European Review for Medical and Pharmacological Sciences 2023; 27: 7470-7484

Identification of diagnostic biomarkers and immuno-infiltration analysis for rheumatoid arthritis based on biological information and WGCNA

Y. ZHANG[1], Q.-Y. YANG[1], L.-S. YANG[2], J. ZHANG[1], K. ZHANG[1], Y. LIU[1], X.-D. XU[1], X.-Y. YANG[1], J. WANG[3], B. YAN[2]

[1]First Clinical College, Shandong University of Traditional Chinese Medicine, Jinan, Shandong, China
[2]Department of Orthopedics, Affiliated Hospital of Shandong University of Traditional Chinese Medicine, Jinan, Shandong, China
[3]Department of Acupuncture, Affiliated Hospital of Shandong University of Traditional Chinese Medicine, Jinan, Shandong, China

Abstract. – OBJECTIVE: Rheumatoid arthritis (RA), as an autoimmune disease, poses a huge social and economic burden worldwide. Although the diagnosis of RA has been gradually improved, there is still a need to discover accurate and rapid biomarkers for diagnosis and therapy with a precise understanding of the disease. This study aimed to screen diagnostic biomarkers and analyze immune infiltration in RA based on weighted gene co-expression network analysis (WGCNA).

MATERIALS AND METHODS: Firstly, we screened the experimental and validation sets associated with RA from the GEO database. Crossover genes were obtained using differential genes (DEGs) and key modules in WGCNA. Subsequently, the crossover genes were constructed into protein-protein interaction (PPI) networks and screened to obtain hub genes. The receiver operating characteristic (ROC) curve assessment was performed to identify diagnostic biomarkers. In addition, we used the Cibersort algorithm for immuno-infiltration analysis and the DGidb database to search for drugs associated with diagnostic biomarkers.

RESULTS: In the end, 377 DEGs were identified, and the enrichment analysis revealed significant associations with the immune system. Blue modules in the WGCNA analysis were positively associated with the disease and were identified as key modules. ROC curves evaluated the four hub genes, which significantly differentiated RA from healthy controls and could be used as diagnostic biomarkers. In further analysis, we found that RA is closely related to immunity, and the search identified multiple drugs that hold promise for treating RA.

CONCLUSIONS: *BCL2A1, PTGS2, FAS*, and *LY96* may be used as diagnostic biomarkers, which is significant for diagnosing and treating RA.

Key Words:

Rheumatoid arthritis, Weighted gene co-expression network analysis, Biological information, Diagnostic biomarker, Immuno-infiltration analysis.

Introduction

Rheumatoid arthritis (RA) is a systemic autoimmune disease that causes damage to bones and joints[1]. It is a multifactorial disease primarily caused by the interaction between genetic and environmental factors[2]. The incidence of RA is widely variable worldwide, with rates as high as 0.5-1% in Northern Europe and North America. Although studies[3] have found that the prevalence and incidence of RA have decreased in recent decades, the expected survival of RA patients is reduced by 3-10 years due to the association with increased mortality. As it continues to cause an increasing global burden and tends to increase year by year in the future, RA remains a disease of worldwide concern[4]. Early diagnosis and treatment with urgency will help reduce the burden of RA. However, the pathogenesis of RA has still not been fully discovered. So, in order to create an effective treatment strategy, it is essential to investigate the underlying mechanisms of RA.

As a chronic progressive inflammatory disease, RA is initially symptomless and progresses to a symmetrical polyarthritis of large and small joints, which may eventually damage joints and peri-joint structures and systemic inflammatory consequences[5]. For this reason, it is necessary to start the best treatments as early as possible to

Corresponding Authors: Bing Yan, MD; e-mail: yanb007@126.com; Jian Wang, MD; e-mail: wangjianj7@126.com

avoid a severe impact. Various disease biomarkers have been proposed to achieve the aim of timely diagnosis. Anti-cyclic citrullinated peptide (anti-CCP) antibody detection may provide a way to identify the risk of RA and potentially delay or prevent its onset when there are no specific symptoms initially. Nevertheless, further research is needed to assess and reduce this detection's potential harms and economics[6]. The blood marker Citrullinated vimentin (VICM) is probably superior to anti-CCP antibody detection for diagnosis and monitoring of treatment but requires verification and validation in prospective studies[7]. Therefore, identifying relevant diagnostic biomarkers for RA is both a current requirement and a challenge for the future, which is significant for the subsequent treatment of RA.

In this study, we downloaded the RA datasets based on the Gene Expression Omnibus (GEO), screened for differentially expressed genes (DEGs), and carried out gene ontology (GO) analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis. In the meantime, weighted Gene Co-expression Network Analysis (WGCNA) identified important key module genes. Then a protein-protein interaction (PPI) network was constructed based on the crossover genes obtained, and the hub genes were screened by the CytoHubba plugin in Cytoscape Tool. Lastly, we used receiver operating characteristic (ROC) curves to evaluate the screened hub genes. The study will contribute to the identification of relevant diagnostic biomarkers for RA patients. In addition, we performed an immune infiltration analysis of RA and explored the interaction between the hub gene and drugs.

Materials and Methods

Data Download and Preprocessing

Information was obtained from the extracted dataset of RA patients in the GEO database (available at: https://www.ncbi.nlm.nih.gov/geo/). The keywords "Rheumatoid arthritis" and "Homo sapiens" were used as selection criteria for the search. The GEO dataset was required to meet the following conditions: (1) The dataset must contain both RA and normal specimens; (2) Each sample is assigned to a group label; (3) The platform type is limited to "microarray"; (4) The number of samples in the dataset is greater than 15. In the end, the dataset GSE93272 with platform GPL570 was selected as the experimental set, including 43 healthy controls and 232 RA patients, and the dataset GSE77298 with platform GPL570 was selected as the verification set, including 7 healthy controls and 16 RA patients, details of the dataset are shown in Table I. Subsequently, the two raw datasets were read and preprocessed by R software (available at: https://www.r-project.org), and the preprocessed gene sets were subjected to the subsequent analysis. The study procedures are summarized in Figure 1.

Identification of Differentially Expressed Genes

To discover the differences in gene expression between RA and healthy controls, we performed a DEG screen using the "limma" R package (available at: http://bioconductor.org/). The adj. p-value<0.05 and $|\log_2 FC| \geq 0.58496$ were screening criteria for further analysis. Also, the "ggplot2" R package and the "heatmap" R package were used to create the relevant volcano plots and expression heatmaps to visualize the DEGs better.

GO and KEGG Enrichment Analysis of DEGs

GO is a database created by the Gene Ontology Consortium which aims to qualify and describe gene and protein functions. It contains three parts: Biological Process (BP), Cellular Component (CC), and Molecular Function (MF). KEGG aims to give functional meaning to genes and genomes at the molecular and higher levels[8]. Two enrichment analysis methods are extremely important components of biological information, and we used the SangerBox tool (version 3.0, Hangzhou

Table I. Detailed statistics of the Datasets.

Dataset	Platform	Manufacturer	Group		Tissue type
			Normal	RA	
GSE93272	GPL570	[HGU133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array	43	232	Blood
GSE77298	GPL570	[HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array	7	16	Synovial membrane

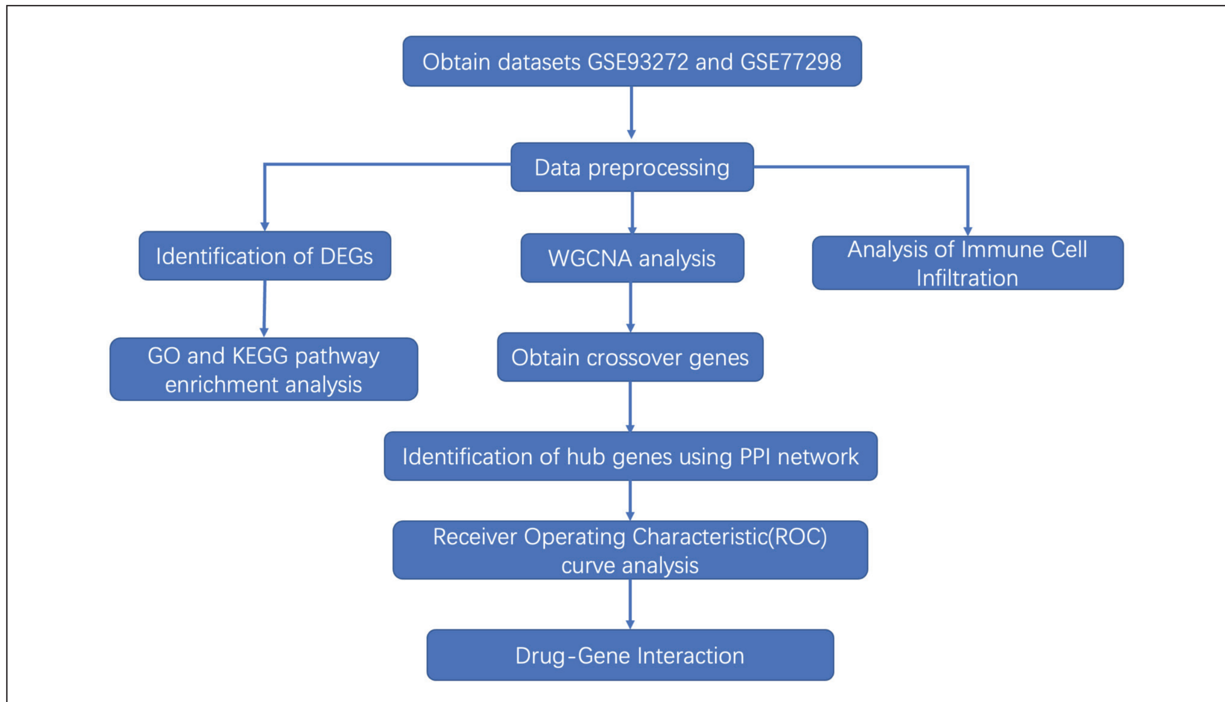

Figure 1. Detailed flow chart of the study.

