



## OPEN ACCESS

## EDITED BY

Ana Belén Arroyo,  
Santa Lucía University General Hospital, Spain

## REVIEWED BY

Asmaa ShamsEldeen,  
Cairo University, Egypt  
Maryam Azlan,  
University of Science Malaysia (USM), Malaysia

## \*CORRESPONDENCE

Chen Chen

✉ chenchen@sxhqh.com.cn

Hua Wu

✉ wuhua@hust.edu.cn

†These authors have contributed  
equally to this work and share  
first authorship

RECEIVED 20 March 2025

ACCEPTED 01 August 2025

PUBLISHED 20 August 2025

## CITATION

Li T, Wei J, Wu H and Chen C (2025) Analysis  
and validation of diagnostic biomarkers and  
immune cell infiltration characteristics in  
osteoarthritis by integrating bioinformatics  
and machine learning.  
*Front. Immunol.* 16:1596912.  
doi: 10.3389/fimmu.2025.1596912

## COPYRIGHT

© 2025 Li, Wei, Wu and Chen. This is an open-  
access article distributed under the terms of  
the [Creative Commons Attribution License](#)  
(CC BY). The use, distribution or reproduction  
in other forums is permitted, provided the  
original author(s) and the copyright owner(s)  
are credited and that the original publication  
in this journal is cited, in accordance with  
accepted academic practice. No use,  
distribution or reproduction is permitted  
which does not comply with these terms.

# Analysis and validation of diagnostic biomarkers and immune cell infiltration characteristics in osteoarthritis by integrating bioinformatics and machine learning

Tianyang Li<sup>†</sup>, Jinpeng Wei<sup>†</sup>, Hua Wu<sup>\*</sup> and Chen Chen<sup>\*</sup>

Department of Orthopedics, Third Hospital of Shanxi Medical University, Shanxi Bethune Hospital,  
Shanxi Academy of Medical Sciences, Tongji Shanxi Hospital, Taiyuan, China

**Objective:** Exosomes as important carriers of intercellular communication have frequently appeared in recent studies related to osteoarthritis (OA), while the specific mechanism of exosome action in osteoarthritis remains unclear. The aim of this study was to identify potential exosome-related biomarkers in osteoarthritis, to explore the role and mechanism of exosome-related genes in articular cartilage.

**Methods:** The data on exosome related genes and normal and OA cartilage genes were obtained through online databases. The potential mechanisms of these genes were revealed by multiple gene enrichment analysis algorithms. Machine learning methods were utilized to identify exosome-related differential genes (ERDEGs) with highly correlated OA features (Hub OA-ERDEGs). In addition, we created a nomogram to assess the ability of Hub OA-ERDEGs to diagnose OA. Single-sample gene set enrichment analysis (ssGSEA) was used to observe the infiltration characteristics of immune cells in OA and their relationship with Hub OA-ERDEGs.

**Results:** The results of screening Hub OA-ERDEGs using machine learning algorithms show that: TOLLIP, ALB, HP, RHOBTB3, GSTM2, S100A8 and AKR1B1 were significantly up-regulated or down-regulated in OA samples and verified by qRT-PCR for validation. Using the ssGSEA algorithm, we discovered that 8 types of immune cell infiltration and 5 types of immune cell activation.

## KEYWORDS

osteoarthritis, exosome, machine learning, immune infiltration, biomarkers

## Introduction

Osteoarthritis (OA) is one of the most common musculoskeletal diseases, and the global prevalence of OA is increasing year by year, and the age of onset of the disease is showing a trend of younger people, according to statistics, more than 22% of adults over the age of 40 years suffer from knee OA (1). OA is characterized by degeneration of articular cartilage, accompanied by subchondral bone remodeling, synovial inflammation, and osteophyte formation, leading to joint pain, stiffness, and limited motion (2, 3). Despite the exciting advances we have made in the pathophysiology of OA, various challenges remain in its diagnosis and early prevention. Current diagnostic methods rely heavily on clinical functional assessment and imaging techniques, while these methods do not allow for the early diagnosis of OA and may also lead to diagnostic discrepancies due to subjective judgments (4). Therefore, a pressing need exists for innovative diagnostic markers and diagnostic strategies, which are important for the prevention and disease management of OA.

Exosomes are small extracellular vesicles, usually between 30 and 150 nanometers in diameter, secreted extracellularly by a variety of cell types. They facilitate intercellular communication by transferring proteins, lipids, and RNA between cells (5, 6). Exosomes from different cellular sources may act on articular chondrocytes through multiple pathways; exosomes from mesenchymal stem cells (MSCs) can activate the Nrf2 signaling pathway through TGF $\beta$ 1, a key gene, to promote cartilage regeneration and inhibit the formation of neutrophil extracellular traps (NETs) (7). Other studies have shown that exosomes can activate the PI3K-AKT-mTOR pathway, promote macrophage polarization towards anti-inflammatory M2-type macrophages, and enhance the antioxidant capacity and survival of chondrocytes (8). Meanwhile, microRNA-26b-5p in M2 macrophage-derived exosomes could further protect articular cartilage by targeting Toll-like receptor 3 (TLR3) and Collagen type X alpha 1 (COL10A1) (9). In contrast, endocytosis of exosomes derived from inflammatory fibroblast-like synoviocytes (FLS) promotes macrophage polarization towards pro-inflammatory M1 macrophages, inducing synovitis and exacerbating the progression of an OA model after injection of exosomes into the joint cavity of mice (10). In conclusion, exosomes of various cellular origins play important roles in influencing inflammation genesis, chondrocyte activity, and cartilage degradation, and highlight the importance of exosome research in elucidating the complex cellular interactions involved in the pathogenesis of OA and provide promising avenues for identifying novel diagnostic and therapeutic targets.

Incorporating bioinformatics and machine learning into the study of OA and exosomes can significantly enhance our understanding of this complex disease. Machine learning techniques are able to analyze large and complex data sets generated from genomic, proteomic, and metabolomic studies, allowing researchers to uncover complex patterns and relationships that may have been overlooked by traditional analytical approaches (11). In addition, machine learning can facilitate disease modeling, allowing for personalized treatment

plans based on individual patient characteristics (12). In this study, machine learning algorithms were used to screen characteristic genes of OA, and single-sample gene set enrichment analysis (ssGSEA) algorithm was used to study the immune infiltration of OA and its relationship with hub genes. It provides guidance for advancing OA research, guiding clinical applications, and ultimately improving patient survival quality through targeted therapy.

## Methods

### Data preprocessing

The gene expression profile microarrays (GSE113825, GSE117999, GSE169077, GSE178557) of normal and OA cartilage samples were downloaded from the Gene Expression Omnibus database (GEO, <https://www.ncbi.nlm.nih.gov/geo/>). The datasets were first transformed into gene symbols, followed by background correction and normalization of each dataset in R software using the “limma” package, and integration of the two cartilage datasets using the “sva” package to eliminate batch effects (13). A two-dimensional PCA clustering plot was used to visualize the differences before and after removing the batch effect from the sample. Genes associated with exosomes were obtained from GeneCards Version 5.23 (<https://www.genecards.org/>), and we searched for exosome related genes (ERGs) and selected genes with correlations greater than 2 to be used for subsequent analysis (14).

Expression matrices of all genes in normal and OA cartilage samples were extracted from the training set and differences were analyzed using the limma package, using a threshold  $|\log FC| > 0.25$ ;  $P$ -value  $< 0.05$  to obtain differential genes (DEGs) (15). Subsequently, the “ggplot2” package in R Version 4.2.3 was used to visualize the differential genes. The “ggvenn” package was used to intersect the DEGs and ERGs in the training set for subsequent analysis, and the genes obtained were named exosome-related differential genes (ERDEGs).

### Constructing protein interaction networks

To assess gene interactions among DEGs, a protein-protein interaction (PPI) network was constructed using the Interacting Genes Search Tool (STRING Version:12.0, <https://cn.string-db.org/>) database, with a confidence score of  $>0.7$  as a cut-off criterion (16). Subsequently, the PPI network was visualized using Cytoscape software (version 3.8.2).

### Functional enrichment analysis

Gene Ontology (GO) is an organized knowledge base designed to provide standardized descriptions of genes and gene products. It mainly covers the annotations of Molecular Function (MF),

Biological Process (BP), Cellular Component (CC) (17). The Kyoto Encyclopedia of Genes and Genomes (KEGG) integrates all kinds of biological information, especially data related to genomics and systems biology. The main feature is visualization using pathway maps to make complex biological processes easier to understand (18). Disease Ontology (DO) is a knowledge base for systematically describing and categorizing a variety of diseases and their associated characteristics that can enhance the effectiveness of disease research (19). To further understand the biological functions, signaling pathway enrichment and disease similarities of DEGs, we analyzed DEGs using the “clusterprofiler” package.

## Gene set enrichment analysis

Gene Set Enrichment Analysis (GSEA) is a bioinformatics method that is commonly used to analyze transcriptomic data to assess whether changes in gene set expression levels under different conditions are significant or not (20). This study utilized the GSEA package (v1.66.0) to conduct gene set enrichment analysis. The `c2.cp.kegg_legacy.v2025.1.Hs.symbols.gmt` and `c5.c5.go.v2025.1.Hs.symbols.gmt` gene sets from the Molecular Signatures Database (MSigDB), along with the immune infiltration cell signature gene set defined by Yu et al (21), were applied to generate signature enrichment scores. To assess enrichment in OA cartilage samples, we used the “clusterprofiler” package to score GO and KEGG pathways and screened for the five pathways with the most significantly upregulated expression levels in OA.

