New Classification of Rheumatoid Arthritis Based on Immune Cells and Clinical Characteristics

Jiaqian Wang1, Yuan Xue2, Liang Zhou3

1Department of Orthopaedic, Zhongshan Hospital, Fudan University, Shanghai, 200032, People’s Republic of China; 2Department of Orthopaedic, Wuxi Ninth People’s Hospital of Soochow University, Wuxi, 214000, People’s Republic of China; 3Department of Orthopaedic, Lianshui County People’s Hospital, Hualian, People’s Republic of China

Correspondence: Jiaqian Wang, Email 1580443814@qq.com

Background: Rheumatoid arthritis (RA) is a chronic systemic immune disease characterized by joint synovitis, but there are differences in clinical manifestations and serum test results among different patients. Methods: This is a bioinformatics study. We first obtained the gene expression profile of RA and normal synovium from the database, and screened the differentially expressed immune related genes for enrichment analysis. Subsequently, we classified RA into three subtypes by unsupervised clustering of serum gene expression profiles based on immune enrichment scores. Then, the enrichment and clinical characteristics of different subtypes were analyzed. Finally, according to the infiltration of different subtypes of immune cells, diagnostic markers were screened and verified by qRT-PCR.

Results: C1 subtype is related to the increase of neutrophils, C-reactive protein and erythrocyte sedimentation rate, and joint pain is more significant in patients. C2 subtype is related to the expression of CD8+ T cells and Tregs, and patients have mild joint pain symptoms. The RF value of C3 subtype is higher, and the expression of various immune cells is increased. CD4 T cells, NK cells activated macrophages M1 and neutrophils are immune cells significantly infiltrated in synovium and serum of RA patients. IFNGR1, TRAC, IFITM1 can be used as diagnostic markers of different subtypes.

Conclusion: In this study, RA patients were divided into different immune molecular subtypes based on gene expression profile, and immune diagnostic markers were screened, which provided a new idea for the diagnosis and treatment of RA.

Keywords: rheumatoid arthritis, molecular subtype, immune cells, immunodiagnostic markers

Introduction

Rheumatoid arthritis (RA) is a chronic systemic autoimmune disease characterized by joint synovitis. Persistent synovitis can lead to irreversible damage of intra-articular cartilage and bone, joint dysfunction, and even disability.1 The Global Burden of Disease research report showed that the prevalence of RA increased by 7.4% and the incidence rate increased by 8.2% between 1990 and 2017.2 However, the exact cause of RA has not yet been clarified, environmental, infectious, genetic and other factors are closely related.3 Among them, it is widely recognized that immune factors play a crucial role in the whole process.4 Studies have shown that the infiltration of immune cells in synovium seriously affects the occurrence and progression of RA.5 For example, macrophages can secrete a large number of cytokines, chemokines and degrading enzymes, which are positively correlated with the severity of the disease.6 Activated CD4+ T cells can initiate specific immune responses, leading to joint inflammation and bone destruction.7 Other immune cells, such as natural killer cells, also play a regulatory role in RA.8 Therefore, it is very important to evaluate the infiltration of immune cells in synovium and determine the differences of immune cell components in order to clarify the pathogenesis of RA.

In addition, although the basic pathological change of RA is synovitis, some patients show alternate onset and relief of joint pain, while others show persistent swelling and joint pain, with almost no relief process.9 Even some patients have extra-articular manifestations, with 14.7% of patients developing interstitial lung disease.10 Therefore, early diagnosis and classification of RA is of great significance for the treatment of RA.11 However, the current role of serological diagnostic markers is limited. Even if rheumatoid factor and anti-cyclic citrulline antibody are positive, RA cannot be diagnosed, and it still needs to be based on
long-term clinical symptoms. C-reactive protein (CRP) and erythrocyte sedimentation rate (ESR) increased in the acute stage of RA, but there was no specificity. This brings great difficulties to the early diagnosis of RA, but the first change in RA patients is usually at the molecular level. With the development and widespread use of microarray and high-throughput sequencing technology, we can early identify changes in gene expression and screen out immune diagnostic markers.

