AHEAD OF PRINT

Original papers

Basic sciences; orthopedics

Identification and validation of novel diagnostic biomarkers across multiple tissues in osteoarthritis

Guihao Zheng[1,2,A,D], Yulong Ouyang[1,A], Shuilin Chen[1,B], Bei Hu[1,2,C], Shuai Xu[1,2,B], Guicai Sun[1,E,F]

[1] Department of Sports Medicine, Orthopedic Hospital, The First Affiliated Hospital, Jiangxi Medical College, Nanchang University, China
[2] Graduate School of Jiangxi Medical College, Nanchang University, China

A – research concept and design; B – collection and/or assembly of data; C – data analysis and interpretation;
D – writing the article; E – critical revision of the article; F – final approval of the article

Advances in Clinical and Experimental Medicine, ISSN 1899–5276 (print), ISSN 2451–2680 (online) *Adv Clin Exp Med*. 2026

Address for correspondence
Guicai Sun
E-mail: ndsfy0740@ncu.edu.cn

Funding sources
None declared

Conflict of interest
None declared

Received on June 4, 2025
Reviewed on September 1, 2025
Accepted on September 26, 2025

Published online on July 3, 2026

Abstract

Background. Osteoarthritis (OA) is a degenerative joint disease characterized by synovial inflammation, cartilage degradation, and subchondral bone remodeling. Currently, no universally accepted or clinically validated biomarkers exist for OA diagnosis and treatment, highlighting the need to identify potential genetic biomarkers to support early detection and therapeutic research.

Objectives. This study aimed to identify and preliminarily validate potential diagnostic biomarkers for OA using integrated bioinformatics and machine learning (ML) approaches to provide molecular insights supporting early diagnosis and therapeutic strategies.

Materials and methods. Gene expression data from OA patients and normal controls were obtained from the Gene Expression Omnibus (GEO) database. Differentially expressed genes (DEGs) were identified using R. Protein–protein interaction (PPI) network analysis and functional module analysis were performed to screen for hub genes. Gene Ontology and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were conducted to determine the biological roles and pathways of DEGs, along with transcription factor prediction. Three ML models (least absolute shrinkage and selection operator (LASSO), support vector machine recursive feature elimination (SVM-RFE), and Random Forest) were used to identify OA-specific characteristic genes. Key diagnostic genes were identified from the intersection of the ML results and validated using external datasets of cartilage, synovium, and blood. IRS2 expression was further validated via in vivo (animal model) and in vitro (cell culture) experiments.

Results. Fifteen OA-related characteristic genes were identified, including *IRS2*, *ADM*, *SIK1*, *PTN*, *CX3CR1*, *WNT5A*, *IL21R*, *APOD*, *CRLF1*, *FKBP5*, *PNMAL1*, *NPR3*, *RARRES1*, *ASPN*, and *POSTN*. Functional enrichment analysis suggested involvement in extracellular matrix (ECM) organization, interleukin-17 (IL-17) and relaxin signaling, and the AGE-RAGE pathway in diabetic complications. Three genes showed strong diagnostic potential. IRS2 was downregulated, while WNT5A and PTN were upregulated in OA samples. Immunohistochemistry (IHC), real-time quantitative polymerase chain reaction (qPCR), and western blotting (WB) confirmed the downregulation of IRS2 in OA samples, consistent with bioinformatics predictions, in both chondrocytes and mouse joint tissues.

Conclusions. *IRS2*, *WNT5A*, and *PTN* may serve as potential diagnostic biomarkers for osteoarthritis, supporting early or tissue-specific diagnosis and offering insights for future clinical and therapeutic applications.

Key words: machine learning, osteoarthritis, biomarkers, bioinformatics, tissue-specific expressed genes

Cite as
Zheng G, Ouyang Y, Chen S, Hu B, Xu S, Sun G. Identification and validation of novel diagnostic biomarkers across multiple tissues in osteoarthritis [published online as ahead of print on July 3, 2026]. *Adv Clin Exp Med*. 2026. doi:10.17219/acem/211377

DOI
10.17219/acem/211377

Highlights

- Integrated bioinformatics and machine learning identified IRS2, WNT5A, and PTN as key diagnostic biomarkers for osteoarthritis (OA).
- IRS2 was significantly downregulated in OA cartilage, synovium, and blood, confirmed through qPCR, immunohistochemistry, and western blot analyses.
- Functional enrichment linked OA pathogenesis to IL-17, relaxin, and AGE-RAGE signaling pathways, emphasizing inflammation and extracellular matrix remodeling.
- Machine learning algorithms (LASSO, SVM-RFE, Random Forest) demonstrated strong diagnostic accuracy, offering promising tools for early OA detection and personalized therapy.

Background

Osteoarthritis (OA) remains a global health challenge,[1] affecting an estimated 600 million people worldwide, or 7.6% of the world's population,[2] with a higher prevalence among women.[3] The incidence and disability rates of OA have been rising since 1990,[4] and its prevalence is expected to nearly double in the next decade.[5] Additionally, similar to rheumatoid arthritis, OA imposes a substantial socioeconomic burden, with costs estimated to have reached 80 billion USD in 2016 and growing at 5.7% per year.[4,6] In the elderly, OA is a leading cause of mobility impairment and disability.[7,8]

Osteoarthritis is clinically characterized by chronic joint pain, stiffness, swelling, reduced range of motion, and progressive loss of function, which can significantly impair quality of life.[9] Pathologically, it involves extracellular matrix (ECM) degradation, altered type II collagen (COL2A1) synthesis, elevated matrix metalloproteinase 13 (MMP13) expression, and cyclooxygenase-2 (COX2)-mediated inflammation triggered by pro-inflammatory cytokines.[10] These factors act on cartilage, synovium, and subchondral bone, which interact and contribute to disease progression.[11–13] The etiology of OA is multifactorial, involving age, obesity, joint injury, genetics, and mechanical stress.[7,8] However, the exact mechanisms remain unclear, and effective early diagnostic tools are still lacking.

With the rapid development of high-throughput sequencing technology, an increasing number of OA-related datasets have been generated and utilized.[14–18] Although some studies have focused on OA biomarkers, most have been based on single tissues or individual datasets.[19] The occurrence and development of OA are closely related to interactions among cartilage, subchondral bone, and synovitis. Despite some controversy regarding the order of pathological events, it is undeniable that these components are all involved in disease progression,[20,21] which may contribute to the limited reliability of current research findings.[22–25] The application of machine learning (ML) in OA biomarker development has garnered growing research interest. For instance, Chen et al.[25] applied synovial tissue microarray datasets and identified *ZBTB16*, *TNFSF11*, *SCRG1*, and *KDELR3* as diagnostic biomarkers for OA. A study by Han et al.[23] also reported that *TLR7*, *CSF1R*, *APOE*, *C1QA*, and *CCL5* are key genes with high diagnostic value for OA after applying weighted gene co-expression network analysis (WGCNA) and the cytoHubba algorithm. Although these findings may contribute to a better understanding of OA and provide new insights into its diagnosis and treatment, the limited sample sizes and the small number of algorithms used may undermine the reliability and robustness of these studies to some extent.

Objectives

This study aimed to identify and validate reliable diagnostic biomarkers for OA using integrated bioinformatics and ML approaches to improve early diagnosis and inform therapeutic strategies

Materials and methods

Data collection

The datasets GSE51588, GSE12021, GSE55457, GSE56409, GSE114007, GSE168505, GSE169077, GSE55235, GSE129147, and GSE48556 were obtained from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo). GSE51588 is a whole-genome expression profile of subchondral bone. GSE12021, GSE55457, GSE56409, and GSE55235 contain data obtained from synovial samples of OA patients and normal controls. GSE114007, GSE168505, GSE169077, and GSE129147 are RNA sequencing datasets derived from joint cartilage tissues collected from both OA patients and healthy individuals, while GSE48556 consists of mononuclear cells isolated from the blood of OA patients and normal controls. All these datasets were derived from human samples.

Identification of differentially expressed genes (DEGs)

Identification of DEGs using the Robust Rank Aggregation method

The 7 raw datasets were processed using the limma package in Bioconductor, including normalization and log transformation. Subsequently, DEGs were analyzed using the limma package (https://bioconductor.org/packages/release/bioc/html/limma.html), and ranked lists of upregulated and downregulated DEGs were generated for each dataset based on fold-change values. Finally, we integrated the results of these 7 datasets using the R package "RobustRankAggreg" based on the Robust Rank Aggregation (RRA) method to identify the most significant DEGs. In the RRA analysis, genes with a log fold change (logFC) >1 and an adjusted p-value <0.05 were selected as significant DEGs.

Identification of DEGs using the Surrogate Variable Analysi) method

Before merging the 7 microarray datasets, we used the ComBat function in the Surrogate Variable Analysis (SVA) package to correct for batch effects, aiming to minimize experimental variance before subsequent analyses. Box plots, principal component analysis (PCA), and other methods were used to assess the data before and after correction (Supplementary Fig. 1). The final merged dataset consisted of 167 samples, including 66 normal control samples and 101 OA samples. In the SVA analysis, genes with a log fold change (logFC) >1 and an adjusted p-value <0.05 were selected as significant DEGs. Finally, the genes identified by both the RRA and SVA methods were considered key genes for OA.

Identification of gene clusters and construction of protein–protein interaction networks

Gene Ontology (GO) functional annotation, Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis, and Disease Ontology (DO) enrichment analysis were performed using R packages (clusterProfiler: https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html; DOSE: https://bioconductor.org/packages/release/bioc/html/DOSE.html; org.Hs.eg.db: https://bioconductor.org/packages/release/data/annotation/html/org.Hs.eg.db.html; enrichplot: https://bioconductor.org/packages/release/bioc/html/enrichplot.html).

An adjusted p-value <0.05 was considered statistically significant for enrichment.

The STRING database (https://string-db.org) was used to construct the protein–protein interaction (PPI) network, which was then imported into Cytoscape v. 3.9.1 software (https://cytoscape.org/index.html), to visualize the key nodes in the molecular interaction network. The cytoHubba algorithms were employed to analyze the PPI network. Only genes identified as hub genes by all cytoHubba algorithms ("MCC", "DMNC", "MNC", "Degree", "EPC", "BottleNeck", "EcCentricity", "Closeness", "Radiality", "Betweenness", "Stress", and "ClusteringCoefficient") were included in the final selection. The interacting genes and functions of these hub genes were predicted, and relevant PPI networks were generated using the GeneMANIA database (https://genemania.org). Enrichment analysis of the hub genes was further conducted using GO functional terms and KEGG pathway analysis, with $p < 0.05$ considered statistically significant.

Machine learning

We used the `train` function from the caret package to identify the best-performing model using various ML methods, including least absolute shrinkage and selection operator (LASSO), Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), eXtreme Gradient Boosting (XGBoost), and Generalized Linear Model (GLM). LASSO regression is a linear regression technique that incorporates L1 regularization to constrain model complexity, thereby achieving feature selection and preventing overfitting.[26] Support Vector Machine is a powerful supervised learning model that can be used for classification and regression tasks. It has shown wide applicability and superior performance in high-dimensional data, complex data structures, and small-sample learning.[27] Decision Tree is a supervised learning algorithm that splits nodes by selecting the features that best partition the data.[28] Random Forest is an ensemble learning method that improves model performance and stability by voting (classification) or averaging (regression) the results of multiple decision trees. It is particularly effective in identifying subtle patterns in complex datasets.[29] XGBoost is a gradient boosting framework that uses tree-based models to provide efficient and flexible ML algorithms. It is widely recognized for its excellent performance and efficiency and is extensively used in ML competitions and real-world applications.[30] Generalized Linear Model is an extension of linear models that can be applied to a wide range of data types, offering strong interpretability and broad applicability.

