

Transcriptomic analyses and machine-learning methods reveal dysregulated key genes and potential pathogenesis in human osteoarthritic cartilage

Cite this article:
Bone Joint Res 2024;13(2):
66–82.

DOI: 10.1302/2046-3758.
132.BJR-2023-0074.R1

Correspondence should be
sent to Jun Liu liujun3040@126.com

D. Zhao,^{1,2} L. Zeng,^{2,3} G. Liang,^{2,3} M. Luo,^{2,3} J. Pan,^{2,3} Y. Dou,^{2,3} F. Lin,^{1,2} H. Huang,^{2,3} W. Yang,^{2,3} J. Liu^{2,4,5}

¹Second Clinical Medical College of Guangzhou University of Chinese Medicine, Guangzhou, China

²Bone and Joint Research Team of Degeneration and Injury, Guangdong Provincial Academy of Chinese Medical Sciences, Guangzhou, China

³Department of Orthopedics, The Second Affiliated Hospital of Guangzhou University of Chinese Medicine, Guangzhou, China

⁴Guangdong Second Traditional Chinese Medicine Hospital, Guangdong Province Engineering Technology Research Institute of Traditional Chinese Medicine, Guangzhou, China

⁵Fifth Clinical Medical College of Guangzhou University of Chinese Medicine, Guangzhou, China

Aims

This study aimed to explore the biological and clinical importance of dysregulated key genes in osteoarthritis (OA) patients at the cartilage level to find potential biomarkers and targets for diagnosing and treating OA.

Methods

Six sets of gene expression profiles were obtained from the Gene Expression Omnibus database. Differential expression analysis, weighted gene coexpression network analysis (WGCNA), and multiple machine-learning algorithms were used to screen crucial genes in osteoarthritic cartilage, and genome enrichment and functional annotation analyses were used to decipher the related categories of gene function. Single-sample gene set enrichment analysis was performed to analyze immune cell infiltration. Correlation analysis was used to explore the relationship among the hub genes and immune cells, as well as markers related to articular cartilage degradation and bone mineralization.

Results

A total of 46 genes were obtained from the intersection of significantly upregulated genes in osteoarthritic cartilage and the key module genes screened by WGCNA. Functional annotation analysis revealed that these genes were closely related to pathological responses associated with OA, such as inflammation and immunity. Four key dysregulated genes (cartilage acidic protein 1 (CRTAC1), iodothyronine deiodinase 2 (DIO2), angiopoietin-related protein 2 (ANGPTL2), and MAGE family member D1 (MAGED1)) were identified after using machine-learning algorithms. These genes had high diagnostic value in both the training cohort and external validation cohort (receiver operating characteristic > 0.8). The upregulated expression of these hub genes in osteoarthritic cartilage signified higher levels of immune infiltration as well as the expression of metalloproteinases and mineralization markers, suggesting harmful biological alterations and indicating that these hub genes play an important role in the pathogenesis of OA. A competing endogenous RNA network was constructed to reveal the underlying post-transcriptional regulatory mechanisms.

Conclusion

The current study explores and validates a dysregulated key gene set in osteoarthritic cartilage that is capable of accurately diagnosing OA and characterizing the biological alterations in osteoarthritic cartilage; this may become a promising indicator in clinical decision-making. This study indicates that dysregulated key genes play an important role in the development and progression of OA, and may be potential therapeutic targets.

Article focus

- To explore the biological and clinical importance of dysregulated key genes in osteoarthritis (OA) patients at the cartilage level to find potential biomarkers and targets for diagnosis and treatment.

Key messages

- We identified a dysregulated key gene set that is capable of accurately diagnosing OA and characterizing the biological alterations in osteoarthritic cartilage.
- These dysregulated key genes are associated with the activation of related pathways and biological processes such as immunoinflammatory processes, extracellular matrix hypermetabolism, and bone formation and mineralization, which implies a high risk for cartilage damage and aggravation.
- A competing endogenous RNA (ceRNA) regulatory network was constructed to provide a new research direction for a clear understanding of the post-transcriptional regulatory mechanism and targeted therapy.

Strengths and limitations

- Various algorithms were combined in this study to multidimensionally analyze the critical genes in the pathogenesis of OA.
- To reduce the deviation introduced by a single method, both microarray and RNA-sequencing methods were used to validate the dysregulated key gene set in OA.
- The biological roles of the four crucial genes in OA and the ceRNA regulatory network are worth further analysis and verification by more experiments.

Introduction

Osteoarthritis (OA) is a common degenerative joint disease with a high prevalence in the elderly population, and the knee is a common site of involvement. The global prevalence of knee OA is 22.9% in individuals aged 40 years and over. In 2020, worldwide, there were approximately 86.7 million individuals with knee OA aged 20 years and older.¹ The typical symptoms of OA include pain, swelling, and stiffness, often accompanied by dysfunction and limited activity. Without timely intervention and treatment, OA can lead to disability and pose a serious threat to human health. Articular cartilage injury and degeneration are the major pathological manifestations of OA, and are accompanied by a series of complex pathological changes, such as synovitis, subchondral bone sclerosis, and osteophyte formation. However, the exact pathogenesis of OA remains unclear. Reversing or reducing the progressive destruction of cartilage is the primary purpose

of OA treatment. The current clinical therapy for OA is mainly based on anti-inflammatory and pain-relieving drugs. These drugs can reduce synovial inflammation and improve the symptoms of patients. Nevertheless, they cannot reverse the destruction of cartilage, and some may even increase the degradation and reduce the differentiation of cartilage to a certain extent,^{2,3} implying that disease progression is not prevented. At present, there are no effective clinical drugs that can prevent cartilage destruction and progression of this disease.⁴⁻⁷ Therefore, for an in-depth understanding of the pathological characteristics of OA, it is particularly important to systematically explore the expression profile and activation status of related key characteristic genes at the cartilage level. These studies will also be conducive to advances in research on molecular diagnosis and therapeutic targets.

In recent years, microarray and RNA-seq technology have made great progress in OA molecular diagnosis, classification, and potential therapeutic target discovery. Transcriptomics analysis combined with machine-learning methods can reveal potentially critical genes and signalling pathways closely related to the development of OA. These analyses will not only help in understanding the complex pathogenesis of OA, but also providing more valuable options and research directions for the diagnosis of the disease.^{8,9}

Quick and accurate diagnosis of OA is important for improving its prognosis. A recent large study identified CRTAC1 in plasma as a novel promising candidate biomarker that is both associated with the development of OA and can predict the progression of arthroplasty.¹⁰ This study also supports previous findings suggesting that COMP is a likely biomarker for OA,^{11,12} although cartilage oligomeric matrix protein (COMP) is a significantly weaker biomarker than cartilage acidic protein 1 (CRTAC1). In addition, urinary C-terminal cross-linked telopeptides of type II collagen (CTX-II), a degradation product of type II collagen, was also found to be a potential biomarker for OA and correlated with both disease occurrence and progression.^{13,14} CRTAC1 is a glycosylated extracellular matrix (ECM) protein found in the interterritorial matrix of articular deep zone cartilage; COMP is a non-collagenous protein of articular cartilage that plays a role in the structural integrity of cartilage; and CTX-II is a fragment of type II collagen that mainly exists in cartilaginous tissues. These candidate biomarkers are all molecules and/or fragments produced by cartilage tissue and released into joint fluid, blood, or eventually urine, which have the unique properties of dynamic change and high sensitivity that may overcome some limitations of the current methods for assessing OA. Therefore, screening key molecules at the cartilage level may be helpful for the molecular diagnosis of OA, and may be applied to clinical practice decisions.

However, due to ethical issues, normal human cartilage is usually difficult to obtain, resulting in a number of studies with small sample sizes. By contrast, integrating multiple datasets through databases can increase the sample size of studies and provide more accurate results.

In this study, we applied transcriptomic analyses and machine-learning methods to screen the key gene set closely related to the development of OA through microarray and RNA-seq data. In addition, protein-protein interaction (PPI) network analysis, functional annotation analysis, immune infiltration analysis, and the construction of a competing endogenous RNA (ceRNA) network targeting critical genes were combined to provide a reference for exploring the pathogenesis of, finding suitable targets for, understanding the post-transcriptional regulatory mechanism of, and identifying potential molecular therapeutic options for OA.

Methods

Data collection and processing

Six datasets, namely the microarray datasets GSE169077, GSE57218,^{15,16} and GSE174049, and the RNA-seq datasets GSE114007,¹⁷ GSE168505,¹⁸ and GSE143514,¹⁹ were collected from the Gene Expression Omnibus (GEO) database (Supplementary Table i). The GSE174049 dataset was applied for the long noncoding RNA (lncRNA) study, and the AnyProbe R package was used for annotation. Due to the lack of OA cartilage datasets for micro RNA (miRNA) study, synovial tissue from OA was selected for analysis (GSE143514). All other datasets were genome expression data for human knee and hip OA cartilage and healthy cartilage tissue. Because of the different sources of data production, we used microarray data for further analysis to screen for dysregulated key gene sets, and RNA-seq data were applied as an external validation set. Microarray data were quantile normalized and log₂ transformed if necessary. RNA-seq data were normalized to variance stable transformation values via the DESeq2 R package. The limma R package was used to identify differentially expressed genes (DEGs) between disease and control samples in the integrated dataset, and the threshold for DEGs was set at a log₂-fold change (logFC) > 1 and adjusted p-value < 0.05, indicating upregulated expression of DEGs. Genes that were significantly upregulated between the two groups were investigated.

Weighted gene coexpression network analysis

Weighted gene coexpression network analysis (WGCNA) is used to effectively explore the relationship between genes and phenotypes.²⁰ The WGCNA package in R was applied to establish the weighted coexpression network of the expression data for the integrated dataset. The optimum soft threshold of the scale-free network was detected and determined by the PickSoftThreshold function. The matrix data were transformed into an adjacency matrix and clustered, and then modules were found by topological overlap. The module eigengene was calculated, and the related modules in the module eigengene-based tree were merged to generate a clustering dendrogram. The criteria for key module selection were designated as the module with the highest module member correlation and $p < 0.05$. The gene information of the corresponding module was used for further investigation.

Functional annotation analysis

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses are widely used methods for annotating genes and gene products that can be used to understand information about gene function and biological pathways.^{21,22} The ClusterProfiler R package was used for GO and KEGG annotation analysis of the overlapping genes obtained from the intersection of DEGs and key module genes. Terms and pathways with a false discovery rate (FDR) < 0.05 were considered statistically significant. In addition, we performed gene set enrichment analysis (GSEA) to analyze whether a particular gene set had statistically significant differences in biological processes between different subgroups. The cutoff criterion for statistical significance was set as the absolute value of normalized enrichment score > 1, $p < 0.05$, and FDR < 0.25 according to the GSEA user guide.

