

A lactate metabolism-related gene signature to diagnose osteoarthritis based on machine learning combined with experimental validation

Jianhua Yang^{1,2}, Wenjun Li^{1,2}, Xuemei Lin³, Wei Liang^{1,2}

¹Department of Pain Medicine, Yuebei People's Hospital, Wujiang, Shaoguan 512000, Guangdong Province, China

²Department of Traditional Chinese Orthopedics and Traumatology, Yuebei People's Hospital, Wujiang, Shaoguan 512000, Guangdong Province, China

³Department of Pediatric Orthopedics, Guangzhou Women and Children's Medical Center, Tianhe, Guangzhou 510623, Guangdong Province, China

Correspondence to: Wei Liang; **email:** beiji3feng@163.com, <https://orcid.org/0009-0002-0771-6899>

Keywords: lactate metabolism, osteoarthritis, diagnosis, machine learning

Received: October 10, 2023

Accepted: March 18, 2024

Published: October 16, 2024

Copyright: © 2024 Yang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/) (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

Background: Lactate is gradually proved as the essential regulator in intercellular signal transduction, energy metabolism reprogramming, and histone modification. This study aims to clarify the diagnosis value of lactate metabolism-related genes in osteoarthritis (OA).

Methods: Lactate metabolism-related genes were retrieved from the MSigDB. GSE51588 was downloaded from the Gene Expression Omnibus (GEO) as the training dataset. GSE114007, GSE117999, and GSE82107 datasets were adopted for external validation. Genomic difference detection, protein-protein interaction network analysis, LASSO, SVM-RFE, Boruta, and univariate logistic regression (LR) analyses were used for feature selection. Multivariate LR, Random Forest (RF), Support Vector Machine (SVM), and XGBoost (XGB) were used to develop the multiple-gene diagnosis models. 12 control and 12 OA samples were collected from the local hospital for re-verification. The transfection assays were conducted to explore the regulatory ability of the gene to the apoptosis and vitality of chondrocytes.

Results: Through the bioinformatical analyses and machine learning algorithms, SLC2A1 and NDUFB9 of the 273 lactate metabolism-related genes were identified as the significant diagnosis biomarkers. The LR, RF, SVM, and XGB models performed impressively in the cohorts (AUC > 0.7). The local clinical samples indicated that SLC2A1 and NDUFB9 were both down-regulated in the OA samples (both $P < 0.05$). The knockdown of NDUFB9 inhibited the viability and promoted the apoptosis of the CHON-001 cells treated with IL-1 β (both $P < 0.05$).

Conclusions: A lactate metabolism-related gene signature was constructed to diagnose OA, which was validated in multiple independent cohorts, local clinical samples, and cellular functional experiments.

INTRODUCTION

Osteoarthritis (OA) is the most common orthopedic disorder and is characterized by progressive damage of articular cartilage and its surrounding structures, such as subchondral bone and synovial tissues [1, 2]. The epidemiological study revealed that there are about

300 million patients with hip and/or knee OA, and OA serves as one of the leading causes of disability around the world [3]. To date, plenty of pathological mechanisms have been disclosed, including mechanical stress, immune response, programmed cell death, and energy metabolism, providing therapeutic opportunities [4–7]. Based on these efforts, some clinical trials

associated with the therapy strategies targeting the disease-causing genes, including TGF-beta, IFN-beta, IL10, IL-1R, and NKX3, have been performed [8]. These bracing advancements enlighten the researchers to seek more gene therapy targets, which is the aim of this study as well.

Lactate was considered the waste substance of anaerobic glycolysis in the past. However, emerging evidence has disclosed its pivotal biological functions and regulatory ability in different cellular processes [9]. Lactate was able to regulate adipocyte differentiation [10], tumor immune escape [11], and neuronal network activity [12] by interacting with its receptor GPR81. Histones are vital components in the chromosome, and their epigenetic modification plays important roles in the regulation of diverse cellular activities. Lactylation is a newly discovered histone modification type and has been proven as the regulator in M2 macrophage polarization, cell metabolic processes, and so forth [13]. The tight association between lactate and OA has also been reported. Compared with the normal hip, the concentration of lactate exhibited a 50% increase in the OA hip [14]. Hurter et al. found that the levels of lactate dehydrogenase in the synovial fluid could act as a potential diagnostic biomarker for OA [15]. Nevertheless, the number of studies focusing on the interaction between lactate metabolism and the pathogenesis of OA is still limited currently. Comprehensive analyses of lactate metabolism-related genes to unearth novel biomarkers of OA and uncover the underlying pathological mechanisms are urgently demanded.

Herein, the present study aims to develop a lactate metabolism-related gene signature as the diagnosis biomarker in OA. Diverse machine learning algorithms, including least absolute shrinkage and selection operator (LASSO), Boruta, supporter vector machine-recursive feature elimination (SVM-RFE), logistic regression (LR), random forest (RF), supporter vector machine (SVM), and XGBoost (XGB) were used for feature selection and model construction. 24 clinical samples collected from the local hospital were utilized to re-confirm our findings. The transfection assays were conducted to elucidate the regulatory relationship between the identified gene and the proliferation and apoptosis of chondrocytes.

MATERIALS AND METHODS

Data collection

273 lactate metabolism-related genes were retrieved from the Molecular Signatures Database (MSigDB, <https://www.gsea-msigdb.org/gsea/msigdb/index.jsp>), as

displayed in Supplementary Table 1. GSE51588 dataset [16], which included the transcriptome sequencing data of the subchondral bone tissues isolated from 10 control and 40 OA patients, was obtained from the Gene Expression Omnibus (GEO, <https://ncbi.nlm.nih.gov/geo/>) as the training cohort. GSE114007 [17], GSE117999, and GSE82107 [18] datasets, which were also downloaded from the GEO, were selected for external validation. It should be stated that GSE114007 and GSE117999 datasets contained the transcriptome sequencing of the cartilage tissues, while the GSE82107 dataset included the transcriptome sequencing of synovial tissue. Therefore, GSE114007 and GSE117999 datasets were merged as one dataset, namely “Validation-cartilage” cohort. GSE82107 dataset was called “Validation-synovial” in this study. To further detect the ability of the genes to distinguish OA from rheumatoid arthritis (RA), GSE89408 dataset [19], which included the transcription sequencing data of the joint synovial biopsies from 22 OA and 152 RA samples, was also downloaded from GEO and named as “OA vs. RA” cohort. The batch effects across these experiments were reduced via the sva package in R as possible. More detailed information on these public cohorts can be seen in Supplementary Table 2.

Gene expression difference detection and protein-protein interaction (PPI) network construction

The differentially-expressed genes between the control and OA samples were identified via the limma package in R with $|\log \text{fold change [FC]}| > 0.5$ and false discovery rate (FDR) < 0.05 filtering. Subsequently, the lactate metabolism-related genes showing expression differences were included in the PPI network based on the STRING database (<https://cn.string-db.org/>) with a confidence level = 0.4. The cytoHubba app in the Cytoscape (version 3.8.0) was harnessed to quantify the importance of the gene in the PPI network, where the degree algorithm was adopted. The top 25 genes exhibiting the highest degree were selected for further analysis.

Functional enrichment and gene set enrichment analysis (GSEA)

The functional annotation of the differentially-expressed genes was performed through the ClueGo plug-in in the Cytoscape software and the Metascape database (<https://metascape.org/gp/index.html#/main/step1>). The Gene Ontology (GO) terms with $P < 0.05$ were displayed in the network. GSEA was conducted through the GSEA software (version 4.3.2), and the Hallmark signatures, which were downloaded from the MSigDB, were chosen as the reference datasets. The signatures with nominal $P < 0.05$ and FDR < 0.05 were considered to be statistically significant.

Feature selection

Multiple machine learning algorithms were simultaneously adopted to identify the genes as the significant diagnosis biomarkers of OA. LASSO regression with 10-fold cross-validation was implemented via the glmnet package. The caret package in R was utilized to carry out the SVM-RFE algorithm. Boruta algorithm rendered each variable labeled “Tentative,” “Rejected,” or “Confirmed,” which represented the influence of the variable on the outcomes [20], and the Boruta package in R was used for the algorithm implementation. Univariate LR was performed to clarify the diagnosis value of the variables using the rms package, and $P < 0.01$ was set as the filtering threshold. At last, the genes co-determined by the PPI network analysis, LASSO, SVM-RFE, Boruta, and univariate LR were identified as the significant diagnosis biomarkers of OA and then included in the diagnosis models.

Unsupervised clustering

The consensus clustering was performed to divide the cartilage samples into different clusters using the ConsensusClusterPlus package in R software, and then the clustering was validated through the Principal Component Analysis (PCA) by the “prcomp” function in R. The differentially-expressed genes among the clusters were detected by the limma package, and $|\log_{2}FC| > 2$ and $FDR < 0.05$ were set as the filtering thresholds.

Diagnosis model construction and validation

Based on the identified genes, we used multiple algorithms, including multivariate LR, RF, SVM, and XGB, to develop the diagnosis models using the train function of the caret package. A nomogram was drawn to visualize the LR model via the rms package. We also utilized the calibrate function of the rms package to conduct the calibration analysis. The pROC package was used to draw the receiver operating characteristic (ROC) curves to quantify the performance of the models in the training and external validation cohorts. Decision curve analysis (DCA) was performed to clarify the predictive ability of the models through the dcurves package of R.

Meta-analyses

Meta-analyses were conducted to pool the odds ratios (ORs) to better clarify the diagnosis value using the meta package in R. The fixed- or random-effects model would be adopted according to the results of heterogeneity test.

Clinical sample collection

The knee cartilage tissues extracted from 12 subjects undergoing traumatic amputation without rheumatoid arthritis or OA and 12 OA subjects going through total knee replacement were obtained from the Yuebei People’s Hospital between 2021 and 2023. The Ethics Committee of Yuebei People’s Hospital reviewed and approved this research project according to the principles of the Declaration of Helsinki. The knee cartilage samples were stored in liquid nitrogen for RNA isolation. These samples collected from our hospital were included in “Validation-local” cohort in this study. The baseline clinicopathological features of these subjects can be found in Supplementary Table 3.

Enzyme-linked immunosorbent assay (ELISA)

The levels of lactate in the knee cartilage tissues from the control and OA subjects were measured by the L-lactate assay kit (Abcam, The Netherlands) following the manufacturer’s instructions. As previously mentioned [21], the samples were treated with lactate assay buffer (Abcam, The Netherlands) and the endogenous lactate dehydrogenase was removed using the deproteinizing sample preparation kit-TCA (Abcam, The Netherlands). The lactate concentration of the samples was then determined using a microplate reader.

Cell culture and treatment

Human immortalized chondrocyte CHON-001 cell line was purchased from the American Type Culture Collection (USA) and maintained in RPMI-1640 medium (Thermo Fisher Scientific, USA) supplemented with 10% FBS and 1% penicillin-streptomycin at 37° C in a humidified atmosphere with 5% CO₂. 10 ng/mL IL-1 β (Sigma-Aldrich, China) was used to treat the CHON-001 cells for 48 hours to mimic OA, as previously described [22, 23].

The transient transfection assays in the CHON-001 cells were conducted following the manufacturer’s protocol of Lipofectamine 2000 reagent (Invitrogen, USA). The small interfering RNA (siRNA) targeting NDUFB9 and the negative control siRNA (NC) were designed and synthesized by the Biosyntech company (Suzhou, China). The siRNA sequences can be found in Supplementary Table 4. Real-time quantitative PCR (RT-qPCR) experiments were used to measure the knockdown efficacy of the siRNAs.

RT-qPCR

The total RNA isolation of the CHON-001 cells and human cartilage tissues was conducted using the

TRIzol reagent (Thermo Fisher Scientific, USA). Transcriptor First Strand cDNA Synthesis Kit (Roche, Switzerland) was used for the synthesis of cDNA, and the SYBR Premix ExTaq kit (TaKaRa, China) was implemented for the RT-qPCR experiments based on the ABI 7500 system (Life Technology, USA). GAPDH was chosen as the reference gene, and the results were normalized via the $2^{-\Delta\Delta Ct}$. The primer sequence in these RT-qPCR experiments is shown in Supplementary Table 5.

Cell viability and apoptosis

The cell viability of CHON-001 cells was assessed with the Cell Counting Kit-8 (CCK-8, Sigma-Aldrich, USA) on a 96-well plate at a density of 1×10^5 cells/well. After the treatment, 10 μ l CCK-8 reagent was added to the platelets and then co-incubated with the cells for 2 hours at 37° C. The optimal density (OD) value of each well was measured by a microplate reader (Bio-Rad, USA) at an absorbance of 450 nm.

The apoptosis levels of the CHON-001 cells treated with IL-1 β were measured with the Hoechst 33342 reagent (Invitrogen, USA) according to the manufacturer's suggestions. The images were then taken by fluorescence microscopy. The ImageJ software (version 1.54d) was used to analyze the fluorescence intensity to quantify the apoptosis levels. Additionally, to further verify our finding, the flow cytometry apoptosis assay was also conducted. The flow cytometry apoptosis analysis was conducted using a fluorescein isothiocyanate-conjugated annexin V apoptosis detection kit I (BD Biosciences, USA) following the manufacturer's instructions. The analysis of the apoptosis levels of these samples was based on a flow cytometry system (BD Biosciences, USA).

Statistical analyses

The statistical analyses of the whole study were based on the R software (version 4.2.0). Welch's corrected t-test was applied to compare the difference of the values detected by the RT-qPCR, CCK-8, and apoptosis staining. Unless otherwise specified, $P < 0.05$ was accepted as the significance threshold. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

Data availability

The data that support the findings of this study are openly available in GEO (<https://www.ncbi.nlm.nih.gov/geo/>) and MSigDB (<https://www.gsea-msigdb.org/gsea/msigdb/index.jsp>). The R codes used in this study can be obtained from the corresponding author upon reasonable request.

RESULTS

74 of 273 lactate metabolism-related genes were differentially expressed

Figure 1 graphically describes the workflow of this study. First, we evaluated the lactate levels in the knee cartilage tissues from the control and OA subjects via ELISA assay. The results indicated that the levels of lactate in the OA samples were higher than those in control cases ($P < 0.05$, Supplementary Figure 1), implying that lactate might exert important functions in the pathogenesis of OA. As stated above, increasing evidence has uncovered that lactate exerts important functions in energy metabolism, signal transduction, and lactylation (Figure 2A), enlightening us to investigate the roles of lactate metabolism-related genes in OA. Based on the GSE51588 cohort, a sum of 74 of 273 lactate metabolism-related genes were differentially expressed between the control and OA samples (Figure 2B and Supplementary Table 6). The expression levels of these 74 genes are visualized in the heatmap (Figure 2C). The GO enrichment analysis indicated that these genes were mainly involved in the energy metabolic process, glycoprotein synthesis, lactate transport, and cell homeostasis maintenance (Figure 2D), indicating the underlying functions of lactate metabolism-related genes in the pathogenesis of OA.