Through training and testing these models, we selected the 3 best-performing models to analyze the differentially expressed genes and identify the optimal feature genes. The resulting genes were intersected with the hub genes and used as diagnostic genes for the disease.

Validation of feature genes with receiver operating characteristic curve and construction of clinical diagnostic model

To validate the diagnostic importance of the candidate genes, we used the "rms" R package (https://hbiostat.org/R/rms) to construct a nomogram and to predict and

interpret the resulting model. Box plots and violin plots were used to visualize differential gene expression between the control and OA groups. We further evaluated the diagnostic value of the candidate genes through receiver operating characteristic (ROC) analysis, obtaining the area under the ROC curve (AUC) and the corresponding 95% confidence interval (95% CI). An AUC > 0.7 was considered indicative of good diagnostic performance.

Validation of feature genes using western blot and qPCR

Human SW1353 chondrocytes, commonly used for OA modeling,[10,31] were cultured in an incubator under standard conditions (37°C, 5% CO_2) using specialized SW1353 culture medium (Dulbecco's modified Eagle's medium (DMEM); Pricella, Wuhan, China), with the medium changed every 2–3 days. When SW1353 cells reached 50% confluence, they were stimulated with 10 ng/mL interleukin (IL)-1β (PeproTech, Rocky Hill, USA) for 24 h to model OA.[32–34]

For IRS2 overexpression, the IRS2-overexpressing plasmid was obtained from HanHeng Biotechnology (Shanghai, China). Transfection was performed in 6-well plates using the plasmid and Lipofectamine 3000 (Thermo Fisher Scientific, Waltham, USA), following the manufacturer's instructions. SW1353 cells were transfected at 70–90% confluence, incubated for 48 h, and then treated with 10 ng/mL IL-1β for an additional 24 h. Successful overexpression of IRS2 was confirmed before proceeding with subsequent experiments.

The treated chondrocytes were lysed in radioimmunoprecipitation assay (RIPA) buffer (Solarbio, Beijing, China) containing 1% protease and phosphatase inhibitors (GlpBio, Montclair, USA) for 30 min. The lysates were collected by thoroughly scraping the adherent cells and then incubated on a shaker at 4°C for 1 h before centrifugation to collect the supernatant for subsequent experiments. The protein concentration in the supernatant was measured using a bicinchoninic acid (BCA) assay kit (Solarbio) on a microplate reader (Thermo Fisher Scientific) at a wavelength of 562 nm.

Subsequently, equal amounts of sample protein and protein markers were separated by 10% sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) and transferred to polyvinylidene difluoride (PVDF) membranes (MilliporeSigma, St. Louis, USA). After blocking with 5% bovine serum albumin (BSA; Servicebio, Wuhan, China) for 1 h, the membranes were incubated with primary antibodies overnight at 4°C. After washing 3 times with Tris-buffered saline with Tween (TBST), the membranes were incubated with secondary antibodies at room temperature for 2 h. Details of the primary and secondary antibodies are listed in Table 1. Protein bands were visualized using an enhanced chemiluminescence reagent (Yeasen, Shanghai, China) and imaged with a Bio-Rad scanner (Bio-Rad, Hercules, USA). *GAPDH* was used as the housekeeping gene. Band analysis was performed using ImageJ v. 1.8.0 (National Institutes of Health (NIH), Bethesda, USA).

Table 1. Antibody information

Name	Type	Species	Source
Collagen II	primary antibody	rabbit	Proteintech
IRS2	primary antibody	mouse	Proteintech
MMP13	primary antibody	rabbit	Proteintech
SOX9	primary antibody	rabbit	Zenbio
COX2	primary antibody	rabbit	Zenbio
GAPDH	primary antibody	mouse	Proteintech
Goat anti-Rabbit IgG	secondary antibody	–	Proteintech
Goat anti-Mouse IgG	secondary antibodies	–	Proteintech

IgG – immunoglobulin G; MMP – matrix metalloproteinase.

Table 2. Primer information (all primers were designed for human genes)

Primer name	Primer sequence (5′-3′)
Collagen II	forward: TGGACGATCAGGCGAAACC
	reverse: GCTGCGGATGCTCTCAATCT
IRS2	forward: CGGTGAGTTCTACGGGTACAT
	reverse: TCAGGGTGTATTCATCCAGCG
MMP13	forward: ACTGAGAGGCTCCGAGAAATG
	reverse: GAACCCCGCATCTTGGCTT
SOX9	forward: AGCGAACGCACATCAAGAC
	reverse: CTGTAGGCGATCTGTTGGGG
COX2	forward: GCACCCCGACATAGAGAGC
	reverse: CTGCGGAGTGCAGTGTTCT
GAPDH	forward: TGTGGGCATCAATGGATTTGG
	reverse: ACACCATGTATTCCGGGTCAAT

Similarly, total RNA was extracted from the treated chondrocytes using TRIzol reagent (Takara, Shiga, Japan) and purified with an RNA extraction kit (TransGen, Beijing, China). The total RNA concentration of the samples was measured using a micro-UV spectrophotometer (Thermo Fisher Scientific). The RNA samples were then reverse transcribed into cDNA using a reverse transcription kit (TransGen). Finally, the target mRNA was amplified on a real-time quantitative polymerase chain reaction (qPCR) instrument (Bio-Rad). The final quantitative results were normalized to *GAPDH* and calculated using the $2^{-\Delta\Delta Ct}$ method. The primer sequences for the target genes are shown in Table 2.

Animal experiments

All animal procedures were approved by the Institutional Animal Care and Use Committee (IACUC) of The First Affiliated Hospital of Nanchang University, China (approval No. CDYFY-IACUC-202505GR093). All experiments involving animals were conducted in strict accordance with the guidelines set forth by the IACUC and the Guide for the Care and Use of Laboratory Animals. Mice were maintained under

specific pathogen-free (SPF) conditions in a 12 h light/dark cycle, with ad libitum access to food and water.

To investigate the differential expression of IRS2 in OA, a total of 12 ten-week-old male C57BL/6 mice were subjected to destabilization of the medial meniscus (DMM) surgery on the right knee.[35] Briefly, 12 mice were used, with 6 mice per group (sham and OA). The OA model was established by performing DMM surgery on the right knee. Briefly, mice were anesthetized with sodium pentobarbital, and the DMM procedure involved transecting the medial meniscotibial ligament (MMTL) to destabilize the meniscus. Sham-operated mice underwent identical surgical exposure without MMTL transection. The DMM model is a validated and widely used method for inducing OA in mice. Mice were euthanized 8 weeks after surgery, and whole knee joints were harvested from both groups for subsequent histological and molecular analyses.

Immunohistochemistry

Bone tissues were fixed in formalin and decalcified in ethylenediaminetetraacetic acid (EDTA) solution for 1 month, with the solution refreshed every 3–7 days. After decalcification, tissues were dehydrated, embedded in paraffin, and sectioned at a thickness of 5 µm. Sections were deparaffinized in xylene, rehydrated through graded ethanol, and subjected to antigen retrieval by boiling in 0.01% sodium citrate buffer. Endogenous peroxidase activity was blocked using hydrogen peroxide, followed by blocking with 5% BSA for 1 h at room temperature.

Primary antibodies diluted in 5% BSA were incubated overnight at 4°C, including anti-IRS2, anti-COX2, and anti-type II collagen (all from Sanying, Wuhan, China; dilution 1:50). The next day, sections were incubated with a horseradish peroxidase (HRP)-conjugated goat anti-mouse/rabbit IgG secondary antibody (Golden Bridge Biotechnology, Beijing, China; 1:50 dilution), followed by 3,3′-diaminobenzidine (DAB) staining and hematoxylin counterstaining to visualize nuclei. Finally, slides were mounted with neutral resin and coverslipped. Images of the stained sections were captured using a digital imaging system (Shunyu Instruments, Zhejiang, China).

Statistical analyses

Data were analyzed using R 4.3.0 (R Foundation for Statistical Computing, Vienna, Austria), Cytoscape 3.9.1, Perl 5.32.1 (https://www.perl.org), and GraphPad Prism 9.7 (GraphPad Software, San Diego, USA). Continuous data were expressed as mean ± standard deviation (SD). For comparisons between 2 groups, if the continuous data followed a normal distribution and had equal variances, Student's t-test was used. If the data did not follow a normal distribution, a nonparametric test (Mann–Whitney U test) was employed. All experiments were independently repeated 3 times, and $p < 0.05$ was considered statistically significant.

Results

Identification of DEGs

Table 3 summarizes the characteristics of the included datasets. Using the RRA method, we identified 202 upregulated and 49 downregulated DEGs (Fig. 1A). A heatmap showing the top 50 upregulated and top 10 downregulated genes is presented in Supplementary Fig. 2A. Figure 1B shows the differential gene expression profile in the merged dataset, revealing 64 upregulated and 30 downregulated genes. By intersecting the RRA and SVA results, we identified 83 key DEGs associated with OA (Fig. 1C).

Functional enrichment and pathway analysis of key genes

We performed GO, KEGG, and DO enrichment analyses to explore the biological roles of the key genes. The GO results showed that these genes are mainly involved in ECM

Table 3. Characteristics of the included datasets

GEO dataset	Platform	Tissue	Total samples	Normal	OA	Group
GSE51588	GPL13497	subchondral bone	50	10	40	training
GSE12021	GPL96	synovium	19	9	10	training
GSE55457	GPL96	synovium	20	10	10	training
GSE56409	GPL570	synovium	22	11	11	training
GSE114007	GPL11154/GPL18573	cartilage	38	18	20	training
GSE168505	GPL16791	cartilage	7	3	4	training
GSE169077	GPL96	cartilage	11	5	6	training
GSE55235	GPL96	synovium	20	10	10	test
GSE129147	GPL6947	cartilage	40	7	33	test
GSE48556	GPL6947	blood	139	33	106	test

GEO – Gene Expression Omnibus; OA – osteoarthritis.

