



## Research article

# Screening and identification of potential hub genes and immune cell infiltration in the synovial tissue of rheumatoid arthritis by bioinformatic approach

Zhi-wei Feng<sup>a,b,c,d</sup>, Yu-chen Tang<sup>a,b,c</sup>, Xiao-yun Sheng<sup>a,b,c</sup>, Sheng-hong Wang<sup>a,b,c</sup>, Yao-bin Wang<sup>a,b,c</sup>, Zhong-cheng Liu<sup>a,b,c</sup>, Jin-min Liu<sup>a,b,c</sup>, Bin Geng<sup>a,b,c</sup>, Ya-yi Xia<sup>a,b,c,\*</sup>

<sup>a</sup> Department of Orthopaedics, Lanzhou University Second Hospital, Lanzhou, China

<sup>b</sup> Gansu Province Orthopaedic Clinical Medicine Research Center, Lanzhou, China

<sup>c</sup> Gansu Province Intelligent Orthopedics Industry Technology Center, Lanzhou, China

<sup>d</sup> Department of Orthopaedics, Nanchong Central Hospital, The Second Clinical Institute of North Sichuan Medical College, Nanchong, China

## ARTICLE INFO

**Keywords:**

Bioinformatics approach  
Hub genes  
WGCNA  
Rheumatoid arthritis  
Immune infiltration

## ABSTRACT

**Background:** Rheumatoid arthritis (RA) is an autoimmune disease that affects individuals of all ages. The basic pathological manifestations are synovial inflammation, pannus formation, and erosion of articular cartilage, bone destruction will eventually lead to joint deformities and loss of function. However, the specific molecular mechanisms of synovitis tissue in RA are still unclear. Therefore, this study aimed to screen and explore the potential hub genes and immune cell infiltration in RA.

**Methods:** Three microarray datasets (GSE12021, GSE55457, and GSE55235), from the Gene Expression Omnibus (GEO) database, have been analyzed to explore the potential hub genes and immune cell infiltration in RA. First, the LIMMA package was used to screen the differentially expression genes (DEGs) after removing the batch effect. Then the clusterProfiler package was used to perform functional enrichment analyses. Second, through weighted coexpression network analysis (WGCNA), the key module was identified in the coexpression network of the gene set. Third, the protein-protein interaction (PPI) network was constructed through STRING website and the module analysis was performed using Cytoscape software. Fourth, the CIBERSORT and ssGSEA algorithm were used to analyze the immune status of RA and healthy synovial tissue, and the associations between immune cell infiltration and RA-related diagnostic biomarkers were evaluated. Fifth, we used the quantitative reverse transcription-polymerase chain reaction (qRT-PCR) to validate the expression levels of the hub genes, and ROC curve analysis of hub genes for discriminating between RA and healthy tissue. Finally, the gene-drug interaction network was constructed using DrugCentral database, and identification of drug molecules based on hub genes using the Drug Signature Database (DSigDB) by Enrichr.

**Results:** A total of 679 DEGs were identified, containing 270 downregulated genes and 409 upregulated genes. DEGs were primarily enriched in immune response and chemokine signaling pathways, according to functional enrichment analysis of DEGs. WGCNA explored the co-

\* Corresponding author. No. 82 Cuiyingmen, Chengguan District, Lanzhou City, Gansu Province, China.;

E-mail addresses: [946678828@qq.com](mailto:946678828@qq.com) (Z.-w. Feng), [tangych20@lzu.edu.cn](mailto:tangych20@lzu.edu.cn) (Y.-c. Tang), [shengxy20@lzu.edu.cn](mailto:shengxy20@lzu.edu.cn) (X.-y. Sheng), [wangshh15@lzu.edu.cn](mailto:wangshh15@lzu.edu.cn) (S.-h. Wang), [348776369@qq.com](mailto:348776369@qq.com) (Y.-b. Wang), [liuzhch\\_14@163.com](mailto:liuzhch_14@163.com) (Z.-c. Liu), [276458898@qq.com](mailto:276458898@qq.com) (J.-m. Liu), [cxxxf@qq.com](mailto:cxxxf@qq.com) (B. Geng), [xiayy@lzu.edu.cn](mailto:xiayy@lzu.edu.cn) (Y.-y. Xia).

<https://doi.org/10.1016/j.heliyon.2023.e12799>

Received 18 July 2022; Received in revised form 29 December 2022; Accepted 2 January 2023

Available online 10 January 2023

2405-8440/© 2023 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

expression network of the gene set and identified key modules, the blue module was selected as the key module associated with RA. Seven hub genes are identified when PPI network and WGCNA core modules are intersected. Immune infiltration analysis using CIBERSORT and ssGSEA algorithms revealed that multiple types of immune infiltration were found to be upregulated in RA tissue compared to normal tissue. Furthermore, the levels of 7 hub genes were closely related to the relative proportions of multiple immune cells in RA. The results of the qRT-PCR demonstrated that the relative expression levels of 6 hub genes (CD27, LCK, CD2, GZMB, IL7R, and IL2RG) were up-regulated in RA synovial tissue, compared with normal tissue. Simultaneously, ROC curves indicated that the above 6 hub genes had strong biomarker potential for RA (AUC >0.8).

**Conclusions:** Through bioinformatics analysis and qRT-PCR experiment, our study ultimately discovered 6 hub genes (CD27, LCK, CD2, GZMB, IL7R, and IL2RG) that closely related to RA. These findings may provide valuable direction for future RA clinical diagnosis, treatment, and associated research.

## 1. Introduction

Rheumatoid arthritis (RA), a systemic autoimmune disease, with synovitis and cartilage degradation as primary clinical features [1-4], afflicting approximately 0.5%–2% population of the world, especially women, but the exact cause remains unclear. Patients with RA can eventually develop joint deformities and disabilities, reducing their quality of life to a great extent. Although drug therapy can improve RA symptoms, the side effects of long-term medication are very serious [5]. Inflammatory cytokines, chemokines, proteases, and matrix lyases can disrupt the immune homeostasis of body, contributing to the development of RA [6]. Consequently, identifying the hub genes of RA may be possible to clarify its pathogenesis and to develop new molecular targets for diagnosis and treatment.

Abundance of microarray data from the GEO database allow us to readily investigated the correlation between genes and diseases. Recent years have seen bioinformatics analysis become an important method to screen disease hub genes and explore pathogenesis. Hub genes are defined as genes that play a crucial role in gene regulation and biological processes, which can interact with many other genes in a gene network [7], and usually closely related to specific diseases. Therefore, the hub genes obtained in this study may play a crucial role in the pathogenesis of RA, and may become potential drug targets for RA targeted therapy. For instance, using bioinformatics analysis, FADD, CXCL8, and CXCL2 were identified as potential hub genes in RA [8]. However, Bioinformatics analysis results have yet to be further validated by experiments. According to another study, bioinformatic analysis identified LCK, SOCS3, STAT1, JAK2, and EGFR as hub genes associated with RA [9], and verified the hub genes by qRT-PCR in the synovial tissue of the mouse model. Since animal models cannot fully replicate the pathogenesis of RA in humans, they can only partially explain the pathological manifestations of RA. In our study, human synovial tissue samples were collected from RA patients and healthy controls to verify the expression of the hub genes.

Bioinformatics encompasses the fields of genomics, biology, mathematics, computer science, and statistics [10], which has important implications for some disease diagnosis, treatment or improved understanding of disease mechanisms. In this study, three microarray datasets of RA and normal synovial tissues downloaded from GEO database were analyzed by bioinformatics methods. After merging and batch correction, differential expression analysis, functional enrichment analysis, WGCNA analysis, and immune infiltration analysis were performed. Then combined with PPI network analysis to explore the hub genes closely related to RA progression. To ensure the accuracy of the analysis, the CIBERSORT and ssGSEA algorithm were used to evaluate the differences in immune infiltration between RA and normal synovium. The correlation between immune cell infiltration and hub gene expressions was then calculated. Finally, the synovial tissue of RA patients was collected and the hub genes were verified by qRT-PCR, and ROC curves were used to determine if hub genes could distinguish RA patients from normals. In the future work, we will select hub gene for further mechanism study in fibroblast-like synoviocytes (FLS).

## 2. Materials & methods

### 2.1. Data processing and technology roadmap

Microarray datasets of RA were downloaded from the GEO database, “Rheumatoid Arthritis”, “Expression profiling by array”, “Homo sapiens”, and “sample count >20” were used as criteria. Three gene expression profiles (GSE12021, GSE55457, and GSE55235) derived from the GPL96 platform were screened for further study. GSE12021 includes 12 samples of synovial tissue from RA patients and 9 samples from healthy knee joints. GSE55457 is made up of 23 samples, 10 of which are synovial tissue samples from healthy knee joints and 13 of which are synovial tissue samples from RA patients. GSE55235 consists of 20 samples, 10 of which are synovial tissue samples taken from healthy knee joints and 10 from RA patients.

The R program (version 4.1.2) and the Bioconductor packages website (<http://www.bioconductor.org/>) were utilized to analyze the data. The computer code utilized in this study exists at [GitHub.com](https://github.com/withfeng/RA) (<https://github.com/withfeng/RA>). Converted gene probe ID to gene symbol using perl software (version 5.30) according to GPL96 platform of Affymetrix Human Genome U133A Array, matched the probe with GPL96 platform genes to perform microarray annotation, when a probe corresponds to multiple genes, was omitted,

when multiple probes correspond to one gene, the median value were calculated. Three datasets were merged by R software merge () function. Removed the batch effects and normalized the merged data with the “SVA” package [11] (version 3.42). Fig. 1 depicts the technology roadmap.

## 2.2. Identification of DEGs

The data set was downloaded from the GEO database in TXT format. Limma package [12] (version: 3.5) of R software (version: 4.1.2) was utilized to identify the DEGs. Analyzed the adj. *P*-value to correct the false positive results of GEO datasets. Set the thresholds to “adj. *P*-value < 0.05 and |logFC| > 1” for the identification of DEGs.