SLC2A1 and NDUFB9 were identified as significant diagnosis biomarkers in OA

A PPI network was constructed to investigate the underlying interactions of the 74 differentially-expressed genes, and the Top 25 genes sharing the highest degree in the network were chosen for further analysis (Figure 3A). 13 genes were identified by the LASSO regression, including XK, GATA1, HMOX1, SLC2A1, CYC1, HAGH, NDUFB9, HTRA2, ISCA1, HIBCH, FKTN, PDP1, and SLC16A8 (Figure 3B and Supplementary Table 7). Meanwhile, the SVM-RFE algorithm determined 13 genes as significant factors (Figure 3C and Supplementary Table 8). The Boruta algorithm helped to identify 44 genes (Figure 3D and Supplementary Table 9), and univariate LR analysis indicated that 64 genes were of high diagnosis value ($P < 0.01$, Supplementary Table 10). Taken together, SLC2A1 and NDUFB9 were co-determined by the PPI network analysis, LASSO, SVM-RFE, Boruta, and univariate LR (Figure 3E), according to which the diagnosis models were constructed.

The unsupervised clustering based on the expressions of SLC2A1 and NDUFB9

According to the expressions of SLC2A1 and NDUFB9, 50 cartilage samples were grouped into Cluster 1 (C1)

and Cluster 2 (C2), as displayed in Figure 4A and Supplementary Table 11. PCA indicated the distinct genomic patterns of these two clusters (Figure 4B). Importantly, we noticed that about 95% C2 cases belong

to OA subjects, and none of the C1 cases exhibited OA phenotype ($P < 0.001$, Figure 4C), suggesting that the clustering was associated with the OA characteristics. The expressions of SLC2A1 ($P < 0.001$) and NDUFB9

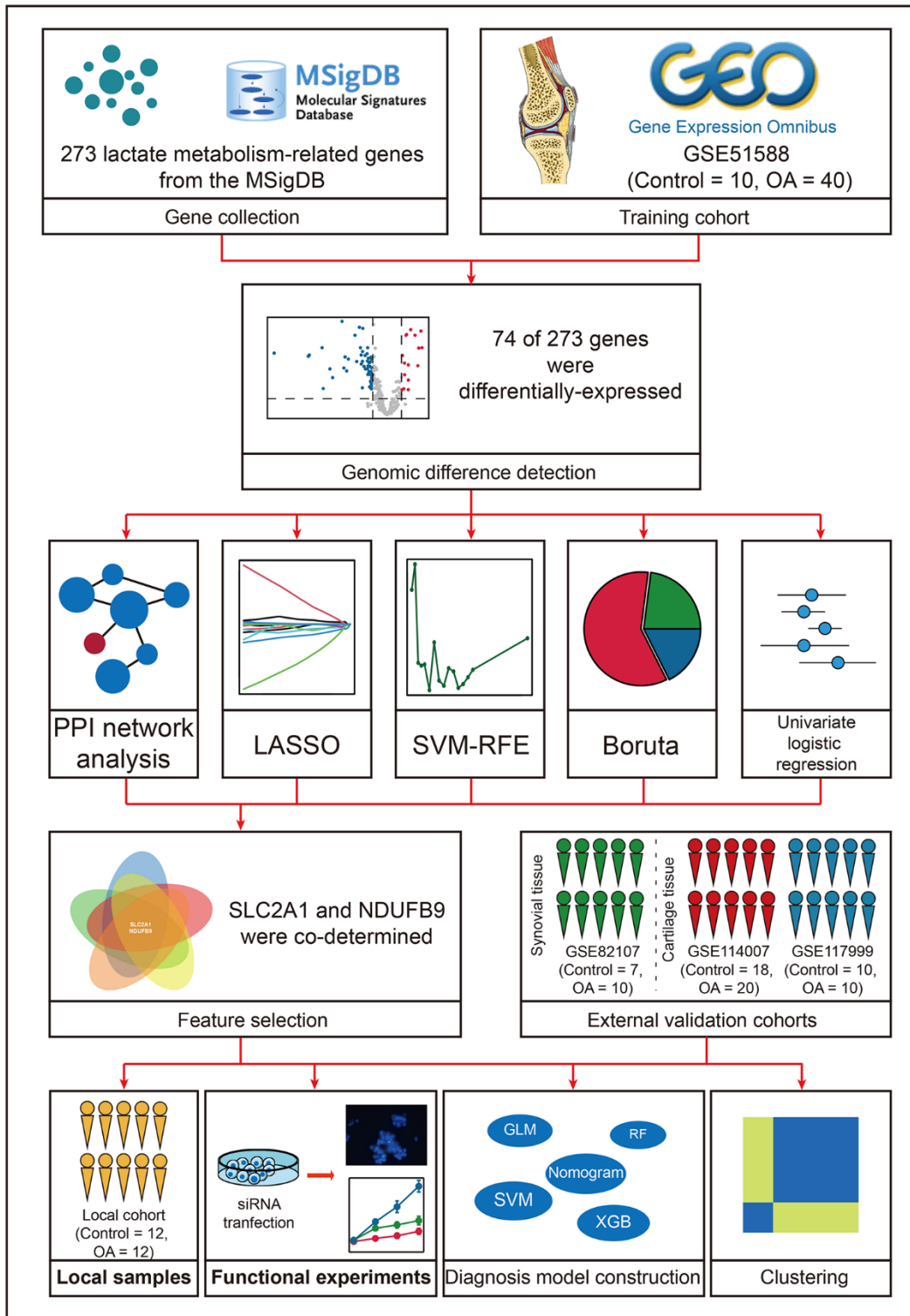


Figure 1. The workflow of the present study.

($P < 0.001$) were significantly down-regulated in C2 subjects (Figure 4D). Subsequently, the differentially-expressed genes between the C1 and C2 subjects were analyzed, and a sum of 190 genes were screened (Figure

4E and Supplementary Table 12). The functional enrichment displayed that these genes were mainly associated with the cell activation and differentiation, immune and inflammatory response, and metabolic

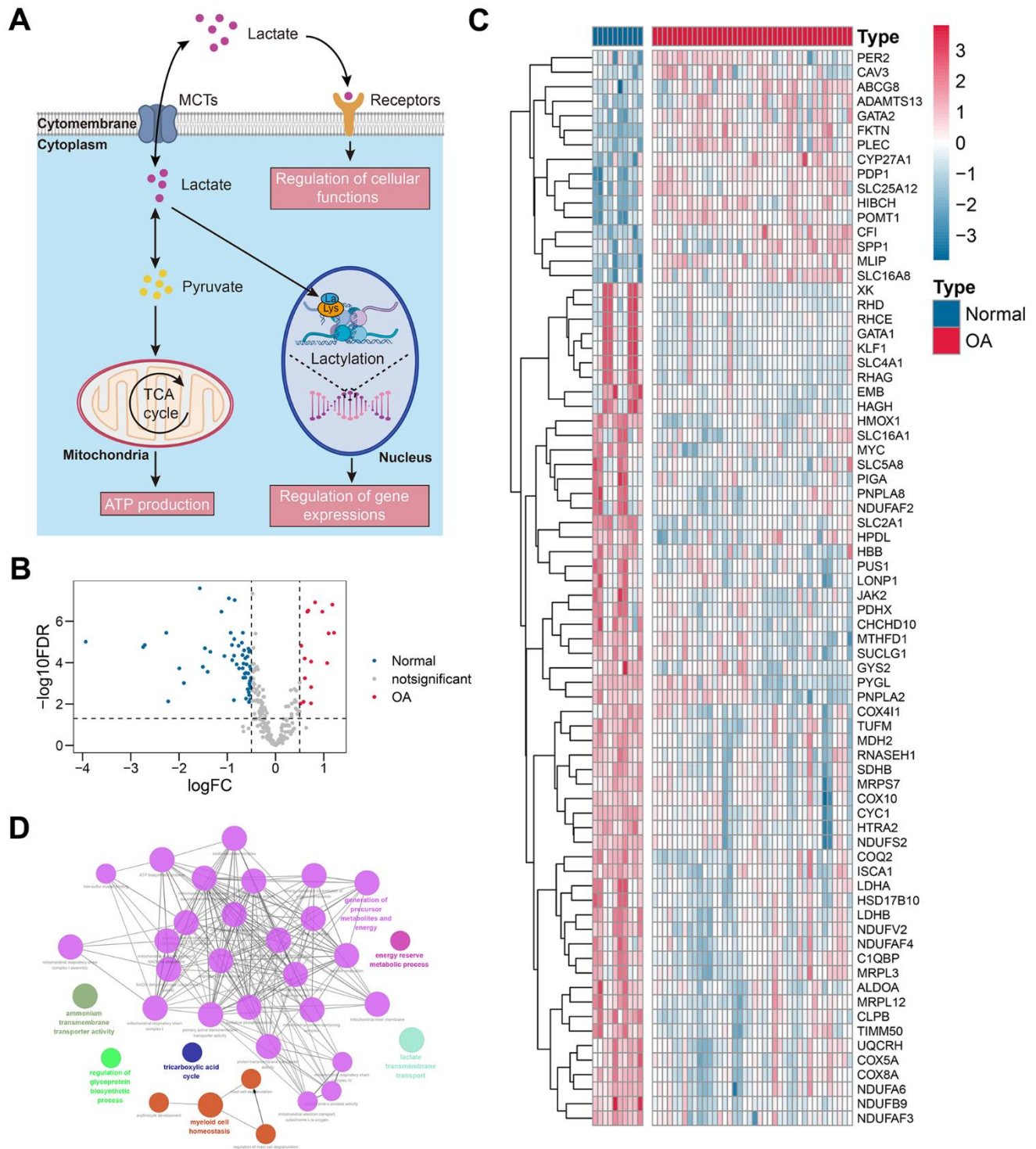


Figure 2. Differentially-expressed lactate metabolism-related genes. (A) The schematic summary of the biological functions of lactate. (B, C) The volcano plot (B) and the heat map (C) indicated that 74 of 273 lactate metabolism-related genes were differentially expressed between the control and OA tissues. (D) The functional annotation of the 74 genes. Abbreviation: OA, osteoarthritis.

processes (Figure 4F), implying the possible functions of SLC2A1 and NDUFB9.

SLC2A1 and NDUFB9 were reliable diagnosis biomarkers in the meta-analyses

The diagnostic performance of SLC2A1 and NDUFB9 in the training, GSE82107, GSE114007, and GSE117999 cohorts was displayed in Figure 5A–5D, respectively. Despite the fact that these genes performed poorly in some cohorts, the meta-analyses indicated that NDUFB9 (pooled OR = 0.33, 95% confidence interval [CI] = 0.15-0.74, Figure 5E) and SLC2A1 (pooled OR = 0.24, 95% CI = 0.11-0.53, Figure 5F) both served as significant diagnosis biomarkers in OA subjects.

The performance of the machine learning-based diagnosis models

Multiple-gene diagnosis models were developed using LR, RF, SVM, and XGB based on the mRNA expression levels of SLC2A1 and NDUFB9. First, a nomogram was constructed using the established LR model (Figure 6A). ROC analyses were performed to evaluate the predictive performance of the nomogram in the training (area under curve [AUC] = 1.000, 95%CI = 1.000-1.000, Figure 6B), validation-cartilage (AUC = 0.726, 95%CI = 0.627-0.830, Figure 6C), and validation-synovial (AUC = 0.571, 95%CI = 0.350-0.786, Figure 6D) cohorts. The calibration plots demonstrated that the nomogram had acceptable predictive capability in the training (Figure 6E),

