



OPEN ACCESS

EDITED BY

Musie Ghebremichael,
Harvard University, United States

REVIEWED BY

Feng Jiang,
Fudan University, China
Jun Jiang,
Fudan University, China

*CORRESPONDENCE

Pengjie Song
✉ 675680048@qq.com
Haiyu Zhou
✉ zhouhy@lzu.edu.cn

†These authors have contributed equally to this work

RECEIVED 31 May 2024

ACCEPTED 11 November 2024

PUBLISHED 04 December 2024

CITATION

Lin M, Li S, Wang Y, Zheng G, Hu F, Zhang Q, Song P and Zhou H (2024) Machine learning-based diagnostic model of lymphatics-associated genes for new therapeutic target analysis in intervertebral disc degeneration. *Front. Immunol.* 15:1441028. doi: 10.3389/fimmu.2024.1441028

COPYRIGHT

© 2024 Lin, Li, Wang, Zheng, Hu, Zhang, Song and Zhou. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Machine learning-based diagnostic model of lymphatics-associated genes for new therapeutic target analysis in intervertebral disc degeneration

Maoqiang Lin^{1,2†}, Shaolong Li^{1†}, Yabin Wang^{1,2†}, Guan Zheng^{1,2}, Fukang Hu^{1,2}, Qiang Zhang^{1,2}, Pengjie Song^{1*} and Haiyu Zhou^{1,2*}

¹Department of Orthopedics, The Second Hospital & Clinical Medical School, Lanzhou University, Lanzhou, Gansu, China, ²Key Laboratory of Bone and Joint Disease Research of Gansu Province, Lanzhou, Gansu, China

Background: Low back pain resulting from intervertebral disc degeneration (IVDD) represents a significant global social problem. There are notable differences in the distribution of lymphatic vessels (LV) in normal and pathological intervertebral discs. Nevertheless, the molecular mechanisms of lymphatics-associated genes (LAGs) in the development of IVDD remain unclear. An in-depth exploration of this area will help to reveal the biological and clinical significance of LAGs in IVDD and may lead to the search for new therapeutic targets for IVDD.

Methods: Data sets were obtained from the Gene Expression Omnibus (GEO) database. Following quality control and normalization, the datasets (GSE153761, GSE147383, and GSE124272) were merged to form the training set, with GSE150408 serving as the validation set. LAGs from GeneCards, MSigDB, Gene Ontology, and KEGG database. The Venn diagram was employed to identify differentially expressed lymphatic-associated genes (DELAGs) that were differentially expressed in the normal and IVDD groups. Subsequently, four machine learning algorithms (SVM-RFE, Random Forest, XGB, and GLM) were used to select the method to construct the diagnostic model. The receiver operating characteristic (ROC) curve, nomogram, and Decision Curve Analysis (DCA) were used to evaluate the model effect. In addition, we constructed a potential drug regulatory network and competitive endogenous RNA (ceRNA) network for key LAGs.

Results: A total of 15 differentially expressed LAGs were identified. By comparing four machine learning methods, the top five genes of importance in the XGB model (MET, HHIP, SPRY1, CSF1, TOX) were identified as lymphatics-associated gene diagnostic signatures. This signature was used to predict the diagnosis of IVDD with strong accuracy and an area under curve (AUC) value of 0.938. Furthermore, the diagnostic model was validated in an external dataset (GSE150408), with an AUC value of 0.772. The nomogram and DCA further prove that the diagnosis model has good performance and predictive value. Additionally, drug regulatory networks and ceRNA networks were constructed, revealing potential therapeutic drugs and post-transcriptional regulatory mechanisms.

Conclusion: We developed and validated a lymphatics-associated genes diagnostic model by machine learning algorithms that effectively identify IVDD patients. These five key LAGs may be potential therapeutic targets for IVDD patients.

KEYWORDS

intervertebral disc degeneration, machine learning, diagnostic model, lymphatic-associated gene, therapeutic target

1 Introduction

The Global Burden of Disease study has reported that low back pain is one of the major causes of disability-adjusted life-years (DALYs) worldwide (1). Low back pain is a multifactorial outcome, and intervertebral disc degeneration (IVDD) is considered to be one of the major causes of low back pain. As evidenced by statistical analysis, discogenic low back pain constitutes approximately 26%–42% of cases (2). At present, the pathogenic mechanisms underlying intervertebral disc degeneration remain unclear. Identifying the various predisposing factors and developing diagnostic markers with etiologic specificity are the most pressing research priorities at this stage. These biomarkers can provide effective strategies for precise prevention and targeted treatment of IVDD.

The lymphatic system is part of the circulatory system of the body that is primarily responsible for the discharge of extracellular fluid containing cells, high- and low-molecular-weight proteins, and other molecules in the tissues (3). The normality of the lymphatic system determines the structure and reflux function of the lymphatic vessels (4, 5). Many diseases are associated with the function of the lymphatic system. For instance, the role of lymphatic reflux has been demonstrated in musculoskeletal disorders, such as osteoarthritis. Modulating lymphatic reflux has been shown to significantly reduce the progression of inflammatory arthritis (6, 7). Furthermore, an increasing number of studies have

indicated that lymphatic vessels are closely linked to the development of IVDD. It has been demonstrated that with the progression of IVDD, lymphatic vessels comprising LYVE1 +/podoplanin+ accompany the growth of vascularized fibrous tissue into the interior of the intervertebral disc (IVD) (8, 9). In addition, during IVDD, a variety of proteases (e.g., matrix metalloproteinases and aggrecan) and cytokines in the degenerating IVD stroma show elevated expression in response to the lack of lymphatic vessel supply, accelerating IVD prolapse (10, 11). At the same time, these factors also stimulate the proliferation and migration of blood and lymphatic vessels into the protruding IVD tissue in an autocrine manner (9). In addition, the lymphatic vasculature also plays an immunomodulatory role (12), and there is an apparent correlation between the progression of IVDD and immune cell infiltration (13). Thus, it is evident that lymphatic vessels play a crucial role in the IVDD process. However, the study of its underlying pathological mechanisms is still in its early stages. Further research into the biological significance of lymphatic-associated genes (LAGs) in IVDD may help accelerate the progress of molecular diagnosis and targeted therapy for this condition.

