Identifying Immune Cell Infiltration and Effective Diagnostic Biomarkers in Rheumatoid Arthritis by Bioinformatics Analysis

Sheng Zhou[1†], Hongcheng Lu[2†] and Min Xiong[3]*

[1] *Department of Orthopedics, The Affiliated Changzhou No.2 People's Hospital of Nanjing Medical University, Changzhou, China,* [2] *Department of Urology, The First Affiliated Hospital of Nanjing Medical University, Nanjing, China,* [3] *Department of Orthopedics, Sinopharm Dongfeng General Hospital, Hubei University of Medicine, Shiyan, China*

Background: Rheumatoid arthritis (RA) is a chronic systemic autoimmune disorder characterized by inflammatory cell infiltration, leading to persistent synovitis and joint destruction. The pathogenesis of RA remains unclear. This study aims to explore the immune molecular mechanism of RA through bioinformatics analysis.

Methods: Five microarray datasets and a high throughput sequencing dataset were downloaded. CIBERSORT algorithm was performed to evaluate immune cell infiltration in synovial tissues between RA and healthy control (HC). Wilcoxon test and Least Absolute Shrinkage and Selection Operator (LASSO) regression were conducted to identify the significantly different infiltrates of immune cells. Differentially expressed genes (DEGs) were screened by "Batch correction" and "RobustRankAggreg" methods. Functional correlation of DEGs were analyzed by Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG). Candidate biomarkers were identified by cytoHubba of Cytoscape, and their diagnostic effectiveness was predicted by Receiver Operator Characteristic Curve (ROC) analysis. The association of the identified biomarkers with infiltrating immune cells was explored using Spearman's rank correlation analysis in R software.

Results: Ten significantly different types of immune cells between RA and HC were identified. A total of 202 DEGs were obtained by intersection of DEGs screened by two methods. The function of DEGs were significantly associated with immune cells. Five hub genes (CXCR4, CCL5, CD8A, CD247, and GZMA) were screened by R package "UpSet". CCL5+CXCR4 and GZMA+CD8A were verified to have the capability to diagnose RA and early RA with the most excellent specificity and sensitivity, respectively. The correlation between immune cells and biomarkers showed that CCL5 was positively correlated with M1 macrophages, CXCR4 was positively correlated with memory activated CD4+ T cells and follicular helper T (Tfh) cells, and GZMA was positively correlated with Tfh cells.

Conclusions: CCL5, CXCR4, GZMA, and CD8A can be used as diagnostic biomarker for RA. GZMA-Tfh cells, CCL5-M1 macrophages, and CXCR4- memory activated CD4+ T

cells/Tfh cells may participate in the occurrence and development of RA, especially GZMA-Tfh cells for the early pathogenesis of RA.

INTRODUCTION

Rheumatoid arthritis (RA) is a chronic systemic autoimmune disease (1). The main clinical manifestation of RA is chronic synovitis of the affected joint characterized by persistent synovitis, synovial hyperplasia, and pannus formation, which destroy the bone tissue and gradually lead to the damage of joint function (2, 3). RA is common in middle-aged women aged 40–60 years, with a prevalence rate of 0.5%–2% (4). The underlying mechanism of RA is complex, which is caused by the interaction of genetic, environmental, and immune factors, in which immune factor plays an essential role in the entire process, especially in the early stage (5, 6). Therefore, exploring the diagnostic biomarkers and revealing the immune mechanism of RA is the key to the early prevention and treatment of RA and is the focus of current research.

Immune cells in synovial membrane, including resident and infiltrating immune cells, play a vital role in the occurrence and development of RA, an autoimmune disorder (7). Studies have shown that macrophages play an important role in promoting the development of RA. These cells secrete abundant cytokines, chemokines, and degrading enzymes, which lead to joint inflammation and bone destruction. These cells can cooperate with other immune cells to aggravate the formation of arthritis (8). In the past decade, a great deal of research was conducted to understand the role of T cells, especially activated Th17 and Th1 cells, in RA (9). Autoimmune Th17 cells induce synovial stroma and innate lymphocytes to secrete the cytokine granulocyte-macrophage colony-stimulating factor (MG-CSF) to initiate and enhance autoimmune arthritis in RA (10). In addition, other immune cells, such as dendritic cells (DCs) and natural killer (NK) cells, play an important regulatory role in the pathogenesis of RA (11). However, the immune mechanisms of RA in synovium have not been investigated thoroughly. Therefore, a systematic approach is urgently needed to evaluate the contribution of immune cells and explore key genes related to immune cells.

With the development and widespread use of microarray and high-throughput sequencing technology, bioinformatics analysis can be used to identify novel genes and biomarkers for many diseases, including autoimmune disease (12–14). A previous bioinformatics study suggested that GRB10 and E2f3 could be used as diagnostic markers of osteoarthritis (15). Immune cell infiltration plays an important role in the pathogenesis of many diseases, including tumor and non-tumor diseases. Currently, cell-type identification has been utilized in many diseases; this method estimates the relative subsets of RNA transcripts (CIBERSORT) and analyzes 22 immune cell subsets in complex tissues by normalized bulk transcriptome profiles (16). Thus, this method can help us systematically explore immune cell infiltration in RA synovial tissues.

In this study, CIBERSORT was performed to analyze immune cell infiltration in RA by using five microarray datasets. Wilcoxon test and LASSO regression were conducted to identify the significantly different infiltrates of immune cells in RA and HC. After DEGs were screened, functional correlation was analyzed by GO and KEGG, and hub genes were identified by R package "UpSet." ROC logistic regression was conducted to analyze the predictive of biomarkers. Spearman's rank correlation in R software was also used to analyze the correlation between biomarkers and significantly different infiltrates of immune cells. This work not only systematically analyzed the infiltration of immune cells in the synovial membrane of RA but also screened novel and effective diagnostic biomarkers for RA. Results provide a new method for the diagnosis and treatment of RA.

MATERIALS AND METHODS

GEO Dataset Collection
Gene expression profiling of RA was screened using the GEO (http://www.ncbi.nlm.nih.gov/geo) database. The inclusion criteria were as follows: (1) expression profiling by array or high-throughput sequencing of mRNA; (2) availability of the synovial tissues of patients with RA or HC from joint in the datasets; and (3) ten or more synovial specimens in the dataset. Six eligible datasets were selected, including GSE1919, GSE12021, GSE55235, GSE55457, GSE77298, and GSE89408. The details of all data are shown in **Table 1**.

Data Preprocessing and Study Design
Raw and series matrix files of the five test microarray datasets, namely, GSE1919, GSE12021, GSE55235, GSE55457, and GSE77298, were downloaded. For raw data, probe expression matrix was extracted and normalized by Robust Multiarray Average (RMA) based on R package "affy." Platform annotation file was used to convert the probe expression matrix into a gene expression matrix. For the case of multiple probes corresponding to one gene, the average value was obtained. After five gene matrixes were merged by Perl script, R package "sva" was applied to eliminate heterogeneity caused by different experimental batches and platforms. Finally, we obtained one merging normalized gene expression matrix and used R package "limma" for analysis. For series matrix files, Perl script was used to extract each probe expression matrix one by one and then convert them into gene expression matrices,

TABLE 1 | Basic information of selected datasets.

GEO	Platform	Tissue	Samples (number)			Experiment	Attribute	Author/Reference
		(Homo sapiens)	Total	HC	RA	type		
GSE1919	GPL91	Synovium	10	5	5	Array	Test	U. Ungethuem (17)
GSE12021	GPL96	Synovium	21	9	12	Array	Test	R. Huber (18)
GSE55235	GPL96	Synovium	20	10	10	Array	Test	D. Woetzel (19)
GSE55457	GPL96	Synovium	23	10	13	Array	Test	D. Woetzel (19)
GSE77298	GPL570	Synovium	23	7	16	Array	Test	M.G. Broeren (20)
GSE89408	GPL11154	Synovium	180	28	57/95*	RNA-Seq	Validation	Y. Guo (21)

Means 57 early RA and 95 established RA in GSE89408 dataset.

respectively. R package "limma" was used to analyze each gene expression matrix. The validation RNA-Seq dataset, GSE89408, was downloaded by the form of gene expression matrix. R package "edgeR" was employed for subsequent analysis. The flow diagram of this study is shown in **Supplementary Figure S1**.

Evaluation of Immune Cell Subtype Distribution

CIBERSORT algorithm was performed to evaluate immune cell infiltration in synovial tissues between RA and HC. This algorithm can transform the normalized gene expression matrix into the composition of infiltrating immune cells. After data were submitted to the CIBERSORT web portal (http://CIBERSORT.stanford.edu/), LM22 was used as a reference expression signature with 1000 permutations. The LM22 signature matrix defined 22 infiltrating immune cell components, including subsets of macrophages (M0 macrophages, M1 macrophages, and M2 macrophages), T cells (CD8+ T cells, naïve CD4+ T cells, memory resting CD4+ T cells, memory activated CD4+ T cells, Tfh cells, regulatory T cells, and gamma delta T cells), natural killer (NK) cells (resting NK cells and activated NK cells), mast cells (resting mast cells and activated mast cells), B cells (naïve B cells and memory B cells), dendritic cells (resting dendritic cells and activated dendritic cells), monocytes, plasma cells, neutrophils, and eosinophils. The p-values and root mean squared errors were determined for each expression file in CIBERSORT. Only data with a CIBERSORT p value < 0.05 was filtered and reserved for the following analysis. The output was directly integrated to generate an entire matrix of immune cell fractions. The results from CIBERSORT were visualized using the R packages "corplot", "vioplot", "ggplot2", and "glment".

