Identification of immune-related genes in diagnosing atherosclerosis with rheumatoid arthritis through bioinformatics analysis and machine learning

Fuze Liu[1], Yue Huang[1], Fuhui Liu[2] and Hai Wang[1*]

[1]Department of Orthopaedic Surgery, Peking Union Medical College Hospital, Peking Union Medical College and Chinese Academy of Medical Sciences, Beijing, China, [2]School of Clinical Medical, Weifang Medical University, Weifang, China

Background: Increasing evidence has proven that rheumatoid arthritis (RA) can aggravate atherosclerosis (AS), and we aimed to explore potential diagnostic genes for patients with AS and RA.

Methods: We obtained the data from public databases, including Gene Expression Omnibus (GEO) and STRING, and obtained the differentially expressed genes (DEGs) and module genes with Limma and weighted gene co-expression network analysis (WGCNA). Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) enrichment analysis, the protein–protein interaction (PPI) network, and machine learning algorithms [least absolute shrinkage and selection operator (LASSO) regression and random forest] were performed to explore the immune-related hub genes. We used a nomogram and receiver operating characteristic (ROC) curve to assess the diagnostic efficacy, which has been validated with GSE55235 and GSE57691. Finally, immune infiltration was developed in AS.

Results: The AS dataset included 5,322 DEGs, while there were 1,439 DEGs and 206 module genes in RA. The intersection of DEGs for AS and crucial genes for RA was 53, which were involved in immunity. After the PPI network and machine learning construction, six hub genes were used for the construction of a nomogram and for diagnostic efficacy assessment, which showed great diagnostic value (area under the curve from 0.723 to 1). Immune infiltration also revealed the disorder of immunocytes.

Conclusion: Six immune-related hub genes (NFIL3, EED, GRK2, MAP3K11, RMI1, and TPST1) were recognized, and the nomogram was developed for AS with RA diagnosis.

1 Introduction

Rheumatoid arthritis (RA) is a systemic autoimmune disease characterized by chronic inflammation that commonly affects individuals aged 50–60 years (1). Patients with RA experience symmetrical joint pain and swelling, which may lead to joint deformity and progressive joint damage (2).

RA patients also have an increased risk of cardiovascular morbidity and mortality (3). Atherosclerosis (AS), the accumulation of a fibrofatty lesion in the artery wall with the infiltration of immunocytes such as macrophages, T cells, and mast cells, is a potential reason for coronary and carotid artery disease (4, 5). Recent evidence suggests that there are similar pathological processes and risk factors in both RA and AS, with chronic inflammation and immune dysfunction being the most significant (5–9).

While the underlying mechanism linking RA and AS is still unknown, it is clear that both conditions involve chronic inflammation and immune infiltration. For example, AS is an inflammatory process that can lead to plaque rupture, thrombosis, and vessel occlusion (10, 11). In patients with RA, immunological processes can occur many years before diagnosis, during the pre-RA phase (12). Furthermore, many pathological processes of the artery wall in AS are reflected in RA synovial inflammation, including the infiltration of macrophages and type 1 T helper cells, which have secondary effects on the artery *via* mediators produced in the synovium (7). Therefore, identifying immune infiltration and associated inflammatory molecules may have early diagnostic efficacy for RA patients with AS, which is significant in avoiding severe cardiovascular consequences.

In this study, we downloaded RA and AS datasets from the Gene Expression Omnibus (GEO) database and screened for differentially expressed genes (DEGs) using Limma. We identified significant module genes *via* weighted co-expression network analysis (WGCNA) and performed functional enrichment analysis. We constructed a protein–protein interaction (PPI) network for the intersection genes and identified candidate genes using machine learning algorithms, including the least absolute shrinkage and selection operator (LASSO) and random forest (RF), and immune cell infiltration analysis. We evaluated the key immune-associated diagnostic genes for AS with RA using nomogram and receiver operating characteristic (ROC) curve assessments. This study is useful in screening immune-related diagnostic biomarkers for AS in RA patients.

2 Materials and methods

2.1 Data collection and data processing

We retrieved four gene expression datasets from the GEO database (https://www.ncbi.nlm.nih.gov/geo/), namely, GSE55457, GSE55235, GSE100927, and GSE57691 (13). The GSE55457 dataset included 11 control samples and 12 RA samples, while GSE55235 included 10 control samples and 10 RA samples. The GSE100927 dataset contained 35 control samples and 69 AS samples, and GSE57691 contained 10 control samples and 9 AS samples. We

normalized the gene expression data using the R package "optparse." The study procedures are summarized in Figure 1.

2.2 Differentially expressed gene screening

We obtained DEGs between RA and the control group with p_{adj} < 0.05 and |$\log_2$Fold change (FC)| > 1.2 in GSE55457, and between AS and the control group with p_{adj} < 0.05 and |$\log_2$FC| > 1.2 in GSE100927. The R software package Limma was used in this analysis. The DEGs were visualized *via* the Sangerbox platform (http://vip.sangerbox.com/).

2.3 Weighted gene correlation network analysis

In this study, we utilized the "WGCNA" package in R software to investigate the association between genes and phenotypes by constructing a gene co-expression network (14). Firstly, we removed 50% of genes with the smallest median absolute deviation (MAD). Secondly, we calculated Pearson's correlation matrices for all pairwise genes and constructed a weighted adjacency matrix using the average linkage method and a weighted correlation coefficient. The "soft" thresholding power (β) was then used to calculate the adjacency, which was converted into a topological overlap matrix (TOM). To group genes with similar expression profiles into modules, we performed average linkage hierarchical clustering based on the TOM-based dissimilarity measure with a minimum gene group size of 50. Finally, we calculated the dissimilarity of module eigengenes, selected a cut line for the module dendrogram, and merged several modules. WGCNA was employed to identify significant modules in AS, and a visualized eigengene network was created.

2.4 Function enrichment analysis

To explore the biological functions of genes, we utilized the "clusterProfile" package in R software (15). First, we conducted Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses, using a *p*-value < 0.05 (16, 17). The results were visualized using the Sangerbox platform (http://vip.sangerbox.com/). We then identified the intersection of DEGs in both AS and the critical module genes of RA, as well as the intersection of DEGs in RA and the critical module genes of RA. We performed GO and KEGG analyses based on these intersections.

2.5 Protein–protein intersection network construction

To investigate the interaction among proteins, pathways, and co-expression, we utilized the STRING database (https://cn.string-db.org/) to construct the protein–protein intersection (PPI) network of the DEGs for AS and the critical module genes (18). Cytoscape software

FIGURE 1
Workflow of the analysis.
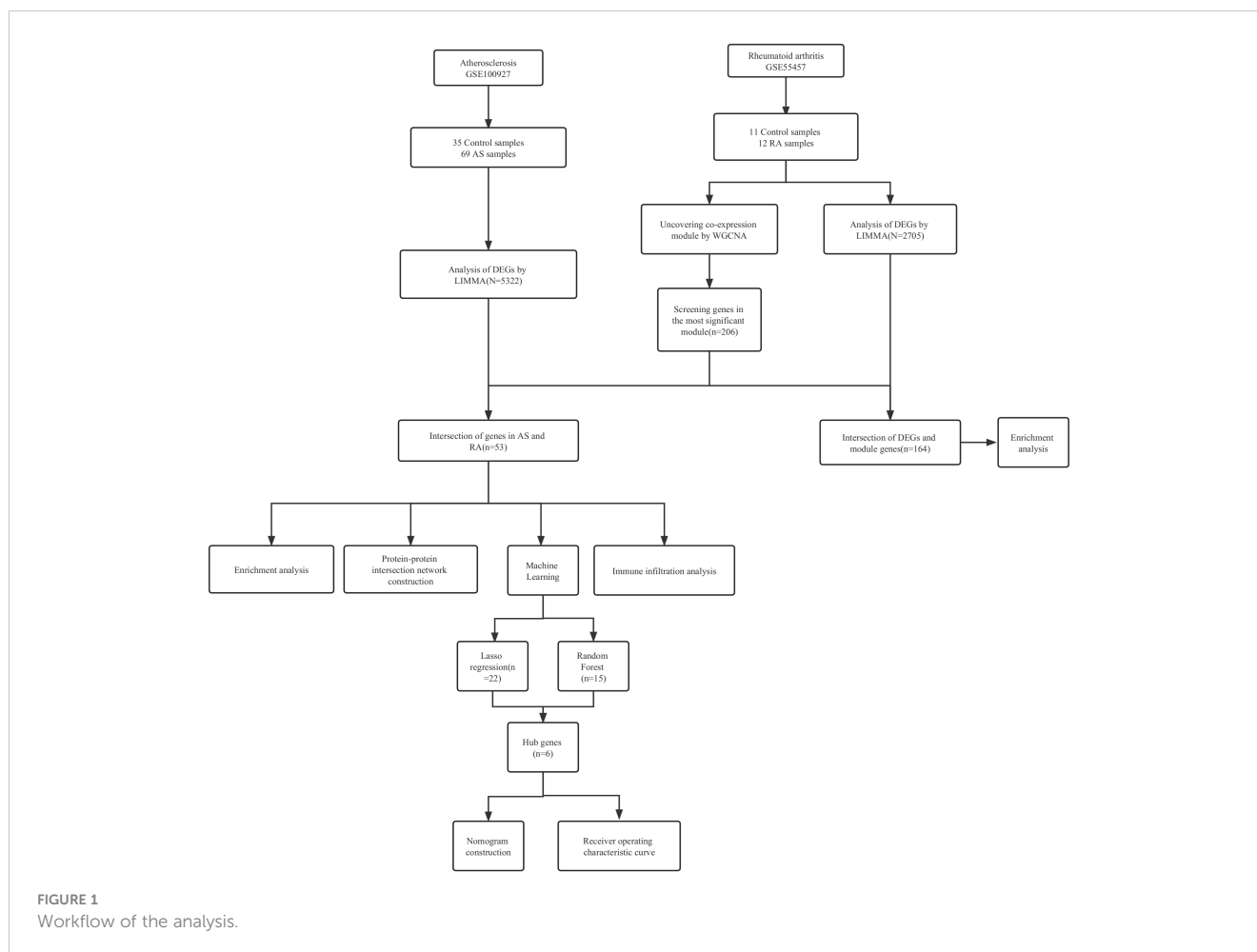
was used to identify the significant interacted genes (19). Only the genes that interacted with each other were chosen for further analysis.