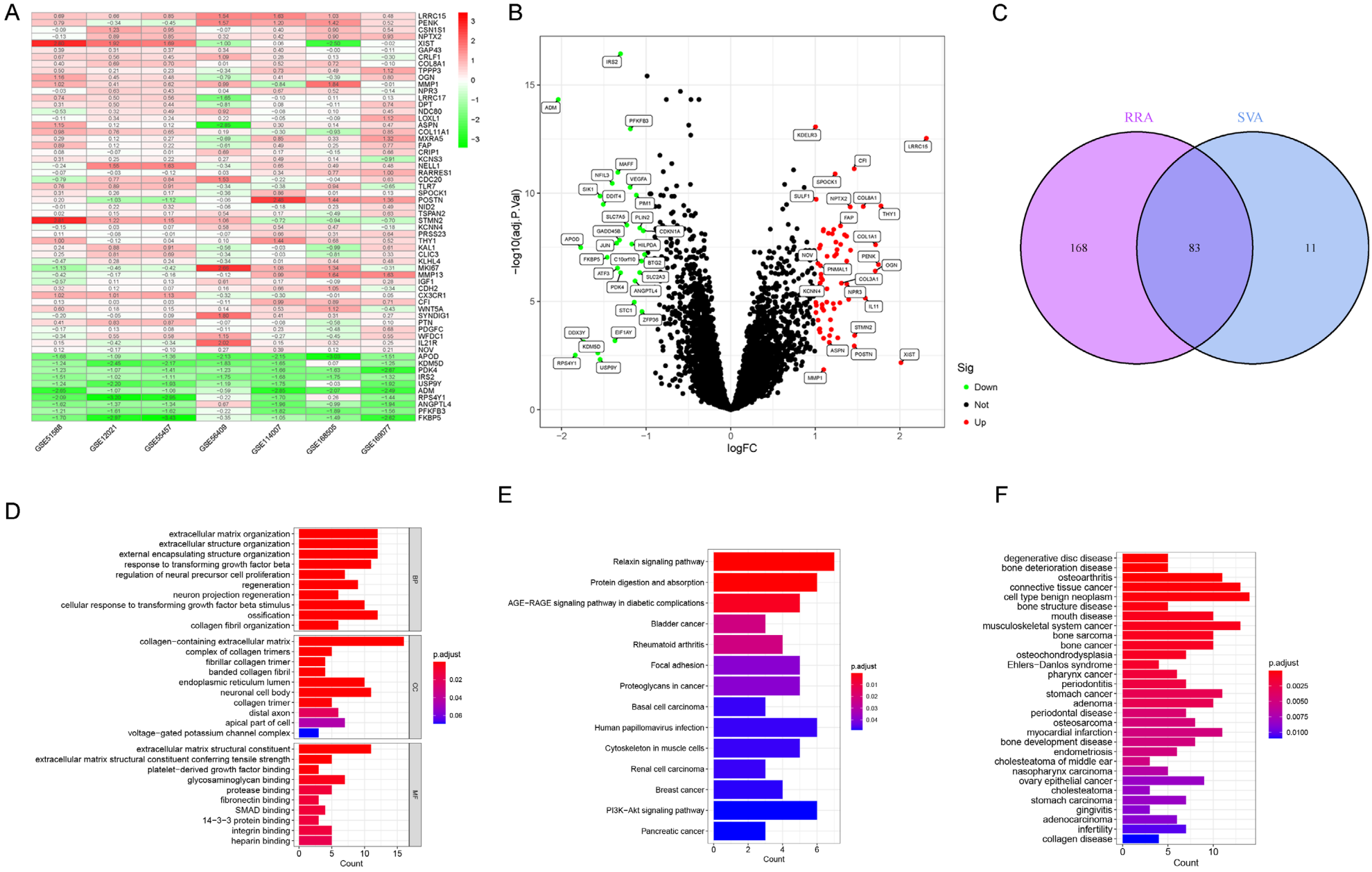

Fig. 1. Identification and enrichment analysis of differentially expressed genes (DEGs) in osteoarthritis (OA). A. Robust Rank Aggregation (RRA) heatmap of the top 50 upregulated and top 10 downregulated genes across 7 datasets (red: upregulated; green: downregulated); B. Volcano plot of DEGs; C. Venn diagram showing the overlap of DEGs identified by RRA and Surrogate Variable Analysis (SVA); D. Gene Ontology (GO) enrichment bar plots; E. Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment bar plots; F. Disease Ontology (DO) enrichment bar plots

and structural organization (BP), are localized to collagen-containing extracellular components (CC), and function as structural constituents of the ECM (MF) (Fig. 1D, Supplementary Fig. 2B). KEGG analysis revealed enrichment in pathways such as relaxin signaling, protein digestion and absorption, advanced glycation end products–receptor for advanced glycation end products (AGE-RAGE) signaling in diabetic complications, and rheumatoid arthritis (Fig. 1E, Supplementary Fig. 2C). DO analysis indicated associations with diseases including OA, degenerative disc disease, connective tissue cancer, and musculoskeletal system tumors (Fig. 1F, Supplementary Fig. 2D).

Construction of PPI network for key genes

We used the key genes to construct a PPI network in STRING (minimum interaction score = 0.4, $p < 0.001$) to explore potential protein interactions. The resulting network was visualized using Cytoscape (Fig. 2A). To identify hub genes, we applied all algorithms within the cytoHubba plugin, and only genes identified by all methods were retained, resulting in 25 hub genes (Fig. 2B), including *COL1A2*, *COL3A1*, *POSTN*, *WNT5A*, and *IRS2*, among others. We also generated volcano plots and heatmaps of hub gene expression (Fig. 3A,B).

Further enrichment analysis showed that the hub genes participate in biological processes such as ECM organization, structural organization, axon development, and regeneration. They are localized to collagen-containing ECM components, including the fibrillar collagen trimer and endoplasmic reticulum (ER) lumen. Molecular functions included ECM structural constituent activity, integrin binding, and growth factor binding (Fig. 3C). KEGG analysis revealed enrichment in the relaxin, IL-17, AGE-RAGE, and PI3K-Akt signaling pathways, as well as protein digestion and absorption (Fig. 3D).

Identifying diagnostic biomarkers through machine learning

To identify key diagnostic biomarkers for OA, we conducted a thorough feature selection process. Initially, we employed the `train` function from the caret package to identify the optimal ML algorithms. The dataset was divided into training and validation sets, with training set proportions of either 0.7 or 0.8. Among the 6 ML methods evaluated, XGBoost, SVM, and LASSO regression consistently ranked highly in terms of residuals and root mean square error (Supplementary Fig. 3A,B), with AUCs exceeding 0.9 (Supplementary Fig. 3C,D), indicating superior diagnostic performance and learning capability. Consequently, we selected XGBoost, SVM, and LASSO regression to screen for key diagnostic biomarkers for OA.

Using the XGBoost algorithm, we identified 23 candidate genes (Fig. 4A). The SVM algorithm identified 23 key genes (Fig. 4B,C), and LASSO regression selected 26 candidate genes (Fig. 4D,E). By intersecting the feature genes identified by these 3 ML methods, we identified 15 candidate genes: *IRS2*, *ADM*, *SIK1*, *PTN*, *CX3CR1*, *WNT5A*, *IL21R*, *APOD*, *CRLF1*, *FKBP5*, *PNMAL1*, *NPR3*, *RARRES1*, ASPN, and *POSTN*. Further intersecting these candidate genes with the hub genes, we identified *IRS2*, *PTN*, *POSTN*, and *WNT5A* as diagnostic biomarkers for OA (Fig. 4F).

The functional annotations, interaction partners, and network features of the final diagnostic biomarkers (*IRS2*, *PTN*, *POSTN*, and *WNT5A*) are summarized in Supplementary Table 1.

Next, we constructed box plots for these candidate genes, which demonstrated differential expression within the dataset. We used ROC curves to evaluate the specificity and sensitivity of these 4 features for distinguishing OA from normal tissues in the test set. The diagnostic values of the 4 genes were as follows: *IRS2* (AUC = 0.879, 95% CI: 0.826–0.927), *PTN* (AUC = 0.784, 95% CI: 0.707–0.852), *POSTN* (AUC = 0.668, 95% CI: 0.583–0.748), and *WNT5A* (AUC = 0.783, 95% CI: 0.706–0.857) (Fig. 5A). Since the AUC of POSTN in the test dataset was 0.668, which is below 0.7, we ultimately selected IRS2, PTN, and WNT5A as diagnostic biomarkers for OA. We then constructed a diagnostic nomogram based on the training dataset to develop a clinically applicable diagnostic model for OA (Fig. 5B). The calibration curve (Fig. 5C) and decision curve (Fig. 5D) clearly demonstrated the model's high predictive ability for OA (AUC = 0.913, 95% CI: 0.866–0.953; Fig. 5E).

Validation of relevant screening genes

To better validate the clinical diagnostic capability of the candidate genes, we used the cartilage tissue dataset GSE129147, the synovial tissue dataset GSE55235, and the blood sample dataset GSE48556 from OA patients for validation, with sample sizes of 40, 20, and 139, respectively. We constructed violin plots for the candidate genes, which demonstrated differential expression across the datasets (Fig. 6A–C). Additionally, we plotted ROC curves to evaluate the diagnostic value of each gene in different tissue samples (Fig. 6D–F).

Among these genes, *WNT5A* had lower diagnostic efficiency in cartilage and blood samples, *PTN* had poor diagnostic efficiency in blood samples, and only *IRS2* exhibited high differential expression ($p < 0.01$) in all sample types, together with good diagnostic performance, indicating high clinical value.

Experimental validation of *IRS2* gene

To validate the reliability and potential role of the candidate gene *IRS2* in OA, we established a DMM-induced OA model in C57BL/6 mice. Successful model

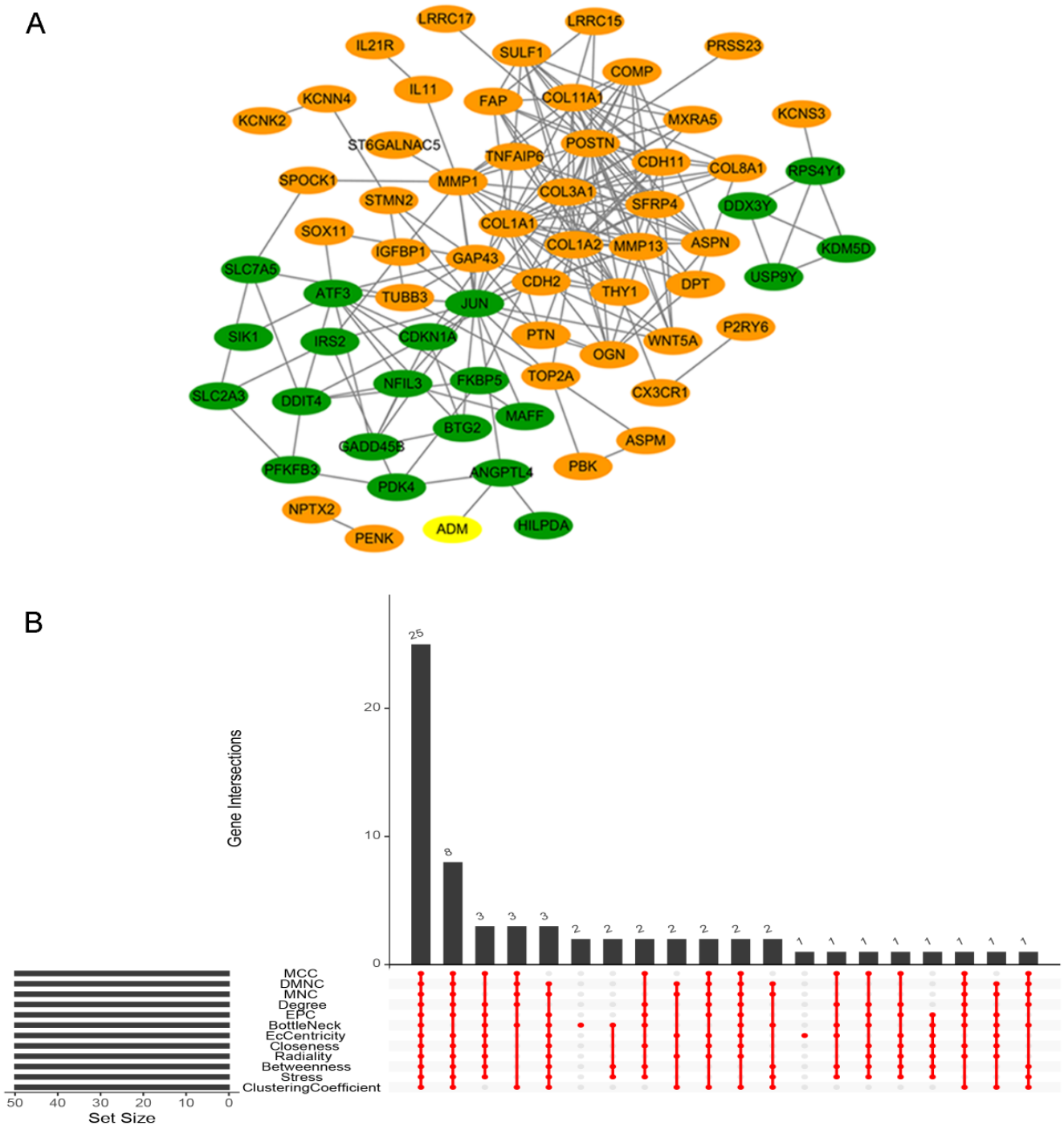

Fig. 2. Protein–protein interaction (PPI) network and hub gene identification. A. PPI network of hub genes (orange: upregulated; green: downregulated); B. Twenty-five hub genes identified using cytoHubba, with scores based on algorithm overlap

establishment was confirmed through Safranin O–Fast Green and hematoxylin and eosin (H&E) staining, as well as by assessing the expression levels of the OA-related markers COX2 and type II collagen. Immunohistochemical staining for *IRS2* (Fig. 7A,B) revealed a significant downregulation of *IRS2* in OA tissues, consistent with our initial hypothesis.