Mugu Technology Co., Ltd, Zhejiang, China) to analyze the biological function of DEGs online. $p<0.05$ and a false discovery rate (FDR) <0.1 were considered to be significantly different. Finally, the top 10 results of each item of the enrichment analysis are visualized separately.

WGCNA Construction and Key Module Screening

The top 10,000 genes in the experimental set are initially screened for WGCNA after processing. We built co-expression networks with the "WGCNA" R package. To check the study's accuracy, we conducted a sample clustering analysis to verify the association of all data in the experimental set. To ensure that gene interactions follow a scale-free distribution, a study called soft threshold selection analysis is used with soft thresholds which range from 1 to 20. Then the soft threshold is determined based on the lowest value near the scale-free network. Following the best soft thresholds, the relationship matrix is converted into an adjacency matrix and a topological overlap matrix (TOM). Then the dynamic tree-cutting algorithm was applied to identify modules by hierarchical clustering with no less than 30 genes in each module. Later, we calculated the correlation coefficients between expression profiles and groups, as well as the similarity of module signature genes (MEs) in the module tree, and merged some modules (MEs<0.2) to obtain the final module. Ultimately, we calculated the correlation between different modules and traits to obtain the Key Module.

Obtaining Crossover Genes

Using the SangerBox tool to obtain the crossover genes from the DEGs and the key module genes from WGCNA analysis, we visualized them with the Venn diagram.

Construction of PPI Network and Identification of Hub Genes

The molecular network is based on text mining, experiments, databases, co-expression, neighborhoods, gene fusion, and symbiosis. PPI and molecular interaction networks are predicted and visualized by the STRING online database (version 11.5, available at: http://string-db.org) and Cytoscape software (version 3.9.1, available at: https://cytoscape.org/) platform. We input the crossover genes into the STRING database and combined them with Cytoscape software for PPI production. Then we used the 12 methods of calculating gene scores [Maximal clique centrality (MCC), Density of Maximum Neighborhood Component (DMNC), Maximum neighborhood component (MNC), Degree, Edge Percolated component (EPC), BottleNeck,

EcCentricity, Closeness, Radiality, Betweenness, Stress, ClusteringCoefficient] in the Cytoscape software CytoHubba plugin to select the 10 highest scoring genes in each of the 12 groups. By calculating the frequency of occurrence of each gene, we obtained the top 10 genes as candidate diagnostic biomarkers.

Receiver Operating Characteristic (ROC) Curve Analysis

We performed an expression analysis of 10 candidate biomarkers on the verification set GSE77298 to verify the predictor value of candidate diagnostic biomarkers. ROC curves are also plotted using the "pROC" R package, and the area under the curve (AUC) is calculated to assess the predictive power of the identified hub genes. The AUC>0.8 and $p<0.05$ indicate that these biomarkers highly predict RA diagnosis.

Analysis of Immune Cell Infiltration

The Cibersort algorithm is a method of deconvolution of the expression matrix of 22 human immune cells using the linear support vector regression principle, which can be used to evaluate the infiltration of immune cells in the gene expression profile between RA and healthy controls. We used the Cibersort algorithm in the SangerBox tool to perform immuno-infiltration analysis. The bar chart was used to visualize the proportion of the 22 immune cells in the different samples. The violin chart compares the differences in the infiltration levels of immune cells between RA and healthy controls. In addition, inter-immune cell correlations and diagnostic biomarker-immune cell correlations are visualized using Pearson's coefficients.

Drug-Gene Interaction

The screening of hub genes is significant for diagnosis and guides the treatment of the disease in most cases. We used the DGidb database (Washington University School of Medicine, Seattle, Washington, WA, USA) to analyze drug-gene interactions for diagnostic biomarkers. Drugs with an interaction type of inhibitor and an interaction score>3 were selected as potential drugs for the treatment of RA, which will provide a reference for future studies.

Statistical Analysis

R software (version 4.2.2, The R Foundation for Statistical Computing, Vienna, Austria) and the SangerBox tool were used for statistical analysis and data processing. Detailed statistical methods employed in processing transcriptome data are described in the Materials and Methods section. In this study, we considered adj. p-value<0.05 statistically significant in the differential genes analysis and $p<0.05$ statistically significant for the remaining parts involving statistical analysis.

Results

Data Preprocessing

After reading the experimental set GSE93272, 152 samples treated with drugs (such as methotrexate, infliximab and golimumab) that could impact this study were removed by the selection, and 43 healthy control samples and 80 RA samples without drug intervention were reserved. The validation set GSE77298 was read to obtain 7 healthy control and 16 RA samples. Firstly, we performed log2 processing on the GSE77298 data and then normalized the data using the "limma" R package. Then we plotted box plots of the sample gene distribution levels of the two datasets together with principal component analysis (PCA) plots, which showed that the datasets were well-formed (Figure 2A-D). Finally, the two gene sets were converted to gene symbols for each probe's name according to the GPL570 platform, and those probes with blank gene symbols were removed. When multiple probes corresponded to the same gene, the average expression value of the multiple probes was used as the expression value of the gene, and the obtained pre-processed data set was taken for future analysis.

Screening for DEGS

The "limma" R package was used to perform differential expression on the pretreated experimental set GSE93272 (43 healthy controls and 80 RAs). Compared to healthy controls, 377 DEGs were obtained in RA samples, including 359 DEGs with up-regulated expression and 18 DEGs with decreased expression (Figure 3A). The heatmap (Figure 3B) showed the top 10 up-regulated genes and the top 10 down-regulated genes ranked by $|\log_2 FC|$.

GO and KEGG Enrichment Analysis of DEGs

To explore the potential biological mechanisms of RA progression, we used multiple enrichment methods to explore DEG information. GO enrichment involves three main items. Among these, biological process (BP) was mainly enriched in responses to biotic stimuli, other organisms,

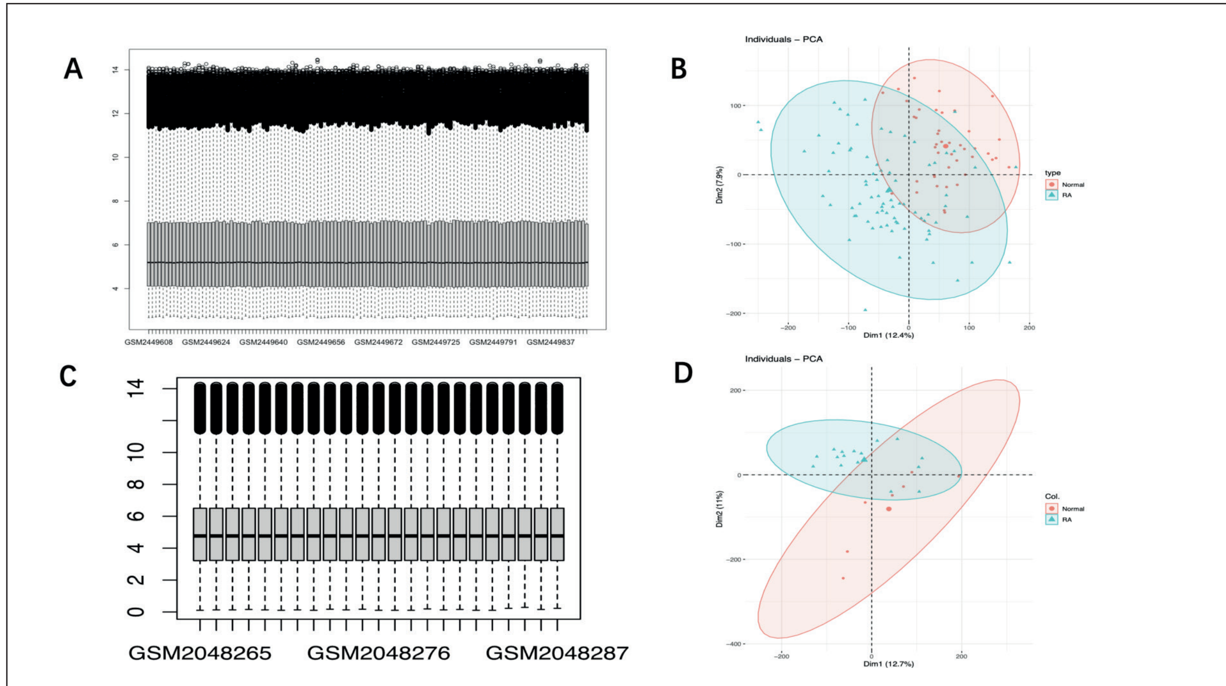

Figure 2. **A**, The box plot of gene expression levels of GSE93272 after the removal of ineligible samples; (**B**) The principal component analysis (PCA) plot of GSE93272 after the removal of ineligible samples; (**C**) The box plot of gene expression levels of GSE77298 after $\log_2$ processing and normalize; (**D**) The principal component analysis (PCA) plot of GSE77298 after $\log_2$ processing and normalization.

