ORIGINAL RESEARCH

Investigating Angiogenesis-Related Biomarkers in Osteoarthritis Patients Through Transcriptomic Profiling

Yang Zheng [1], Miaojia Fang[2], Shriya Sanan[1], Xi-Hui Meng[1], Jie-Feng Huang [1], Yu Qian[1]

[1]Zhejiang Provincial Hospital of Traditional Chinese Medicine, Hangzhou, People's Republic of China; [2]Institute of Forensic Science, Yuhang Public Security Department, Hangzhou, People's Republic of China

Correspondence: Yu Qian, Zhejiang Provincial Hospital of Traditional Chinese Medicine, 54 Youdian Road, Shangcheng District, Hangzhou, 310006, People's Republic of China, Tel +86-571-87068001, Fax +86-571-87034117, Email 20223022@zcmu.edu.cn

Background: Osteoarthritis (OA) is a common age-related joint disease characterized by joint destruction and impaired quality of life. Angiogenesis plays a vital role in OA progression. This study aimed to identify key angiogenesis-related genes (ARGs) in OA using transcriptomic and machine learning methods.

Methods: The GSE55235 dataset (10 OA and 10 healthy synovial tissue samples) was analyzed for differentially expressed genes (DEGs), integrated with weighted gene co-expression network analysis (WGCNA), and ARGs to identify differentially expressed ARGs (DE-ARGs). Candidate genes were identified through three machine learning algorithms and evaluated using ROC curve analysis. Gene set enrichment analysis (GSEA), immune cell infiltration analysis, and therapeutic agent prediction were performed. Synovial samples from 5 OA patients and 5 matched controls were collected for RT-qPCR validation of biomarkers.

Results: From 1552 DEGs, 11 DE-ARGs were identified, and six candidate genes were selected using machine learning. Four genes —COL3A1, OLR1, STC1, and KCNJ8—showed AUC >0.8 in both GSE55235 and GSE1919, indicating strong diagnostic value. GSEA linked biomarkers to the "lysosome" pathway, and eosinophils and Th2 cells were significantly associated with biomarkers. Potential therapeutic agents included bisphenol A, tetrachlorodibenzo-p-dioxin, and valproic acid. Clinical validation confirmed that COL3A1, OLR1, and STC1 expression levels were consistent with database findings.

Conclusion: The study identified COL3A1, OLR1, STC1, and KCNJ8 as key angiogenesis-related biomarkers in osteoarthritis, which could serve as potential diagnostic tools and therapeutic targets. The research underscores the importance of angiogenesis in osteoarthritis progression and suggests that targeting angiogenesis-related pathways may offer new treatment strategies.

Keywords: osteoarthritis, angiogenesis, biomarker, immune infiltration, regulatory network

Introduction

Osteoarthritis (OA) (ICD-10: M19) is a common degenerative joint disease marked by the progressive breakdown of joint cartilage, osteophyte formation, subchondral bone sclerosis, and synovial inflammation.[1] It results from a combination of mechanical stress, inflammation, and biochemical factors, causing pain, stiffness, and reduced joint mobility. In individuals aged 40 to 60, the prevalence of OA is approximately 10–17% and around 50%, respectively. Among those aged 75 and older, the prevalence rises significantly to about 80%.[2] Factors like gender, geography, and physical labor intensity influence its prevalence. In China, women are 1–2 times more susceptible than men, and colder, humid regions and occupations involving heavy labor or prolonged standing show higher OA rates.[3,4] Current OA management focuses on symptom relief through nonsteroidal anti-inflammatory drugs (NSAIDs), intra-articular hyaluronic acid injections, and physical therapy to reduce pain and improve function.[2–4] However, these treatments cannot stop disease progression or repair damaged cartilage. Recent research seeks new biomarkers and therapeutic targets to better understand OA's underlying mechanisms, aiming to develop more effective interventions that delay or prevent joint degeneration.[3–5]

Emerging research identifies angiogenesis as a critical pathological mechanism in OA progression. Unlike healthy avascular cartilage, OA cartilage experiences abnormal vascularization driven by increased osteoclast activity, which creates channels between subchondral bone and cartilage.[5] This process facilitates the influx of sensory nerves, exacerbating joint pain and accelerating cartilage degradation. Targeting angiogenesis could offer a promising therapeutic strategy for OA. However, significant gaps persist in identifying angiogenesis-related biomarkers to improve diagnosis and treatment. While prior studies have emphasized symptom management and structural changes, the molecular mechanisms underlying vascularization remain underexplored.[4,5]

This study aims to bridge existing gaps by identifying angiogenesis-related biomarkers in OA through transcriptomic analysis. Utilizing machine learning and immune cell profiling, it seeks to pinpoint biomarkers and therapeutic targets for early diagnosis and innovative treatments to slow OA progression. By targeting these biomarkers, the study aspires to move beyond symptom-based therapies, improve patient outcomes, and support personalized medicine. Clinically, these findings hold potential to mitigate OA's impact, enhance patient quality of life, and reduce the healthcare burden associated with advanced disease stages.

Materials and Methods

Data Source

The OA (ICD-10: M19) datasets GSE55235 and GSE1919 were obtained from the Gene Expression Ontology (GEO) database (https://www.ncbi.nlm.nih.gov/geo/). The GSE55235 dataset, used as the training set, includes high-throughput sequencing data of synovial tissue from 30 samples on the GPL96 platform. After removing 10 samples of rheumatoid arthritis, the sequencing data of synovial tissue from 10 OA patients and 10 control groups were analyzed.[6] The GSE1919 dataset (GPL91) was used as a validation set, with 5 rheumatoid arthritis samples removed, and contained five OA patients and five controls synovial tissue samples (Supplementary Table 1).[7] All 36 angiogenesis related genes (ARGs) were retrieved from the Molecular Signatures database (MSigDB).[8]

Selection of Differentially Expressed ARGs (DE-ARGs)

Differentially expressed genes (DEGs) between OA and control samples in the GSE55235 dataset were identified using the "limma" package (v3.56.2),[9] with selection criteria of |log2 fold change (FC)| > 0.5 and p-value < 0.05. Volcano and heatmaps were generated using the "ggplot2" (v3.4.2)[10] and "circlize" (v0.4.15)[11] packages, respectively. Weighted gene co-expression network analysis (WGCNA) was then applied to identify modules most strongly associated with OA.[12] The process involved removing outlier samples via clustering analysis and determining the soft threshold (β) by achieving a scale-free R^2 of ~0.9 with mean connectivity near zero. Genes with similar expression patterns were grouped into modules using the dynamic tree-cutting method (minModuleSize = 200). Pearson's correlation analysis was used to identify the key modules most correlated with OA, and their genes were designated as key module genes. Finally, DE-ARGs were identified by intersecting DEGs, key module genes, and ARGs using the "ggvenn" package (v0.1.9).

Explored the Biological Functions of DE-ARGs and Their Interactions at the Protein Level

To explore the biological functions and signaling pathways of DE-ARGs, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway annotations were performed using the "clusterProfiler" (version 4.8.2) ($p < 0.05$).[13,14] In addition, a protein-protein interaction (PPI) network of DE-ARGs was constructed in the search tool for recurring instances of neighbouring genes (STRING) (interaction score greater than 0.4), which was designed to observe whether there is an interaction relationship between DE-ARGs at the protein level.[15,16]

Recognition and Confirmation of Biomarkers

To identify characteristic genes of OA, DE-ARGs were analyzed using the least absolute shrinkage and selection operator (LASSO) method with the "glmnet" package (v4.1–6), performing 3-fold cross-validation with a regularization parameter of 4e-04 to determine optimal coefficients and lambda values.[17,18] The "e1071" package (v1.7–12)[19] was used to

apply the support vector machine recursive feature elimination (SVM-RFE) algorithm with a linear kernel function (kernel = "linear") and a cost parameter of 10. The most relevant features were selected by eliminating less important feature vectors.[20] Additionally, the Boruta algorithm was employed to calculate the features of DE-ARGs. The exp1 dataset was used with group.v as the dependent variable, setting the tracking level (doTrace = 2) and sorting features to filter the most significant ones. The candidate genes were derived by intersecting results from these three machine learning algorithms. Receiver operating characteristic (ROC) curves were constructed using the "pROC" package (v1.18.4) based on the GSE55235 and GSE1919 datasets to assess the diagnostic performance of the candidate genes.[21] Genes with an area under the curve (AUC) greater than 0.8 were considered to have good diagnostic efficacy for OA and were designated as biomarkers. Additionally, the expression profiles of these biomarkers in OA and control samples were validated in the GSE55235 and GSE1919 datasets.

Construction of a Neural Network Model for Biomarkers

To further assess the ability of biomarkers to differentiate between OA and control samples, a back-propagation neural network (BPNN) model was constructed based on the GSE55235 dataset using "neuralnet" (version 0.0.2).[22] After that, the discriminative effect of the model was verified with the help of ROC curves.

Biomarker Correlation Analysis and Subcellular Localisation Analysis

Spearman correlation analysis between biomarkers was performed using the "psych" package (version 2.3.12).[23] To further observe the subcellular localisation of the biomarkers, the proteins of the biomarkers were subcellularly localised using the Hum-mPLoc 3.0.[24]

Biological Pathways Involved in Biomarkers and the Construction of Gene Interaction Networks

Based on the GSE55235 dataset, Spearman correlation analysis of the relationships between biomarkers and other genes were performed separately using the "psych".[25] Then, gene set enrichment analysis (GSEA) was conducted for individual biomarkers using the "clusterProfiler" package ($|NES| > 1$, p.adjust < 0.05).[26] The background gene set "c2.cp.kegg. v2023.1.Hs.symbols" was downloaded from the GSEA website.[27] To further explore the interactions between biomarkers and other genes with similar gene functions, it was necessary to have explored them in GeneMANIA (http://genemania.org/) and constructed co-expression networks between them.[28]

Analysis of the N6-Methyladenosine (m6A) Binding Site Prediction

To explore whether m^6A affected the translational stability of biomarkers, it was necessary to predict the biomarker m^6A binding sites. First, the genome sequence for the biomarkers were retrieved from the National Center for Biotechnology Information (NCBI).[29] Then the m^6A binding sites of the biomarkers were predicted from the sequence-based RNA adenosine methylation site predictor database (SRAMP).[30]

Immune Infiltration Analysis

In terms of immune infiltration, each sample in the GSE55235 dataset was analysed separately for immune cell infiltration by the ssGSEA algorithm of the "GSVA" (version 1.44.5), and scores for 28 immune cells in the samples were calculated.[31] Then, the distribution of immune cells between OA and controls was demonstrated using the "pheatmap" (version 1.0.12).[32] Next, immune cells that were significantly different between OA and controls were compared using the Wilcoxon's test. Additionally, Spearman correlation analysis was used to evaluate the correlation between 28 immune cells. Using the same approach, correlations between biomarkers and differential immune cells were explored.