2.6 Machine learning

To further investigate the potential candidate genes for the diagnosis of AS with RA, we performed LASSO and RF analyses. LASSO, a machine learning technique that combines variable selection and regularization, can enhance predictive accuracy (20). On the other hand, RF is a predictive algorithm that does not impose restrictions on variable conditions, making it capable of providing predictions without apparent variations (21). We employed the R software's "glmnet" and "randomforest" packages to conduct LASSO and RF analyses, respectively. The intersection of the two results can serve as the candidate hub genes for diagnosis (22, 23).

2.7 Nomogram construction and receiver operating characteristic evaluation

In order to determine the importance of the candidate genes for the diagnosis of AS with RA, we constructed a nomogram using the "rms" R package. The nomogram consisted of "Points," which indicated the score of the candidate genes, and "Total Points," which showed the

total sum of all gene scores. The nomogram was an important tool for predicting the diagnosis of AS with RA. We further evaluated the prognostic value of the candidate genes and the nomogram by performing ROC analysis. The ROC analysis generated the area under the curve (AUC) and 95% confidence interval (CI), and an AUC value > 0.7 was considered to have great diagnostic efficacy.

2.8 Immune infiltration analysis

To estimate the infiltration of immune cells based on gene expression profiles, we utilized CIBERSORT, an analytical tool. We evaluated the proportion of immune cells in AS and control groups using this platform (24). The bar plot was used to visualize the proportion of various immune cells, while the vioplot was used to compare the proportions of these cells between the AS and control groups. The heatmap with Sangerbox platform was used to depict the association of immunocytes (25).

2.9 Statistical analysis

Statistical analysis was conducted to analyze the data obtained in this study. The ROC curve and AUC were constructed using SPSS

3 Results

3.1 Identification of differentially expressed genes

A total of 2,705 DEGs were identified from the RA combined dataset with a p-value < 0.05 and |log2FC| > 1.2. The volcano plot and heatmap presented in Figures 2A, B, respectively, illustrate the differential expression pattern of these DEGs. Similarly, for AS, a total of 5,322 DEGs were identified using the same cutoff criteria of p-value < 0.05 and |log2FC| > 1.2. Figures 3A, B depict the differential expression pattern of these DEGs for AS.

3.2 Weighted gene co-expression network analysis and critical module identification

We constructed a scale-free co-expression network using the weighted gene co-expression network analysis (WGCNA) to identify the most associated module in RA. A "soft" threshold β of 7 was chosen based on the scale independence and average connectivity (Figures 3C, D). The clustering dendrogram of RA and control was generated, and 26 gene co-expression modules in different colors were obtained with a module merge threshold of 0.25 and a minimum size of 50, as shown in Figures 3E–G. Clinical correlation analysis results showed that the pink module had the highest association ($r = 0.73$, p-value < 0.001) with RA Figure 3H. Thus, we selected the pink module, which consisted of 206 genes, for further analysis. We conducted a correlation analysis between module membership and gene significance, and found a significant positive correlation between them (correlation coefficient = 0.64, p-value < 0.001) Figure 3I. These results indicated that the genes in the pink module were most closely related to RA.
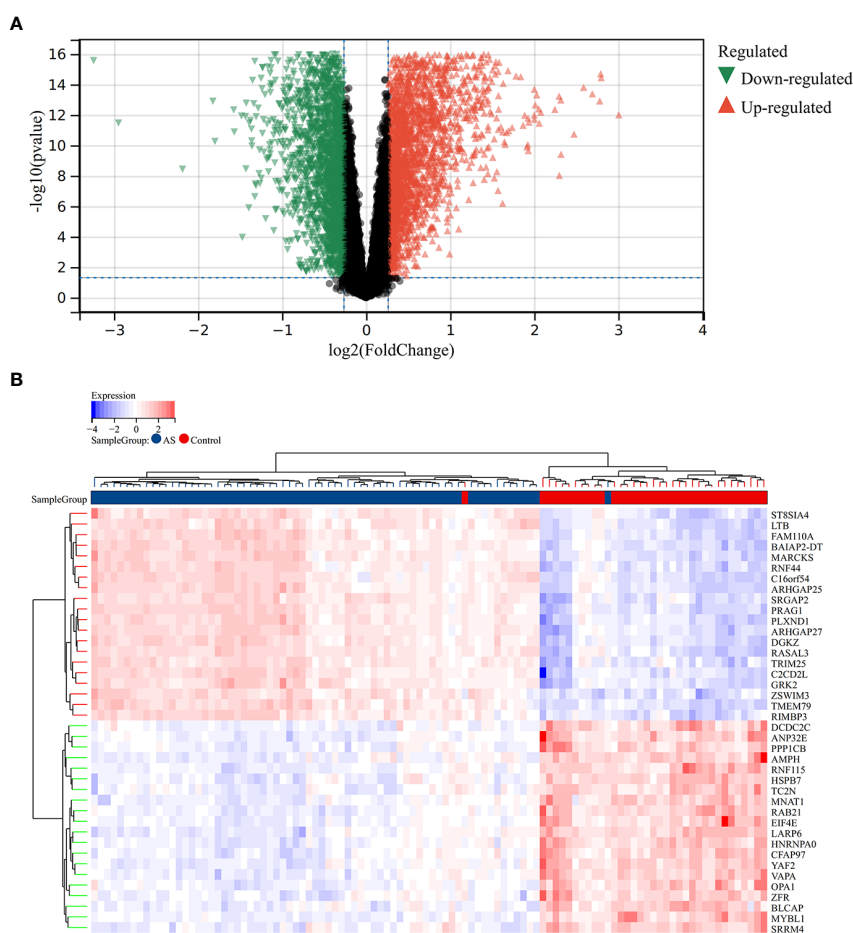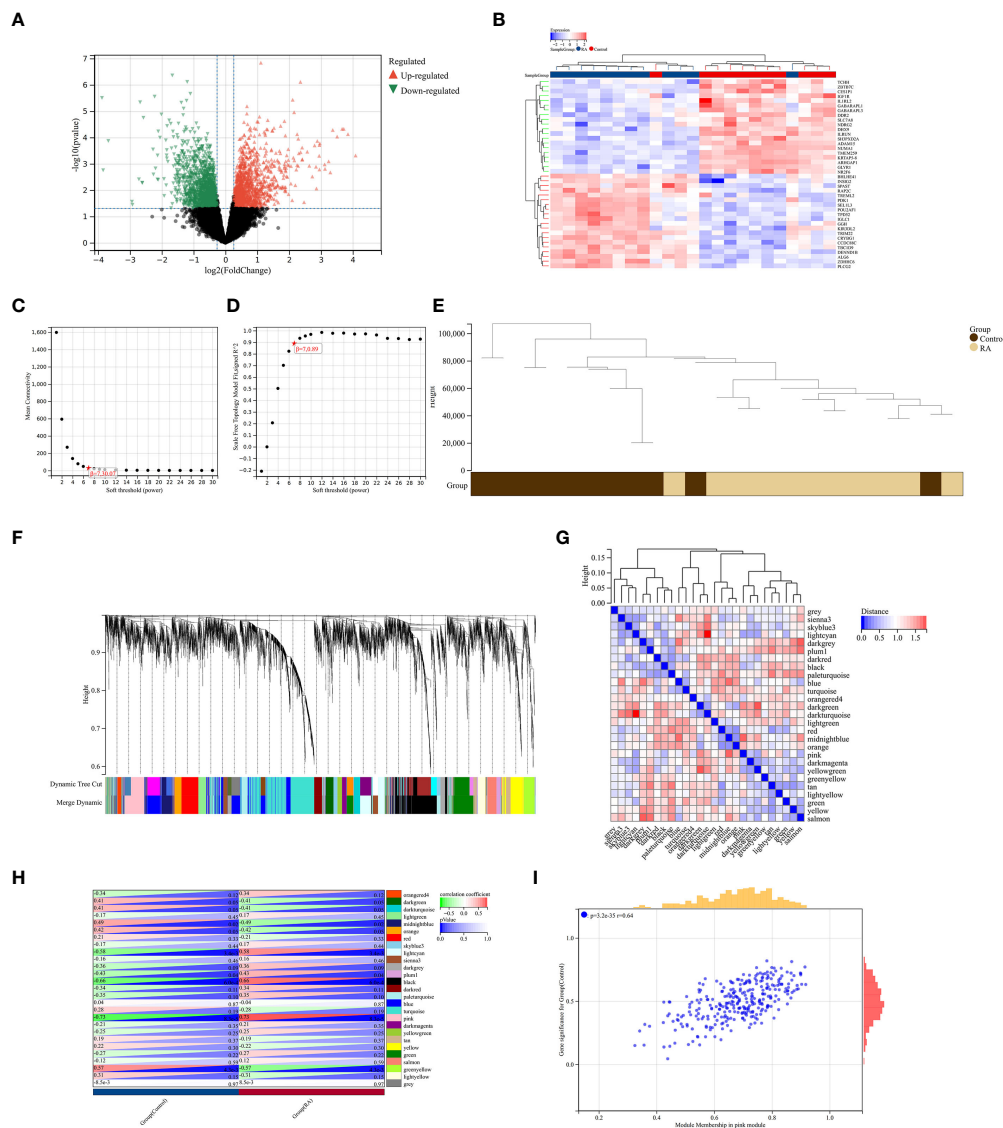