## 2.3. Functional enrichment analysis of DEGs

To investigate the potential biological roles of DEGs in more detail, the “ClusterProfiler” package (version: 3.18.0) [13] and “org.Hs.eg.db” package in R software were utilized to perform Gene Ontology (GO) function enrichment analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis, with *q* value < 0.05 as the enrichment threshold. GO analysis was primarily used to annotate gene functions, especially biological pathways (BP), cellular components (CC) and molecular functions (MF). KEGG analysis was used to detect pathway enrichment in DEGs. Gene Set Enrichment Analysis (GSEA) [14] was then performed, also using the “clusterProfiler” R package and immunologic signature gene set (C7 gene sets) obtained from the Molecular Signatures Database (MSigDB, <https://www.gsea-msigdb.org/>).

## 2.4. Constructing Co-expression network and identifying module

Gene clusters highly associated with RA were identified by the WGCNA method [15] using R software. Determined the soft threshold power ( $\beta$ ) by using the pickSoftThreshold R function, then performed hierarchical clustering into colored modules based on the weighted co-expression network. Gene significance (GS) and module membership (MM) were calculated using the Heatmap plugin to assess the relationship between modules and RA, further function analyses will focus on the module most correlated with RA.

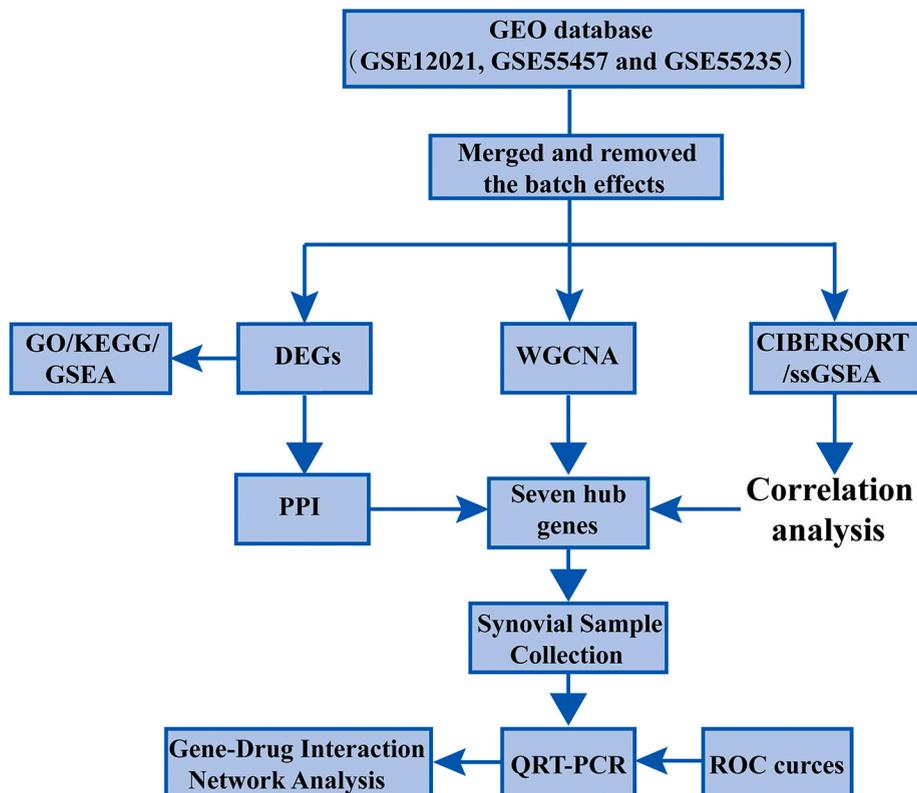


Fig. 1. A flowchart of the studies that were considered for inclusion in the analysis.

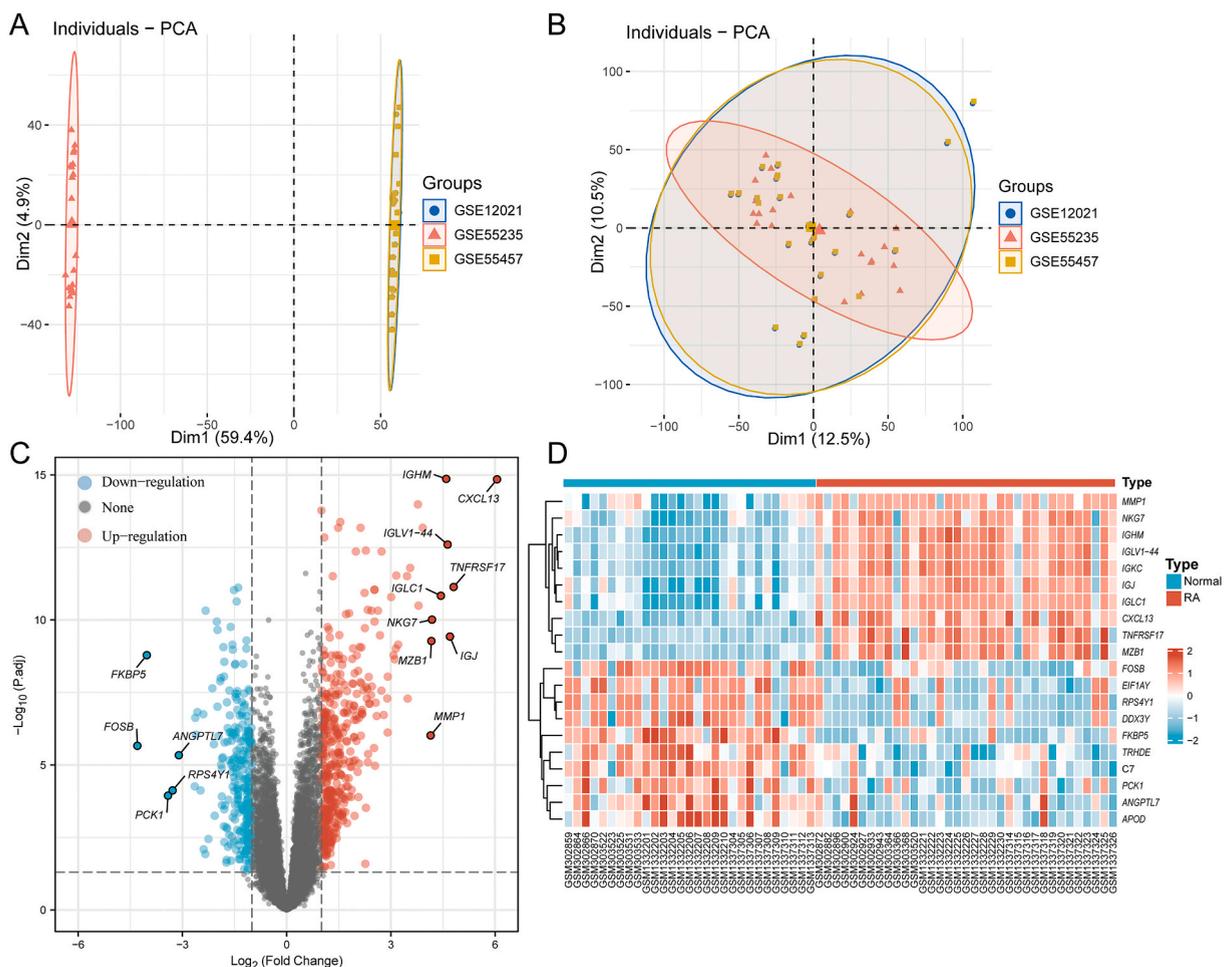
## 2.5. PPI network construction and module analysis

To explore the potential relationship among those DEGs, we applied STRING website (<https://string-db.org/>, Version: 11.5) [16] to construct the PPI network and mapped the DEGs into STRING (confidence score = 0.4, max number of interactors = 0). Furthermore, we used the Molecular Complex Detection (MCODE) [17] plugin (Degree Cutoff = 2, Node Score Cutoff = 0.2 and k-core = 2) and CytosHubba [18] plugin with default parameters to determine hub genes of the PPI network in Cytoscape (version 3.8.2) [19].

## 2.6. Immune infiltration analysis

CIBERSORT [20], as a versatile computational method, was used to quantify immune cell fractions from the massive tissue gene expression datasets based on immune cell signatures, which includes 547 genes with high sensitivity and specificity for 22 phenotypes of human immune cells. The samples were filtered if the  $P$  value < 0.05, and the proportion of each kind of immune cell was calculated. Correlations involving 22 immune cell populations were displayed with the “corrplot” package (version 0.92). To determine whether the immune cell infiltration in the synovial tissue of RA patients differed significantly from that in healthy controls, we used principal component analysis (PCA) on all samples. The “vioplot” package (version 0.3.7) was used to compare immune infiltration levels between the two groups.

In addition, single-sample gene set enrichment analysis (ssGSEA) was used to calculate the infiltration scores of 16 immune cells and the activities of 13 immune-related pathways [21].



**Fig. 2.** (A) PCA plot before removing batch effects. (B) PCA plot after removing batch effects. Clusters of sample points with different distances are from different batches and sequencing platforms. Whereas plot B shows a reduced difference in distance between batches after removing the batch effect. (C) Fold-change values and an adjusted P-value were used to create volcano graphs. The over-expressed mRNAs are shown by the red point in the figure, while the down-expressed mRNAs are represented by the blue point. (D) According to logFC, the top 20 downregulated genes and upregulated genes that were differentially expressed between RA and normal tissues are represented in the heatmap. The over-expressed mRNAs are shown by the red point in the figure, while the down-expressed mRNAs are represented by the blue point. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

## 2.7. Correlation analysis of hub genes and the immune status

The R “ggcorrplot” package was used to conduct Spearman’s rank correlation analysis to illustrate the relationships between hub genes expression levels and the immune status.

## 2.8. Sample collection and QRT-PCR

The synovial tissue of three meniscus injury individuals undergoing arthroscopic surgery and three patients with RA was collected for qRT-PCR verification in order to confirm the outcomes of the bioinformatics analysis. The Ethics Committee of the Second Hospital of Lanzhou University had approved the research protocol (approval number: 2022A-205), and all patients signed an informed consent form. Total RNA from synovial tissue was extracted using the Trizol reagent (Invitrogen, USA). The Prime Script RT Kit (TaKaRa, Japan) was used to synthesize cDNA. PCR amplification was performed by a CFX96 real-time PCR instrument (BIO-RAD, USA) with SYBR Premix (TaKaRa, Japan). PCR data was normalized by GAPDH expression. The method of quantification for PCR data was calculated with the  $2^{-\Delta\Delta C_q}$  computation method [22], then subjected to one-way analysis of variance using Graphpad Prism7 software, and  $P$  value  $< 0.05$  was considered a significant difference. Supplementary information, Table S1, lists all primers.