Construction of Molecular Regulatory Networks and Prediction of Potential Drugs

The miRNAs of the biomarkers were predicted in the miRNA target gene prediction database (miRTarBase) and starBase database, respectively, and the intersection of the expected findings in the two databases was used to identify the miRNAs

that corresponded to the biomarkers.[33,34] Meanwhile, the ChIP-X Enrichment Analysis 3 (ChEA3) was used to predict the transcription factors (TFs) corresponding to the biomarkers, and TFs supported by the ENCODE database in ChIP-seq were selected.[35] Also, the comparative toxicogenomics database (CTD) was used to predict biomarker target drugs, which were screened based on the expression of the biomarker in OA (drugs with reversible effects on biomarker expression) and the amount of data on drug therapy (Reference Count $\geq$ 2).[36]

Reverse Transcription Quantitative Polymerase Chain Reaction (RT-qPCR)

To further confirm the expression levels of biomarkers between the OA group and the control group, this study recruited 5 knee OA patients and 5 healthy controls matched by age, gender, and body mass index from August 1, 2023, to March 31, 2024 (Supplementary Table 2), and collected their synovial samples for RT-qPCR to evaluate the expression of biomarkers. Eligible patients were aged 50 years or older, had experienced knee pain on most days in the past month, had an average knee pain score of 4 or higher on the visual analogue scale in the past week, and had moderate radiographic tibiofemoral OA (Kellgren and Lawrence grade 3 or 4). Exclusion criteria included systemic or inflammatory diseases; use of anti-inflammatory drugs or corticosteroids in the past 3 months; hyaluronic acid injections in the past 6 months; prior treatment with autologous blood products or stem cell preparations; bleeding disorders or current anticoagulant therapy. All participants were followed up, and no one dropped out. Total RNA of 10 samples was separated by the TRIzol (Ambion, Austin, USA) based on the manufacturer's guidance. The inverse transcription of total RNA into cDNA was conducted using the SureScript-First-strand-cDNA-synthesis-kit (Servicebio, Wuhan, China) based on the producer's indication. Then, qPCR was carried out utilizing the 2xUniversal Blue SYBR Green qPCR Master Mix (Servicebio, Wuhan, China) under the direction of the manual. The primer sequences for PCR were tabulated in (Supplementary Table 3). The expression was uniformized to the internal reference GAPDH and computed employing the $2^{-\Delta\Delta Ct}$ method.[37]

Statistical Analysis

This study was conducted using the R programming language (version 4.3.1), and Wilcoxon's test was employed to evaluate the data from various groups. A p-value of less than 0.05 was statistically significant, unless otherwise stated.

Results
## Gene Screening of Data Set

A total of 1552 DEGs between OA and controls were identified in the GSE55235 dataset, of which a total of 954 genes were up-regulated, and 598 genes were down-regulated (Figure 1A). Expression levels of DEGs between different subgroups were confirmed (Figure 1B). Also, the samples in the GSE55235 dataset were clustered well enough to be used in subsequent analyses (Figure 1C). Immediately thereafter, β was 9 for a vertical coordinate R^2 close to 0.9, and the mean connectivity of the network converges to 0 (Figure 1D). Afterwards, the 13 co-expressed gene modules of different colours were obtained (Figure 1E). Subsequently, the MEturquoise (|cor| = 0.97, $p < 0.001$) had the highest correlation with OA, and as such, and it was recorded as a key module for OA and contained 3014 key module genes (Figure 1F).

Identification of DE-ARGs and Its Functional Annotation

The 11 DE-ARGs were obtained at the intersection of 1552 DEGs, 3014 key module genes, and 36 ARGs (Figure 2A). The Gene Ontology (GO) analysis of DE-ARG showed that 380 significant entries were obtained in biological processes, 15 significant entries were enriched in cellular components, and 33 significant entries were obtained in molecular functions. The most important of these were "cartilage development", "connective tissue development", "bone morphogenesis" and so on ($p < 0.05$) (Figure 2B). DE-ARGs were significantly enriched for "protein digestion and absorption" and "human T-cell leukemia virus 1 infection" (Figure 2C). Meanwhile, the PPI network consisted of 11 DE-ARGs and 9 edges, among which the strongest interaction at the protein level was between COL5A2 and COL3A1 (Figure 2D).
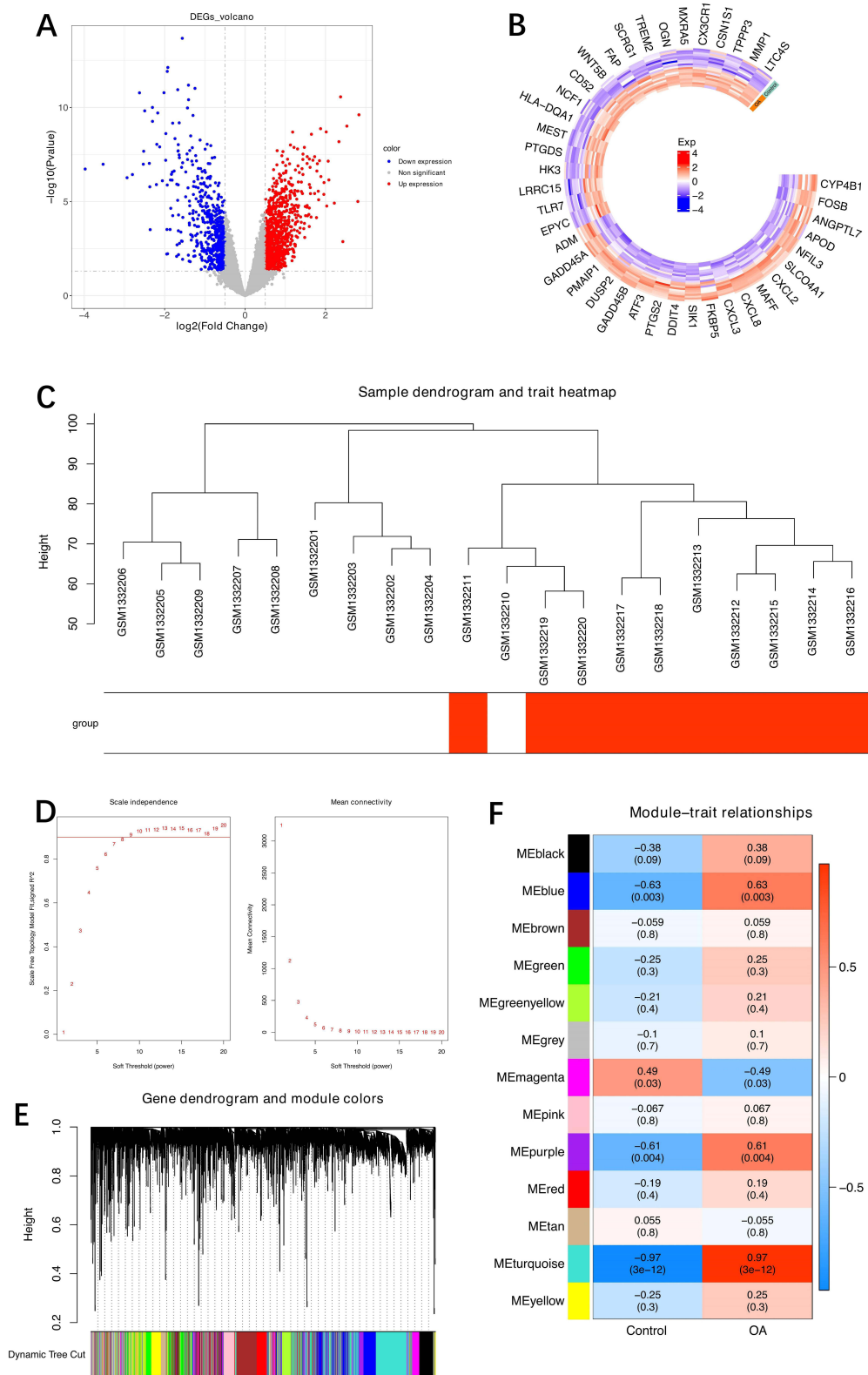
Figure 1 Identification and analysis of DEGs. (**A**) Volcano plot of DEGs. The red dots represented up - regulated differentially expressed genes, the blue dots represented down - regulated differentially expressed genes, and the grey ones were genes without significant statistical significance. Each dot represented a gene. (**B**) Heatmap of DEGs. The abscissa of the circle represented the samples. Above it, green represented the Control samples and Orange represented the OA samples. The ordinate represented the genes. In the figure, red represented highly - expressed genes and blue represented low - expressed genes. (**C**) Sample clustering diagram. The branches represented the samples, the ordinate represented the height of hierarchical clustering, and in the lower part of the figure, red represented OA samples and white represented Control samples. (**D**) Soft threshold screening. (**E**) Module dynamic cutting tree. (**F**) Heatmap of module-disease correlation. Red represented positive correlation, blue represented negative correlation, and the stronger the correlation was, the darker the color was.
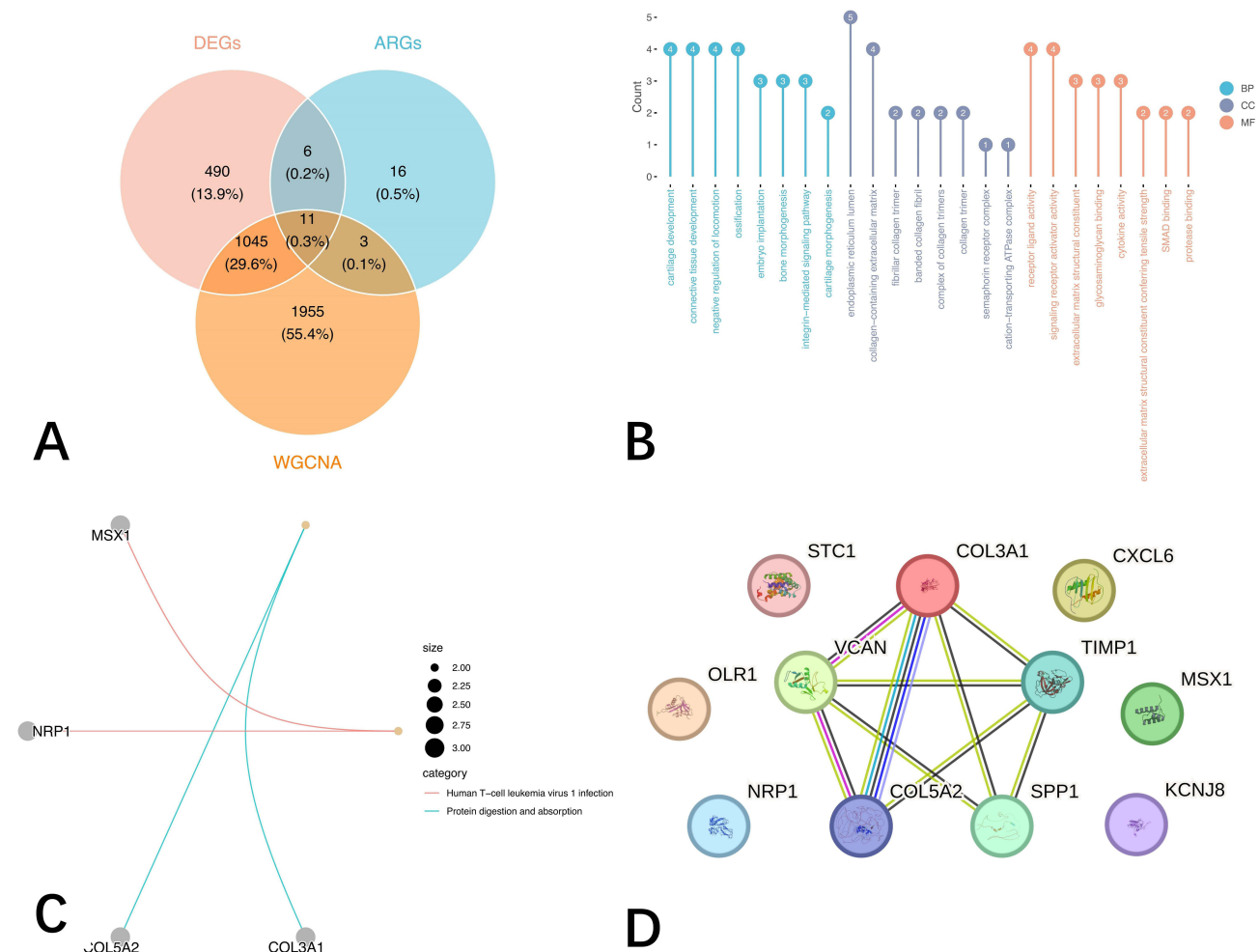
Figure 2 Identification of DE-ARGs and its functional annotation. (**A**) Venn diagram of DE-ARGs. (**B**) GO enrichment analysis of DE-ARGs. The vertical axis, labeled "Count", represented the number of genes associated with each GO term, while the horizontal axis showed the name of each GO term. (**C**) KEGG enrichment analysis of DE-ARGs. The outer circle was the names of genes and pathways, and the size of the dots represented the number of genes on the pathways. (**D**) PPI network.