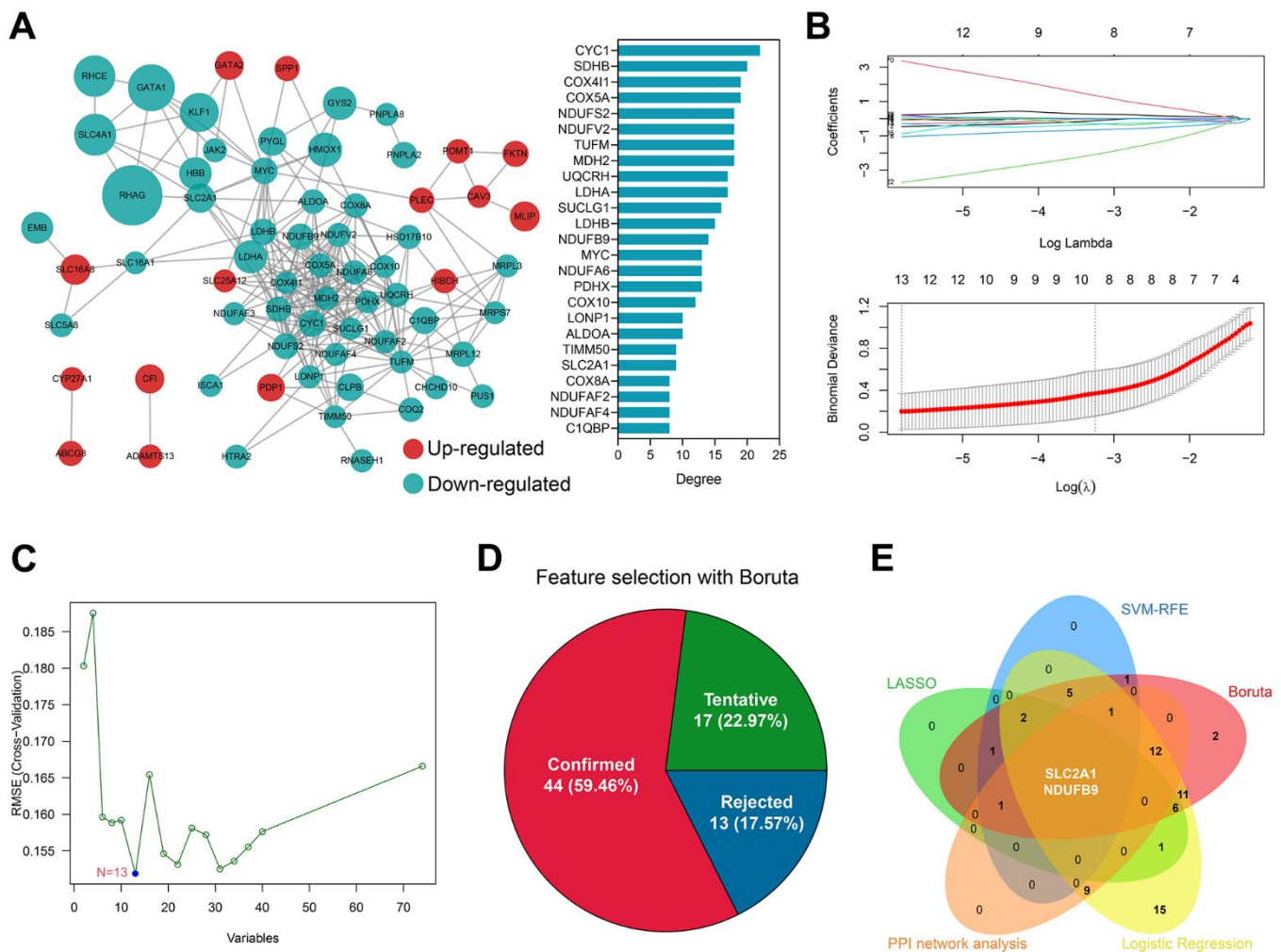
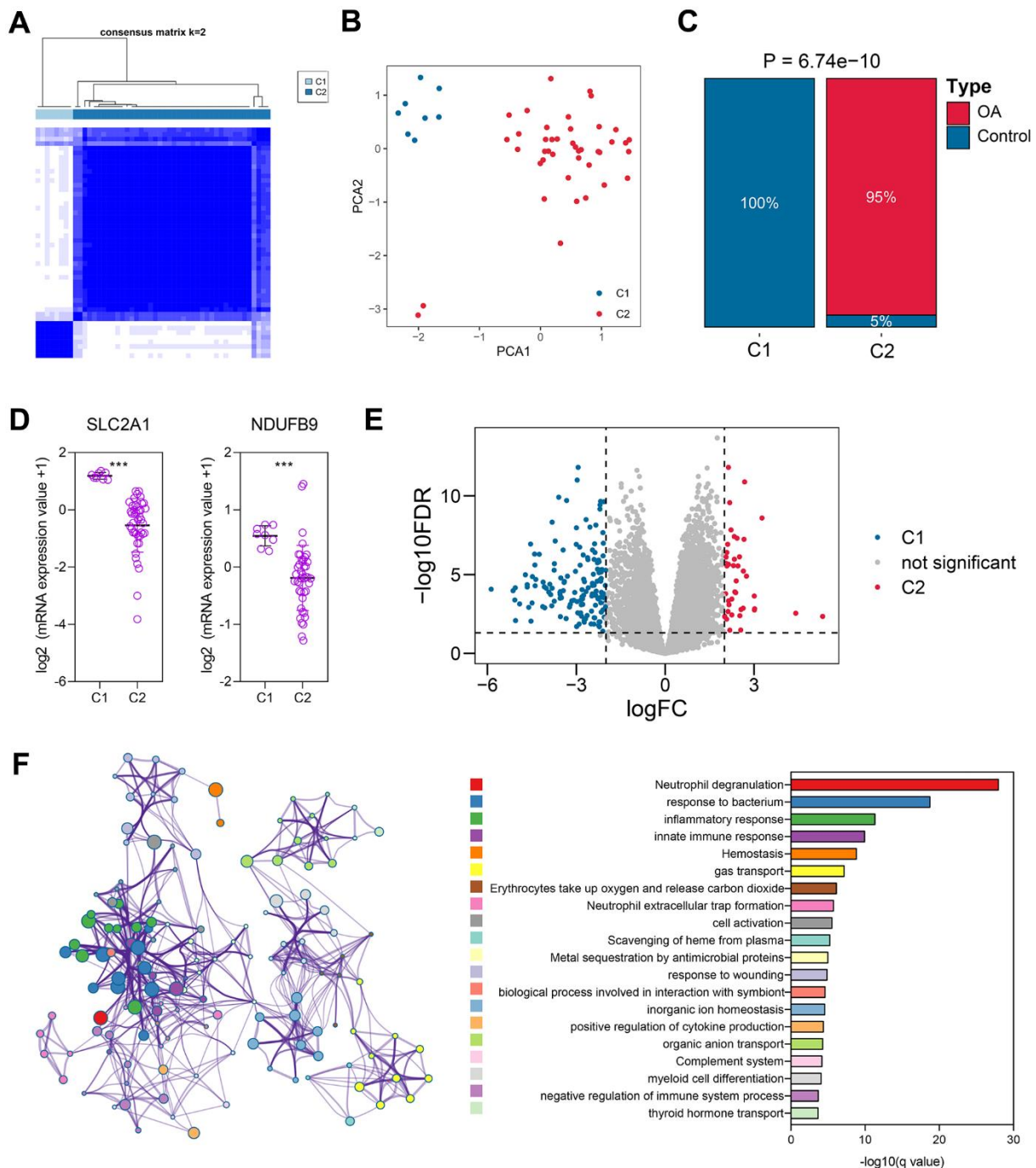


Figure 3. SLC2A1 and NDUFB9 were co-determined by bioinformatical analyses and machine learning algorithms. (A) The construction and analyses of the PPI network. (B) 13 genes were identified by the LASSO regression. (C) 13 genes were determined by the SVM-RFE algorithm. (D) Boruta algorithm showed that 44 genes were of high diagnosis value. (E) SLC2A1 and NDUFB9 were co-determined by the LASSO, SVM-RFE, Boruta, univariate logistic regression, and PPI network analysis. Abbreviations: LASSO, least absolute shrinkage and selection operator; SVM-RFE, supporter vector machine-recursive feature elimination; PPI, protein-protein interaction.

validation-cartilage (Figure 6F), and validation-synovial (Figure 6G) cohorts. Furthermore, decision curve analysis (DCA) was conducted to assess the net benefit of the nomogram at different decision thresholds in these cohorts (Figure 6H–6J).

In addition to the nomogram, alternative machine learning-based models were also developed, as previously mentioned. ROC analyses demonstrated that the RF model exhibited AUCs of 1.000 (95%CI = 1.000-1.000), 0.656 (95% CI = 0.524-0.780), and 0.736



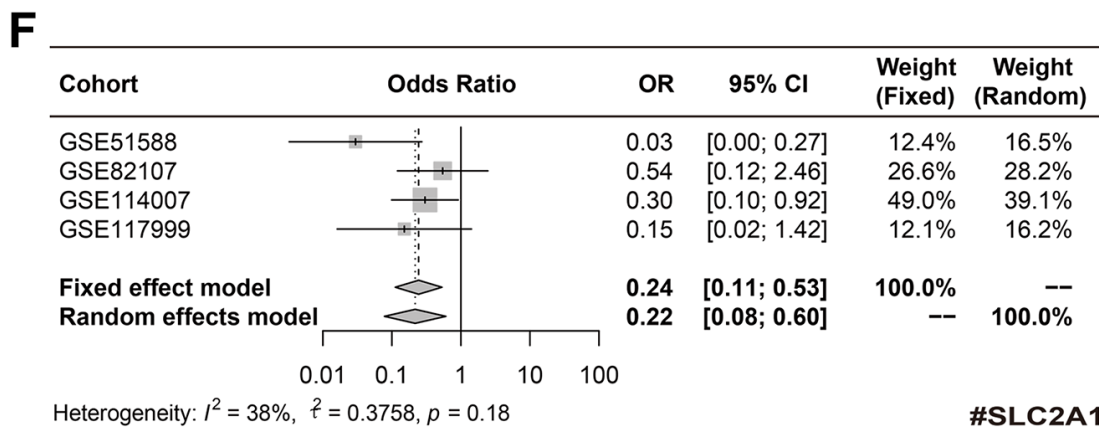
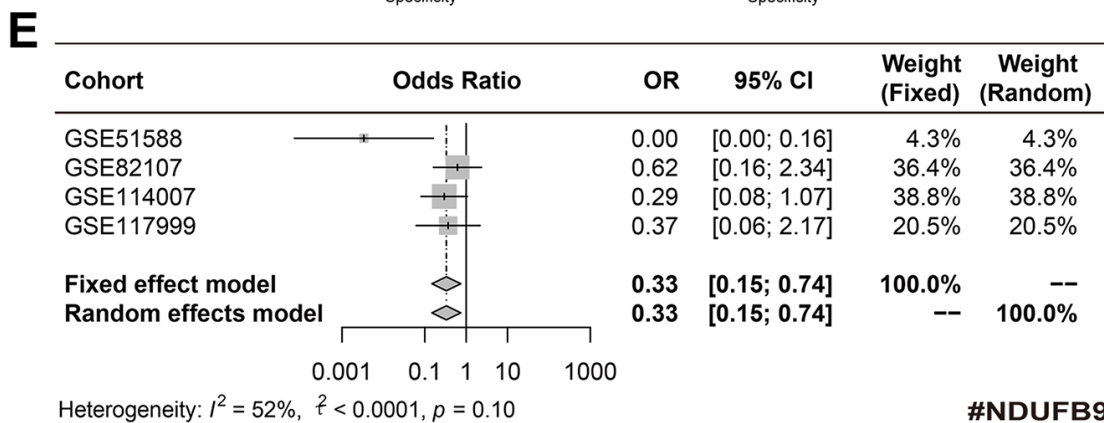
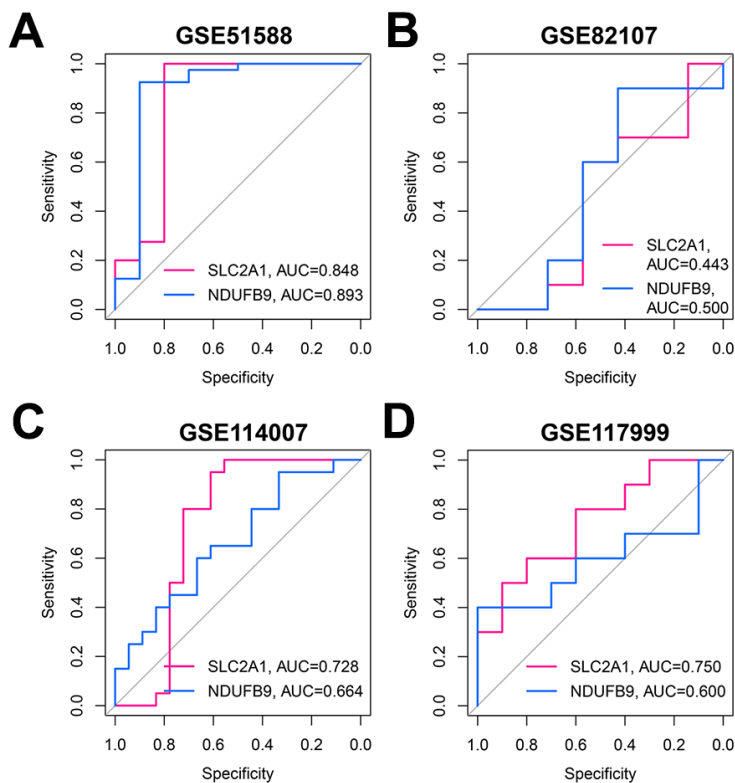


Figure 5. The diagnosis value of NDUFB9 and SLC2A1. (A–D) The diagnostic performance of NDUFB9 and SLC2A1 in the training (A), GSE82107 (B), GSE114007 (C), and GSE117999 (D) cohorts. (E, F) The meta-analyses were used to pool the ORs of NDUFB9 (E) and SLC2A1 (F). Abbreviation: OR, odds ratio.

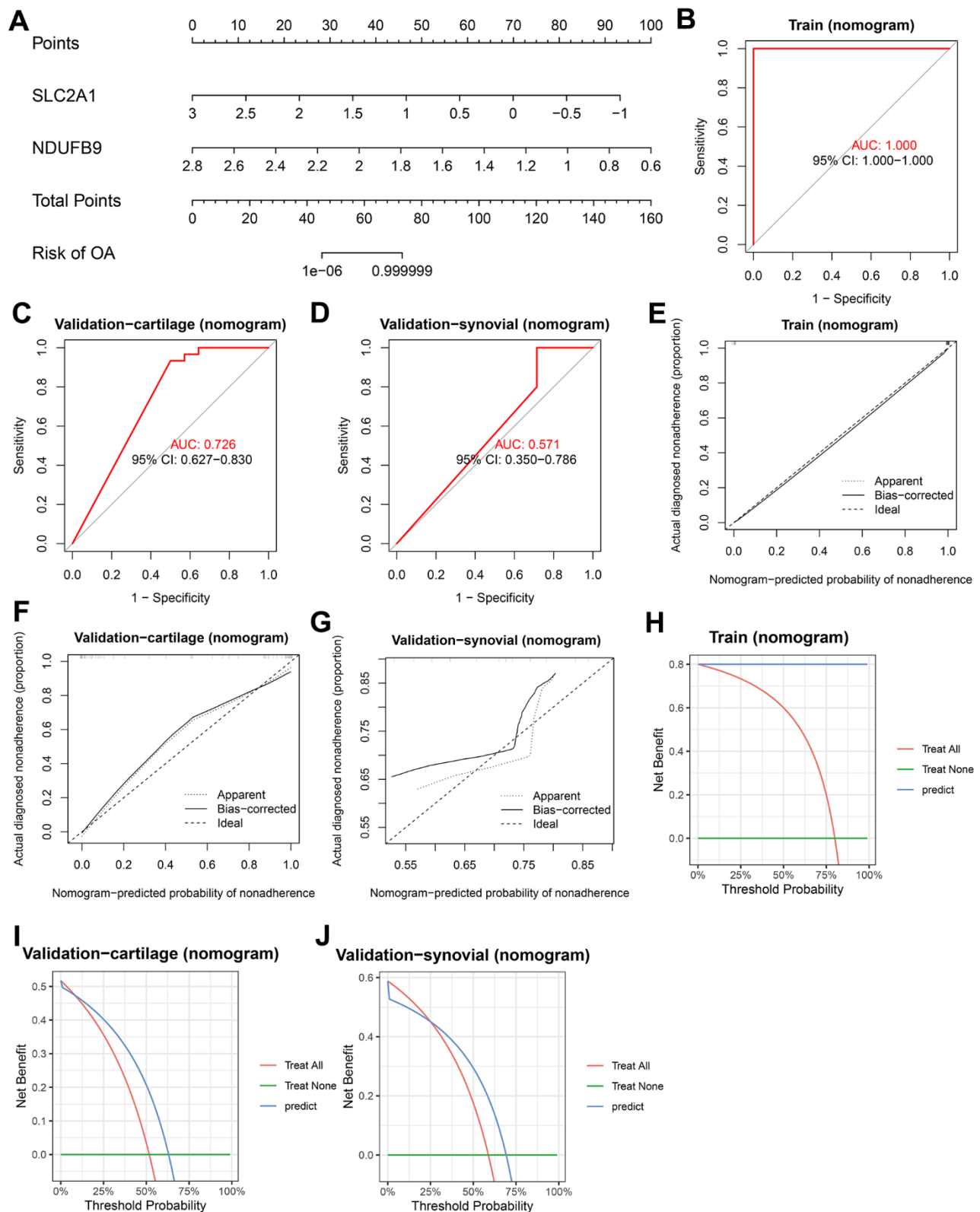


Figure 6. Construction and validation of the diagnostic nomogram. (A) A nomogram was constructed to visualize the LR model. (B–D) ROC analysis shows the predictive ability of the nomogram in the training (B), validation-cartilage (C), and validation-synovial (D) cohorts. (E–G) Calibration analysis shows the predictive ability of the nomogram in the training (E), validation-cartilage (F), and validation-synovial (G) cohorts. (H–J) DCA analysis shows the predictive ability of the nomogram in the training (H), validation-cartilage (I), and validation-synovial (J) cohorts. Abbreviations: ROC, receiver operating characteristic; DCA, decision curve analysis; LR, logistic regression.