Principal Component Analysis (PCA)

Intra-group data repeatability in each group was verified by Pearson's correlation test. The intra-group data repeatability of the dataset was tested by sample clustering analysis. Statistical analysis was performed by R programming language, and the results were presented by R package "ggplot2".

Screening of DEGs

Two methods were performed to obtain accurate DEGs from multiple microarray datasets. The first method ("Batch correction") merged the downloaded five raw datasets into an expression matrix and then analyzed the DEGs with R package "limma" after batch correction and normalization. The second method ("RobustRankAggreg, RRA") used the R package "limma" to analyze the DGEs of the downloaded gene expression matrices. The DEGs of each dataset were integrated with R package "RobustRankAggreg." DEGs obtained by the two methods were intersected by Venn diagram to extract final DEGs. The threshold points for DEGs were adj.P.Val < 0.05 and |log fold change (FC)| > 1.

Functional Enrichment Analysis

The gene names of DEGs were converted to gene ID by R package "org.Hs.eg.db". Analyses of Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) for DEGs were performed by R package "clusterProfiler." The significantly different GO terms and signal pathways were screened by the threshold p value <0.05 and q value <0.05. The results were visualized by R package "enrichplot" and "ggplot2".

Screening of Hub Genes

STRING (https://string-db.org) was performed to analyze the PPI network of DEGs with a highly reliable filtering condition (score>0.7). The interaction file (string_interactions.tsv) was downloaded. Perl was conducted to obtain the network file. The cytoHubba of Cytoscape (v 3.7.2) was conducted to score each node gene by top 10 algorithms, namely, MCC (Maximal Clique Centrality), DMNC (Density of Maximum Neighborhood Component), MNC (Maximum Neighborhood Component), Degree, EPC (Edge Percolated Component), BottleNeck, EcCentricity, Closeness, Radiality, and Betweenness. The top 60 node genes scored by each algorithm were used to screen hub genes by R package "UpSet". The RNA-Seq dataset GSE89408 was used to validate hub genes.

Analysis of the Predictive Value of Biomarkers

ROC analysis was performed to predict the diagnostic effectiveness of biomarkers by SSPA Statistics 23. The area under the ROC curve (AUC) value was utilized to determine the diagnostic effectiveness in discriminating RA from control samples in the GSE89408 dataset.

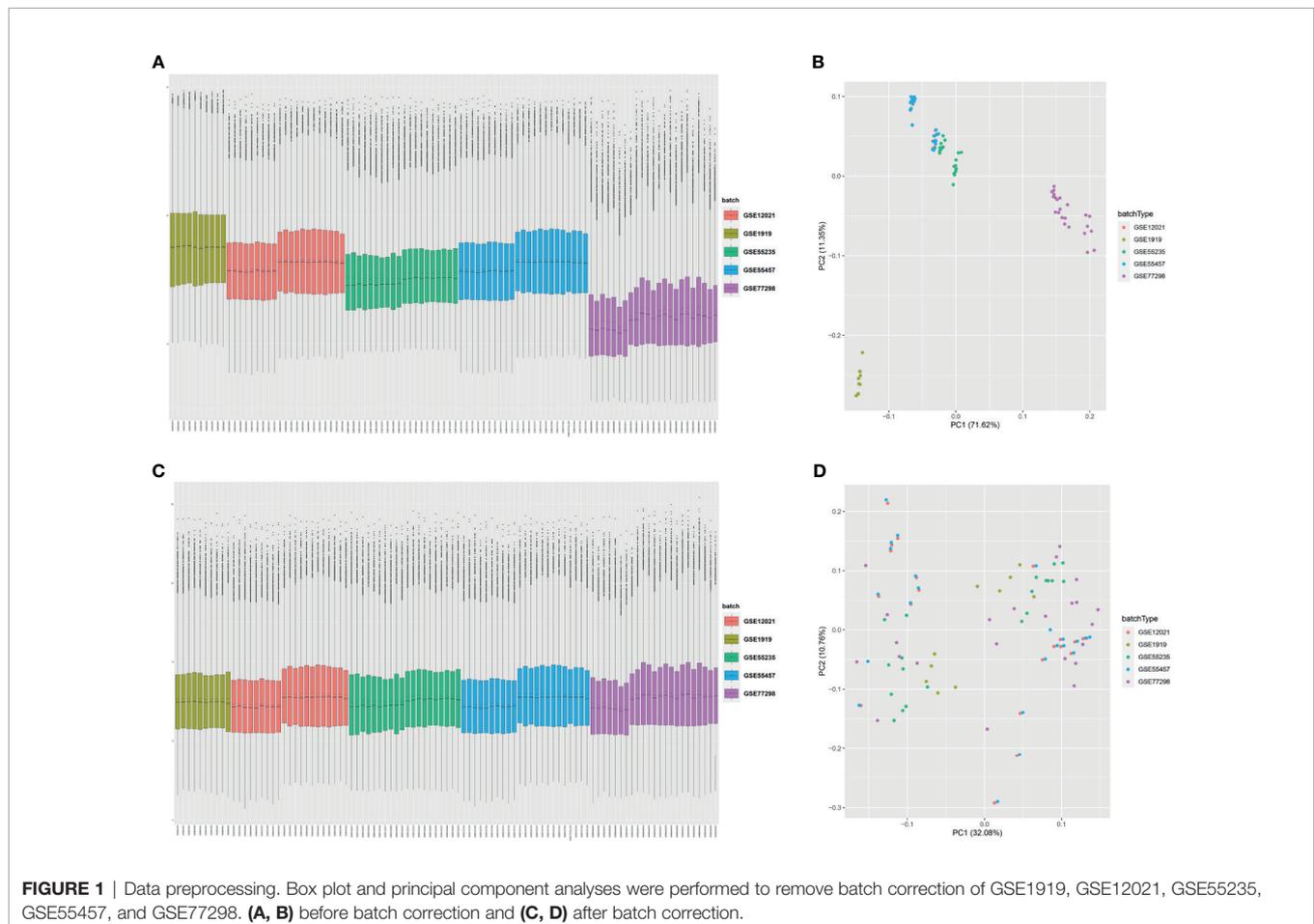## Correlation Analysis Between Biomarkers and Immune Cells

The association of the identified gene biomarkers with the levels of infiltrating immune cells was explored using Spearman's rank correlation analysis in R software. The resulting associations were visualized using the chart technique with the "ggplot2" package.

RESULTS

Immune Cell Infiltration in RA and Normal Synovial Tissues

Five microarray raw datasets, including 56 RA and 41 normal synovial tissues, were selected for the study of immune cell infiltration. The data before (A and B) and after (C and D) batch correction was presented in **Figure 1**, which indicated that the batch effect of the merged data was removed successfully. A total of 55 RA and 29 normal synovial tissues were found to be eligible for the analysis of CIBERSORT (p < 0.05). First, the composition of 22 kinds of immune cells in each sample was presented in a histogram (**Figure 2A**) and a heatmap (**Figure 2B**). In the histogram, the color represents the percentage of different immune cells in each sample, and the

sum is 1. In the heatmap, immune cells in each sample are shown with the normalized absolute abundance. The results indicated that M2 macrophages, CD8+ T cells, resting mast cells, naïve B cells, Tfh cells, M0 macrophages, and M1 macrophages were the main infiltrating immune cells. The correlation of 22 types of immune cells in RA synovial tissues was then evaluated (**Figure 2C**). For example, Tfh cells were positively correlated with memory activated CD4+ T cells and M1 macrophages. γδ T cells were positively related to memory activated CD4+ T cells and naïve B cells. Neutrophils were positively associated with resting NK cells, monocytes, and naïve CD4+ T cells. M2 macrophages were negatively correlated with naïve B cells. However, the above correlation of immune cells decreased or disappeared in HC (**Supplementary Figure S2**). Third, based on immune cell infiltration in synovial tissues, we could completely distinguished RA from normal HC by PCA analysis (**Figure 2D**). Two different algorithms, namely, Wilcoxon test and LASSO regression, were used to identify the significantly different infiltrates of immune cells in RA and HC. The results of Wilcoxon test are shown in a violin diagram in **Figure 3A**, which presented 13 types of immune cells with p < 0.05. The results of Lasso are presented in **Figures 3B, C**, which contained 12 significantly different types of immune cells. The 10 intersecting immune cells of the two methods were extracted.