For the above reasons, we performed unsupervised clustering on the genes expression profile of RA based on the immune related gene set. Enrichment analysis was conducted on different subtypes of RA, and the relationship between clinical characteristics and immune cells of different subtypes was explored, and key genes for different subtypes were further screened. This classification based on immune molecular mechanism and screening of immune markers provide new ideas for the diagnosis and treatment of RA.

Materials and Methods

Datasets Collection and Preprocessing
We searched the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/) with “rheumatoid arthritis” as the keyword. GEO is an international public repository that contains the most non-tumor genomic data uploaded by researchers worldwide. If the data meets the following criteria, it will be included in the analysis: (1) Gene expression profile of homo sapiens; (2) Case–control studies with complete data; (3) Tissue sample size of RA patients is greater than 100. Based on these conditions, a joint synovial data set (GSE89408) and a whole blood dataset (GSE93272) were screened for further study. GSE89408 contains 28 normal control samples and 152 joint synovial samples from RA patients. GSE93272 contains 232 whole blood tissue samples from Japanese Keio University Hospital RA patients. After downloading the raw data, we use the “impute” package to complete the missing values, and use the median to standardize the data.

Evaluation of Immune Cells Infiltration
The infiltration of immune cells in joint synovium and whole blood was evaluated by CIBERSORT algorithm. The algorithm uses LM22 characteristic matrix to define 22 kinds of infiltrating immune cell components, and only retains the data with p value <0.05. The score of immune infiltrating cells of each sample was calculated based on the expression profile of the dataset, and the results were visualized by violin diagram and histogram. The correlation between clinical characteristics and immune cells was analyzed by Pearson method and visualized by correlation heatmap.

Screening of Differential Expressed Immune Related Genes
By comparing the synovial expression datasets GSE89408 of RA patients and normal controls, the differentially expressed genes (DEGs) were identified by “limma” package. The volcanic map shows the difference level of DEGs, and the false discovery rate (FDR) <0.05 and |log2FC| >1 was considered to be statistically significant. The genes with the most significant differences were visualized by heatmap. A total of 1436 immune related genes were downloaded from ImmPort database, and the final DE-IRGs were extracted by Venn diagram intersection.

Functional Enrichment Analysis
The Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis of DE-IRGs were performed by “clusterProfiler” package. Ten significantly different biological processes, cellular components, molecular functions and pathways were screened by p value <0.05.

Gene set enrichment analysis (GSEA) compared the differential expression of different subtypes in whole blood, and screened four significantly different GO biological processes and KEGG pathways in different subtypes by p value <0.05.

Unsupervised Cluster Analysis of Different Subtypes
In order to clarify the different molecular immune subtypes of RA patients, unsupervised clustering was used to analyze the whole blood dataset. We first calculated the enrichment score of each sample in all immune gene sets by using ssGSEA, and obtained the enrichment score matrix. “ConsenseClusterPlus” based on “pam” method was used to cluster the matrix, and the optimal number of clusters was determined according to the empirical cumulative distribution
function diagram and UMAP diagram. Finally, all tissue samples were divided into three subtypes and the clinical characteristics of different subtypes were extracted.

**Weighted Gene Co-Expression Network Analysis**

In order to extract hub genes of different subtypes, we used weighted gene co-expression network analysis (WGCNA). Simply put, we used the goodSamplesGenes method to remove the outliers and samples, and further constructed the scale free co-expression network. After selecting the appropriate soft-threshold β, adjacency is transformed into topological overlap matrix (TOM). Average linkage hierarchical clustering was carried out according to the dissimilarity measure based on TOM, and the most relevant co-expression modules with different subtypes were obtained. The correlation between module and gene expression was calculated to obtain module membership (MM). According to | MM | >0.8, genes with high connectivity in important modules are identified as hub genes.

**Predictive Immunodiagnostic Biomarkers**

The common genes of DE-IRGs in synovium and hub genes of different subtypes in whole blood were used as immunodiagnostic markers. Venn diagram was used to visualize the results. ROC analysis is carried out through the “pROC” package to obtain the area under the curve (AUC). AUC > 0.75 indicates that biomarkers have good diagnostic value.