external biotic stimuli, immune effector processes, and defense responses to other organisms (Figure 4A). Cellular component (CC) was mainly enriched in the organelle membrane, organelle envelope, envelope, mitochondrial membrane, and mitochondrial envelope (Figure 4B). Molecular function (MF) was mainly enriched in the structural constituent of ribosome, oxidoreductase

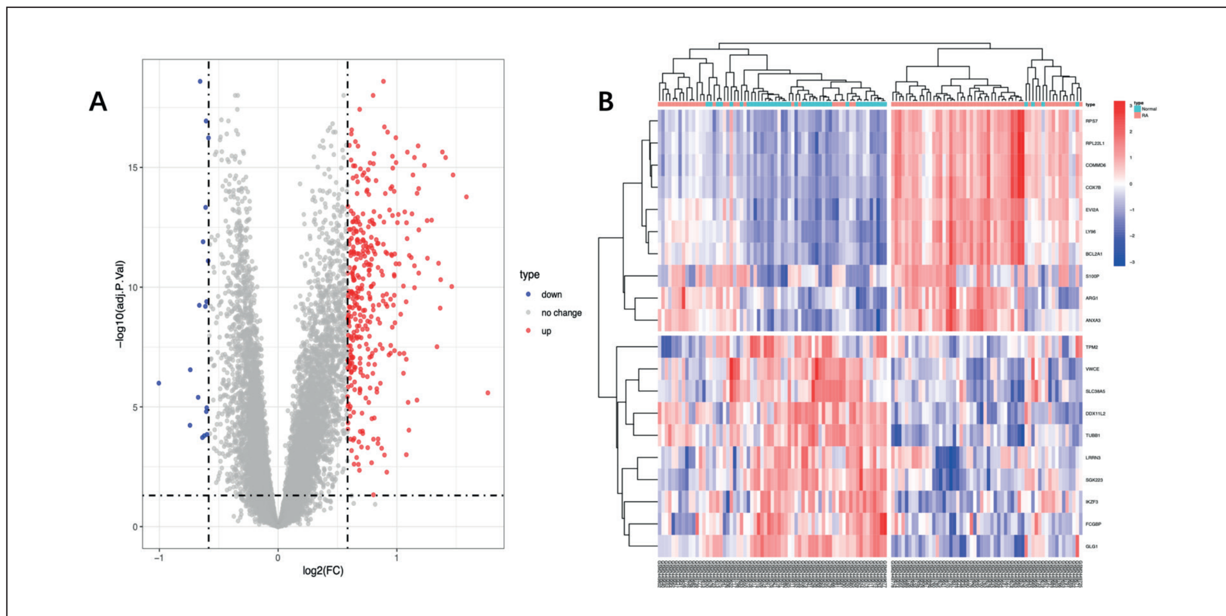

Figure 3. Screening map of RA differential genes: (**A**) Volcano map of DEGs, in which the red dots represent up-regulated genes, and blue dots represent down-regulated genes; (**B**) Heatmap shows the top 10 up-regulated genes and the top 10 down-regulated genes in DEGs. Red represents up-regulated genes, and blue represents down-regulated genes.

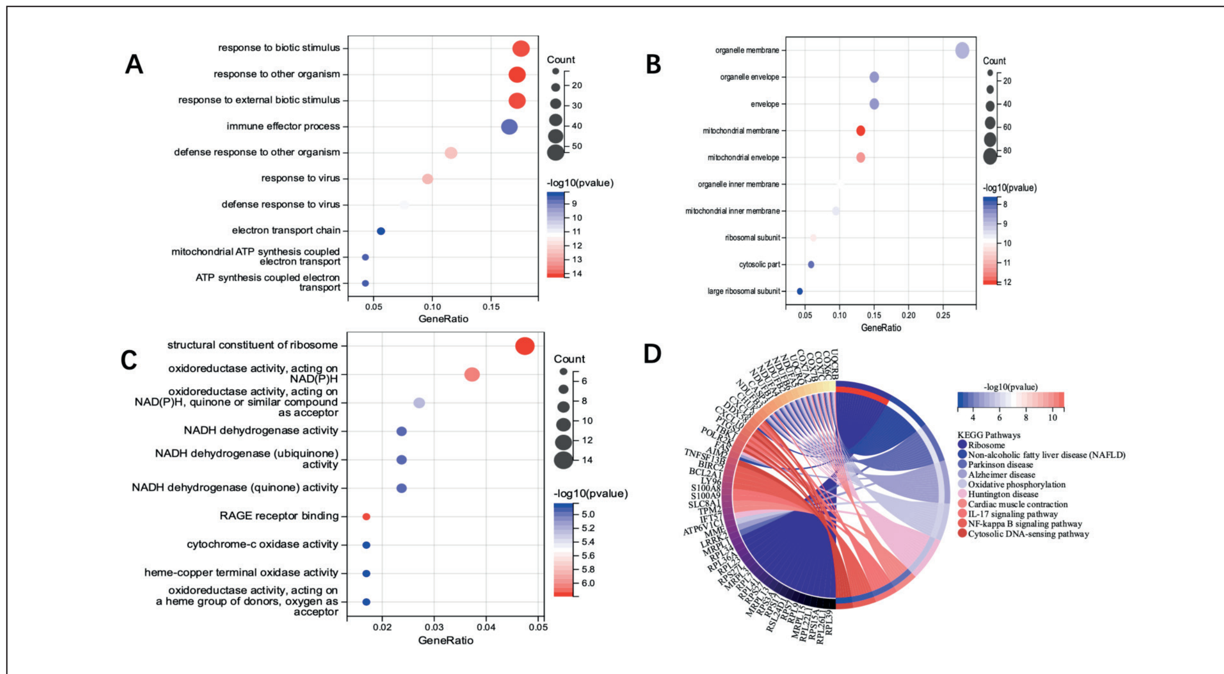

Figure 4. Plots of GO and KEGG enrichment analysis of DEGs: (**A**) The biological processes of GO analysis; (**B**) The cellular composition of GO analysis; (**C**) The molecular function of GO analysis; (**D**) The circle diagram shows the top 10 pathways in the KEGG.

activity acting on nicotinamide adenine dinucleotide phosphate [NAD(P)H], NADH dehydrogenases activity, and NADH dehydrogenase (Ubiquinone and Quinone) activity (Figure 4C). KEGG pathway analysis, mainly enriched in the ribosome, non-alcoholic fatty liver disease, Parkinson's disease, Alzheimer's disease, oxidative phosphorylation, Huntington disease, cardiac muscle contraction, interleukin (IL)-17 signaling pathway, NF-κB signaling pathway, cytosolic DNA-sensing pathway (Figure 4D).

Screening of Key Module based on WGCNA

We first filtered the top 10,000 genes with gene expression values to further correlate clinical information with key genes to perform WGCNA. The dataset was well-clustered, and no abnormal samples were found. The topological calculation used soft thresholds ranging from 1 to 20; the best soft threshold was determined to be 2 (Figure 5A). Then the relationship matrix was converted to an adjacency matrix and a topological overlap matrix (TOM) to determine the average link hierarchical clustering. Related modules were sorted according to the TOM, each consisting of no less than 30 genes. A total of 11 final modules were finally generated by merging related gene modules (Figure 5B). Figure 5C shows the correlation of the various modules with clinical traits. We found that the blue module with 1,849 genes had the highest positive correlation with RA (r=0.68), and the red module with 989 genes had the highest negative correlation with RA (r=-0.58). The blue module, which is highly positively correlated with RA, was selected as the key for screening potential pivotal genes. The scatter plot of module blue had the strongest positive correlation with RA in the GSE93272 (Figure 5D).

Obtaining Crossover Genes

One hundred forty-four crossover genes were obtained by combining the DEGs with the blue key module genes in the WGCNA with the Sangerbox tool and drawing the Venn diagram in Figure 6.