FIGURE 2

Different expression genes between AS and control groups. **(A)** Red and green represent DEGs with significantly higher and lower expression level in AS groups, respectively. **(B)** The heatmap showed the top 20 genes that were significantly expressed in the RA and control groups.

FIGURE 3

Identification of DEGs *via* Limma and WGCNA module genes in RA. **(A)** The volcano plot represents DEGs, of which the red and green triangles refer to significant genes. **(B)** The heatmap shows the top 20 upregulated and downregulated DEGs from the RA dataset, which are shown in red and blue colors. **(C, D)** $\beta = 7$ is chosen as the soft threshold based on the scale independence and average connectivity. **(E)** Clustering dendrogram of the RA and control samples. **(F)** Gene co-expression modules with different colors under the gene tree. **(G)** Heatmap of eigengene adjacency. **(H)** Heatmap of correlation between module genes and RA shows that the pink module has the highest association with RA. For each pair, the top left triangle is colored to represent the correlation coefficient; the bottom right one is colored to indicate the *p*-value. **(I)** Correlation plot between module membership and gene significance of magenta module genes.

3.3 Functional enrichment analysis of RA

To validate the reliable extent of GSE55457, we implemented enrichment analysis for the intersection of genes from Limma and module genes. A total of 164 common genes were obtained, as shown in Figure 4A.

KEGG analysis elucidated that common genes were involved in "p53 signaling pathway" and "Apoptosis", as shown in Figure 4B. The results of GO analysis revealed that common genes were enriched in biological process (BP) terms, including "immune system process", "immune response", and "regulation of immune response", as shown in Figure 4C. For cellular component (CC) ontology, the common genes are involved in "cytosol", "nuclear

part", and "nuclear lumen", as shown in Figure 4D. For molecular function (MF), the results showed that "drug binding" was the most significant term in common genes, as shown in Figure 4E.

The results showed that the common genes for RA were associated with immune response, which were highly related to the pathogenesis of RA.

3.4 Enrichment analysis of AS with RA and screening node genes *via* the protein–protein interaction network

The intersection of the DEGs for AS and the module genes for RA included 53 genes, as seen in Figure 5A. To explore the

FIGURE 4

Function enrichment analysis of the intersection of genes for RA. **(A)** The intersection of DEGs *via* Limma and WGCNA module genes includes 164 genes, which were shown in the Venn diagram. **(B)** KEGG analysis of the intersection of genes. Various significant pathways and associated genes are represented with different colors. **(C–E)** The GO analysis includes biological process, cellular component, and molecular function. The *y*-axis represents GO terms, and the *x*-axis represents gene ratio involved in corresponding GO terms. The size of circles represents gene numbers, and their color refers to *p*-value.

relationship between RA-related genes with the pathogenesis of AS, enrichment analysis was performed based on these genes. The KEGG analysis showed that 53 genes mainly enriched in "NF-kappaB signaling pathway" and "Neurotrophin signaling pathway", which were all closely associated with the immune system, as shown in Figure 5B. GO analysis revealed that genes were involved in "NF-kappaB signaling pathway", "I-kappaB phosphorylation" (BP), "cytosol", "cytoskeleton" (CC), and "transferase activity" (MF), as shown in Figures 5C–E.

A PPI network was constructed, in which 22 genes can interact with each other, as shown in Figure 5F. The node genes were ranked by node numbers in Figure 5G.

3.5 Identification of candidate hub genes *via* machine learning

LASSO regression and RF machine learning algorithms were utilized to identify potential candidate genes associated with the diagnosis of AS with RA. LASSO regression analysis identified 22 genes that were closely associated with the disease (Figures 6A, B). In the RF algorithm, we evaluated the importance of genes based on indicators such as mean decrease accuracy (MDA) and mean decrease gini (MDG) (Figures 6C, D). The AUC and 95% CI of these genes in LASSO regression and the intersection of MDA and MDG in RF machine learning algorithms were calculated, and the ROC curves were plotted. The results showed high accuracy for the LASSO regression (AUC 0.999, CI 0.971–1) and RF machine
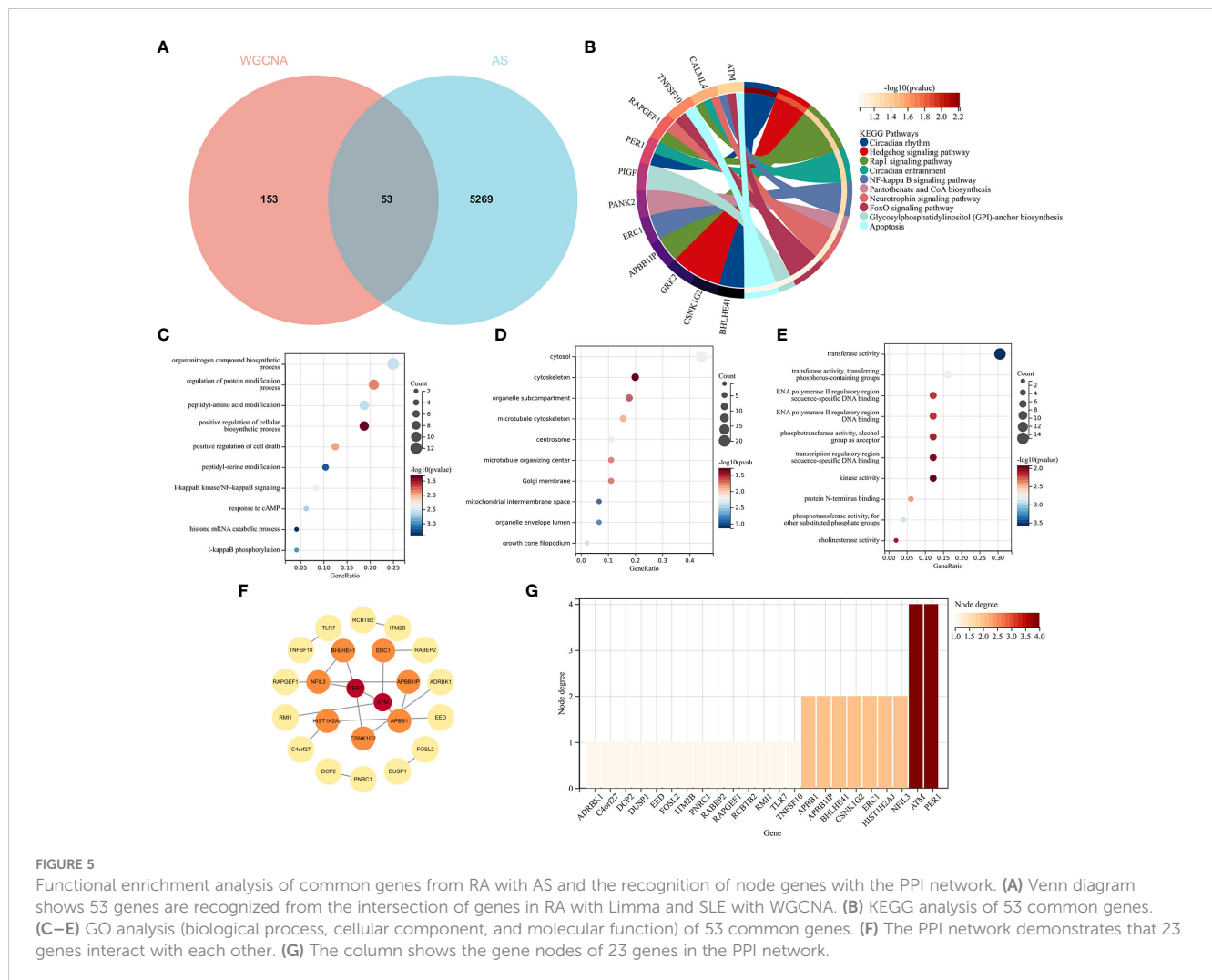
learning algorithms (AUC 0.995, CI 0.971–0.986) (Figures 6E, F). The intersection of the top 15 most important genes from RF and 22 genes from LASSO were visualized in Figure 6E, which identified six genes (NFIL3, EED, GRK2, MAP3K11, RMI1, and TPST1) as key diagnosis genes for the final validation.