## Machine learning algorithms screening hub characteristic genes

To further identify the characteristic genes, we used linear and nonlinear machine learning algorithms for the analysis. Linear models include Single-factor logistic regression (SFLR), which eliminates factors in the variables that are considered insignificant through the “glmnet” package, and the acquired gene expression data are used for the subsequent three nonlinear machine learning analyses (22). Least absolute shrinkage and selection operator (LASSO) logistic regression imposes L1 penalties on the regression coefficients by means of the “glmnet” package, which automatically selects significant features associated with the target variable (23). Support vector machine-recursive feature elimination (SVM-RFE) reduces the number of features efficiently by using the SVM function of the “e1071” package, reducing the computational cost and improving the interpretability of the model (24). Random Forest (RF) automatically evaluates the importance of features through the “random Forest” package, helping to identify the features that have the greatest impact on the target variable, facilitating feature selection and data analysis (25). Finally, using the “ggvenn” package, the intersection of the filtered feature genes of the three nonlinear machine learning models is defined as Hub OA-ERDEGs. The expression of Hub OA-ERDEGs in the training

set and their location on the chromosome were subsequently visualized using the “ggpubr” and “circlize” package.

## Mouse chondrocyte culture and processing

Primary chondrocytes were isolated from the knee cartilage of four 5-day-old C57BL/6 male mice as follows: mice were disinfected by immersion in 75% alcohol. Under aseptic conditions, cartilaginous tissues of the knee joints of mice were dissected and isolated, and digested with 0.25% trypsin (Boster, China) at 37 degrees for 30 minutes. After centrifugation to remove trypsin, 0.2% collagenase type II (BioFroxx, Germany) was added and incubation was continued at 37 degrees for 4–6 hours. After centrifugation, the cells were resuspended in DMEM/F12 (HyClone, USA) medium containing 10% fetal bovine serum (FBS, Newzerum, New Zealand) and incubated at 5% CO<sub>2</sub> and 37°C in an incubator. When the fusion rate reached 80–90%, the cells were digested with 0.25% trypsin and collected, and then transferred to 6-well plates at an appropriate density. After the chondrocytes were attached to the wall, the cells in the OA group were treated with 10 ng/mL concentration of IL-1 $\beta$  (R&D Systems, USA) for 24 hours, and the control group was only changed the culture medium. If not specified, first or second-generation chondrocytes were used for subsequent experiments. Animal studies are reported according to ARRIVE guidelines 2.0. Male C57BL/6 mice, 5-day-old, weighing approximately 3.5 g each, were purchased from Hubei Bainte Biotechnology Co., Ltd. All animal experiments were performed in accordance with a protocol approved by the Institutional Animal Care and Use Committee of Huazhong University of Science and Technology ([2024] IACUC Number: 4741), and strictly followed the recommendations of the “Laboratory Animals of the National Institutes of Health”.

## Quantitative reverse transcription-PCR

Total RNA was extracted using the EZNA Total RNA kit (Omega Bio-Tek, USA) according to manufacturer’s instructions. Then we used 2  $\mu$ g RNA to synthesize first strand cDNA with Reverse Transcription kit (Toyobo Life Science, Japan). PCR amplification was performed with SYBR Green Real-Time PCR Master Mix (Toyobo Life Science, Japan) on CFX96 (Bio-Rad Laboratories, USA) system. The relative gene expression level was calculated by comparative Ct method and normalized to internal control  $\beta$ -Actin. All PCR reactions were repeated in triplicates. Primer sequences for Hub OA-ERDEGs are shown in the [Supplementary Table S13](#).

## Construction and evaluation of a nomogram

The nomogram will combine multiple variables of a predictive model in a simple visual way, providing an easy to understand and

use method for assessing the risk or outcome of a patient-specific event. Therefore, we utilized the “rms” package to create a multivariate logistic regression-based outcome to generate a decision curve analysis (DCA) to assess the clinical utility of the nomogram (26). The “pROC” package was used to generate ROC curves to verify the reliability of the model,  $AUC > 0.75$  indicates that this model has good performance (27).

## Immunological features of osteoarthritis

Single-sample gene set enrichment analysis (ssGSEA) is a method used to assess the enrichment of specific gene sets in a single sample. Unlike traditional gene set enrichment analysis, ssGSEA focuses on a single sample and is able to reveal biological features and pathway activities in the sample in greater detail (28). We calculated enrichment scores for normal and OA cartilage samples in immune cells using the “GSVA” and “GSEABase” packages, and used the “Ggpubr” package to plot box line plots, and the “reshape2” and “tidyverse” packages to plot heat maps of differential gene and immune cell correlations. Spearman correlation analysis was used to correlate Hub OA-ERDEGs with immune cells.

## Animal model

A destabilization of the medial meniscus (DMM) model was established in C57BL/6 mice aged six weeks. Mice were anesthetized with sodium pentobarbital before surgery. Following anesthesia, the knee joint area was prepared by shaving, disinfecting, and making an incision from the distal patellar to the proximal tibial region. The joint capsule was then opened, the patellar tendon retracted, and the medial meniscal ligament was exposed and transected. The Sham group underwent a similar incision procedure without transection of the medial meniscotibial ligament. Postoperative management included prophylactic penicillin administration and unrestricted ambulation in climate-controlled housing with ad libitum access to feed and water. At 12-week endpoint, bilateral joint specimens were harvested following humane euthanasia of three model group and three sham group animals.

## Histological evaluation

Knee tissues underwent post-fixation with 4% paraformaldehyde (Sigma-Aldrich) for 48-hour immobilization, followed by decalcification in 10% EDTA solution (Sigma-Aldrich). After complete demineralization, specimens were processed through paraffin embedding and sectioned sagittally at 6  $\mu\text{m}$  thickness. Tissue sections underwent deparaffinization in xylene baths with subsequent graded ethanol rehydration. Histomorphological analysis was performed using H&E staining according to standardized protocols (Solarbio). For protein localization studies, immunohistochemical detection was implemented with specific antibodies targeting INOS (Abcam) and ARG1 (Abcam).

## Results

### Identification and PPI analysis of ERDEGs in osteoarthritis

The flow chart of this study is shown in Figure 1. The calibrated dataset information is shown in Supplementary Table S1. The PCA results showed a significant difference between the samples from patients with OA and controls, and the batch effect was significantly removed, suggesting that the expression matrix was suitable for subsequent analyses (Figure 2A). 861 exosome-associated genes with correlations greater than 2 were obtained from the GeneCards database for subsequent analysis. (Supplementary Table S2). A total of 1813 DEGs in the dataset of OA-related genes were shown in the volcano map (Figure 2B). We subsequently intersected the DEGs with the ERGs by taking the intersection of the DEGs, and found 105 ERDEGs using the “VennDiagram” package, of which 73 genes were up-regulated and 32 genes were down-regulated in OA, and the Supplementary Table S3 contains a detailed list of these related genes. PPI protein network Interaction analysis showed that ERDEGs interacted closely at the protein level (Figure 2C). The results of the PPI protein network interaction analysis are shown in Supplementary Table S4.

### Functional enrichment analysis of ERDEGs

In order to better understand the potential mechanisms of ERDEGs in OA, we performed GO, KEGG, and DO enrichment analyses of ERDEGs using the “clusterProfiler” package. GO enrichment analyses showed that the first five ERDEGs enrichments were mainly involved in regulation of peptidase activity, ameboidal-type cell migration, viral process, viral life cycle and regulation of endopeptidase activity. The top five ERDEGs enriched in cellular components and molecular functions are shown in Figure 3A. In addition, KEGG pathway analysis showed that these ERDEGs were enriched in the fluid shear stress and atherosclerosis, salmonella infection, regulation of actin cytoskeleton, human cytomegalovirus infection and pathogenic Escherichia coli infection, and that these pathways interacted closely with each other (Figure 3B). DO enrichment analysis revealed disease types with similar pathogenic mechanisms to ERDEGs in OA, such as atherosclerosis, arteriosclerotic cardiovascular disease, arteriosclerosis, breast carcinoma and familial hyperlipidemia (Figure 3C). Supplementary Tables S5–7 show the detailed results of GO, KEGG and DO enrichment of ERDEGs.