Identification of Hub Genes using PPI Network

The activity of protein-protein interaction is considered a major target of cell biology research and the prerequisite for systems biology. To further analyze protein interactions among crossover genes and improve the perspective on protein function in RA, we constructed PPIs using STRING. Further, the molecular interaction network was mapped using Cytoscape software, and the results showed

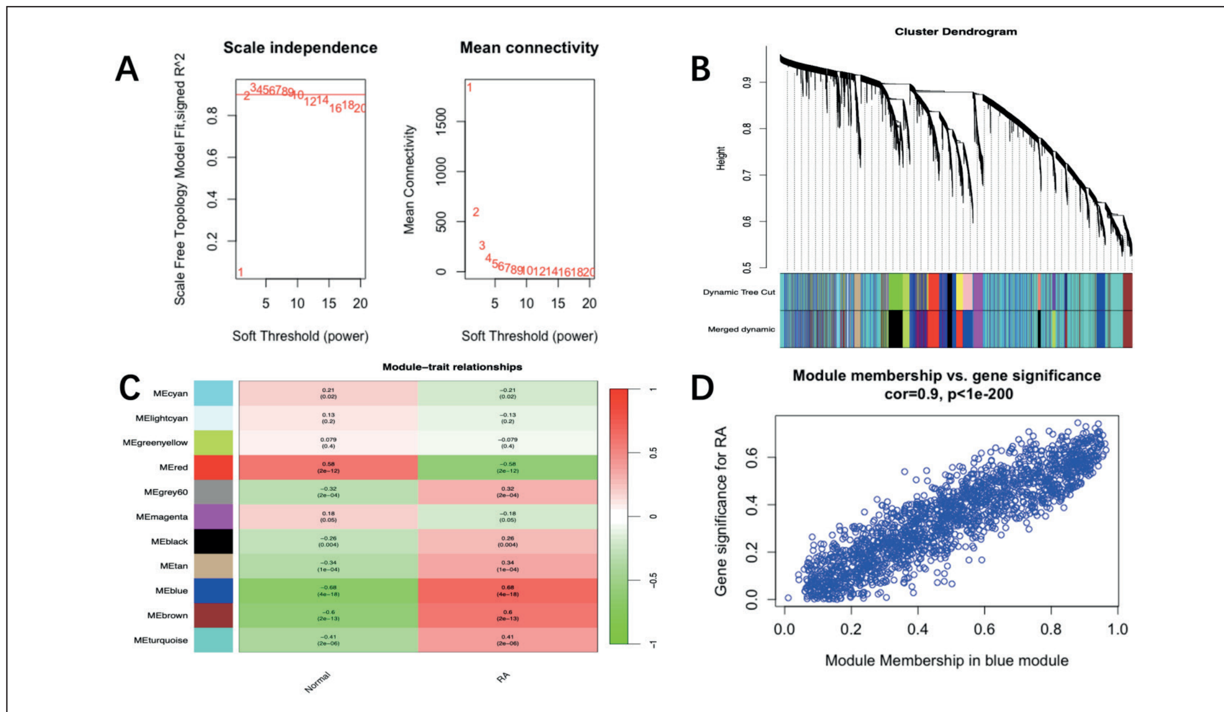

Figure 5. Construction of WGCNA for GSE93272: (**A**) Soft threshold selection process. β=2 is chosen as the soft threshold based on the scale independence and average connectivity; (**B**) Gene co-expression modules with different colors under the gene tree; (**C**) Heatmap of correlation between module genes and RA shows that the blue module has the highest association with RA; (**D**) Correlation plot between module membership and gene significance of blue module genes.

that a complex network of connections existed between these molecules (Figure 7A). Finally, we analyzed the 12 methods of calculating gene scores with the Cytoscape software Cytohubba plugin separately and statistically obtained the top 10 genes in order of frequency: *BCL2A1, PTGS2, FAS, LY96, PMAIP1, BIR2, RB1CC1, CHUK, TBK1,* and *C9orf72* (Figure 7B). These molecules will be used as candidate diagnostic biomarkers to distinguish between RA and healthy individuals.

ROC Curve Analysis in RA

We used the validation set GSE77298, which contained 7 healthy control samples and 16 RA samples, to check the diagnostic efficacy of the hub genes in RA. We then used the "pROC" R package and the "ggplot2" R package to generate ROC curves for 10 key pivotal genes (Figure 8). The promising biomarker should demonstrate both high sensitivity (percentage of true positives correctly identified) and specificity (percentage of true negatives correctly identified), which are reflected by the area under the curve (AUC) values in the ROC curve. We found that four hub genes, including *BCL2A1* (AUC=0.875), *PTGS2* (AUC=0.8482), *FAS* (AUC=0.8482), and *LY96* (AUC=0.8304), all had AUC values greater than 0.8, indicating that these have high diagnostic accuracy and can be used as diagnostic biomarkers.

Analysis of Immune Cell Infiltration in RA

Immune cell infiltration analysis is an invaluable guide for diagnosing RA disease and selecting

Figure 6. Venn diagram of DEGs with blue module genes in WGCNA. 144 crossover genes in total for later analysis.

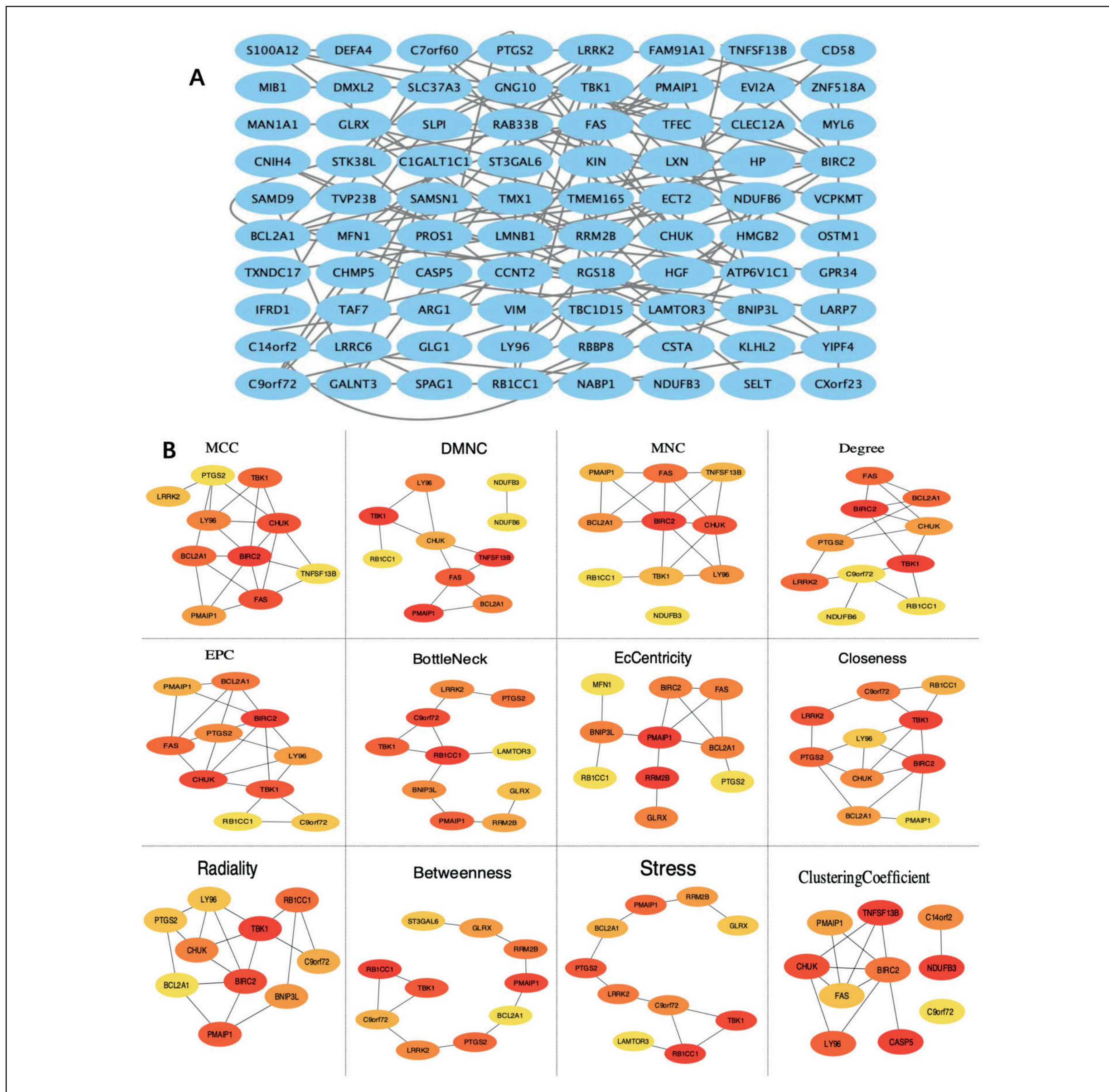

Figure 7. Construction of crossover genes interaction networks: (**A**) PPI network of the related gene in 144 crossover genes; (**B**) Top 10 rated genes obtained from 12 methods of calculating hub genes.