3.6 Diagnosis value evaluation

We constructed the nomogram with six key diagnosis genes, as shown in Figure 7B. The AUC and 95% CI of these genes were calculated with the construction of ROC curves to evaluate the diagnostic efficacy as shown in Figure 7A. The results were as follows: NFIL3 (AUC 0.907, CI 0.8515–0.9622), EED (AUC 0.915, CI 0.8582–0.9712), GRK2 (AUC 0.986, CI 0.9669–1), MAP3K11 (AUC 0.954, CI 0.9089–0.9984), RMI1 (AUC 0.953, CI 0.9157–0.9903), TPST1 (AUC 0.815, CI 0.723–0.9076), and nomogram (AUC 0.996, CI 0.9839–1). We validated the model with GSE55235 and GSE57691, as shown in Figure 7A. All genes and nomogram showed a high value of diagnosis for AS with RA.

3.7 Immune infiltration analysis

Because the key diagnosis genes that were correlated with RA can regulate the pathogenesis of AS and be mainly enriched in immunity, the immune infiltration analysis can better explore the effect of immunity in AS. For AS and the control groups, the
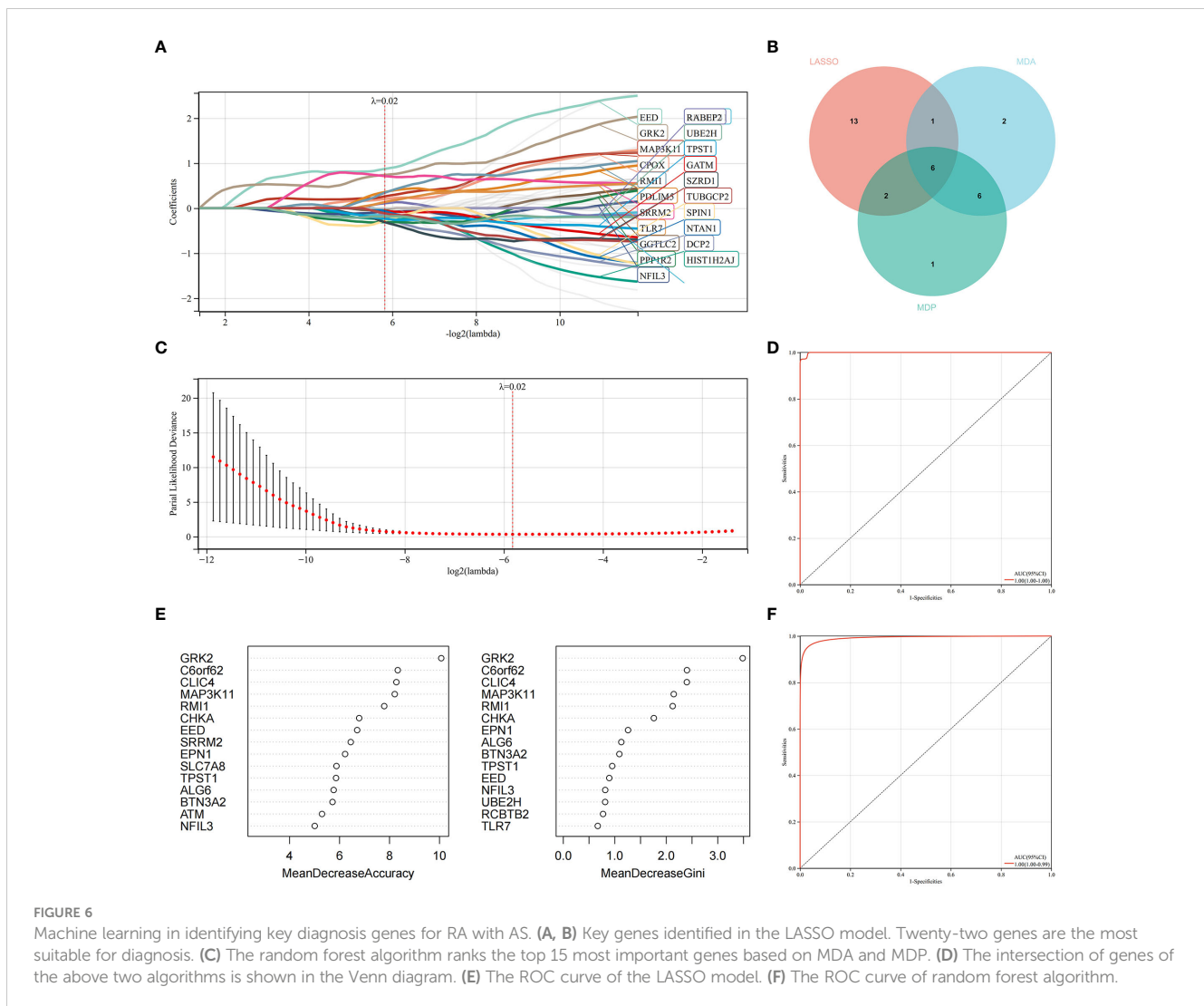
FIGURE 5

Functional enrichment analysis of common genes from RA with AS and the recognition of node genes with the PPI network. **(A)** Venn diagram shows 53 genes are recognized from the intersection of genes in RA with Limma and SLE with WGCNA. **(B)** KEGG analysis of 53 common genes. **(C–E)** GO analysis (biological process, cellular component, and molecular function) of 53 common genes. **(F)** The PPI network demonstrates that 23 genes interact with each other. **(G)** The column shows the gene nodes of 23 genes in the PPI network.

proportion of 22 kinds of immunocytes are shown in Figure 8A. The box plot presented that compared with the control group, naïve B cells, plasma cells, CD4+ naïve T cells, CD4+ memory-activated T cells, follicular helper T cells, activated NK cells, monocytes, M0 macrophage, M1 macrophage, M2 macrophage, resting mast cells, and activated dendritic cells had a lower level in the AS group, while memory B cells, regulatory T cells, gamma delta T cells, and activated mast cells had a high level, as shown in Figure 8B. The correlation of 22 types of immunocytes demonstrates that CD4+ memory resting T cells were positively related to monocytes ($r = 0.55$), monocytes were negatively related to M0 macrophage ($r = -.64$), CD4+ memory resting T cells were negatively related to M0 macrophage, ($r = -0.72$), and all the associations are shown in Figure 8C. In summary, the different level of infiltration of immunocytes in RA patients may serve as a potential treatment target.

