



CD3D, GZMK, and KLRB1 Are Potential Markers for Early Diagnosis of Rheumatoid Arthritis, Especially in Anti-Citrullinated Protein Antibody-Negative Patients

Junqin Lu^{1†}, Yihui Bi^{2†}, Yapeng Zhu³, Shi Huipeng⁴, Wenxiu Duan^{3*} and Jian Zhou^{1*}

¹Department of Orthopedics, The First Affiliated Hospital of Anhui Medical University, Hefei, China, ²Department of Orthopedics, The Second Affiliated Hospital of Anhui Medical University, Hefei, China, ³Department of Interventional Radiology, The First Affiliated Hospital of USTC, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, China, ⁴Department of Orthopedics, Shanghai Sixth People's Hospital, Shanghai, China

OPEN ACCESS

Edited by:

Tao Xu,
Anhui Medical University, China

Reviewed by:

Yang Yang,
University of Texas Health Science
Center at Houston, United States
Xiaogang Wang,
Shanxi Eye Hospital, China

*Correspondence:

Wenxiu Duan
wxduan9@mail.ustc.edu.cn
Jian Zhou
Drzhoujian2021@163.com

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Inflammation Pharmacology,
a section of the journal
Frontiers in Pharmacology.

Received: 17 June 2021

Accepted: 16 August 2021

Published: 16 September 2021

Citation:

Lu J, Bi Y, Zhu Y, Huipeng S, Duan W
and Zhou J (2021) CD3D, GZMK, and
KLRB1 Are Potential Markers for Early
Diagnosis of Rheumatoid Arthritis,
Especially in Anti-Citrullinated Protein
Antibody-Negative Patients.
Front. Pharmacol. 12:726529.
doi: 10.3389/fphar.2021.726529

Early diagnosis and monitoring of rheumatoid arthritis (RA) progress are critical for effective treatment. In clinic, the detection of rheumatoid factor (RF) and anti-citrullinated protein antibodies (ACPA) are usually combined to diagnose early RA. However, the poor specificity of RF and high heterogeneity of ACPA make the early diagnosis of RA still challenging. Bioinformatics analysis based on high-throughput omics is an emerging method to identify novel and effective biomarkers, which has been widely used in many diseases. Herein, utilizing an integrated strategy based on expression correlation analysis and weighted gene coexpression network analysis (WGCNA), we identified 76 RA-trait different expression genes (DEGs). Combined with protein-protein interaction (PPI) network construction and clustering, new hub genes associated in RA synovia, CD3D, GZMK, and KLRB1, were identified. We verified the specificity of these genes in the synovium of RA patients through three external datasets. We also observed high sensitivity and specificity of them for ACPA-negative patients. CD3D, GZMK, and KLRB1 are potentially key mediators of RA pathogenesis and markers for RA diagnosis.

Keywords: rheumatoid arthritis, bioinformatics, biomarker, ACPA-negative, CD3D, GZMK, KLRB1

INTRODUCTION

Rheumatoid arthritis (RA) is a systemic autoimmune disease characterized by chronic inflammation in the synovium tissue of joints (McInnes and Schett, 2011; Gibofsky, 2014; Malmstrom et al., 2017; Smolen et al., 2018). A variety of immune cells and cytokines are involved in synovial inflammation, which ultimately leads to the destruction of soft tissues, cartilage, and bones around the joints (Zhang et al., 2019). Early RA clearly begins months to years when autoimmune response persists and is seronegative before it becomes a manifest polyarthritis and this is known as “preclinical RA.” The treatment of RA is usually only effective in the early stage, and many patients gradually lose their drug response as the disease progresses (Guo et al., 2018). Thus, early diagnosis is pivotal to optimal therapeutic success. Some serum biomarkers like rheumatoid factor (RF) (MacGregor et al., 2000; Onuora, 2012), anti-citrullinated protein antibody (ACPA) (Padyukov et al., 2011; Stahl et al., 2012), anti-cyclic citrullinated peptide (anti-CCP) antibody (Zendman et al., 2006), C-reactive protein

(CRP) (Lee et al., 2016), and erythrocyte sedimentation rate (ESR) (Orr et al., 2018) have good performance to discriminate part of typical early RA patients, but approximately 30% of patients remain seronegative using current immunoassays. Other biomarkers, such as miRNA (Cunningham et al., 2021), calprotectin (Jarlborg et al., 2020), anti-RA33 (Syed Mohamed Suhail et al., 2019), and anti-carbamylated protein (anti-CarP) antibodies (Lamacchia et al., 2021), may be effective in the diagnosis of early RA, but these markers are not widely accepted in clinical application, and their prognostic significance remains controversial owing to the lower sensitivity and specificity of supplementary diagnosis compared with RF and ACPA. Exploring hub genes and their expression status in the inflamed synovium is a critical step in defining new targets for the diagnosis or treatment of RA, especially for the ACPA-negative RA patients whose condition could not be well managed at the developing stage due to negative index of clinical symptoms and serologic testing.

Application of transcriptomic or microarray analysis to whole synovial tissue has already identified specific genes associated with RA (Lewis et al., 2019). However, in most studies, there are many shortcomings such as small sample size, high sample heterogeneity, and only using a single technology platform and other drawbacks exist in most studies. By effectively integrating new high-throughput data, especially gene expression and proteomic-profiling data, bioinformatics analysis is expected to deliver novel clinical diagnostic insights and therapeutic options at high resolution in an unbiased fashion (Sirota and Butte, 2011; Okada et al., 2014; Stephenson et al., 2018). This strategy has already been used to discover distinct transcriptomic features of synovial fibroblasts and found that MTF1 and RUNX1 may be crucial future targets for RA therapy (Tsuchiya et al., 2020). Bioinformatics analysis also defined anti-PTX3 and anti-DUSP11 autoantibodies as newly identified biomarkers for ACPA-negative RA diagnosis (Li et al., 2021).

Herein, aiming to investigate some new and effective biomarkers for improving the management of early diagnosis of RA, especially for ACPA-negative patients, we integrated multiple databases from different GEO platforms to deeply analyze the characteristic genes of synovial tissue for RA patients. Combining differential expression analysis and weighted gene coexpression network analysis (WGCNA), we screen out 76 RA-trait different expression genes (DEGs). Followed by protein-protein interaction (PPI) network construction and hub genes selection, we identified CD3D, GZMK, and KLRB1 as three novel hub genes with RA characteristics. Independent dataset verification indicated that CD3D, GZMK, and KLRB1 can well distinguish RA patients from normal and osteoarthritis (OA) patients and has high sensitivity and specificity for ACPA-negative RA patients.

MATERIALS AND METHODS

Microarray Data Preparation

Gene expression profiling datasets were obtained from the NCBI-GEO database (<https://www.ncbi.nlm.nih.gov/gds>). Datasets GSE55235 and GSE55457 originating from GEO platform 96

(GPL96) were combined with GSE77298 and GSE153015 datasets originating from GEO platform 570 (GPL570) for subsequent analysis. The four datasets contained 100 joint synovial tissue samples, including 27 normal joint synovial samples, 24 OA joint synovial samples, and 49 RA joint synovial samples. Additionally, three datasets were chosen for verification, which included GSE39340 dataset originated from GPL10558, GSE55584 dataset originated from GPL96, and GSE89408 dataset originating from GPL11154. These three datasets contained 202 joint synovial tissue samples, of which 28 are normal, 22 OA, and 152 RA synovial tissues (Table 1).

Integration and Normalization of Microarray Data

The raw data were downloaded and preprocessed before analysis. Background correction, prosummarization, and missing values supplement for the matrix data of each GEO dataset were performed systematically by the “affy” package and the “impute” package in R/Bioconductor software (version 4.0.3). The combat function of “sva” package was used to correct the batch effects between different datasets. Principal component analysis (PCA) construction via “factoextra” package was performed to discover whether the batch effect was eliminated.

