

# Identification of Energy Metabolism-Related Subtypes and Diagnostic Biomarkers for Osteoarthritis by Integrating Bioinformatics and Machine Learning

Sheng Xu, Jie Ye, Xiaochong Cai

Department of Orthopaedics, Jinhua Wenrong Hospital, Jinhua City, Zhejiang, 321000, People's Republic of China

Correspondence: Sheng Xu, Department of Orthopaedics, Jinhua Wenrong Hospital, No. 768 Donglai Road, Sanjiang Street, Wucheng District, Jinhua City, Zhejiang, 321100, People's Republic of China, Tel +86 15957919375, Email shengx202408@163.com

**Background:** Osteoarthritis (OA) is a chronic and complex degenerative joint disease that increasingly burdens and affects the elderly population. Abnormal energy metabolism is closely associated with the pathological mechanisms of OA. This study aims to identify key genes related to energy metabolism that are closely linked to the treatment and diagnosis of OA.

**Methods:** The transcriptomic data for OA were collected from the Gene Expression Omnibus (GEO), with GSE51588 and GSE63359 serving as the training and validation datasets, respectively. Differential expression analysis was conducted to identify key energy metabolism-related genes. Unsupervised clustering was performed to classify molecular subtypes. Three machine learning algorithms were employed to identify key diagnosis genes, specifically Least Absolute Shrinkage and Selection Operator (LASSO), Support Vector Machine Recursive Feature Elimination (SVM-RFE), and Random Forest (RF). We construct a comprehensive nomogram, and the diagnostic performance of both the nomogram and feature genes was evaluated using operating characteristic curve (ROC) and calibration curves. We assessed the immune infiltration levels in OA samples using the IOBR platform and the CIBERSORT algorithm.

**Results:** We classified OA patients into two molecular subtypes, which exhibited distinct differences in immune infiltration levels. Subsequently, we successfully identified two characteristic genes, NUP98 and RPIA, and demonstrated statistically significant differences ( $P < 0.05$ ) and diagnostic performance in the validation cohort. Evaluation using ROC and calibration curves demonstrated that these characteristic genes exhibit robust diagnostic performance. Multiple immune cells may be involved in the development of OA, and all characteristic genes may be associated with immune cells to varying degrees.

**Conclusion:** In conclusion, NUP98 and RPIA have the potential to serve as diagnostic biomarkers for OA and may offer opportunities for therapeutic intervention in OA.

**Keywords:** osteoarthritis, energy metabolism, diagnosis, immunology

## Introduction

Osteoarthritis (OA) is a chronic degenerative joint disease and one of the leading causes of disability, affecting approximately 500 million people worldwide.<sup>1</sup> The clinical pathological features of OA primarily include articular cartilage erosion, synovial hyperplasia, synovial inflammation, abnormal angiogenesis, subchondral bone disruption, ligament and tendon instability, and joint stiffness.<sup>2</sup> Flaviu et al<sup>3</sup> found that the postoperative neutrophil/lymphocyte ratio (NLR) and the aggregate inflammatory systemic index (AISI) are reliable inflammatory biomarkers involved in orthopedic pathological inflammatory processes. Additionally, some studies have shown that tissue inhibitor of metalloproteinases 3 (TIMP-3) can effectively inhibit cartilage degradation in osteoarthritis and may aid in predicting early disease progression and treatment response.<sup>4</sup> Overall, identifying effective biomarker targets is crucial for early diagnosis and the development of molecular-level therapeutic strategies for OA patients.

Metabolism is the process by which cells obtain energy from nutritional substrates and sustain vital cellular functions.<sup>5</sup> Metabolic dysregulation can lead to various diseases, including cardiovascular diseases and OA.<sup>6,7</sup> Clinical evidence shows that OA frequently coexists with metabolic-related diseases, such as diabetes and cardiovascular conditions, which accelerate the rapid deterioration of OA-affected joints.<sup>8</sup> Moreover, chondrocyte metabolic dysfunction has been demonstrated to play a pivotal role in cartilage degeneration and OA progression.<sup>9</sup> Wu et al<sup>10</sup> reported that chondrocyte energy metabolism disorders result in significantly reduced mitochondrial respiration, decreased ATP production, impaired mitochondrial membrane potential, and morphological damage, ultimately compromising the function of OA chondrocytes. Another study found that energy metabolism disruption in OA chondrocytes leads to the accumulation of reactive oxygen species (ROS), contributing to chondrocyte degradation and apoptosis, thus promoting OA development.<sup>11</sup> In summary, metabolism plays a critical role in the onset and progression of OA, particularly energy metabolism, which is closely linked to mitochondrial dysfunction and is key in the degeneration of chondrocytes leading to OA.

Based on this context, the present study utilizes bioinformatics and machine learning techniques to conduct a thorough analysis of the role of energy metabolism-related genes (EMRGs) in the development of OA. We developed a comprehensive nomogram and evaluated the diagnostic efficacy of the feature genes associated with OA. Furthermore, our research delves into how feature genes interact with immune cell infiltration. This study seeks to identify biomarkers that play a crucial role in OA pathogenesis and establish a gene signature based on molecular diagnostics for OA.

## Materials and Methods

### Accessing Analysing Data

This is a data-driven study. In this research, gene expression profile data for subchondral bone were initially acquired from the Gene Expression Omnibus (GEO, <https://www.ncbi.nlm.nih.gov/geo/>) database, categorized as the microarray dataset (GSE51588, comprising 26 samples from the normal group and 40 from the abnormal group). Subsequently, genes linked to OA were sourced from the GeneCards database (<https://www.ncbi.nlm.nih.gov/geo/>). By comparing the differentially expressed genes identified in the microarray data with the OA-related genes, a selection of the top 20 genes was made based on their relevance scores. For validation purposes, peripheral blood expression profile data from OA patients were utilized (GSE63359, normal group: 26, abnormal group: 46). Moreover, the “msigdb” package was harnessed to gather gene data pertinent to energy metabolism.<sup>12</sup>

### Identification and Preliminary Analysis of EMRGs Associated with OA

The analysis of differential expression for the GSE51588 dataset was carried out using the “limma” package. We established our selection parameters at  $|\log \text{ fold change (FC)}| > 0.585$  and  $\text{adjust.p} < 0.05$ , which allowed us to pinpoint the differentially expressed genes (DEGs, [Supplementary Table 1](#)). To illustrate these findings, we created both volcano plots and heat maps. By comparing these DEGs with EMRGs, we discovered a subset of differentially expressed energy metabolism-related genes (DEEMRGs). We subsequently carried out enrichment analyses for Gene Ontology (GO) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) on this group of DEEMRGs. Furthermore, we constructed a protein-protein interaction (PPI) network for the DEEMRGs using the STRING database (<https://string-db.org/>), applying a confidence score threshold greater than 0.4, and subsequently visualized the network with Cytoscape v3.10.2 software (The Cytoscape Consortium, USA).

### Consensus Cluster Analysis of DEEMRGs

Consensus clustering analysis serves as a robust technique for identifying the ideal number of clusters, thereby enhancing the reliability and precision of the findings.<sup>13</sup> In our study, we employed the “ConsensusClusterPlus” package to examine the expression profiles of DEEMRGs, allowing us to pinpoint specific OA subtypes. The assessment utilized Euclidean distance metrics, incorporating 1000 resampling iterations to bolster the analysis’s resilience. Principal Component Analysis (PCA) was conducted using the “FactoMineR” package to reduce data dimensionality and evaluate the clustering distribution of samples in high-dimensional space.

## Analysis and Identification of Differentially Expressed Genes in Molecular Subtypes

This study carried out a differential expression analysis between the two subtypes employing the “limma” package, establishing criteria of  $|\log FC| > 0.585$  and  $\text{adjust.p} < 0.05$  ([Supplementary Table 2](#)). To visualize the genes that were differentially expressed, we created volcano and heat maps. Furthermore, we performed KEGG enrichment analysis utilizing the “clusterProfiler” package. The top 15 pathways, based on the number of differentially expressed genes, were subsequently depicted in a bubble plot for enhanced visualization.

## Single-Sample Gene Set Enrichment Analysis of Molecular Subtypes

We kicked off our analysis by employing the single-sample Gene Set Enrichment Analysis (ssGSEA) method on the samples, utilizing the “GSVA” package for the task. For delving into the gene expression levels, we turned to the “ggpubr” package to parse the data and then depicted our findings through box plots. To ascertain the statistical significance of the disparities in expression levels, we resorted to the Wilcoxon test.

## Screening of Diagnostic Biomarker Signature Genes for OA

The current research utilized three different machine learning techniques to pinpoint gene features that might act as diagnostic biomarkers for OA. In our quest to identify the most promising OA biomarkers, We employed Least Absolute Shrinkage and Selection Operator (LASSO) regression using the “glmnet” package to refine the dataset. ([Supplementary Table 3](#)). Additionally, we executed Support Vector Machine Recursive Feature Elimination (SVM-RFE) analysis via the “caret” package ([Supplementary Table 4](#)). To further our analysis, we also conducted a Random Forest (RF) evaluation utilizing the “randomForest” package, selecting the top 10 genes based on their MeanDecreaseGini importance rankings ([Supplementary Table 5](#)).

## Receiver Operating Characteristic Curve and Expression Level Validation

To evaluate the diagnostic potential of the feature genes, we utilized the “pROC” package to create ROC curves for these genes within both the training dataset (GSE51588) and the validation dataset (GSE63359). Additionally, we applied the Wilcoxon test to assess the statistical significance of variations in gene expression levels.