Identified as Biomarkers of OA: COL3A1, OLR1, STC1, and KCNJ8

There were COL3A1, OLR1, STC1, NRP1, KCNJ8, and CXCL6 with special significance for OA that were screened by LASSO regression analysis (Figure 3A and B). The SVM-RFE algorithm identified STC1, COL3A1, OLR1, KCNJ8, COL5A2, CXCL6, MSX1, TIMP1, VCAN, and NRP1 as important characteristic genes of OA (Figure 3C). The Boruta algorithm obtained COL3A1, OLR1, COL5A2, TIMP1, STC1, VCAN, NRP1, MSX1, KCNJ8, and CXCL6 as characteristic genes of OA (Figure 3D). The intersection of results from LASSO, SVM-RFE, and Boruta algorithms yielded COL3A1, OLR1, STC1, NRP1, KCNJ8, and CXCL6 as candidate OA biomarkers (Figure 3E). Among them, the AUC values of COL3A1, OLR1, STC1, and KCNJ8 were all greater than 0.8 in both the GSE55235 and GSE1919 datasets, indicating that these genes had good diagnostic significance. Accordingly, they were recorded as biomarkers of OA (Figure 3F and G). Additionally, the expressions of COL3A1 and OLR1 were significantly upregulated in OA, whereas the expressions of STC1 and KCNJ8 were significantly downregulated in OA on the GSE55235 dataset. And their expression trends were consistent with the GSE1919 dataset (Figure 3H and I).

The BPNN Model Confirmed That the Biomarkers Were Effective in Distinguishing OA from Controls

The BPNN model as shown in the figure, includes input, hidden and output layers. The input layers compromised COL3A1, OLR1, STC1, and KCNJ8, and the output layers were OA and controls (Figure 4A). Subsequently, in the GSE55235 dataset, the AUC was 1, indicating that the biomarkers were able to effectively predict OA and controls (Figure 4B).
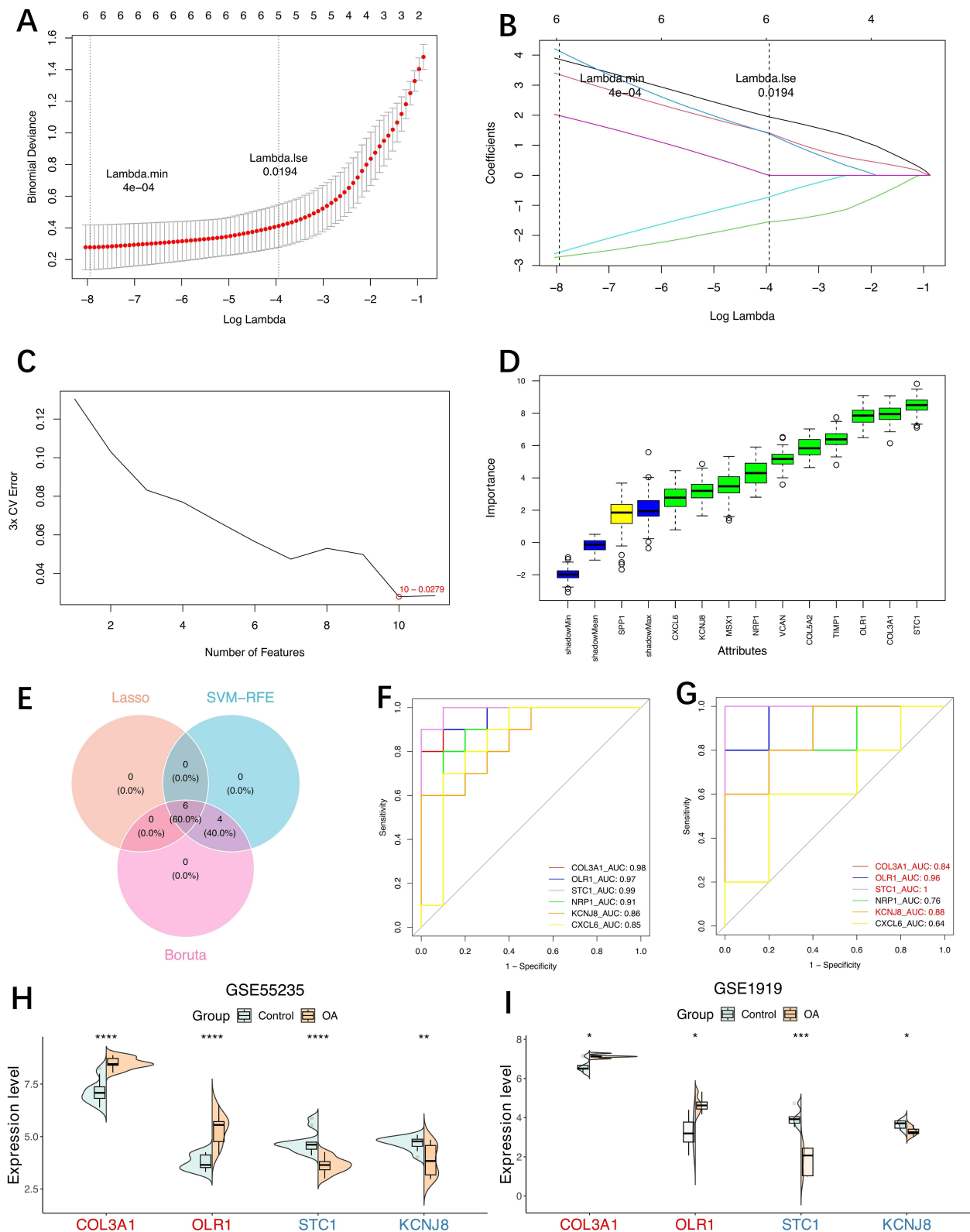
Figure 3 Selection and validation of key genes using machine learning. (**A** and **B**) Lasso regression analysis. In the left figure: the abscissa was log(Lambda), and the ordinate represented the cross - validation error; in the right figure: the abscissa "deviance" represented the proportion of residuals explained by the model, and the ordinate was the coefficient of the gene. (**C**) SVM-RFE selection of candidate genes. The abscissa represented the number of genes, and the ordinate represented the error rate. (**D**) Boruta selection of candidate genes. Green represented important features, yellow represented features with unclear importance, and the blue box showed the minimum, average and maximum Z - values of the shaded attribute. (**E**) Venn diagram of three machine learning algorithms. (**F**) ROC curve of key genes in training set GSE55235. (**G**) ROC curve of key genes in validation set GSE1919. (**H**) The expression levels of key genes in training set GSE55235. (**I**) The expression levels of key genes in validation set GSE1919. *$P<0.05$. **$P<0.01$. ***$P<0.001$. ****$P<0.0001$.
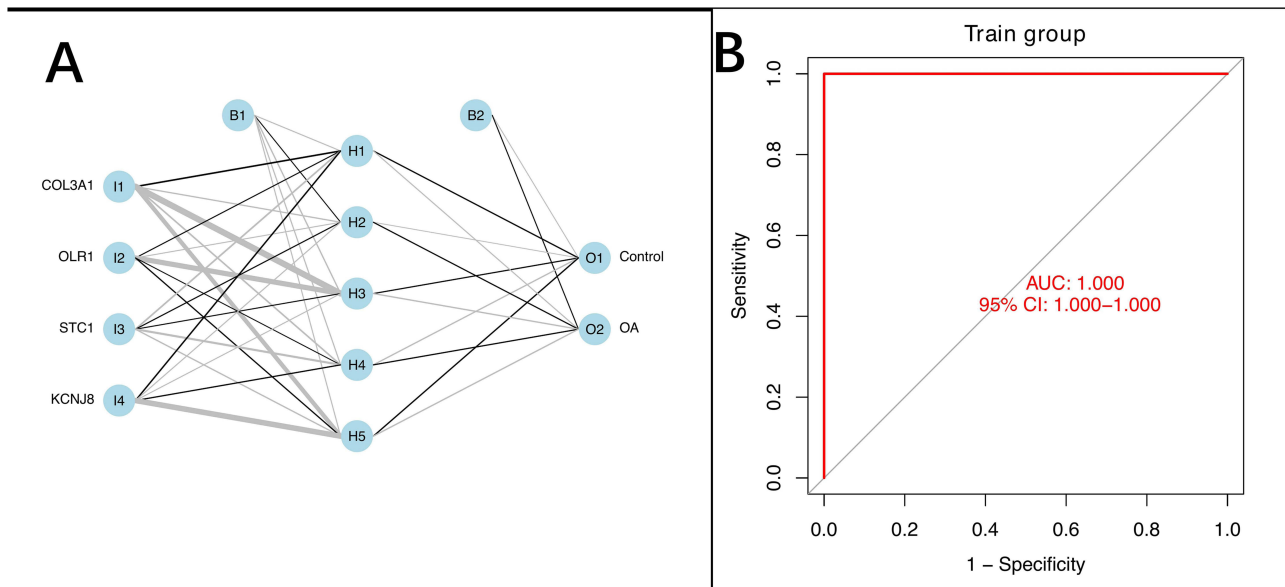
Figure 4 Construction and performance of BP neural network model. (**A**) BP neural network model diagram. (**B**) BP neural network ROC curve.