Identification and Analysis of DEGs

Based on the comparison of gene expression values from the RA vs normal groups and the RA vs OA groups, the “limma” package in R was used to identify DEGs. The overlapped DEGs were outlined as RA-trait DEGs when adjusted $p < 0.05$ and $|\log_{2}FC| > 1$ were used as the filter criteria. Furthermore, the DEG expression levels were visualized in heatmaps and volcano plots by the “pheatmap” and “RColorBrewer” packages in R.

Construction of WGCNA

The merged gene matrix was loaded and checked to exclude abnormal samples which might be escape from sample clustering. The merged dataset contains 100 samples and 12,412 genes for further WGCNA analysis. After selecting an appropriate threshold, adjacency and topological overlap matrix (TOM) were established to be a nearly no-scale network. Then, different modules were recognized through dynamically tree cutting with the calculation of cluster dendrogram. According to the significance of the gene coexpression network, gene significance (GS) and module membership (MM) were obtained to investigate specific modules highly related to clinical traits of RA. Finally, according to the threshold values $|GS| > 0.5$ and $|MM| > 0.7$, significant RA-correlated genes were selected as RA-trait module genes.

Identification of RA-Trait DEGs and Gene Ontology/Kyoto Encyclopedia of Genes and Genomes (GO/KEGG) Enrichment Analyses

The overlapped genes calculated by TBtools (version 1.089) between RA-related DEGs and RA-trait module genes were featured as RA-trait DEGs. GO and KEGG pathway analyses

TABLE 1 | Selected GEO datasets for bioinformatics analysis.

GEO datasets	Platform	Samples			Data category	Application
		Normal	OA	RA		
GSE55235	GPL96	10	10	10	Expression profiling by array	Analysis
GSE55457	GPL96	10	10	13	Expression profiling by array	Analysis
GSE77298	GPL570	7	0	16	Expression profiling by array	Analysis
GSE153015	GPL570	0	4	10	Expression profiling by array	Analysis
GSE39340	GPL10558	0	7	10	Expression profiling by array	Verification
GSE55584	GPL96	0	6	10	Expression profiling by array	Verification
GSE89408	GPL11154	28	22	152	Expression profiling by high-throughput sequencing	Verification

were performed via “clusterProfiler” package in R to acquire the enriched cellular component (CC), biological process (BP), molecular function (MF) categories, and functional pathways. The significant enriched functions and pathways were filtered with adjusted $p < 0.05$ and visualized in bubble plots executed by the “ggplot2” package in R.

Construction of PPI Network and Key Genes Analysis

The PPI network was constructed via the online tool STRING (<https://string-db.org/>) based on the RA-trait DEGs. Cytoscape (version 3.8.2) was applied for the better presentation and visualization of the whole interaction information. The most important cluster in the PPI network was identified by Minimal Common Oncology Data Elements (MCODE), and the key genes were further screened by cytoHubba (Cytoscape plugin), which provide 12 different algorithms to rank the importance or core degree of genes.

Box Plot Drawing and Statistical Methods

Gene expression box plots in different data sets were drawn through “ggplot2” package. Wilcoxon test was used between two variables, and the Kruskal–Wallis test was used between multiple variables.

Analysis of Hub Genes Expression and Drawing Receiver Operating Characteristic (ROC) Curve in External Databases

The “ggplot2” and “ggpubr” packages in R were applied for genes expression analysis via drawing boxplots according to the expression values in different validation datasets. The “pROC” package was used to draw ROC curves.

RESULTS

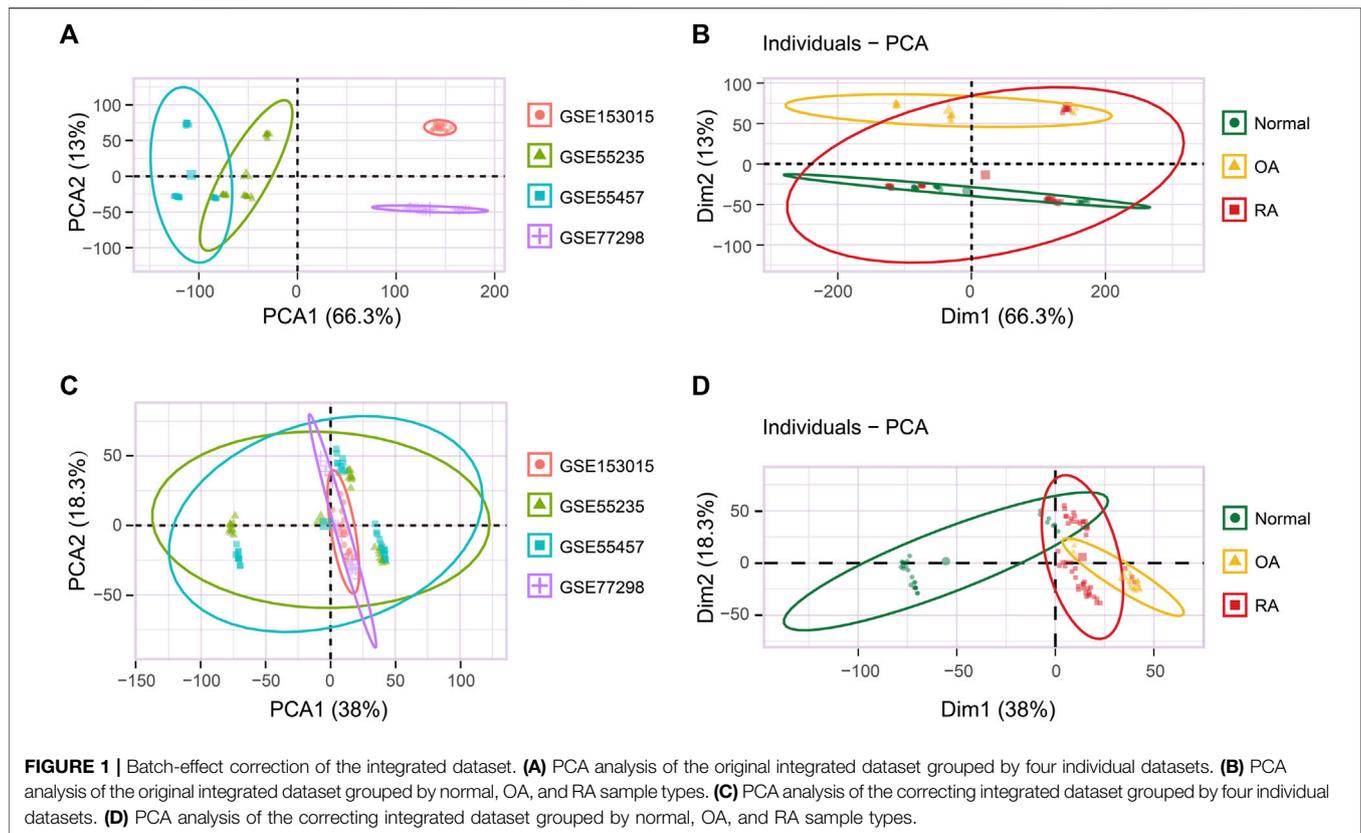
Quality Control of Gene Expression Datasets

We determined the targeted analysis dataset integrating four independent GEO datasets originating from two GEO platforms, each employing RA, OA, and/or normal joint synovial tissue samples (Table 1). Among the GSE55235, GSE55457, GSE77298, and

GSE153015 datasets for which the background correction, missing values supplement, and mean expression value calculation were performed, the batch effect was clearly observed (Figure 1A). Multivariate PCA showed that when classified according to the sample type, the integrated dataset was staggered with poor discrimination (Figure 1B). Then, the batch-effect correction was performed by the “sva” package using R, and the final data showed lower heterogeneity (Figure 1C). PCA analysis spanned 100 individuals of normal ($n = 27$), OA ($n = 24$), and RA ($n = 49$), of which three groups were clearly separated (Figure 1D). Subsequent analysis follows the principles of Supplementary Figure S1.