4 Discussion

Accumulation of plaque in the artery wall, known as, is a primary cause of cardiovascular diseases and is closely associated with complications of the heart, brain, and kidney (26–28). Due to the difficulty in diagnosing and treating AS, finding an appropriate diagnostic biomarker is crucial to improve the prognosis (29). AS and RA share similar pathological processes, and the mortality rate of AS in RA patients is significantly increasing (30). Therefore, we performed bioinformatics analysis and machine learning methods to construct a nomogram to evaluate the diagnostic efficacy of AS in RA patients. We identified six key immune-related candidate genes (NFIL3, EED, GRK2, MAP3K11, RMI1, and TPST1) and constructed a nomogram.

Nuclear-factor interleukin 3 (NFIL3), also known as E4BP4, is a new biomarker for diagnosing AS in RA patients. NFIL3 exerts a transcriptional repressing function by binding to an activation transcription factor (ATF) DNA consensus sequence site (31). As a crucial transcription factor in the immune system, the expression level of NFIL3 is regulated by cytokines and mainly found in natural killer cells, B lymphocytes, T lymphocytes, and other immune cells (31–33). As a crucial transcription factor in the immune system, the expression level of NFIL3 is regulated by cytokines and mainly found in natural killer cells, B lymphocytes, T lymphocytes, and other immune cells (34–36). Inhibition of NFIL3 expression in CD4 + T cells decreases the level of IL10, worsening autoimmune

FIGURE 6

Machine learning in identifying key diagnosis genes for RA with AS. **(A, B)** Key genes identified in the LASSO model. Twenty-two genes are the most suitable for diagnosis. **(C)** The random forest algorithm ranks the top 15 most important genes based on MDA and MDP. **(D)** The intersection of genes of the above two algorithms is shown in the Venn diagram. **(E)** The ROC curve of the LASSO model. **(F)** The ROC curve of random forest algorithm.

encephalomyelitis (37). NFIL3 promotes the Th2 lineage while inhibiting the Th17 lineage and suppresses the production of IL-12 p40 in macrophages, which is associated with the progression of colitis (37, 38). Additionally, the anti-inflammatory effect of NFIL3 in immunity plays a crucial role in autoimmune diseases. NFIL3 has a high expression level in CD4+ T cells of patients with systemic lupus erythematosus (SLE) and suppresses the activation and self-reactivity of T cells and subsequent autoimmune response by downregulating CD40L (39). T follicular helper cells in patients with SLE also show a high level of NFIL3 but a low level of phosphorylation (40). Furthermore, the deficiency of NFIL3 is associated with juvenile idiopathic arthritis and induces more severe arthritis (41). The significant increase in NFIL3 in patients with RA may be associated with the production of multiple pro-inflammatory cytokines and RA progression (42). However, the association of NFIL3 with AS is still unclear. Due to the pro-inflammatory effect of NFIL3 in patients with RA, and the inflammation being a crucial factor in plaque rupture and stability, we suggest that NFIL3 could be a candidate diagnostic gene for AS in RA patients.

Embryonic ectoderm development (EED) is a nuclear factor and a transcriptional repressor. It is a member of the polycomb repressive complex and is involved in the proliferation and differentiation of lymphocytes as well as embryonic development (43–45). WAIT-1, a protein cloned from EED, interacts with integrins at the plasma membrane and plays a crucial role in immunity (46, 47). The activation of the integrin receptor can recruit EED to the plasma membrane, where it participates in the antigen receptor transduction in T cells (44, 48). EED also interacts with the neutral sphingomyelinase 2, which is involved in inflammation, heart failure, AS, and other biological processes (45, 49, 50). The production of ceramide *via* sphingomyelin hydrolysis is involved in the formation of atherogenic plaques, making the sphingomyelinase an important target in the treatment of AS (50). The production of ceramide *via* sphingomyelin hydrolysis is involved in the formation of atherogenic plaques, making the sphingomyelinase an important target in the treatment of AS.

G protein-coupled receptor (GPCR) kinase 2 (GRK2) is a key node in multiple signaling networks and interacts with various cellular proteins associated with signal transduction. This
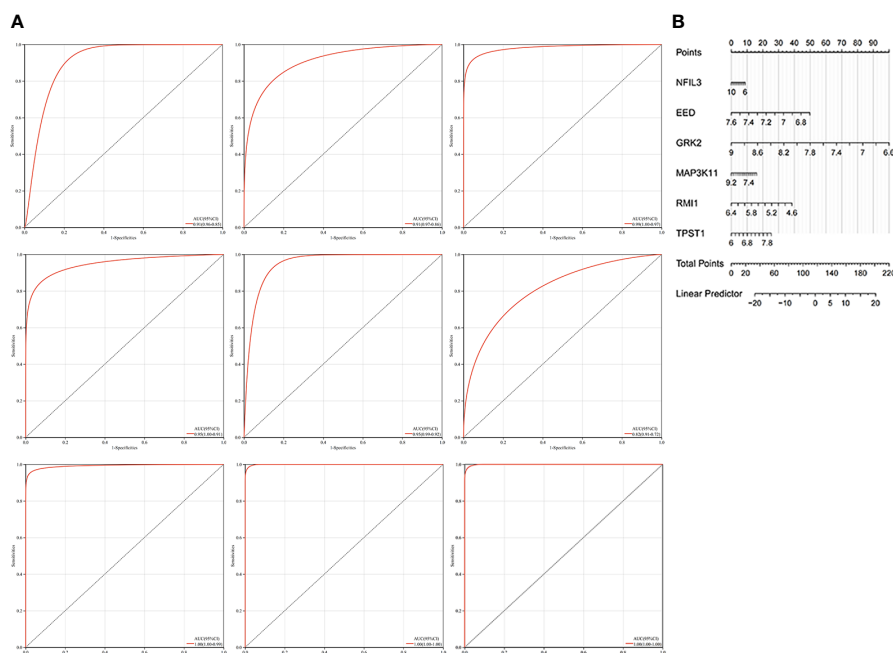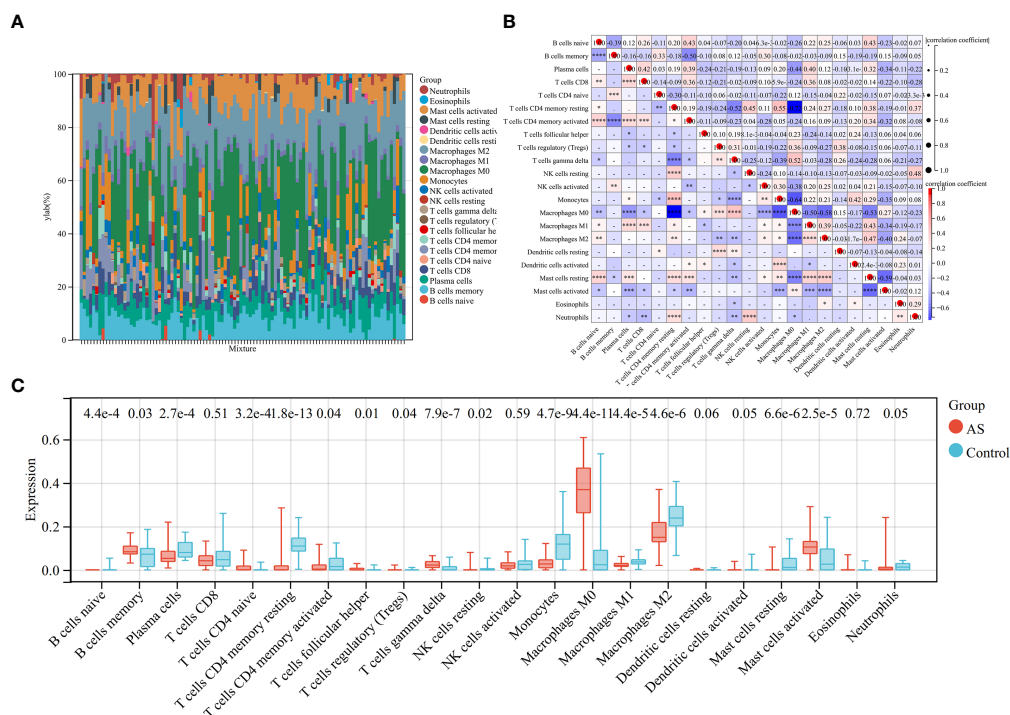
FIGURE 7