## 2.9. Gene-drug interaction network analysis and applicant drug evaluation

DrugCentral [23], a pharmaceutical information resource website, developed by combining information related to active pharmaceutical ingredients (APIs), bioactivity profiles, regulatory information, drug mechanism of action (MoA), pharmaceutical products, pharmacological action, and indications. The hub gene-drug interaction network was constructed using the DrugCentral, and was visualized using Cytoscape (version 3.8.2). Enrichr [24], with a large library of diverse gene sets, can be used to explore gene set enrichment on a genome-wide scale, DSigDB [25] is a drug profile database for identifying drugs whose targets are hub genes, drug molecules identified based on hub genes using DSigDB by Enrichr.

## 2.10. Statistical analysis

Graphpad Prism7 software was used to perform the statistical analysis. The mean  $\pm$  SD of at least three separate experiments were used to represent all data. Two sets of t-tests were employed to compare synovial tissue of RA to healthy tissue. The statistical significance was indicated by  $P < 0.05$ .

## 3. Results

### 3.1. Identification of DEGs

After removing batch effects using the sva package (Fig. 2A,B), DEGs were identified in GSE12021, GSE55457, and GSE55235. Table 1 summarizes the characteristics of the three datasets. A total of 679 genes were identified to be differently expressed, with 270 genes downregulated and 409 genes upregulated. A volcano plot of every gene that was up- or down-regulated was used to corroborate the findings (Fig. 2C). The top 20 downregulated and upregulated DEGs based on logFC are shown in the heatmap (Fig. 2D).

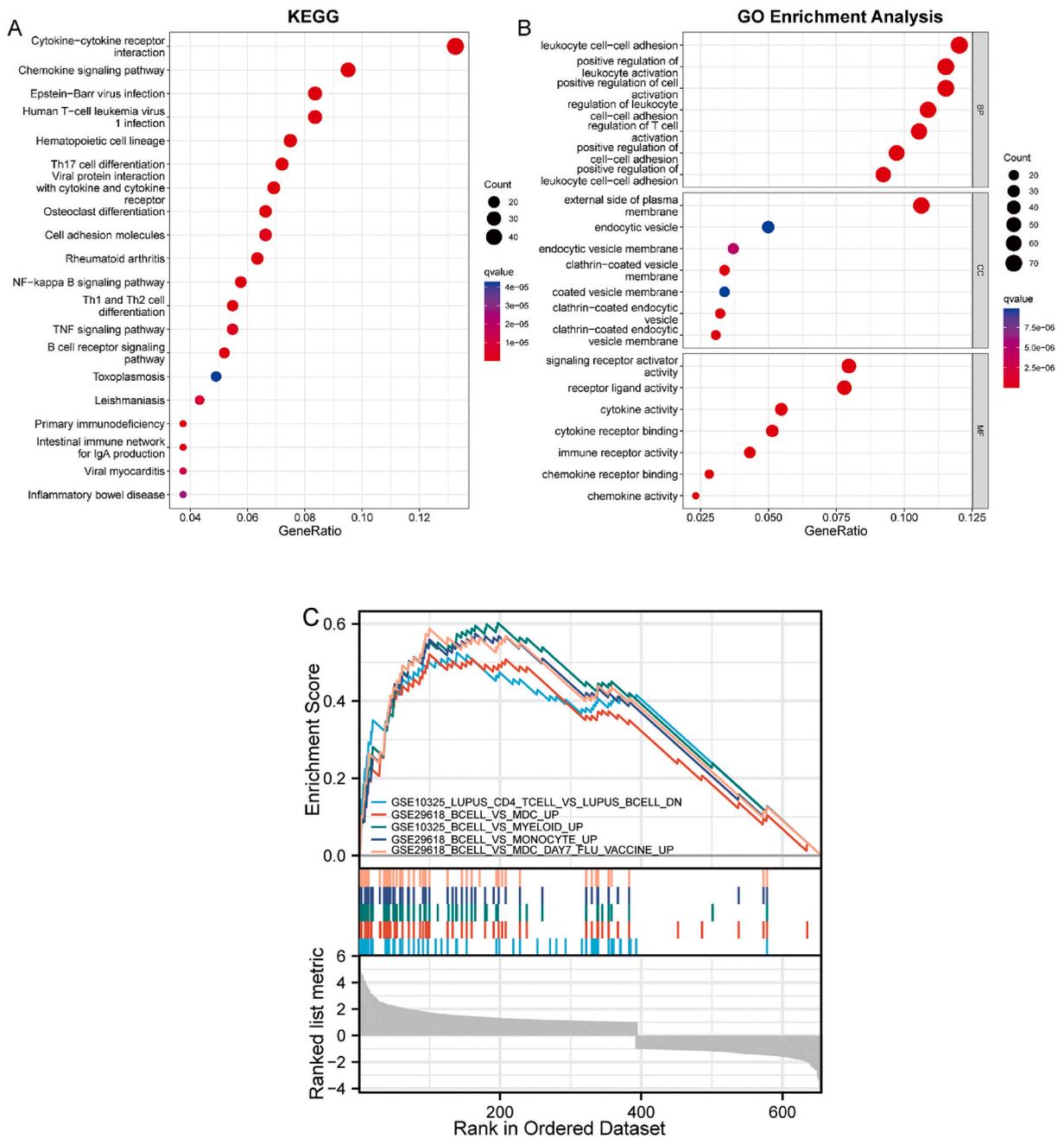
### 3.2. Function enrichment analysis of the DEGs

DEGs were most abundant in hematopoietic cell lineage, Th17 cell differentiation, viral protein interactions with cytokines and their receptors, and cytokine-cytokine receptor interaction, chemokine signaling pathway, rheumatoid arthritis, primary immunodeficiency, osteoclast differentiation, Epstein-Barr virus infection, and B cell receptor signaling pathway, according to KEGG pathway analysis (Fig. 3A). The DEGs played a significant role in the activation of T cells, leukocyte cell-cell adhesion, regulation of leukocyte cell-cell adhesion, positive regulation of leukocyte cell-cell adhesion, positive regulation of cell-cell adhesion, regulation of T cell activation, positive regulation of cell activation, positive regulation of leukocyte activation, regulation of cell-cell adhesion, and differentiation of mononuclear cells, depending on GO functional enrichment analysis (Fig. 3B).

Additionally, the relationship between these DEGs and immunity was discovered using GSEA. Fig. 3C shows that B cells and CD4 T cells, which are crucial in RA immunity, were significantly enriched in the GSEA results.

**Table 1**  
Descriptive statistics.

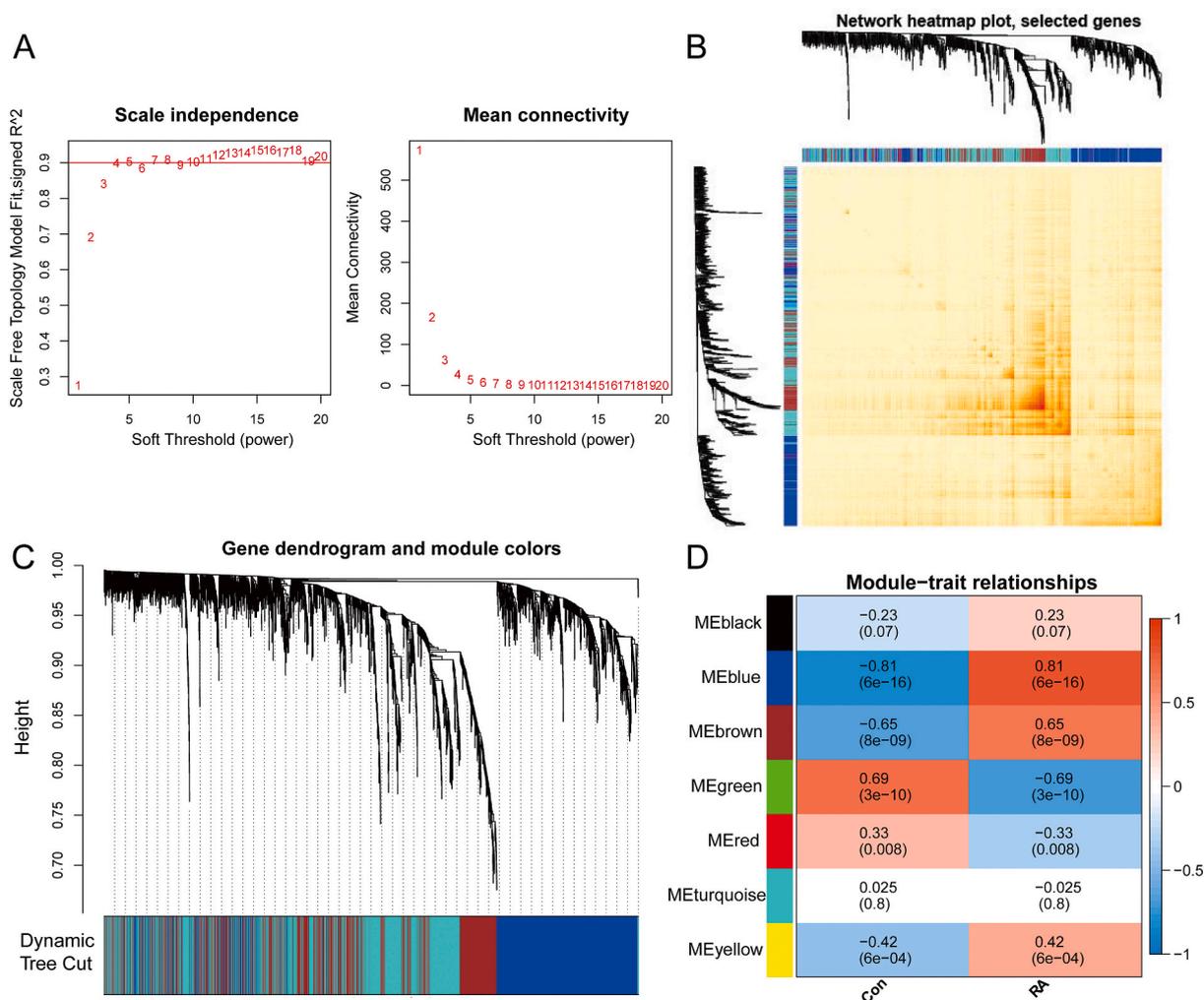
Datasets	Control group		RA group		Affymetrix platform	Species
	Sex (male/female)	Age (years)	Sex (male/female)	Age (years)		
GSE12021	6/3	50.2 $\pm$ 21.9	3/9	64.8 $\pm$ 10.4	GLP96	Homo sapien
GSE55457	8/2	51.0 $\pm$ 19.7	3/9	64.9 $\pm$ 9.9	GLP96	Homo sapien
GSE55235	NA	NA	NA	NA	GLP96	Homo sapien