(95%CI = 0.493-0.936) in the training, validation-cartilage, and validation-synovial cohorts, respectively (Figure 7A). Furthermore, the DCA showcased the net benefit of the RF model at various thresholds in these cohorts (Figure 7B). Likewise, the SVM diagnostic model demonstrated AUCs of 1.000 (95%CI = 1.000-1.000), 0.614 (95%CI = 0.452-0.756), and 0.500 (95%CI = 0.214-0.829) in the training, validation-cartilage, and validation-synovial cohorts (Figure 7C). DCA was also employed to further elucidate the predictive capability of the SVM model in these datasets (Figure 7D). Similarly, the XGB model exhibited impressive performance in the training (AUC = 1.000, 95%CI = 1.000-1.000, Figure 7E), validation-cartilage (AUC = 0.693, 95%CI = 0.568-0.810, Figure 7E), and validation-synovial (AUC = 0.571, 95%CI = 0.271-0.843, Figure 7E) cohorts. The results of DCA for the XGB model in these cohorts are illustrated in Figure 7F. Additionally, the calibration analyses of the RF (Supplementary Figure 2A), SVM (Supplementary Figure 2B), and XGB (Supplementary Figure 2C) in the training, validation-cartilage, and validation-synovial cohorts were also performed. Overall, despite the suboptimal performance of these models in certain cohorts, we have established that the gene signature, encompassing SLC2A1 and NDUFB9, represents a potential diagnostic biomarker for OA, irrespective of whether it is present in cartilage or synovial tissues.

Validation in the local cohort

To further confirm our findings, we collected the cartilage tissues isolated from 12 control and 12 OA subjects in our local hospital. RT-qPCR experiments were conducted to measure the expression levels of SLC2A1 (Figure 8A) and NDUFB9 (Figure 8B), and the results indicated that SLC2A1 ($P < 0.05$, Figure 8C) and NDUFB9 ($P < 0.05$, Figure 8D) were both down-regulated in the OA samples. The ROC analyses displayed the diagnosis performance of SLC2A1 (AUC = 0.833, 95%CI = 0.639-0.979) and NDUFB9 (AUC = 0.743, 95% CI = 0.528-0.917) in the local cohort (Figure 8E).

Furthermore, in order to provide further insights into the diagnostic capabilities of the established models, we conducted an analysis to assess the predictive performance of the nomogram, RF, SVM, and XGB models in the local cohort. The AUCs of the nomogram, RF, SVM, and XGB models were found to be 0.667 (95%CI = 0.542-0.792, Supplementary Figure 3A), 0.715 (95%CI = 0.510-0.889, Supplementary Figure 3B), 0.580 (95%CI = 0.333-0.826, Supplementary Figure 3C), and 0.674 (95%CI = 0.486-0.837, Supplementary Figure 3D), respectively. Additionally, DCA plots were generated to evaluate the performance of the nomogram, RF, SVM,

and XGB models in the local cohort (Supplementary Figure 3E–3H). Moreover, calibration plots were constructed to demonstrate the predictive ability of these models in the local cohort (Supplementary Figure 3I–3L). It is worth noting that the training cohort utilized transcriptome sequencing technology, while the local cohort relied on RT-qPCR experiments. The discrepancy arising from the utilization of distinct gene expression detection platforms could potentially explain the suboptimal performance of the predictive models in the local cohort.

Knockdown of NDUFB9 inhibited the proliferation and promoted the apoptosis of cartilage cells

Human Immortalized chondrocyte CHON-001 cells were treated with IL-1beta to construct the OA cell model. We observed that the expressions of SLC2A1 and NDUFB9 were both down-regulated in the OA cell models (both $P < 0.05$, Figure 8F). Since the roles of SLC2A1 in OA have been reported by previous efforts [24, 25], we then selected NDUFB9 to conduct the functional investigation. Two different siRNAs targeting NDUFB9 were used for the gene knockdown, and siRNA-2 exhibited higher efficacy ($P < 0.01$, Figure 8G) and was chosen for further analysis. The knockdown of NDUFB9 significantly suppressed the proliferation ($P < 0.05$, Figure 8H) and enhanced the apoptosis ($P < 0.05$, Figure 8I) of the CHON-001 cells treated with IL-1beta. In order to gain further insights into the impact of NDUFB9 on the IL-1beta-induced apoptosis of CHON-001 cells, we conducted additional flow cytometry apoptosis analysis. The results obtained from the flow cytometry assay provide additional validation for our findings ($P < 0.05$, Supplementary Figure 4).

The signal pathways associated with SLC2A1 and NDUFB9 in OA

The OA samples extracted from the GSE51588 cohort were equally divided into the low- and high-gene expression subgroups, and then the GSEA was conducted (Figure 9A). The signal pathways associated with NDUFB9 and SLC2A1 were shown in Figure 9B and Figure 9C, respectively. The most relevant pathways of NDUFB9 and SLC2A1 are oxidative phosphorylation (NES= 3.930, FDR < 0.001, Figure 9D) and E2F targets (NES = 2.719, FDR < 0.001, Figure 9E), which indicated the underlying mechanisms of these genes in the pathogenesis of OA.

The ability of SLC2A1 and NDUFB9 to distinguish OA from RA

OA and RA are both prevalent joint disorders, characterized by overlapping clinical manifestations

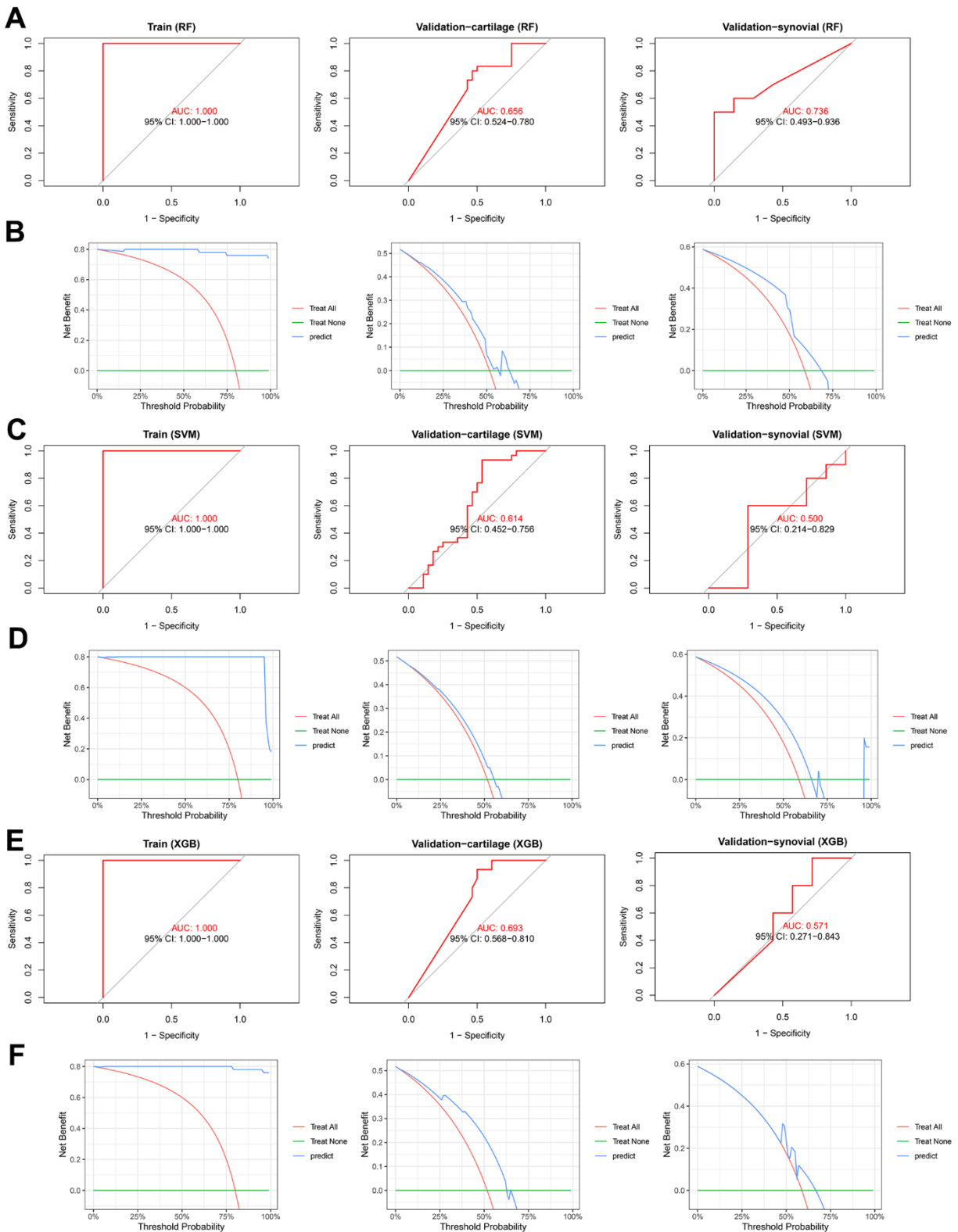


Figure 7. The predictive performance of the RF, SVM, and XGB diagnosis models. (A, B) ROC (A) and DCA (B) analysis indicates the predictive performance of RF model in the training (left), validation-cartilage (middle), and validation-synovial (right) cohorts. (C, D) ROC (C) and DCA (D) analysis indicates the predictive performance of SVM model in the training (left), validation-cartilage (middle), and validation-synovial (right) cohorts. (E, F) ROC (E) and DCA (F) analysis indicates the predictive performance of XGB model in the training (left), validation-cartilage (middle), and validation-synovial (right) cohorts. Abbreviations: RF, random forest; SVM, supporter vector machine; XGB, XGBoost; AUC, area under curve; CI, confidence interval.

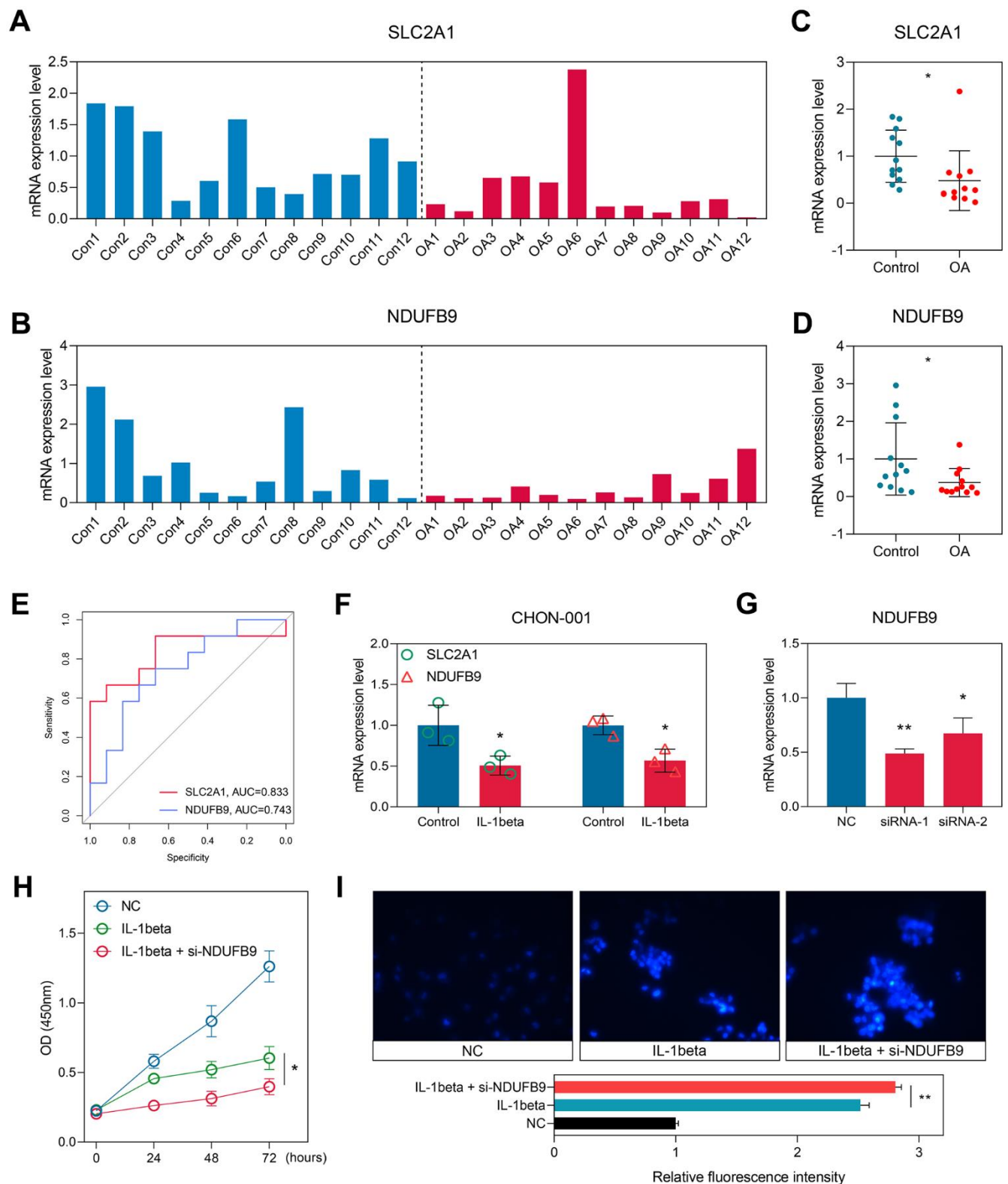


Figure 8. Experimental validation in the clinical samples and cellular experiments. (A, B) The expression levels of SLC2A1 (A) and NDUFB9 (B) in the 12 control and 12 OA samples collected from the local hospital. (C, D) The levels of SLC2A1 (C) and NDUFB9 (D) were down-regulated in the OA tissues. (E) The diagnostic performance of SLC2A1 and NDUFB9 in the local cohort. (F) SLC2A1 and NDUFB9 were both down-regulated in the CHON-001 cells treated with IL-1beta. (G) Two different siRNAs were used to construct the NDUFB9-knockdown CHON-001 cells. (H) The proliferation rate was inhibited in the NDUFB9-knockdown CHON-001 cells. (I) The CHON-001 cells with the knockdown of NDUFB9 showed higher levels of apoptosis after the IL-1beta treatment. * $P < 0.05$; ** $P < 0.01$.

including joint pain, swelling, stiffness, and morning stiffness. These shared symptoms pose a challenge in accurately diagnosing these conditions in clinical settings. Consequently, we aimed to investigate the discriminative potential of SLC2A1 and NDUFB9 in distinguishing between OA and RA, utilizing a sample cohort of 22 OA and 152 RA samples obtained from the GSE82107 dataset. Upon analysis, it was observed that RA tissues exhibited significantly higher levels of NDUFB9 compared to OA samples ($P < 0.05$, Supplementary Figure 5A), while no significant expression difference was observed for SLC2A1 ($P > 0.05$, Supplementary Figure 5A). The area under the curve (AUC) values for NDUFB9 and SLC2A1 were calculated as 0.648 (95%CI = 0.510-0.781,

Supplementary Figure 5B) and 0.580 (95%CI = 0.418-0.728, Supplementary Figure 5C), respectively. These findings suggest that NDUFB9 may serve as a potential biomarker for distinguishing between OA and RA.