## Gene Set Enrichment Analysis (GSEA) of OA Characteristic Genes

We performed pathway enrichment analysis on the characteristic genes utilizing the “clusterProfiler” package along with the reference dataset (c2.cp.kegg.v7.5.symbols.gmt). We set our criteria at  $\text{adjust.p} < 0.05$  and  $|\text{Normalized Enrichment Score (NES)}| > 1$ . For every characteristic gene, we showcased the top five pathways that yielded the most significant P-values.

## Establishment of Competitive Endogenous RNA (ceRNA) Networks

The present study employed the “multiMiR” package to pinpoint miRNAs that interact with specific genes, sifting through results from both TargetScan and miRDB. Next, we determined the overlap of predictions yielded by these two sources. Interaction data concerning lncRNAs and miRNAs was retrieved from the ENCORI database (<https://rnasysu.com/encori/>), selecting lncRNAs with a clipExpNum greater than 10. Finally, the findings were illustrated using the “ggsankey” package.

## Construction of an OA Diagnostic Model

Employing the “rms” package, we developed a nomogram and calibration curve for the feature genes. To visualize the performance of these nomograms and characteristic genes, ROC curves were created utilizing the “pROC” package. Furthermore, we produced decision curve analysis (DCA) curves for both the characteristic genes and the nomograms through the “rmda” package.

## Analysis of Immune Infiltration

To evaluate the infiltration levels of immune cells in OA, we employed the “IOBR” package to create stacked bar charts illustrating the distribution of these immune cells. The CIBERSORT algorithm was applied to calculate immune infiltration levels, which were visualized using box plots. Furthermore, the “ggcorrplot” package was employed to visualize the correlations among differentially expressed immune cells. This approach also generated a heatmap that highlights the connections between characteristic genes and these immune cells.

## Statistical Analysis

All analyses were conducted using R (version 4.1.3, R Foundation for Statistical Computing, Austria), and related packages. The Wilcoxon test was used to determine statistical differences between groups. ROC curves were plotted using the “pROC” package. Data visualization was primarily performed using the “ggplot2” package. P-values were summarized as follows: \*\*\*\*  $P < 0.0001$ ; \*\*\*  $0.0001 < P < 0.001$ ; \*\*  $0.001 < P < 0.01$ ; \*  $0.01 < P < 0.05$ ; ns  $P > 0.05$ .

## Results

### Identification and Preliminary Analysis of DEEMRGs

To investigate the role of energy metabolism in OA, this study first identified 1561 DEGs from the GSE51588 dataset by comparing normal and OA tissue samples. As shown in Figure 1A, 815 genes were downregulated, while 746 were upregulated. Figure 1B presents a heatmap of the top 20 DEGs with the highest relevance scores. By intersecting the DEGs with EMRGs, 49 DEEMRGs were obtained (Figure 1C). A PPI network analysis of these DEEMRGs revealed 41 nodes and 157 edges, as illustrated in Figure 1D. The darker color and larger size of the nodes indicate a higher number of interactions with other proteins, reflecting the tight connections among the DEEMRGs. GO enrichment analysis revealed that these DEEMRGs were significantly involved in biological processes such as the monosaccharide metabolic process, carbohydrate biosynthetic process, cellular carbohydrate metabolic process, hexose metabolic process, and ribose phosphate metabolic process (Figure 1E). KEGG pathway analysis further demonstrated that these genes were primarily enriched in glycolysis/gluconeogenesis, carbon metabolism, the HIF-1 signaling pathway, the glucagon signaling pathway, and central carbon metabolism in cancer (Figure 1F).

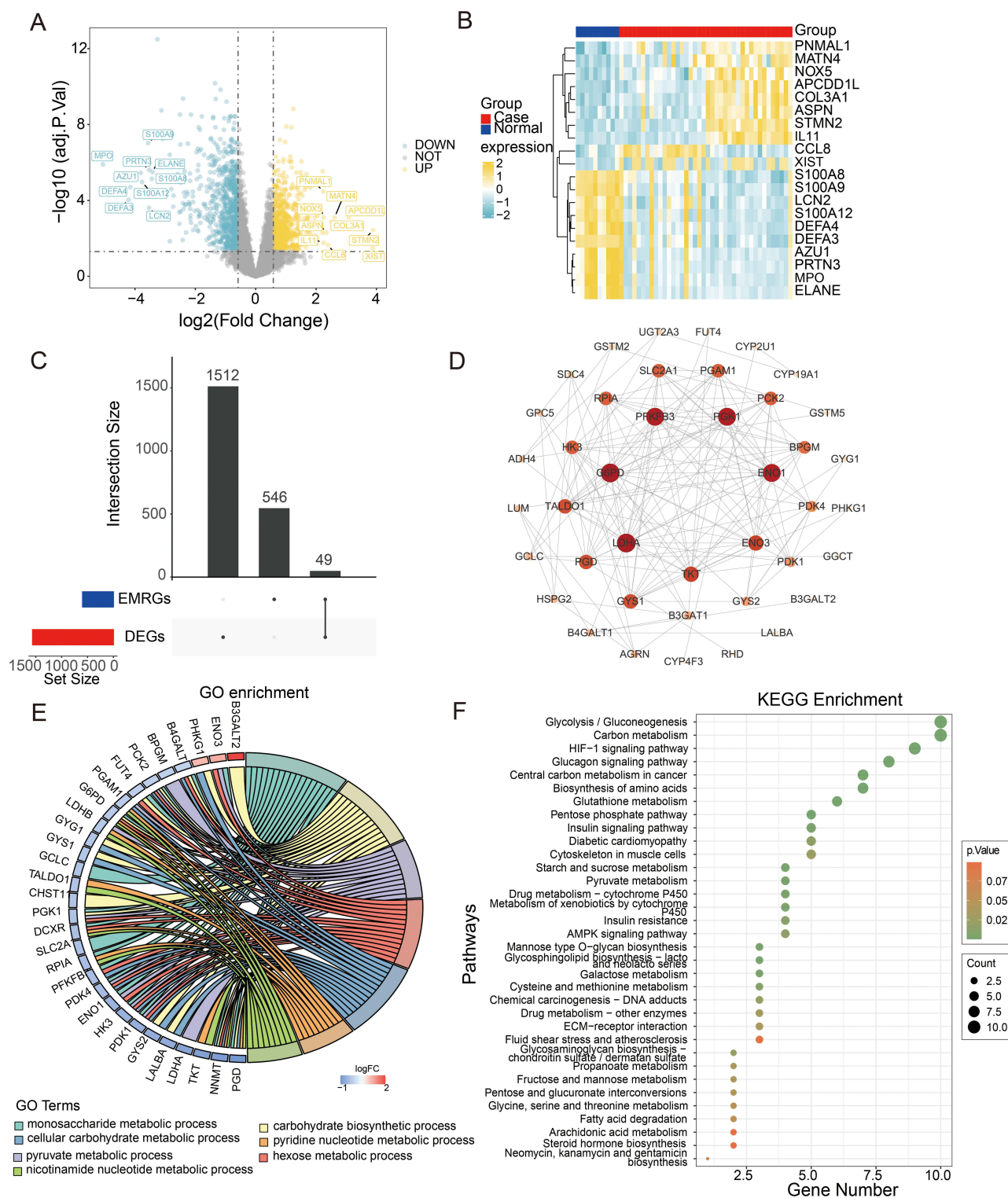
### Consensus Clustering Analysis Based on DEEMRGs

To further investigate the DEEMRGs, we conducted consensus clustering analysis using the “ConsensusClusterPlus” package. As illustrated in Figure 2A–D, when  $K=2$ , the within-group variability was minimal, while the between-group differences were more pronounced. The changes in the cumulative distribution function (CDF) also provided clear insight into the quality of the clustering results. As shown in Figure 2F, the smallest change in the CDF occurred at  $K=2$ , suggesting that dividing the samples into two subtypes was the most appropriate. This conclusion is further supported by Figure 2E, which confirms that  $K=2$  provided the optimal clustering effect. Consequently, the 40 OA samples in the training set were divided into two subtypes. The PCA plot of the subtypes, presented in Figure 2G, demonstrates distinct separation between the two groups.

