Identification of CIRBP and TRPV4 as Immune-Related Diagnostic Biomarkers in Osteoarthritis

Chengyang Tang,1,* Qiang Liu,2,* Yaxuan Zhang,1 Guihu Liu,1 Guangsi Shen1

1Department of Orthopaedics, The Second Affiliated Hospital of Soochow University, Suzhou, People’s Republic of China; 2Institution of Sports Medicine, Peking University Third Hospital, Beijing Key Laboratory of Sports Injuries, Beijing, People’s Republic of China

*These authors contributed equally to this work

Purpose: Osteoarthritis (OA) is the most common chronic joint disorder in elderly individuals. This study aimed to identify immune-related diagnostic gene signatures for OA.

Methods: First, we performed single-sample gene set enrichment analysis (ssGSEA) to evaluate the infiltration of immune cells in OA expression data from the Gene Expression Omnibus (GEO) database. Then, weighted gene coexpression network analysis (WGCNA) was performed to identify hub modules and genes related to immune cell types with significant infiltration. Finally, we screened diagnostic markers from the differentially expressed genes (DEGs) in both the OA group and the hub module using least absolute shrinkage and selection operator (LASSO) logistic regression.

Results: Immune filtration analysis showed that immature B cells, mast cells, natural killer T cells, myeloid-derived suppressor cells (MDSCs), and type 2 T helper cells were dysregulated in OA samples. In WGCNA, a total of 120 genes were selected as hub genes associated with mast cell infiltration. The enrichment analysis showed that spliceosome, positive regulation of cell migration, and response to mechanical stimulus were mainly involved. The LASSO regression model for the GSE117999 dataset revealed 15 DEGs for predicting OA. Finally, two genes were obtained by intersection for further investigation.

Conclusion: Cold-inducible RNA-binding protein (CIRBP) and transient receptor potential vanilloid 4 (TRPV4) were identified as diagnostic biomarkers for OA, and both were positively correlated with mast cell infiltration.

Keywords: osteoarthritis, bioinformatics analysis, immune infiltration, WGCNA

Introduction

Osteoarthritis (OA) is a common degenerative arthritic disease that is increasingly prevalent in elderly individuals and leads to pain, joint destruction, and disability.1 OA imposes a tremendous economic burden on individuals and society, mainly due to the clinical inability of early disease diagnosis.2 Traditionally, the diagnosis of OA is based on clinical symptoms and radiography findings. However, both methods lack insensitivity and have limited value in diagnosis in the early stages of OA. Although joint replacement surgery can be effectively applied at the advanced stage of OA, there are relatively few interventions for addressing earlier stages of OA.3 Therefore, the investigation of diagnostic biomarkers is an important focus to identify patients with early-stage OA.

The critical role of immune cell infiltration in the pathophysiology of OA has been intensively investigated.4 Macrophages are the most common cells in OA synovial...
tissues, and the quantity of activated macrophages is directly corrected with OA severity. Macrophages promote the release of proinflammatory cytokines, growth factors, and matrix metalloproteinases and subsequently induce cartilage breakdown and joint deterioration. Additionally, many studies have described the pathological role of T cells in OA. In OA synovial tissues, CD4+ T cell infiltration is frequent, and increased CD4+/CD8+ ratios are observed. To gain more insight, Li et al investigated the role of each T cell subtype, including Th1 cells, Th2 cells, Th9 cells, Th17 cells, Th22 cells, regulatory T cells, follicular helper T cells, cytotoxic T cells, and T memory cells, in the pathogenesis of OA. These results strongly suggest that immune cells have great potential for the diagnosis and treatment of OA patients.

Extending upon findings from our previous studies, where we demonstrated the promoting roles of OA-related noncoding RNAs in cartilage degradation, the current study aimed to identify potential diagnostic biomarker candidates for OA and provide insights into the critical importance of immune infiltration in OA. To begin with, we conducted single-sample gene set enrichment analysis (ssGSEA) to calculate the immune cell proportions in samples. Then, we performed weighted gene coexpression network analysis (WGCNA) to identify modules and genes correlated with the infiltration levels of 28 immune cell types. Moreover, we used least absolute shrinkage and selection operator (LASSO) regression analysis to identify diagnostic biomarkers in differentially expressed genes (DEGs). Finally, we screened the genes in both LASSO and immune cell infiltration modules for further analysis. To date, we are the first to combine ssGSEA, WGCNA, and LASSO regression analysis to identify immune-related biomarkers in OA. Our research not only provides promising targets for OA but also extends the knowledge of the underlying mechanisms.