The selection of effective features is a crucial step in the discovery of biomarkers, which represents a fundamental task in this field of research. Machine learning is an effective strategy to address this problem, with its enormous resources to handle large, complex and diverse data, and has been applied to genomics research (14). Meanwhile, machine learning has enhanced our capacity to extract pertinent features from copious amounts of high-dimensional data pertaining to gene expression profiles (15), and has been extensively employed in medical biomarker screening (16, 17). However, there is still an unmet need for machine learning in the field of diagnosis and treatment of IVDD. In this study, based on the transcriptome data set and lymphatic vessel-related genes of IVDD patients obtained from the public database, we identified differentially expressed LAGs by bioinformatics analysis, and then used four machine learning algorithms to screen out lymphatic vessel gene biomarkers related to IVDD diagnosis, and used them as model genes to construct a diagnostic model. In addition, we also constructed the drug regulatory network and competitive endogenous RNA (ceRNA) network of model genes. The ultimate goal is to reveal the correlation between lymphatic vessel-related

Abbreviations: IVDD, intervertebral disc degeneration; LV, lymphatic vessel; LAGs, Lymphatic-associated genes; GEO, Gene Expression Omnibus; KEGG, Kyoto Encyclopedia of Genes and Genomes; XGB, Extreme Gradient Boosting; SVM, Support vector machines; GLM, Generalized linear models; DELAGs, Differentially expressed lymphatics-associated genes; ROC, Receiver operating characteristic; AUC, Area under Curve; ceRNA, competitive endogenous RNA; GO, Gene Ontology; DO, Disease Ontology; GSEA, Gene Set Enrichment Analysis; GSVA, Gene set variation analysis; DGIdb, Drug-Gene Interaction Database; ISG, interferon-stimulated gene; lncRNA, long non-coding RNA; miRNAs, microRNA; MMP, Matrix metalloproteinase; AF, annulus fibrosus; NP, nucleus pulposus; IVD, intervertebral disc; NR2F1-AS1, nuclear receptor subfamily 2 group F member 1 antisense RNA 1; MET, Mesenchymal to epithelial transition factor; HHIP, Hedgehog interacting protein; CSF1, Colony-stimulating factor-1; TOX, Thymocyte Selection Associated High Mobility Group Box.

genes and the pathogenesis of IVDD and to provide a reference for potential drugs and candidate targets in IVDD treatment.

2 Materials and methods

2.1 Dataset download and data preprocessing

The original gene expression profile data is derived from the Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/>), which is an international public repository. We identified three training sets (GSE124272, GSE147383, and GSE153761) and one validation set (GSE150408). GSE124272 included eight patients with IVDD and eight normal controls, GSE147383 included four patients with IVDD and four normal controls, and GSE153761 included three patients with cervical spondylotic myelopathy and three healthy subjects. GSE150408 included seventeen patients with sciatica and seventeen healthy volunteers. This data was mainly used to confirm the analysis results.

The GSE124272, GSE147383, and GSE153761 datasets were integrated by the “limma” and “sva” packages in R (4.4.0), and the combined dataset was used as our training set. The “normalizeBetweenArrays” function in the “limma” package of R language was used to normalize the expression matrix of the training set and the validation set, and the gene probes were annotated. A total of 302 LAGs were obtained from GeneCards (<https://www.genecards.org/>), MSigDB (<https://www.gsea-msigdb.org/>), Gene Ontology (<https://geneontology.org/>) and KEGG (<https://www.kegg.jp/kegg/>).

2.2 Characterization of immune infiltrating cells in the training set

We used twenty-two immune cells identified in previous studies (18, 19), analyzed the immune cell content of each sample in the training set using the “CIBERSORT” function in the “preprocessCore” package, and then The number of immune infiltrating cells, differential expression level, and correlation were visualized for each sample.

2.3 Identification and visualization of DEGs

DEGs within the training set were identified using the “limma” package, with screening criteria set at $|\log \text{ fold change (FC)}| > 0.5$ and $P \text{ value} < 0.05$. Subsequently, visualizations of volcano and heat maps were created with the R package using ggplot2 and pheatmap, respectively.

2.4 Identification and functional enrichment of differentially expressed LAGs

The overlap of DEGs and LAGs within IVDD was obtained using the “VennDiagram” package in R, defining them as

differentially expressed lymphatics-associated genes (DELAGs). Subsequently, the DELAGs underwent Gene Ontology (GO) analysis, Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis, Disease Ontology (DO) analysis, and Gene Set Enrichment Analysis (GSEA) using the “clusterProfiler” package in R. Enrichment results with a P -value less than 0.05 were considered statistically significant.

2.5 Screening of model genes by machine learning algorithms

Fifteen LAGs exhibiting differential expression were analyzed via machine learning algorithms. Four machine learning algorithms, including Extreme Gradient Boosting (XGB), Random Forest (RF), Support Vector Machine (SVM), and the Generalized Linear Model (GLM), were employed for the selection of key genes in IVDD diagnosis. We trained RF, SVM, XGB, and GLM models using the train function from the R “caret” package, in conjunction with the “randomForest”, “kernlab”, “xgboost”, and “stats” packages, respectively. Meanwhile, we leveraged the built-in grid search mechanism of the “caret” package to explore the optimal hyperparameter combinations (Table 1) and evaluated the performance of each combination through cross-validation. Additionally, residual boxplots, feature importance plots, reverse cumulative distribution of residuals, and Receiver Operating Characteristic (ROC) curves were constructed for XGB, RF, SVM, and GLM models to ascertain model genes.