Identification of DEGs in RA vs OA and RA vs Normal, Respectively

Next, we explored DEGs of the RA vs normal group (49 RA vs 27 normal) and the RA vs OA group (49 RA vs 24 OA). Using a p -value filter under 0.05 with a difference of twofold or more, we identified 287 DEGs between the RA and normal groups, of which 203 were upregulated and 84 were downregulated genes in the RA group (Figure 2A and Supplementary Table SA). In addition, we also identified 1,564 DEGs between RA and OA group, which contained 792 upregulated genes and 772 downregulated genes in the RA group (Figure 2D and Supplementary Table SB). As expected, these samples could be distinguished in an unsupervised cluster analysis based on the overall expression trend of DEGs (Figures 2B,E). GO enrichment analyses revealed that the MF of DEGs among the RA and normal groups were primarily related to amide binding, immune receptor activity, and peptide binding, whereas the main BP involved immune response-activating cell surface receptor signaling pathway, immune response-activating signal transduction, and T-cell activation. As for the CC term in the GO analysis, DEGs were mainly localized in the external side of the plasma membrane, collagen-containing extracellular matrix, and endocytic vesicle (Figure 2C and Supplementary Figures S2A,B). For the DEGs between the RA and OA groups, the MF were highly related to chemokine activity, chemokine receptor binding, and CXCR chemokine receptor binding, whereas the main BP contained leukocyte chemotaxis, myeloid leukocyte migration, and granulocyte migration. In CC analysis, DEGs were mainly localized in lipoprotein particle, plasma lipoprotein particle, and intracellular transport particle (Figure 2F and Supplementary Figures S2C,D).



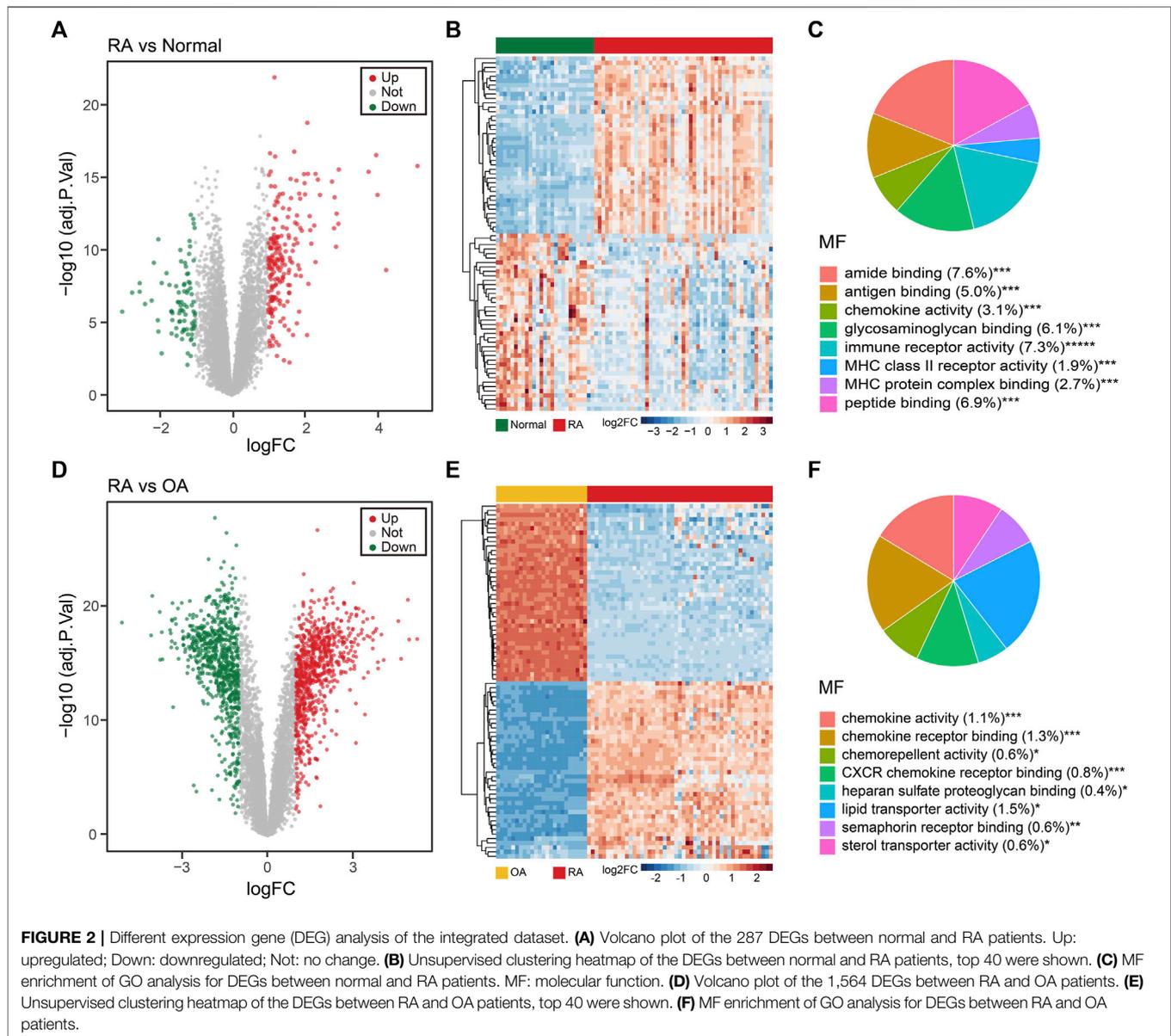
WGCNA Identifies Critical Modules Correlating With RA

To further investigate the changed genes in the RA group, we performed an unbiased gene expression analysis to identify coexpressed genes and modules in our dataset. WGCNA is a systems biology method used to decipher coexpression patterns among genes across different samples (Langfelder and Horvath, 2008). In our WGCNA analysis, the integrated dataset was firstly clustered to screen out whether there were any outliers, and the sample clustering heatmap revealed that no sample was excluded (Figure 3A). The soft-thresholding power β was selected as 7, which could make the scale-free network evaluation coefficient R^2 equal to 0.85 (Figure 3B). Then, we identified 5 gene coexpression modules except the grey module which incorporated uncategorized modular genes through TOM matrix hierarchical clustering and dynamic tree cut (Figures 3C,D). Moreover, the blue module ($R^2 = 0.82$; $P = 1e-25$), the yellow module ($R^2 = -0.61$; $P = 1e-11$), and the brown module ($R^2 = -0.58$; $P = 3e-10$) were highly associated with RA. In addition, the values of GS and MM in three modules were calculated and presented with a scatter diagram (Figure 3E). Finally, by setting the threshold $|GS| > 0.5$ and $|MM| > 0.7$, we screened out 397 RA-trait module genes for further study (Supplementary Table SC).

DEGs Correlating With RA-Trait Module Genes for Hub Gene Selection

To eliminate those invalid genes which have no expression change among the RA-trait module genes, we overlapped the DEGs and RA-trait module genes by TBtools (Chen et al., 2020) (Figure 4A). 76 target genes were obtained and displayed by an unsupervised clustering analysis (Figure 4B). GO and KEGG analyses indicated that these RA-trait genes were closely related to T lymphocyte activation and regulation (Supplementary Figures S3A,B and Supplementary Tables SD, SE). The PPI network of these genes was acquired by STRING (Szklarczyk et al., 2021) and further visualized and analyzed through Cytoscape (Agg et al., 2019). As shown in Figure 4C, the PPI network in Cytoscape included 58 nodes and 287 edges (p -value $< 1.0e-16$). Then, three significant clusters were predicted through plugin MCODE (degree cutoff = 2, node score cutoff = 0.2, K-core = 2, and max. depth = 100). We observed the most significant cluster (score: 12.8) including 16 nodes and 96 edges (Figure 4D).

To sift hub genes that play a pivotal role in the related pathways, we sorted out 8 kinds of algorithms including cytoHubba, namely, BottleNeck, Closeness, Degree, Eccentricity, EPC, MCC, MNC, and Radiality to analyze the most significant cluster. The top 10 genes of each algorithm were shown in Supplementary Table SF. We observed enrichment of 6 genes including GZMB, KLRB1, CD2, CD3D, CD8A, and GZMK



in all 8 algorithms, where CD2 (Raychaudhuri et al., 2009), CD8A (Carvalho et al., 2015; Souto-Carneiro et al., 2020), and GZMB (Bao et al., 2018) had already been reported as important genes involved in the progression of RA (Figure 4E). Finally, the three novel hub genes, CD3D, GZMK, and KLRB1 uniquely exhibited RA associations in our dimensionality reduction analysis.