**qRT-PCR Validation and Statistical Analysis**

In order to further verify the effectiveness of immunodiagnostic markers of different subtypes, whole blood and synovial samples from three patients with confirmed RA who underwent total knee replacement were obtained. Three patients who underwent total knee arthroplasty due to osteoarthritis served as the normal control group. All participants gave written informed consent. All procedures were approved by the medical ethics committee of Lianshui County People’s Hospital and followed the tenets of the Declaration of Helsinki. Quantitative real-time polymerase chain reaction (qRT-PCR) was carried out as the method previously described. All experiments were performed independently in triplicate. Paired t-test and Mann–Whitney U-test were used to assess the significant differences between two groups. GraphPad Prism software (version 5.01; USA) was used for statistical analyses.

**Results**

**Immune Cells Infiltration in Synovium**

We first analyzed the composition of 22 immune cells in RA and normal synovial tissues. T cells CD4 memory resting, macrophage M2, mast cell resting are the main infiltrating immune cells (Figure 1A). Compared with normal samples, RA synovial samples contain more B cells memory, T cells CD4 memory activated, macrophage M1 and neutrophils, while T cells CD4 memory resting, T cells regulatory (Tregs), NK cells, mast cell resting are relatively less, p values < 0.01 (Figure 1B). The correlation heatmap showed that the increased expression of B cells memory was positively correlated with the T cells CD4 memory activated, macrophage M1 was negatively correlated with NK cell activated and mast cell resting, neutrophils were positively correlated with T cells CD4 naive, and were negatively correlated with macrophage M2, correlation coefficient > 0.4 (Figure 1C).

**Screening of DE-IRGs and Enrichment Analysis in Synovium**

In synovial tissue, 8007 up-regulated DEGs and 1741 down-regulated DEGs were found in normal control group and RA patients (Figure 2A). The heat map shows 20 up-regulated genes and 20 down-regulated genes with the most obvious differences (Figure 2B). Through the intersection of DEGs and immune related genes, a total of 467 DE-IRGs were obtained (Figure 2C).

GO enrichment analysis showed that the changes of biological process (BP) of synovial DE-IRGs were mainly concentrated in the cellular response to stimulus and cytokine related immune response (Figure 3A). Cellular components (CC) were concentrated in plasma membranes, and molecular functions (MF) were concentrated in receptor activity and cytokine activity...
Figure 1 Immune cell infiltration in RA and normal synovium. (A) Histogram shows the composition of 22 immune cells in RA and normal synovium. (B) Violin diagram shows the difference of immune cells between RA patients and normal. (C) Correlation heatmap shows the relationship between immune cells. (means no significance, statistically significant difference was *p<0.05, **p<0.01, ***p<0.001, ****p<0.001).

Figure 2 Identification of differential expressed immune related genes. (A) Volcano map shows DEGs of RA and normal in synovium. (B) Heatmap shows the top twenty up-regulated and down-regulated DEGs. (C) DE-IRGs screened by Venn diagram.

https://doi.org/10.2147/JIR.S395566
DovePress
Journal of Inflammation Research 2024:17
Figure 3. Function enrichment analysis of DE-IRGs. (A) Biological process of DE-IRGs. (B) Molecular function of DE-IRGs. (C) Cellular component of DE-IRGs. (D) KEGG pathway of DE-IRGs.

(Figure 3B and C). KEGG pathway analysis showed that NK cell mediated cytotoxicity, T cell receptor signaling pathway, Th17 cell differentiation, B cell receptor signaling pathway and rheumatoid arthritis were the main enrichment pathways (Figure 3D). These results prove the effectiveness of immune enrichment analysis and the importance of immune cells in RA.

Identification of RA Molecular Subtypes in Whole Blood

RA is a systemic immune disease, and diagnosis based on whole blood samples is simpler and more convenient. Therefore, we used immune enrichment scores based on whole blood gene expression profiles for unsupervised clustering. According to the area under the cumulative distribution function curve and the average consistency evaluation within the cluster group, the consistency is better when the specific cluster number k = 3 (Figure 4A and B). A total of 232 RA tissue samples were divided into three subtypes: Cluster 1 (n = 81), Cluster 2 (n = 80), Cluster 3 (n = 71) (Figure 4C). Umap analysis showed that there were differences among the three subtypes, which again proved the reliability of clustering (Figure 4D).