Construction of the nomogram and the diagnosis value assessment. **(A)** The ROC curve of each candidate gene (NFIL3, EED, GRK2, MAP3K11, RMI1, and TPST1), nomogram, and the validation in GSE55235 and GSE57691. **(B)** Nomogram for diagnosis RA with AS.

interaction further promotes signal propagation after GPCR activation (51). The signal transduction involves various cells' activation, including endothelial cells. Excessive angiogenesis is an important factor in the development of inflammatory diseases, such as RA (52, 53). A high expression level of GRK2 has been detected in the synovial tissues of RA patients (51). It has been proven that GRK2 participates in the progression of AS. The mouse with a GRK2 deficiency demonstrates defective angiogenesis and

FIGURE 8

Immune infiltration analysis between AS and control. **(A)** The proportion of 22 immunocytes in all samples visualized from the bar plot. **(B)** Comparison of the proportion of 22 kinds of immunocytes between AS and control groups shown in the vioplot. **(C)** Association of 22 immunocyte-type compositions. *p < 0.05; **p < 0.01; ***p < 0.001.

increasing chemokine and adhesion molecules as AS progresses (54). Moreover, GRK2 is a potential upstream kinase for vinculin *via* mediating phosphorylation of vinculin, which further induces the disruption of the VE-cadherin/catenin complex, promoting the generation of atherogenesis (55). In this study, GRK2 is identified as one of the candidate diagnosis biomarkers for AS with RA.

Mitogen-activated protein kinase 11 (MAP3K11) is a potential target for immune treatment due to its expression in T cells and its regulatory role in T-cell activation and cytotoxicity (56). In addition, MAP3K11 is upregulated by mechanical stress and is associated with the differentiation of bone marrow stromal cells (57, 58). MAP3K11 has also been identified as a target for AS, as its inhibition can reduce the expression of key genes in coronary artery disease and the migration of vascular smooth muscle cells (59–61). It can also be used as a diagnosis marker.

RecQ-Medoayed Genome Instability 1 (RMI1) is crucial for maintaining genomic stability and regulates adipocyte hyperplasia to maintain energy stability (62). RMI1 is upregulated by obesity and high-glucose conditions and plays a role in maintaining genome integrity during replicative stress (63, 64).

Protein-tyrosine sulfotransferase 1 (TPST1) catalyzes the sulfuration of tyrosine residues within the acidic motif of polypeptides (65). It has been proven that TPST1 can regulate immune and inflammatory response through catalyzing sulfation (66) involved in regulating immune and inflammatory responses through tyrosine sulfation (67). Additionally, tyrosine sulfation contributes to monocyte recruitment, a major factor in AS development, making drugs inhibiting TPST1 favorable in AS treatment (68, 69). In this study, TPST1 is selected as a candidate diagnosis biomarker.

It has been identified that immune cells and inflammation play a crucial role in the pathogenesis of AS (70). The interactions between immunocytes and the production of pro-inflammatory and anti-inflammatory chemokines have an important influence in the plaque rupture (5, 71, 72). In AS patients and the animal models of AS, it has been observed that circulating monocytes are associated with the size and stage of plaque (73, 74). Monocytes can further differentiate into macrophages, the key component of plaque, and become foam cells after the accumulation (70). Dendritic cells also participate in the adaptive immune response to AS-associated antigens and the formation of foam cells, further promoting the development of AS (75, 76). Furthermore, Th1 cells are the main type of CD4+ T cells in AS, which produce a large number of pro-inflammatory cytokines, while Th2 cells can produce IL-13 and IL-5 to antagonize atherosclerosis (77–79). The expression level of Tregs has decreased with the progression of AS (80, 81). Moreover, B2 cells, which participate in antibody production, dependent on T cells, promote the progression of AS. In our study, naïve B cells, plasma cells, CD4+ naïve T cells, CD4+ memory-activated T cells, follicular helper T cells, activated NK cells, monocytes, M0 macrophage, M1 macrophage, M2 macrophage, resting mast cells, and activated dendritic cells had a lower level in AS patients, while memory B cells, regulatory T cells, gamma delta T cells, and activated mast cells had a high level in AS patients, consistent with previous studies. In summary, the study on immune manifestation and inflammatory cytokines can favor the diagnosis and treatment for AS.

5 Conclusion

In this study, we have successfully identified six immune-related hub genes (NFIL3, EED, GRK2, MAP3K11, RMI1, and TPST1) using bioinformatics analysis and machine learning algorithms. These genes have shown a potential to serve as diagnostic candidate genes for AS in RA patients. Furthermore, our study has also highlighted the immune dysfunction in AS with RA. We have also constructed a nomogram for diagnosing AS with RA, which can aid in clinical decision-making. Overall, our findings may provide new insights into the pathogenesis and diagnosis of AS with RA. Further validation studies are warranted to confirm the clinical relevance of these genes in AS with RA.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

Author contributions

YH, FHL and HW participated in reviewing the articles. FZL wrote the manuscript. All authors contributed to the article and approved the submitted version.

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2023.1126647/full#supplementary-material

References

1. Sparks JA. Rheumatoid arthritis. *Ann Intern Med* (2019) 170:ITC1–ITC16. doi: 10.7326/AITC201901010

2. Burmester GR, Pope JE. Novel treatment strategies in rheumatoid arthritis. *Lancet* (2017) 389:2338–48. doi: 10.1016/S0140-6736(17)31491-5

3. Reiss AB, Silverman A, Khalfan M, Vernice NA, Kasselman LJ, Carsons SE, et al. Accelerated atherosclerosis in rheumatoid arthritis: Mechanisms and treatment. *Curr Pharm Des* (2019) 25(9):969–86. doi: 10.2174/1381612825666190430113212

4. Falk E. Pathogenesis of atherosclerosis. *J Am Coll Cardiol* (2006) 47:C7–12. doi: 10.1016/j.jacc.2005.09.068

5. Hansson GK, Hermansson A. The immune system in atherosclerosis. *Nat Immunol* (2011) 12:204–12. doi: 10.1038/ni.2001

6. Ozbalkan Z, Efe C, Cesur M, Ertek S, Nasiroglu N, Berneis K, et al. An update on the relationships between rheumatoid arthritis and atherosclerosis. *Atherosclerosis* (2010) 212:377–82. doi: 10.1016/j.atherosclerosis.2010.03.035

7. Skeoch S, Bruce IN. Atherosclerosis in rheumatoid arthritis: is it all about inflammation? *Nat Rev Rheumatol* (2015) 11:390–400. doi: 10.1038/nrrheum.2015.40

8. Hannawi S, Hannawi H, Al Salmi I. Cardiovascular disease and subclinical atherosclerosis in rheumatoid arthritis. *Hypertens Res* (2020) 43(9):982–4. doi: 10.1038/s41440-020-0483-4

9. Karpouzas GA, Bui VL, Ronda N, Hollan I, Ormseth SR. Biologics and atherosclerotic cardiovascular risk in rheumatoid arthritis: a review of evidence and mechanistic insights. *Expert Rev Clin Immunol* (2021) 17(4):355–74. doi: 10.1080/1744666X.2021.1899809

10. Libby P. Role of inflammation in atherosclerosis associated with rheumatoid arthritis. *Am J Med* (2008) 121(10 Suppl 1):S21–31. doi: 10.1016/j.amjmed.2008.06.014

11. Everett BM, Pradhan AD, Solomon DH, Paynter N, Macfadyen J, Zaharris E, et al. Rationale and design of the cardiovascular inflammation reduction trial: a test of the inflammatory hypothesis of atherothrombosis. *Am Heart J* (2013) 166(2):199–207.e15. doi: 10.1016/j.ahj.2013.03.018

12. Radu AF, Bungau SG. Management of rheumatoid arthritis: An overview. *Cells* (2021) 10:(11)2587. doi: 10.3390/cells10112857

13. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data sets–update. *Nucleic Acids Res* (2013) 41:D991–995. doi: 10.1093/nar/gks1193

14. Langfelder P, Horvath S. WGCNA: an r package for weighted correlation network analysis. *BMC Bioinf* (2008) 9:559. doi: 10.1186/1471-2105-9-559

15. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an r package for comparing biological themes among gene clusters. *OMICS* (2012) 16:284–7. doi: 10.1089/omi.2011.0118

16. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* (2000) 28(1):27–30. doi: 10.1093/nar/28.1.27

17. The Gene Ontology C. The gene ontology resource: 20 years and still GOing strong. *Nucleic Acids Res* (2019) 47:D330–8. doi: 10.1093/nar/gky1055

18. Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, et al. The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res* (2021) 49: D605–12. doi: 10.1093/nar/gkaa1074

19. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* (2003) 13:2498–504. doi: 10.1101/gr.1239303

20. Yang C, Delcher C, Shenkman E, Ranka S. Machine learning approaches for predicting high cost high need patient expenditures in health care. *BioMed Eng Online* (2018) 17:131. doi: 10.1186/s12938-018-0568-3

21. Ellis K, Kerr J, Godbole S, Lanckriet G, Wing D, Marshall S. A random forest classifier for the prediction of energy expenditure and type of physical activity from wrist and hip accelerometers. *Physiol Meas* (2014) 35:2191–203. doi: 10.1088/0967-3334/35/11/2191

22. Alderden J, Pepper GA, Wilson A, Whitney JD, Richardson S, Butcher R, et al. Predicting pressure injury in critical care patients: A machine-learning model. *Am J Crit Care* (2018) 27:461–8. doi: 10.4037/ajcc2018525

23. Li YR, Meng K, Yang G, Liu BH, Li CQ, Zhang JY, et al. Diagnostic genes and immune infiltration analysis of colorectal cancer determined by LASSO and SVM machine learning methods: a bioinformatics analysis. *J Gastrointest Oncol* (2022) 13:1188–203. doi: 10.21037/jgo-22-536

24. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods* (2015) 12:453–7. doi: 10.1038/nmeth.3337

25. Shen W, Song Z, Zhong X, Huang M, Shen D, Gao P, et al. Sangerbox: A comprehensive, interaction-friendly clinical bioinformatics analysis platform. *iMeta* (2022) 1:e36. doi: 10.1002/imt2.36

26. Cai W, Lang M, Jiang X, Yu Q, Zhou C, Zou W, et al. *Correlation among high salt intake, blood pressure variability, and target organ damage in patients with essential hypertension: Study protocol clinical trial (SPIRIT compliant)* (2020) Medicine (Baltimore) 99(14):e19548. doi: 10.1097/MD.0000000000019548

27. Lusis AJ. Atherosclerosis. *Nature* (2000) 407(6801):233–41. doi: 10.1038/35025203

28. Frostegård J. Immunity, atherosclerosis and cardiovascular disease. *BMC Med* (2013) 11:117. doi: 10.1186/1741-7015-11-117

29. Lu Y, Zhang X, Hu W, Yang Q. The identification of candidate biomarkers and pathways in atherosclerosis by integrated bioinformatics analysis. *Comput Math Methods Med* (2021) 2021:6276480. doi: 10.1155/2021/6276480

30. Adawi M, Firas S, Blum A. Rheumatoid arthritis and atherosclerosis. *Isr Med Assoc J* (2019) 21(7):460–3.

31. Male V, Nisoli I, Gascoyne DM, Brady HJ. E4BP4: an unexpected player in the immune response. *Trends Immunol* (2012) 33:98–102. doi: 10.1016/j.it.2011.10.002

32. Kashiwada M, Levy DM, Mckeag L, Murray K, Schroder AJ, Canfield SM, et al. IL-4-induced transcription factor NFIL3/E4BP4 controls IgE class switching. *Proc Natl Acad Sci U.S.A.* (2010) 107:821–6. doi: 10.1073/pnas.0909235107

33. Yin J, Zhang J, Lu Q. The role of basic leucine zipper transcription factor E4BP4 in the immune system and immune-mediated diseases. *Clin Immunol* (2017) 180:5–10. doi: 10.1016/j.clim.2017.03.013

34. Kamizono S, Duncan GS, Seidel MG, Morimoto A, Hamada K, Grosveld G, et al. Nfil3/E4bp4 is required for the development and maturation of NK cells *in vivo*. *J Exp Med* (2009) 206:2977–86. doi: 10.1084/jem.20092176

35. Kashiwada M, Pham Nl Fau - Pewe LL, Pewe Ll Fau - Harty JT, Harty Jt Fau - Rothman PB, Rothman PB. NFIL3/E4BP4 is a key transcription factor for CD8α$^+$ dendritic cell development. *Blood* (2011) 117(23):6193–7. doi: 10.1182/blood-2010-07-295873

36. Seillet C, Rankin LC, Groom JR, Mielke LA, Tellier J, Chopin M, et al. Nfil3 is required for the development of all innate lymphoid cell subsets. *J Exp Med* (2014) 211:1733–40. doi: 10.1084/jem.20140145

37. Motomura Y, Kitamura H, Hijikata A, Matsunaga Y, Matsumoto K, Inoue H, et al. The transcription factor E4BP4 regulates the production of IL-10 and IL-13 in CD4+ T cells. *Nat Immunol* (2011) 12:450–9. doi: 10.1038/ni.2020

38. Yu X, Rollins D, Ruhn KA, Stubblefield JJ, Green CB, Kashiwada M, et al. TH17 cell differentiation is regulated by the circadian clock. *Science* (2013) 342(6159):727–30. doi: 10.1126/science.1243884

39. Zhao M, Liu Q, Liang G, Wang L, Luo S, Tang Q, et al. E4BP4 overexpression: a protective mechanism in CD4+ T cells from SLE patients. *J Autoimmun* (2013) 41:152–60. doi: 10.1016/j.jaut.2013.01.004

40. Wang Z, Zhao M, Yin J, Liu L, Hu L, Huang Y, et al. E4BP4-mediated inhibition of T follicular helper cell differentiation is compromised in autoimmune diseases. *J Clin Invest* (2020) 130:3717–33. doi: 10.1172/JCI129018

41. Schlenner S, Pasciuto E, Lagou V, Burton O, Prezzemolo T, Junius S, et al. NFIL3 mutations alter immune homeostasis and sensitise for arthritis pathology. *Ann Rheum Dis* (2019) 78:342–9. doi: 10.1136/annrheumdis-2018-213764

42. Du J, Zheng L, Chen S, Wang N, Pu X, Yu D, et al. NFIL3 and its immunoregulatory role in rheumatoid arthritis patients. *Front Immunol* (2022) 13:950144. doi: 10.3389/fimmu.2022.950144

43. Morin-Kensicki EM, Faust C, Lamantia C, Magnuson T, Magnuson T. Cell and tissue requirements for the gene eed during mouse gastrulation and organogenesis. *Genesis* (2001) 31(4):142–6. doi: 10.1002/gene.10017

44. Witte V, Laffert B, Rosorius O, Lischka P, Blume K, Galler G, et al. HIV-1 nef mimics an integrin receptor signal that recruits the polycomb group protein eed to the plasma membrane. *Mol Cell* (2004) 13(2):179–90. doi: 10.1016/S1097-2765(04)00004-8

45. Philipp S, Puchert M, Adam-Klages S, Tchikov V, Winoto-Morbach S, Mathieu S, et al. The polycomb group protein EED couples TNF receptor 1 to neutral sphingomyelinase. *Proc Natl Acad Sci U.S.A.* (2010) 107:1112–7. doi: 10.1073/pnas.0908486107

46. Rietzler M, Bittner M, Kolanus W, Schuster A, Holzmann B. The human WD repeat protein WAIT-1 specifically interacts with the cytoplasmic tails of beta7-integrins. *J Biol Chem* (1998) 273:27459–66. doi: 10.1074/jbc.273.42.27459

47. Du D, Xu D, Zhu L, Stazi G, Zwergel C, Liu Y, et al. Structure-guided development of small-molecule PRC2 inhibitors targeting EZH2-EED interaction. *J Med Chem* (2021) 64:8194–207. doi: 10.1021/acs.jmedchem.0c02261

48. Su IH, Dobenecker MW, Dickinson E, Oser M, Basavaraj A, Marqueron R, et al. Polycomb group protein ezh2 controls actin polymerization and cell signaling. *Cell* (2005) 121:425–36. doi: 10.1016/j.cell.2005.02.029

49. Clarke CJ, Hannun YA. Neutral sphingomyelinases and nSMase2: bridging the gaps. *Biochim Biophys Acta* (2006) 1758:1893–901. doi: 10.1016/j.bbamem.2006.06.025

50. Pavoine C, Pecker F. Sphingomyelinases: their regulation and roles in cardiovascular pathophysiology. *Cardiovasc Res* (2009) 82:175–83. doi: 10.1093/cvr/cvp030

51. Kuai J, Han C, Wei W. Potential regulatory roles of GRK2 in endothelial cell activity and pathological angiogenesis. *Front Immunol* (2021) 12:698424. doi: 10.3389/fimmu.2021.698424