2.6 Model construction and validation

Diagnostic model nomograms were constructed using model genes and the “rms” package in R software, with calibration curves employed to validate their accuracy. The efficacy of the Nomoclinical grading was assessed through Decision Curve Analysis (DCA). ROC curves were plotted to evaluate the diagnostic value of model genes in IVDD. Subsequently, an external validation cohort (GSE150408) was introduced to assess the diagnostic capability and robustness of the model.

TABLE 1 Hyperparameter tuning settings for each machine learning algorithm.

Machine learning algorithm	Hyperparameters
RF	mtry: 2, ntree: 500
SVM	sigma: 0.04, c: 0.25
XGB	nrounds: 100, max_depth: 2, eta: 0.4, gamma: 0, colsample_bytree: 0.8, min_child_weight: 1, subsample: 0.5
GLM	no specific

2.7 GSVA enrichment analysis of model genes and its correlation with immune cells

To unravel the potential biological functions of the model genes, we performed Gene set variation analysis (GSVA) by using the “clusterprofiler” package in the GO and KEGG genomes. Reference gene sets included c2.cp.kegg.Symbols.gmt and c5.go.Symbols.gmt. Enrichment results with a P-value less than 0.05 were deemed statistically significant. The “GSVA” package in R was employed to investigate the association between model genes and twenty-eight immune cells, with visualization of the results carried out using the “ggplot2” package.

2.8 Construction of drug regulatory networks and ceRNA networks

Interactions between model genes and drugs were obtained using the Drug-Gene Interaction database (DGIdb) (<https://www.dgldb.org/>), and the results were visualized using Cytoscape software. The results were analyzed using miRanda (<http://www.microrna.org/>), miRDB (<http://mirdb.org/>), miWalk (<http://mirwalk.umm.uni-heidelberg.de/>), and TargetScan (<http://www.targetscan.org/>) databases and predicted microRNAs (miRNAs) targeting model genes using Score=1 as a screening condition. SpongeScan database (<http://spongescan.rc.ufl.edu/>) was employed for predicting lncRNA-miRNA interactions (20).

Subsequently, Cytoscape software was utilized for the creation and visualization of lncRNA-miRNA-mRNA regulatory networks.

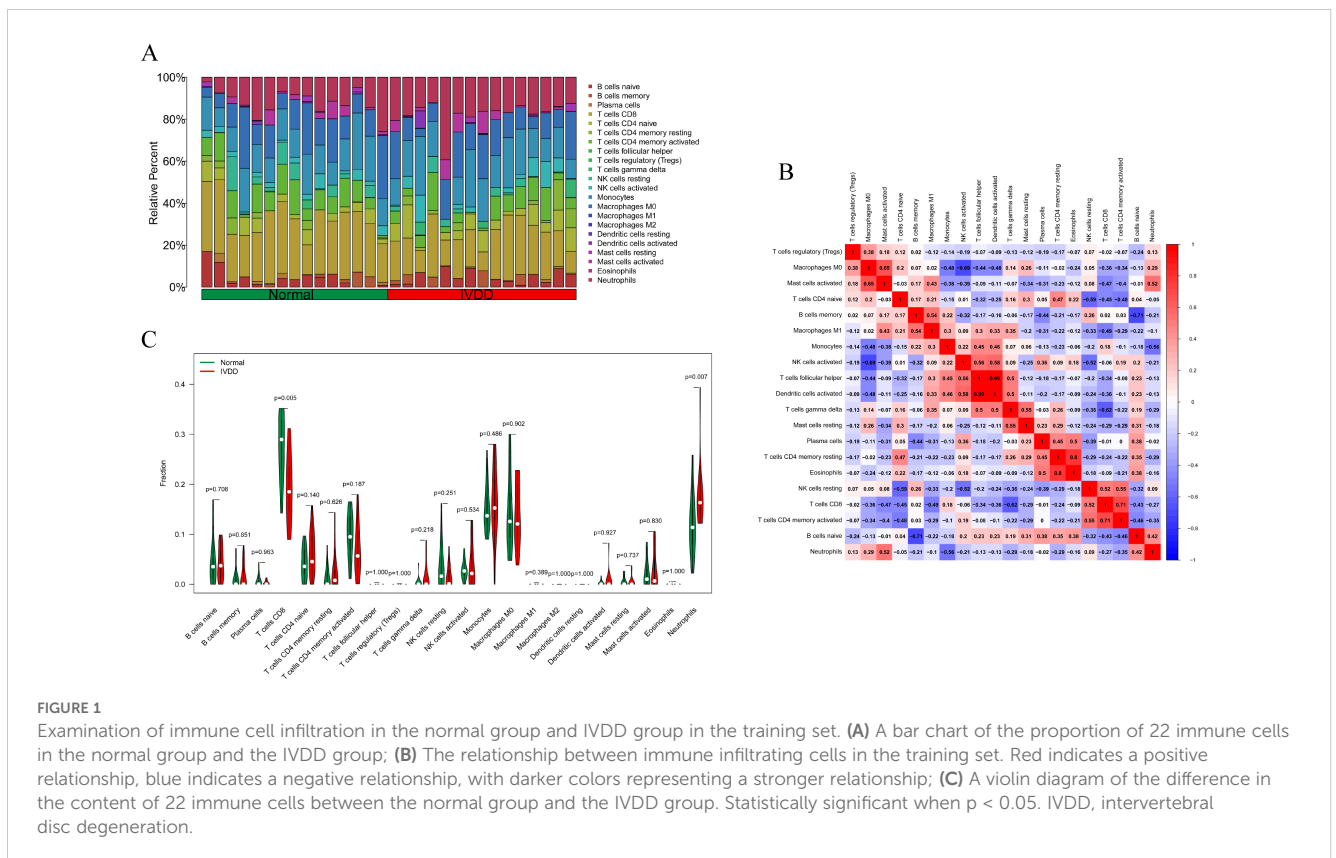
2.9 Statistical analysis

All data processing, plotting, and statistical analysis are performed in R software (4.4.0). Differences between independent variables and non-normally distributed variables were analyzed using the Wilcoxon test. The correlation between the two continuous variables was determined by Spearman correlation analysis. ROC curve analysis was performed using the “pROC” package in R software. $P < 0.05$ was considered statistically significant.