### GSEA enrichment analysis

We investigated the enrichment of GO and KEGG pathways in OA by GSEA method and showed the five gene functions and pathways that were most significantly up-regulated in OA, respectively. GO enrichment results showed that cell chemotaxis,



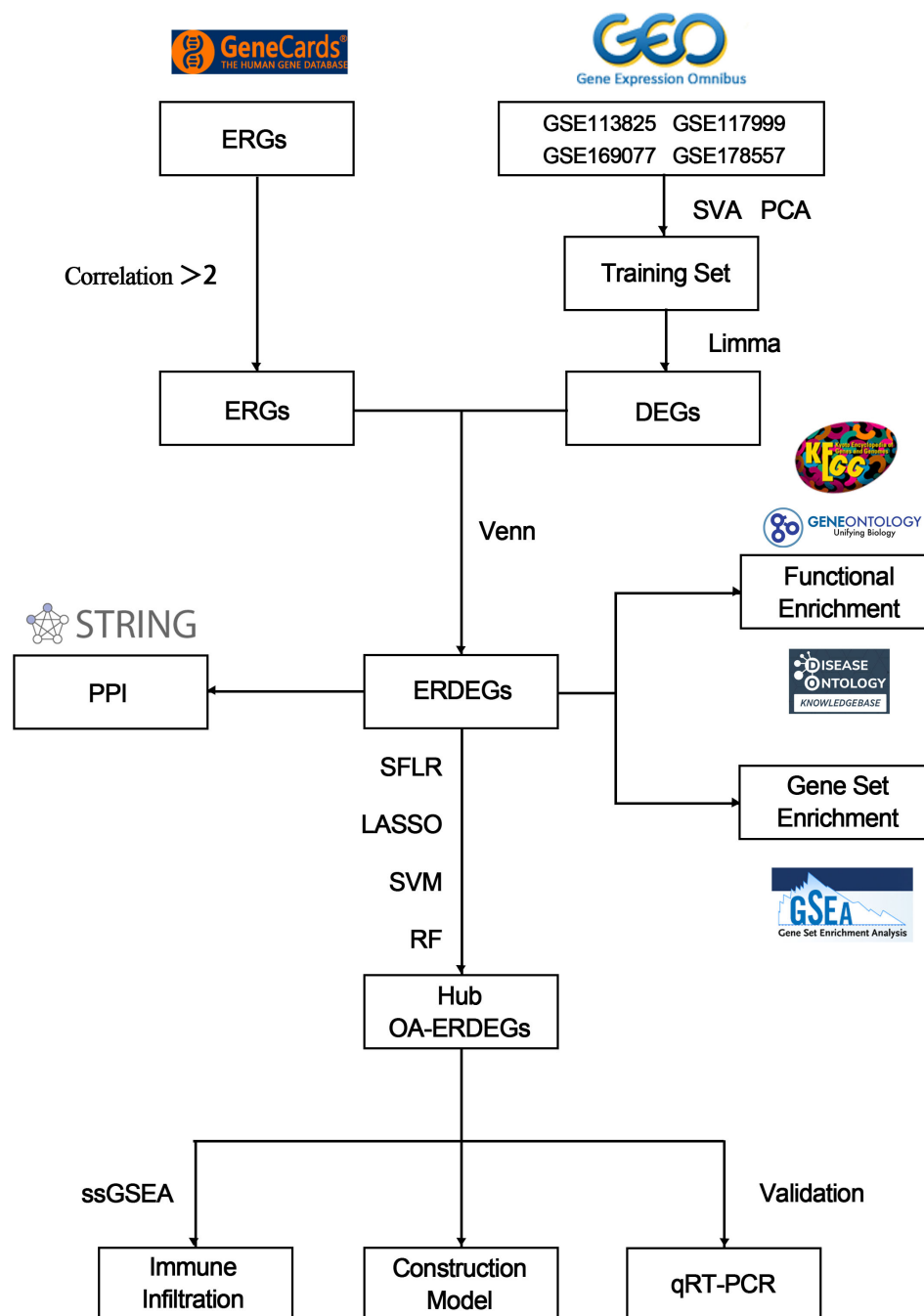
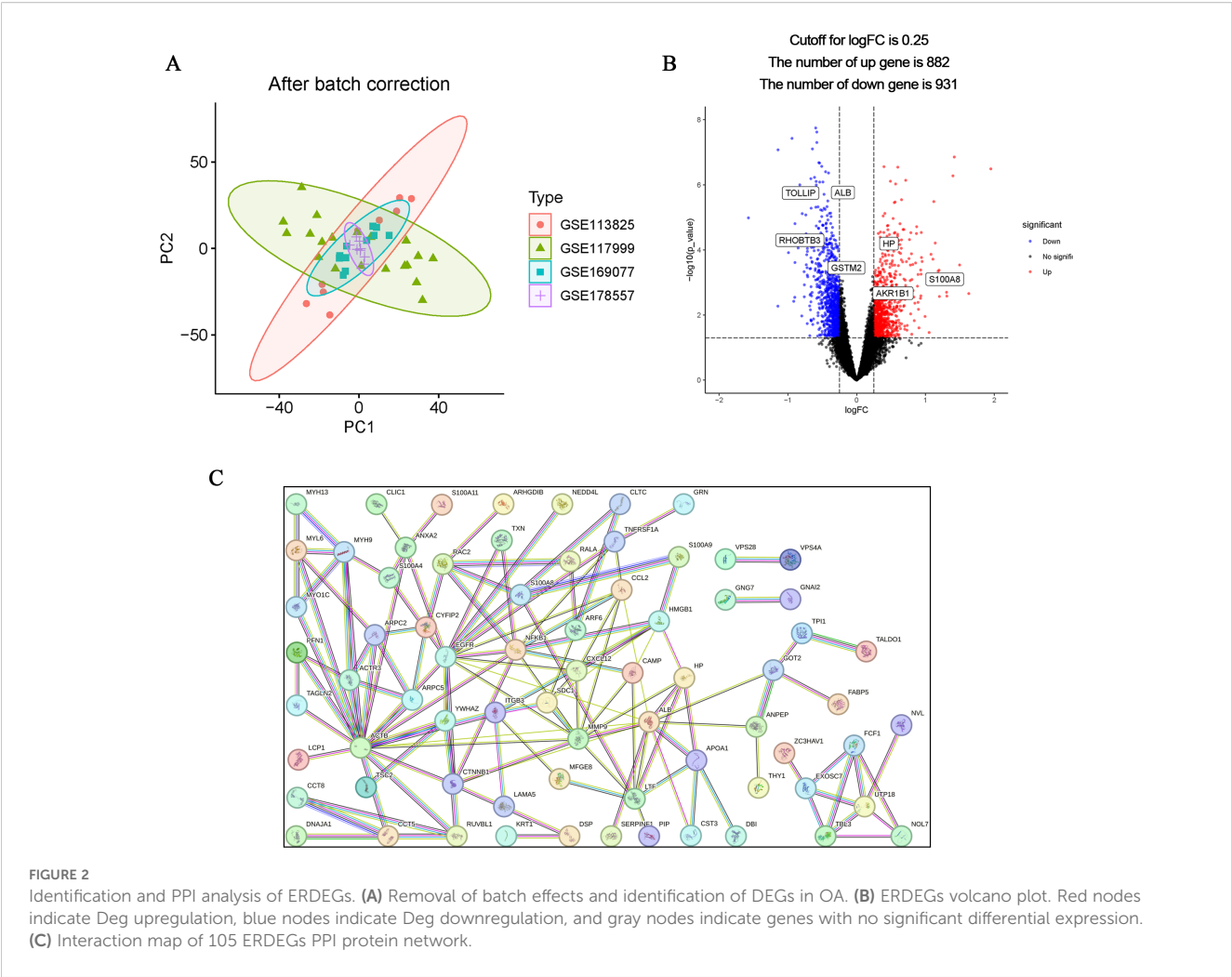


FIGURE 1  
Flowchart for comprehensive analysis of exosome-related genes in OA.

external encapsulating structure organization, collagen containing extracellular matrix, external encapsulating structure and extracellular matrix structural constituent were significantly upregulated (Figure 3D). KEGG pathway enrichment results showed that cytokine receptor interaction, ecm receptor interaction, focal adhesion, hematopoietic cell lineage and prion diseases were significantly upregulated (Figure 3E). Detailed results of the GSEA are shown in Supplementary Tables S8, 9.

## Identification and validation of Hub OA-ERDEGs

To further understand the role of 105 ERDEGs in the diagnosis and prognosis of OA, we aimed to identify hub genes from these 105 ERDEGs to construct a diagnostic prediction model. We used three nonlinear machine learning algorithms, lasso (Figure 4A), SVM-RFE (Figures 4C, D), and Random Forest (Figures 4E, F) to



screen OA-ERDEGs. The results of the three machine learning algorithms for recognizing Hub OA-ERDEGs are shown in [Supplementary Table S10](#). Finally, by combining the results of the three algorithms, a total of nine Hub OA-ERDEGs were obtained, namely, TOLLIP, ALB, HP, RHOBTB3, GSTM2, S100A8 and AKR1B1 ([Figure 4B](#)). TOLLIP, ALB, RHOBTB3, and GSTM2 were down-regulated in OA samples and HP, GSTM2, and S100A8 were up-regulated in OA samples ([Figure 5A](#)). The location of each gene in the chromosome is shown in [Figure 5B](#).

qRT-PCR

We performed qRT-PCR assays to further validate the mRNA expression levels of Hub OA-ERDEGs. The results showed that four genes were expressed meaningfully, of which AKR1B1, S100A8 and HP were significantly upregulated in OA cartilage samples, and RHOBTB3 was significantly downregulated in cartilage samples (p-value less than 0.05), which was consistent with the expression in the training set ([Figures 5C–F](#)).

Construction of Hub OA-ERDEGs risk prediction model

We constructed a nomogram of OA diagnosis based on the expression of Hub OA-ERDEGs to obtain a more applicable OA diagnosis model. By constructing the clinical calibration curves ([Figure 6A](#)) and clinical decision curves ([Figure 6B](#)) of the model, it is obvious that the model has a high predictive ability for OA. The score of each gene expressed in the column line graph accurately predicted the risk of osteoarthritis disease ([Figure 6C](#)).