FIGURE 1 | Data preprocessing. Box plot and principal component analyses were performed to remove batch correction of GSE1919, GSE12021, GSE55235, GSE55457, and GSE77298. **(A, B)** before batch correction and **(C, D)** after batch correction.
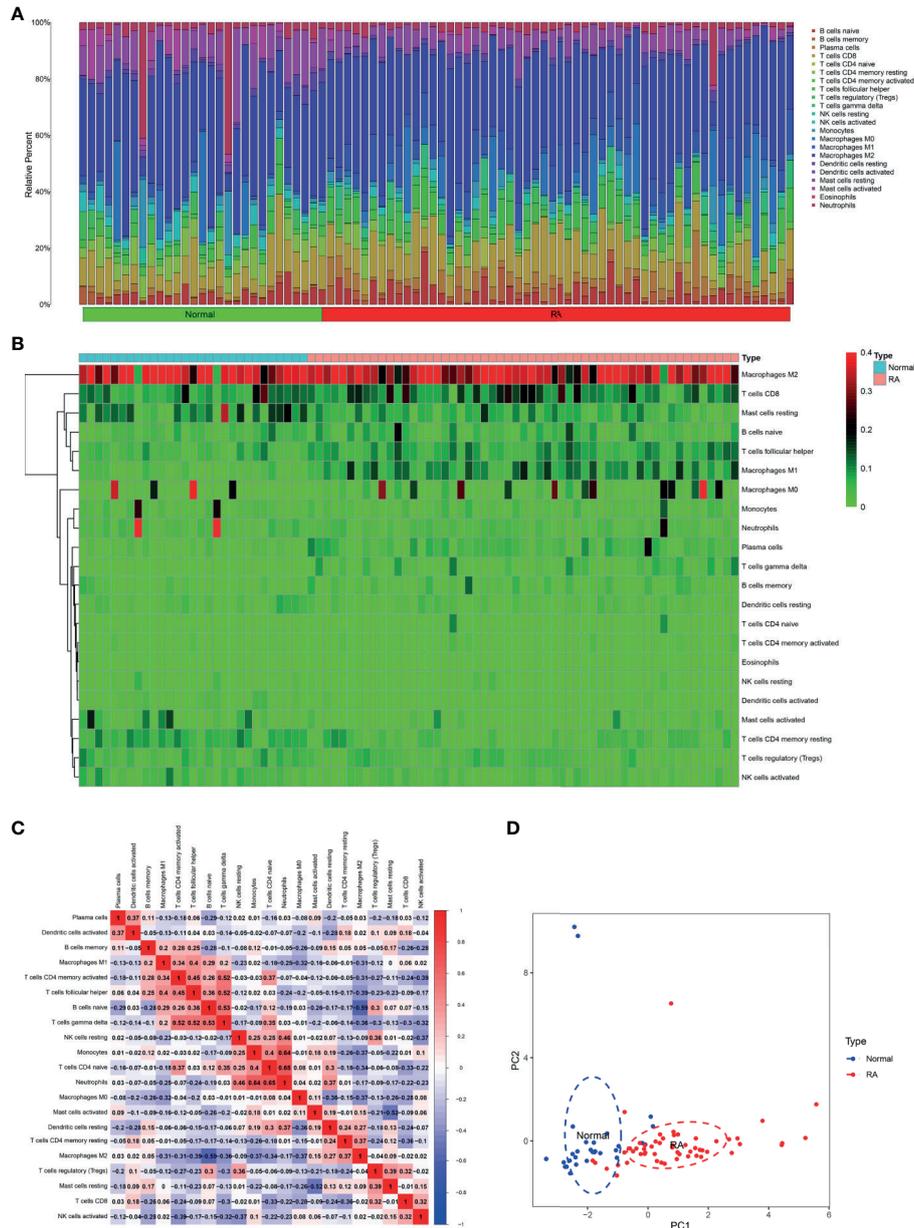
FIGURE 2 | Immune cell infiltration in RA and normal synovial tissues. The composition of 22 kinds of immune cells in each sample was showed in a histogram (A) and a heatmap (B). (C) The correlation of 22 types of immune cells in RA synovial tissues was evaluated. Red: positive correlation; blue: negative correlation. (D) PCA analysis was performed to classify infiltrating immune cells between RA and normal synovial tissues.

Compared with HC M1 macrophages, Tfh cells, the numbers of memory activated CD4+ T cells, and plasma cells were significantly higher in RA synovial tissues, while those of regulatory T cells, activated dendritic cells, activated NK cells, memory resting CD4+ T cells, resting mast cells, and activated mast cells were significantly lower in RA synovial tissues.

Identification of DEGs

Five microarray datasets, including 57 RA and 41 normal samples, were used by the two methods to identify DEGs. A total of 360 DEGs were obtained by "Batch correction" and included 335 upregulated and 25 downregulated genes. A total of 461 DEGs were obtained by "RRA" and included 298 upregulated and 163 downregulated genes. Some of DEGs are shown in **Figure 4A**. A total of 202 DEGs, including 179 upregulated and 23 downregulated genes, were obtained through the intersection of the DEGs screened by the two methods (**Figure 4B** and **Supplementary Table S1**). The final DEGs were visualized by the volcano map (**Figure 4C**) and heatmap (**Figure 4D**).
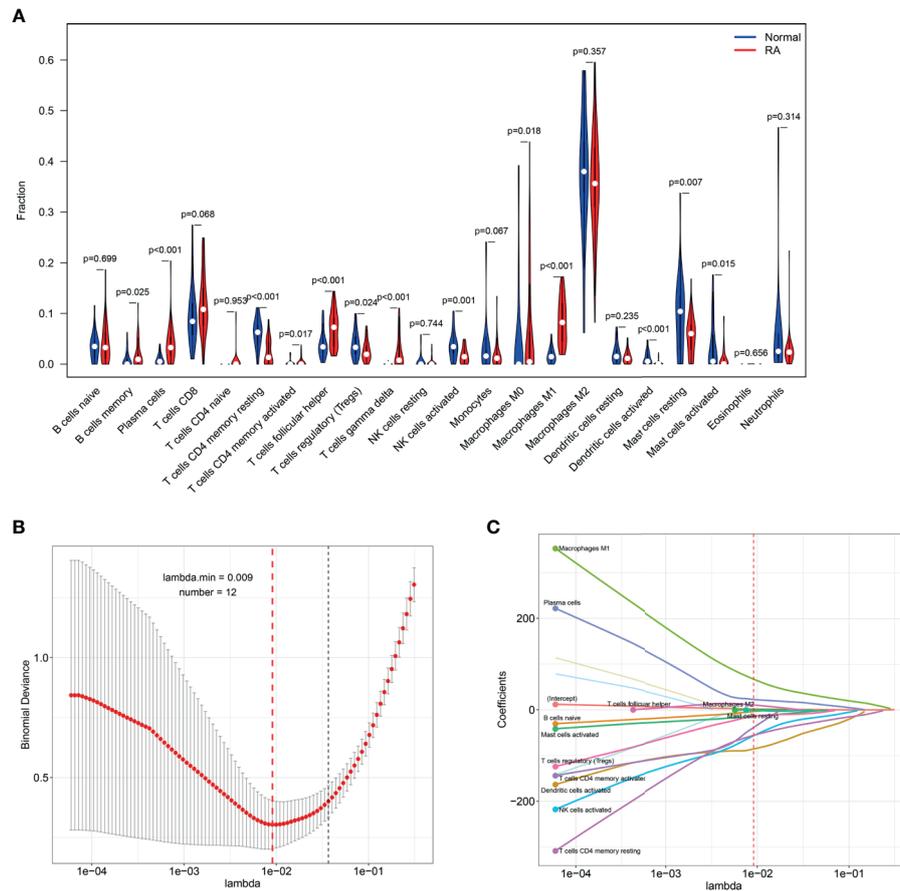
FIGURE 3 | Identifying the significantly different infiltrates of immune cells in RA. **(A)** Wilcoxon test and **(B, C)** LASSO regression were conducted to analyze the different infiltrates of immune cells in RA and HC.