Characteristic gene selection and model construction

The Boruta algorithm is a wrapper based on the random forest algorithm, and the results of the Boruta algorithm usually require multiple iterations and therefore are more stable than the results produced by feature selection methods based on a single random forest algorithm. The goal of Boruta feature selection is to select all the feature sets related to the dependent variable, rather than selecting the feature set that can minimize the cost function of the model for a specific model. This process helps identify the influencing factors of the dependent variable more comprehensively for better and more efficient feature selection.²³ Support vector machine-recursive feature elimination (SVM-RFE) is a regression or classification supervised machine-learning technique that avoids overfitting by training a subset of features from different classes to shrink down the feature set and filter out the most predictive features.²⁴ LASSO is a regression-based method that allows for a large number of covariates in the model and, importantly, has the unique feature of penalizing the absolute values of regression coefficients.²⁵ We first used the Boruta algorithm to calculate all the characteristic gene sets related to the study variables from the overlapping genes obtained from the intersection of DEGs and key module genes. SVM-RFE and LASSO regression were then performed to select the key characteristic genes. The key characteristic genes obtained were applied as the feature variables to construct the LASSO model. The model index for each sample, defined as a risk score, was used to weight the expression values of the selected genes using regression coefficients from the LASSO analysis. $\text{Index} = \text{Exp_Gene1} * \text{Coef_Gene1} + \text{Exp_Gene2} * \text{Coef_Gene2} + \dots + \text{Exp_Gene} * \text{Coef_Gene}$. "Exp" represents the expression value of the gene, and "Coef", derived from LASSO regression, is the regression coefficient of the gene. Finally, to evaluate the diagnostic ability of each candidate genetic biomarker, we evaluated the area under the receiver operating characteristic (ROC) curve, sensitivity, and specificity. Decision curve analysis (DCA) was also performed, and clinical impact curves were drawn to evaluate the potential clinical utility.

Protein-protein interaction network

The protein-protein interaction (PPI) network system analyzes the interaction between proteins in biological systems, which

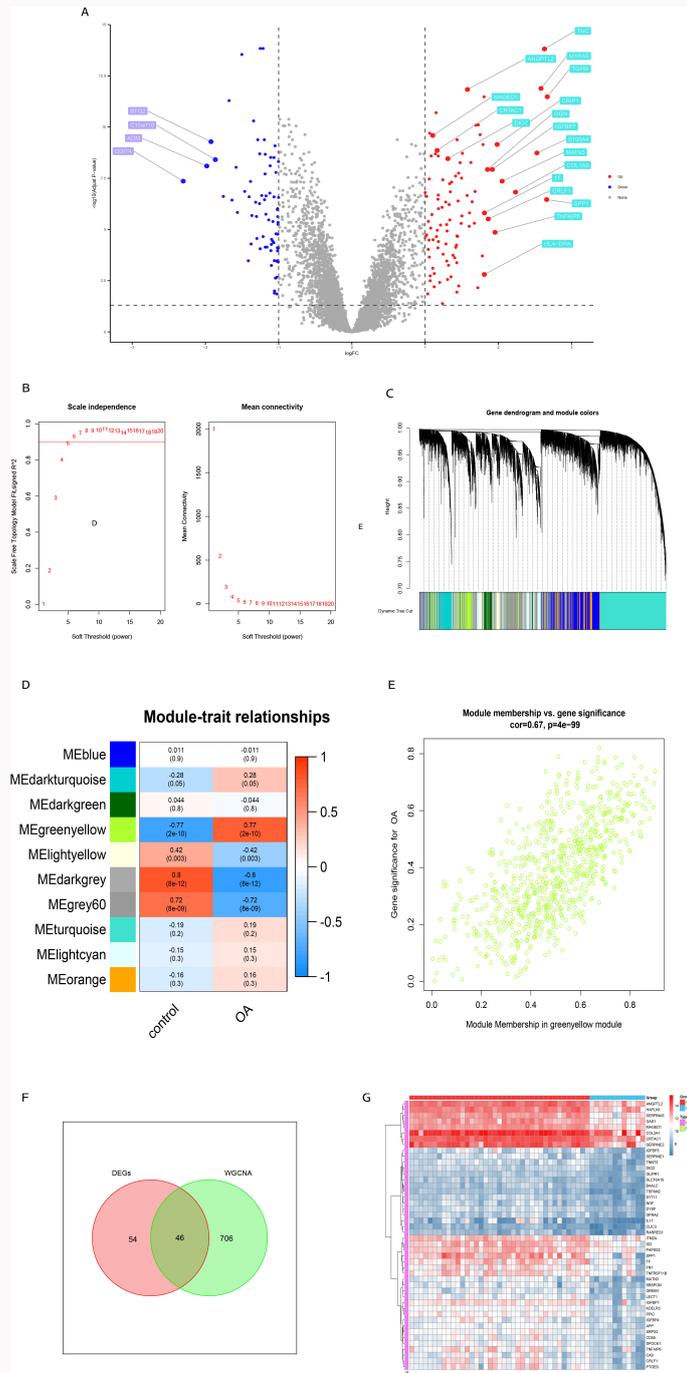


Fig. 1

Identification of overlapping characteristic genes. a) Volcano plot of differentially expressed genes (DEGs) between osteoarthritis (OA) samples and control samples ($\log_{2}FC > 1$ and adjusted p -value < 0.05). b) Soft-threshold power for weighted gene coexpression network analysis (WGCNA). The red line in the left panel indicates scale-free topological fit index = 9. c) Clustering dendrograms of all expressed genes with dissimilarity based on the topological overlap along with the different assigned modules. d) Heatmap of the correlations between modules and clinical traits. Red represents positive correlations, and blue represents negative correlations. e) A scatterplot of gene significance versus module membership in the green-yellow module. f) Venn diagram of the intersection of the overlapping characteristic genes obtained from upregulated DEGs and key module genes. g) Heatmap of the overlapping characteristic genes.

is helpful for understanding the reaction mechanism of biological signals and energy and substance metabolism under special physiological conditions such as diseases. This system can also provide a better understanding of the functional connections between proteins. GeneMANIA is a website for constructing PPI networks that can be used to generate hypotheses about gene function, analyze gene lists, and prioritize genes for functional analysis.²⁶ This website can

be used to discover functionally similar genes based on a large amount of genomic and proteomic data. Therefore, a PPI network was constructed for the identified key genes.

Assessment and correlation of disease immune infiltration cells

Analysis of immune cell infiltration plays an important guiding role in predicting the course of the disease and treatment

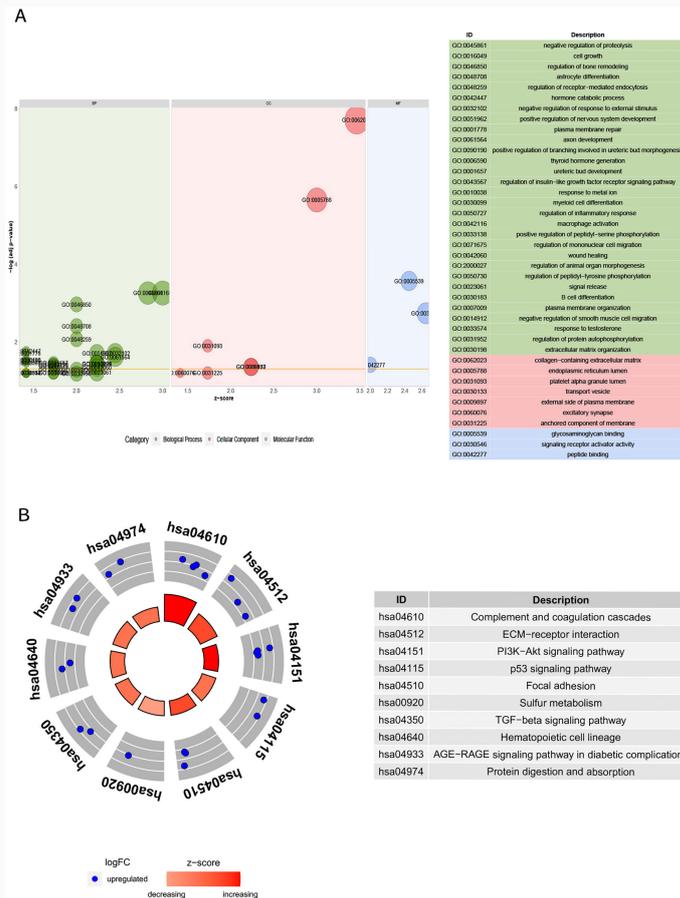


Fig. 2

Functional annotation of the overlapping characteristic genes. a) Gene Ontology functional analysis of the overlapping characteristic genes. b) Kyoto Encyclopedia of Genes and Genomes pathway analysis of the overlapping characteristic genes.

response. First, the estimated immune score was used to evaluate whether there was a difference in the level of immune infiltration between OA and healthy tissues. Then, the single-sample gene set enrichment analysis (ssGSEA) algorithm was used to estimate the infiltration abundance of immune cells in OA and normal cartilage. To further investigate the association between key genes and immune cells, Spearman correlation analysis was performed. A p-value < 0.05 was considered statistically significant.

Construction of a competing endogenous RNA regulatory network

First, differentially expressed lncRNAs and miRNAs were identified by analyzing the GSE174049 and GSE143514 datasets. An absolute value of logFC greater than 1 and a p-value less than 0.05 were set as cutoff criteria. Second, miRNAs targeting key genes were predicted by miR-Walk²⁷ with a threshold of miRWalk Score = 1. Third, the significantly downregulated miRNAs were intersected with the miRNAs predicted via miRWalk to obtain miRNA-mRNA interaction pairs. Fourth, the upstream lncRNAs of these miRNAs were predicted according to LncBase v.3,²⁸ filtering with direct as the validation type and *Homo sapiens* as the species. Fifth, the significantly upregulated lncRNAs were intersected with the lncRNAs predicted by LncBase to obtain lncRNA-miRNA interaction pairs. Finally, the miRNA-mRNA and lncRNA-miRNA interaction pairs obtained from the above

steps were constructed to form a ceRNA network consisting of multiple lncRNA-miRNA-mRNA regulatory axes that were visualized using Cytoscape (v3.9.0).

Statistical analysis

R4.1.0 software (R Foundation for Statistical Computing, Austria) was used for data processing, statistical analysis, and plotting. Independent-samples *t*-test was used for normally distributed variables, and the Mann-Whitney U test was applied for non-normally distributed variables. A p-value < 0.05 was considered statistically significant.

Results

Identification of differentially expressed genes in OA tissue

The overall key steps of this study are illustrated in Supplementary Figure a. The datasets obtained from GEO were standardized, and then the batch effects were corrected to obtain the integrated dataset (Supplementary Figure b). A total of 173 DEGs were identified between OA and healthy samples, including 100 that were upregulated and 73 that were downregulated (Figure 1a).