In vitro, we mimicked the OA pathological environment by stimulating the human chondrocyte cell line SW1353 with IL-1β and examined the protein and mRNA expression levels of IRS2 using western blotting (WB) and qPCR. Additionally, to ensure the fidelity of OA modeling, we assessed key phenotypic markers associated with OA. The results showed that, compared to the control group, COL2A1 and SOX9 levels were significantly decreased, whereas MMP13 and COX2 levels were markedly increased in the OA model group, consistent with the typical characteristics of OA. Importantly, IRS2 expression at both

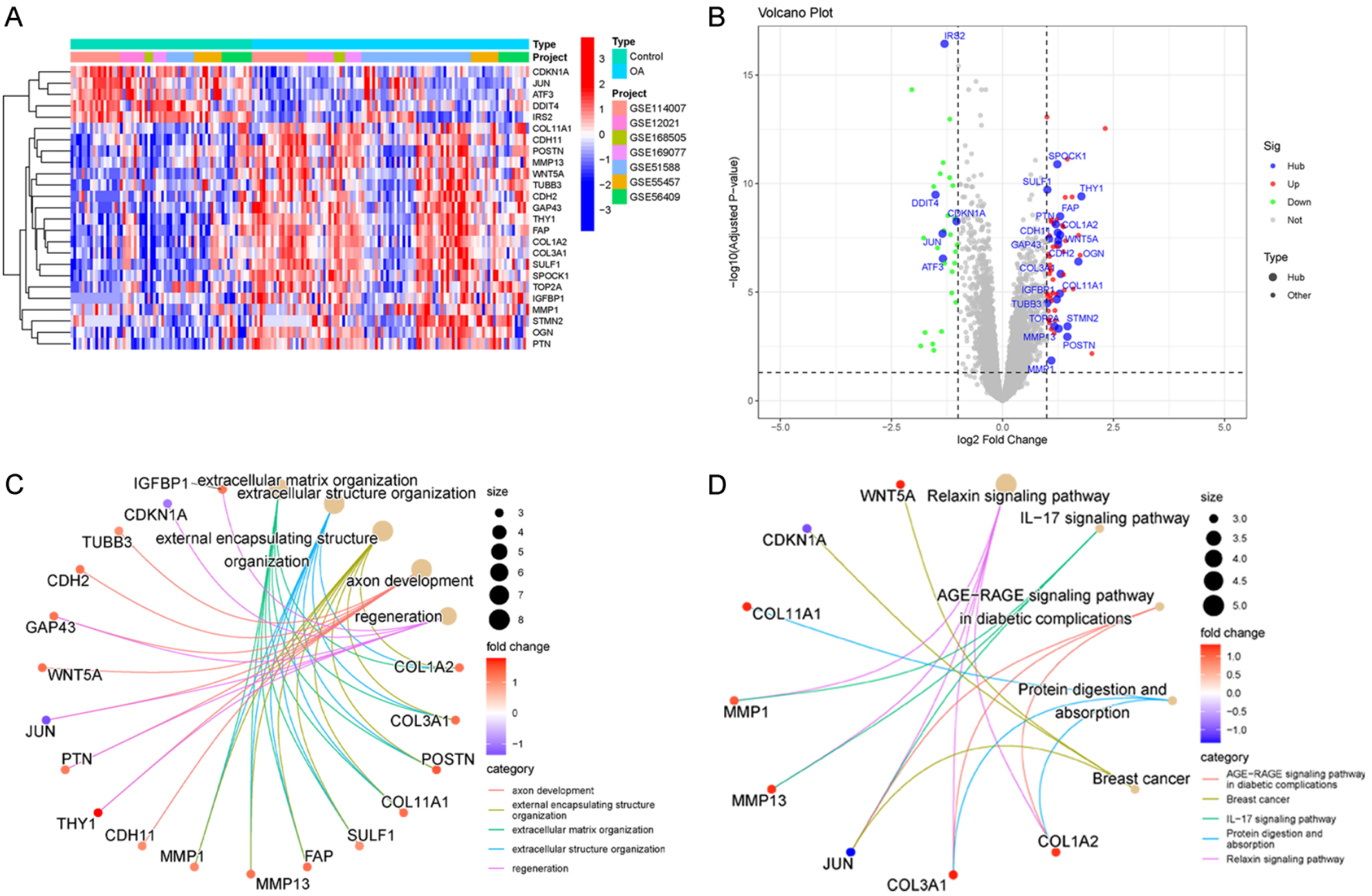

Fig. 3. Expression and functional enrichment analysis of hub genes. A. Heatmap of hub gene expression; B. Volcano plot of hub genes; C. Gene Ontology (GO) network plot; D. Kyoto Encyclopedia of Genes and Genomes (KEGG) network plot. In the network plots, term node size represents gene count; gene color indicates regulation (red: upregulated; purple: downregulated); node size reflects term involvement

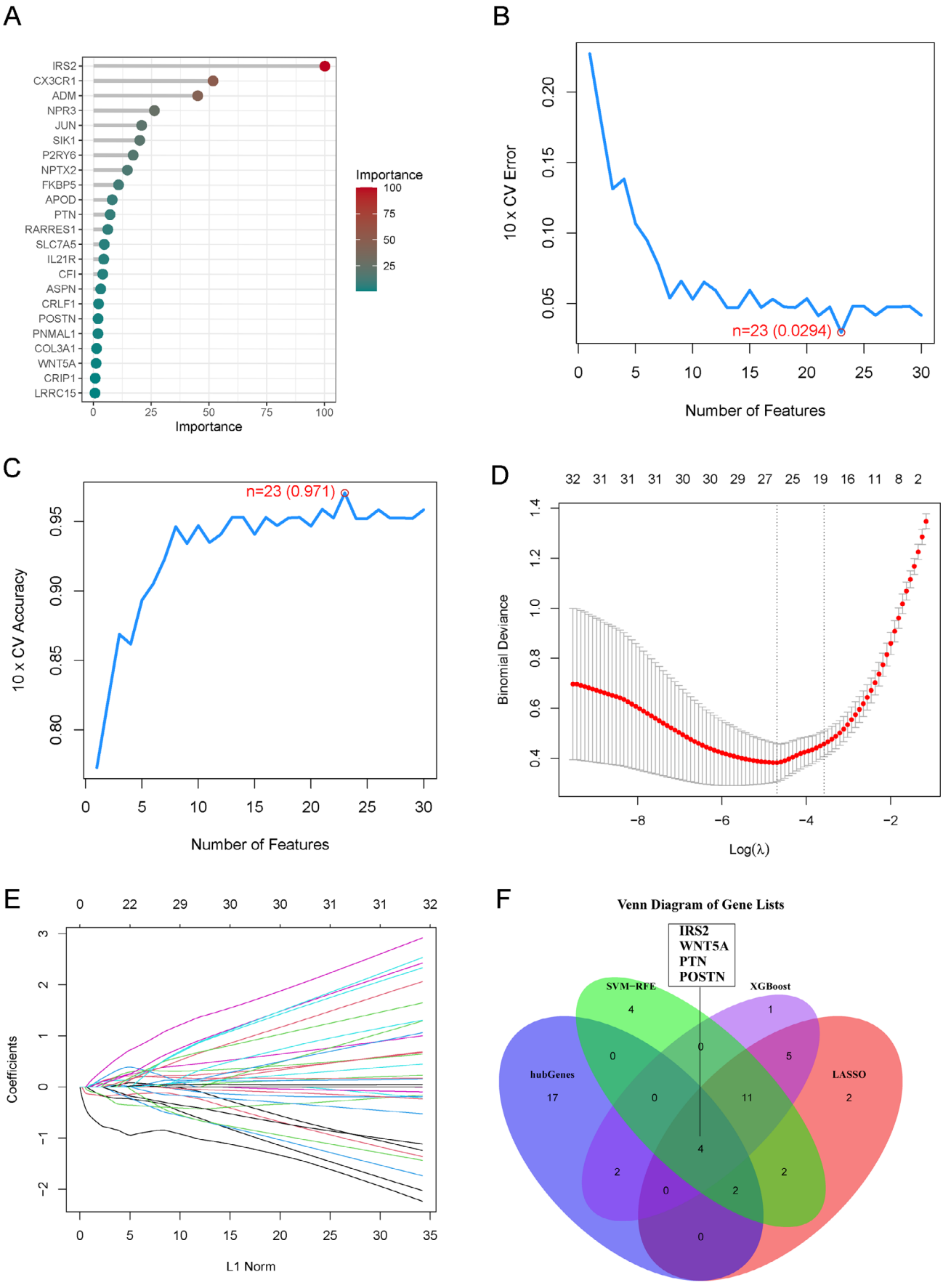

Fig. 4. Machine learning-based screening of osteoarthritis (OA) feature genes. A. Gene importance ranking by XGBoost; B,C. Support vector machine recursive feature elimination (SVM-RFE) accuracy and error plots; D. Least absolute shrinkage and selection operator (LASSO) coefficient analysis with the optimal lambda value; E. LASSO cross-validation curves; F. Venn diagram of feature genes identified by XGBoost, LASSO, and SVM-RFE