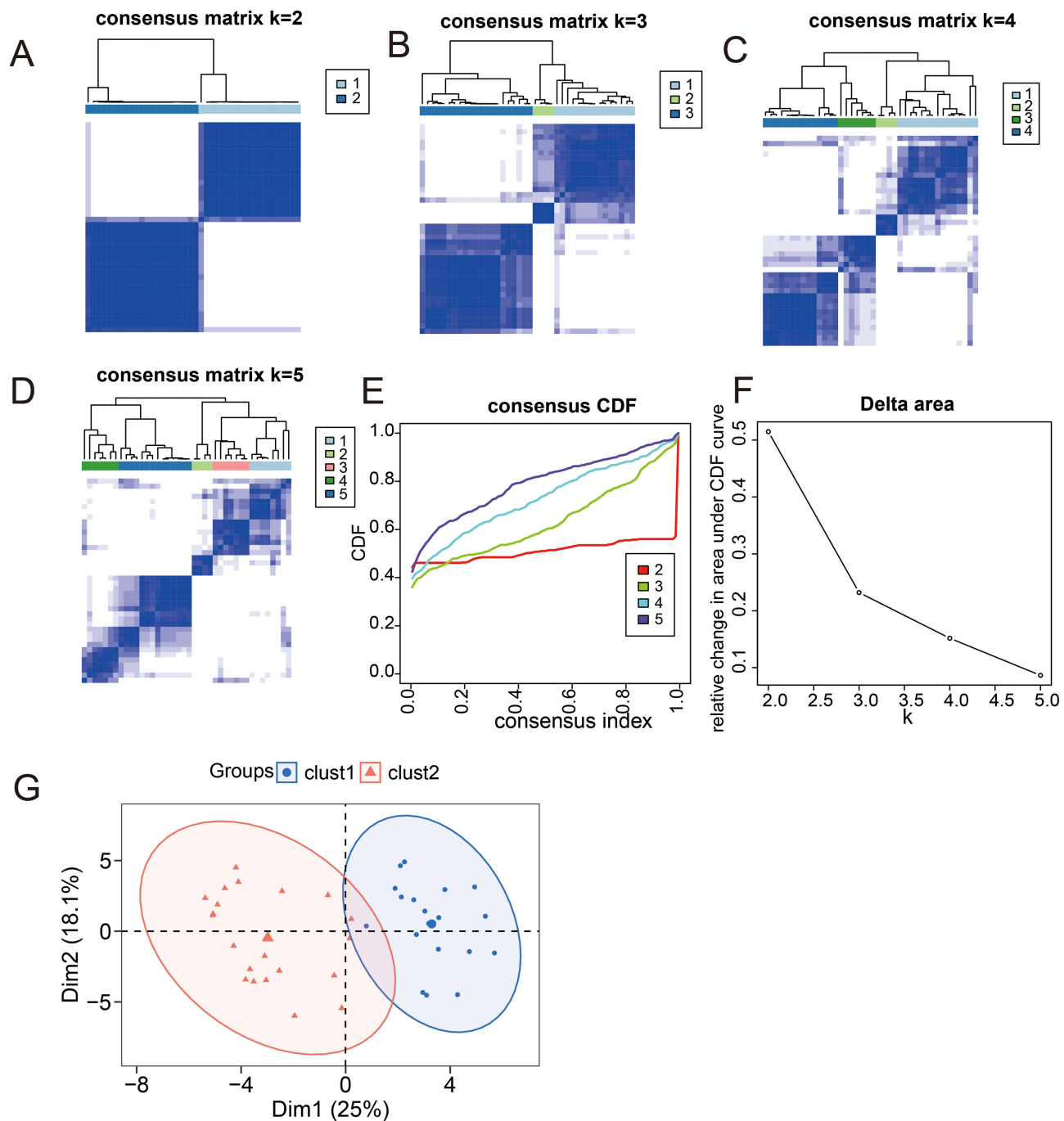
### Differential Expression Analysis and GSEA of Molecular Subtypes

Following an in-depth analysis of differential expression between the two molecular subtypes, a total of 1,785 differential genes were identified. As illustrated in Figure 3A, there are 930 genes upregulated in clust1 relative to clust2, and 855 genes downregulated in clust1 relative to clust2. Figure 3B displays a heatmap of the expression levels for the top 20 differential genes based on relevance score. To further elucidate the molecular mechanisms underlying the subtypes, GSEA was conducted. The enrichment analysis results for upregulated differential genes, shown in Figure 3C, indicate significant enrichment in pathways such as the cytoskeleton in muscle cells, human papillomavirus infection, cell cycle, motor proteins, and protein digestion and absorption. Conversely, downregulated genes were predominantly enriched in pathways associated with cytokine-cytokine receptor interaction, lipid metabolism and atherosclerosis, the MAPK signaling pathway, Chagas disease, and the chemokine signaling pathway (Figure 3D).





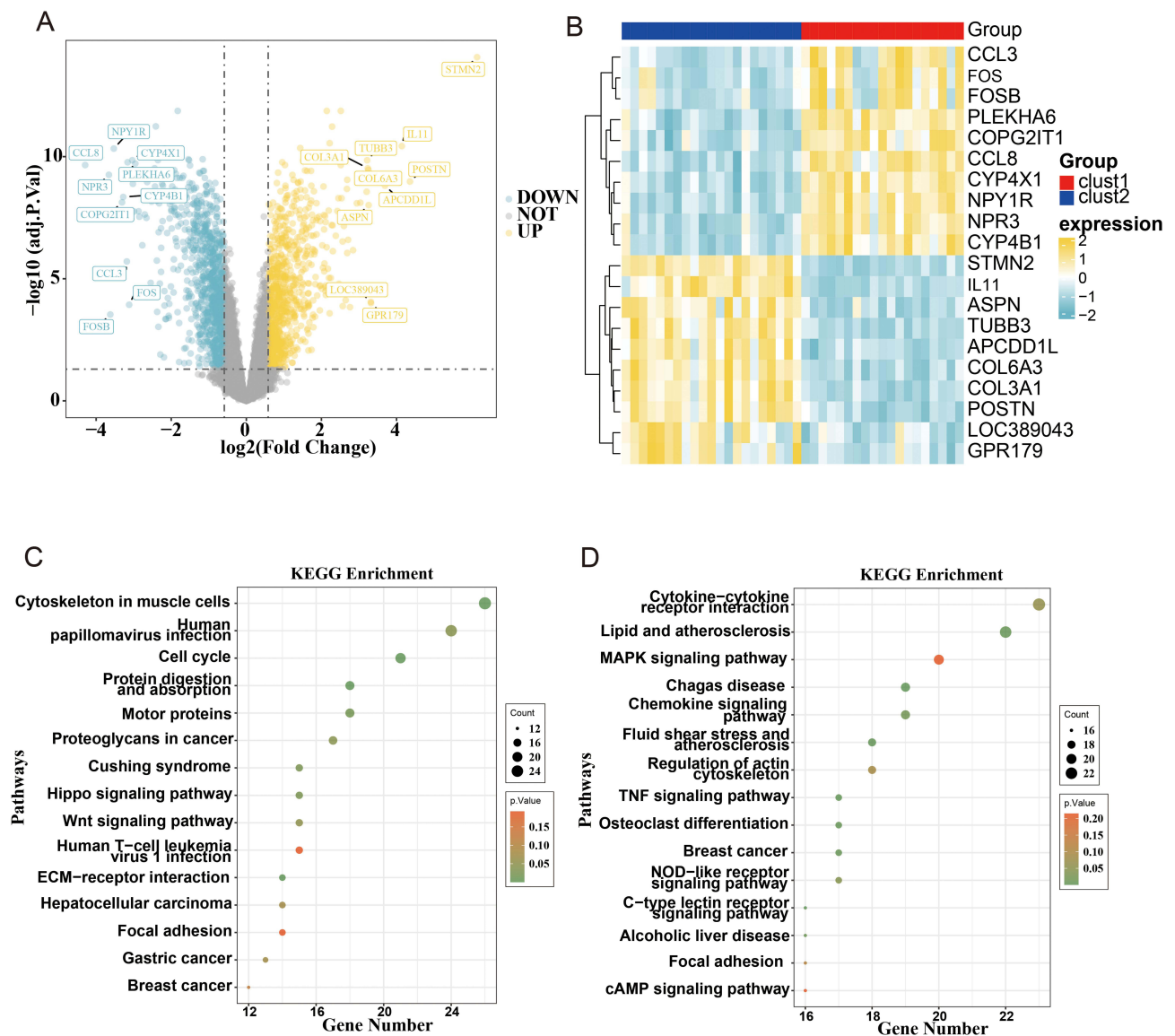
**Figure 1** Identification and preliminary analysis of DEEMRGs. **(A)** The volcano plot illustrates the DEGs between normal tissues and OA tissues, with 815 downregulated genes and 746 upregulated genes. **(B)** The heatmap displays the expression levels of the top 20 DEGs by relevance score in different tissues, with red representing OA tissues and blue representing normal tissues. Values greater than 0 indicate high expression, while values less than 0 indicate low expression. **(C)** The Upset plot shows that the number of DEGs is 1,512, the number of EMRGs is 546, and the number of intersecting genes is 49. **(D)** The PPI network diagram of the 49 DEEMRGs. **(E)** GO enrichment analysis of the biological processes involving DEEMRGs. **(F)** KEGG pathway enrichment analysis of DEEMRGs.



**Figure 2** Consensus clustering analysis based on DEEMRGs. Consensus matrix heatmap at  $K=2$  (A),  $K=3$  (B),  $K=4$  (C),  $K=5$  (D). (E) Consensus cumulative distribution plot of the consensus clustering analysis. The X-axis represents the consensus index, and the Y-axis represents the change in the CDF value. CDF: cumulative distribution function. (F) Delta area plot of the consensus clustering analysis.  $K=2$  indicates the optimal number of clusters. (G) The PCA plot of molecular subtypes shows significant differences between the two molecular subtypes. Blue represents molecular subtype group 1, and red represents molecular subtype group 2.