Weighted gene coexpression network analysis

To identify the key modules associated with OA, all genes were analyzed in the integrated dataset using WGCNA. When the soft threshold was set to 5, the scale-free topological model fitting index (R^2) nearly reached 0.9, and the mean

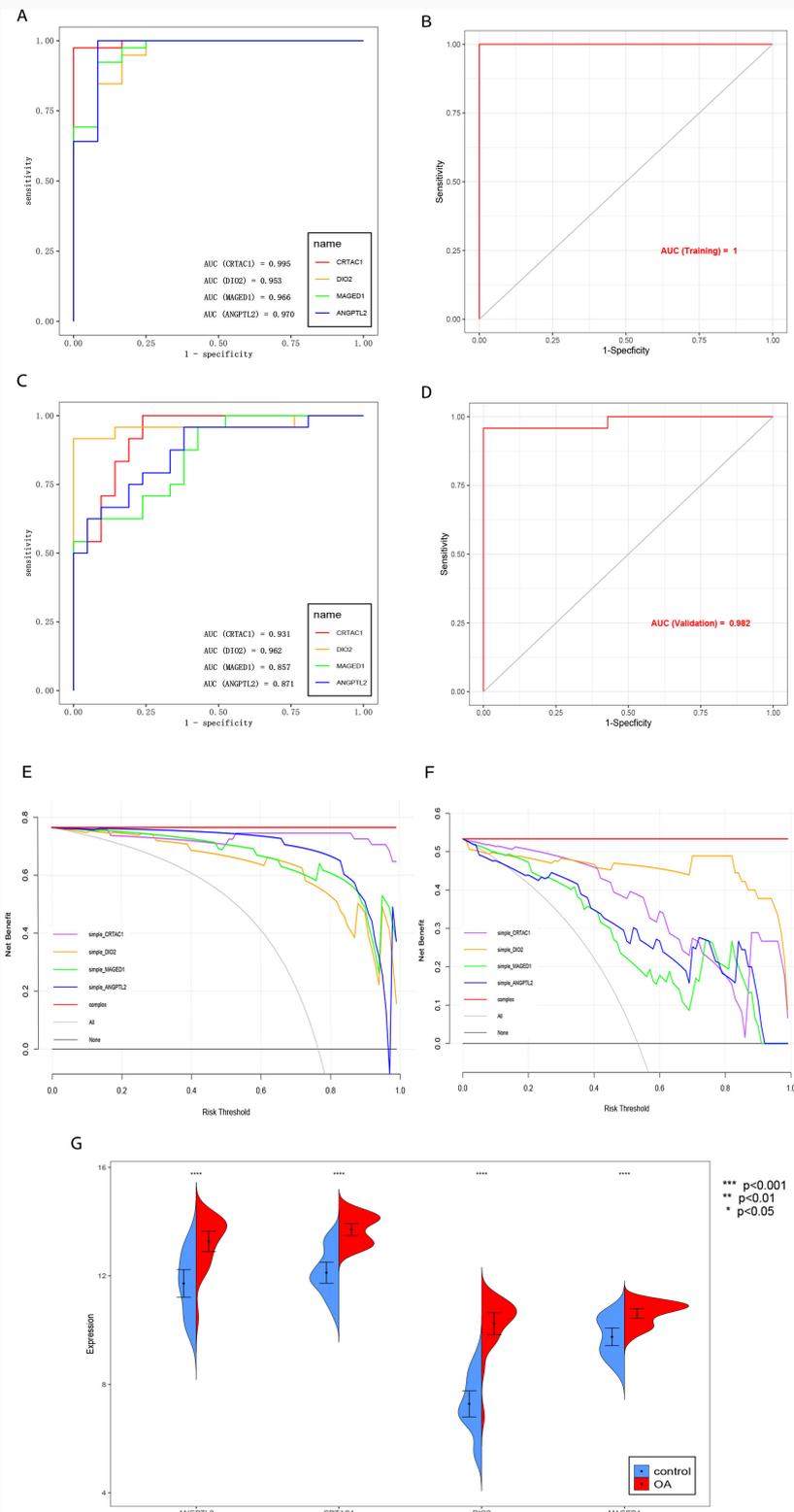


Fig. 4

Diagnostic value of the key characteristic genes. a) Receiver operating characteristic (ROC) analysis of the independent diagnostic efficacy of the four key genes in the training cohort (GSE169077, GSE57218). b) ROC analysis of the overall diagnostic efficacy of the four key genes in the training cohort. c) ROC analysis of the independent diagnostic efficacy of the four key genes in the external validation cohort (GSE89408, GSE143514). d) ROC analysis of the overall diagnostic efficacy of the four key genes in the external validation cohort. e) Decision curve analysis to evaluate the potential clinical value in the training cohort. f) Decision curve analysis to evaluate the potential clinical value in the external validation cohort. g) Differential expression levels of the four key genes between the osteoarthritis (OA) and normal groups in the external validation cohort (Mann-Whitney U test).

of variables was equal to 11, the classification accuracy reached 0.983 for the first time; thus, 11 of 38 variables were selected as optimal genes (Figure 3b). Through LASSO

regression analysis, the lambda value that minimized the error in cross validation was determined, and nine of 38 variables were obtained as candidate genes (Figure 3c). Ultimately, by

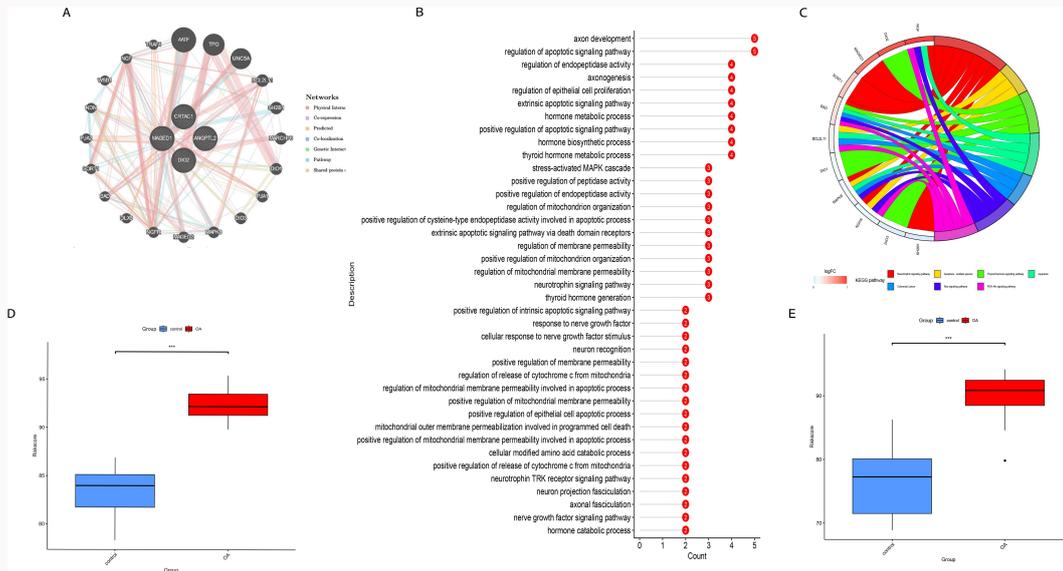


Fig. 5 Potential effect of key genes in osteoarthritis (OA). a) Coexpression network of the characterized genes. b) Gene Ontology functional (GO) analysis of the coexpressed genes. c) Kyoto Encyclopedia of Genes and Genomes pathway analysis (KEGG) of the coexpressed genes. d) Boxplots of the risk score between OA and healthy controls in the training cohort (Mann-Whitney U test).

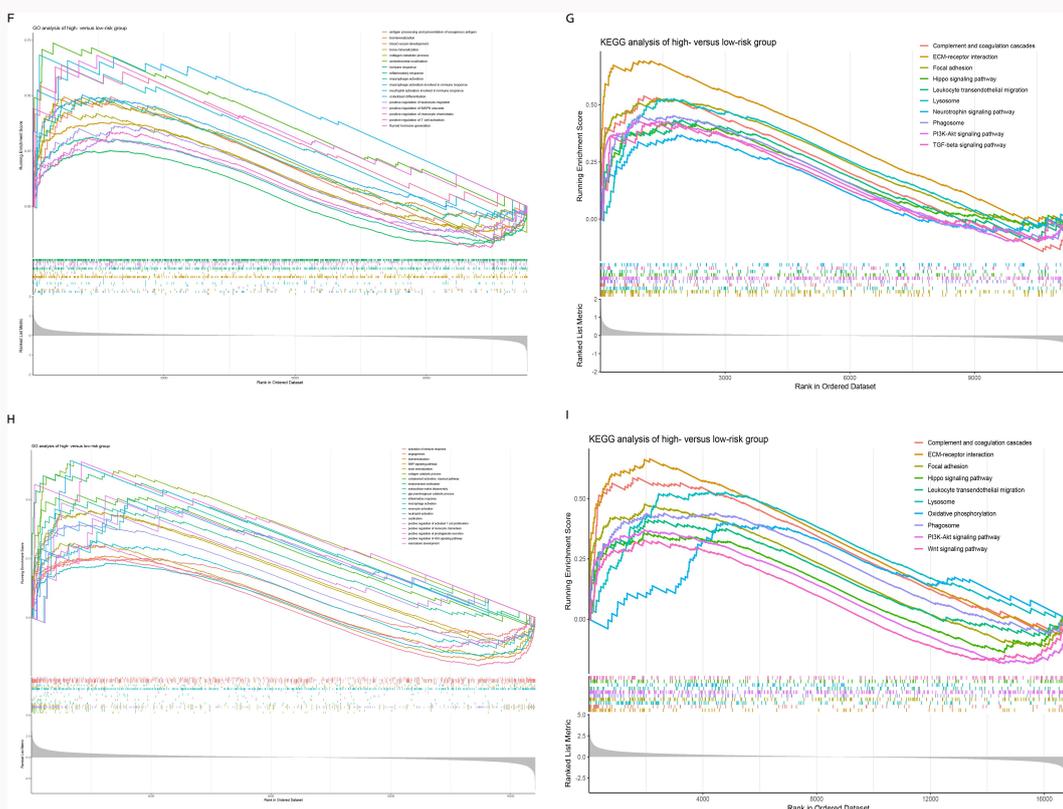


Fig. 6 e) Boxplots of the risk score between osteoarthritis (OA) and healthy controls in the external validation cohort (Mann-Whitney U test). f) to i) Gene set enrichment analysis conducted between high- and low-risk groups. Specific f) Gene Ontology (GO) biological processes and g) Kyoto encyclopedia of Genes and Genomes (KEGG) pathways in the high-risk group of the training cohort (GSE169077, GSE57218). Specific h) GO biological processes and i) KEGG pathways in the high-risk group of the external validation cohort (GSE89408, GSE143514).

intersecting the optimal genes extracted via SVM-RFE and candidate genes selected by LASSO regression, we identified a total of four key characteristic genes (CRTAC1, iodothyronine deiodinase 2 (DIO2), angiotensin-related protein 2

(ANGPTL2), and MAGE family member D1 (MAGED1)) (Figure 3d).