Diagnostic value of Hub OA-ARDEGs

ROC curve analysis showed that 7 Hub OAARDEGs and column line graphs had high diagnostic value for OA. ALB and column line graphs had the highest diagnostic value (AUC=0.866), and the diagnostic values of other genes were: TOLLIP (AUC=0.851), GSTM2 (AUC=0.816), and HP (AUC=0.814), RHOBTB3 (AUC=0.807), S100A8 (AUC=0.784) and AKR1B1

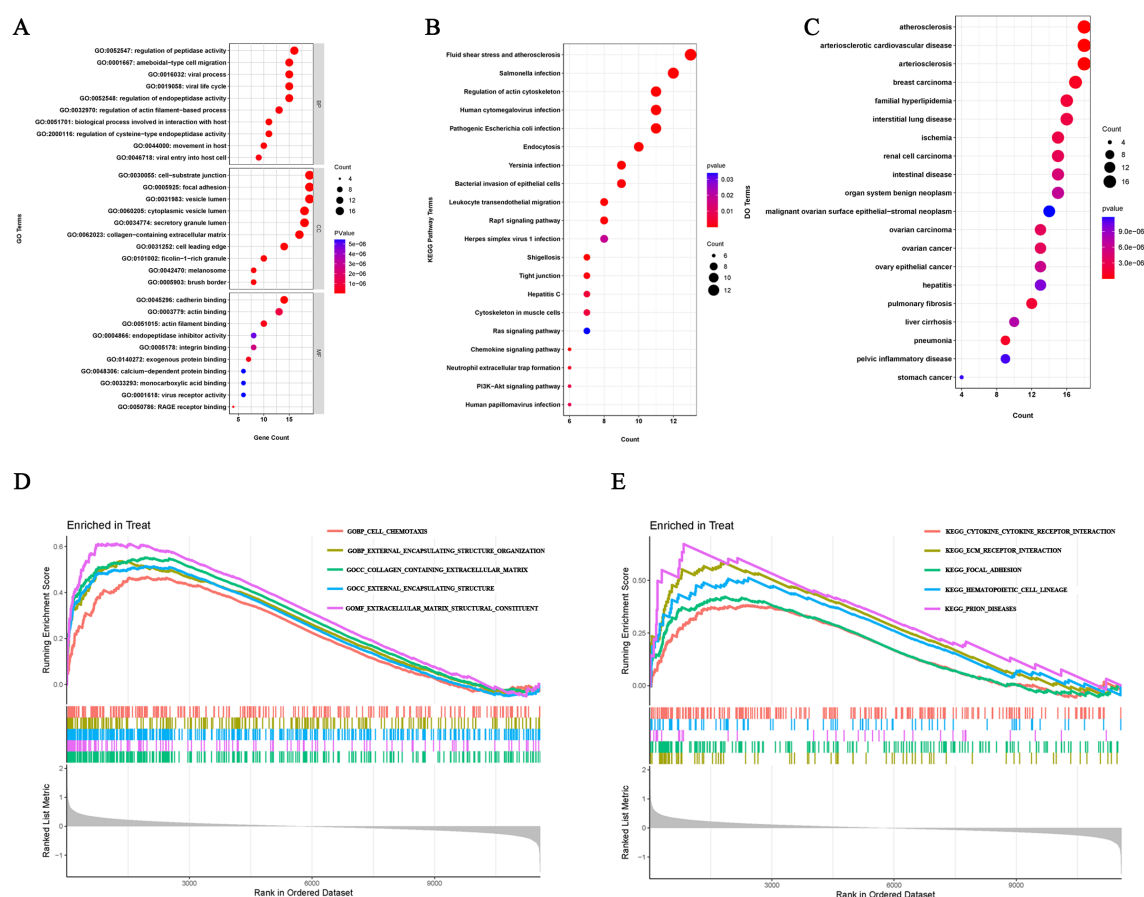


FIGURE 3

ERDEGs functional enrichment analysis. (A) GO enrichment analysis with BP, CC, and MF included. The bubble plots depict the ten most significantly enriched functions, where the size of the bubbles represents the number of DEGs (the larger the circle, the greater the number of DEGs) and the color represents the corrected p-value (the redder the color, the smaller the corrected p-value). (B) Analysis of KEGG enrichment, with bubble plots displaying the top 20 most significant pathway enrichments. (C) Analysis of DO enrichment, with bubble plots displaying the top 20 most similar diseases. (D) Gene set enrichment analysis of GO showing the top five most significant biological processes. (E) Gene set enrichment analysis of KEGG, showing the top five most significant pathways.

(AUC=0.761) (Figure 6D). Therefore, the seven Hub OA-ERDEGs can be used as reliable biomarkers with high diagnostic value for the diagnosis of OA.

## Immune infiltration analysis

We used the ssGSEA algorithm to find myeloid-derived suppressor cells, T follicular helper cell, Macrophage, central memory CD8 T cell, activated CD8 T cell, type 1 T helper cell, gamma delta T cell and activated dendritic cell infiltration were significantly increased (Figures 7A, B). Subsequently, to investigate histological infiltration of immune cells, an OA animal model was established through DMM surgery. During OA progression, characteristic pathological manifestations including cartilage degradation and subchondral bone sclerosis were observed (Figure 7C). Furthermore, upregulated expression of immune cell markers such as INOS and ARG1 was detected in joint tissues of OA group mice (Figure 7D). Spearman correlation was used to analyze the interactions between immune cells. The results showed

significant correlations between most immune cells, for example, effector memory CD8 T cell was significantly positively correlated with MDSC ( $r=0.88$ ) (Figure 8A). Para inflammation, check-point, type II IFN response and other immune functions were significantly activated in OA samples (Figure 8B). AKR1B1, TOLLIP, RHOBTB3, HP and S100A8 were well correlated with multiple immune functions (Figures 8C, D). Data of results were shown in the Supplementary Tables S11, 12.

## Discussion

In recent years, the role of exosomes in OA, as an important mediator of intercellular communication, has gradually gained attention in research. Recent studies have shown that exosomes not only carry a variety of bioactive molecules, but also participate in key signaling pathways in the pathological process of OA (29). For example, microRNAs enriched in exosomes (e.g., miR-146a, miR-140, etc.) have been shown to regulate signaling pathways associated with inflammation and cartilage degradation. These