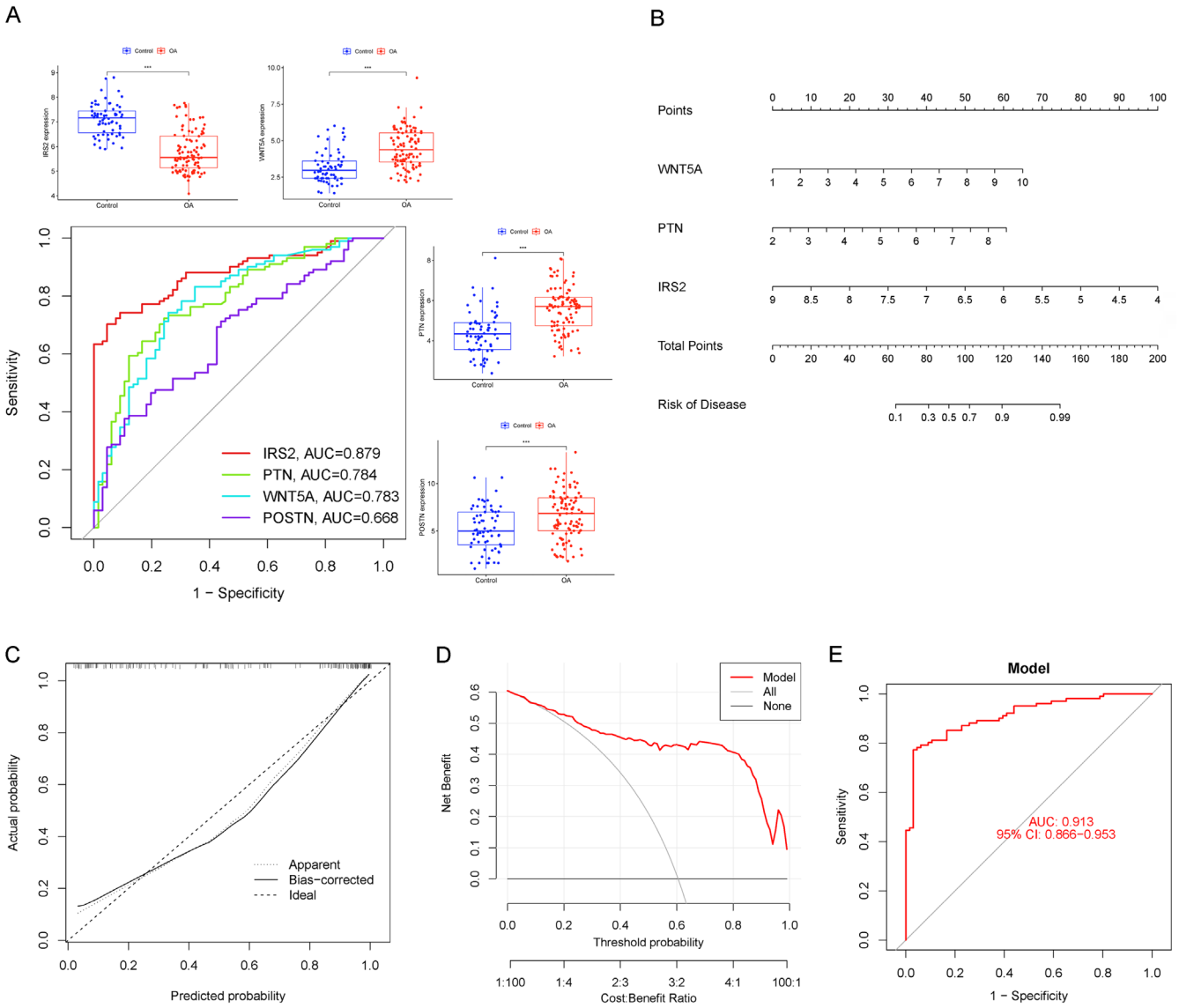

Fig. 5. Risk prediction model based on feature genes. A. Receiver operating characteristic (ROC) curves and expression boxplots of feature genes; B. Nomogram for osteoarthritis (OA) diagnosis; C. Calibration curve (closer to the dashed line = higher accuracy); D. Decision curve analysis (greater distance from the gray line = higher clinical utility); E. ROC curve of the prediction model (*p < 0.05; **p < 0.01; ***p < 0.001)

the protein and mRNA levels was significantly reduced in the OA group (Fig. 8).

To further elucidate the role of *IRS2* in OA, we constructed a plasmid-based *IRS2* overexpression system in SW1353 cells and again measured *IRS2* expression at both the protein and mRNA levels using WB and qPCR. Osteoarthritis-related phenotypic markers were reassessed to confirm the functional impact of *IRS2* overexpression. The data demonstrated that *IRS2* overexpression significantly restored the expression of COL2A1 and SOX9 while suppressing the upregulation of MMP13 and COX2. Compared to the OA group, the *IRS2* overexpression group exhibited significantly increased protein and mRNA levels of *IRS2*, along with a marked recovery of OA-related phenotypes, with statistically significant differences (Fig. 9).

Discussion

Based on the above integrative analysis, we further interpreted the biological significance and potential mechanisms underlying OA. In this study, we explored and validated OA-related biomarkers across multiple tissues using bioinformatics and ML algorithms, aiming to provide new insights into the mechanisms of OA from multiple perspectives. First, we identified 83 DEGs from multiple tissue datasets using the RRA and SVA methods. Functional enrichment, pathway analysis, and DO enrichment analysis of these DEGs revealed their involvement in ECM organization, extracellular structure organization, the relaxin signaling pathway, the PI3K-Akt signaling pathway, and diseases such as connective tissue cancer and musculoskeletal system cancer.

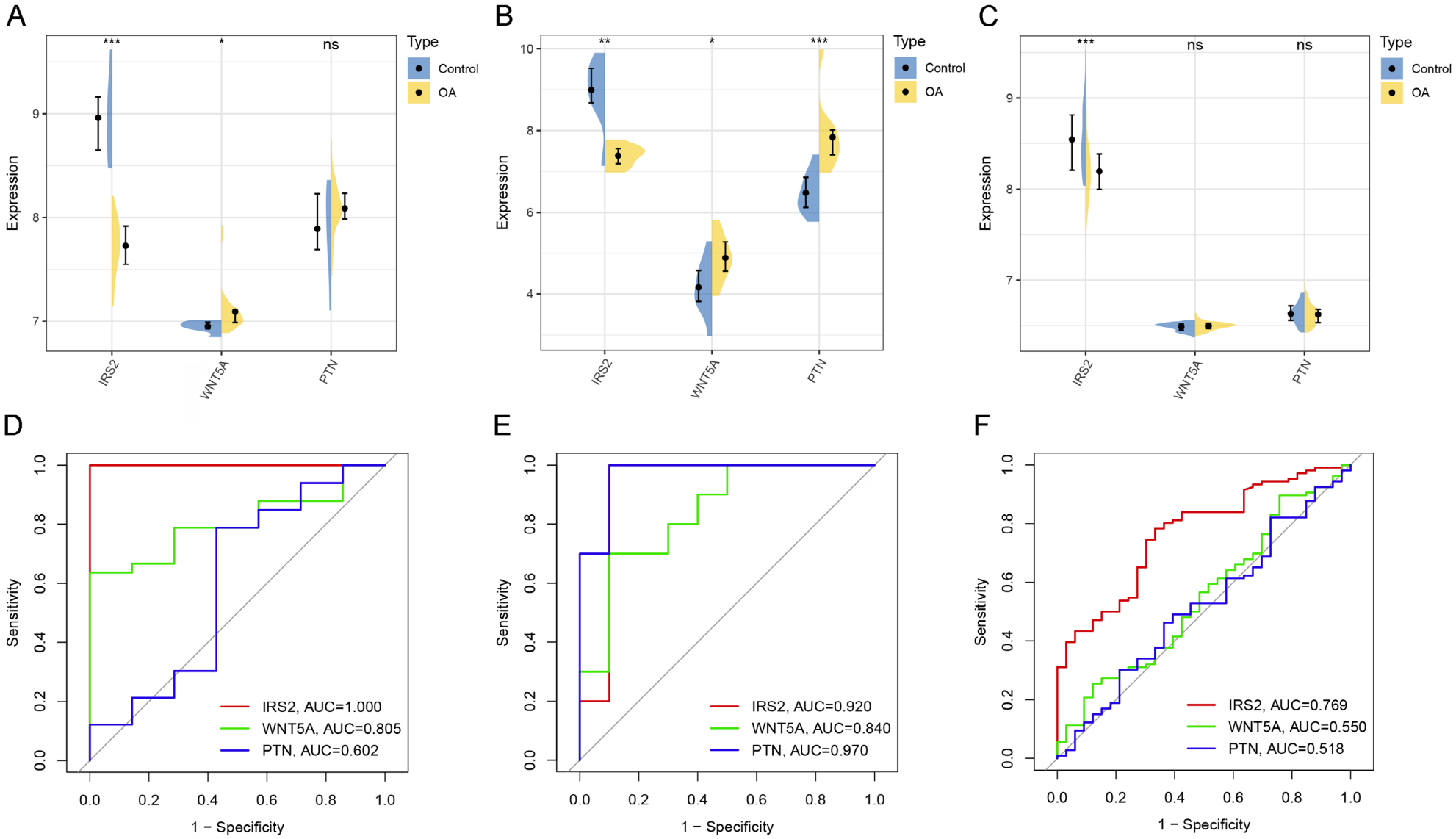

Fig. 6. Validation of feature genes in external datasets. A–C. Violin plots of gene expression in cartilage, synovial, and blood samples (*p < 0.05; **p < 0.01; ***p < 0.001); D–F. Receiver operating characteristic (ROC) curves in the corresponding sample types

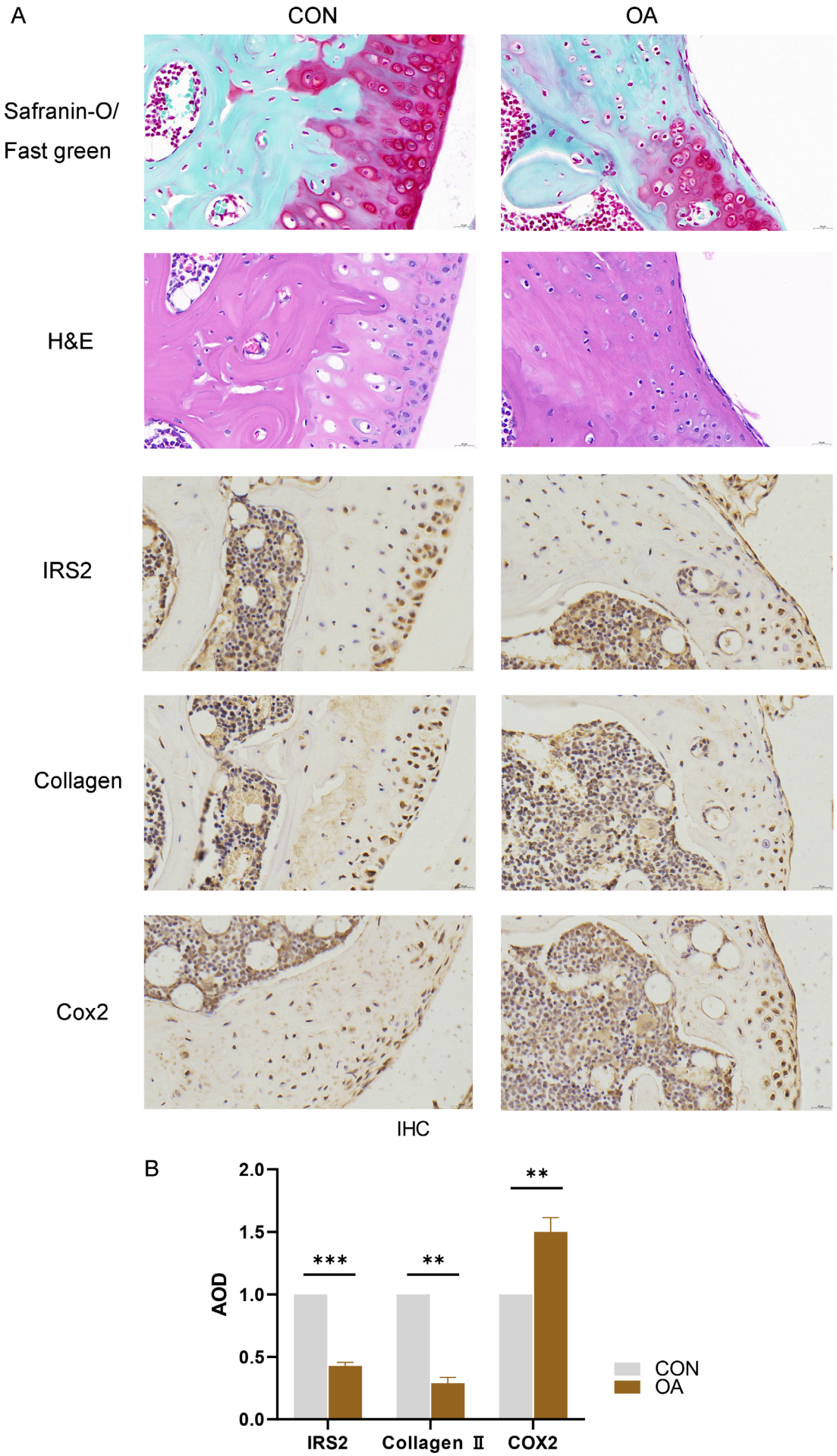

Fig. 7. Histological analysis of IRS2 expression in osteoarthritis (OA). A. Hematoxylin and eosin (H&E), Safranin O–Fast Green staining, and immunohistochemistry (IHC) for collagen II, COX2, and IRS2; B. Quantification of IHC (**p < 0.01; ***p < 0.001; n = 6)