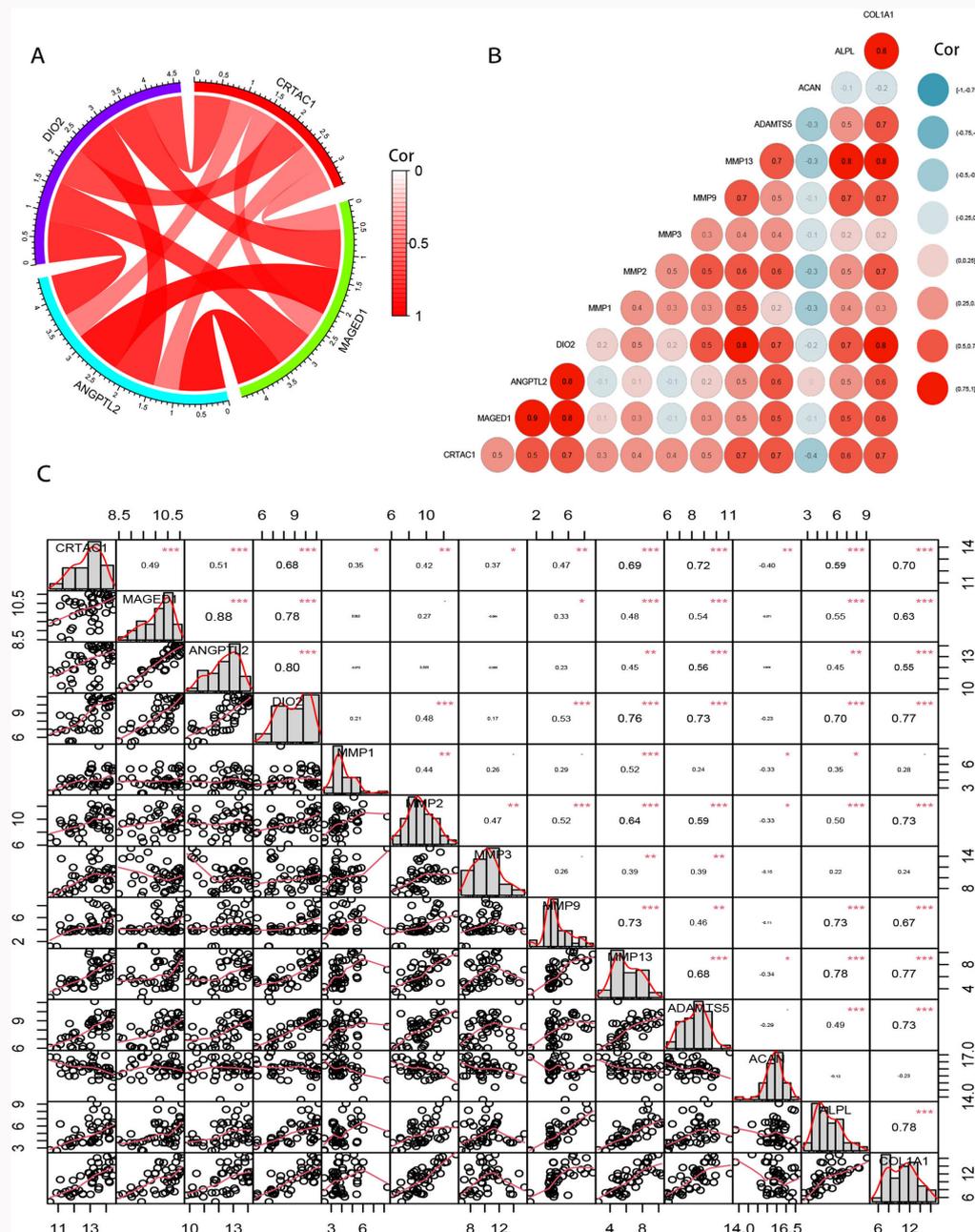


Fig. 7

Correlations of key genes with pathological changes in osteoarthritis (OA). a) Correlation analysis among the four key genes; darker red indicates stronger positive correlations. b) The correlations of four key genes with matrix metalloproteinases (MMPs), a disintegrin and metalloproteinase with thrombospondin motifs 5 (ADAMTS-5), ACAN, alkaline phosphatase (ALPL), and collagen type I alpha 1 (COL1A1); darker red indicates stronger positive correlations and darker blue indicates stronger negative correlations. c) The specific value of the correlation and statistical significance of the four key genes with MMPs, ADAMTS-5, ACAN, ALPL, and COL1A1 (Spearman correlation analysis; * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$).

Diagnostic value of the key gene set

ROC curve analysis was investigated to validate the diagnostic value (GSE169077, GSE57218), and the results indicated the high diagnostic efficacy of all four hub genes (Figures 4a and 4b). Furthermore, RNA-seq datasets (GSE89408, GSE143514) were introduced as an external validation cohort to verify the diagnostic power of the hub genes, and the results also demonstrated a high accuracy for diagnosing the occurrence of OA (Figures 4c and 4d). The results also showed that the ROC curve of the integrated four hub candidate genes was higher than that of a candidate gene alone. According to DCA, the composite genetic model curve and four single

gene model curves were all above the grey line, and the composite genetic model was better than the four single gene models. These results implied that decision-making based on the composite genetic model provides greater clinical benefit and may be more beneficial for patients with OA (Figure 4e). Assessment of the validation cohort also corroborated these findings (Figure 4f). In addition, consistent with the expression trend in the training cohort (GSE169077, GSE57218), the expression levels of four key genes were significantly higher in the OA group than in the normal group in the validation cohort (GSE89408, GSE143514) (Figure 4g), indicating the

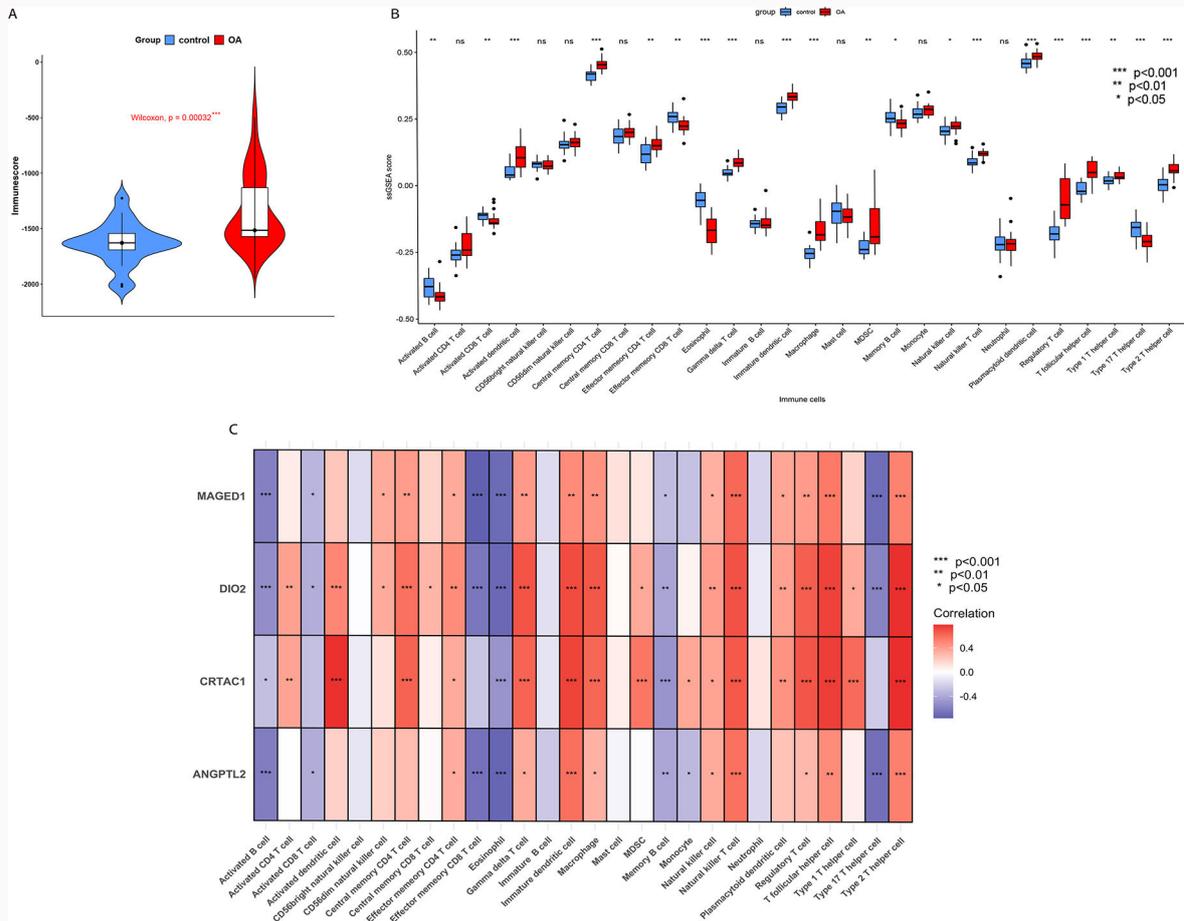


Fig. 8

The landscape of immune infiltration in association with osteoarthritis (OA). a) Violin plot exhibiting the estimated immune score between OA and healthy controls (Mann–Whitney U test). b) Boxplots of the different immune cell infiltration profiles of OA and healthy controls (Mann–Whitney U test). c) The relationship between the four key genes and immune cell abundance (Spearman correlation analysis; *p < 0.05, **p < 0.01, and ***p < 0.001).

good stability and accuracy of the four crucial genes that could play an important role in disease.

Potential effect of key genes in OA

To further explore the underlying molecular mechanisms for the dysregulated key genes, we created a PPI network for the four crucial genes through the GeneMANIA database and analyzed the roles of these functionally similar genes (Figure 5). KEGG/GO analysis indicated that these genes were enriched in the neurotrophin signalling pathway, thyroid hormone signalling pathway, apoptosis, PI3K-Akt signalling pathway, thyroid hormone generation, positive regulation of apoptotic signalling pathway, stress-activated MAPK cascade, positive regulation of mitochondrion organization, and so on. Additionally, based on the median expression value of the four key genes, we introduced GSEA to classify all samples into two categories (high- and low-expression groups). The findings revealed that the four highly expressed gene subgroups activated the PI3K-Akt signalling pathway, ECM-receptor interaction, oxidative phosphorylation, complement and coagulation cascades, lysosome, phagosome, apelin signalling pathway, collagen metabolic process, biomineralization, ossification, and thyroid hormone generation, implying that the roles among these four genes were relatively similar in

OA cartilage (Supplementary Figures c to j). In addition, a risk score was calculated for each patient based on the expression of four hub characteristic genes using the following formula: risk score = CRTAC1*3.870+ DIO2*1.787+ MAGED1*1.232+ ANGPTL2*0.410. The results indicated that risk score was significantly higher in OA patients than healthy controls in both the training cohort and validation cohort (Figure 6). All samples were then divided into high- and low-risk subgroups according to the median risk score, and GSEA was performed on both the training cohort and validation cohort to further explore the potential biological mechanism of dysregulated key genes in OA. Pathways and functions related to ECM metabolism, immune inflammatory activation, and bone formation and mineralization, such as ECM-receptor interaction, PI3K-Akt signalling pathway, complement and coagulation cascades, leucocyte transendothelial migration, Hippo signalling pathway, collagen catabolic process, macrophage activation, neutrophil activation, immune response, inflammatory response, biomineralization, bone mineralization, endochondral ossification, and osteoblast differentiation, were activated in the high-risk group. These results indicated that high expression of hub genes and high risk scores were closely related to the main pathological changes mediating cartilage damage and degeneration in OA. Therefore, further

analyses of the correlation between key genes and immune cells as well as ECM metabolism and mineralization-relevant markers are warranted. MMP1, MMP2, MMP3, MMP9, MMP13, and ADAMTS-5 are vital enzymes related to ECM metabolism in OA,²⁹⁻³³ and aggrecan (ACAN) is the main component of the ECM in cartilage tissue. Through the external validation cohort, we further demonstrated significant positive correlations among the key gene set (Figure 7a). In addition, correlation analysis suggested that there were significant positive correlations between four crucial genes and metalloproteinases, and except for ANGPTL2, all other crucial genes were negatively correlated with ACAN (Figure 7c). This finding indicated that these four crucial genes with significant functional similarity had the capacity to reflect the level of ECM catabolism to a certain extent. In addition, our analysis also indicated that four hub genes were significantly and positively correlated with markers of mineralization (ALPL, COL1A1)³⁴ (Figure 7c), which suggested the ability of four crucial genes to reflect mineralization of articular cartilage during the progression of OA. To further explore whether OA is associated with immune infiltration, we applied the estimated immune score to evaluate the difference in immune infiltration levels between OA and healthy tissues. OA presented elevated immune scores, suggesting that OA might possess a more dominant level of immune infiltration (Figure 8a). The ssGSEA algorithm was used to further evaluate the infiltration abundance of immune cell subpopulations between OA cartilage and healthy controls. The findings revealed that OA cartilage had significantly higher levels of infiltration abundance of most immune cells, which suggested activation of the immune microenvironment (Figure 8b). In addition, correlation analysis revealed positive correlations between infiltration of multiple innate and adaptive immune cell types, such as macrophages, γ δ T cells, immature dendritic cells, natural killer (NK) cells, natural killer T cells, regulatory T cells, T follicular helper cells, type 2 T helper cells and effector memory CD4 T cells, and the expression of all four key genes (Figure 8c), implying that the four key genes were capable of reflecting the increased level of immune cell infiltration during the progression of OA.