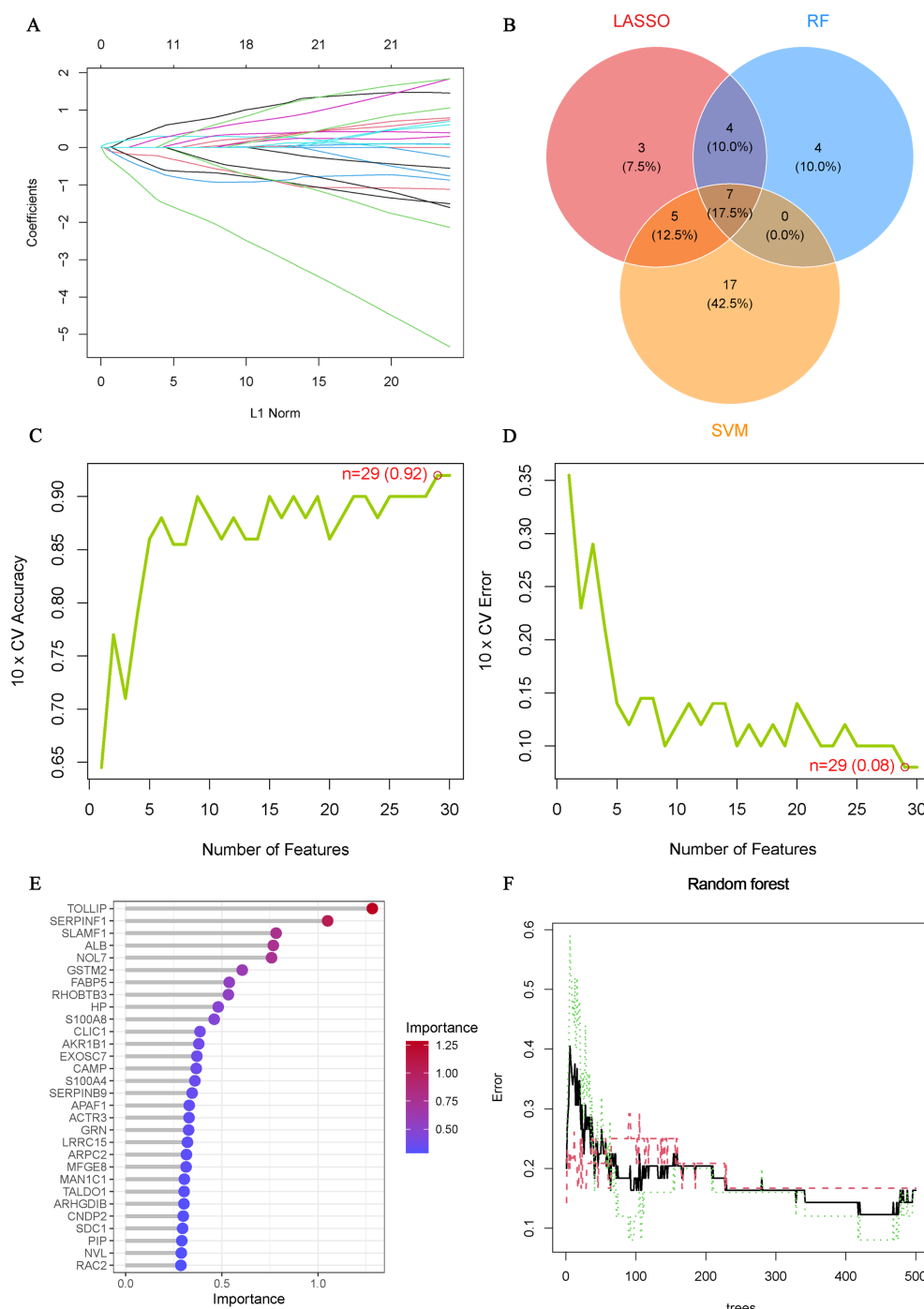


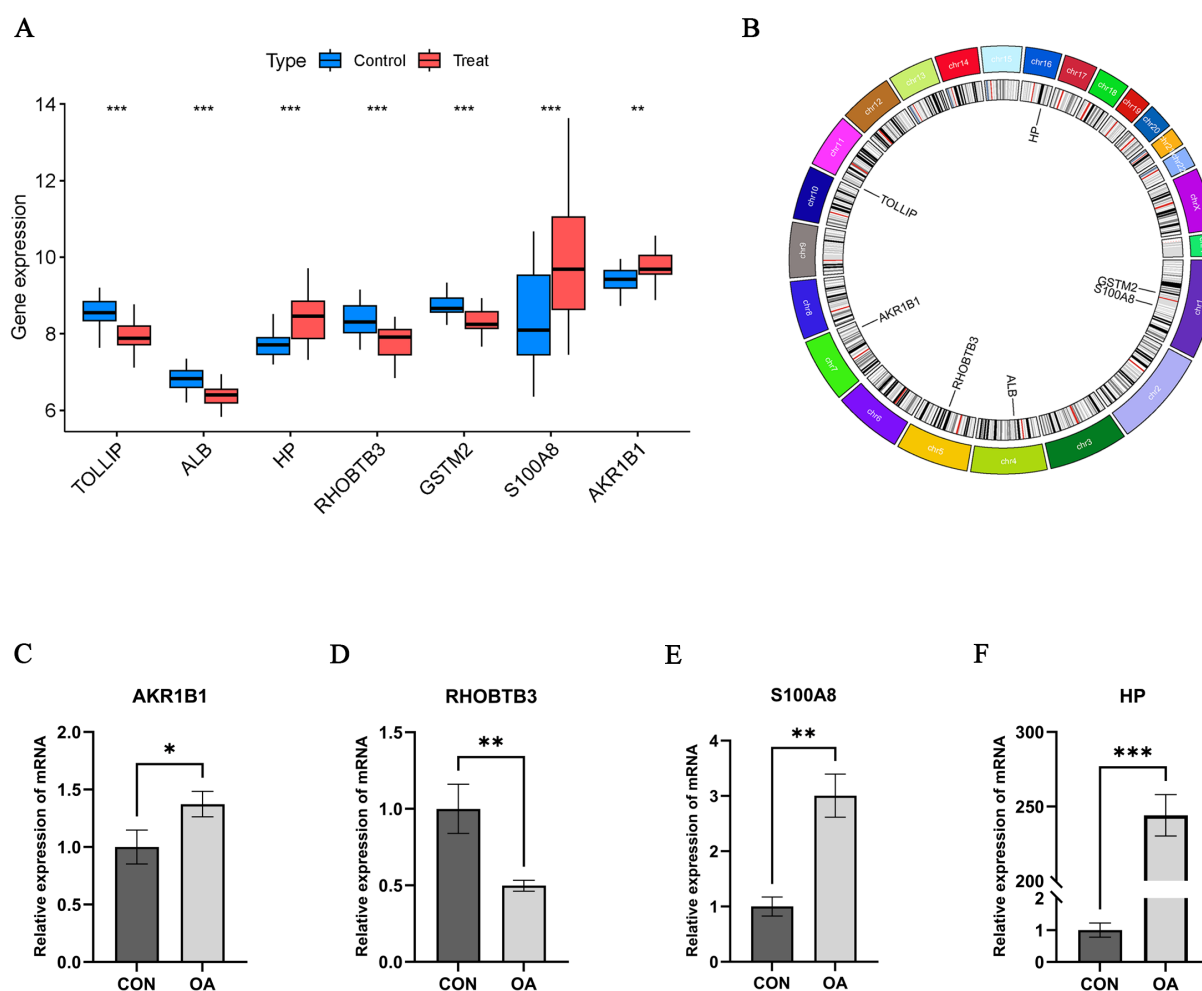
FIGURE 4

Machine Learning Screening Hub OA-ERDEGs. **(A)** Cross-validations of the choice of adjustment parameters in the LASSO model. Each curve corresponds to one gene. **(B)** LASSO, SVM-RFE and Random Forest algorithms for screening Venn diagrams of the OA-ERDEGs. **(C, D)** Maximum accuracy and minimum error plots of the SVM-RFE algorithm for screening optimal OA-ERDEGs. **(E)** Ranking of the relative importance of OA-ERDEGs. **(F)** Relationship between the number of random forest trees and the error rate.

microRNAs are involved in the regulation of classical pathways such as NF- $\kappa$ B, Wnt/ $\beta$ -catenin, etc. by affecting the expression of target genes, which contributes to the pathologic process of OA (30). In addition, it has been shown that exosome-derived proteins also play an important role in the development of OA. Proteins in exosomes are not only associated with intercellular signaling, but

may also affect the proliferation and differentiation of fibroblasts and chondrocytes, which in turn affects the metabolic homeostasis of articular cartilage (31). Notably, the interaction between exosomes and immune cells is also one of the research focuses. Through the release of inflammatory factors and cytokines from exosomes, the activity and infiltration of immune cells can be





**FIGURE 5**  
Hub OA-ERDEGs expression difference and validation. **(A)** Expression of 7 hub genes in OA samples. **(B)** Chromosome distribution of the 7 OA-ERDEGs. **(C-F)** The mRNA expression levels of Hub OA-ERDEGs were detected by qRT-PCR. Among them, AKR1B1, RHOBTB3, S100A8 and HP were significant. \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ .

modulated to influence the inflammatory response in OA. This process may influence the clinical manifestations and course of OA by regulating the accumulation and release of exosomes in the joint cavity (32). Therefore, exploring in depth the role of exosomes and related signaling pathways in OA not only provides new perspectives for understanding the pathogenesis of the disease, but also lays the foundation for seeking new biomarkers and therapeutic targets in the clinic.

In normal and OA cartilage samples, we identified 105 ERDEGs, and further screened by machine learning, 7 Hub OA-ERDEGs (TOLLIP, ALB, HP, RHOBTB3, GSTM2, S100A8, AKR1B1) were identified. Our findings suggest that Hub OA-ERDEGs have excellent diagnostic ability to predict OA and are significantly up- or down-regulated in cartilage samples from OA patients. TOLLIP is considered a negative regulator of the Toll-like receptor (TLRs) signaling pathway. It inhibits the NF- $\kappa$ B and MAPK signaling pathways by binding to TLRs, further inhibiting the activation of downstream signals, thereby reducing the intensity of the inflammatory response (33). The ALB gene encodes albumin, a

plasma protein synthesized primarily by the liver that has a variety of physiological functions, and recent studies have found that the ALB gene and its products play an important role in the inflammatory response. For example, albumin binds and transports antioxidants and reduces oxidative stress, thereby decreasing the inflammatory response. In addition, albumin is able to regulate cytokine secretion and reduce inflammation through interactions with immune cells (34). HP is primarily responsible for encoding acute phase protein-hemoglobin binding proteins in the organism. During tissue injury or inflammation, erythrocyte rupture releases hemoglobin, and free hemoglobin has strong oxidative properties that may lead to cellular damage. HP reduces oxidative stress by binding to free hemoglobin, forming a complex that facilitates its clearance by macrophages (35). RHOBTB3 gene is a gene encoding a member of the Rho GTPase family, and it has been pointed out that the expression level of RHOBTB3 is significantly increased in breast cancer tissues. And it can promote breast cancer progression by regulating the expression of Collagen Type I Alpha 1 Chain (COL1A1) (36). Its high expression in OA samples was also found in this study, so we