Functional Correlation Analysis

After being converted into gene ID, DEGs were analyzed by GO and KEGG. The GO annotations of DEGs consisted of three parts including CC (cellular component), BP (biological process), and MF (Molecular function), which were used to analyze the functional enrichment of DEGs. The DEGs were mainly related to the biological activity of immune cells, such as lymphocyte differentiation, T cell activation, and leukocyte (**Figures 5A, B** and **Supplementary Table S2**). KEGG analysis was conducted to determine the relationship between DEGs and signaling pathway. The DEGs were mainly associated with immune cell-related signaling pathway, such as chemokine signaling pathway, rheumatoid arthritis, and primary immunodeficiency (**Figures 5C, D** and **Supplementary Table S3**). Overall, the function of DEGs were significantly associated with immune cells.
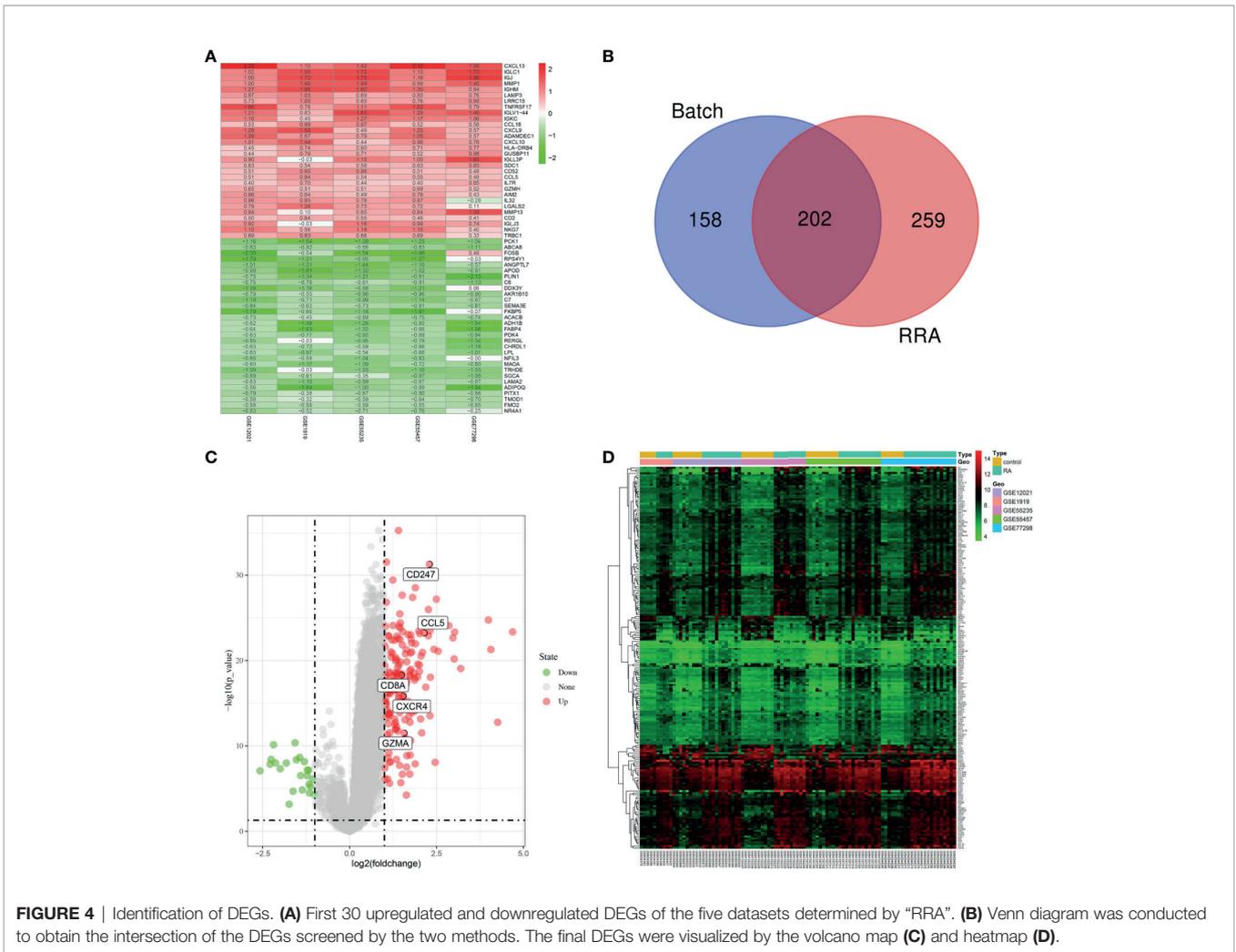
Identification and Validation of Hub Genes

We obtained the PPI network results of DEGs by STRING (**Supplementary Figure S3**). We then used 10 algorithms to calculate the score of each node gene. Finally, we screened five hub genes (CXCR4, CCL5, CD8A, CD247, and GZMA) by R package "UpSet" and marked them with boxes in **Figure 6A**. The

expression levels of the five hub genes were presented by heatmap in merged microarray data (**Figure 6B**). To make the results more reliable, we used the RNA-Seq dataset GSE89408 for validation. The expression levels of CXCR4, CCL5, CD8A, CD247, and GZMA were presented in heatmap (**Figure 7A**). As shown in **Figures 7B–F**, all the five genes had significantly higher expression in RA, both early and established RA, than in the HC (p <0.05). The expression levels of GZMA and CD8A in early RA were significantly higher than those in established RA (p <0.05). Consistent with the above results, the expression of CXCR4, CCL5, CD8A, CD247, and GZMA were also significantly increased in RA compared with osteoarthritis (OA) in dataset GSE89408 (**Figure 7G**). Therefore, GZMA and CD8A might be used as the diagnostic biomarkers for early RA, and CXCR4, CCL5, and CD247 might be used as the diagnostic biomarkers for RA.

Diagnostic Effectiveness of Biomarkers for RA

The GSE89408 dataset was used to validate the diagnostic effectiveness of the biomarkers for RA by ROC analysis. AUC more than 0.800 was considered as having the capability to diagnose

FIGURE 4 | Identification of DEGs. (A) First 30 upregulated and downregulated DEGs of the five datasets determined by "RRA". (B) Venn diagram was conducted to obtain the intersection of the DEGs screened by the two methods. The final DEGs were visualized by the volcano map (C) and heatmap (D).

RA with excellent specificity and sensitivity. As shown in **Figure 8A**, the AUC values of CCL5, CXCR4, and CD247 were 0.835 (95% CI 0.758–0.913), 0.900 (95% CI 0.847–0.953), and 0.797 (95% CI 0.724–0.871), respectively. Moreover, the combined AUC of CCL5 and CXCR4 reached 0.905 (95% CI 0.852–0.957). As shown in **Figure 8B**, the AUC values of GZMA and CD8A were 0.887 (95% CI 0.819–0.956) and 0.821 (95% CI 0.725–0.918), respectively. The combined AUC of GZMA and CD8A reached 0.900 (95% CI 0.837-0.963). Therefore, CCL5+CXCR4 and GZMA +CD8A had the capability to diagnose RA and early RA with excellent specificity and sensitivity, respectively.
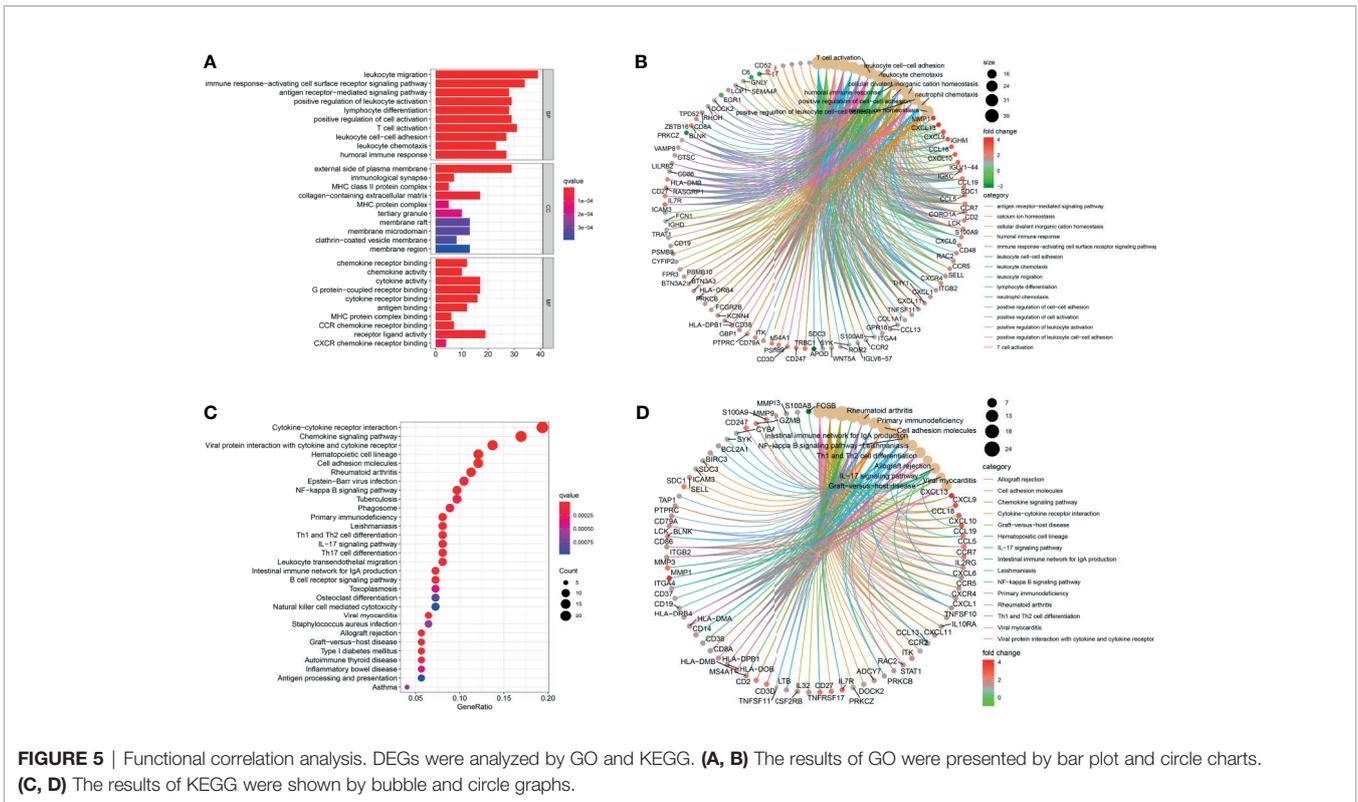
Correlation Between Biomarkers and Differential Immune Cells in RA

The correlation among four effective biomarkers (CCL5, CXCR4, GZMA, and CD8A) and 10 significantly differential immune cells (M1 macrophages, Tfh cells, memory activated CD4+ T cells, plasma cells, regulatory T cells, activated dendritic cells, activated NK cells, memory resting CD4+ T cells, resting mast cells, and activated mast cells) were analyzed in RA synovial tissues. The correlation results are presented in **Figure 9A**.

Significantly related biomarkers and immune cells were screened by R > 0.40 and p < 0.001. The results indicated that CCL5 was positively correlated with M1 macrophages (R = 0.47, p = 0.00038), CXCR4 was positively correlated with memory activated CD4+ T cells (R = 0.44, p = 0.00089) and Tfh cells (R = 0.70, p = 5e-09), and GZMA was positively correlated with Tfh (R = 0.53, p = 5e-05) (**Figures 9B–E**).