drug therapy. We have found that RA involves immune responses in GO and KEGG pathway analyses. Therefore, we further used the Cibersort algorithm to reveal the different levels of immune cell infiltration in the two groups and assess RA's immunological characteristics according to the immune cell infiltration. The proportion of 22 immune cells in each sample of GSE93272 is shown in Figure 9A. Figure 9B showed that compared to healthy controls, RA had higher γδ T cells ($p<0.0001$), monocytes ($p=0.15$), resting mast cells ($p<0.01$), neutrophils ($p<0.001$) and lower memory B cells ($p<0.01$), $CD8^+$ T cells ($p<0.001$), naive $CD4^+$ T cells ($p<0.0001$), regulatory T cells (Tregs) ($p<0.05$), and resting natural killer (NK) cells ($p<0.001$). In the correlation analysis between the 22 immune cells, we found that naive B cells were negatively correlated with memory B cells, $CD8^+$ T cells were negatively correlated with neutrophils, resting NK cells were negatively correlated with neutrophils, M1 macrophages were highly positively correlated with activated dendritic cells, regulatory T cells were positively correlated with

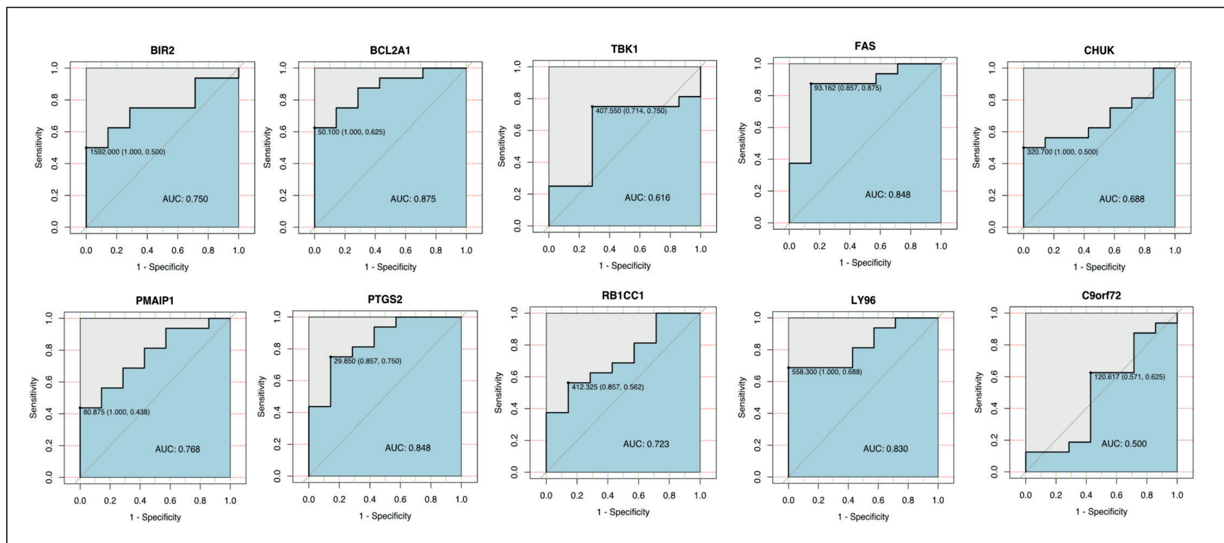

Figure 8. ROC curves for 12 hub genes in GSE77298, showing the diagnostic efficacy of candidate diagnostic biomarkers.

activated dendritic cells, M1 macrophages were positively correlated with regulatory T cells, and regulatory T cells were positively correlated with naïve B cells (Figure 9C). Figure 9D showed the correlation of four diagnostic biomarkers with 22 immune cells, in which *BCL2A1, PAS*, and *LY96* were all positively correlated with γδ T cells and negatively correlated with naive $CD4^+$ T cells, and *PTGS2* was positively correlated with neutrophils and negatively correlated with naive $CD4^+$ T cells.

Hub Gene-Drug Interaction

Based on the current study, we have identified four hub genes highly expressed in RA. It means that the drugs that inhibit the expression of hub genes will be the focus of future analyses. We used the DGidb database and obtained 6 drugs by screening for inhibitor type of interaction and interaction score>3 (Table II). Etoricoxib, Carprofen, Valdecoxib, and Oxaprozin are all non-steroidal anti-inflammatory drugs (NSAIDs) already in clinical use for treating RA. Obatoclax mesylate and Eritoran tetrasodium have been investigated[9,10] in other diseases and are expected to be potential treatments for RA.

Discussion

RA is a common systemic inflammatory disease characterized by painful, swollen joints that can severely damage physical function and quality of life. Compared to the general population, people who have RA are at higher risk of serious infections, breathing disorders, osteoporosis, heart disease, cancer, and death[11]. There have been significant advances in the mechanism of RA pathogenesis, and the introduction of targeted biologics has been significantly effective in the clinic. However, RA remains an incurable, lifelong disease[12]. Therefore, it is necessary to understand the disease's pathogenesis better, which can further discover new treatments to improve the patient's symptoms.

Table II. Table of specific information on drugs that interact with the hub gene.

Gene	Drug	Interaction Type	Interaction Score
BCL2A1	Obatoclax mesylate	inhibitor	10.3
PTGS2	Etoricoxib	inhibitor	8.94
PTGS2	Carprofen	inhibitor	7.45
PTGS2	Valdecoxib	inhibitor	4.84
PTGS2	Oxaprozin	inhibitor	3.35
LY96	Eritoran tetrasodium	inhibitor	30.91

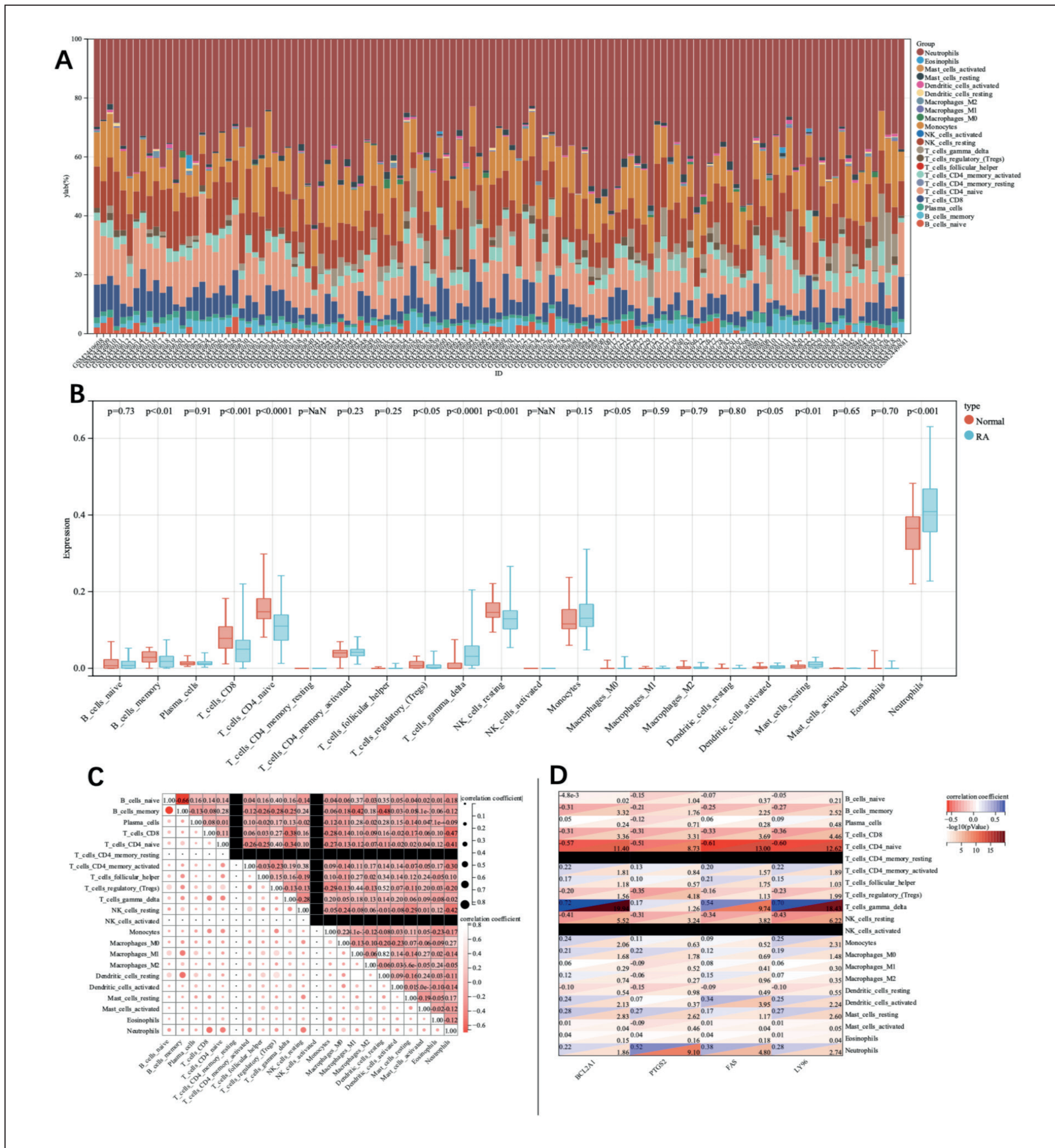

Figure 9. Immune infiltration analysis in RA: (**A**) The proportion of 22 immune cells in the different samples of GSE93272; (**B**) The level of immune infiltration of immune cells in RA and healthy controls; (**C**) The correlation between 22 immune cells; (**D**) The correlation between diagnostic biomarkers and immune cells.