**Fig. 3.** (A) The enriched KEGG signaling pathways were chosen to show the important biological activities of significant potential mRNA. The abscissa represents the gene ratio, whereas the ordinate depicts the enriched pathways. (B) Analysis of probable mRNA targets using the Gene Ontology (GO) system. The ClusterProfiler utility in R software was used to cluster prospective targets by biological pathways (BP), molecular function (MF), and cellular component (CC).  $q$  value  $< 0.05$  was judged to have statistical significance in the enrichment result. (C) The top five GSEA enrichment analysis of DEGs. DEGs are for differentially expressed genes; GO stands for Gene Ontology; KEGG is for Kyoto Encyclopedia of Genes and Genomes; GSEA stands for gene set enrichment analysis.

### 3.3. WGCNA analysis

After normalizing the data, the top 25% genes with the largest fluctuations were selected for WGCNA analysis. A scale-free coexpression network (scale-free  $R^2 > 0.8$ ) was established with a power of soft thresholding = 4 (Fig. 4A). Hierarchical clustering split with minimum module size set to 50 and deepSplit set to 2, then 7 gene co-expression modules were constructed (Fig. 4C). Topological overlap matrix (TOM) of all genes included in the module is visualized as a heatmap (Fig. 4B). Calculated and plotted the



**Fig. 4.** From merged datasets, coexpression modules are identified by WGCNA. (A) Calculation of the soft thresholding value for scale-free coexpression networks. (B) Cluster dendrogram of the modules identified. (C) Gene interactions in coexpression modules. (D) Module correlations with RA patients and controls.

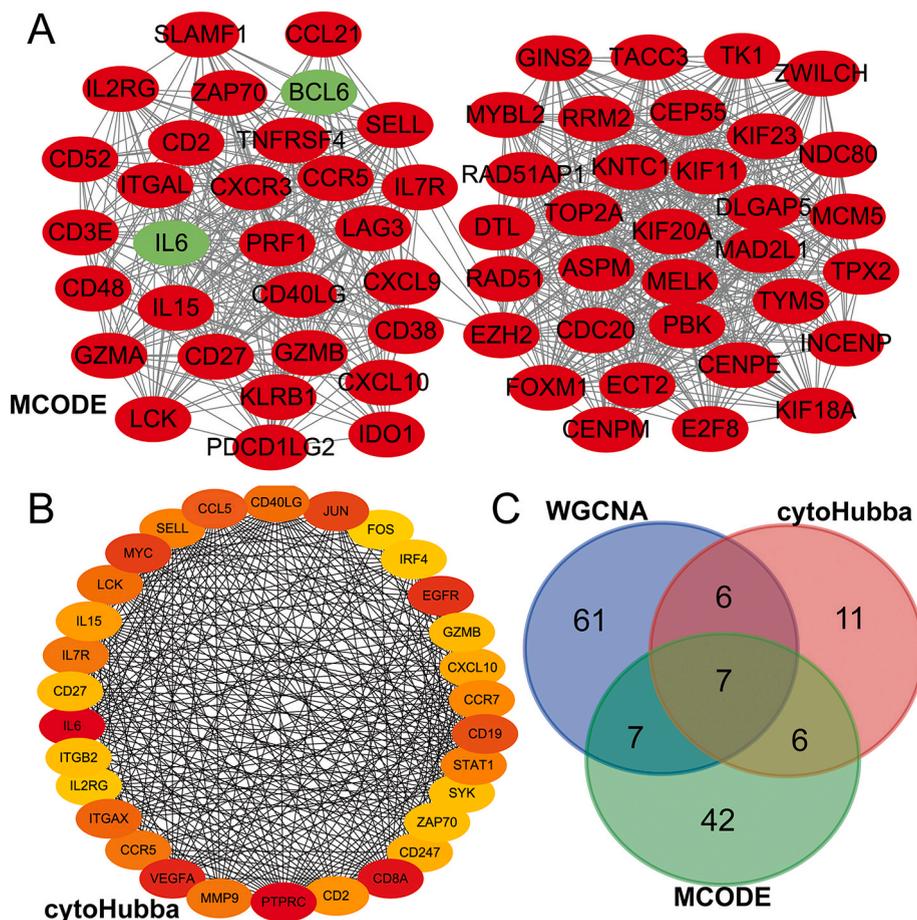
correlation between each module and RA, finally the analysis showed that the blue module containing 81 (Table S2) genes was highly associated with RA positivity ( $r = 0.81$ ,  $P = 6e-16$ ) (Fig. 4D).

### 3.4. Constructing a PPI network and analyzing modules

The PPI network of DEGs was investigated using STRING. From the PPI network, which contained 62 hub genes, the most significant module (PPI score = 26.29) was extracted using the MCODE plugin of Cytoscape (Fig. 5A). The Degree algorithm of cytoHubba was used to extract 30 hub genes from the PPI network (Fig. 5B). Seven common genes were obtained by PPI network and WGCNA (Fig. 5C). Table 2 lists the information of 7 hub genes.

### 3.5. Immune infiltration analyses

Initially, we employed the CIBERSORT approach to examine the differences in immune infiltration between RA and healthy synovial tissues in 22 immune cell subpopulations. The gene expression profiles of 29 healthy individuals and 35 RA patients are presented in Fig. 6A. The proportion of immune cells in the synovial tissues of RA and healthy controls showed significant group-bias clustering and individual differences based on PCA (Fig. 6B). In comparison to normal tissue, RA synovial tissue had higher levels of naive B cells ( $P = 0.023$ ), memory B cells ( $P = 0.005$ ), plasma cells ( $P = 0.001$ ), CD8 T cells ( $P < 0.001$ ), activated memory CD4 T cells ( $P = 0.013$ ), follicular helper T cells ( $P < 0.001$ ), gamma delta T cells ( $P = 0.001$ ), and M1 macrophages ( $P < 0.001$ ), whereas the percentages of resting memory CD4 T cells ( $P < 0.001$ ), resting NK cells ( $P = 0.019$ ), activated NK cells ( $P < 0.001$ ), resting Dendritic



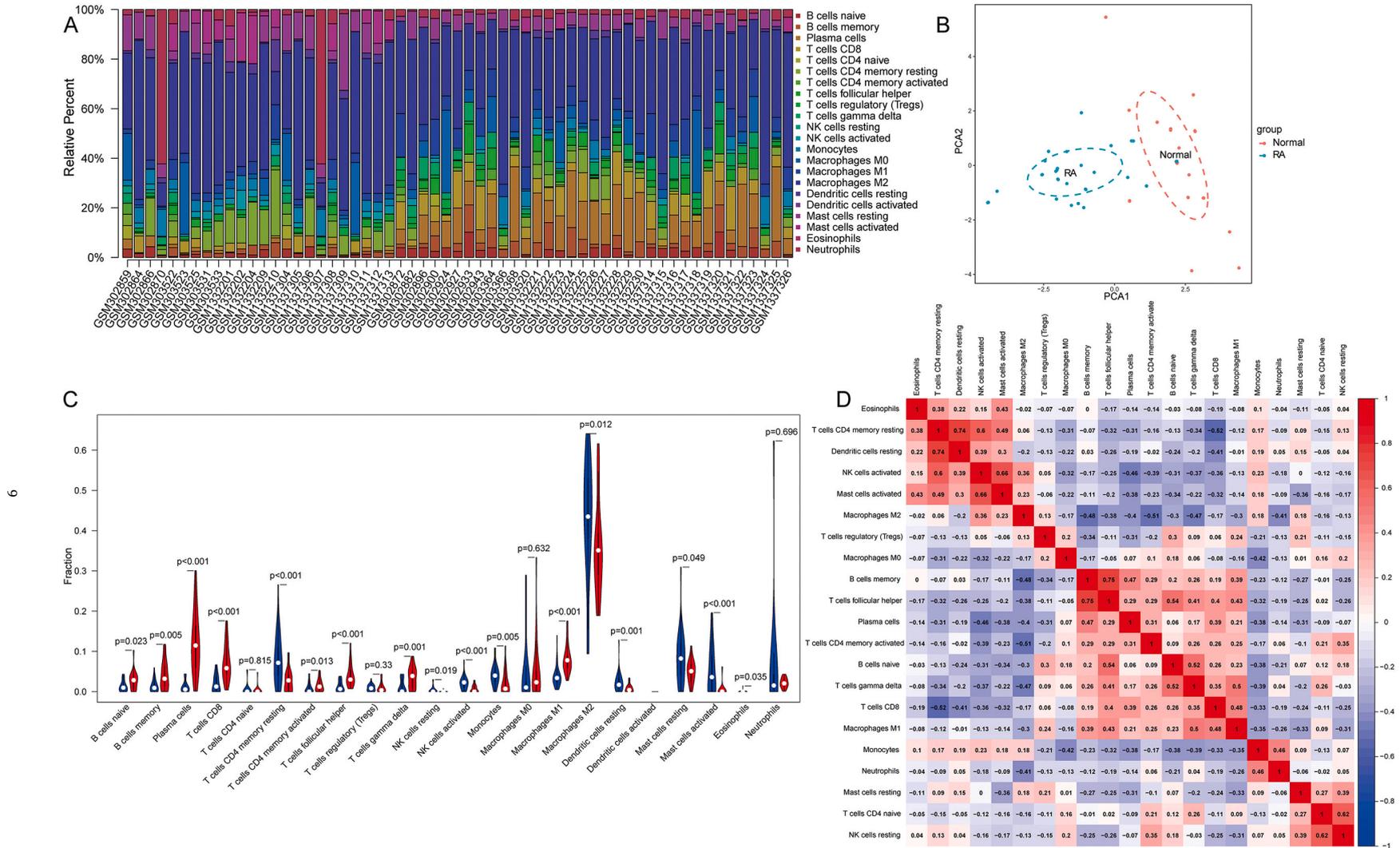
**Fig. 5.** Construction of the PPI network and module analysis of DEGs between Rheumatoid arthritis and normal controls. (A) The most significant module was extracted from the PPI network through the MCODE plugin of Cytoscape. Upregulated genes are marked in red; downregulated genes are marked in green. (B) The hub genes were extracted from the PPI network through Degree algorithms of cytoHubba. (C) MCODE plugin, Degree algorithms of cytoHubba and hub genes of MMBrown module obtain a Venn diagram of common hub genes. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

