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

Y. ZHANG¹, Q.-Y. YANG¹, L.-S. YANG², J. ZHANG¹, K. ZHANG¹, Y. LIU¹, X.-D. XU¹, X.-Y. YANG¹, J. WANG³, B. YAN²

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

Introduction

Rheumatoid arthritis (RA) is a systemic autoimmune disease that causes damage to bones and joints. It is a multifactorial disease primarily caused by the interaction between genetic and environmental factors. The incidence of RA is widely variable worldwide, with rates as high as 0.5-1% in Northern Europe and North America. Although studies 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. 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. 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. 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. 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. 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.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Platform</th>
<th>Manufacturer</th>
<th>Group</th>
<th>Tissue type</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSE93272</td>
<td>GPL570</td>
<td>[HG-U133 Plus 2] Affymetrix Human Genome U133 Plus 2.0 Array</td>
<td>Normal</td>
<td>43</td>
</tr>
<tr>
<td>GSE77298</td>
<td>GPL570</td>
<td>[HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array</td>
<td>7</td>
<td>16</td>
</tr>
</tbody>
</table>
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, etc.].
Identification of diagnostic biomarkers and immuno-infiltration analysis

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 |log2 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,
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 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 processing and normalization; (D) The principal component analysis (PCA) plot of GSE77298 after log processing and normalization.

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.
Identification of diagnostic biomarkers and immuno-infiltration analysis

Assessment of gene expression and 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...
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...
Identification of diagnostic biomarkers and immuno-infiltration analysis

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 \((p=0.001)\), neutrophils \((p<0.001)\) and lower memory B cells \((p<0.001)\), 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...
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 investigated⁹,¹⁰ 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 death¹¹. 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 disease¹². 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.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Drug</th>
<th>Interaction Type</th>
<th>Interaction Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>BCL2A1</td>
<td>Obatoclax mesylate</td>
<td>inhibitor</td>
<td>10.3</td>
</tr>
<tr>
<td>PTGS2</td>
<td>Etoricoxib</td>
<td>inhibitor</td>
<td>8.94</td>
</tr>
<tr>
<td>PTGS2</td>
<td>Carprofen</td>
<td>inhibitor</td>
<td>7.45</td>
</tr>
<tr>
<td>PTGS2</td>
<td>Valdecoxib</td>
<td>inhibitor</td>
<td>4.84</td>
</tr>
<tr>
<td>PTGS2</td>
<td>Oxaprozin</td>
<td>inhibitor</td>
<td>3.35</td>
</tr>
<tr>
<td>LY96</td>
<td>Eritoran tetrasodium</td>
<td>inhibitor</td>
<td>30.91</td>
</tr>
</tbody>
</table>
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
Studies 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. The experiment discovered that RA is associated with levels of immune T cells, which can cause the abnormal proliferation of T cells. IL-17 cytokines are involved in the regulation of 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. When stimulated by an antigen receptor, NF-κB 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. BCL2A1 has been shown 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. A study 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. 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. Th17 cytokines are involved in the regulation of bone destruction in RA. 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. 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. LY96 is expressed explicitly in inflammation-related and immune-related diseases, such as Crohn's disease, rheumatoid arthritis, and inflammatory diabetic cardiomyopathy. LY96 is a valuable potential anti-inflammatory target for various inflammatory settings. 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. In clinical models, 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\textsuperscript{39}. In affected joints, there are tissue-resident memory T cells carried by CD8\textsuperscript{+} cells, and when depletion occurs, their local symptoms improve significantly. Large populations of tissue-resident memory T cells containing CD8\textsuperscript{+} dominance have also been identified in human rheumatoid arthritis joint tissue fluids. These findings\textsuperscript{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\textsuperscript{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\textsuperscript{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\textsuperscript{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\textsuperscript{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\textsuperscript{46}. Etoricoxib also has a better price advantage than non-selective NSAIDs with H2 antagonists\textsuperscript{47}. However, the increased risk of cardiovascular adverse events in patients with Valdecoxib has led to its failure to spread in clinical treatment\textsuperscript{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\textsuperscript{49,50}. Fortunately, pharmacological studies\textsuperscript{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\textsuperscript{52}. Obatoclax mesylate inhibits all anti-apoptotic BCL-2 proteins and can regulate lymphocyte development and promote apoptosis and autophagy\textsuperscript{53}. Obatoclax mesylate plays an important role in cancer disease, and further validation is needed to determine whether it works in RA\textsuperscript{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\textsuperscript{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
5) Cush JJ. Rheumatoid Arthritis: Early Diagnosis and Treatment. Prim Care 2018; 45: 237-255.
Identification of diagnostic biomarkers and immuno-infiltration analysis


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) Taleporo 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.