According to the above analysis, the dysregulation of these key genes is not only associated with the activation of related pathways and biological processes such as immunoinflammatory processes, ECM hypermetabolism, and bone formation and mineralization, but also significantly and positively correlated with multiple metalloproteinases, inflamed immune cells, and markers of mineralization, which implies a high risk for cartilage damage and aggravation. Therefore, these crucial genes may be potential molecular diagnostic biomarkers and therapeutic targets, and more attention should be given.

Construction of a competing endogenous RNA regulatory network for the four hub genes

To understand the potential post-transcriptional regulatory mechanisms of the four key genes in detail and to provide a reference for the selection of potential noncoding RNA therapeutic drugs for the four key genes, a ceRNA regulatory network was constructed. A total of 23 miRNA-mRNA pairs were obtained after intersection with the miRNAs predicted by miRWalk targeting the four key genes and

downregulated differentially expressed miRNAs (Supplementary Figure ka). Overall, 68 lncRNA-miRNA pairs were obtained after intersection with the lncRNAs predicted by LncBase and upregulated differentially expressed lncRNAs (Supplementary Figure kb). Subsequently, miRNA mRNA and lncRNA miRNA binding pairs were integrated to construct a complete ceRNA regulatory network (Supplementary Figure kc, Supplementary Table iii).

Discussion

In the present study, we integrated machine learning and multiple algorithms to analyze the pathological alterations associated with OA cartilage from multiple dimensions, including gene expression, biological pathways, and immune cell infiltration. The high expression of four crucial characteristic genes in OA cartilage was ultimately identified. These dysregulated genes showed not only accurate diagnostic performance but also the potential to characterize the biological characteristics of OA cartilage. A ceRNA network was also constructed targeting the four key genes, contributing to further understanding post-transcriptional regulatory mechanisms and the selection of potential molecular therapeutic agents. Previous studies also analyzed arthritis by integrating public datasets.³⁵⁻⁴⁰ However, a number of studies analyzed OA based on synovial tissues, and several studies, in order to increase the sample size, mixed OA and rheumatoid arthritis or cartilage, meniscus, synovial, and subchondral bone tissues for analysis, which may negatively affect the reliability of the final results. In this study, both microarray and RNA-seq data from OA and normal human cartilage were comprehensively investigated, which increased its accuracy.

In this analysis, a total of 100 upregulated DEGs were identified, and after intersecting the critical module genes obtained from WGCNA, a total of 46 characteristic genes were identified. Functional annotation analysis indicated that pathological responses, such as immune inflammation, were caused by the dysregulation of these characteristic genes in OA cartilage. It is thus necessary to screen critical genes for further investigation. Through multiple iterations of Boruta's algorithm, as well as SVM-SEF and LASSO regression algorithms, CRTAC1, ANGPTL2, DIO2, and MAGED1 were finally identified as key genes in OA cartilage. Evaluation of the external cohort validated that these four key genes were significantly highly expressed in OA cartilage. In addition, through diagnostic value analysis in both the training and validation cohorts, we found that the four key genes had high accuracy and stability for the diagnosis of OA. Notably, the integration of the four crucial genes exhibited higher diagnostic performance than the four candidate genes alone and provided greater clinical benefit, consistent with the multimolecular driving nature of OA. This implies that studies based on the expression and translation levels of these four hub genes in synovial fluid, blood, and urine possess clinical importance, and that changes in these candidate biomarkers may play a vital role in the detection of OA.⁴¹

To further explore the potential molecular mechanisms of the four key genes in OA, we constructed a PPI network based on these crucial genes through the

GeneMANIA database. Annotation of the gene functions through KEGG and GO analyses revealed that the neurotrophin signalling pathway, the thyroid hormone signalling pathway, the PI3K-Akt signalling pathway, thyroid hormone generation, the positive regulation of apoptotic signalling pathway, and the stress-activated mitogen-activated protein kinase (MAPK) cascade were the most important functional categories. In addition, we conducted GSEA and detected that the high-expression subgroup of the four hub genes, and the high-risk subgroup, activated relevant functions and pathways such as ECM metabolism, immune inflammation, and bone formation and mineralization. The activation of these functions and pathways strongly implies the progression of OA. As the marker enzyme of cartilage metabolism, an increase in MMP and ADAMTS-5 expression usually indicates an increase in cartilage catabolism. Decreased expression of ACAN, a major component of the ECM, implies decreased cartilage synthesis. Correlation analysis was conducted for the key characteristic genes and indicated that these four key genes with similar functions were significantly positively correlated with MMPs and ADAMTS-5. CRTAC1, DIO2, and MAGED1 showed a negative correlation with ACAN, among which CRTAC1 showed a significant difference. Furthermore, these four crucial genes were also significantly positively correlated with markers of mineralization. In patients with OA, osteochondral mineralization leads to increased thickening in adjacent subchondral bone as well as loss of cartilage structural and functional integrity.⁴²⁻⁴⁴ The size of cartilage lesions and osteophytes correlated with the severity of joint pain and the progression of OA.⁴⁵ In addition, the estimated immune scores and immune cell infiltration were also significantly increased in OA tissues compared with healthy control tissues. The four key genes also showed significant positive correlations with a variety of immune cells, such as macrophages, regulatory T cells, effector memory CD4 T cells, T follicular helper cells, and NK cells. In OA, immune responses are primarily activated by endogenous stimuli released from tissue damage or stressed cells, leading to sterile inflammation. Under normal circumstances, this immune response can be well controlled, thereby facilitating the repair of the injury. If left unchecked, however, chronic immune responses can lead to pathological manifestations, such as cartilage degeneration and excessive tissue repair (e.g. heterotopic bone formation).^{46,47} It has been suggested that activated macrophages are present in the vast majority of OA cases, and their number is significantly correlated with the severity of OA pain, the formation of osteophytes, and the narrowing of the joint space.⁴⁸ Proinflammatory cytokines produced by macrophages include IL-1 β and TNF- α , which can induce cartilage damage.⁴⁹ The pathological role of T cells in OA remains uncertain, but they may be involved in the production of proinflammatory cytokines and may be related to the recognition of the breakdown products of cartilage matrix proteoglycans.^{50,51} One study found that T follicular helper cells and interleukin (IL)-21 play an important role in the progression of OA, and their expression in OA patients was significantly correlated with the level of inflammation, knee function score, and OA disease activity.⁵² Jaime et al⁵³ indicated that a subset of NK cells (CD56(+)^{bright}CD16(-) cells) is associated with increased levels of proinflammatory cytokines in the synovial fluid of OA patients, which may contribute to the progression

of chronic inflammation. In conclusion, the role of these four key genes in OA deserves further analysis.

CRTAC1, also known as cartilage expressing protein 68 (cep68) or lateral olfactory tract usher substance (LOTUS), was first identified as a marker that could be used to distinguish chondrocytes, osteoblasts, and mesenchymal stem cells.⁵⁴ Recent studies have shown that CRTAC1 in plasma is associated with OA and is a novel promising candidate biomarker for OA.^{10,55,56} A new study found that adding two other proteins, ACAN and NCAN, to CRTAC1 could improve the prediction of OA over CRTAC1 alone.⁵⁶ ACAN is the major proteoglycan in articular cartilage and has important effects on cartilage strength and biology. NCAN, similar to ACAN, is a chondroitin sulfate proteoglycan, but its specific function is uncertain. Therefore, it is of great significance to screen OA-related biomarkers at the cartilage level. In addition, OA is a multimolecule-driven disease, and functionally interconnected genes can play a role together at a certain point in cellular life activity. Therefore, identifying a set of gene signatures may be more conducive to the diagnosis of OA. Our analysis also showed that the diagnostic performance (including ROC curve analysis and DCA) of integrating four key genes (gene set) was better than that of a key gene alone. Apart from the potential value of CRTAC1 in the diagnosis of OA, one study also found that CRTAC1 is an important regulator of OA pathogenesis. Genetic deletion of CRTAC1 in female mice can protect against cartilage degradation, osteophyte formation, and gait abnormalities (reduction of pain response) in a model of post-traumatic OA.⁵⁷ However, at present, there are few studies of the effect of CRTAC1 on the pathogenesis of OA. How CRTAC1 specifically affects OA progression remains unclear. Ge et al⁵⁷ speculated that CRTAC1 presumably mediates the effects of cytokines during OA, including promoting inflammation, increasing catabolism, and inhibiting the anabolic activity of chondrocytes, which is consistent with the results of our correlation analysis, which also found that CRTAC1 was positively correlated with multiple inflammatory immune cells and metalloproteinases and negatively correlated with ACAN. It has been suggested that the NOGO-A/NGR1 pathway, the downstream target of CRTAC1, is essential for osteoclastogenesis.⁵⁸ Therefore, increased levels of CRTAC1 in OA may antagonize NOGO/NGR1 signalling to inhibit osteoclast formation as well as promote subchondral bone sclerosis and cartilage mineralization, which may exacerbate OA progression.⁵⁹⁻⁶² We also found that CRTAC1 was positively correlated with mineralization markers. GSEA in our study indicated that the high-expression subgroup of CRTAC1 could activate relevant functions and pathways such as ECM metabolism, immune inflammation, and bone formation and mineralization. Moreover, CRTAC1 is a glycosylated ECM protein in articular cartilage, and upregulation of this protein in articular cartilage of OA may perturb cartilage homeostasis and matrix turnover, thereby altering the biophysical properties and physical remodelling of the cartilage ECM. Changes in the cartilage ECM may intensify the progression of OA.⁶³⁻⁶⁶ Therefore, synthesizing the above analysis, CRTAC1 probably serves as a potential key therapeutic target for OA, but further research is warranted.