Enrichment Analysis and Clinical Characteristics of Different Subtypes

GSEA analyzed the differences of gene expression profiles among different subtypes, and the results showed that the function of C1 subtype was enriched in neutrophil-related regulatory processes (Figure 5A and B). C2 subtype is closely related to T cells and participates in T cell receptor signaling pathways (Figure 5C and D). C3 subtype is related to
dendritic cell activated and participates in a variety of immune signaling pathways such as systemic lupus erythematosus, autoimmune thyroid disease, intestinal immune network (Figure 5E and F).

According to the clinical characteristic data of different samples, we further compared the differences between different subtypes. Rheumatoid factor, anti-cyclic citrullinated peptide antibody and antinuclear antibody of C3 subtype were increased, among which rheumatoid factor was statistically significant (Figure 6A and B). The C-reactive protein (CRP), erythrocyte sedimentation rates (ESR) and matrix metalloproteinase 3 (MMP3) of C1 subtype were the highest, followed by C3, and C2 subtype was the lowest. The differences were statistically significant (Figure 6C). The visual analog score (VAS) of C1 subtype was significantly higher than that of C2 subtype (Figure 6D). There was no significant difference in the number of 28 tenderness and swelling joints (TJC28, SJC28), drug use between subtypes (Figure 6E and F).

**Immune Cell Infiltration of Different Subtypes in Whole Blood**

We further analyzed the infiltration of different subtypes of immune cells in the whole blood. T cells CD8, T cells CD4 naive and NK cells resting are the main immune cells in whole blood, especially neutrophils (Figure 7A). We found that the expression of T cells CD8 and T cells CD4 naive of C2 subtype was significantly increased. The expression of T cells CD4 memory activated, T cells gamma delta was significantly higher than that of C1 subtype. The neutrophil expression of C1 subtype was significantly higher than that of the other two subtypes (Figure 7B). The correlation heat map showed that CRP, ESR and MMP3 were positively correlated with neutrophils and negatively correlated with NK cells resting (Figure 7C).
Figure 5 Gene set enrichment analysis. (A) GO enrichment analysis of C1 vs C2 subtype. (B) GO enrichment analysis of C2 vs C3 subtype. (C) GO enrichment analysis of C1 vs C3 subtype. (D) KEGG pathway analysis of C1 vs C2 subtype. (E) KEGG pathway analysis of C2 vs C3 subtype. (F) KEGG pathway analysis of C1 vs C3 subtype.

Figure 6 Differences in clinical characteristics of different subtypes. (A) Expression of RF, CCP, ANA in different subtypes. (B) RF positive in different subtypes. (C) Expression of CRP, ESR, MMP3, CRP.DAS28, ESR.DAS28 in different subtypes. (D) D.VAS, PAIN.VAS, SYMPTOM.VAS scores of different subtypes. (E) TJC28, SJC28, CDAI, HAQ, SDAI scores of different subtypes. (F) Number of drugs used by different subtypes. (means no significance, statistically significant difference was *p<0.05, **p<0.01, ***p<0.001).
Screening and Validation of Immunodiagnostic Markers

In order to screen hub genes of different subtypes, a scale-free network was constructed, and the soft threshold was set to 30 (Figure 8A and B). Based on the weighted gene co-expression correlation, 10 modules were obtained after hierarchical cluster analysis (Figure 8C and D). Among all modules, brown module was significantly positively correlated with C1 subtype, red and grey module was significantly positively correlated with C2 subtype, lightcyan module was significantly positively correlated with C3 subtype, r > 0.5 (Figure 8E).

The diagnostic markers of different subtypes were obtained by overlapping hub genes of different subtype-related modules in whole blood with DE-IRGs in synovium. There are five immune hub genes in C1 subtype, four in C2 subtype and two in C3 subtype (Figure 9A). Among them, IFNGR1, NAMPT, PPP3CA are most related to neutrophils, IL21R, LCK, PRKCQ, TRAC are most related to T cells, and IFITM1, TAP1 are most related to dendritic cells activated (Figure 9B). In all 11 immune hub genes, the AUC value of 6 genes is greater than 0.75 (Figure 9C–H).