3 Results

3.1 The landscape of immune infiltration in IVDD

Based on the merged training set from the GSE124272, GSE147383, and GSE153761 datasets, the relative proportions of 22 immune cell subsets in the normal control group and the IVDD group were assessed using CIBERSORT technology (Figure 1A). Subsequently, further analysis was conducted to decode the interrelationships among these infiltrating immune cells (Figure 1B). The results revealed strong positive correlations between T cells follicular helper and dendritic cells activated ($r =$



0.99), T cells CD4 memory resting and eosinophils ($r = 0.8$), and T cells CD8 and T cells CD4 memory activated ($r = 0.71$). On the contrary, there were notable negative correlations between activated NK cells and M0 macrophages ($r = -0.69$), T cells gamma delta and T cells CD8 ($r = -0.62$), and B cells naïve and B cells memory ($r = -0.71$). As depicted in the violin plot (Figure 1C), CD8 T cell infiltration was significantly lower ($p = 0.005$) and neutrophil infiltration was significantly higher ($p = 0.007$) in the IVDD group compared to the normal control group, indicating that immune cells are involved in the progression of IVDD.

3.2 Identification of DEGs in the IVDD and normal groups

GSE124272 comprises 16 samples, GSE147383 includes 8 samples, and GSE153761 contains 6 samples. After merging these three datasets, the training set consists of 30 samples. Subsequently, the expression matrix data of the 30 samples were normalized (Figures 2A, B). A total of 517 differentially expressed genes (DEGs), including 306 upregulated genes and 211 downregulated genes, were identified between the normal and IVDD groups. The volcano plot depicting the expression of DEGs is shown in

Figure 2C, with red representing higher expression levels, green indicating lower expression levels, and all other genes depicted in gray. Figure 2D displays the heatmap of the top 50 genes with increased (red) and decreased (green) expression levels based on logFC between the normal and IVDD groups.

3.3 Identification and functional analysis of DELAGs

Using keywords such as “lymphatic” and “lymphatic vessel,” we retrieved 302 LAGs from GeneCards, MSigDB, Gene Ontology, and KEGG databases. Intersection with the DEGs in the training set identified 15 DELAGs (Figures 3A, B). To elucidate the biological functions of these DELAGs, we employed GO, KEGG, and DO analysis tools. GO enrichment analysis revealed that DELAGs are mainly enriched in processes such as morphogenesis and regulation of branching epithelium and structure, morphogenesis and development of glands, growth factor activity, transmembrane receptor protein tyrosine kinase activity, and exogenous protein binding (Figure 3C; Supplementary Table S1). The DELAGs also exhibited consistent trends with the Rap1, MAPK, Ras, and PI3K-Akt signaling pathways (Figure 3D; Supplementary Table S2).