Materials and Methods

Data Preprocessing and DEGs Screening

We downloaded the OA expression profile datasets GSE117999 and GSE12021 from the Gene Expression Omnibus (GEO) database. The dataset GSE117999 contains gene transcripts expressed in cartilage tissues obtained from 12 patients with OA and from 12 patients without OA (arthroscopic partial meniscectomy). The data were normalized using the limma package. Probes were annotated to the genes. Due to the multiple matches between genes and probes, the mean value was calculated as the expression level of genes. Then, we performed differential expression analysis to screen DEGs between OA patients and normal controls. The cutoff criteria were set at \( P<0.05 \) and \( |\text{logFC}|>0.5 \).

Evaluation of Immune Cell Infiltration

ssGSEA is a modification of standard gene set enrichment analysis, was performed to identify immune cell types that excessively infiltrated OA tissues using the GSVA package. We obtained published cell type-specific markers compiled from 28 immune cell types. Principal component analysis (PCA) was performed to analyze the immune cell infiltration matrix data and find differences between groups. Then, we used the ggplot2 package to draw violin plots to visualize the relative proportions of immune cell subpopulations from the expression profiles.

Weighted Coexpression Network Construction and Module Detection

The R package WGCNA was used to construct a weighted coexpression network based on the expression values of 4561 genes. First, the proper soft-thresholding power \( \beta \) was chosen based on the criterion of approximate scale-free topology. The function pickSoftThreshold was used to analyze network topology and calculate the soft-thresholding power \( \beta \). The network was constructed with the automatic network construction function. Second, according to the dissimilarity coefficient between genes, a hierarchical clustering tree was built for module detection. Gene significance (GS) and module membership (MM) were defined to quantify the correlation between modules and clinical characteristics. Ranked by the absolute value of module significance (MS), modules that showed a strong association with the specific cell subtypes were selected as hub modules.

Functional Enrichment Analysis

Genes in the identified hub modules were extracted for pathway and process enrichment analyses using the web tool Metascape. The top 20 enriched terms with a \( P<0.01 \), a minimum count of 3, and an enrichment factor > 1.5 are presented in a bar graph.
Identification of Diagnostic Markers
We performed LASSO logistic regression\textsuperscript{19} to select diagnostic markers for OA using the glmnet package.\textsuperscript{20} The overlapping genes in the LASSO logistic regression and hub modules were investigated as hub genes for further analysis.

Correlation Analysis Between Diagnostic Markers and Immune Cells
Spearman correlation analysis between the infiltration levels of immune cells and the expression of diagnostic markers was performed, and the results were visualized using the R package ggplot2.

Results
GEO Dataset and DEGs
We obtained the gene expression profile data for 24 samples from the GEO database, and four of them (GSM3317970, GSM3317972, GSM3317976, and GSM3317980) were eliminated due to having too many missing values. In total, the current study included 10 patients with meniscectomy (5 females, 5 males; mean age, 49.7 years; age range, 37–65 years) and 10 OA patients (7 females, 3 males; mean age, 66 years; age range, 57–80 years). No patients were excluded for additional diseases. After data preprocessing, we extracted a total of 100 DEGs from the gene expression matrix. The distribution of DEGs between OA patients and normal controls was visualized by the volcano plot (Figure 1A). In addition, the heat map shows the differences in gene expression between groups (Figure 1B). DEGs were used for the construction of the LASSO regression model.

Proportion of OA-Infiltrating Immune Cells
To identify the cell types potentially involved in the pathology of OA, we performed ssGSEA, which calculates the rank value of each gene according to the expression profile. In this way, we obtained the proportions of 28 immune cell types in each sample. The results of PCA cluster analysis revealed a significant difference in immune cell infiltration between OA patients and normal controls (Figure 2A). Compared with those in non-OA tissues, the proportions of immature B cells and natural killer T cells were relatively low, while the proportions of mast cells, myeloid-derived suppressor cells (MDSCs), and type 2 T helper cells were high in OA tissues (Figure 2B). Subsequently, the fractions of five immune cell types in each sample were selected as external traits for WGCNA.

Gene Coexpression Network of OA
The expression values of the 4561 genes were used to construct a weighted gene coexpression network. First, we calculated the average linkage and Pearson’s correlation values to cluster the samples from the GSE117999 dataset (Figure 3A). In addition, we chose a height cut of 40 and removed two offending samples (Supplementary Figure S1). To build a scale-free network, we selected the soft-thresholding power $\beta$ as 7. Next, we performed one-step network construction and module detection. A hierarchical clustering tree was used for module identification (Figure 3B), and 21 modules were generated. Last and most importantly, the module eigengenes, the first principal component of a module, was corrected with the external traits. Moreover, to quantify the associations of individual genes with traits, GS and MM were calculated.