Bioinformatics is an emerging interdisciplinary discipline that uses the tools of mathematics, computer science, and biology to analyze and research biological data to understand the biological significance of the data. In recent years, bioinformatics has gained much attention in the medical field due to its ability to identify hub genes in developing diseases on time, discover new causative and therapeutic targets for diseases, and thus improve the understanding of disease development and create new treatments.

This study used bioinformatics screening to obtain differentially expressed genes, some associated with RA. However, some remain to be

further investigated, suggesting that these genes are potential new research targets for RA pathogenesis. Similarly, GO and KEGG pathway analyses also brought new insights into RA. The autoimmune response contributes to the increase in osteoclasts and damage to osteoblasts in RA, so clarifying how the immune system affects synovial fibroblasts and immune cells will help us understand the mechanisms of bone damage[13]. Studies[14] have shown that nicotinamide adenine dinucleotide phosphate (NADPH) in RA is associated with levels of immune T cells, which can cause the abnormal proliferation of T cells. IL-17 cytokine family members have multiple biological functions that promote protective immunity against many pathogens but also drive inflammatory responses during infections and autoimmunity[15]. The experiment discovered that RA has high levels of IL-17, which induces osteoclast maturation and promotes harmful activation of immune and stromal cells, such as synovial and tendon[16,17]. The IL-17 pathway has become a key drug target for many autoimmune and chronic inflammatory diseases. However, what remains to be concerned is how to specifically regulate IL-17-mediated immunopathology in chronic inflammation and autoimmunity so that it does not affect protective immunity against infection. NF-κB is a key transcription factor that mediates the expression of proteins associated with the immune response and can enhance the transcription of target genes under specific stimuli to produce immunomodulatory effects[18].

Roghani et al[19] found that NF-κB is involved in pro- and anti-inflammatory processes in the pathogenesis of RA. Based on this pathway, various drugs, such as Artemisia argyi extract, are used to treat RA, which can reduce the expression of proinflammatory cytokines and the transcription factor NF-κB[20]. Icariin can suppress tumor necrosis factor-alpha (TNF-α)-induced inflammatory responses by regulating the tribbles-1 (TRIB1)/toll-like receptor 2 (TLR2)/NF-κB pathway[21]. It promises to be a potential treatment strategy for RA-related inflammation. In addition, the IL-17 pathway is strongly linked to the NF-κB signaling pathway[22]. Therefore a deeper study of whether the two pathways have a synergistic effect will enlighten us in treating RA.

We further screened four hub genes, including *BCL2A1, PTGS2, FAS*, and *LY96*, as the diagnostic biomarkers for RA based on WGCNA and ROC analysis. *BCL2A1* is induced by tumor necrosis factor and has been identified as a target gene for NF-κB. When stimulated by an antigen receptor transcription occurs[23]. *BCL2A1* can negatively regulate autophagy and induce the proliferation of macrophages and mast cells in allergic reactions, which suggests that it plays a vital role in the immune system[24]. *BCL2A1* has been shown[25] to inhibit proinflammatory cell apoptosis during the immune response, suggesting that suppressing its high expression in RA would be beneficial in treating the disease. *PTGS2* is a subtype of prostaglandin endoperoxide synthase (COX), which produces prostaglandins essential in regulating vascular tone, thrombosis, inflammation, and pain[26]. A study[27] has found that silencing *PTGS2* can reverse the inflammatory response, which holds promise as a therapeutic target for inflammation-related diseases. *FAS* is a member of the death receptor family and is central in initiating cell death, a crucial biological process for immune homeostasis[28]. Studies have confirmed that receptor activators of NF-κB ligands can promote the differentiation of T helper cell 17 (Th17) through *FAS*, which plays a vital role in the defense against pathogens and autoimmune diseases[29]. Th17 cytokines are involved in the regulation of bone destruction in RA[30]. Lymphocyte antigen 96 (*LY96*), also known as myeloid differentiation 2 (*MD2*), is necessary for the activation of TLR4 by lipopolysaccharide (LPS), which also plays an important role in innate immunity and is the first line of defense against microbial infection[31]. In RA synovial tissue, TLR2 and TLR4 expression was significantly higher than normal levels, demonstrating that TLR4 activation is associated with the pathogenesis of RA[32,33]. *LY96* is expressed explicitly in inflammation-related and immune-related diseases, such as Crohn's disease, rheumatoid arthritis, and inflammatory diabetic cardiomyopathy[34]. *LY96* is a valuable potential anti-inflammatory target for various inflammatory settings[35]. In conclusion, all four diagnostic biomarkers involve chronic inflammation and associated immune responses, suggesting that research around immune-related RA is paramount.

In the immune infiltration analysis, we found significant differences in multiple immune cells between RA and healthy controls. It demonstrates that immune cells are closely associated with critical biological processes in RA. The high levels of IL-7 in synovial tissue and body fluid in RA are involved in migrating monocytes into the inflamed joint and remodeling into proinflammatory macrophages and mature osteoblasts in RA[36]. In clinical models[37], blocking IL-7 or IL-7R

can effectively suppress joint inflammation, osteoclast formation, and neovascularization by preventing monocyte and endothelial cell infiltration and inhibiting proinflammatory macrophage and Th1/Th17 cell differentiation. In addition, a mechanism mediating local memory has been found in RA, which is closely related to the presence of tissue-resident memory T cells in the local area[38]. In affected joints, there are tissue-resident memory T cells carried by $CD8^+$ cells, and when depletion occurs, their local symptoms improve significantly. Large populations of tissue-resident memory T cells containing $CD8^+$ dominance have also been identified in human rheumatoid arthritis joint tissue fluids. These findings[39] will inspire us to determine whether the tissue-resident T cells could be used as a therapeutic target for autoimmune arthritis diseases. Regulatory T cells (Tregs) regulate the strength of the immune response, making the harmless foreign antigens tolerable and preventing pathogenic immune responses in various disease environments such as cancer and autoimmunity. A study[40] found that RA is a Th17-driven disease, and Th17/Treg imbalance is a crucial factor in RA. Various substances such as butyrate, tetrandrine, and cinnamon tannin D1 can improve autoimmune arthritis by regulating the balance between Treg/Th17[41-43]. Autoimmune diseases are thought to be caused primarily by the T and B cells. However, with the discovery that NK cells can bridge the gap between innate and adaptive immune responses, they also have an essential role in regulating autoimmune diseases[44]. NF-κB transcription factors play multiple vital roles in Treg cell differentiation and immunomodulatory functions, in which members of the NF-κB family are potential new targets for RA by regulating Treg cell function[45].

In the analysis of hub gene-drug interactions, we found that some drugs already play a significant role in the clinical management of RA. For example, etoricoxib and Valdecoxib are COX-2 selective NSAIDs, which provide significant symptomatic relief in RA while avoiding gastrointestinal reactions[46]. Etoricoxib also has a better price advantage than non-selective NSAIDs with H2 antagonists[47]. However, the increased risk of cardiovascular adverse events in patients with Valdecoxib has led to its failure to spread in clinical treatment[48]. Carprofen and Oxaprozin are non-selective NSAIDs with antipyretic, analgesic, and anti-inflammatory properties. Oxaprozin has been used clinically in treating RA, but its gastrointestinal damage and hepatotoxicity have limited its use in clinical practice[49,50]. Fortunately, pharmacological studies[51] have found that Oxaprozin prodrug has the distinct advantage of ensuring clinical effect without causing any ulceration, which is a distinct advantage. Carprofen is mainly used as a veterinary drug to relieve symptoms of arthritis, inflammation, and pain in animals[52]. Obatoclax mesylate inhibits all anti-apoptotic BCL-2 proteins and can regulate lymphocyte development and promote apoptosis and autophagy[53]. Obatoclax mesylate plays an important role in cancer disease, and further validation is needed to determine whether it works in RA[54]. Eritoran tetrasodium, a competitive inhibitor of *LY96*, inhibits the expression of TLR4 and blocks the NF-κB signaling pathway and the production of inflammatory cytokines[55]. These drugs' practical and quicker application would be expected to improve RA patients' symptoms.