Biomarker Correlation and Subcellular Localization Results

The correlations between the biomarkers showed significantly positive correlations between COL3A1 and OLR1 (cor = 0.72, $p < 0.001$). In contrast, there was a significant negative correlation between COL3A1 and STC1 (cor = −0.66, $p < 0.01$). Similarly, OLR1 was significantly negatively correlated with both KCNJ8 and STC1, with correlation coefficients of −0.45 and −0.66, respectively ($p < 0.05$). And the relationship between STC1 and KCNJ8 was a significant positive correlation (cor = 0.62, $p < 0.05$) (Figure 5A). Meanwhile, the subcellular locations of the biomarkers showed that OLR1 was predominantly in the Golgi apparatus, KCNJ8 was predominantly in the plasma membrane, whereas COL3A1 and STC1 were located predominantly extracellularly (Figure 5B–E).
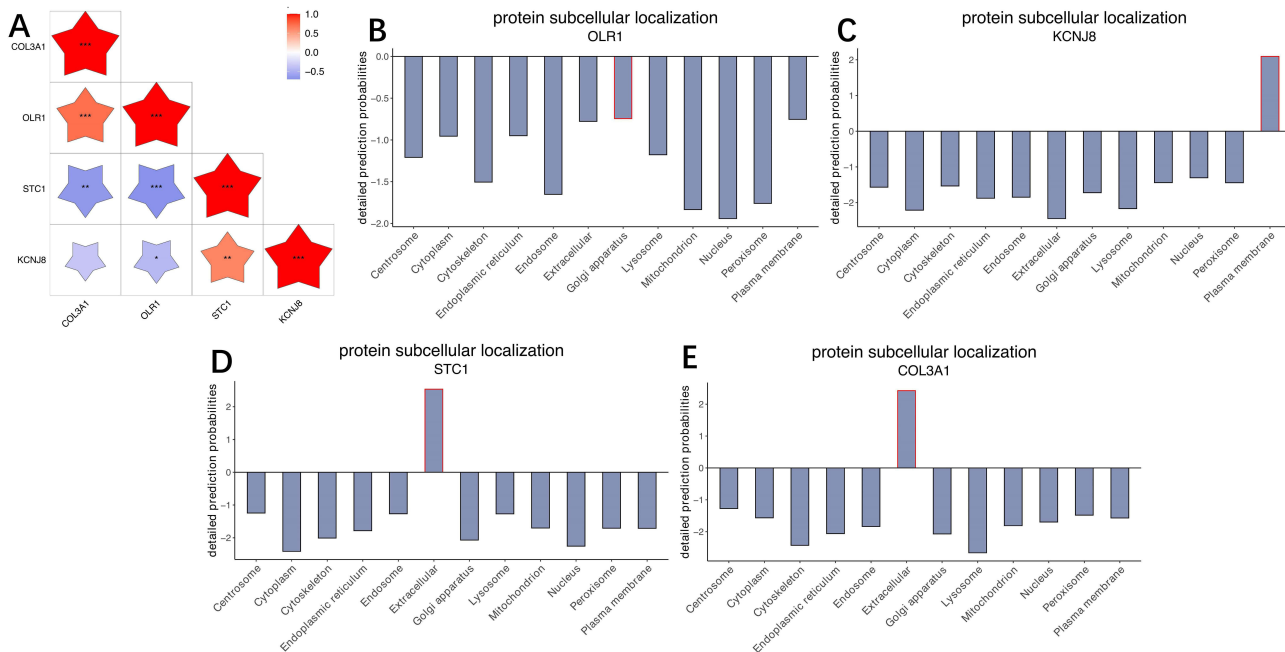
Figure 5 Correlation and subcellular localization of key genes. (**A**) Scatter plot of key gene correlation. The darker the color was, the higher the correlation was. (**B**) Subcellular localization of COL3A1 gene. (**C**) Subcellular localization of OLR1 gene. (**D**) Subcellular localization of STC1 gene. (**E**) Subcellular localization of KCNJ8 gene. * represented $P<0.05$, ** represented $P<0.01$, *** represented $P<0.001$.

10688
DovePress

Journal of Inflammation Research 2024:17

## Functional Correlation and Annotation of Biomarkers

GSEA results showed that COL3A1, OLR1, STC1, and KCNJ8 were enriched in 32, 21, 29, and 31 biological pathways, respectively. Among these, the "lysosome" was the co-enrichment pathway of these biomarkers (Figure 6A–D). Next, in the GeneMANIA results, only COL3A1 was closely linked to other functionally similar genes (Figure 6E). They were mainly involved in biological functions such as "fibrillar collagen trimer" and "extracellular matrix structural constituent".