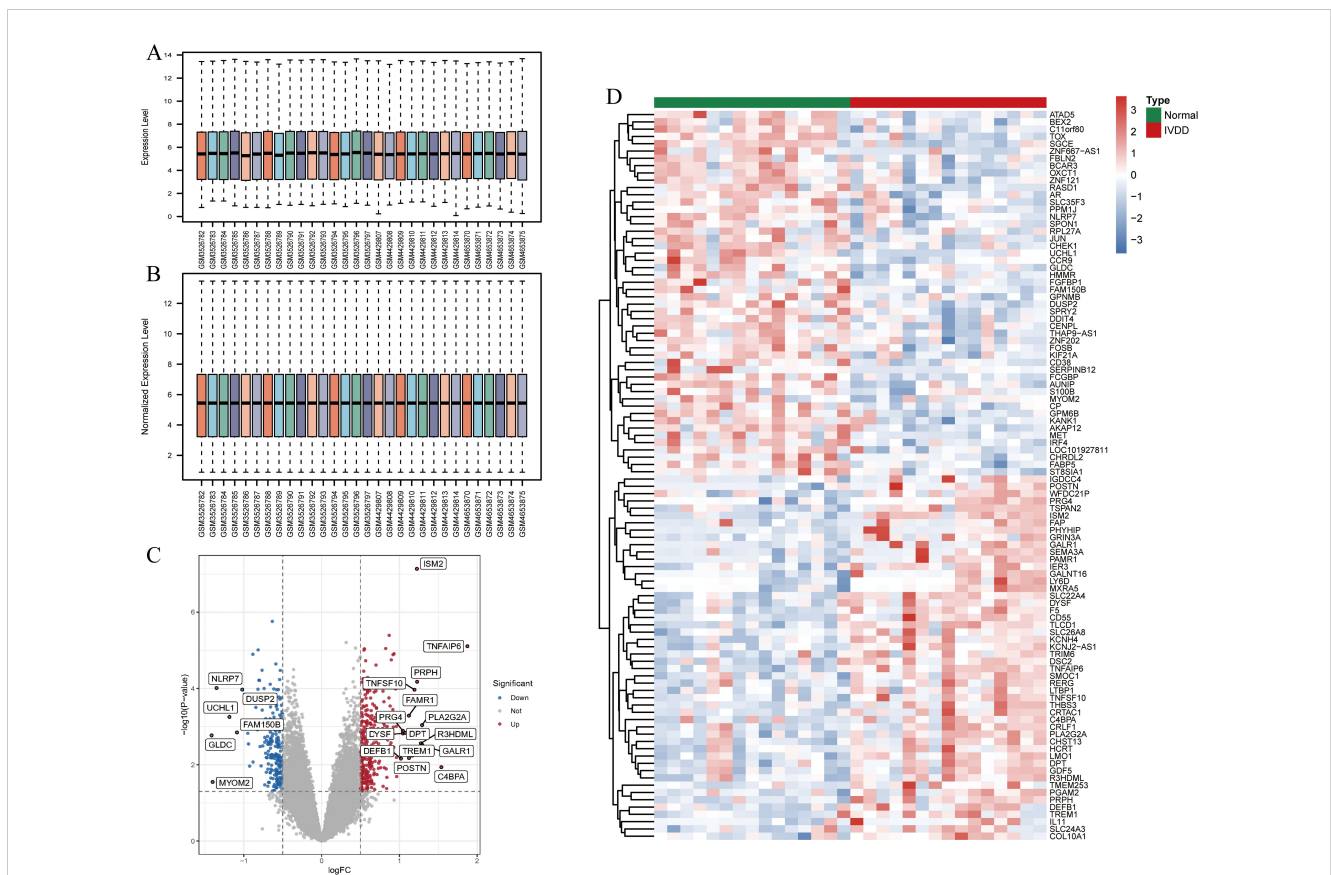
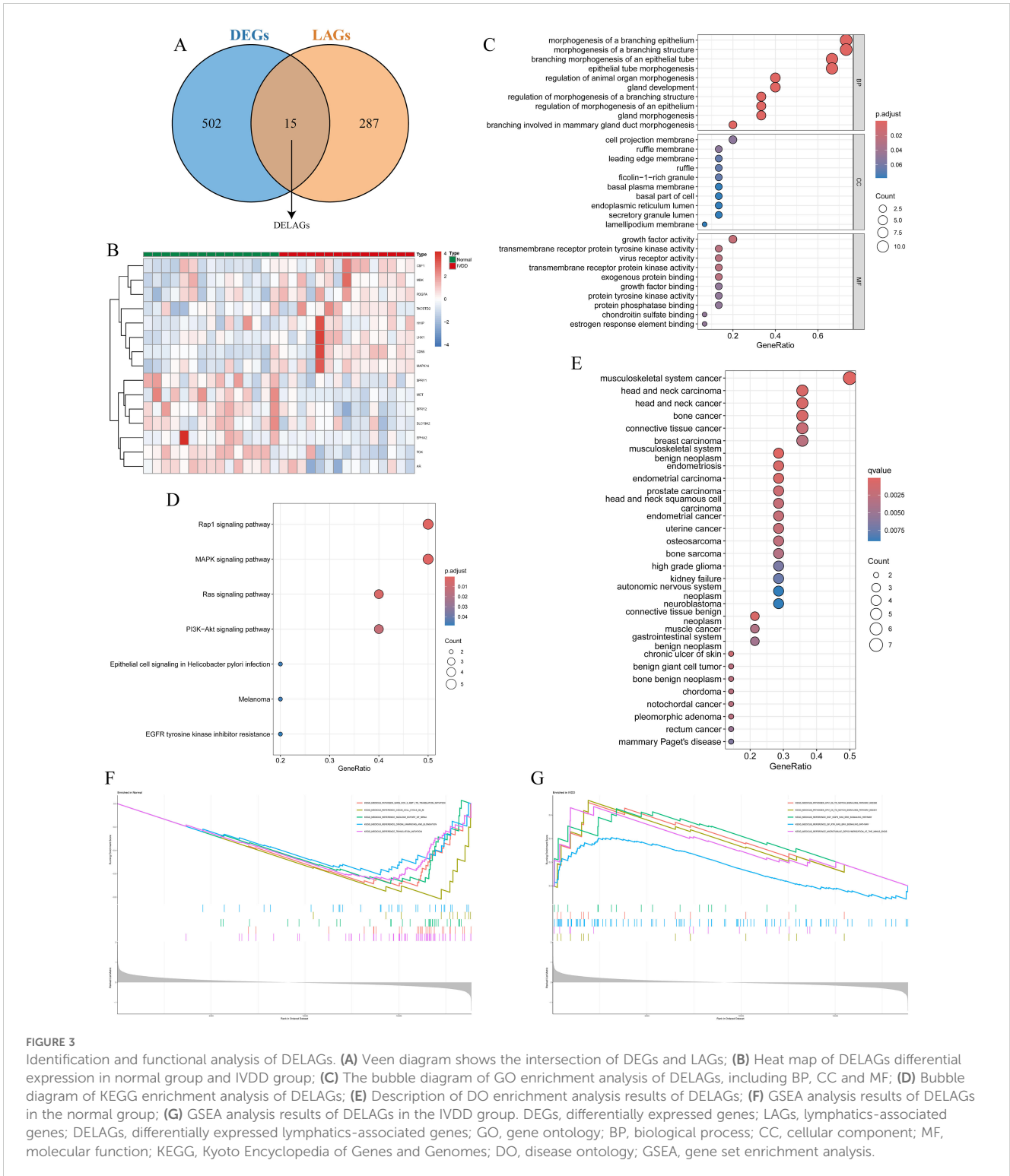


FIGURE 2 Identification and analysis of DEGs in the training set. (A) The bar chart of the expression matrix of 30 samples in the training set before normalization; (B) The bar chart of the expression matrix in 30 samples in the training set after normalization; (C) Volcano plot of DEGs expression. Red indicates genes with increased expression, grey is non-significant, and green indicates genes with decreased expression; (D) Heatmap of DEGs expression. Red is the high expression and green is the low expression. DEGs, differentially expressed genes; IVDD, intervertebral disc degeneration.