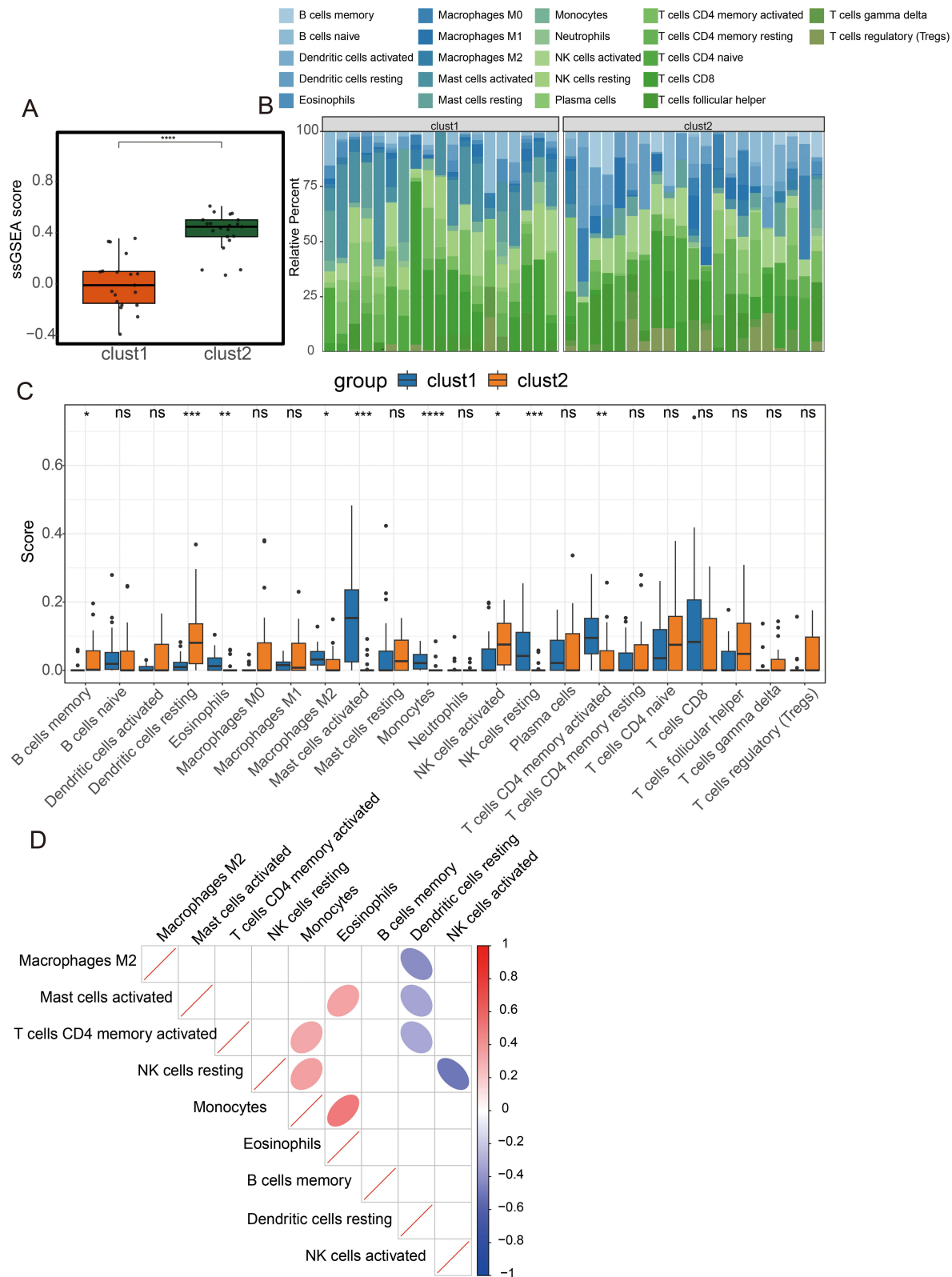
## Immune Infiltration Analysis of Molecular Subtypes

The level of immune cell infiltration is strongly associated with disease onset and progression. Thus, we utilized ssGSEA to analyze the abundance of immune cell infiltration between the two subtypes. The results, as shown in Figure 4A, indicate that subtype 2 exhibits a significantly higher abundance of immune cell infiltration compared to subtype 1. To further explore the infiltration levels of immune cells, we applied the CIBERSORT algorithm to evaluate the infiltration of 22 types of immune cells in each subtype (Figure 4B). The expression levels of these immune cells are presented in



**Figure 3** Differential expression analysis and GSEA of molecular subtypes. **(A)** The volcano plot illustrates the genes that differ between the two molecular subtypes, revealing a total of 930 genes that are upregulated and 855 that are downregulated. **(B)** The heatmap shows the expression levels of the top 20 differential genes in the two molecular subtype groups. **(C)** KEGG pathway enrichment analysis results for upregulated genes. **(D)** Analysis of KEGG for downregulated genes.

Figure 4C, revealing nine differential immune cells. Specifically, Eosinophils, Macrophages M2, Mast cells activated, Monocytes, NK cells resting, and T cells CD4 memory activated are highly expressed in clust1, while B cells memory, Dendritic cells resting, and NK cells activated are more prevalent in clust2. Existing literature highlights the role of Macrophages in immune suppression, immune evasion, and resistance to immunotherapy, whereas NK cells activated are crucial for inhibiting disease progression.<sup>14,15</sup> Therefore, patients in subtype 2 are more likely to benefit from immune activation strategies to suppress disease development. To enhance our understanding of immune cell interactions and assist OA patients in preventing disease exacerbation, we performed Spearman correlation analysis of the differential immune cells between subtypes 1 and 2 and visualized the results using a heatmap. Figure 4D illustrates a notable negative correlation between activated NK cells and resting NK cells. Meanwhile, resting dendritic cells express significant negative correlations with activated M2 macrophages, activated mast cells, and activated CD4 memory T cells. Eosinophils, on the other hand, display significant positive correlations with M2 macrophages and monocytes, whereas monocytes show strong positive correlations with activated CD4 memory T cells and resting NK cells.



**Figure 4** Immune infiltration analysis of molecular subtypes. **(A)** The boxplot represents the ssGSEA scores of the two molecular subtypes. **(B)** The distribution plot of immune cell infiltration levels between molecular subtypes. **(C)** The boxplot shows the differential infiltration levels of immune cells between molecular subtypes. **(D)** The heatmap represents the correlation analysis of immune cell infiltration. For all figures: "ns" represents no significant difference, \* represents  $P < 0.05$ , \*\* represents  $P < 0.01$ , and \*\*\* represents  $P < 0.001$ .

## Screening Feature Genes Based on Machine Learning

To identify diagnostic biomarkers for OA from EMRGs, we first applied LASSO analysis to eliminate highly correlated genes, resulting in the selection of 11 feature genes, as shown in [Figure 5A](#) and [B](#). We then performed further analysis using RF. The residual distribution plot of the RF ([Figure 5C](#)) shows the number of trees on the X-axis and the error on the Y-axis, indicating a gradual decrease in error as the number of trees increases. Based on the MeanDecreaseGini metric, we ranked the top 10 most important genes ([Figure 5D](#)). Additionally, we used SVM-RFE to select core feature genes, with the results shown in [Figure 5E](#), identifying a classifier with the minimum cross-validation error of 3. Finally, by intersecting the genes selected through LASSO, SVM-RFE, and RF, we identified 2 key feature genes: NUP98 and RPIA ([Figure 5F](#)).

## Diagnostic Value and Expression Level Assessment of Feature Genes Based on EMRGs

We further investigated the diagnostic value of the feature genes NUP98 and RPIA for OA using ROC curves. In the training set GSE51588, the AUC values for NUP98 and RPIA were 1 and 0.992, respectively ([Figure 6A](#)). The results shown in [Figure 6B](#) indicate that the expression levels of NUP98 and RPIA in OA tissues are significantly lower than those in normal tissues ( $P < 0.001$ ). In the validation set GSE63359, the AUC values for the ROC curves were 0.748 for NUP98 and 0.703 for RPIA ([Figure 6C](#)). Similarly, the expression levels of these feature genes in OA tissues were significantly lower than those in normal tissues ([Figure 6D](#)). Additionally, [Table 1](#) presents the NPV, PPV, sensitivity, and specificity of the diagnostic gene ROC curves for both the training and validation sets, demonstrating the robust diagnostic performance of these two genes. The ROC curve analysis and expression patterns of these feature genes demonstrate their high accuracy and reliability in OA diagnosis. Notably, after analyzing the expression levels of these feature genes in the two molecular subtypes, we found that NUP98 and RPIA are significantly more expressed in subtype 1 compared to subtype 2 ([Figure 6E](#)). [Figure 6F](#) compares the expression levels of the feature genes between the subtypes and normal tissues, showing that NUP98 and RPIA are significantly more expressed in the control samples compared to subtype 1, and subtype 1 exhibits higher expression levels than subtype 2. The significant differences in NUP98 and RPIA expression between normal and OA tissues provide strong support for their potential as biomarkers for OA diagnosis.