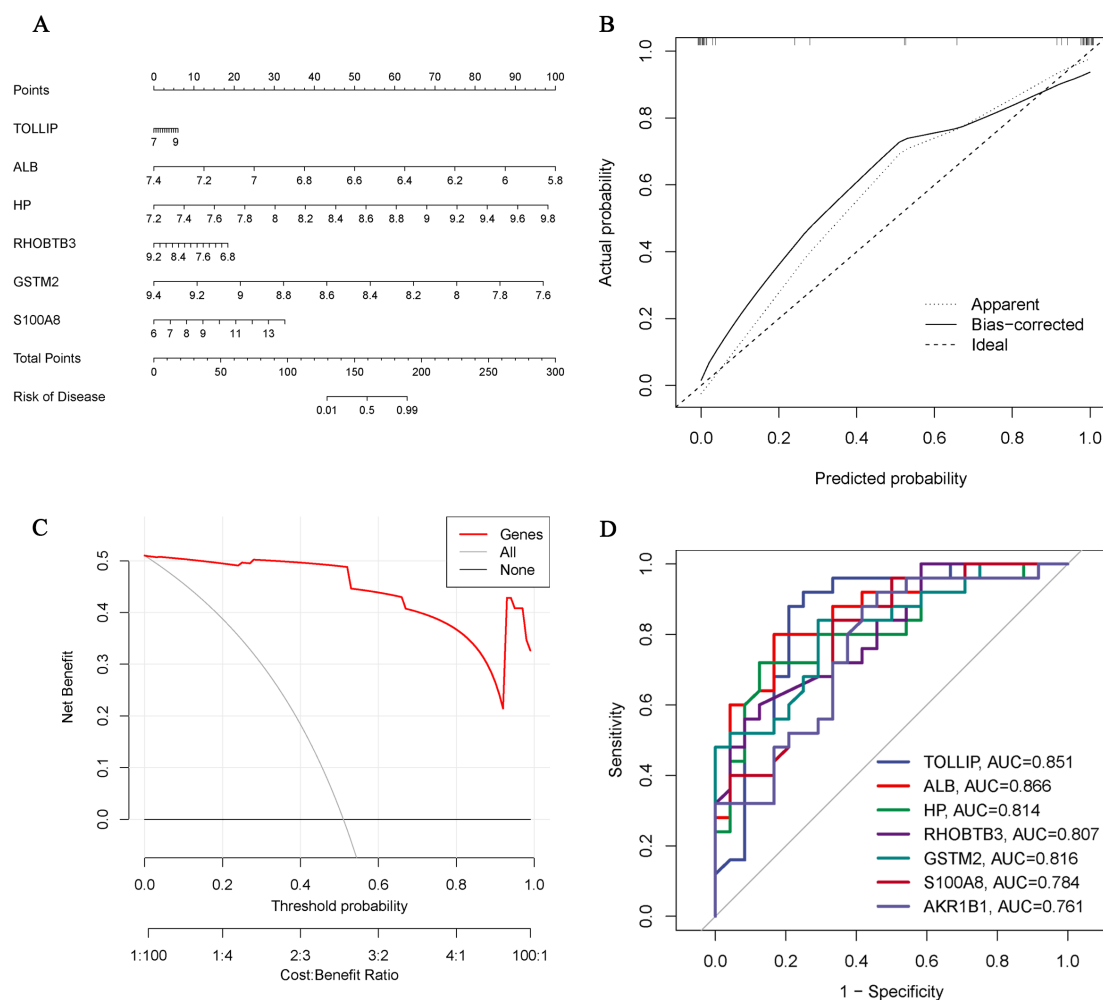


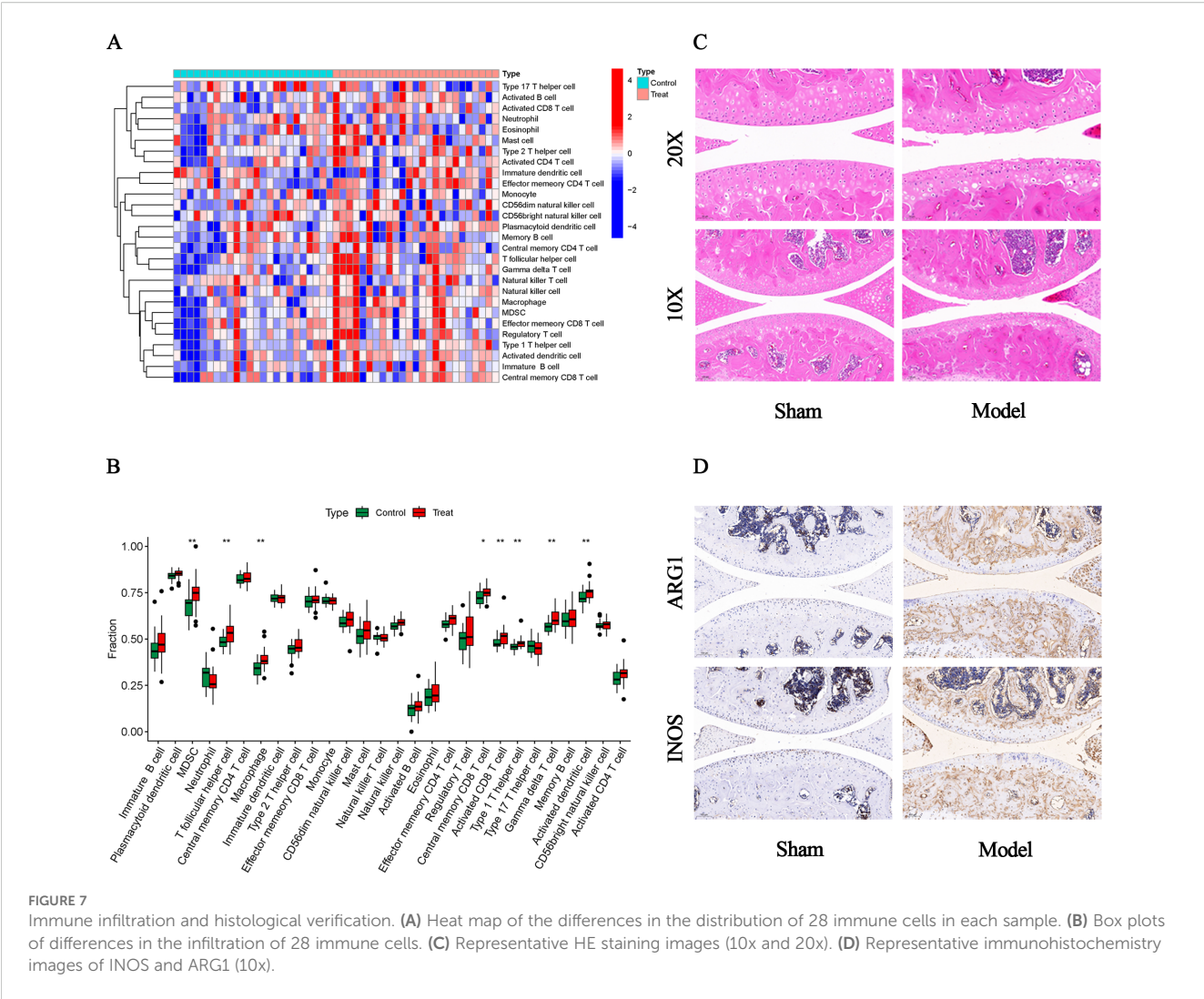
FIGURE 6

Hub OA-ERDEGs risk prediction model. **(A)** Nomogram of Hub OA-ERDEGs in the diagnosis of OA patients. **(B)** Calibration curve used to estimate the predictive accuracy of the nomogram (the closer to the ideal dashed line, the more reliable the result). **(C)** Accuracy of the clinical decision curve detection model (the further the red line endpoints are from the grey line, the higher the accuracy). **(D)** ROC curve analysis of Hub OA-ERDEGs and nomogram in the training set.

hypothesized that RHOBTB3 is also involved in inflammation in osteoarthritis. The GSTM2 gene is a member of the glutathione S-transferase family, which maintains the MAPK signaling pathway by directly interacting with apoptosis signal-regulated kinase 1 (ASK1) to scavenge free radicals and oxidative products *in vivo*, thereby reducing the level of oxidative stress and indirectly decreasing the degree of inflammatory response (37). The enzyme encoded by the AKR1B1 gene is mainly responsible for reducing a variety of aldehydes to their corresponding alcohols, and studies have indicated that Linarin has anti-inflammatory, antioxidant, and anti-apoptotic effects in osteoarthritis, and Linarin reduces the intensity of the inflammatory response by inhibiting endoplasmic reticulum stress and further inhibiting AKR1B1 (38). The S100A8 gene encodes a small protein belonging to the S100 family of proteins, which play important roles in many biological processes, including cell proliferation, differentiation, inflammation, and the immune response. The S100A8 protein usually associates with the S100A9 protein to form a dimer, the S100A8/S100A9 complex,

which promotes the aggregation of monocytes and other inflammatory cells, thereby exacerbating the inflammatory response and tissue damage (39). It has been noted that the mRNA expression levels of S100A8 and S100A9 are significantly increased in the early stages of OA, but their expression in cartilage gradually decreases as the disease progresses. This suggests that these two proteins may promote cartilage degradation by up-regulating matrix metalloproteinases (MMPs) and aggrecans in the early stages of OA, while this sustained effect is lost in the later stages of OA. In contrast, in inflammatory arthritis, S100A8 and S100A9 show more sustained expression and activity (40). S100A8 was significantly up-regulated in OA samples in the present study, which may indicate that the OA model resulting from stimulation of chondrocytes with IL-1 $\beta$  for 24h is at an early stage of the disease, highly consistent with the results of previous studies.