**Table 2**  
The information of 7 hub genes.

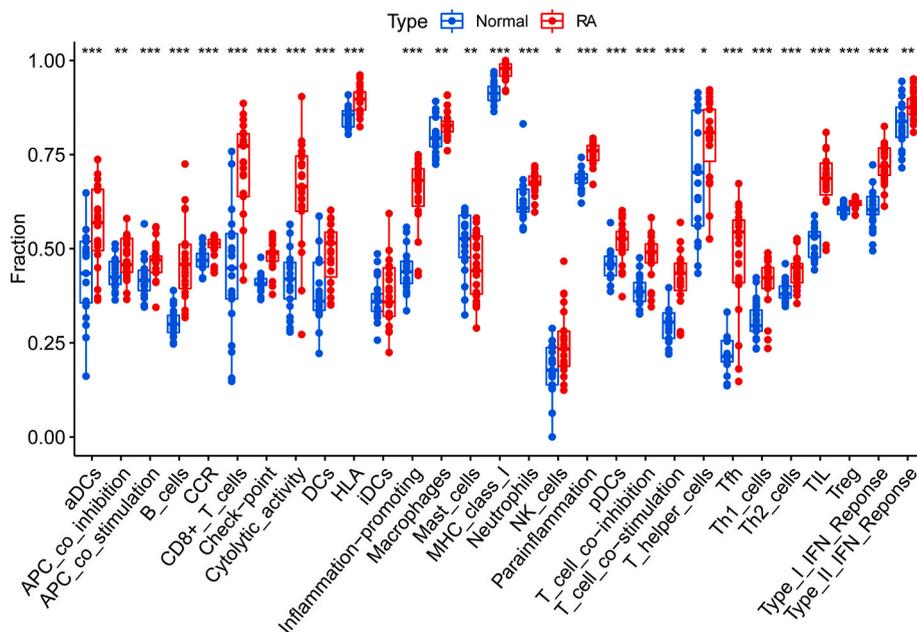
Gene	Gene ID	Full name
IL7R	3575	Interleukin 7 Receptor
CD27	939	CD27 Molecule
CXCL10	3627	C-X-C Motif Chemokine Ligand 10
LCK	3932	LCK Proto-Oncogene, Src Family Tyrosine Kinase
CD2	914	CD2 Molecule
GZMB	3002	Granzyme B
IL2RG	3561	Interleukin 2 Receptor Subunit Gamma

cells ( $P = 0.001$ ), resting Mast cells ( $P = 0.019$ ), activated Mast cells ( $P < 0.001$ ), and M2 macrophages ( $P = 0.012$ ) were relatively lower (Fig. 6C). In the subsequent analysis, the relationships between these immune cell types were analyzed, with heatmaps showing significant positive correlations between memory B cell and activated memory CD4 T cells ( $r = 0.75$ ), resting dendritic cells and resting memory CD4 T cells ( $r = 0.74$ ), and significant negative correlations between CD8 T cells and resting memory CD4 T cells ( $r = -0.52$ ), activated memory CD4 T cells and M2 macrophages ( $r = -0.51$ ) (Fig. 6D).

The ssGSEA algorithm was applied to microarray data from RA and normal samples to assess the distribution of 29 types of immune infiltration. As shown in Fig. 7, all types of immune infiltration were found to be up-regulated in RA tissues compared to normal tissues except for iDCs, suggesting that RA is a highly immunogenic disease type.



**Fig. 6.** The immune infiltration landscape in rheumatoid arthritis and healthy tissue. (A) The proportion of 22 immune cell subpopulations in 64 samples from the GSE12021, GSE55235, and GSE55457 datasets. (B) All samples were subjected to principal component analysis. The first two major components, which account for the majority of data fluctuation, are displayed. (C) The difference between rheumatoid arthritis and healthy controls in terms of immune infiltration. (The normal controls group was color-coded blue, whereas the rheumatoid arthritis group was color-coded red. Statistical significance was defined as a  $P$  value  $< 0.05$ .) (D) Positive and negative correlations among 22 immune cell types, with red and blue indicating positive and negative correlations, respectively. The absence of any association between the mentioned immune cell types is represented by the color white. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)



**Fig. 7.** Immune gene sets identified by ssGSEA. Based on the ssGSEA algorithm, 16 immune cells and 13 immune-related functions were calculated in rheumatoid arthritis and healthy tissue. (\*  $P < 0.05$ , \*\*  $P < 0.01$ , \*\*\*  $P < 0.001$ , ns, non-significant).

### 3.6. Hub genes associations to immune cell infiltration

Initially, we explored the correlation of hub genes expression levels with infiltrating immune cells obtained by the CIBERSORT algorithm in RA, the results showed that the hub genes expression levels were significantly positively correlated with the infiltration levels of immune cells, such as memory B cells, M1 macrophages, CD8 T cells, and follicular helper T cells, and were significantly negative correlated with the immune infiltration level of M2 macrophages (Fig. 8A). Subsequently, we explored the relationship of hub genes expression levels with immune infiltration state obtained by the ssGSEA algorithm in RA, the results revealed that the hub genes expression levels were significantly positively correlated with most immune infiltration states (Fig. 8B). Both of the above two immune infiltration analysis algorithms revealed that the hub genes are closely related to the immune infiltration state in RA.

### 3.7. Hub gene validation by quantitative RT-PCR

The relative expression levels of 7 hub genes were determined using qRT-PCR, including CD27, LCK, CD2, GZMB, IL2RG, and IL7R, were in alignment with the hybridization of microarrays. However, CXCL10 revealed no statistically significant difference (Fig. 9). Table 3 lists all sample details. ROC curves were used to determine if 6 hub genes could distinguish RA patients from normals. Based on the results, several hub genes (CD27, LCK, CD2, GZMB, IL2RG, and IL7R) demonstrated strong biomarker potential for RA (AUC > 0.8) (Fig. 10).

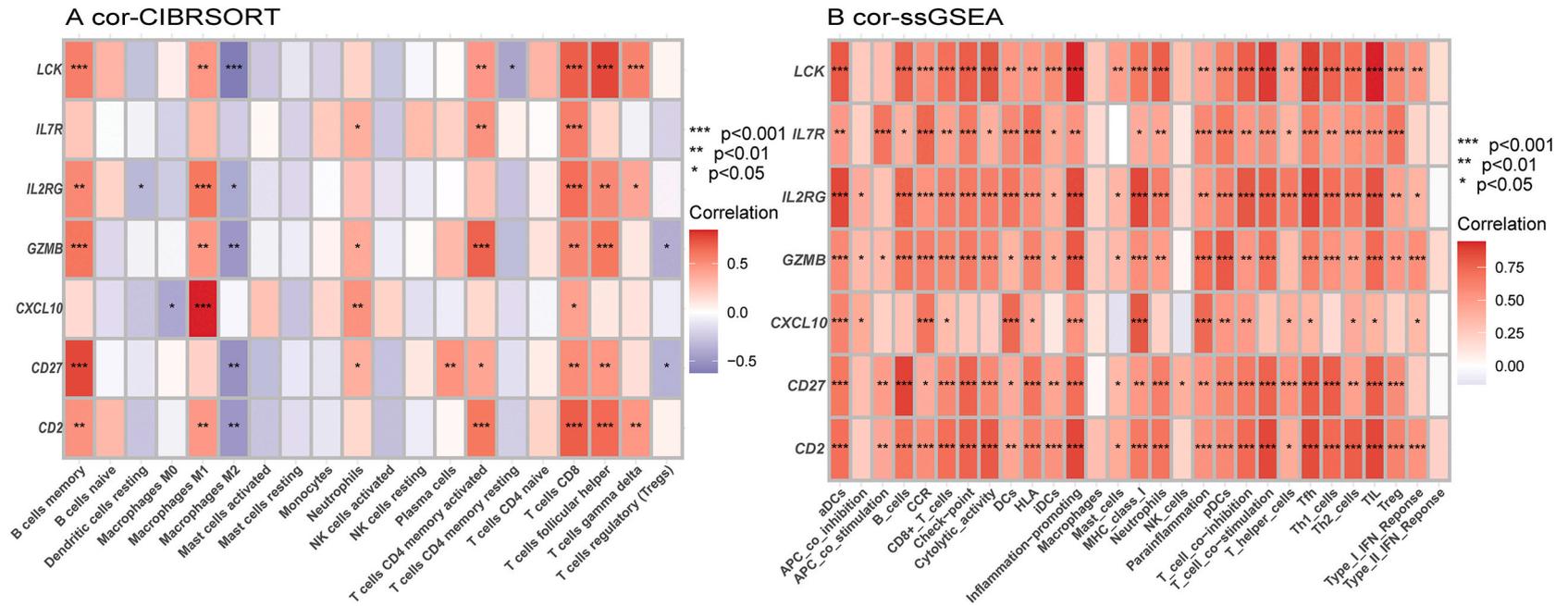
### 3.8. Gene-drug interaction network analysis and candidate drugs identification

To explore the interaction relationship between hub genes and therapeutic drugs, a central gene-drug interaction network was constructed using DrugCentral and visualized by Cytoscape. According to Fig. 11, various drugs can affect the expression of hub gene LCK. For example, astemizole and carbidopa decrease LCK expression level, while nintedanib and sorafenib increase LCK expression level (Fig. 11A), alefacept increase CD2 expression level (Fig. 11B). Drug development increasingly relies on protein-drug interaction networks [26], We identified 21 possible drug molecules based on transcriptome signatures in the DSigDB database using Enrichr. The top 10 drug were extracted based on their Adjusted P-value. These potential drugs are recommended for hub gene. These drugs may be used in the treatment of RA. Table 4 shows the effective drugs of the hub gene in the DSigDB database.