The proteins encoded by DIO2 belong to the iodothyronine deiodinase family. DIO2 is an OA susceptibility gene that encodes a deiodinase type 2 protein, which catalyzes

the conversion of thyroxinogen (T4) to biologically active thyroid hormone (T3) via exocyclic 5'-deiodination.⁶⁷ Active T3 subsequently signals chondrocyte terminal maturation, leading to cell hypertrophy, degradation and mineralization of the cartilage matrix, and bone formation.⁶⁸⁻⁷⁰ This is consistent with the results of our correlation analysis, which indicated that DIO2 showed a positive correlation with a variety of metalloproteinases and mineralization markers. In addition, DIO2 was associated with multiple inflammatory immune cells in our analysis. Studies have indicated that upregulation of DIO2 expression may be responsible for OA pathogenesis by enhancing inflammatory responses and chondrocyte hypertrophy, and the inhibition of deiodinase during in vitro cartilage formation contributes to prolonged endochondral homeostasis, as reflected by significantly attenuated upregulation of matrix-degrading enzymes and increased ECM deposition.^{34,71} Bomer et al found that DIO2(-/-) mice could prevent cartilage damage and significantly reduce synovial inflammatory symptoms under excessive mechanical stress.⁷² A recent study detected no signs of OA in DIO2^{Ala92} mutant mice, suggesting a protective effect of the Ala92 polymorphism.⁷³ The Ala92 variant impaired conversion of the prohormone T4 to the active hormone T3, consequently decreasing local thyroid hormone signalling.⁷⁴ providing further evidence that reducing thyroid hormone signalling can prevent OA.⁷³ Our study also indicated that the high expression of the DIO2 subgroup activated the thyroid hormone signalling pathway, and the high-risk subgroup could upregulate thyroid hormone generation. Together, these analyses demonstrate that DIO2 activity can be a therapeutic target for OA.

ANGPTL2 protein is a secreted glycoprotein homologous to angiopoietin and may act on endothelial cells through autocrine or paracrine pathways. Hyperfunction of ANGPTL2 can cause chronic inflammation, which contributes to the progression of a variety of diseases.^{75,76} ANGPTL2 was reported to promote the nuclear translocation of NF- κ B and induce IL-6 secretion and expression in synovial tissue.⁷⁷ Okada et al⁷⁸ suggested that adipose tissue-derived ANGPTL2 activates the inflammatory cascade of endothelial cells and induces chemotaxis of monocytes/macrophages, leading to the initiation and proliferation of inflammation. Angptl2 mRNA and protein are abundantly expressed in the proliferating rheumatoid synovium of RA patients, especially in fibroblast-like and macrophage-like synoviocytes. Integrin α 5 β 1, a receptor for ANGPTL2, is induced under stress conditions and is involved in the development and progression of inflammation-based pathology.^{79,80} Leucocyte immunoglobulin-like receptor subfamily B member 2 (LILRB2) is also an ANGPTL2 receptor, which is highly expressed on the surface of immune cells, including macrophages, monocytes, and dendritic cells.⁸¹⁻⁸³ An in vitro study found that ANGPTL2 can induce the expression of inflammatory factors in synovial cells through LILRB2.⁸⁴ Another in vitro study showed that ANGPTL2 upregulation accelerated human chondrocyte damage via integrin α 5 β 1 activation of the NF- κ B and p38/MAPK signalling pathways.⁸⁵ These studies also support our findings that ANGPTL2 was highly positively correlated with a variety of immune inflammatory cells and metalloproteinases, and the highly expressed ANGPTL2 subgroup activated the positive regulation of the apoptotic signalling pathway,

collagen metabolic process, and oxidative phosphorylation. In conclusion, ANGPTL2 may be a potential therapeutic target in OA. However, it is worth noting that most of the current studies on ANGPTL2 in arthritis are in vitro studies, and more in vivo studies are necessary for verification.

MAGED1, also known as NRAGE or DLXIN-1, is an X-linked member of the MAGE gene family. MAGED1 is a multifunctional adaptor protein that participates in transcriptional regulation, cell cycle regulation, apoptotic signalling, and cell differentiation.⁸⁶⁻⁸⁸ Matluk et al⁸⁹ showed that NRAGE can activate the nuclear factor kappa-light-chain-enhancer of activated B cells (NF- κ B) signalling pathway through the non-canonical BMP pathway, thus promoting the inflammatory response. Liu et al⁹⁰ demonstrated that MAGED1 can regulate osteogenic differentiation and mineralization, and that MAGED1 deficiency in mice promoted osteopenia and reduced bone mineral density. Therefore, highly expressed MAGED1 may promote cartilage mineralization and osteophyte formation, contributing to the progression of OA. In addition, our study indicated that MAGED1 was not only significantly positively correlated with the other three key genes but also with MMP9, MMP13, and ADAMTS5 as well as multiple immune-inflammatory cell and mineralization markers. Thus, we speculate that MAGED1 might be a potential new therapeutic target for OA. However, there is a lack of research on the pathological mechanism of MAGED1 in OA, and the potential pathogenesis of MAGED1 in OA remains to be elucidated.

Based on the above analysis, the four hub genes not only have good diagnostic ability for OA but also may be potential therapeutic targets for OA. With advances in drug delivery technology, molecular targeted therapy appears to be a potential therapeutic approach for early intervention of OA,^{91,92} where noncoding RNAs (especially lncRNAs and miRNAs) and their regulatory interactors may play an important role in both joint health and disease, affect biological processes and be considered key gene expression modulators.^{93,94} MiR-486 promotes more catabolic phenotypes of chondrocyte-like cells by targeting SIRT6.⁹⁵ MiR-210-3 p can inhibit subchondral angiogenesis by targeting TGFBR1 and ID4 to prevent OA.⁹⁶ MiR-92a-3p regulates chondrocyte-specific gene expression by directly targeting histone deacetylase 2 during chondrogenesis and degradation.⁹⁷ Linc34 can promote abnormal metabolic dysfunction and inflammation in OA chondrocytes by targeting miR-140-5p.⁹⁸ Stelcer et al⁹⁹ identified key miRNAs that could regulate early chondrogenesis, such as hsa-miR-520c-3p and hsa-miR-525-5p. However, there is still a lack of regulatory molecules targeting the four key genes, so we constructed a potential ceRNA regulatory network based on these four genes. Considering the therapeutic potential of noncoding RNAs in preclinical studies of OA, the construction of a ceRNA regulatory network can provide a new reference for a clear understanding of the molecular mechanism and research on targeted therapy.

The fact that we did not obtain a sufficient number of human normal cartilage samples for experiments due to ethical requirements is a limitation of this study. However, our study combined multiple datasets using both microarray and RNA-seq methods to screen and analyze critical genes in OA cartilage to maximize sample size, and reduce the deviation introduced by a single method and exhibited

accurate results. However, these results still need to be verified with large-sample validation. Fortunately, two recent large sample size studies both found plasma CRTAC1 as a specific candidate biomarker for OA and a predictor of OA risk and joint arthroplasty progression,^{10,56} which reflects the reliability and further clinical research value of our screened key genes to a certain degree. However, the efficacy of plasma CRTAC1 in other ethnic populations and the optimal cutoff value need to be investigated further before clinical translation. Additionally, due to the limitation of the OA cartilage datasets, the dataset for studying miRNA was derived from knee synovial tissue. Although cartilage tissue is proximal to synovial tissue in the knee joint, both are covered by synovial fluid, and the expression trends of related molecules in OA and normal tissues may be similar, the accuracy of the final ceRNA results needs to be validated. Finally, we identified four key genes that may be potential therapeutic targets for OA, and this finding was supported by a number of studies. Nevertheless, the biological roles of the four crucial genes in OA and the ceRNA regulatory network are worth further analysis and verification by more experiments.

In conclusion, the current study explores and validates a dysregulated key gene set in osteoarthritic cartilage that is capable of accurately diagnosing OA and characterizing biological alterations in osteoarthritic cartilage; this may become a promising indicator to assist with clinical decision-making. In addition, this study indicates that dysregulated key genes play an important role in the development and progression of OA and are potential therapeutic targets.

Supplementary material

Figures showing the primary flow chart in the study, principal component analysis between datasets before and after debatching, and gene set enrichment analysis conducted for four key genes. Tables detailing the datasets of gene expression profiles, the final decision list of Boruta algorithm, and the list of competing endogenous RNA regulatory networks.

References

- Cui A, Li H, Wang D, Zhong J, Chen Y, Lu H. Global, regional prevalence, incidence and risk factors of knee osteoarthritis in population-based studies. *E Clinical Medicine*. 2020;29–30:100587.
- Abrams GD, Chang W, Dragoo JL. In vitro chondrotoxicity of nonsteroidal anti-inflammatory drugs and opioid medications. *Am J Sports Med*. 2017;45(14):3345–3350.
- Ding C, Cicuttini F, Jones G. Do NSAIDs affect longitudinal changes in knee cartilage volume and knee cartilage defects in older adults? *Am J Med*. 2009;122(9):836–842.
- Reynard LN, Barter MJ. Osteoarthritis year in review 2019: genetics, genomics and epigenetics. *Osteoarthritis Cartilage*. 2020;28(3):275–284.
- Latourte A, Kloppenburg M, Richette P. Emerging pharmaceutical therapies for osteoarthritis. *Nat Rev Rheumatol*. 2020;16(12):673–688.
- Rannou F, Pelletier J-P, Martel-Pelletier J. Efficacy and safety of topical NSAIDs in the management of osteoarthritis: Evidence from real-life setting trials and surveys. *Semin Arthritis Rheum*. 2016;45(4 Suppl): S18–21.
- Savvidou O, Milonaki M, Goumenos S, Flevas D, Papagelopoulos P, Moutsatsou P. Glucocorticoid signaling and osteoarthritis. *Mol Cell Endocrinol*. 2019;480:153–166.
- Orlov YL, Anashkina AA, Klimontov VV, Baranova AV. Medical genetics, genomics and bioinformatics aid in understanding molecular mechanisms of human diseases. *Int J Mol Sci*. 2021;22(18):9962.
- Farrow L, Zhong M, Ashcroft GP, Anderson L, Meek RMD. Interpretation and reporting of predictive or diagnostic machine-learning research in Trauma & Orthopaedics. *Bone Joint J*. 2021;103-B(12):1754–1758.
- Styrkarsdottir U, Lund SH, Saevarsdottir S, et al. The CRTAC1 protein in plasma is associated with osteoarthritis and predicts progression to joint replacement: A large-scale proteomics scan in Iceland. *Arthritis Rheumatol*. 2021;73(11):2025–2034.
- Hoch JM, Mattacola CG, Medina McKeon JM, Howard JS, Lattermann C. Serum cartilage oligomeric matrix protein (sCOMP) is elevated in patients with knee osteoarthritis: a systematic review and meta-analysis. *Osteoarthritis Cartilage*. 2011;19(12):1396–1404.
- Valdes AM, Meulenbelt I, Chassaing E, et al. Large scale meta-analysis of urinary C-terminal telopeptide, serum cartilage oligomeric protein and matrix metalloproteinase-3 as biomarkers for knee and hip osteoarthritis prevalence, incidence and progression of osteoarthritis. *Osteoarthritis Cartilage*. 2014;22(5):683–689.
- van Spil WE, DeGroot J, Lems WF, Oostveen JCM, Lafeber F. Serum and urinary biochemical markers for knee and hip-osteoarthritis: a systematic review applying the consensus BIPED criteria. *Osteoarthritis Cartilage*. 2010;18(5):605–612.
- Hao HQ, Zhang JF, He QQ, Wang Z. Cartilage oligomeric matrix protein, C-terminal cross-linking telopeptide of type II collagen, and matrix metalloproteinase-3 as biomarkers for knee and hip osteoarthritis (OA) diagnosis: a systematic review and meta-analysis. *Osteoarthritis Cartilage*. 2019;27(5):726–736.
- Ramos YFM, den Hollander W, Bovée JVMG, et al. Genes involved in the osteoarthritis process identified through genome wide expression analysis in articular cartilage; the RAAK study. *PLoS One*. 2014; 9(7):e103056.
- den Hollander W, Ramos YFM, Bos SD, et al. Knee and hip articular cartilage have distinct epigenomic landscapes: implications for future cartilage regeneration approaches. *Ann Rheum Dis*. 2014;73(12):2208–2212.
- Fisch KM, Gamini R, Alvarez-Garcia O, et al. Identification of transcription factors responsible for dysregulated networks in human osteoarthritis cartilage by global gene expression analysis. *Osteoarthritis Cartilage*. 2018;26(11):1531–1538.
- Fu W, Hettinghouse A, Chen Y, et al. 14-3-3 epsilon is an intracellular component of TNFR2 receptor complex and its activation protects against osteoarthritis. *Ann Rheum Dis*. 2021;80(12):1615–1627.
- Zhao Y, Lv J, Zhang H, Xie J, Dai H, Zhang X. Gene expression profiles analyzed using integrating RNA sequencing, and microarray reveals increased inflammatory response, proliferation, and osteoclastogenesis in pigmented villonodular synovitis. *Front Immunol*. 2021;12:665442.
- Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9:559.
- Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics*. 2010;26(12):1572–1573.
- Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000;28(1):27–30.
- Kursa MB, Rudnicki WR. Feature selection with the boruta package. *J Stat Softw*. 2010;36(11):1–13.
- Huang ML, Hung YH, Lee WM, Li RK, Jiang BR. SVM-RFE based feature selection and Taguchi parameters optimization for multiclass SVM classifier. *ScientificWorldJournal*. 2014;2014:795624.
- Tibshirani R. Regression shrinkage and selection via the lasso: A retrospective. *Journal of the Royal Statistical Society Series B*. 2011;73(3): 273–282.
- Warde-Farley D, Donaldson SL, Comes O, et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res*. 2010;38(Web Server issue):W214–20.
- Sticht C, De La Torre C, Parveen A, Gretz N. miRWalk: An online resource for prediction of microRNA binding sites. *PLoS One*. - 2018;13(10):e0206239.
- Karagkouni D, Paraskevopoulou MD, Tastsoglou S, et al. DIANA-LncBase v3: indexing experimentally supported miRNA targets on non-coding transcripts. *Nucleic Acids Res*. 2020;48(D1):D101–D110.
- Zeng GQ, Chen AB, Li W, Song JH, Gao CY. High MMP-1, MMP-2, and MMP-9 protein levels in osteoarthritis. *Genet Mol Res*. 2015;14(4):14811–14822.