External Dataset Verification Uncovers Specifically High Expression of CD3D, GZMK, and KLRB1 in the Synovium of RA Patients

We queried three additional datasets to validate the RA associations of CD3D, GZMK, and KLRB1. Datasets

GSE39340 (GPL10558) contained 10 RA and 7 OA synovial tissue samples and another dataset GSE55584 (GPL96) contained 10 RA and 6 OA synovial tissue samples. These two gene profiling datasets both demonstrated that the three hub genes showed obvious statistical significance between RA and OA samples (Figures 5A,B). The third GSE89408 (GPL11154) dataset was a high-throughput expression profiling which contained 28 normal, 22 OA, 57 early RA, and 95 established RA synovial tissue samples. To our surprise, CD3D, GZMK, and KLRB1 not only were highly expressed in the synovial tissue of RA patients but also showed the same trend in early RA patients (Figures 5C-E).

Based on verification results, we hypothesized that CD3D, GZMK, and KLRB1 might be involved in the early evolution of

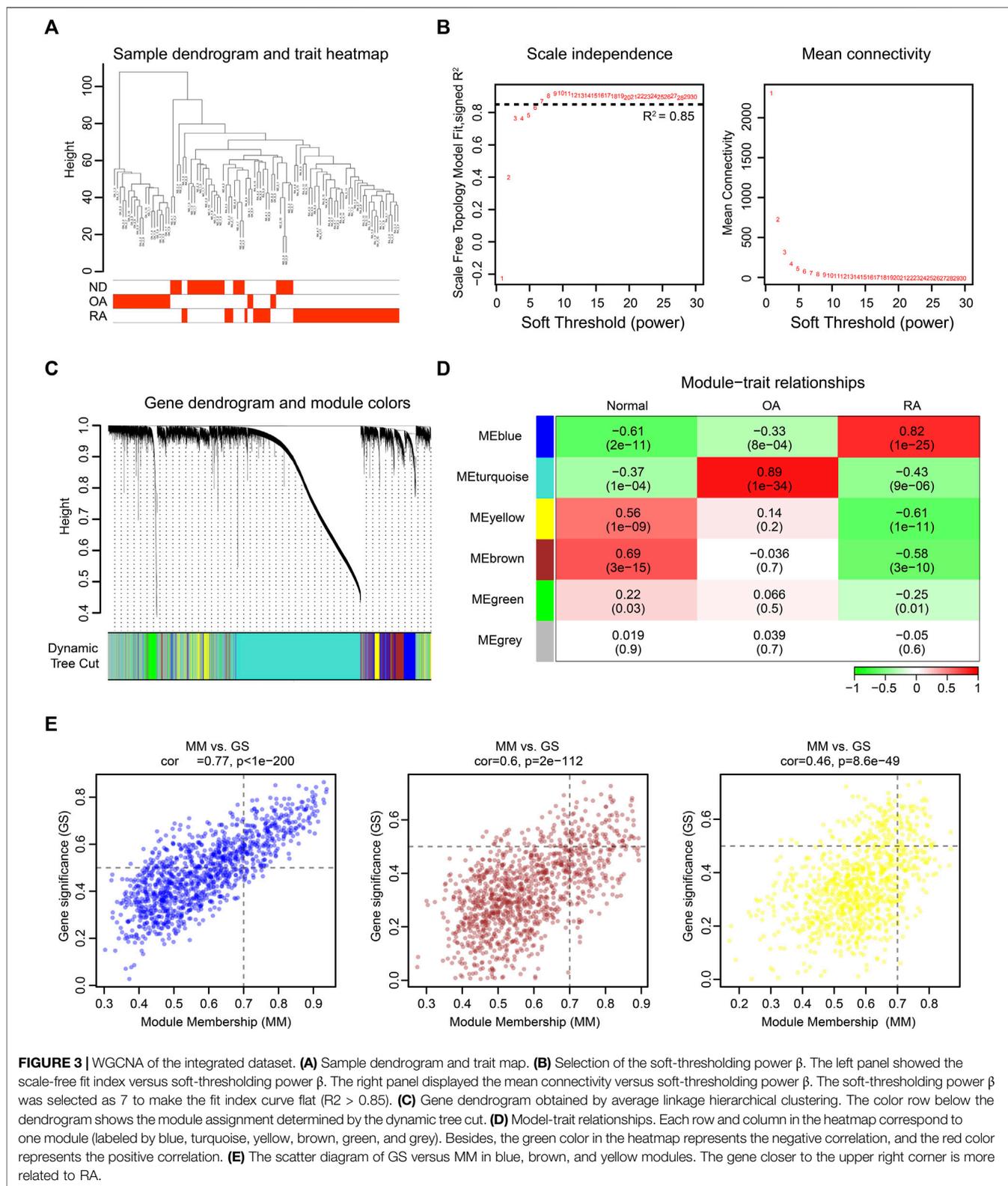
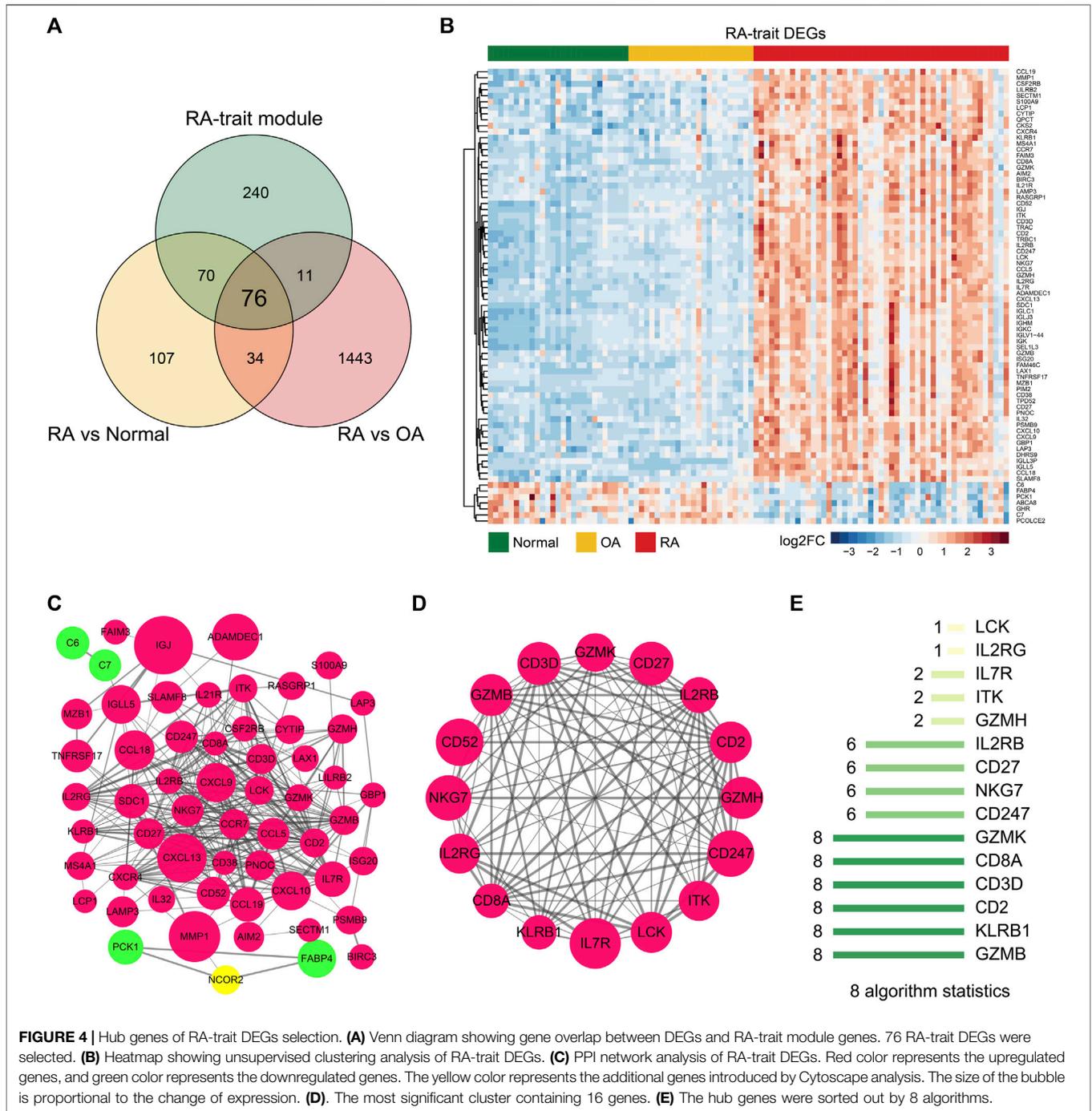


