancy. *Sci Immunol*. 2021;6(56):eabb7800. doi:10.1126/sciimmunol.abb7800
49. Fukui A, Kamoi M, Funamizu A, et al. NK cell abnormality and its treatment in women with reproductive failures such as recurrent pregnancy loss, implantation failures, preeclampsia, and pelvic endometriosis. *Reprod Med Biol*. 2015;14(4):151–157. doi:10.1007/s12522-015-0207-7
50. Braun AS, Vomstein K, Reiser E, et al. NK and T cell subtypes in the endometrium of patients with recurrent pregnancy loss and recurrent implantation failure: implications for pregnancy success. *J Clin Med*. 2023;12(17):5585. doi:10.3390/jcm12175585
51. Barnes ME, Feeney E, Duncan A, et al. Pulmonary haemorrhage in neonates: systematic review of management. *Acta Paediatr*. 2022;111 (2):236–244. doi:10.1111/apa.16127
52. Marcsisin SR, Reichard G, Pybus BS. Primaquine pharmacology in the context of CYP 2D6 pharmacogenomics: current state of the art. *Pharmacol Ther*. 2016;161:1–10. doi:10.1016/j.pharmthera.2016.03.011
53. Bockus LB, Humphries KM. cAMP-dependent protein kinase (PKA) signaling is impaired in the diabetic heart. *J Biol Chem*. 2015;290 (49):29250–29258. doi:10.1074/jbc.M115.681767
54. Maekawa AS, Bennin D, Hartery SA, et al. Maternal loss of 24-hydroxylase causes increased intestinal calcium absorption and hypercalcemia during pregnancy but reduced skeletal resorption during lactation in mice. *J Bone Miner Res*. 2024;39(12):1793–1808. doi:10.1093/jbmr/zjae166
55. Barrera D, Avila E, Hernández G, et al. Estradiol and progesterone synthesis in human placenta is stimulated by calcitriol. *J Steroid Biochem Mol Biol*. 2007;103(3–5):529–532. doi:10.1016/j.jsbmb.2006.12.097