DISCUSSION

RA is a systemic autoimmune disease characterized by synovitis of joints. Innate and adaptive immune responses play an indispensable role in the pathogenesis of RA. Increasing number of studies have shown that the complex interaction and activation of infiltrating immune cells are key factors in the formation of synovitis and persistent joint damage. Effective biomarkers, especially for the early stage, have not been established due to the significant heterogeneity of RA. Early diagnosis and treatment of RA could effectively prevent the disease progression of 90% patients (22). Therefore, scholars

FIGURE 5 | Functional correlation analysis. DEGs were analyzed by GO and KEGG. **(A, B)** The results of GO were presented by bar plot and circle charts. **(C, D)** The results of KEGG were shown by bubble and circle graphs.

have focused on exploring the immune pathogenesis of RA and searching for novel effective biomarkers. In this study, we used comprehensive, objective, and effective bioinformatics methods to explore the role of immune cell infiltration in the synovium and identify effective diagnostic biomarkers for RA.

Compared with normal control M1 macrophages, the numbers of Tfh cells, memory activated CD4+ T cells, and plasma cells were significantly higher in RA synovial tissues, while those of regulatory T cells, activated dendritic cells, activated NK cells, memory resting CD4+ T cells, resting mast cells, and activated mast cells were significantly lower in RA synovial tissues. Macrophages can be polarized into M1 and M2 macrophages under different conditions due to their high degree of heterogeneity and plasticity. In the synovial environment of RA, macrophages mainly differentiate into type M1, which plays a pro-inflammatory role (23). The pathogenicity of M1 cells in RA is mainly realized by secreting cytokines, which in turn promote inflammation by monocyte/neutrophil recruitment, T cell polarization, and synovial fibroblast proliferation and activation (24). Tfh is a subtype of CD4+T cells, whose main functions are assisting B cells and regulating the production of antibodies (25). Tfh cell surface and secreted molecules, including CXCR5, ICOS (costimulatory molecule), and PD1 (programmed death factor 1), are involved in the development of RA (26, 27). Previous studies suggested that mast cells were aberrantly regulated in RA synovium and were mainly activated by TLRs (toll-like receptors), PAMP (pathogen-associated molecular patterns), and FcγR (Fc gamma receptor) (28, 29). Moreover, dendritic cells, NK cells, and regulatory T cells were

demonstrated to play a significant role in the development of RA (30, 31). Therefore, the present results are consistent with previous reports and highlights the importance of these cells in the pathogenesis of RA by bioinformatic analysis.

To improve the accuracy of the results, the DEGs were screened by two independent methods. Further analysis indicated that the function of these DEGs were significantly associated with immune cells. Finally, five hub genes (CXCR4, CCL5, CD8A, CD247, and GZMA) were screened, which were further verified by the validation dataset (GSE89408). Compared with osteoarthritis, all the five hub genes were also significantly higher expressed in RA, which made our result more convincing. Surprisingly, the expression levels of GZMA and CD8A were significantly higher in early RA than in established RA. Therefore, GZMA and CD8A might play an important role in the pathogenesis of early RA, while CXCR4, CCL5, and CD247 might play a vital role in the overall pathogenesis of RA. ROC regression analysis further found that CCL5+CXCR4 and GZMA+CD8A had the capability to diagnose RA and early RA with excellent specificity and sensitivity, respectively.

GZMA is a member of the granzyme family and is mainly secreted by cytotoxic cells (32). In addition to its cytotoxic activity against tumor and virus-infected cells together with perforin, GZMA is involved in innate host during inflammatory and autoimmune disorders in the absence of perforin (33). However, no data have identified the actual function of GZMA. Previous studies reported the elevated expression of GZMA in plasma, synovial fluid, and synovial tissue, indicating that GZMA might be involved in the development of RA (34–36). A recent research further
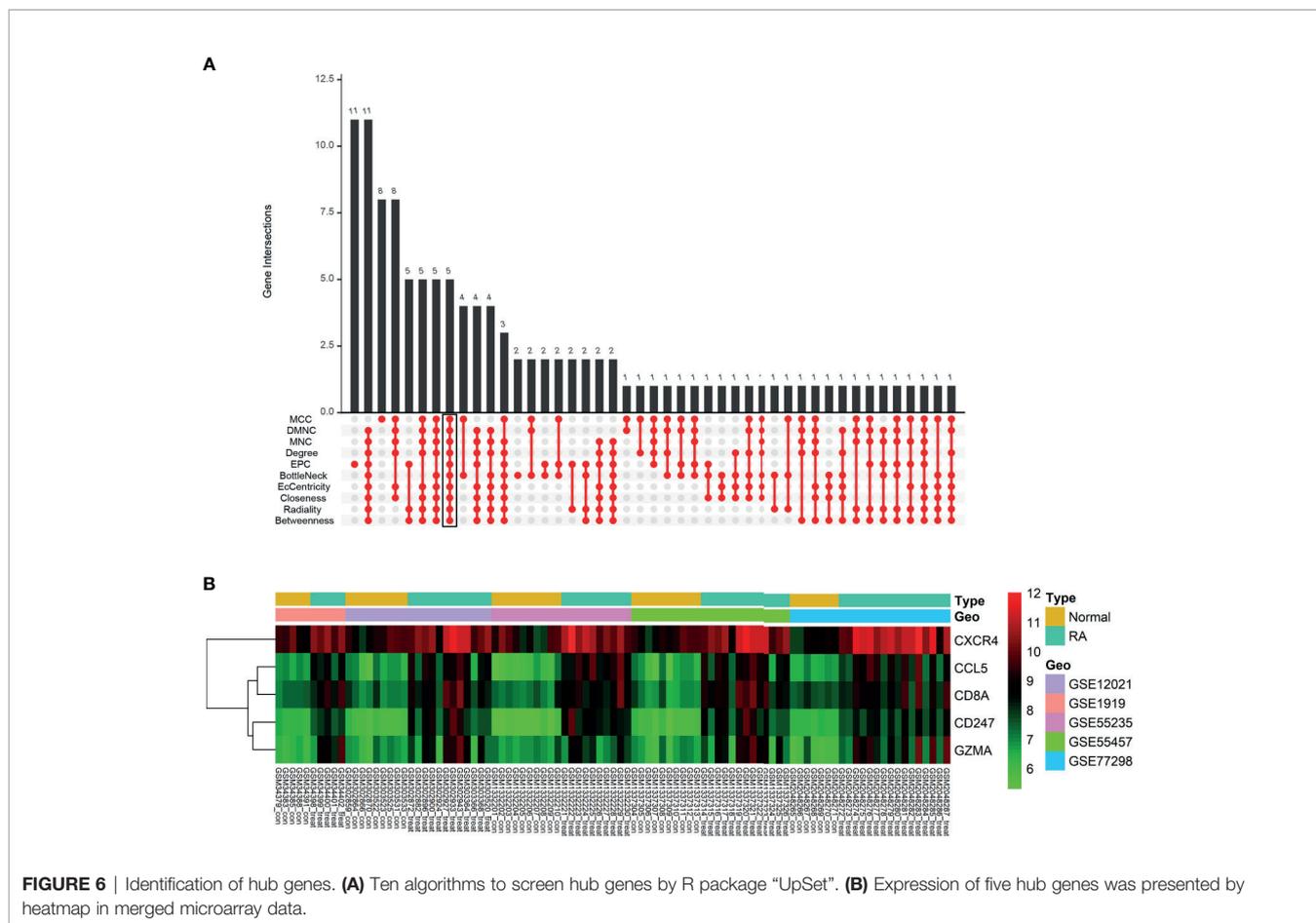
FIGURE 6 | Identification of hub genes. **(A)** Ten algorithms to screen hub genes by R package "UpSet". **(B)** Expression of five hub genes was presented by heatmap in merged microarray data.

suggested that GZMA contributed to the joint destruction in RA in part by promoting osteoclast differentiation (37). The latest mechanism research showed that GZMA from cytotoxic lymphocytes cleaves GSDMB to trigger pyroptosis in target cells (38). We also found that GZMA was mainly highly expressed in early RA. Therefore, GZMA might be used as a diagnostic biomarker because of its essential role in the pathogenesis in early RA. CD8A encodes the CD8α chain of the dimeric CD8 protein, which is mainly involved in cell-mediated immune defense and T-cell development (39, 40). CD8a could act as diagnosis and prognosis biomarker for many diseases, including tumors and inflammatory diseases (41–44). In the present study, CD8A was significantly overexpressed in RA, especially in early RA. The ROC regression analysis further demonstrated that the combination of GZMA and CD8A had the most excellent specificity and sensitivity for diagnosis of early RA. CCL5, belonging to the C-C chemokine family, is mainly secreted by T lymphocytes, macrophages, certain types of tumor cells, and synovial fibroblasts and mainly induces the recruitment and migration of immune cells to inflammatory sites (45, 46). CXCR4, which belongs to the family of 7-trasmembrane receptors, is involved in many pathophysiological processes by binding to CXCL12 and other non-specific ligands (47). Although numerous studies indicated that CCL5 and CXCR4 played an important role in the

progression of many diseases, including chronic liver disease, malignant tumor, and autoimmune disease (48–50), the specific mechanism needs to be further explored. Our study also suggests that CCL5 and CXCR4 are the hub genes of the entire process of RA development and might be used as the diagnose biomarker for this disease.