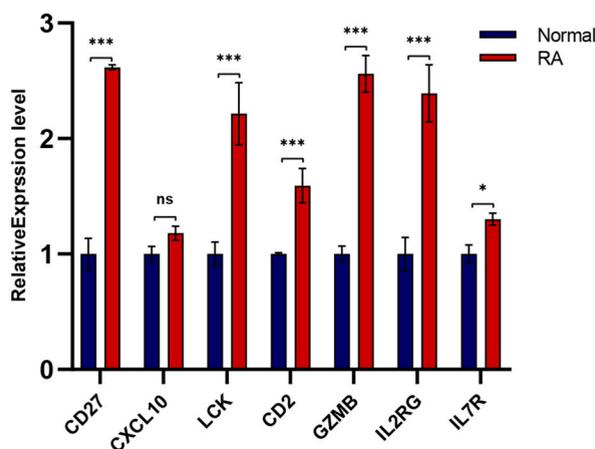
## 4. Discussion

In our study, we attempted to discover critical indicators of improperly expressed genes and immune infiltration associated to synovitis by comparing the differences in synovial gene expression patterns between RA patients and healthy controls, and to give diagnostic and therapeutic targets for RA.

Bioinformatics analysis identified out 679 DEGs, comprising 270 downregulated genes and 409 upregulated genes, and the



**Fig. 8.** The correlation between hub genes and the immune status in RA (A) The proportion of infiltrated immune cells was obtained by CIBERSORT algorithm analysis. (B) The proportion of infiltrated immune cells was obtained by ssGSEA algorithm analysis. Positive or negative correlations between the hub genes and 22 immune cell types, with red and blue indicating positive and negative correlations, respectively. The absence of any association between the mentioned immune cell types is represented by the color white. (\*  $P < 0.05$ , \*\*  $P < 0.01$ , \*\*\*  $P < 0.001$ ). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)



**Fig. 9.** Validation of the hub genes in RA and normal tissue using qRT-PCR. All studies were carried out in triplicate, with the findings provided as  $M \pm SD$ . (\*  $P < 0.05$ , \*\*  $P < 0.01$ , \*\*\* $P < 0.001$ , ns, non-significant).

**Table 3**

Sample details.

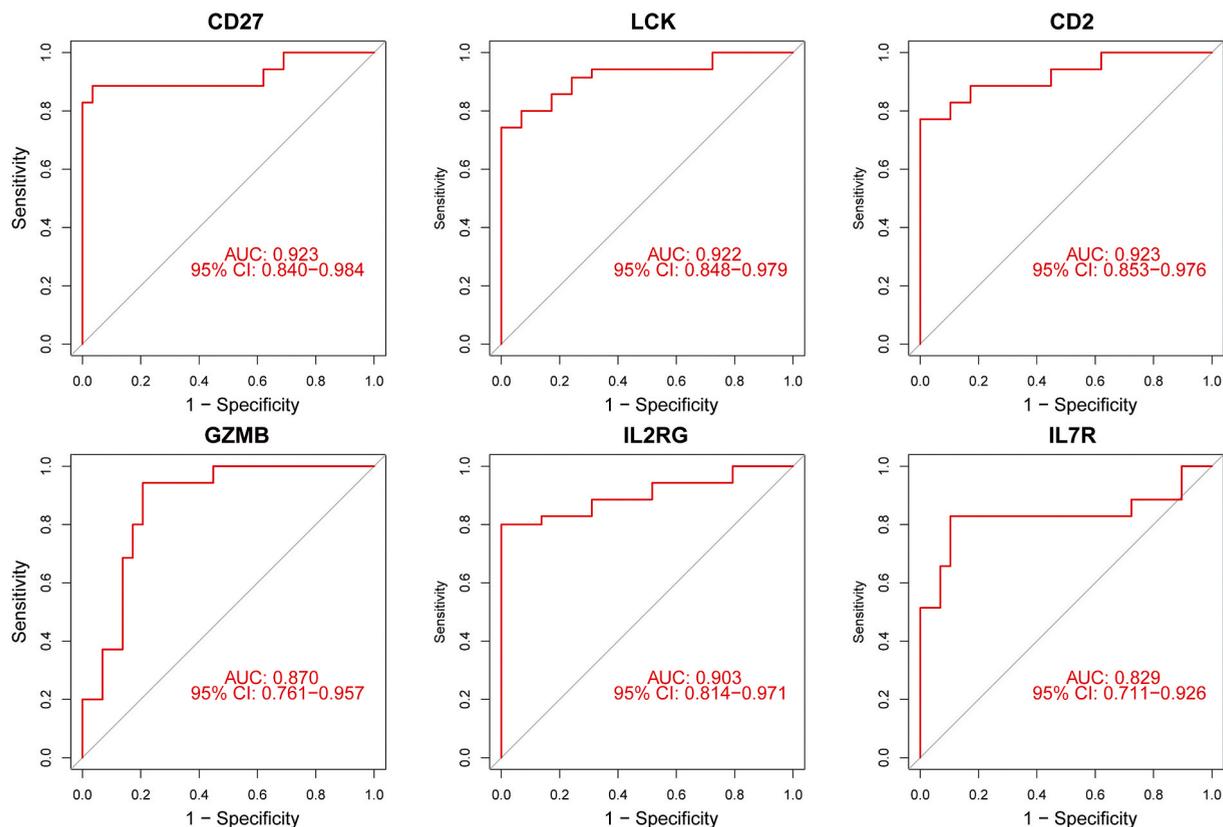
Sample	Age (years)	Sex	Diagnosis	Surgical approach
Control_1	53	Male	Meniscus damage	Arthroscopic surgery
Control_2	36	Female	Meniscus damage	Arthroscopic surgery
Control_3	49	Male	Cruciate ligament injury	Arthroscopic surgery
RA_1	67	Female	RA	Total knee arthroplasty
RA_2	61	Female	RA	Total knee arthroplasty
RA_3	64	Male	RA	Total knee arthroplasty

signaling pathways and biological function involved were investigated in this study. For the genes significantly downregulated, the overexpression of FKBP5 promotes osteoclast differentiation in bone marrow (BM) CD34 (+) cells, and leading to bone destruction in RA [27]. Splice variants of FosB can promote osteoblast differentiation by regulating transcriptional target genes in the osteoblast lineage [28]. In collagen-induced arthritis synovial tissue, PCK1 was found to be down-regulated by qRT-PCR [29]. KEGG analyses revealed these DEGs to be primarily enriched in the Cytokine-cytokine receptor interaction and Th17 cell differentiation signaling pathways, both of which are inflammatory. GO enrichment analyses suggested the DEGs to be mainly correlated with T cell activation and regulation of leukocyte cell-cell adhesion. B cells and CD4 T cells were highly enriched, according to GSEA analysis. Studies have shown that T lymphocytes are key pathogenic effector cells in RA [30].

In this study, we performed WGCNA analysis after merging three RA-related datasets downloaded from the GEO database, seven modules were obtained after hierarchical cluster analysis. Ultimately, 81 genes in the blue module were found to be most closely associated with RA ( $r = 0.81$ ,  $P = 6e-16$ ) (Fig. 4D). In addition, the construction of protein-protein interaction networks, which group and organize all protein-coding genes in a genome, has been demonstrated to be effective in the investigation of a multitude of diseases. The most significant PPI module was investigated using the MCODE and cytoHubba plugins of cytoscape, then take the intersection with the key modules obtained by WGCNA analysis, identifying 7 hub genes including as IL7R, CD27, CXCL10, LCK, CD2, GZMB, and IL2RG. The reliability of 6 hub genes (CD27, LCK, CD2, GZMB, IL7R, and IL2RG) obtained by bioinformatics analysis was supported by qRT-PCR verification. According to the ROC curve, the above 6 hub genes ( $AUC > 0.8$ ) had a strong ability to distinguish RA from control.

In order to further investigate the possibility of the above six hub genes as potential therapeutic targets for RA, we analyzed the interactions between 6 hub genes and existing therapeutic drugs using drug database, and found that multiple drugs may affect the expression of the hub genes. However, whether these hub genes can be used as therapeutic targets for RA still needs further experimental support.

The immune system, which typically fights infection, assaults the joint lining in people with RA, causing the joints to become inflamed, swollen, stiff, and painful [31]. Inflammatory cytokines in serum, including TNF- $\alpha$ , IL-6 and IL-8, are closely associated with the main pathological features of RA [32,33]. LCK, or lymphocyte-specific protein tyrosine kinase, is a critical regulator of T cell activation and differentiation [34,35], autoimmune disease can be caused by mutations in the LCK gene [36]. Inflammatory immune diseases such as RA can be effectively treated with LCK inhibitors [37]. Another study demonstrated that CD27, a kind of cluster differentiation antigens on immune cells, which up-regulated expression suggest the activation of T cells under pathological conditions [38]. CD2, a particular flag protein found on the surface of T and NK cells which discovered by topological analysis [39], was a crucial node gene controlling immunological diseases throughout the pathogenesis of RA. CD2 functioned as a receptor, transmitting signals into cells that activated T and NK cell activity [40,41]. Through the MAPK signaling pathway, GZMB gene silencing protects against