DISCUSSION

Seeking novel biomarkers in OA has always been a hot topic, which helps to perform the early diagnosis, disclose new pathogenesis, and develop novel targeted drugs. Recently, tremendous advancements have been made in OA biomarker identification with the proposal and application of machine learning algorithms, multi-omics sequencing, and big-data mining [26–28].

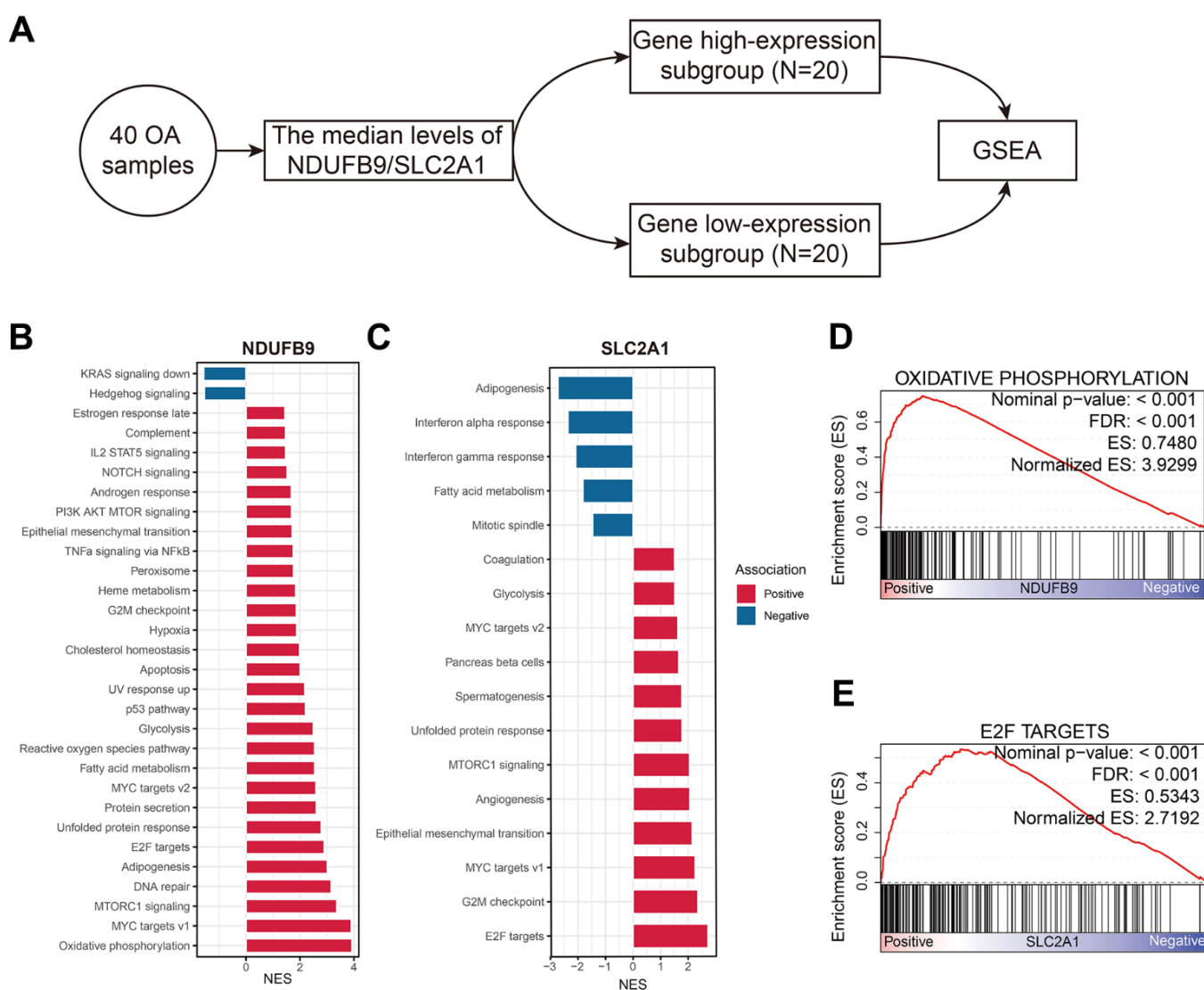


Figure 9. The functional annotation of NDUFB9 and SLC2A1 in the OA samples. (A) GSEA was conducted according to the median expression levels of the genes. (B, C) The biological processes associated with NDUFB9 (B) and SLC2A1 (C). (D, E) The signal pathways most relevant to NDUFB9 (D) and SLC2A1 (E). Abbreviation: GSEA, Gene Set Enrichment Analysis.

Detection of the biomarkers in OA from particular biological aspects, such as immunity [29], cell death [30, 31], RNA-binding proteins [32], and epigenetic regulators [33], is becoming more and more popular. These great efforts expanded our knowledge of the initiation and development of OA and provided the potential tools in clinical practice. However, despite the fact the multiple and pivotal functions of lactate have attracted increasing attention in recent years [34, 35], as discussed above, no comprehensive analysis of the lactate metabolism-related genes as the biomarkers in OA has been performed up to date.

In this study, initially, a total of 12 control and 12 OA samples were collected from our hospital, and the levels of lactate in these tissues were quantified using the ELISA. It was observed that the concentration of lactate in the OA samples was significantly higher than that in the control samples ($P < 0.05$). Subsequently, a comprehensive analysis was conducted using 173 lactate metabolism-related genes obtained from the MSigDB database, and the GSE51588 dataset from the GEO database was selected as the training cohort. Out of the 273 lactate metabolism-related genes, 74 exhibited differential expression between the control and OA samples. Through the integration of the PPI network, LASSO, SVM-RFE, Boruta, and univariate LR, SLC2A1 and NDUF9 were identified as significant diagnostic biomarkers. The diagnostic ability of SLC2A1 and NDUF9 in OA was further validated through unsupervised clustering and meta-analyses. Furthermore, multiple machine learning diagnosis models, including LR, RF, SVM, and XGB, were constructed based on the expressions of SLC2A1 and NDUF9. A nomogram was developed to visualize the LR predictive model. These models demonstrated satisfactory performance across the training, validation-cartilage, and validation-synovial cohorts. Notably, the diagnosis models also exhibited effectiveness in synovial tissues, despite being trained on subchondral bone tissues, indicating the widespread involvement of SLC2A1 and NDUF9 in the pathogenic processes of OA. Subsequently, an independent set of 12 control and 12 OA samples obtained from a local hospital were utilized to confirm our findings (both $P < 0.05$). Cellular functional experiments were conducted, revealing that the knockdown of NDUF9 significantly suppressed the viability and enhanced the apoptosis of CHON-001 cells treated with IL-1 β (both $P < 0.05$). Finally, the ability of SLC2A1 and NDUF9 to distinguish between OA and RA was evaluated, and the results suggested that NDUF9 holds promise as a potential biomarker for distinguishing OA from RA.

In this study, we integrated sequencing data from subchondral bone, cartilage, and synovial tissue of

OA patients, considering the tight relationship of these three tissues in the pathogenesis of OA [36]. The main pathology of OA is the degeneration of articular cartilage, where the cartilage surface becomes rough, with cracks and wear. After cartilage degeneration, the underlying bone experiences additional pressure and friction, leading to increased bone cell proliferation and bone density, eventually resulting in bone sclerosis and osteophyte formation. Osteophytes can further damage the synovium and cartilage, exacerbating the condition of OA. The degradation and destruction of cartilage release certain cellular factors and inflammatory mediators that can stimulate synovial cells to produce an inflammatory response. Synovitis not only causes synovial lesions but also worsens joint destruction. Inflammatory mediators produced by synovitis, such as cytokines and enzymes, can directly affect the metabolism and degeneration process of cartilage, while also stimulating inflammatory responses and metabolic disorders in subchondral bone cells, further exacerbating bone changes and sclerosis [37]. Nevertheless, despite the close relationship of these tissues in the pathogenesis of OA, the distinctions among these tissues should also be acknowledged. In this study, the training dataset primarily consisted of subchondral bone tissues, whereas the validation process focused on cartilage and synovial tissues. This selection leads to the potential divergence of the analysis results, a factor that should be underscored to ensure readers' awareness.

The roles SLC2A1 played in OA have been widely reported [24, 25, 38]. Guan et al. disclosed that the knockout of SLC2A1 promoted the levels of HIF-1 α and apoptosis in chondrocytes [24]. Yao et al. and Zheng et al. reported that SLC2A1 could serve as the diagnosis biomarker in OA [25, 38], which was in accordance with our findings. The current study has identified a positive correlation between SLC2A1 and E2F-related signaling pathways. The E2F gene family plays a pivotal role in the regulation of the cell cycle and cell proliferation. E2F proteins function as transcription factors that govern the expression of genes involved in cell division, DNA replication, and cell differentiation [39]. These proteins are responsible for orchestrating the progression of cells through different phases of the cell cycle, ensuring appropriate cell growth and development. Dysregulation of E2F genes has been linked to various diseases, including osteoarthritis (OA), where abnormal E2F activity can result in uncontrolled cell apoptosis [40]. However, further investigation is required to elucidate the regulatory relationship between SLC2A1 and E2F-related signaling in the pathogenesis of OA.

NDUFB9 encodes a protein serving as a subunit of the mitochondrial oxidative phosphorylation complex I [41]. Therefore, the dysfunction of NDUFB9 leads to the overproduction of mitochondria-derived reactive oxygen species (mtROS) and the disturbance of the NAD⁺/NADH balance in tumor cells [42]. Nevertheless, no investigation of the biological functions of NDUFB9 in OA has been conducted so far. In this study, we reported that NDUFB9 was a significant diagnosis biomarker and could promote vitality and inhibit the apoptosis of chondrocytes in OA for the first time. Besides, the local samples indicated that NDUFB9 was significantly down-regulated in the cartilage tissues isolated from OA patients. The function enrichment analysis revealed a positive correlation between NDUFB9 and the activation of oxidative phosphorylation. Oxidative phosphorylation is a vital metabolic process in cells, primarily responsible for converting the chemical energy of organic molecules into a usable form of cellular energy, specifically adenosine triphosphate (ATP). Oxidative phosphorylation is the predominant pathway through which cells generate ATP and plays a critical role in maintaining cellular survival and function. A previous study reported an upregulation of oxidative phosphorylation activation levels in osteoarthritis (OA) samples, and inhibiting this process significantly impeded OA progression [43]. Our findings suggest a strong association between NDUFB9 and the dysregulation of oxidative phosphorylation in OA, although the underlying mechanisms remain unclear.

The current study developed the diagnostic models based on four common algorithms, namely LR, RF, SVM, and XGB. These algorithms each have their own characteristics [44]. The advantage of the LR model lies in its simplicity and speed, making it suitable for binary classification problems and performing well on linearly separable datasets. It also has strong interpretability. The RF model is effective in handling high-dimensional data and a large number of features, and it has good robustness. The SVM model can handle high-dimensional data and non-linear relationships, and it has good generalization ability. Especially in dealing with small samples and high-dimensional features, the SVM model performs well and can be used for non-linear mapping through kernel functions. The XGB model can handle various types of data and has good accuracy and generalization ability. It can handle high-dimensional data and non-linear relationships, has good robustness and interpretability, and can automatically handle missing and outlier values. In this study, we observed remarkable performance of the LR model in the validation-cartilage cohort (AUC = 0.726), whereas the RF model exhibited strong performance in both the validation-synovial (AUC = 0.736) and validation-

local (AUC = 0.715) cohorts. These findings indicate the versatility of these models across various tissue types.

The limitations of this study should be acknowledged. First, despite the fact that the diagnosis values of SLC2A1 and NDUFB9 have been validated in multiple public cohorts and local clinical samples, a large-scale, multi-center, and prospective clinical trial would be more beneficial to clarify these genes' usefulness. Second, although we have investigated the functions of NDUFB9 *in vitro*, the *in vivo* experiments and the exploration of its specific regulatory mechanisms remain needed.

CONCLUSIONS

Collectively, a lactate metabolism-related gene signature was developed to diagnose OA, which was validated in multiple independent cohorts, local clinical samples, and cellular functional experiments. Our findings provide novel insights into the biological mechanisms of OA and offer a possible tool in clinical practice.

AUTHOR CONTRIBUTIONS

WL designed the whole study and provided the financial support. JY collected and analyzed the data. JY and WL participated in the provision of study materials or patients. JY, XL, and WL conducted the cellular experiments, ELISA, RT-qPCR tests, and flow cytometry apoptosis analysis. All authors participated in manuscript writing and final approval of manuscript.

ACKNOWLEDGMENTS

Special thanks for the contributions of GEO and MSigDB databases.

CONFLICTS OF INTEREST

The authors report no conflicts of interest in this work.

ETHICAL STATEMENT AND CONSENT

The Ethics Committee of Yuebei People's Hospital reviewed and approved this research project (KY-2023006). All methods were performed in accordance with the relevant guidelines and regulations according to the principles of the Declaration of Helsinki. All the participants had signed the informed consent forms prior to enrollment into the study.

FUNDING

This study receives no funding from any agency.