Further Disease Ontology analysis explored the associations of DELAGs with various diseases, including tumors of the musculoskeletal system and head and neck, as well as chronic ulcers of the skin (Figure 3E; Supplementary Table S3). GSEA analysis results revealed the enrichment of pathways such as Cell cycle and translation initiation in samples from the normal control group (Figure 3F), while the Notch and Ras/ERK signaling pathways were enriched in IVDD samples (Figure 3G).

3.4 Based on multiple machine learning algorithms to screen key DELAGs as model genes

We compare the four machine learning algorithms (XGB, RF, SVM, and GLM) by calculating the residual values (Figure 4A), the reverse residual distribution plots (Figure 4B), and the ROC curves (Figure 4C). The residual values are the smallest for XGB and the

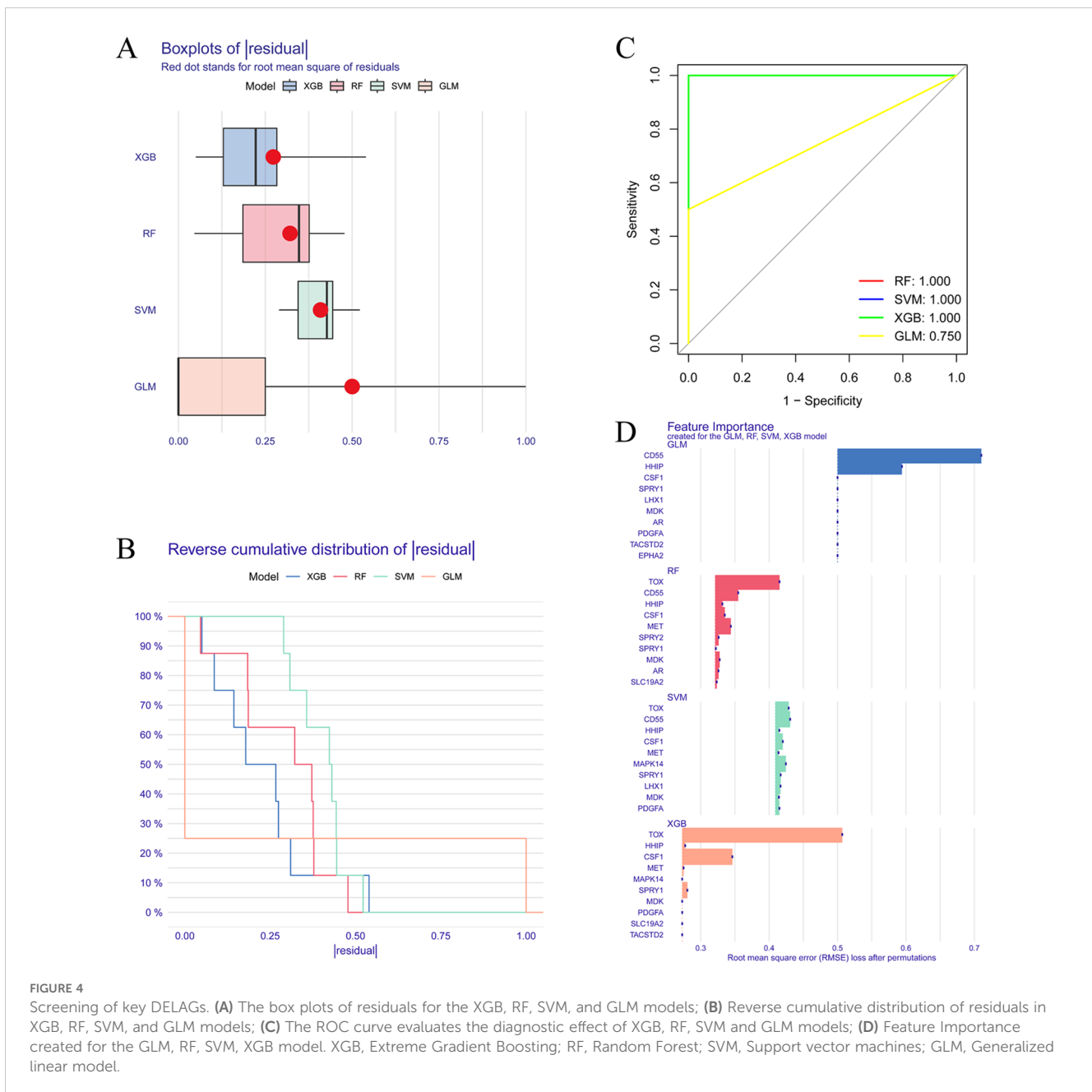


FIGURE 4 Screening of key DELAGs. (A) The box plots of residuals for the XGB, RF, SVM, and GLM models; (B) Reverse cumulative distribution of residuals in XGB, RF, SVM, and GLM models; (C) The ROC curve evaluates the diagnostic effect of XGB, RF, SVM and GLM models; (D) Feature Importance created for the GLM, RF, SVM, XGB model. XGB, Extreme Gradient Boosting; RF, Random Forest; SVM, Support vector machines; GLM, Generalized linear model.