Based on the correlation between immune cells and AUC values, we determined that IFNGR1 was immunodiagnostic markers of C1 subtype, TRAC was diagnostic marker of C2 subtype, and IFITM1 was diagnostic marker of C3 subtype. These three immune diagnostic markers showed good efficacy in synovium, with significant differences, p < 0.0001 (Figure 10A–C). They also have diagnostic value in serum, but the difference is not as significant as that in synovium (Figure 10D–F).

Discussion

RA is a systemic autoimmune disease, with synovitis and erosive joint destruction as the main clinical manifestations. The immune system not only wrongly attacks normal joint tissue, but also involves extra-articular connective tissue such as serosa, heart, lung and eye. Therefore, our study first explored the infiltration mode of RA immune cells from the synovium of joints. The immune dysfunction of RA is closely related to the abnormality of immune cells. We found that RA synovial samples contain more B cells memory, T cells CD4 memory activated, macrophage M1 and neutrophils, while NK cells, mast cell resting are relatively less. Previous studies have shown that B cells can differentiate into plasma...
cells stimulated by antigens. Plasma cells synthesize and secrete antibodies, form antigen antibody complexes, deposit on the vascular wall and other parts, induce inflammatory reactions, and lead to multiple organ damage throughout the body. In our results, B cells memory is positively correlated with T cells CD4 memory activated. Activated CD4+T cells can also initiate specific immune responses and lead to corresponding arthritis symptoms. In addition, macrophages can secrete a large number of cytokines and differentiate into pro-inflammatory M1 subtypes in the synovitis environment. These cytokines can further lead to joint destruction through neutrophil recruitment and T cell activation. However, early RA is often difficult to diagnose because of its atypical clinical symptoms and lack of specific serum diagnostic markers. Existing articles only analyze the infiltration of immune cells in RA synovium, but RA is a systemic immune disease, and many patients have extra-articular symptoms. We combined gene expression profile data, clinical characteristics and immune cell infiltration to classify RA patients, in order to find more accurate diagnostic markers and drug treatment targets. We found that C1 subtype is enriched in neutrophil-related functions, and the neutrophil expression of this subtype is also significantly higher than that of the other two subtypes. Neutrophils have been found to be related to the occurrence and sustainable development of rheumatoid arthritis. Various chemokines can promote the migration of neutrophils to joints. The activation and proliferation of neutrophils promote cartilage destruction and bone absorption. The relief of RA symptoms is also related to the reduction and absorption of neutrophils in synovial fluid. Bont found that the antibody of neutrophil extracellular trap was related to the increase of RF, which may be a biomarker of RA, which was consistent with our experimental results. We also found that CRP, ESR and MMP3 of C1 subtype were significantly increased, and patients’ joint pain scores were higher. Therefore, we describe C1 as neutrophils associated acute joint pain subtype.

C2 subtype is enriched in T cells related functions, the expression of T cells CD8 and T cells CD4 naive is significantly increased, and T cells CD4 memory activated is also significantly higher than C1 subtype, T cells regulatory (Tregs) is higher than C3 subtype. The expression of RF, CRP, ESR, MMP3 and VAS score of C2 subtype were lower.
Figure 9 Screening immunodiagnostic markers of different subtypes. (A) Venn diagram shows the intersection of different subtype hub genes with DE-IRGs. (B) Heatmap of correlation between immunodiagnostic markers and immune cells. (C–H) Immunodiagnostic markers with AUC > 0.75.