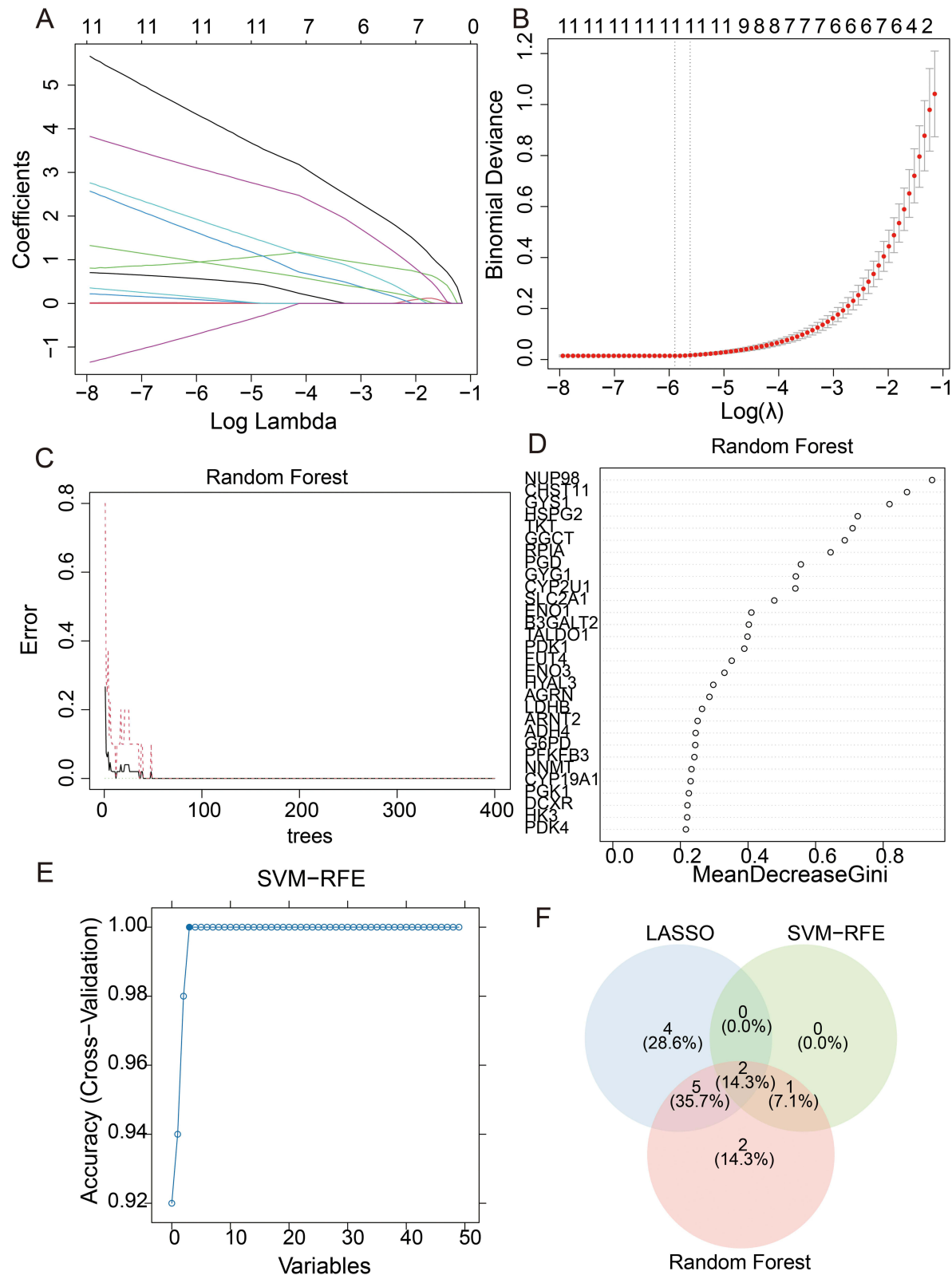
## Construction of an OA Diagnostic Signature Based on EMRGs

To provide a clearer visualization of the diagnostic performance of the characteristic genes, we constructed a comprehensive nomogram ([Figure 7A](#)). The nomogram, incorporating gene expression levels and risk scores, significantly improved our ability to predict OA risk. As shown in [Figure 7B](#), the AUC values for the ROC curves of the nomogram, NUP98, and RPIA were 1, 1, and 0.992, respectively, indicating excellent diagnostic accuracy. Furthermore, the calibration curve ([Figure 7C](#)) demonstrated that the nomogram's predictions closely aligned with actual observations. The DCA was employed to evaluate the clinical utility of this diagnostic model. As illustrated in [Figure 7D](#), the nomogram showed a substantial net benefit, highlighting its strong potential for use in clinical decision process. Overall, the diagnostic model based on EMRGs exhibited outstanding diagnostic efficacy and precision.

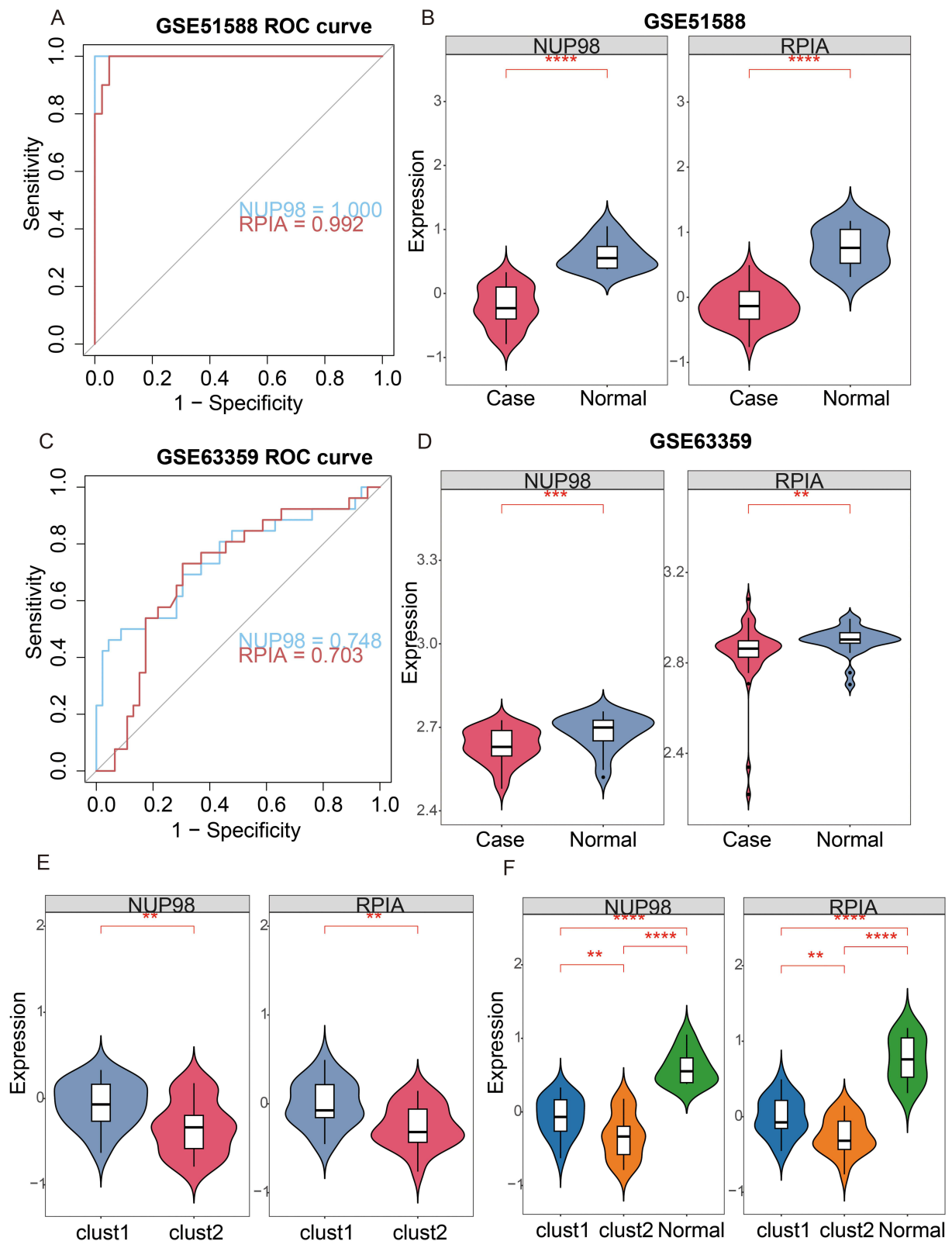
## GSEA Enrichment Analysis and Construction of the ceRNA Network for Characteristic Genes

To further investigate the functional roles of the characteristic genes, we conducted GSEA analysis. The enrichment analysis of NUP98 ([Figure 8A](#)) revealed that it is primarily involved in pathways such as hypertrophic cardiomyopathy, spliceosome, leishmania infection, receptor interaction, and olfactory transduction. Similarly, RPIA was found to participate in olfactory transduction, ribosome function, receptor interaction, basal cell carcinoma, and the hedgehog signaling pathway ([Figure 8B](#)). Additionally, by constructing a ceRNA network, we identified 18 pairs of relationships involving 4 miRNAs and 8 lncRNAs. The interactions between the characteristic genes, lncRNAs, and miRNAs were highly interconnected ([Figure 8C](#)).





**Figure 5** Feature genes were selected based on machine learning. **(A)** Ten cross-validations for selecting adjustment parameters in the LASSO model. Each curve signifies a distinct gene. **(B)** LASSO coefficient evaluation. Vertical dashed lines indicate the optimal lambda. **(C)** The graph produced by the RF tree illustrates the correlation between the quantity of trees represented on the X-axis and the cross-validation error plotted on the Y-axis. **(D)** Feature importance plot of the RF. **(E)** The SVM-RFE feature selection result plot, with the horizontal axis representing variables and the vertical axis representing cross-validation error. **(F)** The Venn diagram illustrates the intersection of gene sets after applying three different machine learning methods.



**Figure 6** Diagnostic value and expression level assessment of feature genes based on EMRGs. **(A)** In the GSE51588 training set, the ROC curve AUC value for NUP98 is 1, while the AUC value for RPIA is 0.992. **(B)** Comparison of the expression levels of feature genes between normal and OA tissues in the training set. **(C)** In the GSE63359 validation set, the ROC curve AUC value for NUP98 is 0.748, and for RPIA, it is 0.703. **(D)** Comparison of the expression levels of feature genes between normal and OA tissues in the validation set. **(E)** Analyzing the expression levels of characteristic genes across various molecular subtypes. **(F)** A thorough analysis of the expression levels of key genes in relation to the two molecular subtypes and normal tissues. For all figures: "ns" represents no significant difference, \*\* represents  $P < 0.01$ , \*\*\* represents  $P < 0.001$ , and \*\*\*\* represents  $P < 0.0001$ .