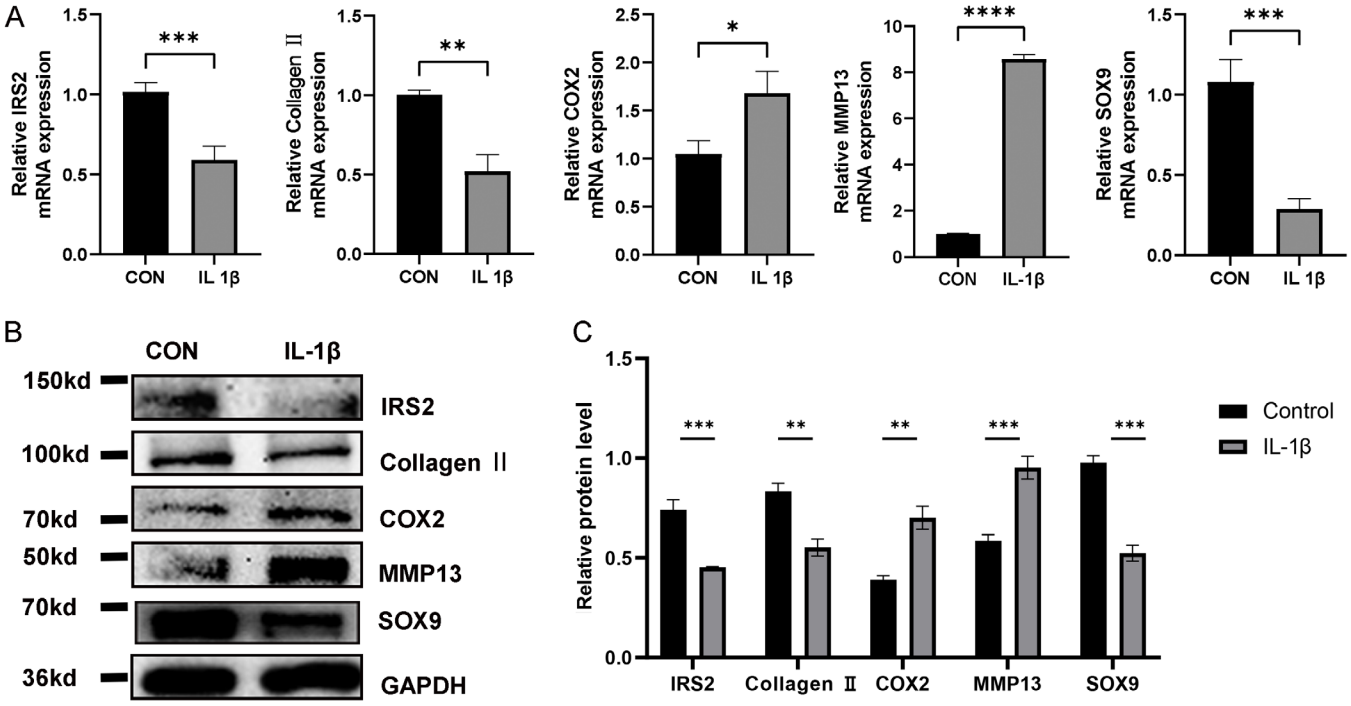

Fig. 8. IRS2 downregulation in osteoarthritis (OA). A. Real-time quantitative polymerase chain reaction (qPCR) analysis of collagen II, SOX9, COX2, and MMP13 in control and OA groups; B. Western blot analysis of the same markers; C. Quantified protein levels (mean ± standard deviation (SD); *p < 0.05; **p < 0.01; ***p < 0.001; n = 3)

Next, we constructed a PPI network and identified hub genes from the DEGs. Our methods identified 25 hub genes associated with OA. Functional enrichment and pathway analysis of these hub genes revealed their involvement in ECM organization, extracellular structure organization, the relaxin signaling pathway, the IL-17 signaling pathway, the AGE-RAGE signaling pathway in diabetic complications, and the PI3K-Akt signaling pathway. These pathways and functions were highly consistent with those identified for the DEGs, demonstrating the representative value of the identified hub genes. Additionally, these pathways and enrichment results are consistent with previous studies.[22,36–38]

To better identify OA-specific biomarkers, we employed ML algorithms for screening. Initially, we divided the dataset into training and test sets and evaluated the performance of 6 commonly used ML methods. We found that, regardless of the proportion of the training set, XGBoost, SVM, and LASSO regression demonstrated superior performance in identifying OA biomarkers and exhibited higher diagnostic accuracy, significantly outperforming the other ML methods. Consequently, we selected these 3 algorithms for screening OA-specific biomarkers and identified 15 candidates: *IRS2*, *ADM*, *SIK1*, *PTN*, *CX3CR1*, *WNT5A*, *IL21R*, *APOD*, *CRLF1*, *FKBP5*, *PNMAL1*, *NPR3*, *RARRES1*, *ASPN*, and *POSTN*. By intersecting these genes with the identified hub genes, we identified *IRS2*, *WNT5A*, *PTN*, and *POSTN* as diagnostic biomarkers for OA.

Although POSTN showed statistically significant differences in OA ($p < 0.01$), its ROC curve yielded an AUC below 0.7, indicating poor diagnostic performance. Thus, we constructed an OA risk prediction model based on the remaining 3 genes. The constructed and validated risk score nomogram was able to distinguish OA from normal tissues, with an AUC of 0.913, indicating high diagnostic accuracy. We validated these diagnostic genes using external datasets of OA cartilage, synovial, and blood samples. In the cartilage dataset, PTN did not show significant expression differences and had an AUC below 0.7. In the synovial tissue dataset, POSTN did not exhibit differential expression compared to the control group. However, in OA blood samples, IRS2 showed statistically significant expression differences and exhibited differential expression across all 3 tissue types ($p < 0.01$). Its AUC was consistently above 0.7, demonstrating good diagnostic performance. Thus, we have reason to believe that IRS2 may be involved in multiple aspects of OA progression, highlighting its potential as a promising diagnostic biomarker for OA.

Periostin (POSTN) is a 90 kDa matricellular protein discovered in 1993,39 involved in the pathogenesis of various diseases, including tumors, pulmonary fibrosis, and allergic diseases, with expression levels increasing as disease progresses in most cases.[39–41] POSTN interacts with ECM proteins to regulate cell–matrix organization, leading to remodeling and fibrosis. The unique characteristics of POSTN can be attributed to highly complex signaling pathways that lead to increased POSTN production.[39] In recent years, its association with orthopedic diseases has also become apparent.[42] Numerous studies have confirmed elevated POSTN expression in OA cartilage,[43–46] although its expression in OA synovial tissue and peripheral blood

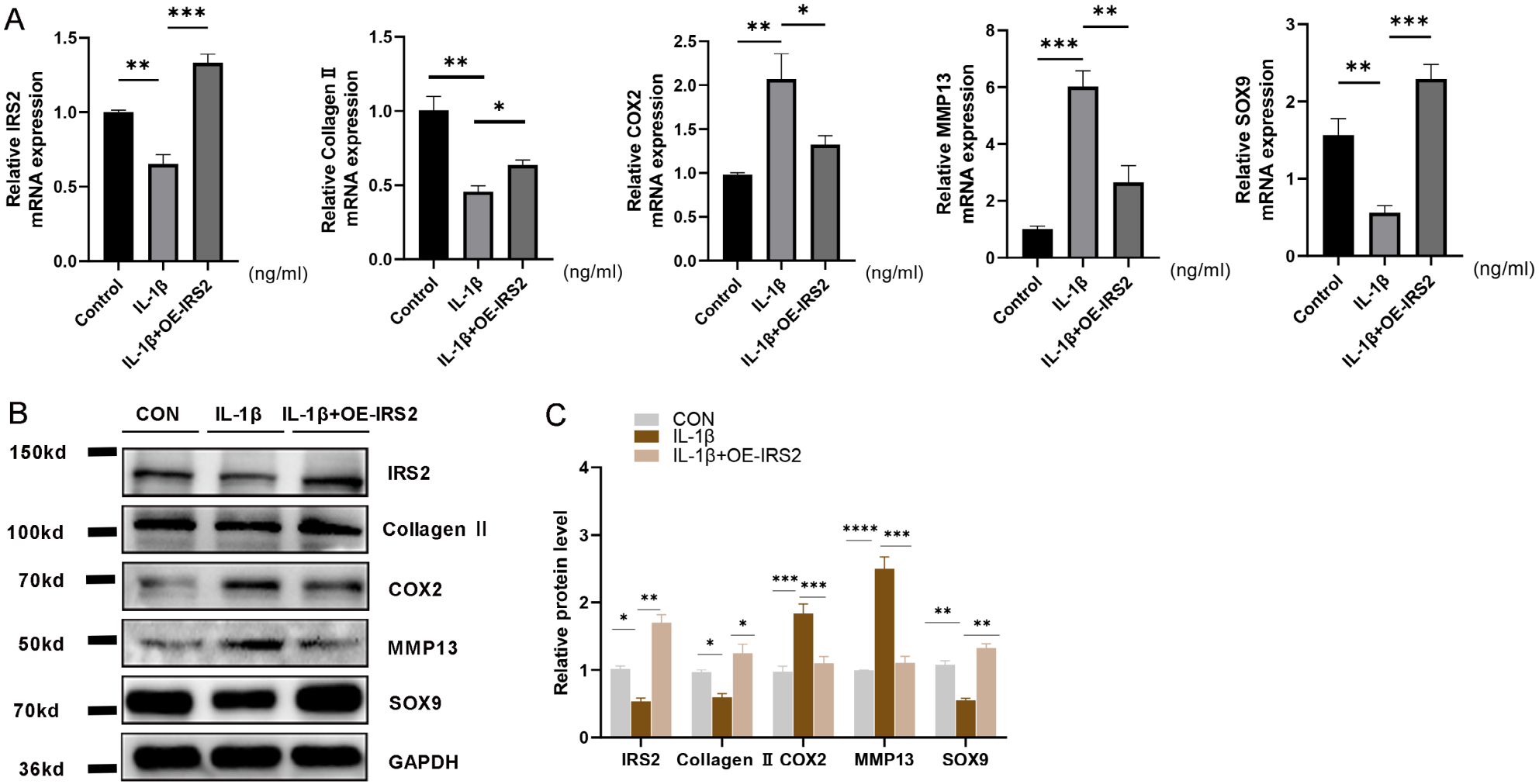

Fig. 9. IRS2 overexpression mitigates the effects of interleukin-1 beta (IL-1β) in chondrocytes. A. Real-time quantitative polymerase chain reaction (qPCR) analysis of markers after IRS2 overexpression and IL-1β treatment; B. Western blot analysis of the same markers; C. Representative blots (mean ± standard deviation (SD); *p < 0.05; **p < 0.01; ***p < 0.001; ****p < 0.0001; n = 3)

remains controversial. For example, Tajika et al.[47] reported high POSTN expression in OA synovial cells, while Attur et al.[48] observed no significant difference in expression between OA and control samples. Similarly, Rousseau et al[49] found serum POSTN to be associated with OA progression in women, but Sittisak Honsawek et al.[50] and Tan et al.[51] suggested that serum POSTN levels lack diagnostic value for OA. In our study, POSTN was consistently upregulated in OA cartilage samples but showed no significant changes in synovial or blood tissues. Although it was not ultimately selected as a diagnostic biomarker due to suboptimal performance (AUC < 0.7), its tissue-specific expression pattern may still reflect pathological remodeling in OA cartilage. These findings suggest that POSTN may play a role in disease progression, particularly in cartilage, and warrant further investigation in larger, well-characterized cohorts.

Pleiotrophin is a member of the midkine family,[52] a secreted heparin-binding peptide expressed during development in mesodermal and neuroectodermal cells but rarely in adult tissues.[53] Pufe et al.[54] found that PTN is almost undetectable in normal cartilage and synovial cells but is highly expressed in OA, with significant expression in the early and middle stages of OA but not in the late stage. Furthermore, Pufe et al.[55] suggested that PTN might participate in the early onset and development of OA by stimulating the activation of the activator protein-1 (AP-1) transcription factor and altering gene expression. In blood, studies by Fadda et al.[56] showed no significant difference in average PTN levels between OA patients and healthy controls, suggesting that while PTN may play an important role in OA, its potential as a disease biomarker requires larger-scale investigation and further research.