FIGURE 3 | WGCNA of the integrated dataset. **(A)** Sample dendrogram and trait map. **(B)** Selection of the soft-thresholding power β . The left panel showed the scale-free fit index versus soft-thresholding power β . The right panel displayed the mean connectivity versus soft-thresholding power β . The soft-thresholding power β was selected as 7 to make the fit index curve flat ($R^2 > 0.85$). **(C)** Gene dendrogram obtained by average linkage hierarchical clustering. The color row below the dendrogram shows the module assignment determined by the dynamic tree cut. **(D)** Model-trait relationships. Each row and column in the heatmap correspond to one module (labeled by blue, turquoise, yellow, brown, green, and grey). Besides, the green color in the heatmap represents the negative correlation, and the red color represents the positive correlation. **(E)** The scatter diagram of GS versus MM in blue, brown, and yellow modules. The gene closer to the upper right corner is more related to RA.

RA and could become potential diagnostic markers. ROC curves were performed to assess the utility of CD3D, GZMK, and KLRB1 to differentiate between OA, early RA,

and established RA. As shown in **Figures 5F–H**, the significant clinical correlation was only presented in early RA and established RA patients ($p < 0.001$), which indicated that

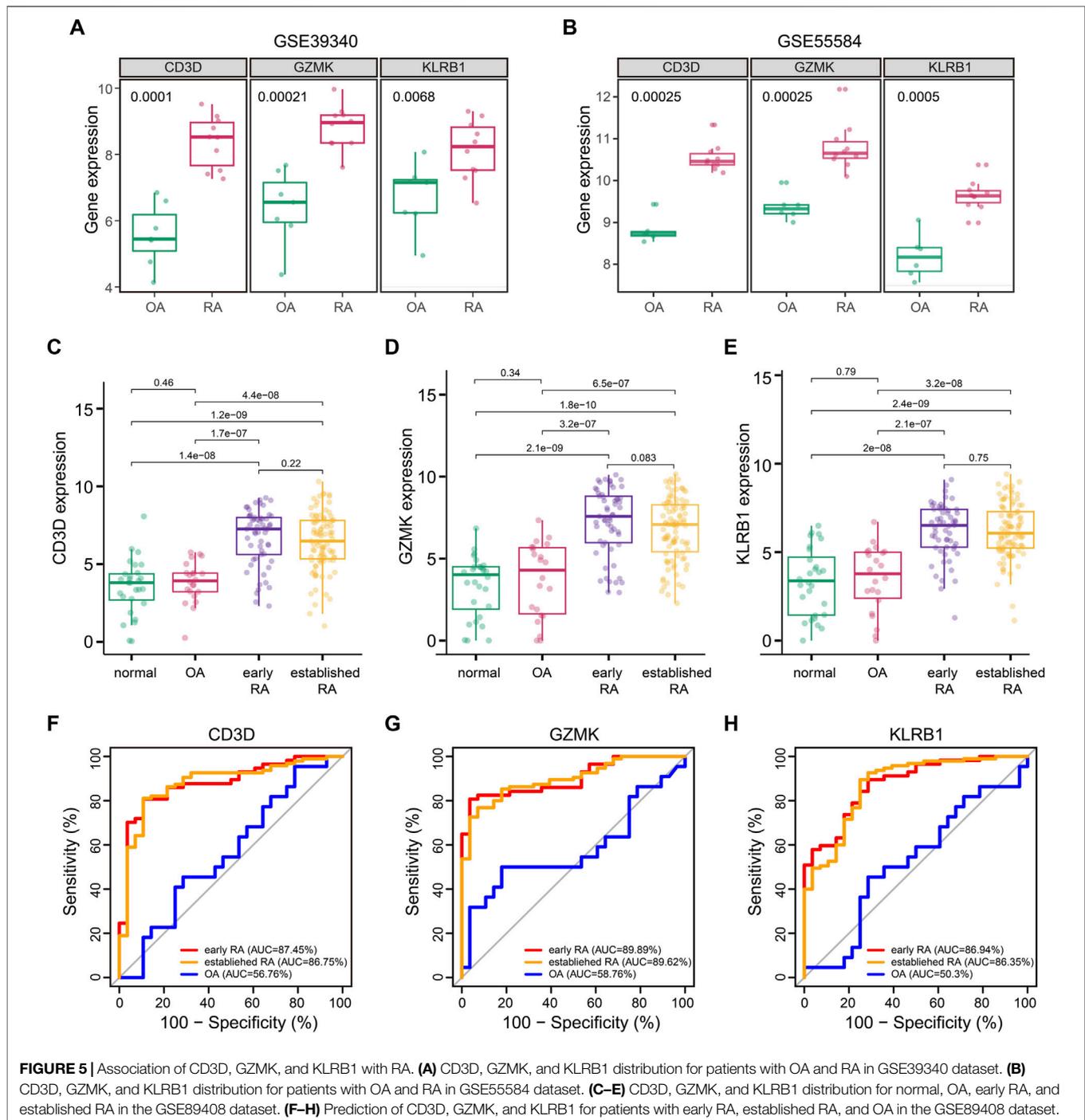


these three hub genes specifically correlated with RA. The area under the curve (AUC) of the ROC curve is an indicator combining sensitivity and specificity, which could demonstrate the intrinsic effectiveness of diagnostic tests. The AUC of CD3D, GZMK, and KLRB1 levels for early RA was 0.869, 0.875, and 0.899, respectively. These three hub genes showed significantly greater predictive power for RA

patients than OA patients but failed to distinguish between early RA and established RA patients.

Performance of CD3D, GZMK, and KLRB1 in ACPA-Negative RA Patients

Clinically, ACPA-negative RA poses a great challenge for early diagnosis (Ohmura et al., 2010; Daha and Toes, 2011; Li et al.,