**Table 1** Detailed Information of the GWAS in Our Analysis

Datasets	Gene	AUC	Sensitivity	Specificity	PPV	NPV	CutOff
Training set	NUP98	1	1	1	1	1	0.355
	RPIA	0.992	1	0.95	1	0.833	0.298
Validation set	NUP98	0.748	0.462	0.957	0.759	0.857	5.544
	RPIA	0.703	0.731	0.696	0.821	0.576	6.422

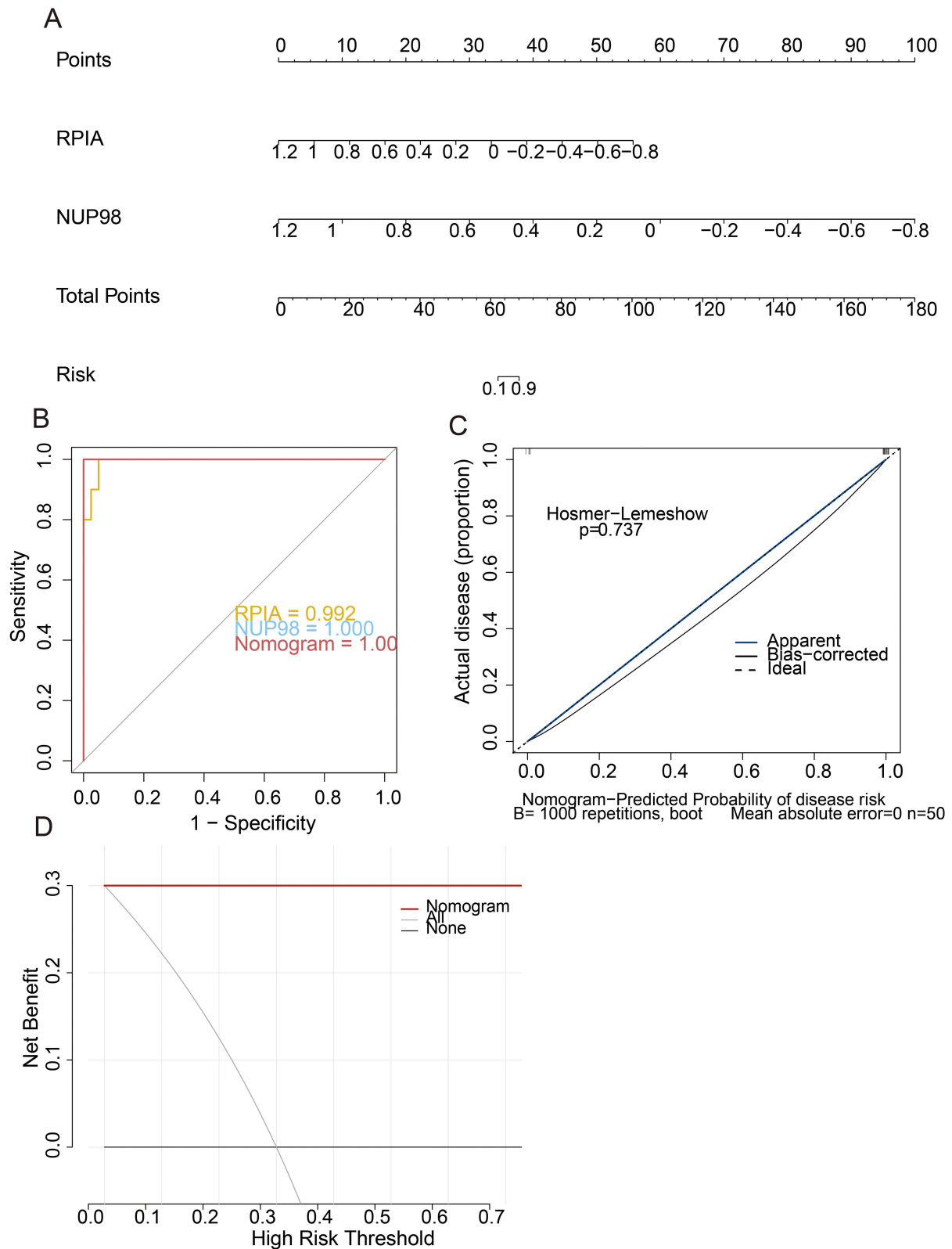
Analysis of the Immunological Landscape

To clearly depict the immune landscape in OA patients, we analyzed the immune cell infiltration status in normal and OA tissues (Figure 9A). Notably, an examination of immune cell infiltration through differential analysis showed considerable variations among 15 different immune cell types across the groups (Figure 9B). Among these, naive B cells, CD8<sup>+</sup> T cells, naive CD4<sup>+</sup> T cells, Tregs, activated NK cells, M1 and M2 macrophages, and resting dendritic cells showed higher infiltration in OA tissues, whereas plasma cells, resting memory CD4<sup>+</sup> T cells, monocytes, activated dendritic cells, resting mast cells, activated mast cells, and neutrophils exhibited lower infiltration in OA tissues. We further conducted a detailed analysis of the interactions between these immune cells to uncover the relationships among immune infiltrates within the immune microenvironment (Figure 9C). Crucially, we also explored the relationship between the expression levels of feature genes and the infiltration of immune cells. Figure 9D indicates a notable negative correlation between NUP98 and naive B cells, CD8<sup>+</sup> T cells, Tregs, activated NK cells, as well as activated dendritic cells. Conversely, it shows a positive correlation with plasma cells, resting memory CD4<sup>+</sup> T cells, resting mast cells, and neutrophils. On the other hand, RPIA displayed a negative correlation with CD8<sup>+</sup> T cells, follicular helper T cells, Tregs, activated NK cells, M1 macrophages, and resting dendritic cells. However, it featured a positive correlation with plasma cells, resting memory CD4<sup>+</sup> T cells, activated memory CD4<sup>+</sup> T cells, monocytes, resting mast cells, and neutrophils.