largest for GLM. The Area under Curve (AUC) values for XGB, RF, SVM, and GLM are 1.00, 1.00, 1.00, and 0.75. At the same time, we also calculated the accuracy, standard deviation of accuracy, Kappa value, and standard deviation of Kappa value of the four models. XGB was 0.853, 0.219, 0.704, and 0.444; SVM was 0.863, 0.237, 0.716, 0.488; RF was 0.807, 0.298, 0.608, 0.6; GLM was 0.57, 0.388, 0.134, 0.768. By performing Friedman’s test and Nemenyi’s *post hoc* test on the accuracy and kappa values, we found that there is no significant difference in the performance of the three models, XGB, SVM, and RF ($p > 0.05$), whereas there is a significant difference between the GLM model and the three, and in particular, there is a significant difference between the XGB and the SVM with it ($p < 0.01$). Although XGB, SVM, and RF are not statistically different, when combined with specific values, the SVM model has the best accuracy and kappa

value, while the XGB model is the most stable. In addition, we plotted the feature importance of the four models (Figure 4D). Based on the results of the above-combined comparison of the four models, we identified the XGB model as the best model. Then, we identified the top five genes in the XGB model as the key DELAGs (MET, HHIP, SPRY1, CSF1, and TOX) and used them as the model genes to construct a diagnostic model for predicting IVDD.

3.5 Model gene-based diagnostic model construction and diagnostic performance evaluation

A nomogram model was built based on five model genes to investigate the risk of IVDD (Figure 5A). Evaluation of the

nomogram model using calibration curves showed good consistency between the five model genes and the ideal model (Figure 5B), indicating the nomogram's strong predictive value. The DCA results showed that the decision curve of the model was always higher than the None and ALL reference lines, indicating that the decision-making based on the nomogram may benefit IVDD patients, and the nomogram model has strong clinical utility (Figure 5C). Finally, separate ROC curves and an overall ROC curve were constructed using the five model genes in both the training and validation sets (Figures 5D–G). The results of the ROC curve analysis demonstrated that the constructed diagnostic model exhibited excellent predictive performance in the training set, as evidenced by an area under the curve (AUC) value of 0.973 (Figure 5E). Furthermore, the sensitivity, specificity, and accuracy of the model, as well as the positive predictive value (PPV) and negative predictive value (NPV), were all 0.933. Meanwhile, the Positive Likelihood Ratio (PLR) was as high as 14, while the Negative Likelihood Ratio (NLR) was 0.07. In the validation set, the AUC value of the diagnostic model was 0.772 (Figure 5G), and its sensitivity was 0.765, specificity 0.706, and accuracy 0.735. In addition, the PPV was 0.722, NPV 0.75, PLR 2.6, and NLR 0.33. Combining the above diagnostic performance metrics, the model demonstrated good accuracy and stability in diagnosing IVDD. Additionally, we visually represented the expression of the five model genes in the training set using boxplots (Figures 5H–L). The results revealed that MET, SPRY1, and TOX exhibited lower expression in the IVDD group compared to the normal control group, while CSF1 and HHIP showed higher expression in the IVDD group.

3.6 GSVA analysis of model genes and its correlation with immune cells

In the training set, GSVA of the five model genes (Figures 6A–J) revealed the following: GO function in the MET low expression group was mainly enriched in the binding of cell adhesion protein involved in communication between the bundle of HIS cells and Purkinje myocytes (Figure 6A), and the high expression group of the KEGG pathway was mainly enriched in metabolism of xenobiotics by cytochrome p450 pathway (Figure 6B). The GO function of the HHIP high expression group was mainly enriched in ISG15 protein conjugation (Figure 6C), and the KEGG pathway was mainly enriched in glycosphingolipid biosynthesis globo series and tryptophan metabolism (Figure 6D). The GO function in the low CSF1 expression group was mainly enriched in the negative regulation of endothelial cell chemotaxis (Figure 6G), and the KEGG pathway was mainly enriched in metabolism of xenobiotics by cytochrome p450 and Glycosaminoglycan biosynthesis keratan sulfate, whereas the KEGG pathway in the high-expression group was mainly enriched in Protein export and circadian rhythms in mammals (Figure 6H). Moreover, the correlation between the five model genes and 28 immune cells was analyzed (Figure 6K). The analysis revealed that CSF1 was positively correlated with natural killer cells, central memory CD8 T cells, and CD56 bright natural killer cells while showing negative

correlations with natural killer T cells and activated CD4 T cells. HHIP exhibited a significant negative correlation with CD56 bright natural killer cells. TOX had the strongest positive correlation with T follicular helper cells, followed by immature B cells. These findings suggest that these model genes are involved in the immune regulation processes during IVDD progression.

3.7 Construction of drug regulatory network and ceRNA network

Based on the five model genes, we identified potential therapeutic drugs using the gene-drug interaction data in DGIdb and demonstrated the interactions between the drugs and model genes using Cytoscape (Figure 7A). Information on the source, interaction type, and score of the target drugs is shown in Supplementary Table S4. According to the 'Interaction Score' in DGIdb, a total of 52 potential drugs were identified, 50 drugs were targeted at MET, and 2 drugs were targeted at CSF1. No potential drugs for HHIP, SPRY1, and TOX were found. Additionally, to uncover the potential post-transcriptional regulatory mechanisms of the five model genes, a lncRNA-miRNA-mRNA regulatory network was constructed. miRNAs targeting the five model genes were predicted using miRanda, miRDB, miRWalk, and TargetScan, selecting those with a score of 1 across all four databases. SpongeScan was then used to predict lncRNA-miRNA interactions, and the complete ceRNA network was constructed using Cytoscape (Figure 7B). Within each lncRNA-miRNA-mRNA regulatory axis, lncRNAs can enhance the expression levels of the model genes by inhibiting the corresponding miRNAs. For instance, LINC00173 may indirectly upregulate CSF1 expression by suppressing miR-939-5p, thereby enhancing its role in regulating the survival, proliferation, and differentiation of hematopoietic progenitor cells. Overall, this ceRNA network helps to elucidate the potential post-transcriptional regulatory mechanisms in IVDD and provides a reference for subsequent therapeutic strategies.