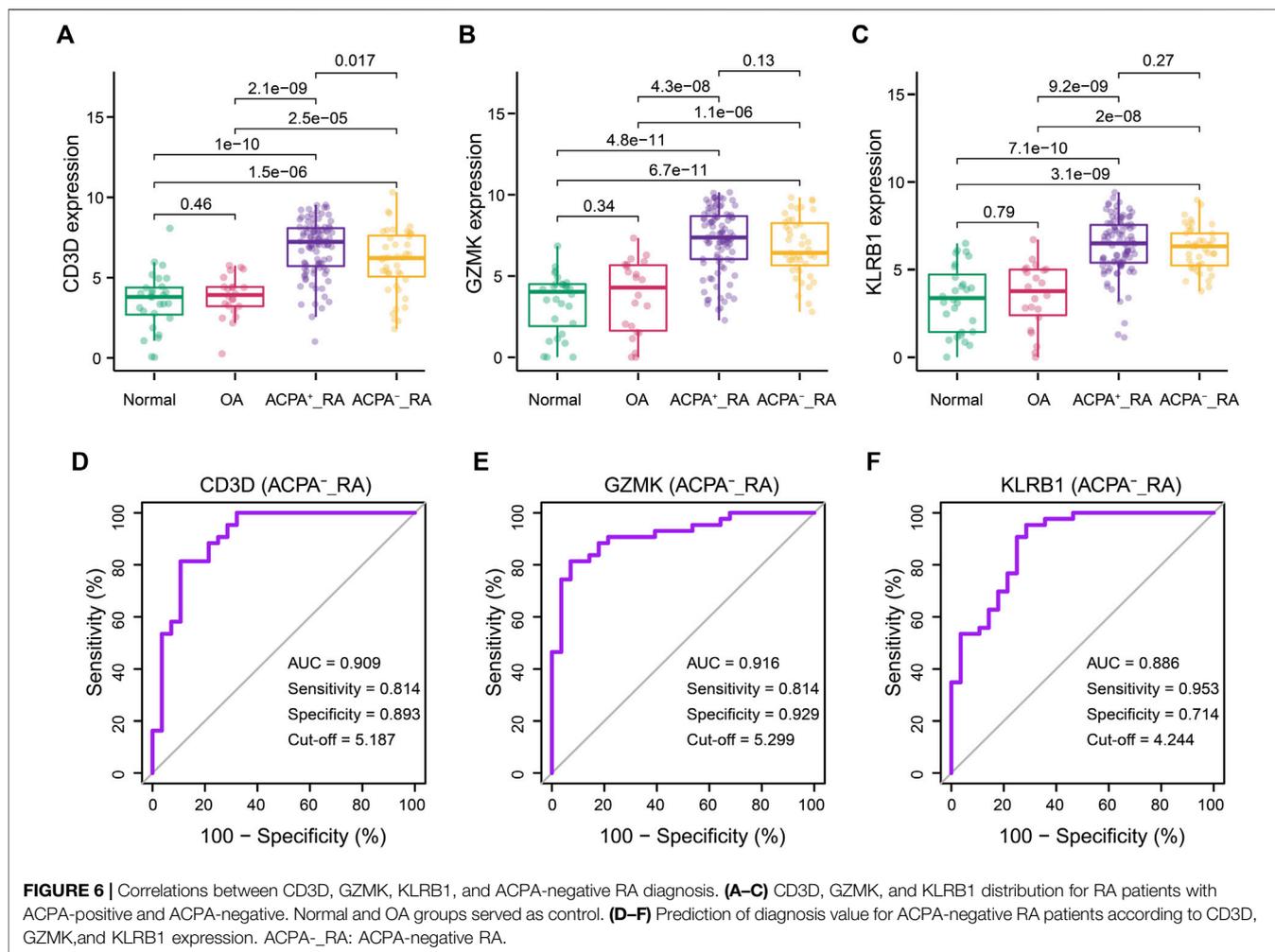


2021). Thus, we reintegrated across the GSE89408 dataset to evaluate the performance of the CD3D, GZMK, and KLRB1 in the diagnosis of ACPA-negative RA. The RA samples in GSE89408 dataset were divided into ACPA-negative ($n = 96$) and ACPA-positive RA ($n = 43$) subgroups for further verification. Surprisingly, similar to the ACPA-positive RA group, the expression levels of CD3D, GZMK, and KLRB1 in the ACPA-negative RA group were significantly higher than those in the normal and OA groups (**Figures 6A–C**). The AUC

of CD3D, GZMK, and KLRB1 expression levels was 0.909, 0.916, and 0.886, respectively (**Figures 6D–F**).

DISCUSSION

Bioinformatics analysis methods are widely used in many studies for various diseases (Kanehisa and Bork, 2003). In this study, we



performed WGCNA and differential expression analysis to obtain RA-trait DEGs. GO and KEGG analyses revealed the pathological mechanism of these genes involved in the pathogenesis of RA. Six hub genes, named CD2, CD3D, CD8A, GZMB, GZMK, and KLRB1, were acquired through 8 algorithms. ROC curve revealed the diagnostic and prognostic significance of those hub genes in this study. By reviewing previous studies, GZMB, CD8A, and CD2 had been proven to have potential diagnosis and application value in RA. The other three hub genes, CD3D, GZMK, and KLRB1, were currently found to be lack of attention in RA. They were further verified in independent GEO datasets and the statistically significant difference was only presented in RA samples when compared with normal and OA samples. Meanwhile, clinical correlation analysis and ROC curve drawing were carried out and it was found that these three hub genes were not only meaningful for the early diagnosis of RA but also have guiding significance for the diagnosis of ACPA-negative patients.

Previous studies have shown that CD3D encodes the δ subunit of transmembrane CD3 antigen complex and forms the T-cell receptor/CD3 complex (TCR/CD3 complex) with the other four CD3 subunits for T-cell development and signal transduction (Gil et al., 2011). The

deficiency of CD3D could cause damage to immunity (Dadi et al., 2003). Additionally, studies have reported that CD3D is a molecular diagnostic marker for immunodeficiency in early infancy (Kwan et al., 2014). Moreover, CD3D has been confirmed to actually participate in the abnormal activation of T lymphocyte immune-related pathways based on epigenetic and genomic analysis (Limbach et al., 2016; Deng et al., 2019). Coincidentally, these pathways were observed to be enriched in the present study (**Supplementary Figure S4A**).

GZMK is a member of the serine-proteases family, which is mainly expressed by T lymphocytes (**Supplementary Figure S4B**). In addition to the cytotoxicity of this family, it also has the effect of promoting proinflammatory cytokines release (Joeckel et al., 2011). Moreover, in human infectious diseases, GZMK has been found to activate protease-activated receptor-1 (PAR-1) in endothelial and fibroblast cells and induce the production of inflammatory cytokines, such as TNF- α , IL1, IL-6, and MCP-1 (Sharma et al., 2016; Herich et al., 2019). All these cytokines could cause inflammation cascades, leading to more inflammatory cells infiltration. Besides, its protease effect can promote the degradation of the extracellular matrix, resulting in

inflammatory cell infiltration and tissue destruction (Turner et al., 2019). These studies suggest that GZMK may trigger the continuous inflammation amplification of RA.

The protein NKRPIA encoded by KLRB1 was a member of the NKRPI family, and it was mainly expressed on T lymphocytes and natural killer (NK) cells (**Supplementary Figure S4C**). NKRPIA plays an inhibitory role in NK cell cytotoxicity (Kurioka et al., 2018). In addition, activation of NKRPIA on T cells had been found to be associated with IL17, IFN- γ , and TNF production (Billerbeck et al., 2010) and is involved in the process of inflammation and the pathogenesis of autoimmune diseases. The signal transduction mechanism of KLRB1 might be related to the activation of PI3 kinase/AKT and ERK1/2 pathways in NK cells and the PI3 kinase/AKT and STAT3 pathways in T cells (Pozo et al., 2006; Bai et al., 2014). Consistently, the activation of these signaling pathways is closely related to the pathogenesis of RA.

The study is the first to directly link the expression change of three genes with RA through unbiased and independent bioinformatics analysis, and we defined these three genes as RA-trait DEGs. We observed heterogeneity of CD3D, KZMK, and KLRB1 detectable in normal, OA, and RA synovial tissues. The gene expression of these RA-trait DEGs can distinguish early RA from OA patients. Lack of effective biomarkers in ACPA-negative RA patients impedes early diagnosis and treatment. Statistically, about one-third of RA patients are ACPA-negative, and the clinical characteristics of ACPA-negative RA are different from those of ACPA-positive RA. Our ROC curves analysis suggested that the identified RA-trait DEGs might be potential markers for ACPA-negative RA diagnosis. However, this study has several limitations. First, the specificity of this gene expression in other inflammatory or immune diseases, such as psoriasis, SLE, and multiple sclerosis, has not been investigated. Second, the sample size of our study was still limited, and it is essential to interrogate the gene expression in synovium by further experiment. Third, it is necessary to explore the possibility of checking these three genes or corresponding proteins through blood or joint fluid. In conclusion, our study could provide a theoretical basis for better management of early diagnosis of RA.