52. Elshabrawy HA, Chen Z, Volin MV, Ravella S, Virupannavar S, Shahrara S. The pathogenic role of angiogenesis in rheumatoid arthritis. *Angiogenesis* (2015) 18:433–48. doi: 10.1007/s10456-015-9477-2

53. Fearon U, Canavan M, Biniecka M, Veale DJ. Hypoxia, mitochondrial dysfunction and synovial invasiveness in rheumatoid arthritis. *Nat Rev Rheumatol* (2016) 12:385–97. doi: 10.1038/nrrheum.2016.69

54. Gambardella J, Sorriento D, Bova M, Rusciano M, Loffredo S, Wang X, et al. Role of endothelial G protein-coupled receptor kinase 2 in angioedema. *Hypertension* (2020) 76:1625–36. doi: 10.1161/HYPERTENSIONAHA.120.15130

55. Shih YT, Wei SY, Chen JH, Wang WL, Wu HY, Wang MC, et al. Vinculin phosphorylation impairs vascular endothelial junctions promoting atherosclerosis. *Eur Heart J* (2022) 44(4):304–18. doi: 10.1093/eurheartj/ehac647

56. Kumar S, Singh SK, Viswakarma N, Sondarva G, Nair RS, Sethupathi P, et al. Mixed lineage kinase 3 inhibition induces T cell activation and cytotoxicity. *Proc Natl Acad Sci U.S.A.* (2020) 117:7961–70. doi: 10.1073/pnas.1921325117

57. Yu W, Qian C, Jiang X, Zhang F, Weng W. Mechanisms of stem cell osteogenic differentiation on TiO2 nanotubes. *Colloids Surf B Biointerf.* (2015) 136:779–85. doi: 10.1016/j.colsurfb.2015.10.019

58. Knackmuss U, Lindner SE, Aneichyk T, Kotkamp B, Knust Z, Villunger A, et al. MAP3K11 is a tumor suppressor targeted by the oncomiR miR-125b in early b cells. *Cell Death Differ* (2016) 23:242–52. doi: 10.1038/cdd.2015.87

59. Ridker PM, Everett BM, Thuren T, Macfadyen JG, Chang WH, Ballantyne C, et al. Antiinflammatory therapy with canakinumab for atherosclerotic disease. *N Engl J Med* (2017) 377:1119–31. doi: 10.1056/NEJMoa1707914

60. Georgakis MK, Gill D, Rannikmae K, Traylor M, Anderson CD, Lee JM, et al. Genetically determined levels of circulating cytokines and risk of stroke. *Circulation* (2019) 139:256–68. doi: 10.1161/CIRCULATIONAHA.118.035905

61. Chignon A, Mathieu S, Rufiange A, Argaud D, Voisine P, Bosse Y, et al. Enhancer promoter interactome and mendelian randomization identify network of druggable vascular genes in coronary artery disease. *Hum Genomics* (2022) 16:8. doi: 10.1186/s40246-022-00381-4

62. Lai MS, Seki M, Ui A, Enomoto T. Rmi1, a member of the Sgs1-Top3 complex in budding yeast, contributes to sister chromatid cohesion. *EMBO Rep* (2007) 8:685–90. doi: 10.1038/sj.embor.7401000

63. Suwa A, Kurama T, Shimokawa T. Adipocyte hyperplasia and RMI1 in the treatment of obesity. *FEBS J* (2011) 278:565–9. doi: 10.1111/j.1742-4658.2010.07980.x

64. Xu C, Fang L, Kong Y, Xiao C, Yang M, Du LQ, et al. Knockdown of RMI1 impairs DNA repair under DNA replication stress. *Biochem Biophys Res Commun* (2017) 494:158–64. doi: 10.1016/j.bbrc.2017.10.062

65. Ippel JH, De Haas CJ, Bunschoten A, Van Strijp JA, Kruijtzer JA, Liskamp RM, et al. Structure of the tyrosine-sulfated C5a receptor n terminus in complex with chemotaxis inhibitory protein of staphylococcus aureus. *J Biol Chem* (2009) 284:12363–72. doi: 10.1074/jbc.M808179200

66. Wang E, Wang Y, Zhou S, Xia X, Han R, Fei G, et al. Identification of three hub genes related to the prognosis of idiopathic pulmonary fibrosis using bioinformatics analysis. *Int J Med Sci* (2022) 19:1417–29. doi: 10.7150/ijms.73305

67. Nakamura N, Shimaoka Y, Tougan T, Onda H, Okuzaki D, Zhao H, et al. Isolation and expression profiling of genes upregulated in bone marrow-derived mononuclear cells of rheumatoid arthritis patients. *DNA Res* (2006) 13:169–83. doi: 10.1093/dnares/dsl006

68. Westmuckett AD, Moore KL. Lack of tyrosylprotein sulfotransferase activity in hematopoietic cells drastically attenuates atherosclerosis in ldlr-/- mice. *Arterioscler Thromb Vasc Biol* (2009) 29(11):1730–6. doi: 10.1161/ATVBAHA.109.192963

69. Westmuckett AD, Thacker KM, Moore KL. Tyrosine sulfation of native mouse psgl-1 is required for optimal leukocyte rolling on p-selectin *in vivo*. *PloS One* (2011) 6: e20406. doi: 10.1371/journal.pone.0020406

70. Schaftenaar F, Frodermann V, Kuiper J, Lutgens E. Atherosclerosis: the interplay between lipids and immune cells. *Curr Opin Lipidol* (2016) 27:209–15. doi: 10.1097/MOL.0000000000000302

71. Legein B, Temmerman L, Biessen EA, Lutgens E. Inflammation and immune system interactions in atherosclerosis. *Cell Mol Life Sci* (2013) 70:3847–69. doi: 10.1007/s00018-013-1289-1

72. Libby P, Lichtman AH, Hansson GK. Immune effector mechanisms implicated in atherosclerosis: from mice to humans. *Immunity* (2013) 38:1092–104. doi: 10.1016/j.immuni.2013.06.009

73. Tacke F, Alvarez D, Kaplan TJ, Jakubzick C, Spanbroek R, Llodra J, et al. Monocyte subsets differentially employ CCR2, CCR5, and CX3CR1 to accumulate within atherosclerotic plaques. *J Clin Invest* (2007) 117:185–94. doi: 10.1172/JCI28549

74. Rogacev KS, Seiler S, Zawada AM, Reichart B, Herath E, Roth D, et al. CD14+ +CD16+ monocytes and cardiovascular outcome in patients with chronic kidney disease. *Eur Heart J* (2011) 32:84–92. doi: 10.1093/eurheartj/ehq371

75. Paulson KE, Zhu SN, Chen M, Nurmohamed S, Jongstra-Bilen J, Cybulsky MI. Resident intimal dendritic cells accumulate lipid and contribute to the initiation of atherosclerosis. *Circ Res* (2010) 106:383–90. doi: 10.1161/CIRCRESAHA.109.210781

76. Subramanian M, Tabas I. Dendritic cells in atherosclerosis. *Semin Immunopathol* (2014) 36:93–102. doi: 10.1007/s00281-013-0400-x

77. Laurat E, Poirier B, Tupin E, Caligiuri G, Hansson GK, Bariéty J, et al. *In vivo* downregulation of T helper cell 1 immune responses reduces atherogenesis in apolipoprotein e-knockout mice. *Circulation* (2001) 104(2):197–202. doi: 10.1161/01.CIR.104.2.197

78. Buono C, Binder CJ, Stavrakis G, Stavrakis G, Witztum JL, Glimcher LH, et al. T-Bet deficiency reduces atherosclerosis and alters plaque antigen-specific immune responses. *Proc Natl Acad Sci U.S.A.* (2005) 102(5):1596–601. doi: 10.1073/pnas.0409015102

79. Tsiantoulas D, Diehl CJ, Witztum JL, Binder CJ. B cells and humoral immunity in atherosclerosis. *Circ Res* (2014) 114:1743–56. doi: 10.1161/CIRCRESAHA.113.301145

80. Mor A, Planer D, Luboshits G, Afek A, Metzger S, Chajek-Shaul T, et al. Role of naturally occurring CD4+ CD25+ regulatory T cells in experimental atherosclerosis. *Arterioscler Thromb Vasc Biol* (2007) 27:893–900. doi: 10.1161/01.ATV.0000259365.31469.89

81. Maganto-Garcia E, Tarrio ML, Grabie N, Bu DX, Lichtman AH. Dynamic changes in regulatory T cells are linked to levels of diet-induced hypercholesterolemia. *Circulation* (2011) 124:185–95. doi: 10.1161/CIRCULATIONAHA.110.006411