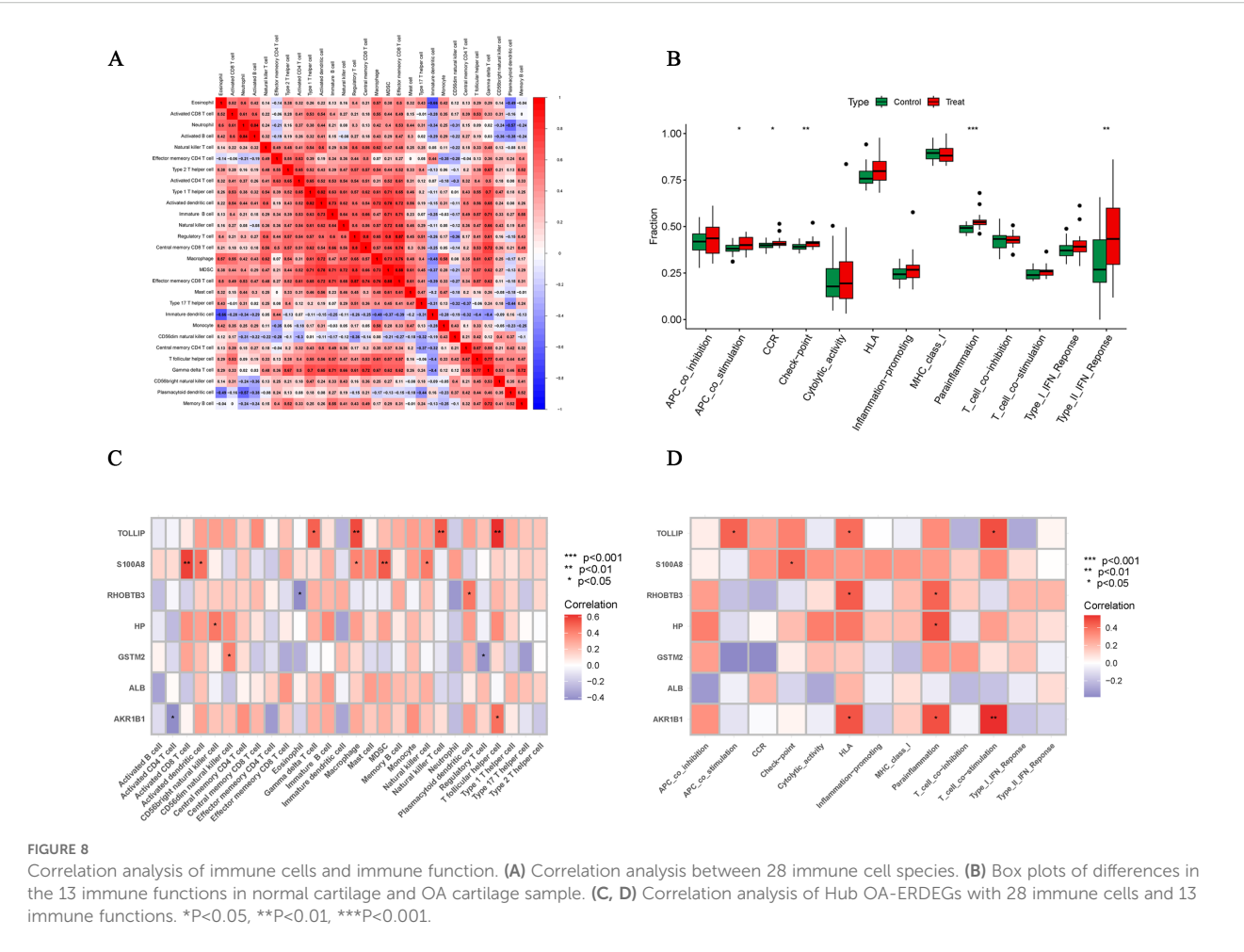
Inflammatory responses in osteoarthritis exacerbate damage to articular cartilage and synovium through the release of pro-inflammatory factors and mediators, leading to joint pain and



dysfunction; while immune deficiencies may deregulate these inflammatory processes, further exacerbating inflammatory processes and degenerative joint disease (41). T follicular helper (TFh) cells are a specific class of CD4<sup>+</sup> T cells that are primarily responsible for promoting B cell maturation and antibody production in the germinal centers of the lymph nodes. TFh cells may be involved in the pathological process of arthritis through the production of cytokines, such as IL-21, and the modulation of B cell activity (42). It has been suggested that the expression of long chain non-coding RNA (lnc RNA) MM2P in monocyte-derived cells promotes the polarization of M2-type macrophages and the delivery of Sox9 mRNA and protein via their secreted exosomes, thereby enhancing the differentiation and function of primary chondrocytes (43). Dendritic cells are important antigen-presenting cells that play a key role in the immune response. Exosomes produced by dendritic cells can deliver mRNAs, miRNAs, and cytokines and interact with immune cells (44). Myeloid-derived suppressor cells (MDSCs) are a class of cells with immunosuppressive functions, mainly derived from bone marrow, capable of suppressing the activity of T cells and other immune cells through multiple mechanisms. In the tumor

microenvironment, MDSCs promote tumor cell growth and metastasis as well as suppress anti-tumor immune responses by secreting exosomes (45). We hypothesized that in some inflammation-related diseases, such as osteoarthritis, exosomes produced by MDSCs could similarly play an immunomodulatory role. We used Spearman correlation analysis to show that nine Hub OA-ERDEGs were reasonably correlated with immune cell function, suggesting that these genes influence immune cell function and play a key role in the development and progression of OA.

This study has several limitations. First, the intersection of non-chondrocyte-derived exosomal genes with OA chondrocyte datasets may result in gene omission due to the absence of cell type-specific exosomal gene references. While data screening and processing enhanced the reliability of candidate genes, future validation through chondrocyte-specific exosome isolation combined with multi-omics analysis remains necessary. Second, algorithm-dependent variability (LASSO/SVM/RF) may affect result stability, though our multi-model integration aligns with current multi-omics standards. Third, the sample size (24 controls/25 OA cases across four datasets), though geographically diverse, remains



modest, necessitating multicenter validation. Finally, while qRT-PCR confirmed expression trends of AKR1B1, RHOBTB3, S100A8, and HP in OA chondrocytes, their diagnostic specificity, functional mechanisms, and clinical utility require further exploration through large-scale cohorts, *in vitro/in vivo* models, and differential diagnosis assays. These limitations highlight the need for expanded validation frameworks to advance the translational potential of our findings.

Ethics

All summary statistical data used for screening Hub genes were generated from previous studies, and all original studies obtained ethical approval and personal consent.

Conclusion

In summary, this study preliminarily explored the potential linkage of differential and exosome-related genes in OA cartilage tissues. In addition, seven Hub OA-ERDEGs have excellent diagnostic ability for OA, these genes may become new targets for

OA diagnosis and treatment. However, more experimental studies are needed to support our findings.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding authors.

Ethics statement

The animal study was approved by Institutional Animal Care and Use Committee of Huazhong University of Science and Technology. The study was conducted in accordance with the local legislation and institutional requirements.

Author contributions

TL: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Project administration, Resources,



Software, Writing – original draft. JW: Conceptualization, Data curation, Methodology, Project administration, Software, Writing – original draft. HW: Supervision, Validation, Visualization, Writing – review & editing. CC: Conceptualization, Supervision, Validation, Visualization, Writing – review & editing.

## Funding

The author(s) declare that no financial support was received for the research and/or publication of this article.

## Acknowledgments

We would like to thank all the online databases used in the article. We would like to thank the help of the proofreaders and editors.