REFERENCES

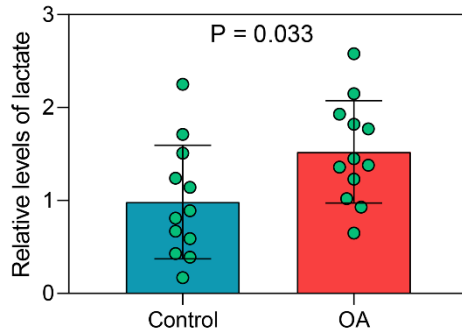
- Giorgino R, Albano D, Fusco S, Peretti GM, Mangiavini L, Messina C. Knee Osteoarthritis: Epidemiology, Pathogenesis, and Mesenchymal Stem Cells: What Else Is New? An Update. *Int J Mol Sci.* 2023; 24:6405.
<https://doi.org/10.3390/ijms24076405>
PMID:[37047377](https://pubmed.ncbi.nlm.nih.gov/37047377/)
- Hunter DJ, Bierma-Zeinstra S. Osteoarthritis. *Lancet.* 2019; 393:1745–59.
[https://doi.org/10.1016/S0140-6736\(19\)30417-9](https://doi.org/10.1016/S0140-6736(19)30417-9)
PMID:[31034380](https://pubmed.ncbi.nlm.nih.gov/31034380/)
- Safiri S, Kolahi AA, Smith E, Hill C, Bettampadi D, Mansournia MA, Hoy D, Ashrafi-Asgarabad A, Sepidarkish M, Almasi-Hashiani A, Collins G, Kaufman J, Qorbani M, et al. Global, regional and national burden of osteoarthritis 1990-2017: a systematic analysis of the Global Burden of Disease Study 2017. *Ann Rheum Dis.* 2020; 79:819–28.
<https://doi.org/10.1136/annrheumdis-2019-216515>
PMID:[32398285](https://pubmed.ncbi.nlm.nih.gov/32398285/)
- Warmink K, Vinod P, Korthagen NM, Weinans H, Rios JL. Macrophage-Driven Inflammation in Metabolic Osteoarthritis: Implications for Biomarker and Therapy Development. *Int J Mol Sci.* 2023; 24:6112.
<https://doi.org/10.3390/ijms24076112>
PMID:[37047082](https://pubmed.ncbi.nlm.nih.gov/37047082/)
- Liu S, Pan Y, Li T, Zou M, Liu W, Li Q, Wan H, Peng J, Hao L. The Role of Regulated Programmed Cell Death in Osteoarthritis: From Pathogenesis to Therapy. *Int J Mol Sci.* 2023; 24:5364.
<https://doi.org/10.3390/ijms24065364>
PMID:[36982438](https://pubmed.ncbi.nlm.nih.gov/36982438/)
- Fang T, Zhou X, Jin M, Nie J, Li X. Molecular mechanisms of mechanical load-induced osteoarthritis. *Int Orthop.* 2021; 45:1125–36.
<https://doi.org/10.1007/s00264-021-04938-1>
PMID:[33459826](https://pubmed.ncbi.nlm.nih.gov/33459826/)
- Zheng L, Zhang Z, Sheng P, Mobasheri A. The role of metabolism in chondrocyte dysfunction and the progression of osteoarthritis. *Ageing Res Rev.* 2021; 66:101249.
<https://doi.org/10.1016/j.arr.2020.101249>
PMID:[33383189](https://pubmed.ncbi.nlm.nih.gov/33383189/)
- Evans CH, Ghivizzani SC, Robbins PD. Osteoarthritis gene therapy in 2022. *Curr Opin Rheumatol.* 2023; 35:37–43.
<https://doi.org/10.1097/BOR.0000000000000918>
PMID:[36508307](https://pubmed.ncbi.nlm.nih.gov/36508307/)
- Wu Y, Ma W, Liu W, Zhang S. Lactate: a pearl dropped in the ocean-an overlooked signal molecule in physiology and pathology. *Cell Biol Int.* 2023; 47:295–307.
<https://doi.org/10.1002/cbin.11975> PMID:[36511218](https://pubmed.ncbi.nlm.nih.gov/36511218/)
- Cai TQ, Ren N, Jin L, Cheng K, Kash S, Chen R, Wright SD, Taggart AK, Waters MG. Role of GPR81 in lactate-mediated reduction of adipose lipolysis. *Biochem Biophys Res Commun.* 2008; 377:987–91.
<https://doi.org/10.1016/j.bbrc.2008.10.088>
PMID:[18952058](https://pubmed.ncbi.nlm.nih.gov/18952058/)
- Feng J, Yang H, Zhang Y, Wei H, Zhu Z, Zhu B, Yang M, Cao W, Wang L, Wu Z. Tumor cell-derived lactate induces TAZ-dependent upregulation of PD-L1 through GPR81 in human lung cancer cells. *Oncogene.* 2017; 36:5829–39.
<https://doi.org/10.1038/onc.2017.188> PMID:[28604752](https://pubmed.ncbi.nlm.nih.gov/28604752/)
- de Castro Abrantes H, Briquet M, Schmuziger C, Restivo L, Puyal J, Rosenberg N, Rocher AB, Offermanns S, Chatton JY. The Lactate Receptor HCAR1 Modulates Neuronal Network Activity through the Activation of G α and G $\beta\gamma$ Subunits. *J Neurosci.* 2019; 39:4422–33.
<https://doi.org/10.1523/JNEUROSCI.2092-18.2019>
PMID:[30926749](https://pubmed.ncbi.nlm.nih.gov/30926749/)
- Fan H, Yang F, Xiao Z, Luo H, Chen H, Chen Z, Liu Q, Xiao Y. Lactylation: novel epigenetic regulatory and therapeutic opportunities. *Am J Physiol Endocrinol Metab.* 2023; 324:E330–8.
<https://doi.org/10.1152/ajpendo.00159.2022>
PMID:[36856188](https://pubmed.ncbi.nlm.nih.gov/36856188/)
- Kiaer T, Grønlund J, Sørensen KH. Subchondral pO₂, pCO₂, pressure, pH, and lactate in human osteoarthritis of the hip. *Clin Orthop Relat Res.* 1988; 149–55.
PMID:[3127099](https://pubmed.ncbi.nlm.nih.gov/3127099/)
- Hurter K, Spreng D, Rytz U, Schawalder P, Ott-Knüseler F, Schmökel H. Measurements of C-reactive protein in serum and lactate dehydrogenase in serum and synovial fluid of patients with osteoarthritis. *Vet J.* 2005; 169:281–5.
<https://doi.org/10.1016/j.tvjl.2004.01.027>
PMID:[15727922](https://pubmed.ncbi.nlm.nih.gov/15727922/)
- Chou CH, Wu CC, Song IW, Chuang HP, Lu LS, Chang JH, Kuo SY, Lee CH, Wu JY, Chen YT, Kraus VB, Lee MT. Genome-wide expression profiles of subchondral bone in osteoarthritis. *Arthritis Res Ther.* 2013; 15:R190.
<https://doi.org/10.1186/ar4380> PMID:[24229462](https://pubmed.ncbi.nlm.nih.gov/24229462/)
- Fisch KM, Gamini R, Alvarez-Garcia O, Akagi R, Saito M, Muramatsu Y, Sasho T, Koziol JA, Su AI, Lotz MK. Identification of transcription factors responsible for dysregulated networks in human osteoarthritis cartilage by global gene expression analysis. *Osteoarthritis Cartilage.* 2018; 26:1531–8.

- <https://doi.org/10.1016/j.joca.2018.07.012>
PMID:[30081074](https://pubmed.ncbi.nlm.nih.gov/30081074/)
18. Broeren MG, de Vries M, Bennink MB, van Lent PL, van der Kraan PM, Koenders MI, Thurlings RM, van de Loo FA. Functional Tissue Analysis Reveals Successful Cryopreservation of Human Osteoarthritic Synovium. *PLoS One*. 2016; 11:e0167076.
<https://doi.org/10.1371/journal.pone.0167076>
PMID:[27870898](https://pubmed.ncbi.nlm.nih.gov/27870898/)
 19. Guo Y, Walsh AM, Fearon U, Smith MD, Wechalekar MD, Yin X, Cole S, Orr C, McGarry T, Canavan M, Kelly S, Lin TA, Liu X, et al. CD40L-Dependent Pathway Is Active at Various Stages of Rheumatoid Arthritis Disease Progression. *J Immunol*. 2017; 198:4490–501.
<https://doi.org/10.4049/jimmunol.1601988>
PMID:[28455435](https://pubmed.ncbi.nlm.nih.gov/28455435/)
 20. Zhou R, Liang J, Chen Q, Tian H, Yang C, Liu C. A 3-Genes Random Forest Model to Diagnose Non-obstructive Azoospermia Based on Transcription Factor-Related Genes. *Reprod Sci*. 2023; 30:233–46.
<https://doi.org/10.1007/s43032-022-01008-8>
PMID:[35715550](https://pubmed.ncbi.nlm.nih.gov/35715550/)
 21. Kubelt C, Peters S, Ahmeti H, Huhndorf M, Huber L, Cohrs G, Hövener JB, Jansen O, Synowitz M, Held-Feindt J. Intratumoral Distribution of Lactate and the Monocarboxylate Transporters 1 and 4 in Human Glioblastoma Multiforme and Their Relationships to Tumor Progression-Associated Markers. *Int J Mol Sci*. 2020; 21:6254.
<https://doi.org/10.3390/ijms21176254>
PMID:[32872409](https://pubmed.ncbi.nlm.nih.gov/32872409/)
 22. Ou D, Liu S, Tong C, Tan H, Yang Y, He C. LIM mineralization protein-1 inhibits IL-1 β -induced human chondrocytes injury by altering the NF- κ B and MAPK/JNK pathways. *Exp Ther Med*. 2022; 23:61.
<https://doi.org/10.3892/etm.2021.10983>
PMID:[34934432](https://pubmed.ncbi.nlm.nih.gov/34934432/)
 23. Liu Y, Li Q, Gao Z, Lei F, Gao X. Circ-SPG11 knockdown hampers IL-1 β -induced osteoarthritis progression via targeting miR-337-3p/ADAMTS5. *J Orthop Surg Res*. 2021; 16:392.
<https://doi.org/10.1186/s13018-021-02526-y>
PMID:[34140036](https://pubmed.ncbi.nlm.nih.gov/34140036/)
 24. Guan Z, Jin X, Guan Z, Liu S, Tao K, Luo L. The gut microbiota metabolite capsiate regulate SLC2A1 expression by targeting HIF-1 α to inhibit knee osteoarthritis-induced ferroptosis. *Aging Cell*. 2023; 22:e13807.
<https://doi.org/10.1111/acer.13807>
PMID:[36890785](https://pubmed.ncbi.nlm.nih.gov/36890785/)
 25. Yao S, Deng M, Du X, Huang R, Chen Q. A Novel Hypoxia Related Marker in Blood Link to Aid Diagnosis and Therapy in Osteoarthritis. *Genes (Basel)*. 2022; 13:1501.
<https://doi.org/10.3390/genes13091501>
PMID:[36140669](https://pubmed.ncbi.nlm.nih.gov/36140669/)
 26. Xuan A, Chen H, Chen T, Li J, Lu S, Fan T, Zeng D, Wen Z, Ma J, Hunter D, Ding C, Zhu Z. The application of machine learning in early diagnosis of osteoarthritis: a narrative review. *Ther Adv Musculoskelet Dis*. 2023; 15:1759720X231158198.
<https://doi.org/10.1177/1759720X231158198>
PMID:[36937823](https://pubmed.ncbi.nlm.nih.gov/36937823/)
 27. Vidal-Bralo L, Lopez-Golan Y, Mera-Varela A, Rego-Perez I, Horvath S, Zhang Y, Del Real Á, Zhai G, Blanco FJ, Riancho JA, Gomez-Reino JJ, Gonzalez A. Specific premature epigenetic aging of cartilage in osteoarthritis. *Aging (Albany NY)*. 2016; 8:2222–31.
<https://doi.org/10.18632/aging.101053>
PMID:[27689435](https://pubmed.ncbi.nlm.nih.gov/27689435/)
 28. Pandey D, Onkara Perumal P. A scoping review on deep learning for next-generation RNA-Seq. data analysis. *Funct Integr Genomics*. 2023; 23:134.
<https://doi.org/10.1007/s10142-023-01064-6>
PMID:[37084004](https://pubmed.ncbi.nlm.nih.gov/37084004/)
 29. Xu J, Chen K, Yu Y, Wang Y, Zhu Y, Zou X, Jiang Y. Identification of Immune-Related Risk Genes in Osteoarthritis Based on Bioinformatics Analysis and Machine Learning. *J Pers Med*. 2023; 13:367.
<https://doi.org/10.3390/jpm13020367>
PMID:[36836601](https://pubmed.ncbi.nlm.nih.gov/36836601/)
 30. Chang B, Hu Z, Chen L, Jin Z, Yang Y. Development and validation of cuproptosis-related genes in synovitis during osteoarthritis progress. *Front Immunol*. 2023; 14:1090596.
<https://doi.org/10.3389/fimmu.2023.1090596>
PMID:[36817415](https://pubmed.ncbi.nlm.nih.gov/36817415/)
 31. Xia L, Gong N. Identification and verification of ferroptosis-related genes in the synovial tissue of osteoarthritis using bioinformatics analysis. *Front Mol Biosci*. 2022; 9:992044.
<https://doi.org/10.3389/fmolb.2022.992044>
PMID:[36106017](https://pubmed.ncbi.nlm.nih.gov/36106017/)
 32. 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 (Albany NY)*. 2023; 15:193–212.
<https://doi.org/10.18632/aging.204469> PMID:[36641761](https://pubmed.ncbi.nlm.nih.gov/36641761/)
 33. Xie X, Zhang Y, Yu J, Jiang F, Wu C. Significance of m⁶A regulatory factor in gene expression and immune function of osteoarthritis. *Front Physiol*. 2022; 13:918270.
<https://doi.org/10.3389/fphys.2022.918270>
PMID:[36160850](https://pubmed.ncbi.nlm.nih.gov/36160850/)