Given the important role of immune infiltrating cells and hub genes in RA, the correlation among four effective biomarkers (CCL5, CXCR4, GZMA and CD8A) and the top 10 significantly differential immune cells was further investigated in RA. CCL5 was positively correlated with M1 macrophages (R =0.47, p =0.00038), and CXCR4 was positively correlated with memory activated CD4+ T cells (R =0.44, p =0.00089) and Tfh cells (R =0.70, p =5e-09). Previous research revealed that CCL5 could directly activate M1 polarization, and CXCR4 had the ability to active memory activated CD4+ T cells and Tfh cells (51–53). Therefore, CCL5 and CXCR4 may participate in the occurrence and development of RA by regulating corresponding immune cells, which should be verified by further experiments. The most intriguing result was that GZMA was positively correlated with follicular helper T cells (R =0.53, p =5e-05). Studies have elucidated that Tfh cells and GZMA play a crucial role in infection and autoimmune diseases (54–56). M. Perreau et al. showed that HIV-infected individuals had a significantly higher frequency of Tfh cells among total CD4+ T cells compared with non-infected controls (57). Exaggerated
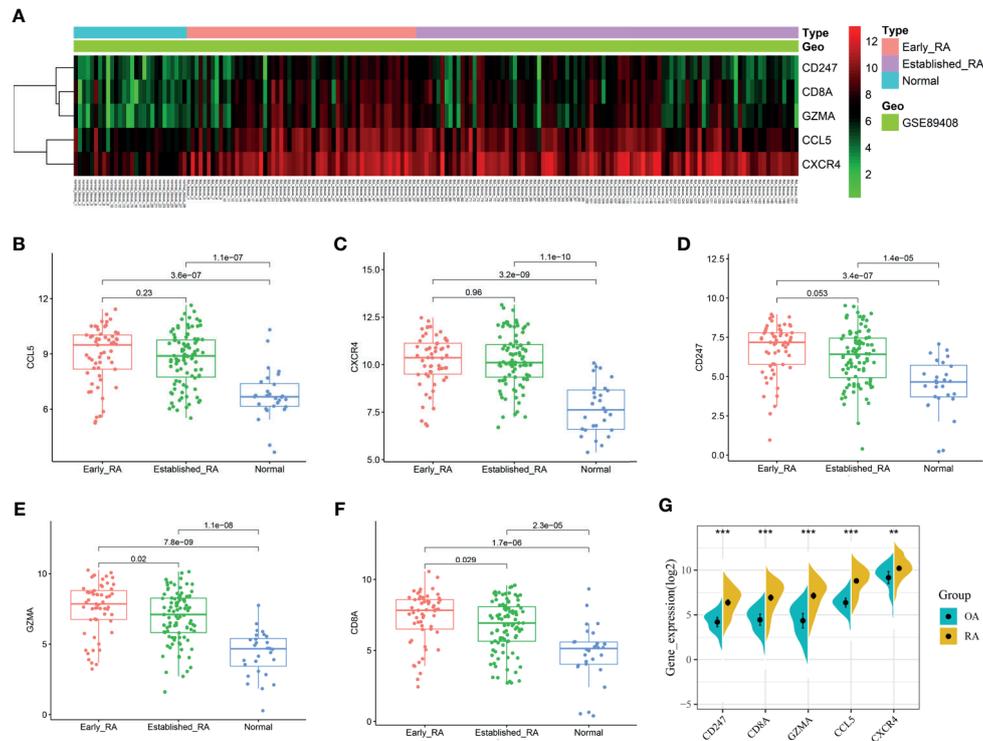
FIGURE 7 | Validation of hub genes. RNA-Seq dataset GSE89408 was used to validate the expression of CXCR4, CCL5, CD8A, CD247, and GZMA, the results of which were presented as heatmap **(A)**. **(B–F)** Detailed expression of five hub genes in early RA, established RA, and HC. **(G)** Detailed expression of five hub genes in RA and HC. **p < 0.01, ***p < 0.001.

expansion of Tfh cells resulted in self-reactive B cell proliferation, and increased long-lived plasma cell differentiation, as well as an overproduction of pathogenic autoantibodies (58). Repeated exposure to exogenous, endogenous, or symbiotic viruses and bacteria can lead to the high levels of pathogenic autoantibodies,

which act as a trigger to promote the occurrence of RA (3). Previous studies have shown that Tfh and GZMA participate in the pathogenesis of RA. Our results also indicated that GZMA was significantly overexpressed in early RA. Therefore, we speculate that infectious agents may play an important role in the early
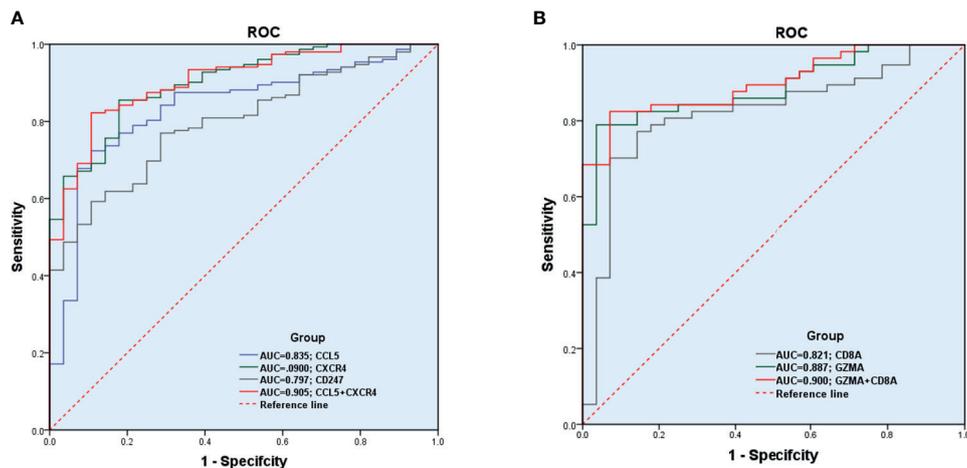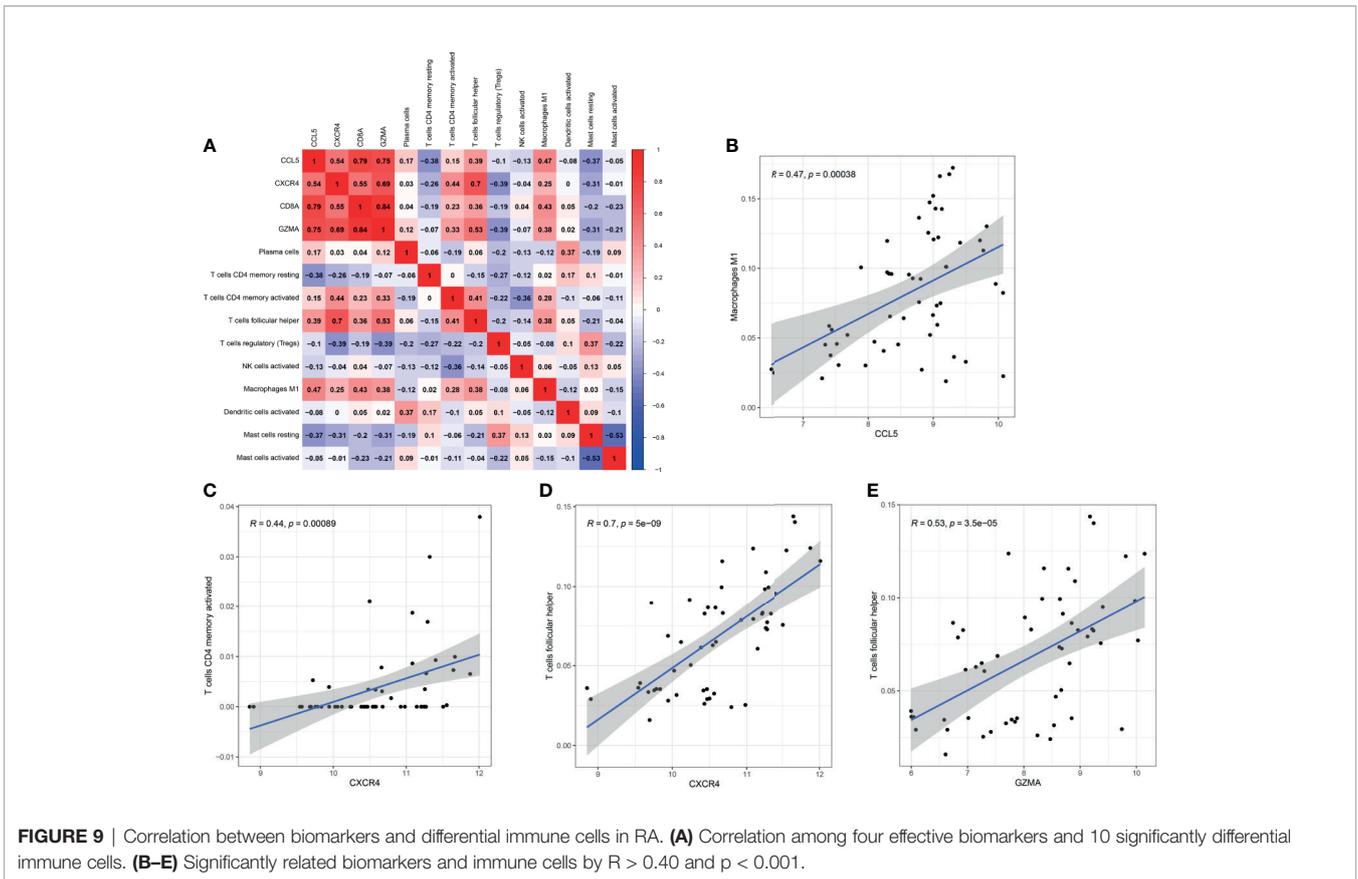

FIGURE 8 | Diagnostic effectiveness of the biomarkers for RA. **(A, B)** The GSE89408 dataset was used to validate the diagnostic effectiveness of the biomarkers for RA by ROC analysis.