Location of Bases at Which m6A Methylation of Biomarkers Occurs

The predicted locations of biomarkers susceptible to m^{6}A methylation showed that COL3A1 was more susceptible to m^{6}A methylation at positions 2196 and 484 of the base positions (Figure 7A). OLR1 was prone to m^{6}A methylation at approximately 107, 148, and 172 base positions (Figure 7B). STC1 was found to be susceptible to m^{6}A methylation at around 769, 1805, 2340, 2702, 3015, and 3047 base positions (Figure 7C). KCNJ8 was susceptible to m^{6}A methylation at around 1570 and 2166 base positions (Figure 7D).
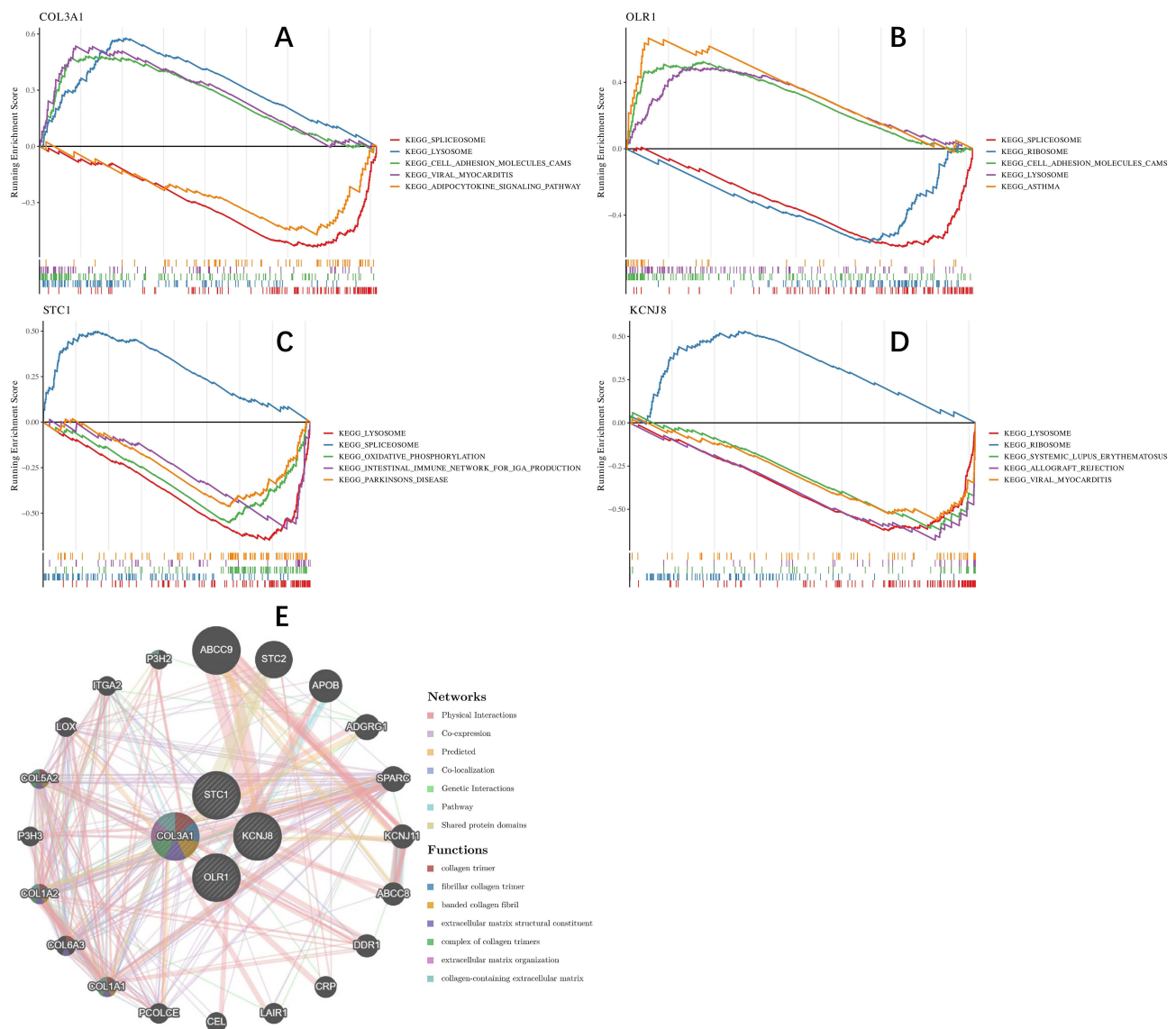
Figure 6 GSEA enrichment analysis and GeneMANIA network of key genes. (**A**) GSEA enrichment analysis of COL3A1 gene. (**B**) GSEA enrichment analysis of OLR1 gene. (**C**) GSEA enrichment analysis of STC1 gene. (**D**) GSEA enrichment analysis of KCNJ8 gene. (**E**) GeneMANIA network diagram of key genes.
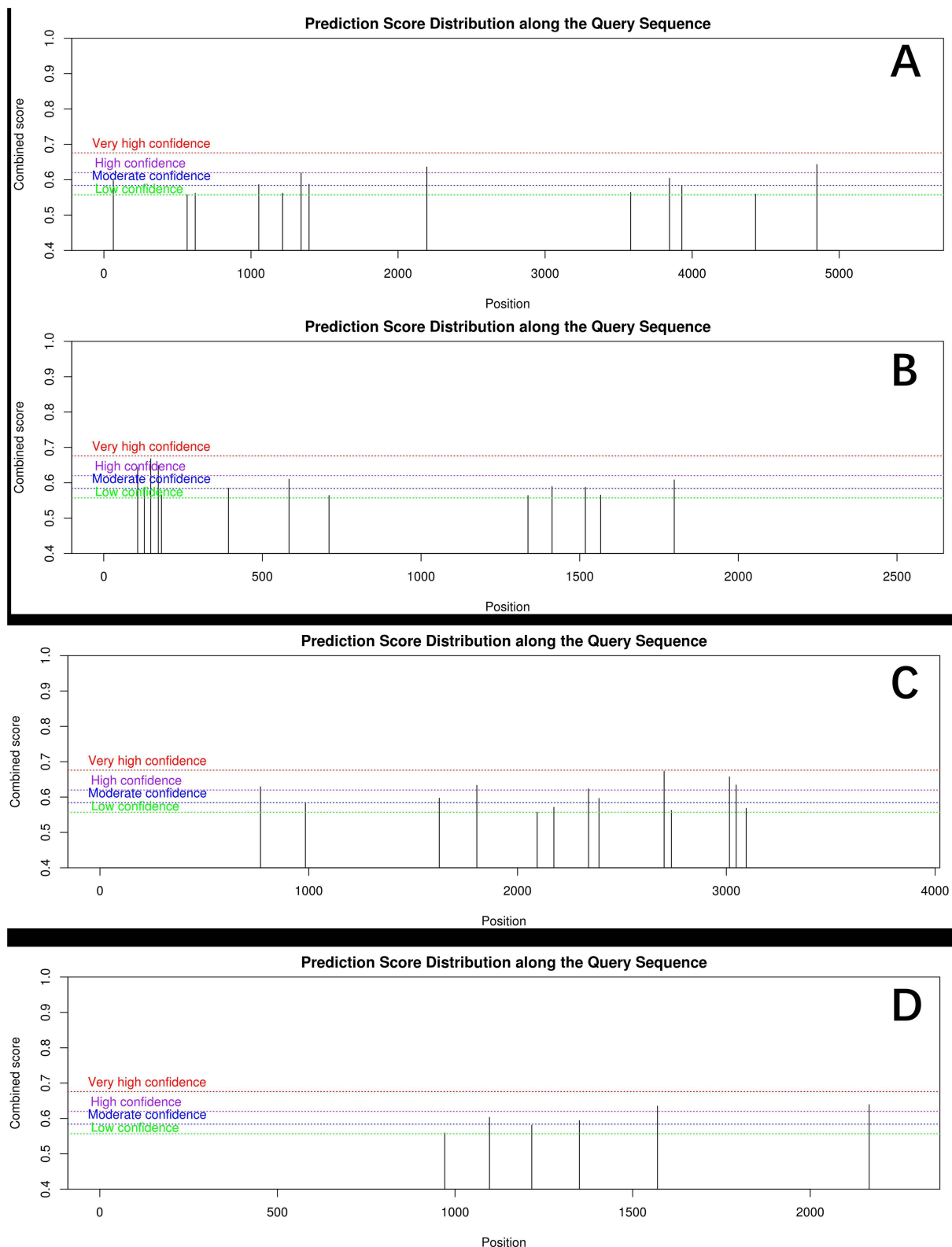
Figure 7 m6A binding site analysis of key genes. (**A**) m6A binding sites of COL3A1 gene. (**B**) m6A binding sites of OLR1 gene. (**C**) m6A binding sites of STC1 gene. (**D**) m6A binding sites of KCNJ8 gene.

Infiltration of Immune Cells in OA

In the GSE55235 dataset, these 28 immune cells were distributed differently (Figure 8A). Among them, the infiltration of eosinophils and type 2 T helper cells (Th2) were significantly reduced in OA. In contrast, 15 types of immune cells, including activated B cells, memory B cells and regulatory T cells, were significantly more infiltrated in OA (Figure 8B). The results of immune cell correlations revealed that immature dendritic cells and muscle-derived stem cells (MDSC) had the strongest positive correlation, with a correlation coefficient of 0.86 ($p < 0.001$). And the highest negative correlation was found between Th2 and immature B cells (cor = −0.79, $p < 0.001$) (Figure 8C). Furthermore, the results of the correlation between the biomarkers and immune cells showed that STC1 and COL3A1 were significantly positively correlated with eosinophils and Th2, respectively, while they were significantly negatively correlated with the other 13 immune cells. In contrast, OLR1 and KCNJ8 were significantly positively correlated with eosinophils and Th2, respectively, and significantly negatively correlated with the other immune cells (Figure 8D).

Construction of TF-miRNA-mRNA Regulatory Network and Drug-mRNA Network

A total of 40 miRNAs were obtained by intersecting the 128 miRNAs predicted in the miRTarBase database with the 300 miRNAs predicted in the starBase database. The ChEA3 database predicted 28 TFs corresponding to the biomarkers, with CBX2 and GATA3 jointly involved in regulating biomarker expression. Based on this, the TF-miRNA-mRNA molecular regulatory network of OA was constructed (Figure 9A). A total of 30 potential drugs for the treatment of OA were identified based on the screening criteria, and a drug-mRNA network was constructed based on these results (Figure 9B).
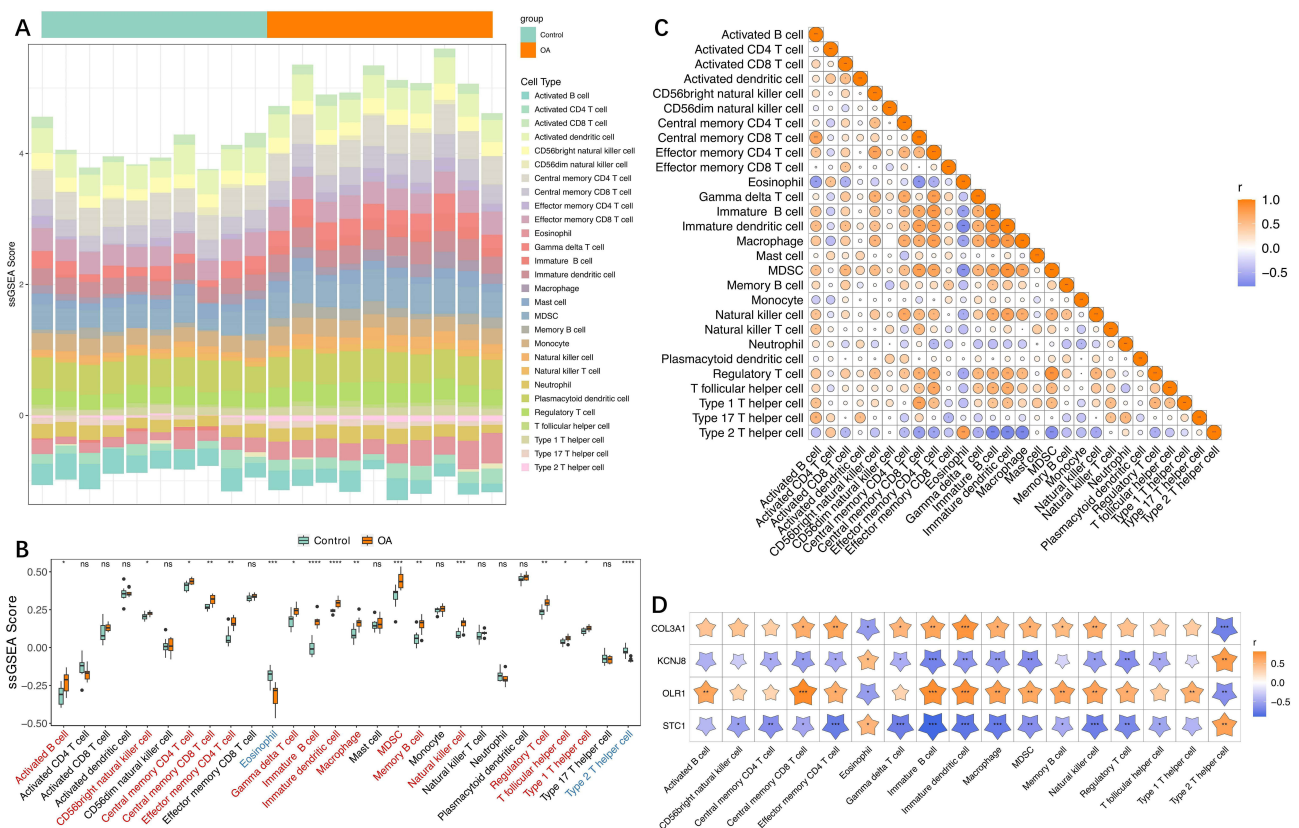
Figure 8 Analysis of immune cell infiltration and correlation with key genes. (**A**) Heatmap of immune cell infiltration. (**B**) Boxplot of 28 types of immune cell infiltration in Control group and OA group. "ns" represented no significantly. (**C**) Heatmap of immune cell correlation. Orange - red indicated positive correlation, and blue indicated negative correlation. The darker the color was, the higher the correlation was. (**D**) Heatmap of key gene and immune cell correlation. The abscissa represented immune cells, and the ordinate represented key genes. In the figure, the upward - pointing five - pointed star indicated positive correlation, and the downward - pointing one indicated negative correlation. The darker the color was, the higher the correlation was, * represented $P<0.05$, ** represented $P<0.01$, *** represented $P<0.001$.
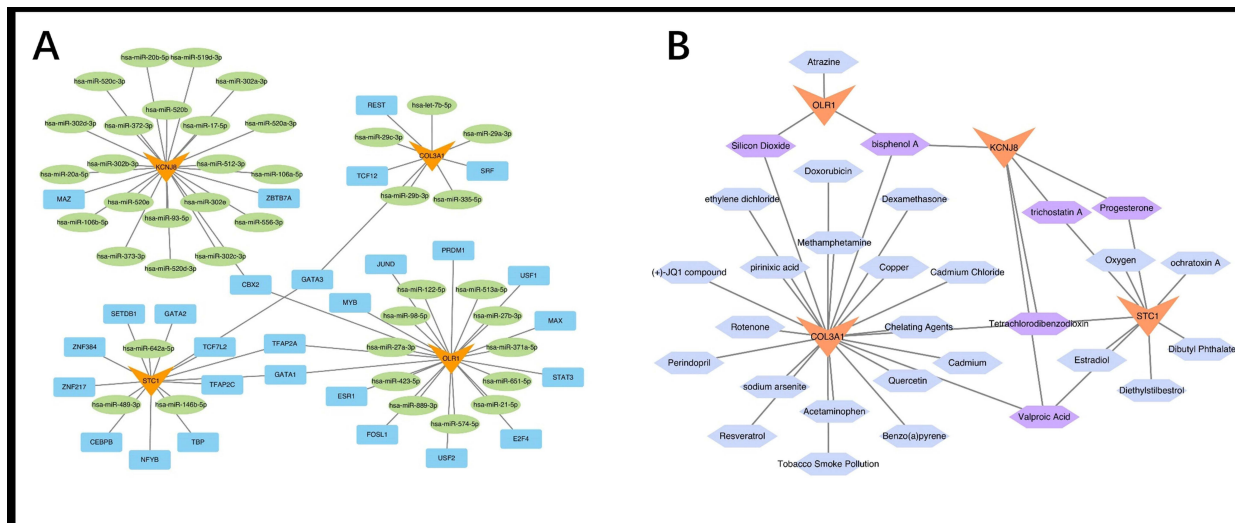
Figure 9 Regulatory networks involving key genes. (**A**) TF-miRNA-mRNA network. Orange triangles represented mRNA, green circles represented miRNA, and blue rectangles represented TF. (**B**) Therapeutic drug network diagram corresponding to key genes. Orange triangles were key genes, blue hexagons were therapeutic drugs, and purple hexagons were drugs that had interactions with two or more genes.