In our study, as shown in Fig. 6, PTN did not show statistically significant differences in the external cartilage tissue dataset, possibly because our cartilage tissue dataset represented late-stage OA. However, its high expression in the OA synovial tissue dataset is consistent with previous findings. Taken together, these results suggest that PTN may play a role in the diagnosis of early synovitis in OA.

WNT5A is a member of the Wnt protein family, which comprises a group of highly conserved signaling proteins that play crucial roles in embryonic development and various cellular processes, such as cell migration, polarity, and differentiation. WNT5A operates in the non-canonical Wnt signaling pathway, particularly influencing cell movement and polarity rather than directly affecting cell proliferation.[57] The high expression of WNT5A in OA cartilage has been confirmed in numerous studies, and its mechanisms of action in OA have been widely explored.[58–62] However, the role of WNT5A in OA synovium is less well understood. Lambert et al.[63] found high expression of WNT5A in OA synovium, regulated via the Wnt signaling pathway. This finding aligns with our conclusions. However, the expression of WNT5A in OA blood remains unstudied, necessitating further investigation.

Insulin receptor substrate 2 (IRS2) is a crucial member of the insulin receptor substrate family, playing a key role in insulin signaling and metabolic regulation. It primarily mediates signaling downstream of insulin and other growth factor receptors.[64] IRS2 is expressed in various cell types, including those in the liver, muscle, and adipose tissue, and is involved in multiple physiological processes.[65] The absence of IRS2 can lead to insulin resistance, which in turn promotes the development of type 2 diabetes.[66] Numerous studies have focused on the role of IRS2 in type 2 diabetes, obesity, and non-alcoholic fatty liver disease, exploring its molecular mechanisms.[65,66] Research has shown that the absence of IRS2 can have detrimental effects in various cell types and disease conditions.[67] However, the role of IRS2 in OA remains underexplored. In our study, we used external datasets from cartilage, synovial tissue, and OA blood samples for validation and observed consistently low expression of IRS2 across these tissues and sample types. Given the lack of in vivo and in vitro studies investigating IRS2 in OA, we conducted further validation through in vitro experiments. Western blotting and qPCR analyses revealed a significant reduction in IRS2 expression in OA chondrocytes, consistent with the findings from the datasets and statistically significant ($p < 0.001$). This suggests that IRS2 may play a role in the pathogenesis of OA.

Currently, the precise pathways through which IRS2 contributes to the progression of OA remain unclear. However, existing research suggests that IRS2 may influence disease progression by regulating inflammatory and apoptotic pathways. For instance, a study by Mayumi et al.[67] found that, under hypoxic conditions, high expression of IRS2 in macrophages contributes to anti-inflammatory effects in pulmonary vascular remodeling. Similarly, Baquedano et al.[68] observed increased oxidative stress and apoptosis in the hypothalamus of diabetic male mice with *IRS2* gene knockout. These studies indicate that *IRS2* might affect the progression of OA by modulating inflammation and apoptosis. Additionally, research by Kaede et al.[69] demonstrated that vitamin C enhances insulin-induced chondrocyte differentiation in ATDC5 cells by upregulating the expression of *IRS2* and other insulin signaling molecules. This further supports the potential role of *IRS2* in OA.

Future research should focus on elucidating the specific functions and regulatory mechanisms of IRS2 in OA and exploring its feasibility as a potential therapeutic target. Through further mechanistic studies and clinical validation, it is hoped that the role of *IRS2* in the pathological process of OA will be elucidated, offering new insights and approaches for the diagnosis and treatment of OA.

Limitations of the study

This study has several limitations. First, it relies on publicly available GEO datasets, which, despite offering rich biological information, may vary in quality, consistency, and

sequencing platforms – potentially affecting the reliability of the results despite batch effect correction (e.g., SVA). Second, although 3 diagnostic genes (*IRS2*, *WNT5A*, *PTN*) were identified and showed good performance in ROC analysis, their clinical utility requires further validation. Third, while multiple ML methods (XGBoost, LASSO, SVM) were applied, other models like deep learning were not explored, and model performance on external datasets needs further verification. Lastly, *IRS2* expression was only validated in chondrocytes; its role in other OA-relevant tissues (e.g., synovium, subchondral bone, blood) remains to be clarified. Future studies should include broader datasets and deeper functional validation to strengthen the findings.

Conclusions

This study integrates bioinformatics analysis and ML algorithms to identify and validate 3 diagnostic biomarkers for OA: *IRS2*, *WNT5A*, and *PTN*. IRS2 demonstrated high and consistent diagnostic performance across multiple tissue types and external datasets, highlighting its potential as a robust biomarker. WNT5A exhibited good discriminatory ability, particularly in cartilage and synovial tissues. Although PTN showed moderate diagnostic performance overall, its expression patterns suggest potential relevance for early OA diagnosis and warrant further investigation.

Supplementary data

The supplementary materials are available at https://doi.org/10.5281/zenodo.17409505. The package contains the following files:

Supplementary Table 1. Final diagnostic biomarkers for OA and their functional annotations, interacting genes, and network characteristics.

Supplementary Fig. 1. Comprehensive visualization of batch-effect correction and sample grouping.

Supplementary Fig. 2. Identification and enrichment analysis of DEGs in osteoarthritis.

Supplementary Fig. 3. Machine-learning screening of OA feature genes.

Data Availability Statement

The data from the GEO dataset can be downloaded here: https://www.ncbi.nlm.nih.gov/geo

Consent for publication

Not applicable

Use of AI and AI-assisted technologies

Not applicable.

Preprint disclosure

This paper was previously made available as a preprint on Research Square at https://www.researchsquare.com/article/rs-4706641/v1.

ORCID iDs

Guihao Zheng https://orcid.org/0000-0002-0850-1569
Yulong Ouyang https://orcid.org/0000-0002-0690-053
Shuilin Chen https://orcid.org/0009-0007-7533-0329
Bei Hu https://orcid.org/0009-0002-4900-3766
Shuai Xu https://orcid.org/0000-0007-1722-451X
Guicai Sun https://orcid.org/0000-0002-1306-7592

References

1. Vos T, Flaxman AD, Naghavi M, et al. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990–2010: A systematic analysis for the Global Burden of Disease Study 2010. *Lancet*. 2012;380(9859):2163–2196. doi:10.1016/S0140-6736(12)61729-2
2. Steinmetz JD, Culbreth GT, Haile LM, et al. Global, regional, and national burden of osteoarthritis, 1990–2020 and projections to 2050: A systematic analysis for the Global Burden of Disease Study 2021. *Lancet Rheumatol*. 2023;5(9):e508–e522. doi:10.1016/S2665-9913(23)00163-7
3. Puig-Junoy J, Ruiz Zamora A. Socio-economic costs of osteoarthritis: A systematic review of cost-of-illness studies. *Semin Arthritis Rheum*. 2015;44(5):531–541. doi:10.1016/j.semarthrit.2014.10.012
4. Peat G, Thomas MJ. Osteoarthritis year in review 2020: Epidemiology & therapy. *Osteoarthritis Cartilage*. 2021;29(2):180–189. doi:10.1016/j.joca.2020.10.007
5. Holt HL, Katz JN, Reichmann WM, et al. Forecasting the burden of advanced knee osteoarthritis over a 10-year period in a cohort of 60–64 year-old US adults. *Osteoarthritis Cartilage*. 2011;19(1):44–50. doi:10.1016/j.joca.2010.10.009
6. Cao M, Li C, Li M, et al. Prevalence and incidence of rheumatoid arthritis in urban China: A national population-based study. *Int J Rheum Dis*. 2025;28(5):e70184. doi:10.1111/1756-185X.70184
7. Guccione AA, Felson DT, Anderson JJ, et al. The effects of specific medical conditions on the functional limitations of elders in the Framingham Study. *Am J Public Health*. 1994;84(3):351–358. doi:10.2105/AJPH.84.3.351
8. Chen D, Shen J, Zhao W, et al. Osteoarthritis: Toward a comprehensive understanding of pathological mechanism. *Bone Res*. 2017;5(1):16044. doi:10.1038/boneres.2016.44
9. Katz JN, Arant KR, Loeser RF. Diagnosis and treatment of hip and knee osteoarthritis: A review. *JAMA*. 2021;325(6):568. doi:10.1001/jama.2020.22171
10. Xiong H, Huang TY, Chang YL, Su WT. Achyranthes bidentate extracts protect the IL-1β-induced osteoarthritis of SW1353 chondrocytes. *J Biosci Bioeng*. 2023;136(6):462–470. doi:10.1016/j.jbiosc.2023.09.008
11. Li X, Chen W, Liu D, et al. Pathological progression of osteoarthritis: A perspective on subchondral bone. *Front Med (Lausanne)*. 2024;18(2):237–257. doi:10.1007/s11684-024-1061-y
12. Yang D, Xu K, Xu X, Xu P. Revisiting prostaglandin E2: A promising therapeutic target for osteoarthritis. *Clin Immunol*. 2024;260:109904. doi:10.1016/j.clim.2024.109904
13. Wen Z, Sun Q, Shan Y, et al. Endoplasmic reticulum stress in osteoarthritis: A novel perspective on the pathogenesis and treatment. *Aging Dis*. 2022;14(2):283–286. doi:10.14336/AD.2022.0725
14. Wang Q, Rozelle AL, Lepus CM, et al. Identification of a central role for complement in osteoarthritis. *Nat Med*. 2011;17(12):1674–1679. doi:10.1038/nm.2543
15. 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
16. Filer A, Antczak P, Parsonage GN, et al. Stromal transcriptional profiles reveal hierarchies of anatomical site, serum response and disease and identify disease specific pathways. *PLoS One*. 2015;10(3):e0120917. doi:10.1371/journal.pone.0120917