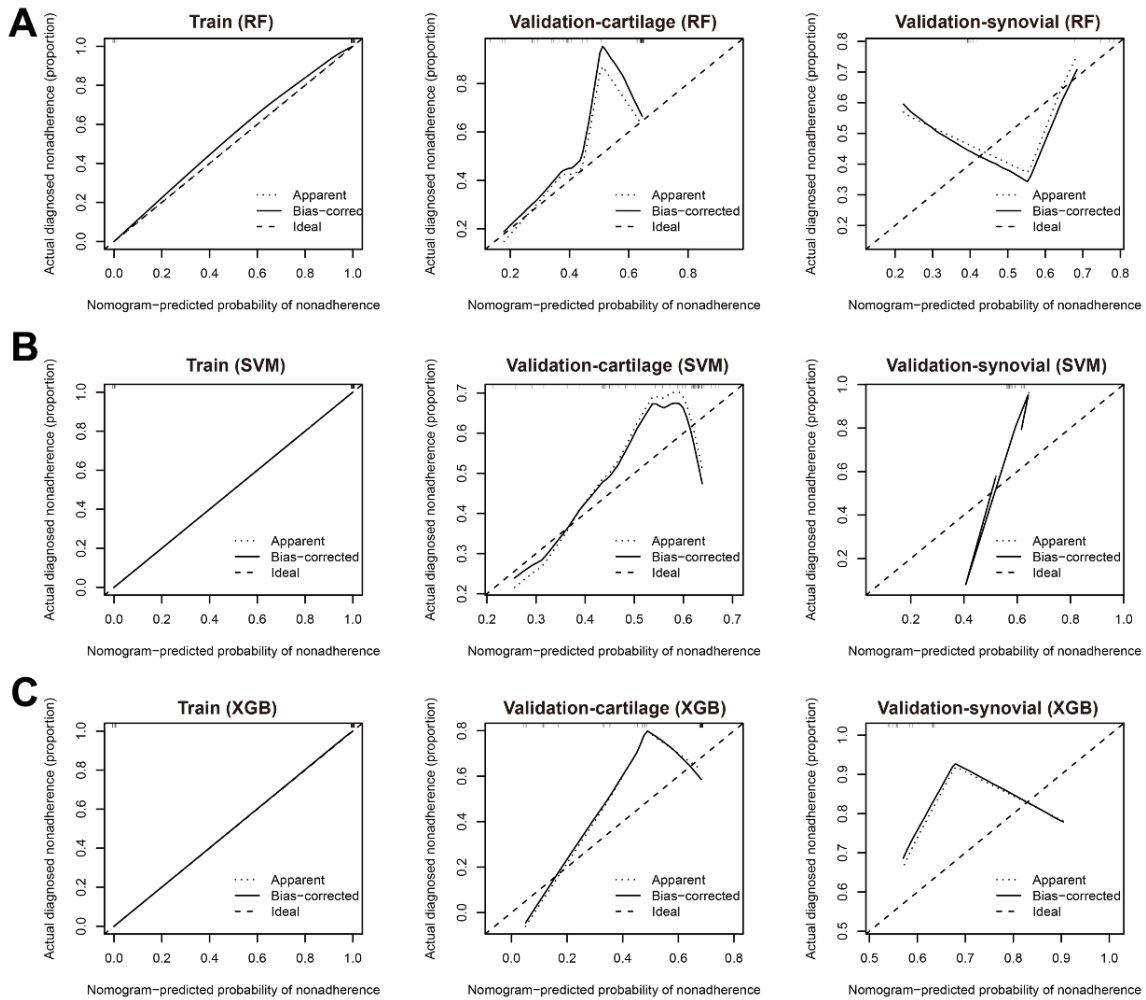
34. Liu W, Wang Y, Bozi LH, Fischer PD, Jedrychowski MP, Xiao H, Wu T, Darabedian N, He X, Mills EL, Burger N, Shin S, Reddy A, et al. Lactate regulates cell cycle by remodelling the anaphase promoting complex. *Nature*. 2023; 616:790–7.
<https://doi.org/10.1038/s41586-023-05939-3>
PMID:[36921622](https://pubmed.ncbi.nlm.nih.gov/36921622/)
35. Ivashkiv LB. The hypoxia-lactate axis tempers inflammation. *Nat Rev Immunol*. 2020; 20:85–6.
<https://doi.org/10.1038/s41577-019-0259-8>
PMID:[31819164](https://pubmed.ncbi.nlm.nih.gov/31819164/)
36. Cao H, Zhou X, Xu B, Hu H, Guo J, Wang M, Li N, Jun Z. Advances in the study of mitophagy in osteoarthritis. *J Zhejiang Univ Sci B*. 2024; 25:197–211.
<https://doi.org/10.1631/jzus.B2300402>
PMID:[38453635](https://pubmed.ncbi.nlm.nih.gov/38453635/)
37. Coaccioli S, Sarzi-Puttini P, Zis P, Rinonapoli G, Varrassi G. Osteoarthritis: New Insight on Its Pathophysiology. *J Clin Med*. 2022; 11:6013.
<https://doi.org/10.3390/jcm11206013> PMID:[36294334](https://pubmed.ncbi.nlm.nih.gov/36294334/)
38. Zheng L, Chen W, Xian G, Pan B, Ye Y, Gu M, Ma Y, Zhang Z, Sheng P. Identification of abnormally methylated-differentially expressed genes and pathways in osteoarthritis: a comprehensive bioinformatic study. *Clin Rheumatol*. 2021; 40:3247–56.
<https://doi.org/10.1007/s10067-020-05539-w>
PMID:[33420869](https://pubmed.ncbi.nlm.nih.gov/33420869/)
39. Li L, Wang S, Zhang Y, Pan J. The E2F transcription factor 2: What do we know? *Biosci Trends*. 2021; 15:83–92.
<https://doi.org/10.5582/bst.2021.01072>
PMID:[33952804](https://pubmed.ncbi.nlm.nih.gov/33952804/)
40. Pellicelli M, Picard C, Wang D, Lavigne P, Moreau A. E2F1 and TFDP1 Regulate PITX1 Expression in Normal and Osteoarthritic Articular Chondrocytes. *PLoS One*. 2016; 11:e0165951.
<https://doi.org/10.1371/journal.pone.0165951>
PMID:[27802335](https://pubmed.ncbi.nlm.nih.gov/27802335/)
41. Lin X, Wells DE, Kimberling WJ, Kumar S. Human NDUFB9 gene: genomic organization and a possible candidate gene associated with deafness disorder mapped to chromosome 8q13. *Hum Hered*. 1999; 49:75–80.
<https://doi.org/10.1159/000022848> PMID:[10077726](https://pubmed.ncbi.nlm.nih.gov/10077726/)
42. Li LD, Sun HF, Liu XX, Gao SP, Jiang HL, Hu X, Jin W. Down-Regulation of NDUFB9 Promotes Breast Cancer Cell Proliferation, Metastasis by Mediating Mitochondrial Metabolism. *PLoS One*. 2015; 10:e0144441.
<https://doi.org/10.1371/journal.pone.0144441>
PMID:[26641458](https://pubmed.ncbi.nlm.nih.gov/26641458/)
43. Yang X, Jiang Q, Luan T, Yu C, Liu Z, Wang T, Wan J, Huang J, Li K. Pyruvate Dehydrogenase Kinase 1 inhibition mediated oxidative phosphorylation enhancement in cartilage promotes osteoarthritis progression. *BMC Musculoskelet Disord*. 2023; 24:597.
<https://doi.org/10.1186/s12891-023-06585-6>
PMID:[37474941](https://pubmed.ncbi.nlm.nih.gov/37474941/)
44. Karmand H, Andishgar A, Tabrizi R, Sadeghi A, Pezeshki B, Ravankhah M, Taherifard E, Ahmadizar F. Machine-learning algorithms in screening for type 2 diabetes mellitus: Data from Fasa Adults Cohort Study. *Endocrinol Diabetes Metab*. 2024; 7:e00472.
<https://doi.org/10.1002/edm2.472> PMID:[38411386](https://pubmed.ncbi.nlm.nih.gov/38411386/)

SUPPLEMENTARY MATERIALS

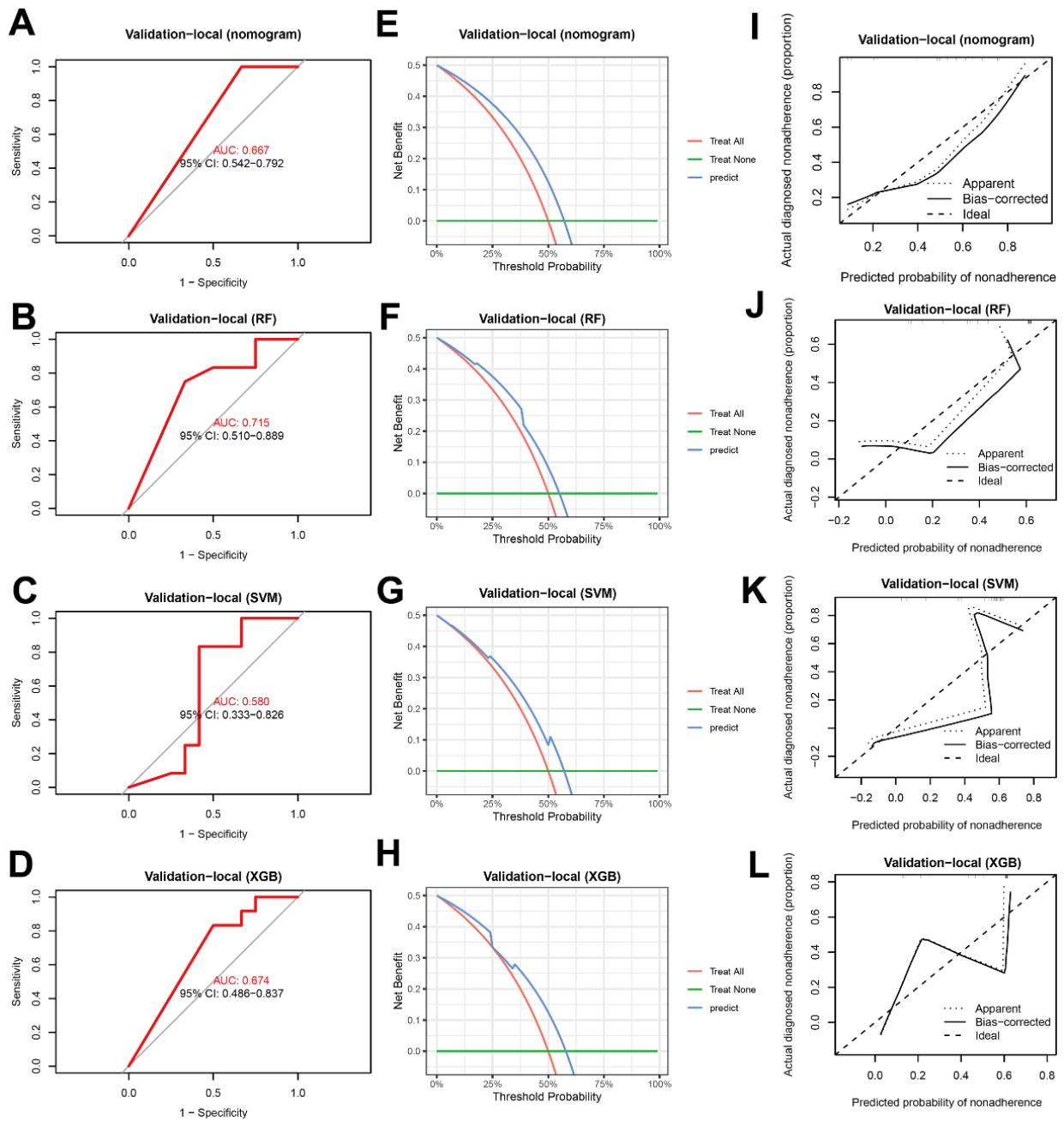
Supplementary Figures



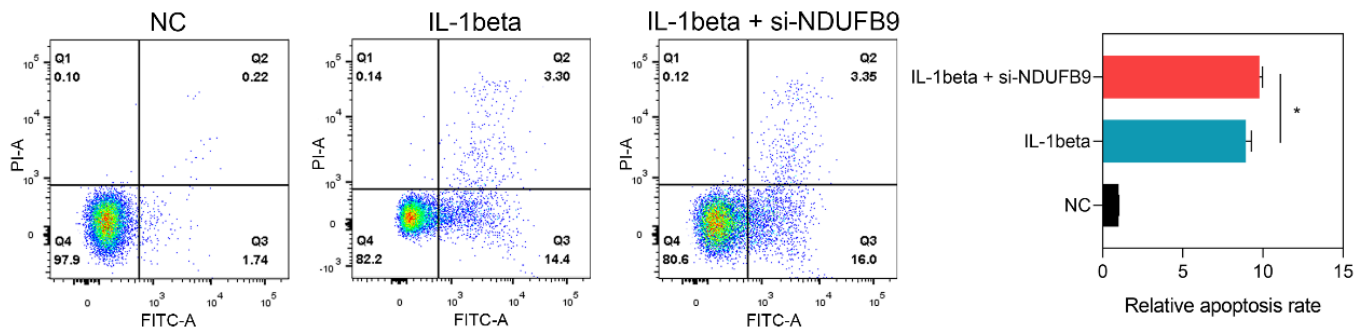
Supplementary Figure 1. The levels of lactate in the knee cartilage tissues collected from the control and OA subjects. Abbreviation: OA, osteoarthritis.



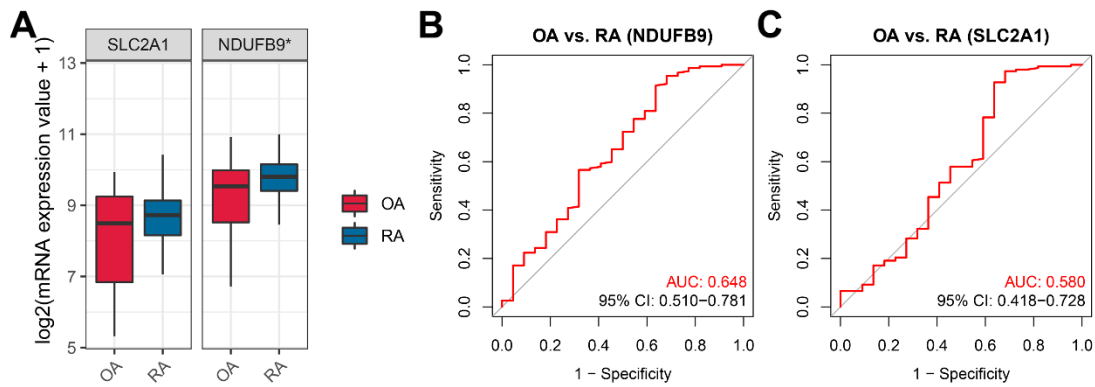
Supplementary Figure 2. The calibration analysis of the RF (A), SVM (B), and XGB (C) models. Abbreviations: RF, random forest; SVM, supporter vector machine; XGB, XGBoost.



Supplementary Figure 3. The predictive performance of the established models in the local cohort. (A–D) The ROC analysis indicates the predictive performance of the nomogram (A), RF (B), SVM (C), and XGB (D) models in the validation-local cohort. (E–H) DCA displays the net benefit under different thresholds of the nomogram (E), RF (F), SVM (G), and XGB (H) models in the local cohort. (I–L) The calibration analysis shows the predictive performance of the nomogram (I), RF (J), SVM (K), and XGB (L) models in the local cohort. Abbreviations: ROC, receiver operating characteristic; DCA, decision curve analysis; AUC, area under curve; CI, confidence interval.



Supplementary Figure 4. Flow cytometry apoptosis analysis.



Supplementary Figure 5. The ability of SLC2A1 and NDUFB9 to distinguish OA from RA. (A) The expression levels of SLC2A1 and NDUFB9 in the joint synovial tissues from 22 OA and 152 RA samples. (B, C) ROC analysis indicates the ability of NDUFB9 (B) and SLC2A1 (C) to distinguish OA from RA.