## Expression Evaluation of Biomarkers

The expression of biomarkers between OA and control was assessed by RT-qPCR. The results showed that the expressions of COL3A1, OLR1, and STC1 were consistent with the public database. The expression levels of COL3A1 and OLR1 were significantly up-regulated in the OA group, while STC1 was significantly down-regulated compared to the control group. However, no statistically significant difference was observed in KCNJ8 expression between the OA and control groups (Figure 10).

Discussion

In recent years, OA treatment has shifted from symptom management to early prevention, aiming to halt disease progression and prevent irreversible joint damage. Targeted gene therapy has emerged as a promising approach, highlighting the need to identify biomarkers and therapeutic targets to understand the underlying mechanisms of OA.[38,39]

In this study, bioinformatics analyses were used to identify angiogenesis-related biomarkers in OA, specifically examining four key genes: COL3A1, OLR1, STC1, and KCNJ8. These genes and their associated biological pathways, immune microenvironment characteristics, regulatory mechanisms, and links to related diseases and potential drugs were explored to provide insights into OA diagnosis and treatment. COL3A1 encodes type III collagen, a structural protein essential for tissues like ligaments, blood vessels, and cartilage.[40] OLR1's role in inflammatory pathways implicates it in promoting immune cell activation, likely increasing tissue inflammation.[41] STC1 encodes stanniocalcin-1, a protein involved in cell proliferation, apoptosis, and oxidative stress, with roles in inflammatory conditions.[41] KCNJ8 is involved in ion channel regulation, influencing cellular responses to mechanical stress and pain perception.[42] These genes likely impact OA by modulating key processes such as angiogenesis, inflammation, and cartilage structure maintenance, which collectively alter disease severity and progression.

Machine learning algorithms are increasingly applied in bioinformatics and genomics, particularly for identifying disease-related feature genes. By analyzing large biological datasets, these methods can uncover key biomarkers, predict disease progression, and assess patient responses. LASSO regression, with its sparsity-promoting penalty, is effective for feature selection and variable reduction, performing well even with small sample sizes.[43] It has been used to construct diagnostic models for osteoarthritis and metabolic syndrome.[44] SVM-RFE, leveraging support vector machines, handles non-linear relationships and considers gene interactions, making it ideal for high-dimensional gene expression data and effective in various disease gene screening.[45] Boruta, a random forest-based algorithm, iteratively determines feature importance, minimizing overfitting and false positives.[46] Combining LASSO, SVM-RFE, and Boruta enhances feature selection accuracy and model generalization, aiding in understanding disease mechanisms and identifying biomarkers.[47]
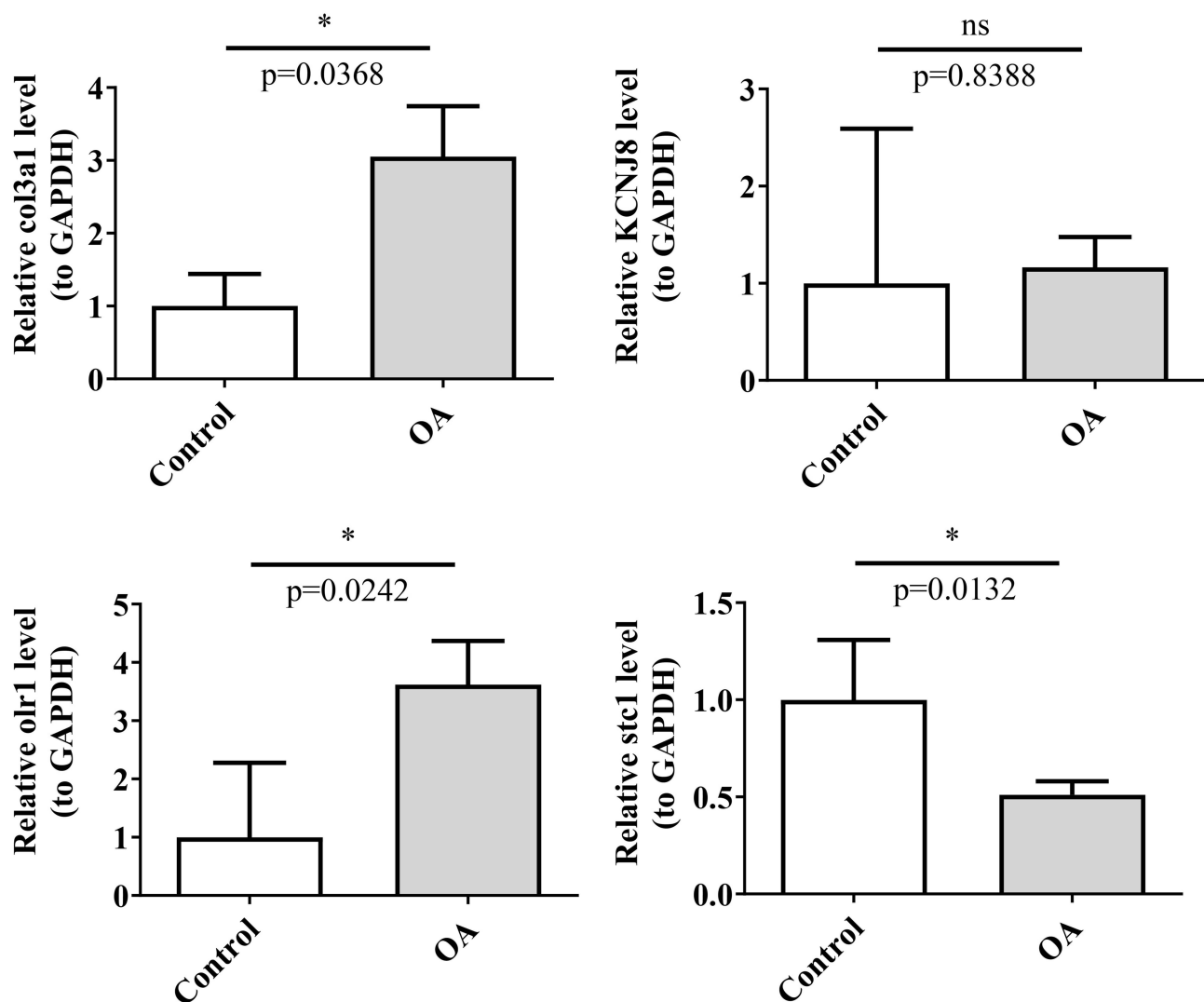
Figure 10 The figure of RT-qPCR verification of key genes. "ns" represented no significantly, * represented *P*<0.05.