17. Fisch KM, Gamini R, Alvarez-Garcia O, et al. Identification of transcription factors responsible for dysregulated networks in human osteoarthritis cartilage by global gene expression analysis. *Osteoarthritis Cartilage*. 2018;26(11):1531–1538. doi:10.1016/j.joca.2018.07.012
18. Xu S, Ye J, Cai X. Identification of energy metabolism-related subtypes and diagnostic biomarkers for osteoarthritis by integrating bioinformatics and machine learning. *J Multidiscip Healthc*. 2025;18: 1353–1369. doi:10.2147/JMDH.S510308
19. Dong S, Xia T, Wang L, Zhao Q, Tian J. Investigation of candidate genes for osteoarthritis based on gene expression profiles. *Acta Orthop Traumatol Turc*. 2016;50(6):686–690. doi:10.1016/j.aott.2016.04.002
20. Coppola C, Greco M, Munir A, et al. Osteoarthritis: Insights into diagnosis, pathophysiology, therapeutic avenues, and the potential of natural extracts. *Curr Issues Mol Biol*. 2024;46(5):4063–4105. doi:10.3390/cimb46050251
21. Luo H, Li L, Han S, Liu T. The role of monocyte/macrophage chemokines in pathogenesis of osteoarthritis: A review. *Int J Immunogenet*. 2024;51(3):130–142. doi:10.1111/iji.12664
22. Zhou J, Huang J, Li Z, 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
23. Han Y, Wu J, Gong Z, et al. Identification and development of a novel 5-gene diagnostic model based on immune infiltration analysis of osteoarthritis. *J Transl Med*. 2021;19(1):522. doi:10.1186/s12967-021-03183-9
24. Yin W, Lei Y, Yang X, Zou J. A two-gene random forest model to diagnose osteoarthritis based on RNA-binding protein-related genes in knee cartilage tissue. *Aging*. 2023;15(1):193–212. doi:10.18632/aging.204469
25. Chen Z, Wang W, Zhang Y, Xue X, Hua Y. Identification of four-gene signature to diagnose osteoarthritis through bioinformatics and machine learning methods. *Cytokine*. 2023;169:156300. doi:10.1016/j.cyto.2023.156300
26. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw*. 2010;33(1):1–22. PMID:20808728. PMCID:PMC2929880.
27. Luts J, Ojeda F, Van De Plas R, De Moor B, Van Huffel S, Suykens JAK. A tutorial on support vector machine-based methods for classification problems in chemometrics. *Anal Chim Acta*. 2010;665(2):129–145. doi:10.1016/j.aca.2010.03.030
28. Decup F, Léger S, Lefèvre S, Doméjean S, Grosgogeat B. Risk factors or indicators for dental caries and tooth wear and their relative importance in adults according to age. *J Dent*. 2024;147:105092. doi:10.1016/j.jdent.2024.105092
29. Hart CR, Wilson DK, Pettit CL, Nykaza ET. Machine-learning of long-range sound propagation through simulated atmospheric turbulence. *J Acoust Soc Am*. 2021;149(6):4384–4395. doi:10.1121/10.0005280
30. Islam SMS, Talukder A, Awal MdA, et al. Machine learning approaches for predicting hypertension and its associated factors using population-level data from three South Asian countries. *Front Cardiovasc Med*. 2022;9:839379. doi:10.3389/fcvm.2022.839379
31. Xiao J, Luo C, Li A, et al. Icariin inhibits chondrocyte ferroptosis and alleviates osteoarthritis by enhancing the SLC7A11/GPX4 signaling. *Int Immunopharmacol*. 2024;133:112010. doi:10.1016/j.intimp.2024.112010
32. Chen Y, Pan X, Zhao J, et al. Icariin alleviates osteoarthritis through PI3K/Akt/mTOR/ULK1 signaling pathway. *Eur J Med Res*. 2022;27(1):204. doi:10.1186/s40001-022-00820-x
33. Mu Y, Wang L, Fu L, Li Q. Knockdown of LMX1B suppressed cell apoptosis and inflammatory response in IL-1β-induced human osteoarthritis chondrocytes through NF-κB and NLRP3 signal pathway. *Mediators Inflamm*. 2022;2022:1870579. doi:10.1155/2022/1870579
34. Xu H, Chen S, Meng C, He Y, Huang XJ, You HB. Inhibition of CC chemokine receptor 1 ameliorates osteoarthritis in mouse by activating PPAR-γ. *Mol Med*. 2024;30(1):74. doi:10.1186/s10020-024-00823-w
35. Glasson SS, Blanchet TJ, Morris EA. The surgical destabilization of the medial meniscus (DMM) model of osteoarthritis in the 129/SvEv mouse. *Osteoarthritis Cartilage*. 2007;15(9):1061–1069. doi:10.1016/j.joca.2007.03.006
36. Sun K, Luo J, Guo J, Yao X, Jing X, Guo F. The PI3K/AKT/mTOR signaling pathway in osteoarthritis: A narrative review. *Osteoarthritis Cartilage*. 2020;28(4):400–409. doi:10.1016/j.joca.2020.02.027
37. Xiao J, Zhang P, Cai FL, et al. IL-17 in osteoarthritis: A narrative review. *Open Life Sci*. 2023;18(1):20220747. doi:10.1515/biol-2022-0747
38. Wang K, Li Y, Lin J. Identification of diagnostic biomarkers for osteoarthritis through bioinformatics and machine learning. *Heliyon*. 2024; 10(6):e27506. doi:10.1016/j.heliyon.2024.e27506
39. Ono J, Takai M, Kamei A, Azuma Y, Izuhara K. Pathological roles and clinical usefulness of periostin in type 2 inflammation and pulmonary fibrosis. *Biomolecules*. 2021;11(8):1084. doi:10.3390/biom11081084
40. Yu Y, Tan CM, Jia YY. Research status and the prospect of POSTN in various tumors. *Neoplasma*. 2021;68(4):673–682. doi:10.4149/neo_2021_210223N239
41. Sonnenberg-Riethmacher E, Miehe M, Riethmacher D. Periostin in allergy and inflammation. *Front Immunol*. 2021;12:722170. doi:10.3389/fimmu.2021.722170
42. Yoshihara T, Morimoto T, Hirata H, et al. Mechanisms of tissue degeneration mediated by periostin in spinal degenerative diseases and their implications for pathology and diagnosis: A review. *Front Med (Lausanne)*. 2023;10:1276900. doi:10.3389/fmed.2023.1276900
43. Han T, Mignatti P, Abramson SB, Attur M. Periostin interaction with discoidin domain receptor-1 (DDR1) promotes cartilage degeneration. *PLoS One*. 2020;15(4):e0231501. doi:10.1371/journal.pone.0231501
44. Chijimatsu R, Kunugiza Y, Taniyama Y, Nakamura N, Tomita T, Yoshikawa H. Expression and pathological effects of periostin in human osteoarthritis cartilage. *BMC Musculoskelet Disord*. 2015;16(1):215. doi:10.1186/s12891-015-0682-3
45. Duan X, Cai L, Pham CTN, et al. Amelioration of posttraumatic osteoarthritis in mice using intraarticular silencing of periostin via nanoparticle-based small interfering RNA. *Arthritis Rheum*. 2021;73(12): 2249–2260. doi:10.1002/art.41794
46. Attur M, Yang Q, Shimada K, et al. Elevated expression of periostin in human osteoarthritic cartilage and its potential role in matrix degradation via matrix metalloproteinase-13. *FASEB J*. 2015;29(10): 4107–4121. doi:10.1096/fj.15-272427
47. Tajika Y, Moue T, Ishikawa S, et al. Influence of periostin on synoviocytes in knee osteoarthritis. *In Vivo*. 2017;31(1):69–78. doi:10.21873/invivo.11027
48. Attur M, Duan X, Cai L, et al. Periostin loss-of-function protects mice from post-traumatic and age-related osteoarthritis. *Arthritis Res Ther*. 2021;23(1):104. doi:10.1186/s13075-021-02477-z
49. Rousseau JC, Sornay-Rendu E, Bertholon C, Garnero P, Chapurlat R. Serum periostin is associated with prevalent knee osteoarthritis and disease incidence/progression in women: The OFELY study. *Osteoarthritis Cartilage*. 2015;23(10):1736–1742. doi:10.1016/j.joca.2015.05.015
50. Honsawek S, Wilairatana V, Udomsinprasert W, Sinlapavilawan P, Jirathanathornnukul N. Association of plasma and synovial fluid periostin with radiographic knee osteoarthritis: Cross-sectional study. *Joint Bone Spine*. 2015;82(5):352–355. doi:10.1016/j.jbspin.2015.01.023
51. Tan Q, Yang Z, Xin X, et al. Serum periostin level is not sufficient to serve as a clinically applicable biomarker of osteoarthritis. *BMC Musculoskelet Disord*. 2022;23(1):1039. doi:10.1186/s12891-022-06017-x
52. Yazihan N. Midkine in inflammatory and toxic conditions. *Curr Drug Deliv*. 2013;10(1):54–57. doi:10.2174/1567201811310010009
53. Mentlein R. Targeting pleiotropin to treat osteoarthritis. *Exp Opin Ther Targets*. 2007;11(7):861–867. doi:10.1517/14728222.11.7.861
54. Pufe T, Bartscher M, Petersen W, Tillmann B, Mentlein R. Pleiotrophin, an embryonic differentiation and growth factor, is expressed in osteoarthritis. *Osteoarthritis Cartilage*. 2003;11(4):260–264. doi:10.1016/S1063-4584(02)00385-0
55. Pufe T, Groth G, Goldring MB, Tillmann B, Mentlein R. Effects of pleiotrophin, a heparin-binding growth factor, on human primary and immortalized chondrocytes. *Osteoarthritis Cartilage*. 2007;15(2): 155–162. doi:10.1016/j.joca.2006.07.005
56. Fadda SMH, Bassyouni IH, Khalifa RH, Elsaid NY. Pleiotrophin, the angiogenic and mitogenic growth factor: Levels in serum and synovial fluid in rheumatoid arthritis and osteoarthritis, and correlation with clinical, laboratory and radiological indices. *Z Rheumatol*. 2018;77(4):322–329. doi:10.1007/s00393-016-0234-8
57. Suthon S, Perkins RS, Bryja V, Miranda-Carboni GA, Krum SA. WNT5B in physiology and disease. *Front Cell Dev Biol*. 2021;9:667581. doi:10.3389/fcell.2021.667581

58. Shao L, Gou Y, Fang J, et al. The protective effects of parathyroid hormone (1-34) on cartilage and subchondral bone through down-regulating JAK2/STAT3 and WNT5A/ROR2 in a collagenase-induced osteoarthritis mouse model. *Orthop Surg*. 2021;13(5):1662–1672. doi:10.1111/os.13019
59. Martineau X, Abed É, Martel-Pelletier J, Pelletier JP, Lajeunesse D. Alteration of Wnt5a expression and of the non-canonical Wnt/PCP and Wnt/PKC-Ca_2^+ pathways in human osteoarthritis osteoblasts. *PLoS One*. 2017;12(8):e0180711. doi:10.1371/journal.pone.0180711
60. Li Y, Xiao W, Sun M, et al. The expression of osteopontin and Wnt5a in articular cartilage of patients with knee osteoarthritis and its correlation with disease severity. *Biomed Res Int*. 2016;2016:9561058. doi:10.1155/2016/9561058
61. Qi Y, Tang R, Shi Z, Feng G, Zhang W. Wnt5a/Platelet-rich plasma synergistically inhibits IL-1β-induced inflammatory activity through NF-κB signaling pathway and prevents cartilage damage and promotes meniscus regeneration. *J Tissue Eng Regen Med*. 2021;15(7):612–624. doi:10.1002/term.3198
62. Ding D, Wang L, Yan J, et al. Zoledronic acid generates a spatiotemporal effect to attenuate osteoarthritis by inhibiting potential Wnt5a-associated abnormal subchondral bone resorption. *PLoS One*. 2022;17(7):e0271485. doi:10.1371/journal.pone.0271485
63. Lambert C, Dubuc J, Montell E, et al. Gene expression pattern of cells from inflamed and normal areas of osteoarthritis synovial membrane. *Arthritis Rheum*. 2014;66(4):960–968. doi:10.1002/art.38315
64. Lee YH, White MF. Insulin receptor substrate proteins and diabetes. *Arch Pharm Res*. 2004;27(4):361–370. doi:10.1007/BF02980074
65. Eckstein SS, Weigert C, Lehmann R. Divergent roles of IRS (insulin receptor substrate) 1 and 2 in liver and skeletal muscle. *Curr Med Chem*. 2017;24(17):1827–1852. doi:10.2174/0929867324666170426142826
66. White MF. IRS2 integrates insulin/GF1 signalling with metabolism, neurodegeneration and longevity. *Diabetes Obes Metab*. 2014;16(Suppl 1):4–15. doi:10.1111/dom.12347
67. Nakahara M, Ito H, Skinner JT, et al. The inflammatory role of dysregulated IRS2 in pulmonary vascular remodeling under hypoxic conditions. *Am J Physiol Lung Cell Mol Physiol*. 2021;321(2):L416–L428. doi:10.1152/ajplung.00068.2020
68. Baquedano E, Burgos-Ramos E, Canelles S, et al. Increased oxidative stress and apoptosis in the hypothalamus of diabetic male mice in the insulin receptor substrate-2 knockout model. *Dis Model Mech*. 2016;9(5):573–583. doi:10.1242/dmm.023515
69. Okita K, Hikiji H, Koga A, et al. Ascorbic acid enhances chondrocyte differentiation of ATDC5 by accelerating insulin receptor signaling. *Cell Biol Int*. 2023;47(10):1737–1748. doi:10.1002/cbin.12067