Supplementary Tables

Please browse Full Text version to see the data of Supplementary Tables 6, 12.

Supplementary Table 1. 273 lactate metabolism-related genes collected from the MSigDB.

Gene symbol
HMOX1, NGLY1, CLPB, NDUFB9, FKTN, MLIP, HIBCH, PDP1, SLC2A1, POMT1, SLC4A1, CYC1, COQ2, SLC16A8, CFI, SLC25A19, LDHB, RHAG, ALDOA, C1QBP, GATA1, SLC25A12, MRPL12, XK, NDUFB8, LDHA, NDUFAF3, TIMM50, NDUFV2, TUFM, EMB, COX8A, MYC, PYGL, MDH2, HTRA2, CYP27A1, HAGH, PLEC, GATA2, HSD17B10, PDSS1, SCO1, NDUFS2, CHCHD10, MTHFD1, MRPS22, GYS2, UQCRH, KLF1, SDHB, DGUOK, ISCA1, NDUFS1, AIFM1, RPS14, NDUFS6, HBB, PUS1, RNASEH1, COX5A, SUCLG1, JAK2, NDUFA6, ADAMTS13, NDUFA2, PIGA, PDHX, STAT2, SLC19A1, RHD, PNPLA8, IRAK1, MRPS7, UQCRC2, HPDL, ABCG8, COX6B1, COX4I1, POMT2, SOD1, NDUFAF2, NDUFA13, SFXN4, LIAS, LETM1, ZNFX1, LONP1, SLC25A10, SLC25A3, NDUFAF4, MRPS34, ACAT2, MRPS2, SLC5A8, COX10, ECHS1, PMPCB, PNPLA2, SLC16A1, ACAT1, PNPT1, NDUFB11, ATAD3A, RHCE, COQ4, NUBPL, PER2, FLI1, MRPL3, TRMT5, LRPPRC, SPP1, CAV3, RRM2B, SDHA, COX15, POMGNT1, DNMI1, NDUFS3, EARS2, MRPS28, LYST, INPP5K, WARS2, TK2, PNPO, NARS2, FASTKD2, YARS2, NDUFS8, SLC25A4, KY, GAA, NDUFA12, NDUFA1, MIPEP, GOT2, CA5A, ACAD9, ALDH4A1, NDUFB3, MPV17, PITRM1, NDUFS7, SLC16A7, AGK, UQCRQ, AARS2, PDHA1, TET2, DLD, SLC13A3, CDAN1, MVK, PHKG2, GTPBP3, PARK7, BCS1L, PLA2G6, LIPT2, FBXL4, OGDH, LYRM4, LIPT1, TRMU, PIK3CG, KIF23, NDUFA8, TXN2, SLC7A7, TSFM, POLRMT, NDUFB7, NFS1, KCNN4, VARS2, CALR, SLC16A3, COX6A2, FARS2, TCIRG1, NDUFA11, SIL1, NDUFB10, LYRM7, SLC19A3, OCRL, STAT4, CD46, LIPA, ATPAF2, DARS2, MPL, POLG2, TTC26, NDUFAF1, NSUN2, GFM2, PDSS2, GFM1, UQCRB, LPIN1, SURF1, COX16, MTO1, SYNJ1, PPCS, HLA-DRB1, SLC25A42, ACTN3, AKR1D1, CHEK2, VPS13A, PC, COL4A1, NDUFC2, MRPS14, SLC25A26, B3GALNT2, NDUFA4, HIF1A, TP53, DTYMK, CPT2, FKR, RARS2, LDHC, FOXRED1, DAG1, RMND1, HMGCL, TMEM126B, COQ9, PIEZO1, TACO1, MECR, TIMMDC1, DNAJC19, NDUFA10, ISCU, SLC39A8, MECP2, TIMM22, HS6ST2, NSUN3, TPK1, HMGCS2, NDUFS4, CFH, ACADM, MTFMT, MRPS16, ALDH6A1, LDHAL6B, LDHD, SCO2, COG8, NFU1, SLC25A13, CARS2, TARS2, PNKD, NDUFV1, UPB1, LDHAL6A, SERAC1, RB1, POLG, NDUFA9, MRPL44, PFKFB2, SLC5A12.

Supplementary Table 2. The information of the datasets from the GEO.

ID	Dataset type	Tissues	Platform	Experimental type	Sample size (Control/OA)	PMID
GSE51588	Training	Cartilage tissue	GPL13497	Expression profiling by array	10/40	24229462
GSE114007	Validation	Cartilage tissue	GPL11154 GPL18573	Expression profiling by high throughput sequencing	18/20	30081074
GSE117999	Validation	Cartilage tissue	GPL20844	Expression profiling by array	12/12	NA
GSE82107	Validation	Synovial tissue	GPL570	Expression profiling by array	7/10	27870898
GSE89408	-	Synovial tissue	GPL11154	Expression profiling by high throughput sequencing	OA: 22 RA: 152	28455435 28863153

Supplementary Table 3. The baseline clinicopathological features of the control and OA subjects in the local cohort.

Feature	Control (N = 12)	OA (N = 12)
Age	54.17 ± 7.8	60.08 ± 8.1
Sex		
Male	8 (66.67%)	6 (50.00%)
Female	4 (33.33%)	6 (50.00%)
BMI	29.17 ± 7.5	32.08 ± 9.4
Kellgren-Lawrence Grade		
I	-	0 (0.00%)
II	-	0 (0.00%)
III	-	3 (25.00%)
IV	-	9 (75.00%)

Supplementary Table 4. The siRNA sequences used in this study.

ID	Sequence (5'-3')
si-NDUFB9-1	GTCCCAGAATGGTGCTTAGAT
si-NC-1	GCGTGTCCAATTCGAGTAGAT
si-NDUFB9-2	GGCAATGTATCCTGATTAATT
si-NC-2	GACTTACTCGGCGTTATAATT

Supplementary Table 5. The primer sequence of the qPCR experiments.

Gene	Sequence (5'-3')
SLC2A1	Forward Primer: GGCCAAGAGTGTGCTAAAGAA Reverse Primer: ACAGCGTTGATGCCAGACAG
NDUFB9	Forward Primer: GTGGTGCCTCCAGAGAGAC Reverse Primer: GGCCTTCGCCATATCCTTTTC
GAPDH	Forward Primer: GGAGCGAGATCCCTCCAAAAT Reverse Primer: GGCTGTTGTCATACTTCTCATGG

Supplementary Table 6. The expression difference of 273 lactate metabolism-related genes between the control and OA samples.

Supplementary Table 7. The coefficients of the genes in the LASSO regression model.

Gene	Coefficients
XK	-0.466813114
GATA1	-0.297177278
HMOX1	-1.044903664
SLC2A1	-0.306222582
CYC1	-0.009444988
HAGH	-0.09719996
NDUFB9	-3.705789774
HTRA2	-0.00717518
ISCA1	-0.870643408
HIBCH	0.211640847
FKTN	0.211721146
PDP1	3.359955862
SLC16A8	0.103277647

Supplementary Table 8. The genes identified by the SVM-RFE algorithm.

Gene
NDUFB9, CYC1, HMOX1, PYGL, CFI, FKTN, MLIP, SLC2A1, CYP27A1, CLPB, NDUFS2, RHAG, HIBCH.

Supplementary Table 9. The results obtained from the Boruta algorithm.

Gene	Status
RHAG	Confirmed
XK	Confirmed
GATA1	Confirmed
SLC4A1	Confirmed
RHCE	Confirmed
KLF1	Confirmed
RHD	Confirmed
HMOX1	Confirmed
LDHA	Confirmed
EMB	Confirmed
SLC2A1	Confirmed
PYGL	Confirmed
CLPB	Confirmed
CYC1	Confirmed
C1QBP	Confirmed
HAGH	Confirmed
MYC	Confirmed
LDHB	Confirmed
NDUFB9	Confirmed
NDUFS2	Confirmed
MRPL12	Confirmed
ALDOA	Confirmed

COX8A	Confirmed
COQ2	Confirmed
NDUFV2	Confirmed
HTRA2	Confirmed
CHCHD10	Confirmed
MDH2	Confirmed
NDUFAF2	Confirmed
NDUFAF3	Confirmed
TUFM	Confirmed
TIMM50	Confirmed
PDHX	Confirmed
MTHFD1	Confirmed
SLC25A12	Confirmed
CYP27A1	Confirmed
ADAMTS13	Confirmed
POMT1	Confirmed
HIBCH	Confirmed
FKTN	Confirmed
PDP1	Confirmed
CFI	Confirmed
MLIP	Confirmed
SLC16A8	Confirmed
PNPLA2	Rejected
HSD17B10	Rejected
JAK2	Rejected
SLC5A8	Rejected
LONP1	Rejected
MRPL3	Rejected
NDUFAF4	Rejected
PNPLA8	Rejected
SUCLG1	Rejected
CAV3	Rejected
PER2	Rejected
SPP1	Rejected
ABCG8	Rejected
GYS2	Tentative
HBB	Tentative
SDHB	Tentative
UQCRH	Tentative
NDUFA6	Tentative
COX5A	Tentative
PUS1	Tentative
RNASEH1	Tentative
MRPS7	Tentative
SLC16A1	Tentative
COX10	Tentative
HPDL	Tentative
ISCA1	Tentative
PIGA	Tentative
COX4I1	Tentative
PLEC	Tentative
GATA2	Tentative

Supplementary Table 10. The univariate logistic regression of the differentially-expressed genes associated with lactate metabolism.

Gene	OR (95%CI)	P-value
SLC2A1	0.037 (0.004-0.173)	<0.001
SLC25A12	143.349 (11.744-4357.33)	0.001
HIBCH	1172.193 (41.919-208455.093)	0.001
LDHB	0.021 (0.001-0.140)	0.001
COQ2	0.003 (0.001-0.054)	0.001
TIMM50	0.002 (0.001-0.047)	0.001
COX8A	0.009 (0.001-0.099)	0.001
MTHFD1	0.004 (0.001-0.068)	0.001
SLC16A8	20.916 (4.512-199.328)	0.001
C1QBP	0.021 (0.001-0.145)	0.001
MDH2	0.001 (0.001-0.037)	0.001
MRPL12	0.017 (0.001-0.135)	0.001
MYC	0.038 (0.003-0.207)	0.002
PLEC	37.541 (5.505-573.250)	0.002
NDUFA6	0.010 (0.001-0.113)	0.002
CYP27A1	258.300 (14.192-19613.825)	0.002
ISCA1	0.005 (0.001-0.089)	0.002
NDUFV2	0.003 (0.001-0.064)	0.002
TUFM	0.001 (0.001-0.021)	0.002
NDUFAF3	0.001 (0.001-0.012)	0.002
SUCLG1	0.013 (0.001-0.147)	0.002
RHAG	0.535 (0.324-0.747)	0.002
HBB	0.241 (0.081-0.538)	0.002
GATA2	10.556 (2.938-69.825)	0.002
CHCHD10	0.025 (0.001-0.190)	0.002
GYS2	0.226 (0.073-0.523)	0.003
PDHX	0.053 (0.006-0.304)	0.003
XK	0.450 (0.236-0.697)	0.003
EMB	0.190 (0.051-0.477)	0.003
ADAMTS13	42.235 (5.112-805.013)	0.003
HAGH	0.106 (0.017-0.375)	0.003
ALDOA	0.003 (0.001-0.063)	0.003
NDUFS2	0.001 (0.001-0.014)	0.003
UQCRH	0.018 (0.001-0.166)	0.003
NDUFB9	0.001 (0.001-0.004)	0.003
SLC4A1	0.293 (0.105-0.565)	0.003
LDHA	0.142 (0.028-0.406)	0.003
HPDL	0.018 (0.001-0.184)	0.003
SDHB	0.013 (0.001-0.135)	0.003
RNASEH1	0.005 (0.001-0.098)	0.004
MLIP	345.260 (17.277-80337.860)	0.004
CLPB	0.001 (0.001-0.039)	0.004
GATA1	0.435 (0.214-0.689)	0.004
PUS1	0.039 (0.003-0.252)	0.004
PYGL	0.009 (0.001-0.116)	0.004
COX4I1	0.021 (0.001-0.206)	0.004

MRPS7	0.002 (0.001-0.063)	0.004
HSD17B10	0.029 (0.001-0.209)	0.004
COX10	0.003 (0.001-0.090)	0.004
PIGA	0.050 (0.004-0.311)	0.005
HTRA2	0.001 (0.001-0.009)	0.005
PDP1	181.034 (10.849-17070.709)	0.005
COX5A	0.020 (0.001-0.188)	0.005
KLF1	0.451 (0.224-0.728)	0.005
JAK2	0.069 (0.007-0.348)	0.006
NDUFAF4	0.113 (0.020-0.481)	0.006
PNPLA8	0.063 (0.005-0.353)	0.006
NDUFAF2	0.112 (0.017-0.464)	0.007
RHD	0.514 (0.279-0.791)	0.008
FKTN	119875.058 (245.767-44103618194.23)	0.009
LONP1	0.051 (0.003-0.341)	0.010
ABCG8	10.071 (2.296-80.359)	0.010
SLC5A8	0.190 (0.045-0.611)	0.010
SLC16A1	0.126 (0.021-0.532)	0.010
MRPL3	0.146 (0.028-0.567)	0.010
SPP1	3.812 (1.445-12.166)	0.012
PER2	5.184 (1.569-22.371)	0.012
PNPLA2	0.147 (0.023-0.521)	0.013
RHCE	0.688 (0.492-0.913)	0.014
POMT1	15336.767 (60.167-493515289.745)	0.014
CAV3	10.665 (2.035-102.238)	0.014
HMOX1	0.001 (0.001-0.021)	0.017
CFI	708510.387 (272.556-4445894389564.95)	0.017
CYC1	0.001 (0.000-0.001)	0.217

Supplementary Table 11. The samples in the C1 and C2 clusters.

Cluster	Sample ID
C1	GSM1248759, GSM1248760, GSM1248761, GSM1248762, GSM1248764, GSM1248765, GSM1248766, GSM1248767.
C2	GSM1248763, GSM1248768, GSM1248769, GSM1248770, GSM1248771, GSM1248772, GSM1248773, GSM1248774, GSM1248775, GSM1248776, GSM1248777, GSM1248778, GSM1248779, GSM1248780, GSM1248781, GSM1248782, GSM1248783, GSM1248784, GSM1248785, GSM1248786, GSM1248787, GSM1248788, GSM1248789, GSM1248790, GSM1248791, GSM1248792, GSM1248793, GSM1248794, GSM1248795, GSM1248796, GSM1248797, GSM1248798, GSM1248799, GSM1248800, GSM1248801, GSM1248802, GSM1248803, GSM1248804, GSM1248805, GSM1248806, GSM1248807, GSM1248808.

Supplementary Table 12. The differentially-expressed genes between C1 and C2 subjects.