FIGURE 9 | Correlation between biomarkers and differential immune cells in RA. (A) Correlation among four effective biomarkers and 10 significantly differential immune cells. (B–E) Significantly related biomarkers and immune cells by R > 0.40 and p < 0.001.

pathogenesis of RA through GZMA-Tfh cells axis, which needs further experimental verification.

This research has some limitations. First, the study lacks clinically relevant information, including the activity of the disease, and the use of drugs for the disease. Second, this study was conducted only from the perspective of gene transcriptome and lacks multi-group trials. Last, only bioinformatics methods were used for data analysis, and subsequent confirmatory experiments *in vivo* and *in vitro* are needed.

In summary, our study not only offers insights into the landscape of immune cells associated with RA but also identifies effective diagnostic biomarkers for RA. GZMA-Tfh cells, CCL5-M1 macrophages, and CXCR4- memory activated CD4+ T cells/Tfh cells may participate in the occurrence and development of RA; in particular, GZMA-Tfh cells may be involved in the early pathogenesis of RA. Therefore, this study may provide a new perspective for the diagnosis and immune cellular-molecular mechanism of RA.

DATA AVAILABILITY STATEMENT

Publicly available datasets (GSE1919, GSE12021, GSE55235, GSE55457, GSE77298 and GSE89408) were analyzed in this study. All the datasets were obtained from the GEO (http://www.ncbi.nlm.nih.gov/geo) database.

AUTHOR CONTRIBUTIONS

SZ analyzed and wrote the manuscript. HL designed the experiments and analyzed the data. MX devised the concept and supervised the study. All authors contributed to the article and approved the submitted version.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2021.726747/full#supplementary-material

Supplementary Figure 1 | The flow diagram of this study.

Supplementary Figure 2 | The correlation of 22 types of immune cells in HC synovial tissues was evaluated.

Supplementary Figure 3 | PPI network results of DEGs by STRING.

REFERENCES

1. Smolen JS, Aletaha D, Barton A, Burmester GR, Emery P, Firestein GS, et al. Rheumatoid Arthritis. *Nat Rev Dis Primers* (2018) 4:18001. doi: 10.1038/nrdp.2018.1

2. Smolen JS, Aletaha D, McInnes IB. Rheumatoid Arthritis. *Lancet* (2016) 388:2023–38. doi: 10.1016/S0140-6736(16)30173-8

3. McInnes IB, Schett G. Pathogenetic Insights From the Treatment of Rheumatoid Arthritis. *Lancet* (2017) 389:2328–37. doi: 10.1016/S0140-6736(17)31472-1

4. Gabriel SE. The Epidemiology of Rheumatoid Arthritis. *Rheum Dis Clin North Am* (2001) 27:269–81. doi: 10.1016/S0889-857X(05)70201-5

5. Karami J, Aslani S, Jamshidi A, Garshasbi M, Mahmoudi M. Genetic Implications in the Pathogenesis of Rheumatoid Arthritis; an Updated Review. *Gene* (2019) 702:8–16. doi: 10.1016/j.gene.2019.03.033

6. Firestein GS, McInnes IB. Immunopathogenesis of Rheumatoid Arthritis. *Immunity* (2017) 46:183–96. doi: 10.1016/j.immuni.2017.02.006

7. Siouti E, Andreakos E. The Many Facets of Macrophages in Rheumatoid Arthritis. *Biochem Pharmacol* (2019) 165:152–69. doi: 10.1016/j.bcp.2019.03.029

8. Elshabrawy HA, Chen Z, Volin MV, Ravella S, Virupannavar S, Shahrara S. The Pathogenic Role of Angiogenesis in Rheumatoid Arthritis. *Angiogenesis* (2015) 18:433–48. doi: 10.1007/s10456-015-9477-2

9. Meednu N, Zhang H, Owen T, Sun W, Wang V, Cistrone C, et al. Production of RANKL by Memory B Cells: A Link Between B Cells and Bone Erosion in Rheumatoid Arthritis. *Arthritis Rheumatol* (2016) 68:805–16. doi: 10.1002/art.39489

10. Hirota K, Hashimoto M, Ito Y, Matsuura M, Ito H, Tanaka M, et al. Autoimmune Th17 Cells Induced Synovial Stromal and Innate Lymphoid Cell Secretion of the Cytokine GM-CSF to Initiate and Augment Autoimmune Arthritis. *Immunity* (2018) 48:1220–32.e5. doi: 10.1016/j.immuni.2018.04.009

11. Yap HY, Tee SZ, Wong MM, Chow SK, Peh SC, Teow SY. Pathogenic Role of Immune Cells in Rheumatoid Arthritis: Implications in Clinical Treatment and Biomarker Development. *Cells* (2018) 7:161. doi: 10.3390/cells7100161

12. Shen-Orr SS, Gaujoux R. Computational Deconvolution: Extracting Cell Type-Specific Information From Heterogeneous Samples. *Curr Opin Immunol* (2013) 25:571–8. doi: 10.1016/j.coi.2013.09.015

13. Qiu L, Liu X. Identification of Key Genes Involved in Myocardial Infarction. *Eur J Med Res* (2019) 24:22. doi: 10.1186/s40001-019-0381-x

14. Zhao X, Zhang L, Wang J, Zhang M, Song Z, Ni B, et al. Identification of Key Biomarkers and Immune Infiltration in Systemic Lupus Erythematosus by Integrated Bioinformatics Analysis. *J Transl Med* (2021) 19:35. doi: 10.1186/s12967-021-02728-2

15. Deng YJ, Ren EH, Yuan WH, Zhang GZ, Wu ZL, Xie QQ. GRB10 and E2F3 as Diagnostic Markers of Osteoarthritis and Their Correlation With Immune Infiltration. *Diagnostics (Basel)* (2020) 10:171. doi: 10.3390/diagnostics10030171

16. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust Enumeration of Cell Subsets From Tissue Expression Profiles. *Nat Methods* (2015) 12:453–7. doi: 10.1038/nmeth.3337

17. Ungethuem U, Haeupl T, Witt H, Koczan D, Krenn V, Huber H, et al. Molecular Signatures and New Candidates to Target the Pathogenesis of Rheumatoid Arthritis. *Physiol Genomics* (2010) 42a:267–82. doi: 10.1152/physiolgenomics.00004.2010

18. Huber R, Hummert C, Gausmann U, Pohlers D, Koczan D, Guthke R, et al. Identification of Intra-Group, Inter-Individual, and Gene-Specific Variances in mRNA Expression Profiles in the Rheumatoid Arthritis Synovial Membrane. *Arthritis Res Ther* (2008) 10:R98. doi: 10.1186/ar2485

19. Woetzel D, Huber R, Kupfer P, Pohlers D, Pfaff M, Driesch D, et al. Identification of Rheumatoid Arthritis and Osteoarthritis Patients by Transcriptome-Based Rule Set Generation. *Arthritis Res Ther* (2014) 16:R84. doi: 10.1186/ar4526

20. Broeren MG, de Vries M, Bennink MB, Arntz OJ, Blom AB, Koenders MI, et al. Disease-Regulated Gene Therapy With Anti-Inflammatory Interleukin-10 Under the Control of the CXCL10 Promoter for the Treatment of Rheumatoid Arthritis. *Hum Gene Ther* (2016) 27:244–54. doi: 10.1089/hum.2015.127

21. Guo Y, Walsh AM, Fearon U, Smith MD, Wechalekar MD, Yin X, et al. CD40L-Dependent Pathway Is Active at Various Stages of Rheumatoid Arthritis Disease Progression. *J Immunol* (2017) 198:4490–501. doi: 10.4049/jimmunol.1601988

22. Aletaha D, Smolen JS. Diagnosis and Management of Rheumatoid Arthritis: A Review. *Jama* (2018) 320:1360–72. doi: 10.1001/jama.2018.13103

23. Di Benedetto P, Ruscitti P, Vadasz Z, Toubi E, Giacomelli R. Macrophages With Regulatory Functions, a Possible New Therapeutic Perspective in Autoimmune Diseases. *Autoimmun Rev* (2019) 18:102369. doi: 10.1016/j.autrev.2019.102369

24. Udalova IA, Mantovani A, Feldmann M. Macrophage Heterogeneity in the Context of Rheumatoid Arthritis. *Nat Rev Rheumatol* (2016) 12:472–85. doi: 10.1038/nrrheum.2016.91

25. Liu R, Zhao P, Zhang Q, Che N, Xu L, Qian J, et al. Adiponectin Promotes Fibroblast-Like Synoviocytes Producing IL-6 to Enhance T Follicular Helper Cells Response in Rheumatoid Arthritis. *Clin Exp Rheumatol* (2020) 38:11–8.

26. Gensous N, Charrier M, Duluc D, Contin-Bordes C, Truchetet ME, Lazaro E, et al. T Follicular Helper Cells in Autoimmune Disorders. *Front Immunol* (2018) 9:1637. doi: 10.3389/fimmu.2018.01637

27. Li Y, Bai S, Zhang J, Shen T, Chen S, Wang Y, et al. [T Follicular Helper Cells and Their Related Molecules in Rheumatoid Arthritis: A Review]. *Xi Bao Yu Fen Zi Mian Yi Xue Za Zhi* (2019) 35:281–6.