## Conflict of interest

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

## References

1. Yao Q, Wu X, Tao C, Gong W, Chen M, Qu M, et al. Osteoarthritis: pathogenic signaling pathways and therapeutic targets. *Signal Transduct Target Ther.* (2023) 8:56. doi: 10.1038/s41392-023-01330-w
2. Katz JN, Arant KR, Loeser RF. Diagnosis and treatment of hip and knee osteoarthritis: A review. *JAMA.* (2021) 325:568–78. doi: 10.1001/jama.2020.22171
3. Hunter DJ, Bierma-Zeinstra S. Osteoarthritis. *Lancet.* (2019) 393:1745–59. doi: 10.1016/S0140-6736(19)30417-9
4. Emery CA, Whittaker JL, Mahmoudian A, Lohmander LS, Roos EM, Bennell KL, et al. Establishing outcome measures in early knee osteoarthritis. *Nat Rev Rheumatol.* (2019) 15:438–48. doi: 10.1038/s41584-019-0237-3
5. Kalluri R, McAndrews KM. The role of extracellular vesicles in cancer. *Cell.* (2023) 186:1610–26. doi: 10.1016/j.cell.2023.03.010
6. Flemming A. RNA exosome essential for early B cell development. *Nat Rev Immunol.* (2022) 22:462–3. doi: 10.1038/s41577-022-00758-5
7. Zhou R, Guo J, Jin Z. Advancing osteoarthritis therapy with GMOCS hydrogel-loaded BMSCs-exos. *J Nanobiotechnol.* (2024) 22:493. doi: 10.1186/s12951-024-02713-z
8. Li Z, Lu H, Fan L, Ma X, Duan Z, Zhang Y, et al. Microneedle-delivered PDA@Exo for multifaceted osteoarthritis treatment via PI3K-Akt-mTOR pathway. *Adv Sci (Weinh).* (2024) 11:e2406942. doi: 10.1002/adv.202406942
9. Qian Y, Chu G, Zhang L, Wu Z, Wang Q, Guo J, et al. M2 macrophage-derived exosomal miR-26b-5p regulates macrophage polarization and chondrocyte hypertrophy by targeting TLR3 and COL10A1 to alleviate osteoarthritis. *J Nanobiotechnol.* (2024) 22:72. doi: 10.1186/s12951-024-02336-4
10. Liu B, Xian Y, Chen X, Shi Y, Dong J, Yang L, et al. Inflammatory fibroblast-like synovocyte-derived exosomes aggravate osteoarthritis via enhancing macrophage glycolysis. *Adv Sci (Weinh).* (2024) 11:e2307338. doi: 10.1002/adv.202307338
11. Machine learning helps to determine the diverse conformations of RNA molecules. *Nature.* (2024) 637(8048):1234–43. doi: 10.1038/d41586-024-04134-2
12. Greener JG, Kandathil SM, Moffat L. A guide to machine learning for biologists. *Nat Rev Mol Cell Biol.* (2022) 23:40–55. doi: 10.1038/s41580-021-00407-0
13. Haghverdi L, Lun A, Morgan MD. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nat Biotechnol.* (2018) 36:421–7. doi: 10.1038/nbt.4091
14. Malta TM, Sokolov A, Gentles AJ, Burzykowski T, Poisson L, Weinstein JN, et al. Machine learning identifies stemness features associated with oncogenic dedifferentiation. *Cell.* (2018) 173:338–54.e15. doi: 10.1016/j.cell.2018.03.034
15. Chen B, Sun X, Huang H, Feng C, Chen W, Wu D, et al. An integrated machine learning framework for developing and validating a diagnostic model of major depressive disorder based on interstitial cystitis-related genes. *J Affect Disord.* (2024) 359:22–32. doi: 10.1016/j.jad.2024.05.061
16. Zhou J, Huang J, Li Z, Song Q, Yang Z, Wang L, et al. Identification of aging-related biomarkers and immune infiltration characteristics in osteoarthritis based on bioinformatics analysis and machine learning. *Front Immunol.* (2023) 14:1168780. doi: 10.3389/fimmu.2023.1168780
17. Gene Ontology Consortium: going forward. *Nucleic Acids Res.* (2015) 43:D1049–1056. doi: 10.1093/nar/gku1179
18. Kanehisa M, Furumichi M, Sato Y. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res.* (2023) 51:D587–92. doi: 10.1093/nar/gkac963
19. Schriml LM, Munro JB, Schor M, Olley D, McCracken C, Felix V, et al. The human disease ontology 2022 update. *Nucleic Acids Res.* (2022) 50:D1255–61. doi: 10.1093/nar/gkab1063
20. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U.S.A.* (2005) 102:15545–50. doi: 10.1073/pnas.0506580102
21. Yu X, Chen YA, Conejo-Garcia JR. Estimation of immune cell content in tumor using single-cell RNA-seq reference data. *BMC Cancer.* (2019) 19:715. doi: 10.1186/s12885-019-5927-3
22. Zeng Y, Liu R, Li S, Wei J, Luo F, Chen Y, et al. Analysis of risk factors and development of a nomogram prediction model for renal tubular acidosis in primary Sjogren syndrome patients. *Arthritis Res Ther.* (2024) 26:151. doi: 10.1186/s13075-024-03383-w
23. Bai J, Huang JH, Price C, Schauer JM, Suh LA, Harmon R, et al. Prognostic factors for polyp recurrence in chronic rhinosinusitis with nasal polyps. *J Allergy Clin Immunol.* (2022) 150:352–61.e7. doi: 10.1016/j.jaci.2022.02.029
24. Zhang WY, Chen ZH, An XX, Li H, Zhang HL, Wu SJ, et al. Analysis and validation of diagnostic biomarkers and immune cell infiltration characteristics in

## Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

## Publisher's note

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

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fimmu.2025.1596912/full#supplementary-material>

- pediatric sepsis by integrating bioinformatics and machine learning. *World J Pediatr.* (2023) 19:1094–103. doi: 10.1007/s12519-023-00717-7
25. Hu J, Szymczak S. A review on longitudinal data analysis with random forest. *Brief Bioinform.* (2023) 24:bbad002. doi: 10.1093/bib/bbad002
26. Van Calster B, Wynants L, Verbeek J, Verbakel J, Christodoulou E, Vickers A, et al. Reporting and interpreting decision curve analysis: A guide for investigators. *Eur Urol.* (2018) 74:796–804. doi: 10.1016/j.eururo.2018.08.038
27. Silva G, Fagundes TP, Teixeira BC. Machine learning for hypertension prediction: a systematic review. *Curr Hypertens Rep.* (2022) 24:523–33. doi: 10.1007/s11906-022-01212-6
28. Chen Y, Feng Y, Yan F, Zhao Y, Zhao H, Guo Y, et al. A novel immune-related gene signature to identify the tumor microenvironment and prognose disease among patients with oral squamous cell carcinoma patients using ssGSEA: A bioinformatics and biological validation study. *Front Immunol.* (2022) 13:922195. doi: 10.3389/fimmu.2022.922195
29. Cheng C, Wu Y, Huang Y, Xue Q, Wang Y, Liao F, et al. Epigenetic modification and exosome effects on autophagy in osteoarthritis. *Biochem Pharmacol.* (2023) 218:115930. doi: 10.1016/j.bcp.2023.115930
30. Qiu B, Xu X, Yi P. Curcumin reinforces MSC-derived exosomes in attenuating osteoarthritis via modulating the miR-124/NF- $\kappa$ B and miR-143/ROCK1/TLR9 signalling pathways. *J Cell Mol Med.* (2020) 24:10855–65. doi: 10.1111/jcmm.15714
31. Yu H, Huang Y, Yang L. Research progress in the use of mesenchymal stem cells and their derived exosomes in the treatment of osteoarthritis. *Ageing Res Rev.* (2022) 80:101684. doi: 10.1016/j.arr.2022.101684
32. Wang LT, Liu KJ, Sytwu HK. Advances in mesenchymal stem cell therapy for immune and inflammatory diseases: Use of cell-free products and human pluripotent stem cell-derived mesenchymal stem cells. *Stem Cells Transl Med.* (2021) 10:1288–303. doi: 10.1002/sctm.21-0021
33. Liu X, Ren X, Zhou L, Liu K, Deng L, Qing Q, et al. Tollip orchestrates macrophage polarization to alleviate intestinal mucosal inflammation. *J Crohns Colitis.* (2022) 16:1151–67. doi: 10.1093/ecco-jcc/jjac019
34. Eckart A, Struja T, Kutz A, Baumgartner A, Baumgartner T, Zurfluh S, et al. Relationship of nutritional status, inflammation, and serum albumin levels during acute illness: A prospective study. *Am J Med.* (2020) 133:713–22.e7. doi: 10.1016/j.amjmed.2019.10.031
35. Shih AW, McFarlane A, Verhovsek M. Haptoglobin testing in hemolysis: measurement and interpretation. *Am J Hematol.* (2014) 89:443–7. doi: 10.1002/ajh.23623
36. Kim K, Kim YJ. RhoBTB3 regulates proliferation and invasion of breast cancer cells via col1a1. *Mol Cells.* (2022) 45:631–9. doi: 10.14348/molcells.2022.2037
37. Lan T, Hu Y, Hu F, Li H, Chen Y, Zhang J, et al. Hepatocyte glutathione S-transferase mu 2 prevents non-alcoholic steatohepatitis by suppressing ASK1 signaling. *J Hepatol.* (2022) 76:407–19. doi: 10.1016/j.jhep.2021.09.040
38. Chen X, Chen C, Hao J, Qin R, Qian B, Yang K, et al. AKR1B1 upregulation contributes to neuroinflammation and astrocytes proliferation by regulating the energy metabolism in rat spinal cord injury. *Neurochem Res.* (2018) 43:1491–9. doi: 10.1007/s11064-018-2570-3
39. Huang X, Liu J, Huang W. Identification of S100A8 as a common diagnostic biomarkers and exploring potential pathogenesis for osteoarthritis and metabolic syndrome. *Front Immunol.* (2023) 14:1185275. doi: 10.3389/fimmu.2023.1185275
40. Zreikat H, Belluoccio D, Smith MM, Wilson R, Rowley LA, Jones K, et al. S100A8 and S100A9 in experimental osteoarthritis. *Arthritis Res Ther.* (2010) 12:R16. doi: 10.1186/ar2917
41. Faust HJ, Zhang H, Han J, Wolf MT, Jeon OH, Sadler K, et al. IL-17 and immunologically induced senescence regulate response to injury in osteoarthritis. *J Clin Invest.* (2020) 130:5493–507. doi: 10.1172/JCI134091
42. Lin M, Huang L, Huang J, Yu J, Yang X, Yang J, et al. Modulation of PKM2 inhibits follicular helper T cell differentiation and ameliorates inflammation in lupus-prone mice. *J Autoimmun.* (2024) 145:103198. doi: 10.1016/j.jaut.2024.103198
43. Bai J, Zhang Y, Zheng X, Huang M, Cheng W, Shan H, et al. LncRNA MM2P-induced, exosome-mediated transfer of Sox9 from monocyte-derived cells modulates primary chondrocytes. *Cell Death Dis.* (2020) 11:763. doi: 10.1038/s41419-020-02945-5
44. Li Z, Wang Y, Xiao K, Xiang S, Li Z, Weng X, et al. Emerging role of exosomes in the joint diseases. *Cell Physiol Biochem.* (2018) 47:2008–17. doi: 10.1159/000491469
45. Gautam SK, Batra SK, Jain M. Molecular and metabolic regulation of immunosuppression in metastatic pancreatic ductal adenocarcinoma. *Mol Cancer.* (2023) 22:118. doi: 10.1186/s12943-023-01813-y