This combined approach has been applied across various diseases, including glioblastoma, Alzheimer's disease, diabetic foot ulcers, and acute ischemic stroke, demonstrating its broad utility and offering valuable insights for future research.[48,49]

GO analysis revealed that DE-ARGs are significantly enriched in "cartilage development", "connective tissue development", and "bone morphogenesis". GSEA results identified "lysosomes" as a common pathway for the four biomarkers, primarily involved in cellular autophagy, chondrocyte apoptosis, and matrix degradation. Srinivas observed autophagic features in OA chondrocytes, suggesting that changes in autophagic activity may relate to cartilage degeneration.[50] Caramés found that as cartilage degeneration progresses, autophagic activity decreases, while apoptotic gene expression increases.[51] GeneMANIA analysis highlighted that COL3A1 shares biological functions with other collagen genes, reinforcing its significant role in OA development. Han et al identified hub genes, including COL6A3, COL1A2, COL1A1, and COL3A1, that are involved in immune responses, apoptosis, inflammation, and bone development.[52] COL3A1 has been linked to inflammation signaling pathways such as PI3K/AKT, NF-κB, and IL-17, as well as ECM-receptor interactions, contributing to OA inflammation.[53,54] GSEA analysis also suggests that OLR1 is closely related to inflammation and vascular abnormalities in OA. Research confirms that IL-1β-producing synovial cells in OA are primarily inflammatory macrophages and dendritic cells expressing OLR1.[55] STC1 is implicated in neovascularization in the OA synovium, and KCNJ8, involved in potassium ion regulation, may relate to neuronal excitability

and pain in OA.[56] Compared to studies by Han et al and Chou et al, our approach integrates machine learning to refine gene selection, increasing diagnostic precision and uncovering novel associations such as the lysosomal pathway in OA chondrocyte regulation, a finding not previously highlighted in OA-related angiogenesis studies.[52–54]

TFs and miRNAs play critical roles in regulating biomarkers and pathways linked to OA. This study identified GATA3 (GATA Binding Protein 3) and CBX2 (Chromobox 2) as key regulators of biomarker expression. GATA3, a transcription factor, influences genes involved in chondrocyte function and bone metabolism. Its reduced expression in OA cartilage suggests a protective role against cartilage degeneration, likely through regulation of COL3A1 expression, though the precise mechanisms require further investigation.[57] Additionally, miRNAs such as miR-29, miR-513a, and miR-155 were found to regulate COL3A1 and OLR1, impacting collagen synthesis and inflammatory signaling, thereby contributing to OA progression.[34]

This study revealed significant differences in immune cell infiltrates between OA and control groups, involving 17 types of immune cells. Eosinophil and Th2 cell infiltration were significantly reduced in OA, while infiltration of immune cells like activated B cells, memory B cells, regulatory T cells, CD56 bright NK cells, natural killer cells, neutrophils, and plasma-like dendritic cells increased, possibly associated with OA progression. Neutrophil infiltration is strongly associated with knee joint damage in OA.[58] Additionally, their interaction with NK cells through CXCL10/CXCR3 further promotes the progression of OA.[59] Plasma-like dendritic cells, crucial in immune responses, secrete high levels of inflammatory cytokines in OA, potentially contributing to its pathogenesis.[58,59] Correlation analysis showed that STC1 and COL3A1 were positively correlated with eosinophils and Th2 cells, while negatively correlated with 13 other immune cell types. In contrast, OLR1 and KCNJ8 showed the opposite pattern, indicating that these biomarkers play a role in regulating immune activity in OA.

The study also identified potential drugs targeting COL3A1, OLR1, STC1, and KCNJ8 based on their expression in OA. Bisphenol A (BPA), an endocrine disruptor, regulates COL3A1, OLR1, and KCNJ8 but may worsen OA by promoting cartilage destruction and inflammation.[60,61] Trichostatin A (TSA), an HDAC inhibitor, regulates STC1 and KCNJ8, showing potential to protect cartilage by reducing inflammation and promoting chondrocyte proliferation. Tetrachlorodibenzo-p-dioxin (TCDD) and valproic acid (VPA) regulate KCNJ8, COL3A1, and STC1; however, TCDD can impair joint health, while VPA, primarily used for neurological conditions, may influence bone cell proliferation and inflammation.[62,63] Further research is needed to validate the safety and efficacy of these drugs in OA treatment.

In conclusion, this study identified COL3A1, OLR1, STC1, and KCNJ8 as key angiogenesis-related biomarkers in OA, offering potential as diagnostic tools and therapeutic targets. Targeting these biomarkers could transform OA management by addressing its underlying mechanisms, enabling more effective treatments, improving patient quality of life, and reducing the healthcare burden associated with advanced OA. The findings emphasize the critical role of angiogenesis in OA progression and suggest novel pathways for therapeutic intervention.

This study has several limitations. The small sample size and the incomplete consideration of individual factors associated with OA progression may introduce bias. While gene expression analysis showed no batch effects, subtle biases cannot be entirely excluded. Additionally, the absence of experimental validation in animal or cellular models limits translational potential. Future research will address these issues by expanding the sample size, performing subgroup analyses to examine demographic and clinical variations, and prioritizing in vivo and in vitro validation to improve reliability and support the development of targeted OA therapies.

Ethics Approval and Informed Consent

This study was approved by the Ethics Committee of Zhejiang Provincial Hospital of Traditional Chinese Medicine (2022K25401). The ethics committee conducted a comprehensive review of the study design, implementation process, and the potential risks and benefits to participants, in accordance with internationally recognized ethical guidelines, including the Declaration of Helsinki, and approved the study. All participants have signed informed consent forms.

Acknowledgment

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

Disclosure

The author(s) report no conflicts of interest in this work.

References

1. Kolasinski SL, Neogi T, Hochberg MC, et al. 2019 American college of rheumatology/arthritis foundation guideline for the management of osteoarthritis of the hand, hip, and knee. *Arthritis Care Res*. 2020;72(2):149–162. doi:10.1002/acr.24131
2. Knights AJ, Redding SJ, Maerz T. Inflammation in osteoarthritis: the latest progress and ongoing challenges. *Curr Opin Rheumatol*. 2023;35 (2):128–134. doi:10.1097/BOR.0000000000000923
3. Li HZ, Han D, Ao RF, et al. Tanshinone IIA attenuates osteoarthritis via inhibiting aberrant angiogenesis in subchondral bone. *Arch Biochem Biophys*. 2024;753:109904. doi:10.1016/j.abb.2024.109904
4. Xu J, He SJ, Xia TT, et al. Targeting type H vessels in bone-related diseases. *J Cell Mol Med*. 2024;28(4):e18123. doi:10.1111/jcmm.18123
5. Cui Z, Wu H, Xiao Y, et al. Endothelial PDGF-BB/PDGFR-β signaling promotes osteoarthritis by enhancing angiogenesis-dependent abnormal subchondral bone formation. *Bone Res*. 2022;10(1):58. doi:10.1038/s41413-022-00229-6
6. Woetzel D, Huber R, Kupfer P, et al. Identification of rheumatoid arthritis and osteoarthritis patients by transcriptome-based rule set generation. *Arthritis Res Ther*. 2014;16(2):R84. doi:10.1186/ar4526
7. Ungethuem U, Haeupl T, Witt H, et al. Molecular signatures and new candidates to target the pathogenesis of rheumatoid arthritis. *Physiol Genomics*. 2010;42A(4):267–282. doi:10.1152/physiolgenomics.00004.2010
8. Huang F, Zhang C, Yang W, et al. Identification of a DNA damage repair-related LncRNA signature for predicting the prognosis and immunotherapy response of hepatocellular carcinoma. *BMC Genomics*. 2024;25(1):155. doi:10.1186/s12864-024-10055-1
9. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47. doi:10.1093/nar/gkv007
10. Steenwyk JL, Rokas A. ggpubfigs: colorblind-friendly color palettes and ggplot2 graphic system extensions for publication-quality scientific figures. *Microbiol Resour Announc*. 2021;10(44):e87121. doi:10.1128/MRA.00871-21
11. Shen G, Wang WL. Circlize package in R and analytic hierarchy process (AHP): contribution values of ABCDE and AGL6 genes in the context of floral organ development. *PLoS One*. 2022;17(1):e261232. doi:10.1371/journal.pone.0261232
12. Wu L, Wang Q, Gao QC, et al. Potential mechanisms and drug prediction of rheumatoid arthritis and primary sjogren's syndrome: a public databases-based study. *PLoS One*. 2024;19(2):e298447.
13. Huckvale E, Moseley HNB. kegg_pull: a software package for the RESTful access and pulling from the Kyoto Encyclopedia of Gene and Genomes. *BMC Bioinf*. 2023;24(1):78. doi:10.1186/s12859-023-05208-0
14. Li C, Lian Y, Lin Y, et al. A network pharmacology and molecular dynamics simulation-based study of Qing run Hua Jie decoction in interstitial pneumonia treatment. *Infect Drug Resist*. 2024;17:605–621. doi:10.2147/IDR.S433755
15. Mobeen SA, Saxena P, Jain AK, et al. Integrated bioinformatics approach to unwind key genes and pathways involved in colorectal cancer. *J Cancer Res Ther*. 2023;19(7):1766–1774. doi:10.4103/jcrt.jcrt_620_21
16. Lihua J, Haodan K, Yuan XU. Efficacy of Buzhong Yiqi decoction on benign prostatic hyperplasia and its possible mechanism. *J Tradit Chin Med*. 2023;43(3):533–541. doi:10.19852/j.cnki.jtcm.2023.03.003
17. Guo Z, Zhao Z, Wang X, et al. Identification of mitophagy-related hub genes during the progression of spinal cord injury by integrated multinomial bioinformatics analysis. *Biochem Biophys Rep*. 2024;38:101654. doi:10.1016/j.bbrep.2024.101654
18. Yang L, Qu Q, Hao Z, et al. Powerful Identification of large quantitative trait loci using genome-wide r/glmnet-based regression. *J Heredity*. 2022;113(4):472–478. doi:10.1093/jhered/esac006
19. Meyer D, Dimitriadou E, Hornik K, et al. _e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien_. R package version 1.7-12. 2022.
20. Sanz H, Valim C, Vegas E, et al. SVM-RFE: selection and visualization of the most relevant features through non-linear kernels. *BMC Bioinf*. 2018;19(1):432. doi:10.1186/s12859-018-2451-4
21. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinf*. 2011;12(1):77. doi:10.1186/1471-2105-12-77
22. Li S, Que Y, Yang R, et al. Construction of osteosarcoma diagnosis model by random forest and artificial neural network. *J Pers Med*. 2023;13 (3):447. doi:10.3390/jpm13030447
23. Robles-Jimenez LE, Aranda-Aguirre E, Castelan-Ortega OA, et al. Worldwide traceability of antibiotic residues from livestock in wastewater and soil: a systematic review. *Animals*. 2021;12(1). doi:10.3390/ani12010060
24. Zhou H, Yang Y, Shen HB. Hum-mPLoc 3.0: prediction enhancement of human protein subcellular localization through modeling the hidden correlations of gene ontology and functional domain features. *Bioinformatics*. 2017;33(6):843–853. doi:10.1093/bioinformatics/btw723

25. Gong HX, Wu B, Xie SY, et al. OCTA characteristics in non-arteritic central retinal artery occlusion and correlation with visual acuity. *Int J Ophthalmol*. 2024;17(2):289–296. doi:10.18240/ijo.2024.02.10

26. Zhang LM, Liang XL, Xiong GF, et al. Analysis and identification of oxidative stress-ferroptosis related biomarkers in ischemic stroke. *Sci Rep*. 2024;14(1):3803. doi:10.1038/s41598-024-54555-2

27. Zheng C, Guo H, Mo Y, et al. Integrating bioinformatics and drug sensitivity analyses to identify molecular characteristics associated with targeting necroptosis in breast cancer and their clinical prognostic significance. *Recent Pat Anticancer Drug Discov*. 2024;19:681–694.