28. Min HK, Kim KW, Lee SH, Kim HR. Roles of Mast Cells in Rheumatoid Arthritis. *Korean J Intern Med* (2020) 35:12–24. doi: 10.3904/kjim.2019.271

29. Kopicky-Burd JA, Kagey-Sobotka A, Peters SP, Dvorak AM, Lennox DW, Lichtenstein LM, et al. Characterization of Human Synovial Mast Cells. *J Rheumatol* (1988) 15:1326–33.

30. Wehr P, Purvis H, Law SC, Thomas R. Dendritic Cells, T Cells and Their Interaction in Rheumatoid Arthritis. *Clin Exp Immunol* (2019) 196:12–27. doi: 10.1111/cei.13256

31. Schwaneck EC, Renner R, Junker L, Tony HP, Kleinert S, Gernert M, et al. T Cells, Natural Killer Cells, and γδt Cells in a Large Patient Cohort With Rheumatoid Arthritis: Influence of Age and Anti-Rheumatic Therapy. *Scand J Rheumatol* (2020) 49:8–12. doi: 10.1080/03009742.2019.1634755

32. Trapani JA, Smyth MJ. Functional Significance of the Perforin/Granzyme Cell Death Pathway. *Nat Rev Immunol* (2002) 2:735–47. doi: 10.1038/nri911

33. Shimizu K, Yamasaki S, Sakurai M, Yumoto N, Ikeda M, Mishima-Tsumagari C, et al. Granzyme A Stimulates pDCs to Promote Adaptive Immunity *via* Induction of Type I IFN. *Front Immunol* (2019) 10:1450. doi: 10.3389/fimmu.2019.01450

34. Griffiths GM, Alpert S, Lambert E, McGuire J, Weissman IL. Perforin and Granzyme A Expression Identifying Cytolytic Lymphocytes in Rheumatoid Arthritis. *Proc Natl Acad Sci USA* (1992) 89:549–53. doi: 10.1073/pnas.89.2.549

35. Kummer JA, Tak PP, Brinkman BM, van Tilborg AA, Kamp AM, Verweij CL, et al. Expression of Granzymes A and B in Synovial Tissue From Patients With Rheumatoid Arthritis and Osteoarthritis. *Clin Immunol Immunopathol* (1994) 73:88–95. doi: 10.1006/clin.1994.1173

36. Tak PP, Spaeny-Dekking L, Kraan MC, Breedveld FC, Froelich CJ, Hack CE. The Levels of Soluble Granzyme A and B are Elevated in Plasma and Synovial Fluid of Patients With Rheumatoid Arthritis (RA). *Clin Exp Immunol* (1999) 116:366–70. doi: 10.1046/j.1365-2249.1999.00881.x

37. Santiago L, Menaa C, Arias M, Martin P, Jaime-Sánchez P, Metkar S, et al. Granzyme A Contributes to Inflammatory Arthritis in Mice Through Stimulation of Osteoclastogenesis. *Arthritis Rheumatol* (2017) 69:320–34. doi: 10.1002/art.39857

38. Zhou Z, He H, Wang K, Shi X, Wang Y, Su Y, et al. Granzyme A From Cytotoxic Lymphocytes Cleaves GSDMB to Trigger Pyroptosis in Target Cells. *Science* (2020) 368:eaaz7548. doi: 10.1126/science.aaz7548

39. Tregaskes CA, Kong FK, Paramithiotis E, Chen CL, Ratcliffe MJ, Davison TF, et al. Identification and Analysis of the Expression of CD8 Alpha Beta and CD8 Alpha Alpha Isoforms in Chickens Reveals a Major TCR-Gamma Delta CD8 Alpha Beta Subset of Intestinal Intraepithelial Lymphocytes. *J Immunol* (1995) 154:4485–94.

40. Xu Q, Chen Y, Zhao WM, Huang ZY, Zhang Y, Li X, et al. DNA Methylation and Regulation of the CD8A After Duck Hepatitis Virus Type 1 Infection. *PloS One* (2014) 9:e88023. doi: 10.1371/journal.pone.0088023

41. Ma K, Qiao Y, Wang H, Wang S. Comparative Expression Analysis of PD-1, PD-L1, and CD8A in Lung Adenocarcinoma. *Ann Transl Med* (2020) 8:1478. doi: 10.21037/atm-20-6486

42. Kristensen LK, Fröhlich C, Christensen C, Melander MC, Poulsen TT, Galler GR, et al. CD4(+) and CD8a(+) PET Imaging Predicts Response to Novel PD-

1 Checkpoint Inhibitor: Studies of Sym021 in Syngeneic Mouse Cancer Models. *Theranostics* (2019) 9:8221–38. doi: 10.7150/thno.37513

43. Kristensen LK, Christensen C, Alfsen MZ, Cold S, Nielsen CH, Kjaer A. Monitoring CD8a(+) T Cell Responses to Radiotherapy and CTLA-4 Blockade Using [(64)Cu]NOTA-CD8a PET Imaging. *Mol Imaging Biol* (2020) 22:1021–30. doi: 10.1007/s11307-020-01481-0

44. Alromaih S, Mfuna-Endam L, Bosse Y, Filali-Mouhim A, Desrosiers M. CD8A Gene Polymorphisms Predict Severity Factors in Chronic Rhinosinusitis. *Int Forum Allergy Rhinol* (2013) 3:605–11. doi: 10.1002/alr.21174

45. Marques RE, Guabiraba R, Russo RC, Teixeira MM. Targeting CCL5 in Inflammation. *Expert Opin Ther Targets* (2013) 17:1439–60. doi: 10.1517/14728222.2013.837886

46. Aldinucci D, Colombatti A. The Inflammatory Chemokine CCL5 and Cancer Progression. *Mediators Inflamm* (2014) 2014:292376. doi: 10.1155/2014/292376

47. Pozzobon T, Goldoni G, Viola A, Molon B. CXCR4 Signaling in Health and Disease. *Immunol Lett* (2016) 177:6–15. doi: 10.1016/j.imlet.2016.06.006

48. Mohs A, Kuttkat N, Reißing J, Zimmermann HW, Sonntag R, Proudfoot A, et al. Functional Role of CCL5/RANTES for HCC Progression During Chronic Liver Disease. *J Hepatol* (2017) 66:743–53. doi: 10.1016/j.jhep.2016.12.011

49. Seo W, Shimizu K, Kojo S, Okeke A, Kohwi-Shigematsu T, Fujii SI, et al. Runx-Mediated Regulation of CCL5 *via* Antagonizing Two Enhancers Influences Immune Cell Function and Anti-Tumor Immunity. *Nat Commun* (2020) 11:1562. doi: 10.1038/s41467-020-15375-w

50. Kawaguchi N, Zhang TT, Nakanishi T. Involvement of CXCR4 in Normal and Abnormal Development. *Cells* (2019) 8;185. doi: 10.3390/cells8020185

51. Li M, Sun X, Zhao J, Xia L, Li J, Xu M, et al. CCL5 Deficiency Promotes Liver Repair by Improving Inflammation Resolution and Liver Regeneration Through M2 Macrophage Polarization. *Cell Mol Immunol* (2020) 17:753–64. doi: 10.1038/s41423-019-0279-0

52. Serre K, Mohr E, Bénézech C, Bird R, Khan M, Caamaño JH, et al. Selective Effects of NF-κb1 Deficiency in CD4⁺ T Cells on Th2 and TFh Induction by Alum-Precipitated Protein Vaccines. *Eur J Immunol* (2011) 41:1573–82. doi: 10.1002/eji.201041126

53. Yu H, Hu W, Song X, Zhao Y. Immune Modulation of Platelet-Derived Mitochondria on Memory CD4(+) T Cells in Humans. *Int J Mol Sci* (2020) 21:6295. doi: 10.3390/ijms21176295

54. Vinuesa CG, Linterman MA, Yu D, MacLennan IC. Follicular Helper T Cells. *Annu Rev Immunol* (2016) 34:335–68. doi: 10.1146/annurev-immunol-041015-055605

55. Schanoski AS, Le TT, Kaiserman D, Rowe C, Prow NA, Barboza DD, et al. Granzyme A in Chikungunya and Other Arboviral Infections. *Front Immunol* (2019) 10:3083. doi: 10.3389/fimmu.2019.03083

56. Arias MA, Jiménez de Bagües MP, Aguiló N, Menao S, Hervás-Stubbs S, de Martino A, et al. Elucidating Sources and Roles of Granzymes A and B During Bacterial Infection and Sepsis. *Cell Rep* (2014) 8:420–9. doi: 10.1016/j.celrep.2014.06.012

57. Perreau M, Savoye AL, De Crignis E, Corpataux JM, Cubas R, Haddad EK, et al. Follicular Helper T Cells Serve as the Major CD4 T Cell Compartment for HIV-1 Infection, Replication, and Production. *J Exp Med* (2013) 210:143–56. doi: 10.1084/jem.20121932

58. Qin L, Waseem TC, Sahoo A, Bieerkehazhi S, Zhou H, Galkina EV, et al. Insights Into the Molecular Mechanisms of T Follicular Helper-Mediated Immunity and Pathology. *Front Immunol* (2018) 9:1884. doi: 10.3389/fimmu.2018.01884