30. **Lipari L, Gerbino A.** Expression of gelatinases (MMP-2, MMP-9) in human articular cartilage. *Int J Immunopathol Pharmacol.* 2013;26(3): 817–823.
31. **Hashizume M, Mihara M.** Desirable effect of combination therapy with high molecular weight hyaluronate and NSAIDs on MMP production. *Osteoarthritis Cartilage.* 2009;17(11):1513–1518.
32. **Vincenti MP, Brinckerhoff CE.** Transcriptional regulation of collagenase (MMP-1, MMP-13) genes in arthritis: integration of complex signaling pathways for the recruitment of gene-specific transcription factors. *Arthritis Res.* 2002;4(3):157–164.
33. **Latourte A, Richette P.** Inhibition of ADAMTS-5: the right target for osteoarthritis? *Osteoarthritis Cartilage.* 2022;30(2):175–177.
34. **Bomer N, den Hollander W, Ramos YFM, et al.** Underlying molecular mechanisms of DIO2 susceptibility in symptomatic osteoarthritis. *Ann Rheum Dis.* 2015;74(8):1571–1579.
35. **Han Y, Wu J, Gong Z, et al.** Identification and development of a novel 5-gene diagnostic model based on immune infiltration analysis of osteoarthritis. *J Transl Med.* 2021;19(1):522.
36. **Li S, Wang H, Zhang Y, et al.** COL3A1 and MMP9 serve as potential diagnostic biomarkers of osteoarthritis and are associated with immune cell infiltration. *Front Genet.* 2021;12:721258.
37. **Hu X, Ni S, Zhao K, Qian J, Duan Y.** Bioinformatics-led discovery of osteoarthritis biomarkers and inflammatory infiltrates. *Front Immunol.* 2022;13:871008.
38. **Li Z, Chen Z, Wang X, et al.** Integrated analysis of miRNAs and gene expression profiles reveals potential biomarkers for osteoarthritis. *Front Genet.* 2022;13:814645.
39. **Liu Y, Lu T, Liu Z, et al.** Six macrophage-associated genes in synovium constitute a novel diagnostic signature for osteoarthritis. *Front Immunol.* 2022;13:936606.
40. **Yang J, Fan Y, Liu S.** ATF3 as a potential diagnostic marker of early-stage osteoarthritis and its correlation with immune infiltration through bioinformatics analysis. *Bone Joint Res.* 2022;11(9):679–689.
41. **Rousseau JC, Chapurlat R, Garnero P.** Soluble biological markers in osteoarthritis. *Ther Adv Musculoskelet Dis.* 2021;13:1759720X211040300.
42. **van der Kraan PM, van den Berg WB.** Osteophytes: relevance and biology. *Osteoarthritis Cartilage.* 2007;15(3):237–244.
43. **Finnilä MAJ, Das Gupta S, Turunen MJ, et al.** Mineral crystal thickness in calcified cartilage and subchondral bone in healthy and osteoarthritic human knees. *J Bone Miner Res.* 2022;37(9):1700–1710.
44. **Lories RJ, Luyten FP.** The bone-cartilage unit in osteoarthritis. *Nat Rev Rheumatol.* 2011;7(1):43–49.
45. **Goldring MB, Goldring SR.** Articular cartilage and subchondral bone in the pathogenesis of osteoarthritis. *Ann NY Acad Sci.* 2010;1192:230–237.
46. **Chen GY, Nuñez G.** Sterile inflammation: sensing and reacting to damage. *Nat Rev Immunol.* 2010;10(12):826–837.
47. **van den Bosch MHJ, van Lent P, van der Kraan PM.** Identifying effector molecules, cells, and cytokines of innate immunity in OA. *Osteoarthritis Cartilage.* 2020;28(5):532–543.
48. **Kraus VB, McDaniel G, Huebner JL, et al.** Direct in vivo evidence of activated macrophages in human osteoarthritis. *Osteoarthritis Cartilage.* 2016;24(9):1613–1621.
49. **Haseeb A, Haqqi TM.** Immunopathogenesis of osteoarthritis. *Clin Immunol.* 2013;146(3):185–196.
50. **Sae-Jung T, Sengprasert P, Apinun J, et al.** Functional and T cell receptor repertoire analyses of peripheral blood and infrapatellar fat pad T cells in knee osteoarthritis. *J Rheumatol.* 2019;46(3):309–317.
51. **de Jong H, Berlo SE, Hombrink P, et al.** Cartilage proteoglycan aggrecan epitopes induce proinflammatory autoreactive T-cell responses in rheumatoid arthritis and osteoarthritis. *Ann Rheum Dis.* 2010;69(1):255–262.
52. **Shan Y, Qi C, Liu Y, Gao H, Zhao D, Jiang Y.** Increased frequency of peripheral blood follicular helper T cells and elevated serum IL-21 levels in patients with knee osteoarthritis. *Mol Med Rep.* 2017;15(3):1095–1102.
53. **Jaime P, García-Guerrero N, Estella R, Pardo J, García-Álvarez F, Martínez-Lostao L.** CD56+/CD16– Natural killer cells expressing the inflammatory protease granzyme A are enriched in synovial fluid from patients with osteoarthritis. *Osteoarthritis Cartilage.* 2017;25(10):1708–1718.
54. **Steck E, Benz K, Lorenz H, Loew M, Gress T, Richter W.** Chondrocyte expressed protein-68 (CEP-68), a novel human marker gene for cultured chondrocytes. *Biochem J.* 2001;353(Pt 2):169–174.
55. **Szilagyi IA, Vallerga CL, Boer CG, et al.** Plasma proteomics identifies CRTAC1 as a biomarker for osteoarthritis severity and progression. *Rheumatology (Oxford).* 2023;62(3):1286–1295.
56. **Styrkarsdottir U, Lund SH, Thorleifsson G, et al.** Cartilage Acidic Protein 1 in Plasma Associates With Prevalent Osteoarthritis and Predicts Future Risk as Well as Progression to Joint Replacements: Results From the UK Biobank Resource. *Arthritis Rheumatol.* 2023;75(4):544–552.
57. **Ge X, Ritter SY, Tsang K, Shi R, Takei K, Aliprantis AO.** Sex-specific protection of osteoarthritis by deleting cartilage acid protein 1. *PLoS One.* 2016;11(7):e0159157.
58. **Lee Y, Kim HJ, Park CK, Kim WS, Lee ZH, Kim HH.** Novel extraneural role of neurite outgrowth inhibitor A: modulation of osteoclastogenesis via positive feedback regulation of nuclear factor of activated T cell cytoplasmic 1. *J Bone Miner Res.* 2012;27(5):1043–1054.
59. **Goldring SR, Goldring MB.** Changes in the osteochondral unit during osteoarthritis: structure, function and cartilage-bone crosstalk. *Nat Rev Rheumatol.* 2016;12(11):632–644.
60. **Sharma AR, Jagga S, Lee SS, Nam JS.** Interplay between cartilage and subchondral bone contributing to pathogenesis of osteoarthritis. *Int J Mol Sci.* 2013;14(10):19805–19830.
61. **Burr DB, Gallant MA.** Bone remodelling in osteoarthritis. *Nat Rev Rheumatol.* 2012;8(11):665–673.
62. **Bertrand J, Cromme C, Umlauf D, Frank S, Pap T.** Molecular mechanisms of cartilage remodelling in osteoarthritis. *Int J Biochem Cell Biol.* 2010;42(10):1594–1601.
63. **Hodgkinson T, Kelly DC, Curtin CM, O'Brien FJ.** Mechanosignalling in cartilage: an emerging target for the treatment of osteoarthritis. *Nat Rev Rheumatol.* 2022;18(2):67–84.
64. **Maldonado M, Nam J.** The role of changes in extracellular matrix of cartilage in the presence of inflammation on the pathology of osteoarthritis. *Biomed Res Int.* 2013;2013:284873.
65. **Vincent TL, McClurg O, Troeberg L.** The extracellular matrix of articular cartilage controls the bioavailability of pericellular matrix-bound growth factors to drive tissue homeostasis and repair. *Int J Mol Sci.* 2022;23(11):6003.
66. **Kim J-H, Lee G, Won Y, et al.** Matrix cross-linking-mediated mechano-transduction promotes posttraumatic osteoarthritis. *Proc Natl Acad Sci U S A.* 2015;112(30):9424–9429.
67. **Meulenbelt I, Min JL, Bos S, et al.** Identification of DIO2 as a new susceptibility locus for symptomatic osteoarthritis. *Hum Mol Genet.* 2008;17(12):1867–1875.
68. **Wang L, Shao YY, Ballock RT.** Thyroid hormone-mediated growth and differentiation of growth plate chondrocytes involves IGF-1 modulation of beta-catenin signaling. *J Bone Miner Res.* 2010;25(5):1138–1146.
69. **Adams SL, Cohen AJ, Lassová L.** Integration of signaling pathways regulating chondrocyte differentiation during endochondral bone formation. *J Cell Physiol.* 2007;213(3):635–641.
70. **Goldring MB.** Insight into the function of DIO2, a susceptibility gene in human osteoarthritis, as an inducer of cartilage damage in a rat model: is there a role for chondrocyte hypertrophy? *Osteoarthritis Cartilage.* 2013;21(5):643–645.
71. **Nagase H, Nagasawa Y, Tachida Y, et al.** Deiodinase 2 upregulation demonstrated in osteoarthritis patients cartilage causes cartilage destruction in tissue-specific transgenic rats. *Osteoarthritis Cartilage.* 2013;21(3):514–523.
72. **Bomer N, Cornelis FMF, Ramos YFM, et al.** The effect of forced exercise on knee joints in Dio2(-/-) mice: type II iodothyronine deiodinase-deficient mice are less prone to develop OA-like cartilage damage upon excessive mechanical stress. *Ann Rheum Dis.* 2016;75(3): 571–577.
73. **Butterfield NC, Curry KF, Steinberg J, et al.** Publisher correction: Accelerating functional gene discovery in osteoarthritis. *Nat Commun.* 2021;12(1):3302.
74. **Jo S, Fonseca TL, Bocco BMLC, et al.** Type 2 deiodinase polymorphism causes ER stress and hypothyroidism in the brain. *J Clin Invest.* 2019;129(1):230–245.
75. **Aoi J, Endo M, Kadomatsu T, et al.** Angiotensin-like protein 2 accelerates carcinogenesis by activating chronic inflammation and oxidative stress. *Mol Cancer Res.* 2014;12(2):239–249.
76. **Tabata M, Kadomatsu T, Fukuhara S, et al.** Angiotensin-like protein 2 promotes chronic adipose tissue inflammation and obesity-related systemic insulin resistance. *Cell Metab.* 2009;10(3):178–188.