Identification of Hub Modules and Enrichment Analysis
As shown in Figure 4A, the green module showed a high correlation with mast cells ($R^2=0.66, P=1e-03$); the brown module was correlated with MDSCs ($R^2=0.57, P=1e-02$) and type 2 T helper cells ($R^2=0.74, P=4e-04$); the tan module was correlated with MDSCs ($R^2=0.66, P=3e-03$) and the light cyan module was related to type 2 T helper cells ($R^2=0.51, P=3e-02$). The correlation coefficient between other modules and immune cells was less than 0.5. The green module that showed a significant correlation with mast cells captured our attention. To further explore the biological functions of genes in the green module, enrichment analyses were performed, and the top 20 enriched terms are presented in a bar graph (Figure 4B).

Screening the Immune-Related Diagnostic Gene Signature
On the one hand, the central genes in interesting modules had the highest association with traits. According to the cutoff criteria ($MM > 0.5$ and $GS> 0.5$), 120 genes were selected as hub genes associated with mast cell infiltration (Figure 5A). On the other hand, the LASSO regression model for the GSE117999 dataset revealed 15 DEGs for predicting OA (Figure 5B and Supplementary Figure S2).
Finally, we screened two overlapping genes from the predictive markers and hub genes: cold-inducible RNA-binding protein (CIRBP) and transient receptor potential vanilloid 4 (TRPV4) (Figure 5C). Additionally, receiver operating characteristic (ROC) curve analysis of each gene was applied. The GSE12021 dataset was set as the validation set. The results suggested that CIRBP and TRPV4 might have outstanding diagnostic value in OA (Figure 5D and E).
Correlation of CIRBP and TRPV4 with Immune Cell Infiltration

We performed Spearman correlation analysis of the expression of CIRBP and TRPV4 in immune cells. As shown in Figure 6, CIRBP was positively correlated with mast cells ($r=0.81$, $P<0.01$) and TRPV4 was positively correlated with mast cells ($r=0.55$, $P=0.01$).

Discussion

OA is a degenerative joint disorder that leads to pain and disability. Conventional imaging techniques have poor detection ability for the early stages of disease, and OA patients often lose the best chance for treatment. Thus, disease prevention and early diagnosis should be the critical focus of OA research. Recent advances in molecular markers present a wide range of possibilities for the early diagnosis and treatment of OA. The LASSO regression model is a popular method for dimension reduction, parameter estimation, and variable selection. WGCNA is an effective method that calculates correlations between the expression levels of genes and external disease traits. Both algorithms can be used to detect biomarkers of disease, and their combination provides accurate and convincing results. To gain more insight into the role of immune cell infiltration in OA, we performed ssGSEA to estimate the relative proportions of OA-infiltrating immune cells.
and increased infiltration of mast cells was observed. The enrichment analysis revealed that hub genes extracted from the green module majorly participated in spliceosome, positive regulation of cell migration, response to mechanical stimulus, and so on. Studies have reported that the pathology of OA is related to the migration of multiple cell types, including chondrocytes, synoviocytes, mesenchymal stem cells, and endothelial cells.22,23 Our results indicated that CIRBP and TRPV4 whose expression levels are strongly associated with mast cell infiltration, can be used as diagnostic markers of OA.

CIRBP belongs to the glycine-rich RNA-binding protein family. It acts as a vital mediator in complex biological processes, including the circadian rhythm, tumorigenesis, the immune response, and inflammation.24 A number of studies have demonstrated that CIRBP plays a critical role in the immune response. Activated by the Toll-like receptor 4 (TLR4)-myeloid differentiation factor

Figure 3 Sample clustering and module detection. (A) Clustering dendrogram. The heat map presents the degree of infiltration of immune cells in each sample. (B) Clustering dendrogram of genes, with dissimilarity based on topological overlap, together with assigned module colors.
2 (MD2) complex, CIRBP could promote the release of tumor necrosis factor-alpha (TNF-α) and interleukin-6 (IL-6) in shock and sepsis. Bolognese et al reported that CIRBP acts as a proinflammatory mediator that predisposes CD4(+) T cells to a Th1 hyperinflammatory response profile and drives CD8(+) T cells to a cytotoxic profile. Moreover, studies have proved that CIRBP induces various inflammatory reactions. It is well established that inflammation has a direct connection with the pathogenesis of OA. Yoo et al observed an increase in the concentrations of CIRBP in rheumatoid arthritis tissues and positive correlations between CIRBP and disease activity. A recent study demonstrated that, in OA patients, the expression of CIRBP in the synovial fluid was markedly higher than that in the serum; additionally, inflammatory molecules such as TNF-α and IL-6 were highly correlated with CIRBP. Thus, CIRBP plays a critical role in the immune response and inflammatory reactions, suggesting that CIRBP might be a potential marker for the detection and diagnosis of OA.