Large-scale studies of high-throughput sequencing technologies and molecular mechanisms have provided insights into the origin and development of RA, but further research is needed to clarify the pathogenesis of RA.

Limitations

The limitations of this study are the limited sample size and the potential heterogeneity that needs to be better addressed. Although we extracted relevant DEGs and obtained diagnostic biomarkers for RA by bioinformatics analysis in combination with WGCNA, there needs to be additional literature support and relevant experiments such as protein blotting and immunohistochemical analysis to verify. In addition, further studies must be carried out *in vivo* and *in vitro* to obtain experimental confirmation. These shortcomings will be the main aspects that we need to resolve in the future.

Conclusions

In summary, RA's complex pathogenesis and heterogeneity make diagnosis and treatment difficult. We screened four key hub genes, *BCL2A1*, *PTGS2*, *FAS*, and *LY96*, which can be used as potential diagnostic biomarkers for RA. In addition, we have again demonstrated by enrichment and immune infiltration analyses that the immune response is closely associated with RA. Therefore, in future research, studying the changing characteristics of the human immune system may

provide new ideas for the prevention, diagnosis, and treatment of RA. Overall, our findings help to improve the diagnosis and treatment of RA. However, the mechanisms underlying the role of these four genes in the onset and progression of RA still need to be further explored.

Conflict of Interest

The author 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.

Funding

This study did not receive funding support.

Acknowledgements

We acknowledge the GEO database platforms and contributors for uploading their meaningful datasets. Meanwhile, we thank Dr. Bing Yan and Dr. Jian Wang, Affiliated Hospital of Shandong University of Traditional Chinese Medicine, for the assistance with this study. We also thank all the researchers in our working group.

Ethics Approval

The research presented in this article did not involve animal experimentation. The primary source of the data used in this study was the GEO database, which contains previously collected data from human participants. All participants provided informed consent for their data to be used for research purposes. The authors further affirm that all research was conducted with the highest standards of ethical conduct, and all necessary permissions and approvals were obtained before beginning the study.

Authors' Contributions

The study's conception and design were contributed by YZ. The first draft of the manuscript was written by YZ, BY, and JW. All authors contributed towards material preparation, data collection, and analysis. The final versions of the manuscript were revised by YZ, QYY and LSY. The final manuscript was read and approved by all authors.

Availability of Data and Materials

The principal authors are very grateful for the data support provided by the GEO databases. The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

ORCID ID

Yu Zhang: 0009-0009-3822-509X.

Informed Consent

Not applicable.

References

1) Finckh A, Gilbert B, Hodkinson B, Bae SC, Thomas R, Deane KD, Alpizar-Rodriguez D, Lauper K. Global epidemiology of rheumatoid arthritis. Nat Rev Rheumatol 2022; 18: 591-602.
2) Tobón GJ, Youinou P, Saraux A. The environment, geo-epidemiology, and autoimmune disease: Rheumatoid arthritis. Autoimmun Rev 2010; 9: A288-A292.
3) Alamanos Y, Drosos AA. Epidemiology of adult rheumatoid arthritis. Autoimmun Rev 2005; 4: 130-136.
4) Cai Y, Zhang J, Liang J, Xiao M, Zhang G, Jing Z, Lv L, Nan K, Dang X. The Burden of Rheumatoid Arthritis: Findings from the 2019 Global Burden of Diseases Study and Forecasts for 2030 by Bayesian Age-Period-Cohort Analysis. J Clin Med 2023; 12: 1291.
5) Cush JJ. Rheumatoid Arthritis: Early Diagnosis and Treatment. Prim Care 2018; 45: 237-255.
6) Siddle HJ, Bradley SH, Anderson AM, Mankia K, Emery P, Richards SH. Opportunities and challenges in early diagnosis of rheumatoid arthritis in general practice. Br J Gen Pract 2023; 73: 152-154.
7) Drobinski PJ, Nissen NI, Sinkeviciute D, Willumsen N, Karsdal MA, Bay-Jensen AC. In Contrast to Anti-CCP, MMP-Degraded and Citrullinated Vimentin (VICM) Is Both a Diagnostic and a Treatment Response Biomarker. Int J Mol Sci 2022; 24: 321.
8) Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res 2017; 45: D353-D361.
9) Chen S, Ren Y, Duan P. Biomimetic nanoparticle loading obatoclax mesylate for the treatment of non-small-cell lung cancer (NSCLC) through suppressing Bcl-2 signaling. Biomed Pharmacother 2020; 129:110371.
10) Mcdonald KA, Huang H, Tohme S, Loughran P, Ferrero K, Billiar T, Tsung A. Toll-like receptor 4 (TLR4) antagonist eritoran tetrasodium attenuates liver ischemia and reperfusion injury through inhibition of high-mobility group box protein B1 (HMGB1) signaling. Mol Med 2015; 20: 639-648.
11) Sparks JA. Rheumatoid Arthritis. Ann Intern Med 2019; 170: ITC1-ITC16.
12) Alivernini S, Firestein GS, McInnes IB. The pathogenesis of rheumatoid arthritis. Immunity 2022; 55: 2255-2270.
13) Komatsu N, Takayanagi H. Mechanisms of joint destruction in rheumatoid arthritis - immune cell-fibroblast-bone interactions. Nat Rev Rheumatol 2022; 18: 415-429.
14) Masoumi M, Alesaeidi S, Khorramdelazad H, Behzadi M, Baharlou R, Alizadeh-Fanalou S, Karami J. Role of T Cells in the Pathogenesis of Rheumatoid Arthritis: Focus on Immunometabolism Dysfunctions. Inflammation 2023; 46: 88-102.

15) Mills KHG. IL-17 and IL-17-producing cells in protection versus pathology. Nat Rev Immunol 2023; 23: 38-54.

16) Kotake S, Udagawa N, Takahashi N, Matsuzaki K, Itoh K, Ishiyama S, Saito S, Inoue K, Kamatani N, Gillespie MT, Martin TJ, Suda T. IL-17 in synovial fluids from patients with rheumatoid arthritis is a potent stimulator of osteoclastogenesis. J Clin Invest 1999; 103: 1345-1352.

17) Gremese E, Tolusso B, Bruno D, Perniola S, Ferraccioli G, Alivernini S. The forgotten key players in rheumatoid arthritis: IL-8 and IL-17 - Unmet needs and therapeutic perspectives. Front Med (Lausanne) 2023; 10: 956127.

18) Wang B, Shen J. NF-κB Inducing Kinase Regulates Intestinal Immunity and Homeostasis. Front Immunol 2022; 13: 895636.

19) Roghani SA, Lotfi R, Soleymani B, Samimi Z, Feizollahi P, Asar S, Abdan Z, Khorasanizadeh A, Taghadosi M. Investigating the correlation of the NF-κB and FoxP3 gene expression with the plasma levels of pro- and anti-inflammatory cytokines in rheumatoid arthritis patients. Clin Rheumatol 2023; 42: 1443-1450.

20) Chen JK, Kuo CH, Kuo WW, Day CH, Wang TF, Ho TJ, Lin PY, Lin SZ, Shih TC, Shih CY, Huang CY, Lu CY. Artemisia argyi extract ameliorates IL-17A-induced inflammatory response by regulation of NF-κB and Nrf2 expression in HIG-82 synoviocytes. Environ Toxicol 2022; 37: 2793-2803.

21) Wu ZM, Xiang YR, Zhu XB, Shi XD, Chen S, Wan X, Guo J. Icariin represses the inflammatory responses and survival of rheumatoid arthritis fibroblast-like synoviocytes by regulating the TRIB1/TLR2/NF-kB pathway. Int Immunopharmacol 2022; 110: 108991.

22) Zhang Q, Li Y, Yin C, Gao M, Yu J, Guo J, Xian X, Hou Z, Wang Q. IL-17A deletion reduces sevoflurane-induced neurocognitive impairment in neonatal mice by inhibiting NF-κB signaling pathway. Bioengineered 2022; 13: 14562-14577.

23) Kathania M, Raje CI, Raje M, Dutta RK, Majumdar S. Bfl-1/A1 acts as a negative regulator of autophagy in mycobacteria infected macrophages. Int J Biochem Cell Biol 2011; 43: 573-585.

24) Vogler M. BCL2A1: the underdog in the BCL2 family. Cell Death Differ 2012; 19: 67-74.Hellmann J, Tang Y, Zhang MJ, Hai T, Bhatnagar A, Srivastava S, Spite M. Atf3 negatively regulates Ptgs2/Cox2 expression during acute inflammation. Prostaglandins Other Lipid Mediat 2015; 116-117: 49-56.