77. Sugimoto K, Nakamura T, Tokunaga T, et al. Angiopoietin-like protein 2 Induces synovial inflammation in the facet joint leading to degenerative changes via interleukin-6 secretion. *Asian Spine J.* 2019;13(3):368–376.
78. Okada T, Tsukano H, Endo M, et al. Synoviocyte-derived angiopoietin-like protein 2 contributes to synovial chronic inflammation in rheumatoid arthritis. *Am J Pathol.* 2010;176(5):2309–2319.
79. Tabata M, Kadomatsu T, Fukuhara S, et al. Angiopoietin-like protein 2 promotes chronic adipose tissue inflammation and obesity-related systemic insulin resistance. *Cell Metab.* 2009;10(3):178–188.
80. Takano M, Hirose N, Sumi C, et al. ANGPTL2 promotes inflammation via integrin $\alpha 5\beta 1$ in chondrocytes. *Cartilage.* 2021;13(2_suppl):885S–897S.
81. Zheng J, Umikawa M, Cui C, et al. Inhibitory receptors bind ANGPTLs and support blood stem cells and leukaemia development. *Nature.* 2012;485(7400):656–660.
82. Wagtmann N, Rojo S, Eichler E, Mohrenweiser H, Long EO. A new human gene complex encoding the killer cell inhibitory receptors and related monocyte/macrophage receptors. *Curr Biol.* 1997;7(8):615–618.
83. Chang CC, Ciubotariu R, Manavalan JS, et al. Tolerization of dendritic cells by T(S) cells: the crucial role of inhibitory receptors ILT3 and ILT4. *Nat Immunol.* 2002;3(3):237–243.
84. Nishiyama S, Hirose N, Yanoshita M, et al. ANGPTL2 induces synovial inflammation via LILRB2. *Inflammation.* 2021;44(3):1108–1118.
85. Shan W, Cheng C, Huang W, et al. Angiopoietin-like 2 upregulation promotes human chondrocyte injury via NF- κ B and p38/MAPK signaling pathway. *J Bone Miner Metab.* 2019;37(6):976–986.
86. Bragason BT, Palsdottir A. Interaction of PrP with NRAGE, a protein involved in neuronal apoptosis. *Mol Cell Neurosci.* 2005;29(2):232–244.
87. Kendall SE, Battelli C, Irwin S, Mitchell JG, Glackin CA, Verdi JM. NRAGE mediates p38 activation and neural progenitor apoptosis via the bone morphogenetic protein signaling cascade. *Mol Cell Biol.* 2005;25(17):7711–7724.
88. Wen C-J, Xue B, Qin W-X, et al. hNRAGE, a human neurotrophin receptor interacting MAGE homologue, regulates p53 transcriptional activity and inhibits cell proliferation. *FEBS Lett.* 2004;564(1–2):171–176.
89. Matluk N, Rochira JA, Karaczyn A, Adams T, Verdi JM. A role for NRAGE in NF-kappaB activation through the non-canonical BMP pathway. *BMC Biol.* 2010;8:7.
90. Liu M, Xu L, Ma X, et al. MAGED1 is a negative regulator of bone remodeling in mice. *Am J Pathol.* 2015;185(10):2653–2667.
91. Cho Y, Jeong S, Kim H, et al. Disease-modifying therapeutic strategies in osteoarthritis: current status and future directions. *Exp Mol Med.* 2021;53(11):1689–1696.
92. Colella F, Garcia JP, Sorbona M, et al. Drug delivery in intervertebral disc degeneration and osteoarthritis: Selecting the optimal platform for the delivery of disease-modifying agents. *J Control Release.* 2020;328:985–999.
93. Ali SA, Peffers MJ, Ormseth MJ, Jurisica I, Kapoor M. The non-coding RNA interactome in joint health and disease. *Nat Rev Rheumatol.* 2021;17(11):692–705.
94. He CP, Jiang XC, Chen C, et al. The function of lncRNAs in the pathogenesis of osteoarthritis. *Bone Joint Res.* 2021;10(2):122–133.
95. Yang J, Zhou Y, Liang X, Jing B, Zhao Z. MicroRNA-486 promotes a more catabolic phenotype in chondrocyte-like cells by targeting SIRT6: possible involvement in cartilage degradation in osteoarthritis. *Bone Joint Res.* 2021;10(7):459–466.
96. Tang H, Zhu W, Cao L, et al. miR-210-3p protects against osteoarthritis through inhibiting subchondral angiogenesis by targeting the expression of TGFBR1 and ID4. *Front Immunol.* 2022;13:982278.
97. Mao G, Zhang Z, Huang Z, et al. MicroRNA-92a-3p regulates the expression of cartilage-specific genes by directly targeting histone deacetylase 2 in chondrogenesis and degradation. *Osteoarthritis Cartilage.* 2017;25(4):521–532.
98. Wei W, He S, Wang Z, et al. LINC01534 Promotes the aberrant metabolic dysfunction and inflammation in il-1 β -simulated osteoarthritic chondrocytes by targeting miR-140-5p. *Cartilage.* 2021;13(2_suppl):898S–907S.
99. Stelcer E, Kulcenty K, Rucinski M, et al. The role of MicroRNAs in early chondrogenesis of human induced pluripotent stem cells (hiPSCs). *Int J Mol Sci.* 2019;20(18):18.

Author information

D. Zhao, PhD, Doctoral Student

F. Lin, PhD, Doctoral Student

Second Clinical Medical College of Guangzhou University of Chinese Medicine, Guangzhou, China; Bone and Joint Research Team of Degeneration and Injury, Guangdong Provincial Academy of Chinese Medical Sciences, Guangzhou, China.

L. Zeng, PhD, Associate Professor

G. Liang, MSc, Assistant Professor

M. Luo, MSc, Associate Professor

J. Pan, PhD, Associate Professor

Y. Dou, PhD, Assistant Professor

H. Huang, PhD, Assistant Professor

W. Yang, PhD, Professor

Bone and Joint Research Team of Degeneration and Injury, Guangdong Provincial Academy of Chinese Medical Sciences, Guangzhou, China; Department of Orthopedics, The Second Affiliated Hospital of Guangzhou University of Chinese Medicine, Guangzhou, China.

J. Liu, MD, Professor, Bone and Joint Research Team of Degeneration and Injury, Guangdong Provincial Academy of Chinese Medical Sciences, Guangzhou, China; Guangdong Second Traditional Chinese Medicine Hospital, Guangdong Province Engineering Technology Research Institute of Traditional Chinese Medicine, Guangzhou, China; Fifth Clinical Medical College of Guangzhou University of Chinese Medicine, Guangzhou, China.

Author contributions

D. Zhao: Conceptualization, Investigation, Formal analysis, Methodology, Writing – original draft.

L. Zeng: Conceptualization, Data curation, Methodology, Validation.

G. Liang: Conceptualization, Investigation, Methodology, Writing – review & editing.

M. Luo: Data curation, Methodology, Validation.

J. Pan: Conceptualization, Formal analysis, Validation.

Y. Dou: Methodology, Software, Visualization.

F. Lin: Conceptualization, Software, Visualization.

H. Huang: Investigation, Formal analysis.

W. Yang: Data curation, Project administration, Supervision, Writing – review & editing.

J. Liu: Data curation, Project administration, Resources, Supervision, Writing – review & editing.

D. Zhao, L-F. Zeng, and G-H. Liang contributed equally to this work.

W-Y. Yang and J. Liu contributed equally to this work.

Funding statement

The authors disclose receipt of the following financial or material support for the research, authorship, and/or publication of this article: the study was supported by the National Natural Science Foundation of China (No. 82305263, No. 81873314, No. 82004386, No. 82004383), Natural Science Foundation of Guangdong Province (2022A1515010385, 2022A1515011700, 2023A1515012626), the Science and Technology Research Project of Guangdong Provincial Hospital of Chinese Medicine (No. YN2022GK05, YN2019ML08), Guangdong Provincial Key Laboratory of Chinese Medicine for Prevention and Treatment of Refractory Chronic Diseases, Health Appropriate Technology Promotion Project of Guangdong Province (No.202303211022372094).

ICMJE COI statement

The study was supported by the National Natural Science Foundation of China (No. 82305263, No. 81873314, No. 82004386, No. 82004383), Natural Science Foundation of Guangdong Province (2022A1515010385, 2022A1515011700, 2023A1515012626), the Science and Technology Research Project of Guangdong Provincial Hospital of Chinese Medicine (No. YN2022GK05, YN2019ML08), Guangdong Provincial Key Laboratory of Chinese Medicine for Prevention and Treatment of Refractory Chronic Diseases, Health Appropriate Technology Promotion Project of Guangdong Province (No.202303211022372094).

Data sharing

The data that support the findings for this study are available to other researchers from the corresponding author upon reasonable request.

Acknowledgements

Yan-hong Han and Nan-jun Xu assisted with data interpretation.

Open access funding

The authors report that they received open access funding for their manuscript from the National Natural Science Foundation of China (No. 82305263, No. 81873314, No. 82004386, No. 82004383) and the Natural Science Foundation of Guangdong Province (2022A1515010385, 2022A1515011700, 2023A1515012626).

© 2024 Liu et al. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial No Derivatives (CC BY-NC-ND 4.0) licence, which permits the copying and redistribution of the work only, and provided the original author and source are credited. See <https://creativecommons.org/licenses/by-nc-nd/4.0/>