Figure 10 The validity of immunodiagnostic markers was verified by qRT-PCR. (A–C) Expression of RA and normal IFNGR1, TRAC, IFITM1 in synovium. (D–F) Expression of RA and normal IFNGR1, TRAC, IFITM1 in whole blood. (Statistically significant difference was * p<0.05, ** p<0.01, *** p<0.001).
than those of the other two groups. Based on the above results, we described the C2 subtype as T cells related low immune subtype. Tregs cells regulate the immune response by secreting cytokines and reduce the pathological damage caused by immune inflammation. CD8+ T cells have anti-inflammatory properties and help reduce the persistent autoimmune response of rheumatoid joints. Clinical characteristics also proved that C2 subtype may have mild inflammation and better prognosis. C3 subtype is enriched in a variety of immune disease pathways throughout the body, and its expression in immune cells such as B cells memory, T cells gamma delta, macrophages M1, dendritic cells activated are significantly higher than that of the other two subtypes. RF is an autoantibody with denatured IgG as the target antigen, which exists in a variety of autoimmune diseases. RF with high titer often indicates poor prognosis. According to the high expression of RF in C3 subtype, we describe it as variety of immune cells related high immune subtypes, which may be related to the extra-articular symptoms of RA. Dana classifies the gene expression data of RA synovium tissue, which distinguishes high inflammatory subtypes characterized by extensive leukocyte infiltration. The CRP, ESR and joint pain levels of this subtype are high, which once again proves the reliability of our classification.

In order to further study the diagnostic markers of each subtype, combining DE-IRGs in synovium and hub genes of different subtypes, we obtained three immune diagnostic markers. Among them, IFNGR1 was immune diagnostic marker of C1 subtype. IFNGR1 expression is associated with the presence of RA. IFNGR1 (interferon-γ receptor 1), after binding of IFN-γ, signal transduction activates macrophages through JAK-STAT pathway and initiates the secretion of proinflammatory cytokines to play a role in RA. In the correlation heat map, IFNGR1 is related to neutrophils, but its relationship with neutrophils is still being explored. TRAC (TCRα subunit constant gene) mutations can cause hereditary immunodeficiency disorders. TRAC mainly exists on the surface of T cells, and its mutation leads to T cells dysfunction. C2 subtype is associated with T cells, so TRAC was immunodiagnostic marker of C2 subtype. IFITM1 is a membrane protein, which can be induced by interferon and plays a key role in immune response. Interferon (IFN) is a pleiotropic cytokine produced as a defense response in innate immunity and has an immunomodulatory effect. Nehmar found that IFN may promote the production of pathogenic autoantibodies. IFN signals describe a subgroup of RA patients and are related to disease activity score or autoantibodies. This also provides further theoretical support for our experiment.

Conclusion
Rheumatoid arthritis is a systemic immune disease involving T cells CD4, NK cells activated, macrophage M1 and neutrophils. We classified them into high CRP, ESR, neutrophils associated acute pain subtypes, T cells associated low immune subtypes, and multiple immune cells associated high immune subtypes. This provides a new idea for the diagnosis and treatment of RA.

Data Sharing Statement
The data used to support the findings of this study are included within the article.

Ethical Review Statement
This study was approved by the research ethics committee of Lianshui County People’s Hospital.

Consent for Publication
Informed consent was obtained from all individual participants included in the study.

Acknowledgments
We would like to thank Qifan Li for communicating data and information.

Author Contributions
All authors contributed to data analysis, drafting or revising the article, have agreed on the journal to which the article will be submitted, gave final approval of the version to be published, and agree to be accountable for all aspects of the work.
Funding
No benefit in any form has been received or will be received from a commercial party related directly or indirectly to the subject of this article.

Disclosure
The authors declare that they have no conflicts of interest in this work.

References


---

**Journal of Inflammation Research**

**Publish your work in this journal**

The Journal of Inflammation Research is an international, peer-reviewed open-access journal that welcomes laboratory and clinical findings on the molecular basis, cell biology and pharmacology of inflammation including original research, reviews, symposium reports, hypothesis formation and commentaries on: acute/chronic inflammation; mediators of inflammation; cellular processes; molecular mechanisms; pharmacology and novel anti-inflammatory drugs; clinical conditions involving inflammation. The manuscript management system is completely online and includes a very quick and fair peer-review system. Visit http://www.dovepress.com/testimonials.php to read real quotes from published authors.

Submit your manuscript here: https://www.dovepress.com/journal-of-inflammation-research-journal