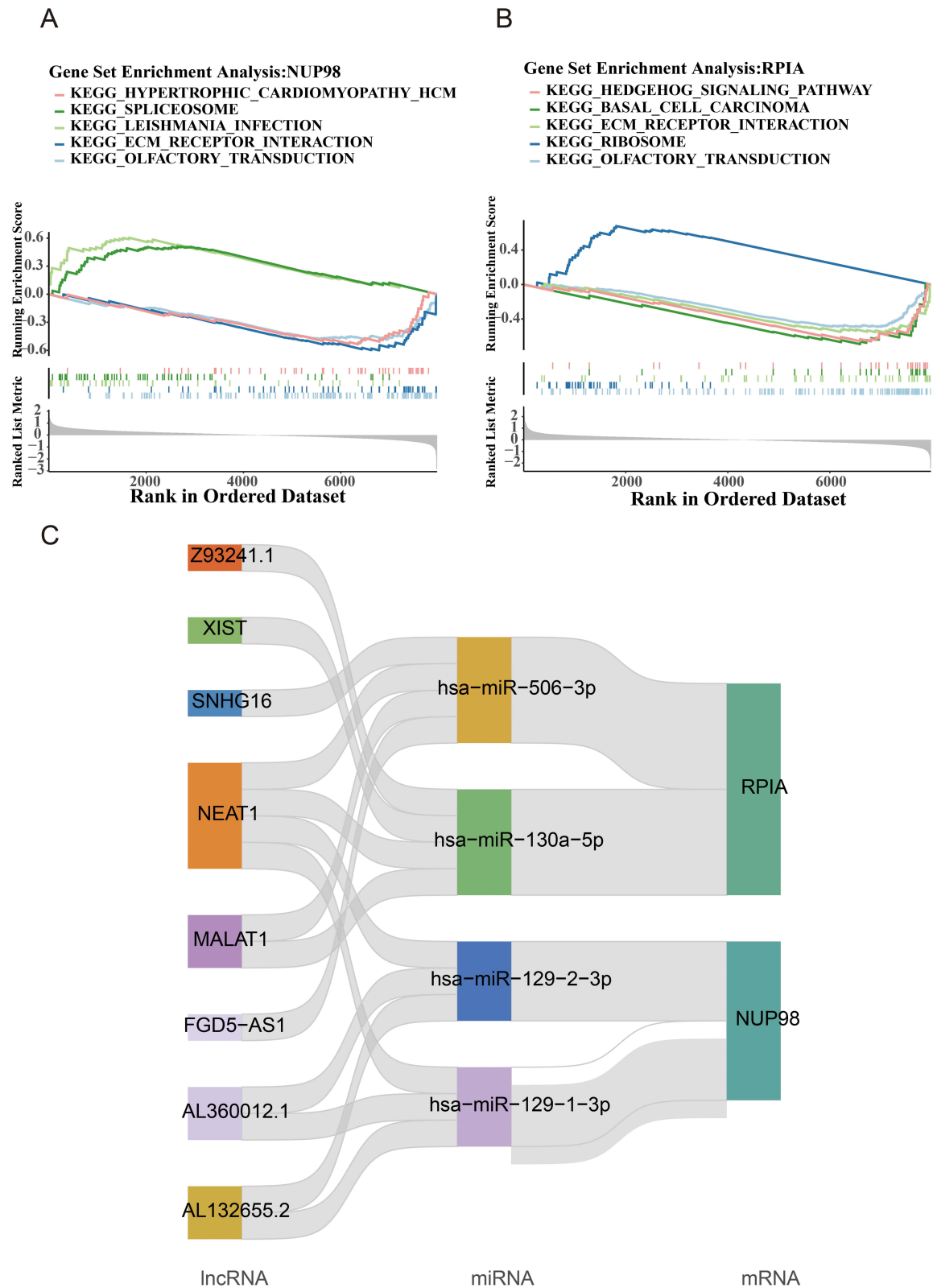
Discussion

OA is a chronic neurodegenerative joint disorder that significantly affects the quality of life in middle-aged and elderly individuals.<sup>16</sup> From a histopathological perspective, OA is marked by a gradual deterioration of the articular cartilage, the emergence of osteophytes, and the formation of subchondral cysts.<sup>16</sup> Under adverse conditions, metabolic activity shifts from a resting state to a highly active one, playing a key role in combating disease onset and progression.<sup>17</sup> Metabolic processes are crucial for the functioning of various organs, including cartilage and synovial joints.<sup>18</sup> Glucose serves as the primary energy source for the body, and its breakdown provides the necessary energy to maintain normal physiological functions. Cartilage, in particular, relies on an adequate energy supply for its proper function.<sup>19</sup> Disruptions in energy metabolism can lead to abnormalities in intracellular biological processes, laying the foundation for disease development.<sup>10,12</sup> The present study, therefore, investigates the diagnostic potential of EMRGs in OA, exploring their functional roles in OA progression and providing a theoretical basis for identifying diagnostic biomarkers and informing clinical decision making in OA.

This study identified two key genes associated with OA progression, NUP98 and RPIA, from the EMRGs gene set using three machine learning methods: LASSO, SVM-RFE, and RF. Nucleoporin 98 (NUP98) encodes a protein that forms part of the nuclear pore complex.<sup>20</sup> Research indicates that nuclear pore proteins (Nups) interact with chromatin to regulate gene expression and chromatin architecture.<sup>21</sup> According to S. Capitanio et al, NUP98 interacts with several DExH/D-box proteins (DBPs) to modulate gene expression and RNA metabolism.<sup>22</sup> RNA metabolism plays a pivotal role in chondrocyte senescence, a key process in OA pathology. Specifically, the long noncoding RNA ELDR is known to accelerate chondrocyte aging and OA progression.<sup>23</sup> Abnormal gene expression forms the basis of disease development and is crucial for identifying disease biomarkers.<sup>24</sup> Our findings demonstrate that NUP98 expression is significantly reduced in OA tissues. ROC and DCA curve analyses further confirmed the high diagnostic accuracy of NUP98 in OA.

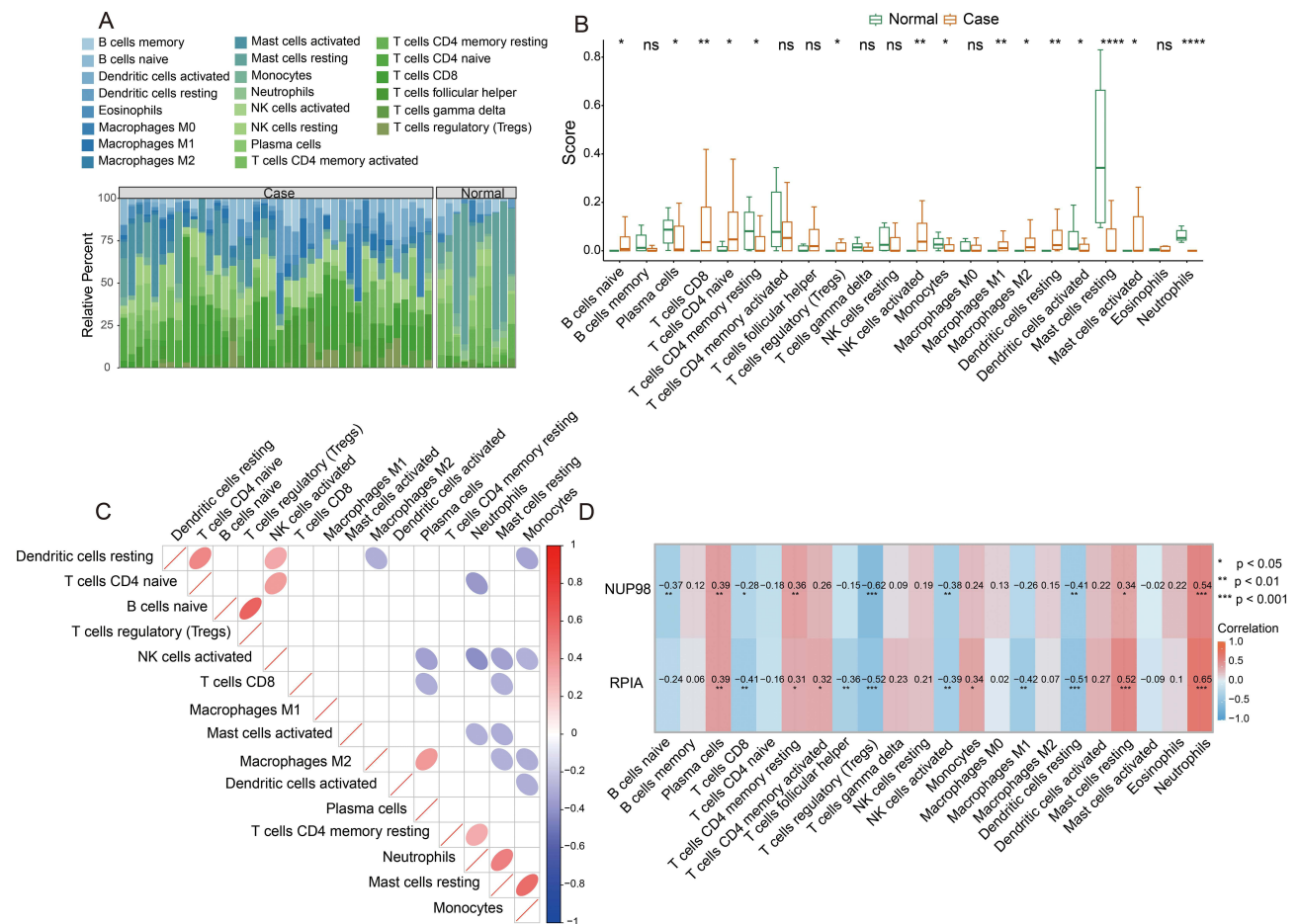


**Figure 7** Establishment of an OA diagnostic model based on EMRGs. **(A)** A nomogram created from the two feature genes. **(B)** The results of the ROC curve show that the AUC values for the nomogram, NUP98, and RPIA are 1, 1, and 0.992, respectively. **(C)** The calibration curve: the horizontal axis represents the predicted probability of different clinical outcomes by the model, and the vertical axis represents the observed probability of patient clinical outcomes. The dashed line represents the ideal curve for comparison. **(D)** The DCA curve for the feature genes and diagnostic model: "None" represents the clinical outcome where none of the OA patients are diagnosed with the disease, and "All" represents the outcome where all OA patients are diagnosed. The X-axis indicates the threshold probability, while the Y-axis shows the net gain.



**Figure 8** GSEA enrichment analysis and construction of the ceRNA network for feature genes. **(A)** The GSEA enrichment results indicate the pathways involving NUP98. **(B)** The GSEA enrichment findings reveal the pathways associated with RPIA. **(C)** The Sankey diagram illustrates the ceRNA network of the feature genes.





**Figure 9** Immune cell infiltration analysis. **(A)** A comparison of the infiltration levels of 22 different immune cell types in both normal and osteoarthritic tissues. **(B)** The box plot illustrates the comparative analysis of immune cell infiltration across various tissues. **(C)** The heatmap depicts the correlation analysis related to differences in immune cell infiltration. **(D)** The examination of the relationship between feature gene expression levels and immune cell infiltration. For all figures: "ns" represents no significant difference, \* represents  $P < 0.05$ , \*\* represents  $P < 0.01$ , \*\*\* represents  $P < 0.001$ , and \*\*\*\* represents  $P < 0.0001$ .

The another gene, ribose-5-phosphate isomerase A (RPIA), has been implicated in the development of several malignancies, including liver cancer<sup>25</sup> and colorectal cancer.<sup>26</sup> Our results suggest that RPIA holds promise as a potential diagnostic biomarker for OA. While RPIA is highly expressed in normal tissues, its levels are significantly decreased in OA tissues. Importantly, further validation through nomogram analysis and ROC curves confirmed the high efficiency of RPIA as a diagnostic marker for OA.

OA is a complex disease driven by multiple factors, including inflammation.<sup>27</sup> Immune cells, particularly macrophages, play a critical role as mediators of the inflammatory response.<sup>27</sup> Macrophages are highly heterogeneous and can be classified into activated M1 and M2 macrophages, as well as classically activated macrophages.<sup>28</sup> M1 macrophages are pro-inflammatory, secreting cytokines such as interleukin-6 (IL-6), inducible nitric oxide synthase (iNOS), and tumor necrosis factor- $\alpha$  (TNF- $\alpha$ ).<sup>29</sup> Conversely, M2 macrophages produce anti-inflammatory factors, including IL-10, IL-1 receptor antagonist, and arginase, which inhibit IL-1 $\beta$  and iNOS activity.<sup>29</sup> Studies have shown that shifting macrophages from the M1 to the M2 phenotype may have therapeutic benefits for OA.<sup>7,30,31</sup> In this study, ssGSEA analysis revealed elevated infiltration of both M1 and M2 macrophages in OA tissues, suggesting that targeted regulation of macrophage polarization could be a promising therapeutic strategy for OA. In addition, our analysis revealed a strong negative correlation between RPIA expression and the infiltration levels of M1 macrophages, which supports the notion of tailoring OA treatments to individual patients.