REFERENCES

- Ágg, B., Császár, A., Szalay-Bekő, M., Veres, D. V., Mizsei, R., Ferdinandy, P., et al. (2019). The EntOptLayout Cytoscape Plug-In for the Efficient Visualization of Major Protein Complexes in Protein-Protein Interaction and Signalling Networks. *Bioinformatics* 35 (21), 4490–4492. doi:10.1093/bioinformatics/btz257
- Bai, A., Moss, A., Kokkotou, E., Usheva, A., Sun, X., Cheifetz, A., et al. (2014). CD39 and CD161 Modulate Th17 Responses in Crohn's Disease. *J. Immunol.* 193 (7), 3366–3377. doi:10.4049/jimmunol.1400346
- Bao, C. X., Chen, H. X., Mou, X. J., Zhu, X. K., Zhao, Q., and Wang, X. G. (2018). 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, 346–354. doi:10.1016/j.biopha.2018.04.023
- Billerbeck, E., Kang, Y. H., Walker, L., Lockstone, H., Grafmueller, S., Fleming, V., et al. (2010). Analysis of CD161 Expression on Human CD8+ T Cells Defines a Distinct Functional Subset with Tissue-Homing Properties. *Proc. Natl. Acad. Sci. U S A.* 107 (7), 3006–3011. doi:10.1073/pnas.0914839107

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

WD and JZ: conceptualization, supervision, and manuscript editing. JL: design, data curation, formal analysis, investigation, methodology, and writing. YB: data curation, formal analysis, and manuscript editing. YZ and SH: supervision, methodology, and resources supervision. All authors have read and agreed to the published version of the manuscript.

FUNDING

This research was funded by the Anhui Provincial Science and Technology Major Project (8212929035).

ACKNOWLEDGMENTS

We are grateful to the National Center for Biotechnology Information (NCBI, National Institutes of Health, United States) for the free GEO database resources. We also thank Raimund W Kinne et al. (GSE55235, GSE55457, and GSE55584), Broeren MG et al. (GSE77298), Patrick Durez et al. (GSE153015), Chang X (GSE39340), and Guo Y et al. (GSE89408) for selfless sharing of the database.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fphar.2021.726529/full#supplementary-material>

- Carvalho, H., Duarte, C., Silva-Cardoso, S., da Silva, J. A., and Souto-Carneiro, M. M. (2015). CD8+ T Cell Profiles in Patients with Rheumatoid Arthritis and Their Relationship to Disease Activity. *Arthritis Rheumatol.* 67 (2), 363–371. doi:10.1002/art.38941
- Chen, C., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y., et al. (2020). TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. *Mol. Plant.* 13 (8), 1194–1202. doi:10.1016/j.molp.2020.06.009
- Cunningham, C. C., Wade, S., Floudas, A., Orr, C., McGarry, T., Wade, S., et al. (2021). Serum miRNA Signature in Rheumatoid Arthritis and "At-Risk Individuals". *Front. Immunol.* 12, 633201. doi:10.3389/fimmu.2021.633201
- Dadi, H. K., Simon, A. J., and Roifman, C. M. (2003). Effect of CD3delta Deficiency on Maturation of Alpha/beta and Gamma/delta T-Cell Lineages in Severe Combined Immunodeficiency. *N. Engl. J. Med.* 349 (19), 1821–1828. doi:10.1056/NEJMoa031178
- Daha, N. A., and Toes, R. E. (2011). Rheumatoid Arthritis: Are ACPA-Positive and ACPA-Negative RA the Same Disease? *Nat. Rev. Rheumatol.* 7 (4), 202–203. doi:10.1038/nrrheum.2011.28