**Fig. 10.** ROC curve for the 6 hub genes that are specifically expressed.

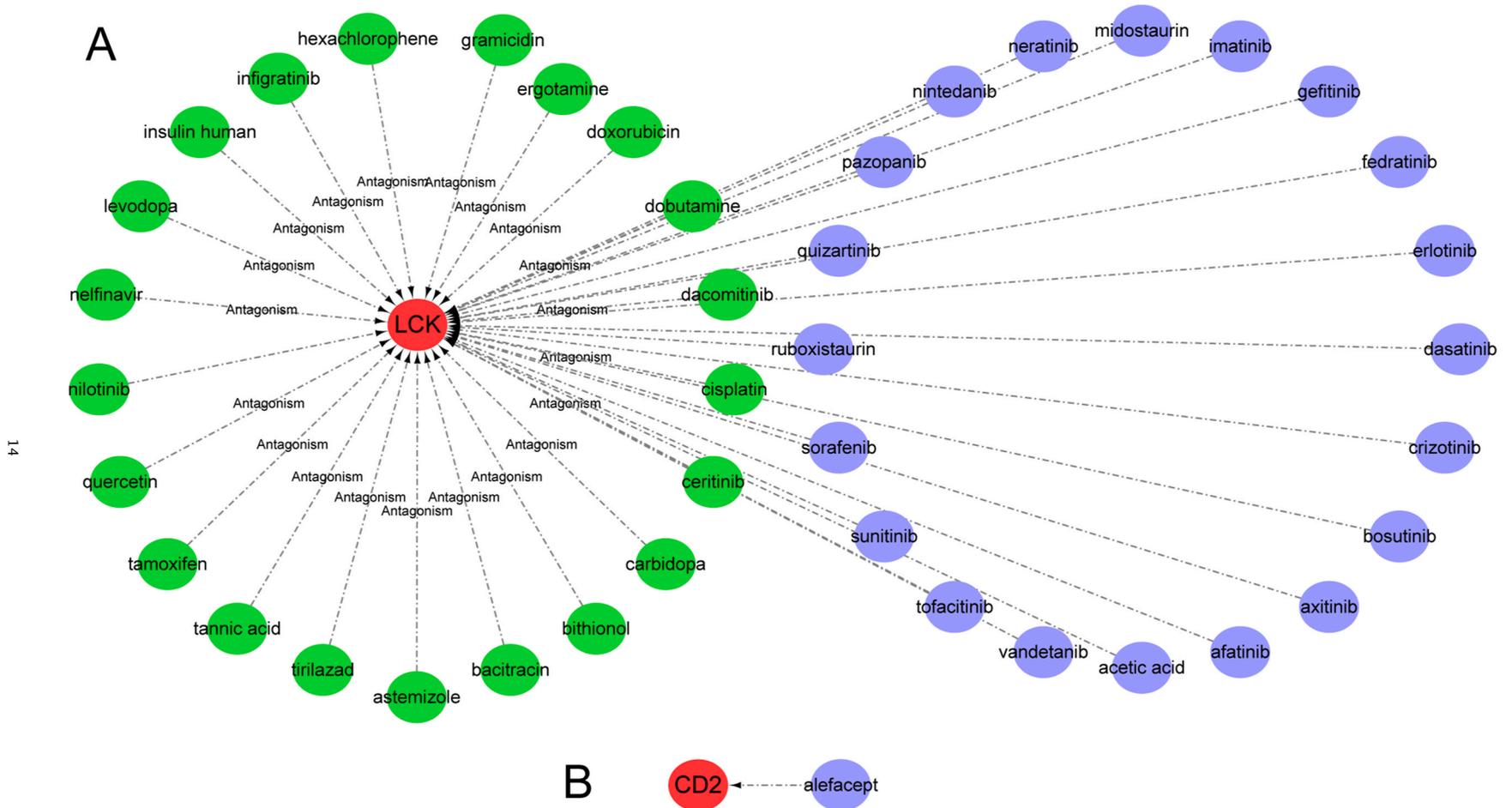
synovial tissue hyperplasia and articular cartilage tissue damage in rheumatoid arthritis [42]. Linkage of IL-7 to IL-7R enhances monocyte homing and angiogenesis, thereby promoting the progression of RA [43]. These studies are consistent with our results obtained through bioinformatics analysis.

The CIBERSORT algorithm was used to analyze immune cell infiltration in RA synovial, and the results showed a significant difference between RA synovial and normal controls in terms of immune cell infiltration. The result showed that there was a significant difference between the synovial in RA patients and healthy controls in terms of the relative cell content of naive B cells, memory B cells, plasma cells, CD4 T cells, activated memory CD4 T cells, follicular helper T cells, gamma delta T cells, and M1 macrophages (Fig. 6C). The ssGSEA algorithm was applied to assess the distribution of 29 types of immune infiltration in RA and normal samples. The result revealed that almost all types of immune infiltration were found to be up-regulated in RA tissues compared to normal tissues except for iDCs (Fig. 7). Interestingly, a single-cell RNA-seq study of RA found four synovial fibroblast (SF) clusters with respective marker genes, and assessed the correlation of SF cluster proportion with RA pathological process and clinical symptoms through deconvolution analysis, whose lymphoid pathotype results were similar to those of the immune infiltration analysis in our study, with significant enrichment of myeloid cells, B cells, T cells, and plasma cells [44]. These findings suggest that RA is a highly immunogenic disease type. Whereas, the specific effect of these differentially expressed chemokines on the immune infiltration of the synovial needs further research.

In addition, our study found a high correlation between the hub genes expression and immune infiltrating cells. Previous studies have found that the immunological interactions of citrulline-containing self-proteins can be activated by interactions between B cells, dendritic cells, and T cells [45]. Studies have found that inflammatory responses and B cells differentiation are regulated by CD4<sup>+</sup> memory T cells in the synovium, which can produce RF or ACPA to exacerbate RA symptoms [46]. Under homeostatic conditions, dendritic cells control the production of inflammatory cytokines and modulate tolerogenic T cell response to modulate the pathological process of RA [47]. These studies suggest that immune cell infiltration plays an important role in the pathogenesis of RA.

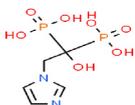
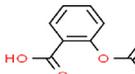
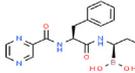
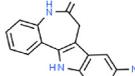
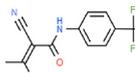
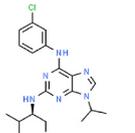
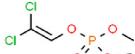
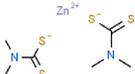
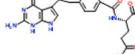
## 5. Conclusions

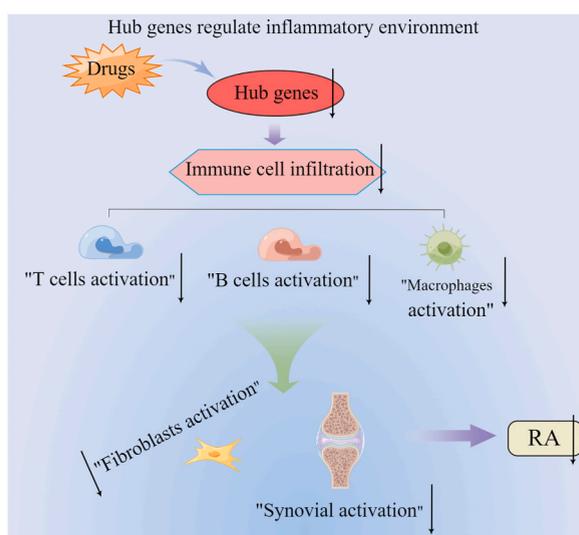
Conclusively, bioinformatics analysis and qRT-PCR validation have revealed that the hub genes LCK, CD27, CD2, IL7R, GZMB, and IL2RG have the potential to be exploited as RA therapeutic targets (Fig. 12). In spite of this, there are some limitations to our study. Firstly, some samples in the GEO database lacked clinical information, resulting in biased bioinformatics analyses. Next, the difficulties in obtaining synovial tissue samples for qRT-PCR validation lead us to only collect three cohorts, which may result in biased results.



**Fig. 11.** Gene-drug interaction network constructed from hub genes and drugs based on DrugCentral. Available drugs that increase or decrease mRNA or protein expression of hub genes are shown in panels A and B. Hub genes are marked in red, drugs increase the expression of hub genes are marked in blue; drugs decrease the expression of hub genes are marked in green. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

**Table 4**  
List of the suggested drugs for RA.

Name	Adjusted P-value	Chemical	Structure	Genes
Zoledronic acid CTD 00003127	0.025887865	C <sub>5</sub> H <sub>10</sub> N <sub>2</sub> O <sub>7</sub> P <sub>2</sub>		GZMB; CD27; IL2RG
aspirin CTD 00005447	0.03192734	C <sub>9</sub> H <sub>8</sub> O <sub>4</sub>		LCK; IL7R; IL2RG
Bortezomib CTD 00003736	0.037342525	C <sub>19</sub> H <sub>25</sub> BN <sub>4</sub> O <sub>4</sub>		GZMB; CD27; IL7R
alsterpaullone CTD 00003709	0.037343	C <sub>16</sub> H <sub>11</sub> N <sub>3</sub> O <sub>3</sub>		LCK
Teriflunomide CTD 00002799	0.037343	C <sub>12</sub> H <sub>9</sub> F <sub>3</sub> N <sub>2</sub> O <sub>2</sub>		LCK
purvalanol A CTD 00004150	0.037342525	C <sub>19</sub> H <sub>25</sub> ClN <sub>6</sub> O		LCK
dichlorvos CTD 00004426	0.037342525	C <sub>4</sub> H <sub>7</sub> Cl <sub>2</sub> O <sub>4</sub> P		GZMB
ZIRAM CTD 00007014	0.037342525	C <sub>6</sub> H <sub>12</sub> N <sub>2</sub> S <sub>4</sub> Zn		GZMB
Cobalt sulfate CTD 00001238	0.037342525	CoO <sub>4</sub> S		IL7R
Pemetrexed CTD 00003054	0.037343	C <sub>20</sub> H <sub>21</sub> N <sub>5</sub> O <sub>6</sub>		IL7R



**Fig. 12.** The role of targeted drugs to antagonize hub genes expression in the pathogenesis of RA. Targeted drugs antagonize hub genes expression to inhibit immune cell infiltration, thereby inhibiting synovial activation and ultimately inhibiting the occurrence of RA.

Besides, the absence of more comprehensive experimental verification of the hub gene, resulting in a lack of credibility. In future work, we will select the hub gene for further functional mechanism studies.

### Authors' contributions

All authors read and approved the final manuscript. Zhi-wei Feng, Yu-chen Tang, Bin Geng and Ya-yi Xia\* - Conceived and designed the experiments; Zhi-wei Feng, Yu-chen Tang, Xiao-yun Sheng, and Sheng-hong Wang- Performed the experiments; Yao-bin Wang, Zhongcheng Liu, and Jin-min Liu- Analyzed and interpreted the data; Bin Geng and Ya-yi Xia\*-Contributed reagents, materials, analysis tools or data; Zhi-wei Feng and Yu-chen Tang- Wrote the paper.