25) Iglesias M, Augustin JJ, Alvarez P, Santiuste I, Postigo J, Merino J, Merino R. Selective Impairment of TH17-Differentiation and Protection against Autoimmune Arthritis after Overexpression of BCL2A1 in T Lymphocytes. PLoS One 2016; 11: e0159714.

26) Hellmann J, Tang Y, Zhang MJ, Hai T, Bhatnagar A, Srivastava S, Spite M. Atf3 negatively regulates Ptgs2/Cox2 expression during acute inflammation. Prostaglandins Other Lipid Mediat 2015; 116-117: 49-56.

27) Li R, Xie J, Xu W, Zhang L, Lin H, Huang W. LPS-induced PTGS2 manipulates the inflammatory response through trophoblast invasion in preeclampsia via NF-κB pathway. Reproductive Biology 2022; 22: 100696.

28) Maeda K, Nakayama J, Taki S, Sanjo H. TAK1 Limits Death Receptor Fas-Induced Proinflammatory Cell Death in Macrophages. J Immunol 2022; 209: 1173-1179.

29) Cui Z, Feng R, Liu Z, Gong Y, Zhang Y. Receptor Activator of Nuclear Factor (Nf)-kb Ligand Promotes T Helper 17 Cell Differentiation through Fas. Immunol Invest 2022; 51: 1385-1397.

30) Kim KW, Kim HR, Kim BM, Cho ML, Lee SH. Th17 cytokines regulate osteoclastogenesis in rheumatoid arthritis. Am J Pathol 2015; 185: 3011-3024.

31) Dou T, Fu M, Wang Y, Zhao Y, Wang Z, Bian Z, Zhou Y. Signatures of positive selection in LY96 gene in vertebrates. J Biosci 2013; 38: 899-904.

32) Radstake TR, Roelofs MF, Jenniskens YM, Oppers-Walgreen B, van Riel PL, Barrera P, Joosten LA, van den Berg WB. Expression of toll-like receptors 2 and 4 in rheumatoid synovial tissue and regulation by proinflammatory cytokines interleukin-12 and interleukin-18 via interferon-gamma. Arthritis Rheum 2004; 50: 3856-3865.

33) Kowalski ML, Wolska A, Grzegorczyk J, Hilt J, Jarzebska M, Drobniewski M, Synder M, Kurowski M. Increased Responsiveness to Toll-Like Receptor 4 Stimulation in Peripheral Blood Mononuclear Cells from Patients with Recent Onset Rheumatoid Arthritis. Mediators Inflamm 2008; 2008: 132732.

34) Nie K, Li J, Peng L, Zhang M, Huang W. Pan-Cancer Analysis of the Characteristics of LY96 in Prognosis and Immunotherapy Across Human Cancer. Front Mol Biosci 2022; 9: 837393.

35) Chen L, Fu W, Zheng L, Wang Y, Liang G. Recent progress in the discovery of myeloid differentiation 2 (MD2) modulators for inflammatory diseases. Drug Discovery Today 2018; 23: 1187-1202.

36) Iwamoto N, Kawakami A. The monocyte-to-osteoclast transition in rheumatoid arthritis: Recent findings. Front Immunol 2022; 13: 998554.

37) Meyer A, Parmar PJ, Shahrara S. Significance of IL-7 and IL-7R in RA and autoimmunity. Autoimmun Rev 2022; 21: 103120.

38) Gao A, Zhao W, Wu R, Su R, Jin R, Luo J, Gao C, Li X, Wang C. Tissue-resident memory T cells: The key frontier in local synovitis memory of rheumatoid arthritis. J Autoimmun 2022; 133: 102950.

39) Chang MH, Levescot A, Nelson-Maney N, Blaustein RB, Winden KD, Morris A, Wactor A, Balu S, Grieshaber-Bouyer R, Wei K, Henderson LA, Iwakura Y, Clark RA, Rao DA, Fuhlbrigge RC, Nigrovic PA. Arthritis flares mediated by tissue-resident memory T cells in the joint. Cell Rep 2021; 37: 109902.

40) Li YJ, Chen Z. Cell-based therapies for rheumatoid arthritis: opportunities and challenges. Ther Adv Musculoskelet Dis 2022; 14: 1759720X221100294.

41) Hui W, Yu D, Cao Z, Zhao X. Butyrate inhibit collagen-induced arthritis via Treg/IL-10/Th17 axis. Int Immunopharmacol 2019; 68: 226-233.

42) Shi C, Zhang H, Wang X, Jin B, Jia Q, Li Y, Yang Y. Cinnamtannin D1 attenuates autoimmune arthritis by regulating the balance of Th17 and treg cells through inhibition of aryl hydrocarbon receptor expression. Pharmacol Res 2020; 151: 104513.

43) Yuan X, Tong B, Dou Y, Wu X, Wei Z, Dai Y. Tetrandrine ameliorates collagen-induced arthritis in mice by restoring the balance between Th17 and Treg cells via the aryl hydrocarbon receptor. Biochem Pharmacol 2016; 101: 87-99.

44) 41.Hojjatipour T, Aslani S, Salimifard S, Mikaeili H, Hemmatzadeh M, Gholizadeh Navashenaq J, Ahangar Parvin E, Jadidi-Niaragh F, Mohammadi H. NK cells - Dr. Jekyll and Mr. Hyde in autoimmune rheumatic diseases. Int Immunopharmacol 2022; 107: 108682.

45) Hövelmeyer N, Schmidt-Supprian M, Ohnmacht C. NF-κB in control of regulatory T cell development, identity, and function. J Mol Med (Berl) 2022; 100: 985-995.

46) Chen YF, Jobanputra P, Barton P, Bryan S, Fry-Smith A, Harris G, Taylor RS. Cyclooxygenase-2 selective non-steroidal anti-inflammatory drugs (etodolac, meloxicam, celecoxib, rofecoxib, etoricoxib, valdecoxib and lumiracoxib) for osteoarthritis and rheumatoid arthritis: a systematic review and economic evaluation. Health Technol Assess 2008; 12: 1-278, iii.

47) Moore A, Phillips C, Hunsche E, Pellissier J, Crespi S. Economic evaluation of etoricoxib versus non-selective NSAIDs in the treatment of osteoarthritis and rheumatoid arthritis patients in the UK. Pharmacoeconomics 2004; 22: 643-660.

48) Iftkhar S, de Sá AGC, Velloso JPL, Aljarf R, Pires DEV, Ascher DB. cardioToxCSM: A Web Server for Predicting Cardiotoxicity of Small Molecules. J Chem Inf Model 2022; 62: 4827-4836.

49) Abdelbasset WK, Elkholi SM, Ismail KA, Alshehri S, Alobaida A, Huwaimel B, Alatawi AD, Alsubaiyel AM, Venkatesan K, Abourehab MAS. Development a novel robust method to enhance the solubility of Oxaprozin as nonsteroidal anti-inflammatory drug based on machine-learning. Sci Rep 2022; 12: 13138.

50) Sharma V, Bhatia P, Alam O, Javed Naim M, Nawaz F, Ahmad Sheikh A, Jha M. Recent advancement in the discovery and development of COX-2 inhibitors: Insight into biological activities and SAR studies (2008–2019). Bioorganic Chemistry 2019; 89: 103007.

51) Peesa JP, Atmakuri LR, Yalavarthi PR, Mandava Venkata BR, Rasheed A, Pachava V. Oxaprozin prodrug as safer nonsteroidal anti-inflammatory drug: Synthesis and pharmacological evaluation. Archiv der Pharmazie 2018; 351: 1700256.

52) Gómez-Segura L, Boix-Montañes A, Mallandrich M, Parra-Coca A, Soriano-Ruiz JL, Calpena AC, Gimeno Á, Bellido D, Colom H. Swine as the Animal Model for Testing New Formulations of Anti-Inflammatory Drugs: Carprofen Pharmacokinetics and Bioavailability of the Intramuscular Route. Pharmaceutics 2022; 14: 1045.

53) Cournoyer S, Addioui A, Belounis A, Beaunoyer M, Nyalendo C, Le Gall R, Teira P, Haddad E, Vassal G, Sartelet H. GX15-070 (Obatoclax), a Bcl-2 family proteins inhibitor engenders apoptosis and pro-survival autophagy and increases Chemosensitivity in neuroblastoma. BMC Cancer 2019; 19: 1018.

54) Li J, Xu J, Li Z. Obatoclax, the pan-Bcl-2 inhibitor sensitizes hepatocellular carcinoma cells to promote the anti-tumor efficacy in combination with immune checkpoint blockade. Transl Oncol 2021; 14: 101116.

55) Talepoor AG, Rastegari B, Kalani M, Doroudchi M. Decrease in the inflammatory cytokines of LPS-stimulated PBMCs of patients with atherosclerosis by a TLR-4 antagonist in the co-culture with HUVECs. Int Immunopharmacol 2021; 101: 108295.