- Deng, Q., Luo, Y., Chang, C., Wu, H., Ding, Y., and Xiao, R. (2019). The Emerging Epigenetic Role of CD8⁺T Cells in Autoimmune Diseases: A Systematic Review. *Front. Immunol.* 10, 856. doi:10.3389/fimmu.2019.00856
- Gibofsky, A. (2014). Epidemiology, Pathophysiology, and Diagnosis of Rheumatoid Arthritis: A Synopsis. *Am. J. Manag. Care.* 20 (7 Suppl. 1), S128–S135.
- Gil, J., Busto, E. M., Garcillán, B., Chean, C., García-Rodríguez, M. C., Díaz-Alderete, A., et al. (2011). A Leaky Mutation in CD3D Differentially Affects $\alpha\beta$ and $\gamma\delta$ T Cells and Leads to a T $\alpha\beta$ -T $\gamma\delta$ +B+NK+ Human SCID. *J. Clin. Invest.* 121 (10), 3872–3876. doi:10.1172/JCI44254
- Guo, Q., Wang, Y., Xu, D., Nossent, J., Pavlos, N. J., and Xu, J. (2018). Rheumatoid Arthritis: Pathological Mechanisms and Modern Pharmacologic Therapies. *Bone Res.* 6, 15. doi:10.1038/s41413-018-0069-9
- Herich, S., Schneider-Hohendorf, T., Rohlmann, A., Khaleghi Ghadiri, M., Schulte-Mecklenbeck, A., Zondler, L., et al. (2019). Human CCR5^{high} Effector Memory Cells Perform CNS Parenchymal Immune Surveillance via GZMK-Mediated Transendothelial Diapedesis. *Brain* 142 (11), 3411–3427. doi:10.1093/brain/awz301
- Jarlborg, M., Courvoisier, D. S., Lamacchia, C., Martinez Prat, L., Mahler, M., Bentow, C., et al. (2020). Serum Calprotectin: a Promising Biomarker in Rheumatoid Arthritis and Axial Spondyloarthritis. *Arthritis Res. Ther.* 22 (1), 105. doi:10.1186/s13075-020-02190-3
- Joeckel, L. T., Wallich, R., Martin, P., Sanchez-Martinez, D., Weber, F. C., Martin, S. F., et al. (2011). Mouse Granzyme K Has Pro-inflammatory Potential. *Cell Death Differ.* 18 (7), 1112–1119. doi:10.1038/cdd.2011.5
- Kanehisa, M., and Bork, P. (2003). Bioinformatics in the post-sequence Era. *Nat. Genet.* 33 Suppl (Suppl. 1), 305–310. doi:10.1038/ng1109
- Kurioka, A., Cosgrove, C., Simoni, Y., van Wilgenburg, B., Geremia, A., Björkander, S., et al. (2018). CD161 Defines a Functionally Distinct Subset of Pro-inflammatory Natural Killer Cells. *Front. Immunol.* 9, 486. doi:10.3389/fimmu.2018.00486
- Kwan, A., Abraham, R. S., Currier, R., Brower, A., Andruszewski, K., Abbott, J. K., et al. (2014). Newborn Screening for Severe Combined Immunodeficiency in 11 Screening Programs in the United States. *JAMA* 312 (7), 729–738. doi:10.1001/jama.2014.9132
- Lamacchia, C., Courvoisier, D. S., Jarlborg, M., Bas, S., Roux-Lombard, P., Möller, B., et al. (2021). Predictive Value of Anti-CarP and Anti-PAD3 Antibodies Alone or in Combination with RF and ACPA for the Severity of Rheumatoid Arthritis. *Rheumatology (Oxford)*. doi:10.1093/rheumatology/keab050
- Langfelder, P., and Horvath, S. (2008). WGCNA: an R Package for Weighted Correlation Network Analysis. *BMC Bioinformatics* 9, 559. doi:10.1186/1471-2105-9-559
- Lee, Y. C., Hackett, J., Frits, M., Iannaccone, C. K., Shadick, N. A., Weinblatt, M. E., et al. (2016). Multibiomarker Disease Activity Score and C-Reactive Protein in a Cross-Sectional Observational Study of Patients with Rheumatoid Arthritis with and without Concomitant Fibromyalgia. *Rheumatology (Oxford)* 55 (4), 640–648. doi:10.1093/rheumatology/kev388
- Lewis, M. J., Barnes, M. R., Blighe, K., Goldmann, K., Rana, S., Hackney, J. A., et al. (2019). Molecular Portraits of Early Rheumatoid Arthritis Identify Clinical and Treatment Response Phenotypes. *Cell Rep.* 28 (9), 2455–2470. doi:10.1016/j.celrep.2019.07.091
- Li, K., Mo, W., Wu, L., Wu, X., Luo, C., Xiao, X., et al. (2021). Novel Autoantibodies Identified in ACPA-Negative Rheumatoid Arthritis. *Ann. Rheum. Dis.* 80, 739–747. doi:10.1136/annrheumdis-2020-218460
- Limbach, M., Saare, M., Tserel, L., Kisand, K., Eglit, T., Sauer, S., et al. (2016). Epigenetic Profiling in CD4⁺ and CD8⁺ T Cells from Graves' Disease Patients Reveals Changes in Genes Associated with T Cell Receptor Signaling. *J. Autoimmun.* 67, 46–56. doi:10.1016/j.jaut.2015.09.006
- MacGregor, A. J., Snieder, H., Rigby, A. S., Koskenvuo, M., Kaprio, J., Aho, K., et al. (2000). Characterizing the Quantitative Genetic Contribution to Rheumatoid Arthritis Using Data from Twins. *Arthritis Rheum.* 43, 30–37. doi:10.1002/1529-0131(200001)43:1<30::AID-ANR5>3.0.CO;2-B
- Malmström, V., Catrina, A. L., and Klareskog, L. (2017). The Immunopathogenesis of Seropositive Rheumatoid Arthritis: from Triggering to Targeting. *Nat. Rev. Immunol.* 17 (1), 60–75. doi:10.1038/nri.2016.124
- McInnes, I. B., and Schett, G. (2011). The Pathogenesis of Rheumatoid Arthritis. *N. Engl. J. Med.* 365 (23), 2205–2219. doi:10.1056/NEJMra1004965
- Ohmura, K., Terao, C., Maruya, E., Katayama, M., Matoba, K., Shimada, K., et al. (2010). Anti-citrullinated Peptide Antibody-Negative RA Is a Genetically Distinct Subset: a Definitive Study Using Only Bone-Erosive ACPA-Negative Rheumatoid Arthritis. *Rheumatology (Oxford)* 49 (12), 2298–2304. doi:10.1093/rheumatology/keq273
- Okada, Y., Wu, D., Trynka, G., Raj, T., Terao, C., Ikari, K., et al. (2014). Genetics of Rheumatoid Arthritis Contributes to Biology and Drug Discovery. *Nature* 506 (7488), 376–381. doi:10.1038/nature12873
- Onuora, S. (2012). Rheumatoid Arthritis: RF Levels Predict RA Risk in the General Population. *Nat. Rev. Rheumatol.* 8 (10), 562. doi:10.1038/nrrheum.2012.159
- Orr, C. K., Najm, A., Young, F., McGarry, T., Biniecka, M., Fearon, U., et al. (2018). The Utility and Limitations of CRP, ESR and DAS28-CRP in Appraising Disease Activity in Rheumatoid Arthritis. *Front. Med. (Lausanne)* 5, 185. doi:10.3389/fmed.2018.00185
- Padyukov, L., Seielstad, M., Ong, R. T., Ding, B., Rönnelid, J., Seddighzadeh, M., et al. (2011). A Genome-wide Association Study Suggests Contrasting Associations in ACPA-Positive versus ACPA-Negative Rheumatoid Arthritis. *Ann. Rheum. Dis.* 70 (2), 259–265. doi:10.1136/ard.2009.126821
- Pozo, D., Valés-Gómez, M., Mavaddat, N., Williamson, S. C., Chisholm, S. E., and Reyburn, H. (2006). CD161 (Human NKR-P1a) Signaling in NK Cells Involves the Activation of Acid Sphingomyelinase. *J. Immunol.* 176 (4), 2397–2406. doi:10.4049/jimmunol.176.4.2397
- Raychaudhuri, S., Thomson, B. P., Remmers, E. F., Eyre, S., Hinks, A., Guiducci, C., et al. (2009). Genetic Variants at CD28, PRDM1 and CD2/CD58 Are Associated with Rheumatoid Arthritis Risk. *Nat. Genet.* 41 (12), 1313–1318. doi:10.1038/ng.479
- Sharma, M., Merkulova, Y., Raithatha, S., Parkinson, L. G., Shen, Y., Cooper, D., et al. (2016). Extracellular Granzyme K Mediates Endothelial Activation through the Cleavage of Protease-Activated Receptor-1. *FEBS J.* 283 (9), 1734–1747. doi:10.1111/febs.13699
- Sirota, M., and Butte, A. J. (2011). The Role of Bioinformatics in Studying Rheumatic and Autoimmune Disorders. *Nat. Rev. Rheumatol.* 7 (8), 489–494. doi:10.1038/nrrheum.2011.87
- Smolen, J. S., Aletaha, D., Barton, A., Burmester, G. R., Emery, P., Firestein, G. S., et al. (2018). Rheumatoid Arthritis. *Nat. Rev. Dis. Primers* 4, 18001. doi:10.1038/nrdp.2018.1
- Souto-Carneiro, M. M., Klika, K. D., Abreu, M. T., Meyer, A. P., Saffrich, R., Sandhoff, R., et al. (2020). Effect of Increased Lactate Dehydrogenase A Activity and Aerobic Glycolysis on the Proinflammatory Profile of Autoimmune CD8⁺ T Cells in Rheumatoid Arthritis. *Arthritis Rheumatol.* 72 (12), 2050–2064. doi:10.1002/art.41420
- Stahl, E. A., Wegmann, D., Trynka, G., Gutierrez-Achury, J., Do, R., Voight, B. F., et al. (2012). Bayesian Inference Analyses of the Polygenic Architecture of Rheumatoid Arthritis. *Nat. Genet.* 44 (5), 483–489. doi:10.1038/ng.2232
- Stephenson, W., Donlin, L. T., Butler, A., Rozo, C., Bracken, B., Rashidfarrokhi, A., et al. (2018). Single-cell RNA-Seq of Rheumatoid Arthritis Synovial Tissue Using Low-Cost Microfluidic Instrumentation. *Nat. Commun.* 9 (1), 791. doi:10.1038/s41467-017-02659-x
- Syed Mohamed Suhail, S. M., Nur Atiqah, I., Nur Salimi Zakirah, Z. A., Lailatul Syazwani, Z., Batis, W. W., Md Monoto, E. M., et al. (2019). Diagnostic Performance of Anti-RA33 Antibody as a Serological Marker for Rheumatoid Arthritis. *Malays J. Pathol.* 41 (3), 259–265.
- Szklarczyk, D., Gable, A. L., Nastou, K. C., Lyon, D., Kirsch, R., Pyysalo, S., et al. (2021). The STRING Database in 2021: Customizable Protein-Protein Networks, and Functional Characterization of User-Uploaded Gene/measurements Sets. *Nucleic Acids Res.* 49 (D1), D605–D612. doi:10.1093/nar/gkaa1074
- Tsuchiya, H., Ota, M., Sumitomo, S., Ishigaki, K., Suzuki, A., Sakata, T., et al. (2020). Parsing Multiomics Landscape of Activated Synovial Fibroblasts Highlights Drug Targets Linked to Genetic Risk of Rheumatoid Arthritis. *Ann. Rheum. Dis.* 80, 440–450. doi:10.1136/annrheumdis-2020-218189
- Turner, C. T., Zeglinski, M. R., Richardson, K. C., Zhao, H., Shen, Y., Papp, A., et al. (2019). Granzyme K Expressed by Classically Activated Macrophages Contributes to Inflammation and Impaired Remodeling. *J. Invest. Dermatol.* 139 (4), 930–939. doi:10.1016/j.jid.2018.09.031
- Zendman, A. J., van Venrooij, W. J., and Pruijn, G. J. (2006). Use and Significance of Anti-CCP Autoantibodies in Rheumatoid Arthritis. *Rheumatology (Oxford)* 45 (1), 20–25. doi:10.1093/rheumatology/kei111

Zhang, F., Wei, K., Slowikowski, K., Fonseka, C. Y., Rao, D. A., Kelly, S., et al. (2019). Defining Inflammatory Cell States in Rheumatoid Arthritis Joint Synovial Tissues by Integrating Single-Cell Transcriptomics and Mass Cytometry. *Nat. Immunol.* 20 (7), 928–942. doi:10.1038/s41590-019-0378-1

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

The handling editor declared a shared affiliation with several of the authors (JZ, JL, YB) at time of review.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article or claim that may be made by its manufacturer is not guaranteed or endorsed by the publisher.

Copyright © 2021 Lu, Bi, Zhu, Huipeng, Duan and Zhou. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.