28. Franz M, Rodriguez H, Lopes C, et al. GeneMANIA update 2018. *Nucleic Acids Res*. 2018;46(W1):W60–W64. doi:10.1093/nar/gky311

29. Shekhova E, Salazar F, Da SDA, et al. Age difference of patients with and without invasive aspergillosis: a systematic review and meta-analysis. *BMC Infect Dis*. 2024;24(1):220. doi:10.1186/s12879-024-09109-2

30. Zhou Y, Zeng P, Li YH, et al. SRAMP: prediction of mammalian N6-methyladenosineN 6 -methyladenosine (m6A)m 6 A) sites based on sequence-derived features. *Nucleic Acids Res*. 2016;44(10):e91. doi:10.1093/nar/gkw104

31. Hanzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinf*. 2013;14(1):7. doi:10.1186/1471-2105-14-7

32. Song Y, Feng T, Cao W, et al. Identification of key genes in nasopharyngeal carcinoma based on bioinformatics analysis. *Comput Intell Neurosci*. 2022;2022:9022700. doi:10.1155/2022/9022700

33. von Grothusen C, Frisendahl C, Modhukur V, et al. Uterine fluid microRNAs are dysregulated in women with recurrent implantation failure. *Hum Reprod*. 2022;37(4):734–746. doi:10.1093/humrep/deac019

34. Shen Y, Zhang Y, Wang Q, et al. MicroRNA-877-5p promotes osteoblast differentiation by targeting EIF4G2 expression. *J Orthop Surg Res*. 2024;19(1):134. doi:10.1186/s13018-023-04396-y

35. Zhou X, Song L, Cong R, et al. A comprehensive analysis on the relationship between BDE-209 exposure and erectile dysfunction. *Chemosphere*. 2022;308(Pt 3):136486. doi:10.1016/j.chemosphere.2022.136486

36. Huang W, Luo T, Lan M, et al. Identification and characterization of a ceRNA regulatory network involving LINC00482 and PRRC2B in peripheral blood mononuclear cells: implications for COPD pathogenesis and diagnosis. *Int J Chron Obstruct Pulmon Dis*. 2024;19:419–430. doi:10.2147/COPD.S437046

37. Shang Y, Zhang Y, Liu J, et al. Decreased E2F2 expression correlates with poor prognosis and immune infiltrates in patients with colorectal cancer. *J Cancer*. 2022;13(2):653–668. doi:10.7150/jca.61415

38. Hao HQ, Zhang JF, He QQ, et al. 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. doi:10.1016/j.joca.2018.10.009

39. Wu Z, Yuan K, Zhang Q, et al. Antioxidant PDA-PEG nanoparticles alleviate early osteoarthritis by inhibiting osteoclastogenesis and angiogenesis in subchondral bone. *J Nanobiotechnology*. 2022;20(1):479. doi:10.1186/s12951-022-01697-y

40. Ren J, Zhao S, Lai J. Role and mechanism of COL3A1 in regulating the growth, metastasis, and drug sensitivity in cisplatin-resistant non-small cell lung cancer cells. *Cancer Biol Ther*. 2024;25(1):2328382. doi:10.1080/15384047.2024.2328382

41. Zhang P, Zhao Y, Xia X, et al. Expression of OLR1 gene on tumor-associated macrophages of head and neck squamous cell carcinoma, and its correlation with clinical outcome. *Oncoimmunology*. 2023;12(1):2203073. doi:10.1080/2162402X.2023.2203073

42. Brownstein CA, Towne MC, Luquette LJ, et al. Mutation of KCNJ8 in a patient with Cantu syndrome with unique vascular abnormalities - support for the role of K(ATP) channels in this condition. *Eur J Med Genet*. 2013;56(12):678–682. doi:10.1016/j.ejmg.2013.09.009

43. Xi LJ, Guo ZY, Yang XK, et al. Application of LASSO and its extended method in variable selection of regression analysis. *Zhonghua Yu Fang Yi Xue Za Zhi*. 2023;57(1):107–111. doi:10.3760/cma.j.cn112150-20220117-00063

44. Jiang X, Zhong R, Dai W, et al. Exploring diagnostic biomarkers and comorbid pathogenesis for osteoarthritis and metabolic syndrome via bioinformatics approach. *Int J Gen Med*. 2021;14:6201–6213. doi:10.2147/IJGM.S325561

45. Lin X, Li C, Zhang Y, et al. Selecting feature subsets based on SVM-RFE and the overlapping ratio with applications in bioinformatics. *Molecules*. 2017;23(1):52. doi:10.3390/molecules23010052

46. Yang Z, He F. An immune cell infiltration landscape classification to predict prognosis and immunotherapy effect in oral squamous cell carcinoma. *Comput Methods Biomech Biomed Engin*. 2024;27(2):191–203. doi:10.1080/10255842.2023.2179364

47. Zeng L, Chen Z. Screening of genes characteristic of pancreatic cancer by LASSO regression combined with support vector machine and recursive feature elimination, and immune correlation analysis. *J Int Med Res*. 2024;52(3):3000605241233160. doi:10.1177/03000605241233160

48. Wu Y, Huang Y, Zhou C, et al. A novel necroptosis-related prognostic signature of glioblastoma based on transcriptomics analysis and single cell sequencing analysis. *Brain Sci*. 2022;12(8):988. doi:10.3390/brainsci12080988

49. Xu D, Qi P, Liu P, et al. Machine learning models reveal the critical role of nighttime systolic blood pressure in predicting functional outcome for acute ischemic stroke after endovascular thrombectomy. *Front Neurol*. 2024;15:1405668. doi:10.3389/fneur.2024.1405668

50. Srinivas V, Shapiro IM. Chondrocytes embedded in the epiphyseal growth plates of long bones undergo autophagy prior to the induction of osteogenesis. *Autophagy*. 2006;2(3):215–216. doi:10.4161/auto.2649

51. Carames B, Olmer M, Kiosses WB, et al. The relationship of autophagy defects to cartilage damage during joint aging in a mouse model. *Arthritis Rheumatol*. 2015;67(6):1568–1576. doi:10.1002/art.39073

52. Han Y, Wu J, Gong Z, et al. Identification and development of the novel 7-genes diagnostic signature by integrating multi cohorts based on osteoarthritis. *Hereditas*. 2022;159(1):10. doi:10.1186/s41065-022-00226-z

53. 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. doi:10.3389/fgene.2021.721258

54. Li J, Zheng J. Theaflavins prevent cartilage degeneration via AKT/FOXO3 signaling in vitro. *Mol Med Rep*. 2019;19(2):821–830. doi:10.3892/mmr.2018.9745

55. Chou CH, Jain V, Gibson J, et al. Synovial cell cross-talk with cartilage plays a major role in the pathogenesis of osteoarthritis. *Sci Rep*. 2020;10(1):10868. doi:10.1038/s41598-020-67730-y

56. Lambert C, Dubuc JE, Montell E, et al. Gene expression pattern of cells from inflamed and normal areas of osteoarthritis synovial membrane. *Arthritis Rheumatol*. 2014;66(4):960–968. doi:10.1002/art.38315

57. Singh P, Yadav US, Azad K, et al. NFIA and GATA3 are crucial regulators of embryonic articular cartilage differentiation. *Development*. 2018;145 (2). doi:10.1242/dev.156554

58. Zhang Y, Wang C, Xia Q, et al. Machine learning-based prediction of candidate gene biomarkers correlated with immune infiltration in patients with idiopathic pulmonary fibrosis. *Front Med Lausanne*. 2023;10:1001813. doi:10.3389/fmed.2023.1001813

59. Hou S, Wang D, Yuan X, et al. Identification of biomarkers co-associated with M1 macrophages, ferroptosis and cuproptosis in alcoholic hepatitis by bioinformatics and experimental verification. *Front Immunol*. 2023;14:1146693. doi:10.3389/fimmu.2023.1146693

60. Wang KC, Lin YF, Qin CH, et al. Bisphenol-A interferes with estradiol-mediated protection in osteoarthritic chondrocytes. *Toxicol Lett*. 2010;198 (2):127–133. doi:10.1016/j.toxlet.2010.06.007

61. Mei L, Zhang Z, Chen R, et al. Identification of candidate genes and chemicals associated with osteoarthritis by transcriptome-wide association study and chemical-gene interaction analysis. *Arthritis Res Ther*. 2023;25(1):179. doi:10.1186/s13075-023-03164-x

62. Han X, Bai F, Li P, et al. Identification of novel potential drugs for the treatment and prevention of osteoarthritis. *Biochem Biophys Rep*. 2024;37:101647. doi:10.1016/j.bbrep.2024.101647

63. Woolley B, Mills J. Versatile valproic acid. *Issues Ment Health Nurs*. 2022;43(11):1072–1074. doi:10.1080/01612840.2022.2122431