However, there are several limitations in this study. First, this research was based on transcriptomic data obtained from publicly available databases, which lacks access to additional clinically relevant data from OA samples, potentially

impacting the results. Secondly, further experimental validation is needed to confirm the inflammatory findings and strengthen our conclusions. In summary, our study provides valuable insights into targeting energy metabolism as a therapeutic approach for OA, and future research could expand and deepen the investigation in this area.

## Conclusion

This study identified two diagnostic genes associated with EMRGs in OA through bioinformatics combined with machine learning algorithms, emphasizing the connection between energy metabolism and OA progression. The diagnostic accuracy of these feature genes was validated via the construction of nomogram and ROC curve. Furthermore, we examined the differences in immune cell infiltration between normal and OA tissues, and explored the correlation between feature genes and immune infiltration patterns. Collectively, this analysis underscores the critical role of energy metabolism in OA, providing theoretical support for personalized treatment strategies and clinical diagnostic decisions for OA patients.

## Data Sharing Statement

The data and materials in the current study are available from the corresponding author on reasonable request.

## Ethics Approval and Consent to Participate

This study involves publicly available human data and complies with the exemption criteria outlined in items 1 and 2 of Article 32 of the Measures for Ethical Review of Life Science and Medical Research Involving Human Subjects (issued on February 18, 2023, China). Specifically:

1. The research involves publicly accessible data that has been legally obtained and does not involve the collection of new biological samples or private information from participants.
2. The research poses no risk to the participants' rights or health and does not violate ethical principles.

Based on the above guidelines, this study is exempt from review by the Institutional Review Board (IRB) or local ethics committee. All data used in this study are de-identified and anonymized prior to public release, ensuring the protection of participants' privacy.

## 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 declare that they have no potential conflicts of interest.

## References

1. Roelofs AJ, De Bari C. Osteoarthritis year in review 2023: biology. *Osteoarthritis Cartilage*. 2024;32(2):148–158. doi:10.1016/j.joca.2023.11.002
2. Yao Q, Wu X, Tao C, et al. Osteoarthritis: pathogenic signaling pathways and therapeutic targets. *Signal Transduct Target Ther*. 2023;8(1):56. doi:10.1038/s41392-023-01330-w
3. Moldovan F. Correlation between peripheric blood markers and surgical invasiveness during Humeral Shaft fracture osteosynthesis in young and middle-aged patients. *Diagnostics*. 2024;14(11):1112. doi:10.3390/diagnostics14111112
4. Green J, Tinson RAJ, Betts JHJ, et al. Suramin analogues protect cartilage against osteoarthritic breakdown by increasing levels of tissue inhibitor of metalloproteinases 3 (TIMP-3) in the tissue. *Bioorg Med Chem*. 2023;92:117424. doi:10.1016/j.bmc.2023.117424

5. Judge A, Dodd MS. Metabolism. *Essays Biochem.* **2020**;64(4):607–647. doi:10.1042/EBC20190041
6. Qi Z, Zhu J, Cai W, Lou C, Li Z. The role and intervention of mitochondrial metabolism in osteoarthritis. *mol Cell Biochem.* **2024**;479(6):1513–1524. doi:10.1007/s11010-023-04818-9
7. Zhao K, Ruan J, Nie L, Ye X, Li J. Effects of synovial macrophages in osteoarthritis. *Front Immunol.* **2023**;14:1164137. doi:10.3389/fimmu.2023.1164137
8. Runge N, Aina A, May S. The benefits of adding manual therapy to exercise therapy for improving pain and function in patients with knee or hip osteoarthritis: a systematic review with meta-analysis. *J Ortho Sports Phys Ther.* **2022**;52(10):675–a613. doi:10.2519/jospt.2022.11062
9. Courties A, Sellam J, Berenbaum F. Metabolic syndrome-associated osteoarthritis. *Curr Opin Rheumatol.* **2017**;29(2):214–222. doi:10.1097/BOR.0000000000000373
10. Wu X, Liyanage C, Plan M, et al. Dysregulated energy metabolism impairs chondrocyte function in osteoarthritis. *Osteoarthritis Cartilage.* **2023**;31(5):613–626. doi:10.1016/j.joca.2022.11.004
11. Sun K, Jing X, Guo J, Yao X, Guo F. Mitophagy in degenerative joint diseases. *Autophagy.* **2021**;17(9):2082–2092. doi:10.1080/15548627.2020.1822097
12. Li C, Li X, Li G, et al. Identification of a prognosis-associated signature associated with energy metabolism in triple-negative breast cancer. *Oncol Rep.* **2020**;44(3):819–837. doi:10.3892/or.2020.7657
13. Nong J, Lu G, Huang Y, et al. Identification of cuproptosis-related subtypes, characterization of immune microenvironment infiltration, and development of a prognosis model for osteoarthritis. *Front Immunol.* **2023**;14:1178794. doi:10.3389/fimmu.2023.1178794
14. Locati M, Curtale G, Mantovani A. Diversity, mechanisms, and significance of macrophage plasticity. *Ann Rev Pathol.* **2020**;15:123–147. doi:10.1146/annurev-pathmechdis-012418-012718
15. Fang F, Xie S, Chen M, et al. Advances in NK cell production. *Cell mol Immunol.* **2022**;19(4):460–481. doi:10.1038/s41423-021-00808-3
16. Millerand M, Berenbaum F, Jacques C. Danger signals and inflammaging in osteoarthritis. *Clin Exp Rheumatol.* **2019**;37(Suppl 120):48–56.
17. Zheng L, Zhang Z, Sheng P, Mobasheri A. The role of metabolism in chondrocyte dysfunction and the progression of osteoarthritis. *Ageing Res Rev.* **2021**;66:101249. doi:10.1016/j.arr.2020.101249
18. Mobasheri A, Rayman MP, Gualillo O, Sellam J, van der Kraan P, Fearon U. The role of metabolism in the pathogenesis of osteoarthritis. *Nat Rev Rheumatol.* **2017**;13(5):302–311. doi:10.1038/nrrheum.2017.50
19. Hollander JM, Zeng L. The emerging role of glucose metabolism in cartilage development. *Curr Osteoporosis Rep.* **2019**;17(2):59–69. doi:10.1007/s11914-019-00506-0
20. Pascual-Garcia P, Little SC, Capelson M. Nup98-dependent transcriptional memory is established independently of transcription. *eLife.* **2022**;11. doi:10.7554/eLife.63404
21. Sump B, Brickner JH. Nup98 regulation of histone methylation promotes normal gene expression and may drive leukemogenesis. *Genes Dev.* **2017**;31(22):2201–2203. doi:10.1101/gad.310359.117
22. Capitanio JS, Montpetit B, Wozniak RW. Nucleoplasmic Nup98 controls gene expression by regulating a DEXH/D-box protein. *Nucleus.* **2018**;9(1):1–8. doi:10.1080/19491034.2017.1364826
23. Ji ML, Li Z, Hu XY, Zhang WT, Zhang HX, Lu J. Dynamic chromatin accessibility tuning by the long noncoding RNA ELDR accelerates chondrocyte senescence and osteoarthritis. *Am J Hum Genet.* **2023**;110(4):606–624. doi:10.1016/j.ajhg.2023.02.011
24. Oliva M, Muñoz-Aguirre M, Kim-Hellmuth S, et al. The impact of sex on gene expression across human tissues. *Science.* **2020**;369(6509). doi:10.1126/science.aba3066
25. Ciou SC, Chou YT, Liu YL, et al. Ribose-5-phosphate isomerase A regulates hepatocarcinogenesis via PP2A and ERK signaling. *Int J Cancer.* **2015**;137(1):104–115. doi:10.1002/ijc.29361
26. Chou YT, Jiang JK, Yang MH, et al. Identification of a noncanonical function for ribose-5-phosphate isomerase A promotes colorectal cancer formation by stabilizing and activating  $\beta$ -catenin via a novel C-terminal domain. *PLoS Biol.* **2018**;16(1):e2003714. doi:10.1371/journal.pbio.2003714
27. Nedunchezhiyan U, Varughese I, Sun AR, Wu X, Crawford R, Prasadam I. Obesity, inflammation, and immune system in osteoarthritis. *Front Immunol.* **2022**;13:907750. doi:10.3389/fimmu.2022.907750
28. McLaughlin T, Ackerman SE, Shen L, Engleman E. Role of innate and adaptive immunity in obesity-associated metabolic disease. *J Clin Invest.* **2017**;127(1):5–13. doi:10.1172/JCI88876
29. Sun AR, Friis T, Sekar S, Crawford R, Xiao Y, Prasadam I. Is synovial macrophage activation the inflammatory link between obesity and osteoarthritis? *Curr Rheumatol Rep.* **2016**;18(9):57. doi:10.1007/s11926-016-0605-9
30. Zhang H, Cai D, Bai X. Macrophages regulate the progression of osteoarthritis. *Osteoarthritis Cartilage.* **2020**;28(5):555–561. doi:10.1016/j.joca.2020.01.007
31. Wu CL, Harasymowicz NS, Klimak MA, Collins KH, Guilak F. The role of macrophages in osteoarthritis and cartilage repair. *Osteoarthritis Cartilage.* **2020**;28(5):544–554. doi:10.1016/j.joca.2019.12.007