TRPV4 encodes a member of the TRPV subfamily of TRP ion channels and is a Ca2+-permeable, nonselective cation channel involved in sensing mechanical or osmotic pressure. Much work thus far has focused on the critical role of TRPV4 in maintaining the health and function of joint tissues, particularly cartilage. By activating TRPV4, mechanical stimuli increase intracellular Ca2+ concentrations and mediate metabolic processes. A previous study confirmed that by mediating Ca2+ signaling, TRPV4 was a vital factor in transducing...
mechanical signals and supporting cartilage extracellular matrix maintenance and joint health. \(^{36}\) Similarly, Xu et al demonstrated that the upregulation of TRPV4 in OA articular cartilage resulted from excessive mechanical stress and induced chondrocyte apoptosis via Ca²⁺ influx. \(^{37}\) Hence, the identification of downstream targets of chondrocyte TRPV4-mediated Ca²⁺ signaling in the cartilage may provide potential targets for OA therapeutics. \(^{38}\) Furthermore, many inflammatory cytokines, such as interleukin-1 and TNF-α, are regulated by TRPV4 activity, and their downstream effectors participate in the pathogenesis of OA. \(^{39–41}\) In conclusion, TRPV4 plays a crucial role in many biological processes, including transducing mechanical signatures, inducing hyperalgesia, \(^{42}\) and mediating inflammatory cytokines in musculoskeletal tissues. We believe that TRPV4 is an attractive biomarker for novel targets or treatments of OA.

The correlation analysis between CIRBP, TRPV4 and immune cells revealed that the expression of both genes was positively correlated with mast cells. Mast cells mainly reside in the synovium and synovial fluids and have a crucial influence on various processes, including cytokine recruitment and activation, inflammation, and cartilage destruction. \(^{43}\) Studies have investigated the increasing importance of mast cells and their mediators in the pathogenesis of OA. \(^{7,44}\) A recent study reported a positive correlation between synovial mast cell numbers and structural damage in OA patients. \(^{45}\) Wang et al demonstrated that IgE-dependent mast cell activation might result in mast cell degranulation and the release of proinflammatory mediators, leading to cartilage destruction and the occurrence of OA. \(^{46}\) These observations suggest a central role for mast cells in the pathophysiology of OA. Based on our findings, it can be speculated that the overexpression of CIRBP and TRPV4 promotes mast cell infiltration and subsequently leads to the occurrence and development of OA.

It should be pointed out that this work is an early exploratory study due to the following limitations. First, the numbers of samples and genetic data were insufficient.

Figure 5 Screening the immune-related gene diagnostic signature. (A) A scatterplot of module eigengenes in the green module. (B) The least absolute shrinkage and selection operator logistic regression algorithm revealed 15 genes associated with diagnosis. (C) Venn diagram of the overlapping genes between the two algorithms. (D) ROC curve analysis of the diagnostic efficacy of identified genes in GSE11799. (E) ROC curve analysis of the diagnostic efficacy of identified genes in GSE12021.

Abbreviation: AUC, area under the receiver operating characteristic curve.
Second, the accuracy of the immune-related diagnostic gene signature needs to be assessed in additional datasets. Last, our research was based on public datasets and requires experimental validation.

**Conclusion**

In conclusion, the current study innovatively combined LASSO logistic regression, ssGSEA, and WGCNA algorithms to screen diagnostic biomarkers associated with specific immune cells for OA. We found that CIRBP and TRPV4 were overexpressed in osteoarthritic tissues, and they were positively correlated with mast cell infiltration. These findings reveal a new strategy for the early diagnosis and immunomodulatory therapies of OA. Future work will focus on determining pathogenic targets and investigating immunopathology in OA.

**Abbreviations**

OA, osteoarthritis; ssGSEA, single-sample gene set enrichment analysis; GEO, Gene Expression Omnibus; WGCNA, weighted gene coexpression network analysis; DEGs, differentially expressed genes; LASSO, least absolute shrinkage and selection operator; MDSCs, myeloid-derived suppressor cells; CIRBP, cold-inducible RNA-binding protein; TRPV4, transient receptor potential vanilloid 4; PCA, principal component analysis; GS, gene significance; MM, module membership; MS, module significance; ROC, receiver operating characteristic; AUC, area under the receiver operating characteristic curve; TLR4, Toll-like receptor 4; MD2, differentiation factor 2; TNF-α, tumor necrosis factor-alpha; IL-6, interleukin-6.

**Data Sharing Statement**

The data in this study can be obtained from the corresponding author upon reasonable request.

**Ethical Approval**

This study has been granted an exemption by Medical Ethics Committee of The Second Affiliated Hospital of Soochow University. The public database mentioned in this study is publicly available for re-analyzing, and no ethical approval was required by the local ethics committees, so that this study does not require ethics approval.

**Acknowledgments**

We would like to acknowledge the reviewers for their invaluable comments on this study.

**Funding**

This work was funded by the Training project of “national tutor system” for young health talents in Suzhou (grant number 20042), Gusu Health Talents (grant number...
Disclosure
The authors declare that they have no competing interests.

References