### Fundings

The National Natural Science Foundation of China (81874017, 81960403 and 82060405); Lanzhou Science and Technology Plan Program (20JR5RA320); Cuiying Scientific and Technological Innovation Program of Lanzhou University Second Hospital (CY2017-ZD02, CY2021-MS-A07).

### Human ethics

Regarding ethical clearances, the following details (including the authorizing body and reference numbers) were provided: The Ethics Committee of the Second Hospital of Lanzhou University had approved the research protocol (2022A-205).

### Declarations

Consent to participate and ethics approval.

### Declaration of competing interest

The authors claim to have no conflicts of interest.

### Availability of data and materials

The datasets used and analyzed during the current study are available from NCBI GEO: GSE12021, GSE55235 and GSE55457. The following datasets from NCBI GEO were used in this work and were analyzed: GSE12021, GSE55235, and GSE55457. The Supplementary Files contain the code.

### Acknowledgments

Zhi-wei Feng and Yu-chen Tang contributed equally to this work as the first authors.

### Appendix A. Supplementary data

Supplementary data related to this article can be found at <https://doi.org/10.1016/j.heliyon.2023.e12799>.

### References

- [1] T. Iwamoto, et al., Molecular aspects of rheumatoid arthritis: chemokines in the joints of patients, *FEBS J.* 275 (18) (2008) 4448–4455.
- [2] G.R. Burmester, et al., Mononuclear phagocytes and rheumatoid synovitis. Mastermind or workhorse in arthritis? *Arthritis Rheum.* 40 (1) (1997) 5–18.
- [3] U. Sack, P. Stiehl, G. Geiler, Distribution of macrophages in rheumatoid synovial membrane and its association with basic activity, *Rheumatol. Int.* 13 (5) (1994) 181–186.
- [4] J.J. Haringman, et al., Synovial tissue macrophages: a sensitive biomarker for response to treatment in patients with rheumatoid arthritis, *Ann. Rheum. Dis.* 64 (6) (2005) 834–838.
- [5] S. Bindu, S. Mazumder, U. Bandyopadhyay, Non-steroidal anti-inflammatory drugs (NSAIDs) and organ damage: a current perspective, *Biochem. Pharmacol.* 180 (2020), 114147.
- [6] Q. Fang, C. Zhou, K.S. Nandakumar, Molecular and cellular pathways contributing to joint damage in rheumatoid arthritis, *Mediat. Inflamm.* 2020 (2020), 3830212.
- [7] D. Yu, et al., Enhanced construction of gene regulatory networks using hub gene information, *BMC Bioinf.* 18 (1) (2017) 186.
- [8] Y. Chen, et al., Machine learning to identify immune-related biomarkers of rheumatoid arthritis based on WGCNA network, *Clin. Rheumatol.* 41 (4) (2022) 1057–1068.
- [9] X. He, et al., Identification and validation of potential hub genes in rheumatoid arthritis by bioinformatics analysis, *Am. J. Transl. Res.* 14 (9) (2022) 6751–6762.
- [10] A.A. Al-Absi, D.K. Kang, Long read alignment with parallel MapReduce cloud platform, *BioMed Res. Int.* 2015 (2015), 807407.
- [11] J.T. Leek, et al., The sva package for removing batch effects and other unwanted variation in high-throughput experiments, *Bioinformatics* 28 (6) (2012) 882–883.
- [12] M.E. Ritchie, et al., Limma powers differential expression analyses for RNA-sequencing and microarray studies, *Nucleic Acids Res.* 43 (7) (2015) e47.

- [13] G. Yu, et al., clusterProfiler: an R package for comparing biological themes among gene clusters, *OMICS* 16 (5) (2012) 284–287.
- [14] A. Subramanian, et al., Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles, *Proc. Natl. Acad. Sci. U. S. A.* 102 (43) (2005) 15545–15550.
- [15] P. Langfelder, S. Horvath, WGCNA: an R package for weighted correlation network analysis, *BMC Bioinf.* 9 (1) (2008) 559.
- [16] D. Szklarczyk, et al., STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets, *Nucleic Acids Res.* 47 (D1) (2019) D607–d613.
- [17] G.D. Bader, C.W. Hogue, An automated method for finding molecular complexes in large protein interaction networks, *BMC Bioinf.* 4 (2003) 2.
- [18] C.H. Chin, et al., cytoHubba: identifying hub objects and sub-networks from complex interactome, *BMC Syst. Biol.* 8 (Suppl 4) (2014) S11.
- [19] P. Shannon, et al., Cytoscape: a software environment for integrated models of biomolecular interaction networks, *Genome Res.* 13 (11) (2003) 2498–2504.
- [20] A.M. Newman, et al., Robust enumeration of cell subsets from tissue expression profiles, *Nat. Methods* 12 (5) (2015) 453–457.
- [21] M.S. Rooney, et al., Molecular and genetic properties of tumors associated with local immune cytolytic activity, *Cell* 160 (1–2) (2015) 48–61.
- [22] K.J. Livak, T.D. Schmittgen, Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method, *Methods* 25 (4) (2001) 402–408.
- [23] O. Ursu, et al., DrugCentral: online drug compendium, *Nucleic Acids Res.* 45 (D1) (2017) D932–d939.
- [24] M.V. Kuleshov, et al., Enrichr: a comprehensive gene set enrichment analysis web server 2016 update, *Nucleic Acids Res.* 44 (W1) (2016) W90–W97.
- [25] M. Yoo, et al., DSigDB: drug signatures database for gene set analysis, *Bioinformatics* 31 (18) (2015) 3069–3071.
- [26] T.T. Ashburn, K.B. Thor, Drug repositioning: identifying and developing new uses for existing drugs, *Nat. Rev. Drug Discovery* 3 (8) (2004) 673–683.
- [27] M. Kimura, et al., Role of FK506 binding protein 5 (FKBP5) in osteoclast differentiation, *Mod. Rheumatol.* 23 (6) (2013) 1133–1139.
- [28] G. Sabatakos, et al., Overexpression of DeltaFosB transcription factor(s) increases bone formation and inhibits adipogenesis, *Nat. Med.* 6 (9) (2000) 985–990.
- [29] Y. Zhao, et al., PGK1, a glucose metabolism enzyme, may play an important role in rheumatoid arthritis, *Inflamm. Res.* 65 (10) (2016) 815–825.
- [30] Z. Wen, et al., N-myristoyltransferase deficiency impairs activation of kinase AMPK and promotes synovial tissue inflammation, *Nat. Immunol.* 20 (3) (2019) 313–325.
- [31] J.A. Singh, et al., Biologic or tofacitinib monotherapy for rheumatoid arthritis in people with traditional disease-modifying anti-rheumatic drug (DMARD) failure: a Cochrane Systematic Review and network meta-analysis (NMA), *Cochrane Database Syst. Rev.* 11 (11) (2016) Cd012437.
- [32] F.M. Brennan, et al., Inhibitory effect of TNF alpha antibodies on synovial cell interleukin-1 production in rheumatoid arthritis, *Lancet* 2 (8657) (1989) 244–247.
- [33] S. Tanida, et al., CCL20 produced in the cytokine network of rheumatoid arthritis recruits CCR6+ mononuclear cells and enhances the production of IL-6, *Cytokine* 47 (2) (2009) 112–118.
- [34] R.J. Brownlie, R. Zamoyska, T cell receptor signalling networks: branched, diversified and bounded, *Nat. Rev. Immunol.* 13 (4) (2013) 257–269.
- [35] R.J. Salmond, et al., T-cell receptor proximal signaling via the Src-family kinases, Lck and Fyn, influences T-cell activation, differentiation, and tolerance, *Immunol. Rev.* 228 (1) (2009) 9–22.
- [36] F.D. Goldman, et al., Defective expression of p56lck in an infant with severe combined immunodeficiency, *J. Clin. Invest.* 102 (2) (1998) 421–429.
- [37] A.K. Farag, et al., Novel LCK/FMS inhibitors based on phenoxyypyrimidine scaffold as potential treatment for inflammatory disorders, *Eur. J. Med. Chem.* 141 (2017) 657–675.
- [38] R.Q. Hintzen, et al., Elevated levels of a soluble form of the T cell activation antigen CD27 in cerebrospinal fluid of multiple sclerosis patients, *J. Neuroimmunol.* 35 (1–3) (1991) 211–217.
- [39] J.J. Yang, et al., Structural biology of the cell adhesion protein CD2: alternatively folded states and structure-function relation, *Curr. Protein Pept. Sci.* 2 (1) (2001) 1–17.
- [40] C. Watzl, E.O. Long, Signal transduction during activation and inhibition of natural killer cells, *Curr. Protoc. Immunol.* (2010). Chapter 11: p. Unit 11.9B.
- [41] J.R. James, R.D. Vale, Biophysical mechanism of T-cell receptor triggering in a reconstituted system, *Nature* 487 (7405) (2012) 64–69.
- [42] C.X. Bao, et al., GZMB gene silencing confers protection against synovial tissue hyperplasia and articular cartilage tissue injury in rheumatoid arthritis through the MAPK signaling pathway, *Biomed. Pharmacother.* 103 (2018) 346–354.
- [43] Z. Chen, et al., The novel role of IL-7 ligation to IL-7 receptor in myeloid cells of rheumatoid arthritis and collagen-induced arthritis, *J. Immunol.* 190 (10) (2013) 5256–5266.
- [44] R. Micheroli, et al., Role of synovial fibroblast subsets across synovial pathotypes in rheumatoid arthritis: a deconvolution analysis, *RMD Open* 8 (1) (2022) e001949.
- [45] I.B. McInnes, G. Schett, The pathogenesis of rheumatoid arthritis, *N. Engl. J. Med.* 365 (23) (2011) 2205–2219.
- [46] J.S. Smolen, et al., Rheumatoid arthritis, *Nat. Rev. Dis. Prim.* 4 (1) (2018), 18001.
- [47] L. Schittenhelm, et al., Dendritic cell integrin expression patterns regulate inflammation in the rheumatoid arthritis joint, *Rheumatology* 60 (3) (2020) 1533–1542.