4 Discussion

IVDD is a primary cause of lower back pain, with its etiology rooted in various factors such as genetic predisposition (21), mechanical stress (22), trauma (23), smoking (24), obesity (25), and aging (26). The extrusion of nucleus pulposus tissue during IVDD, which has autoantigenic properties, can trigger an immune-inflammatory response (27, 28), compressing the nerves or spinal cord, leading to immune cell infiltration of the intervertebral tissues (29–31), and causing low back pain that affects the mobility of both lower limbs (32). The lymphatic system, a crucial component of the circulatory system, is essential for maintaining tissue fluid balance, lipid absorption, immune surveillance, and the transport of immune cells (33). Current research generally agrees that the process of IVDD involves changes in the distribution of lymphatic vessels. Specifically, during the degeneration of a normal IVD, lymphatic vessels are accompanied by blood vessels from the surrounding tissues into the degenerated IVD tissue. The

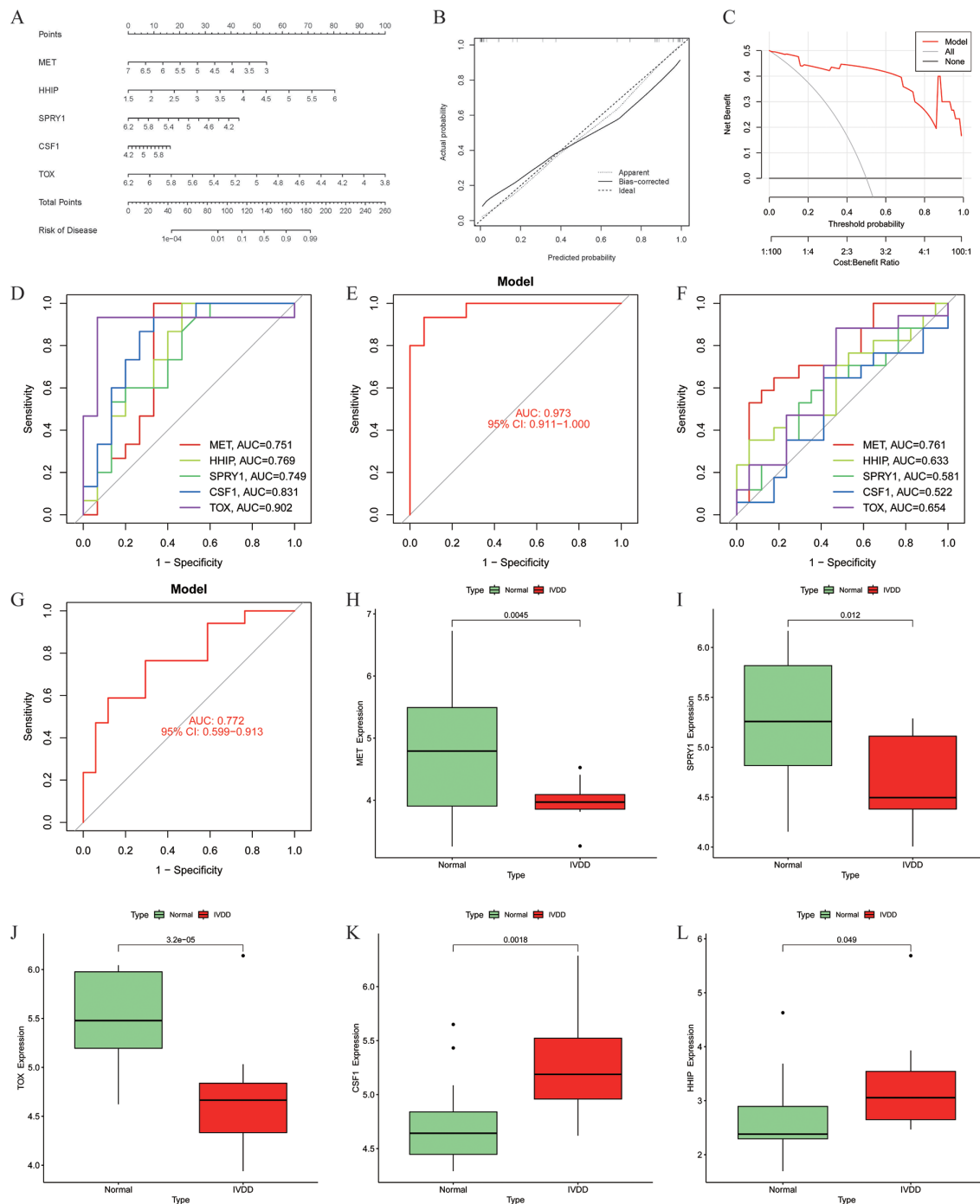


FIGURE 5

Construction and diagnostic value of the diagnostic model. (A) Model gene nomogram for the diagnosis of IVDD; (B) Calibration curve evaluation of the nomogram model; (C) DCA curves of the nomogram prediction; (D) ROC curves evaluating the diagnostic effect of five model genes in the training set; (E) The entire ROC curve for the five model genes in the training set; (F) ROC curves evaluating the diagnostic effect of five model genes in the validation set; (G) The entire ROC curve for the five model genes in the validation set; (H–L) Differential expression of model genes in the training set, MET ($p = 0.0045$), SPRY1 ($p = 0.012$) and TOX ($p = 3.2 \times 10^{-5}$) were lowly expressed in the IVDD group, and CSF1 ($p = 0.0018$) and HHIP ($p = 0.049$) were highly expressed in the IVDD group, $p < 0.05$ was statistically significant. DCA, Decision Curve Analysis; ROC, Receiver operating characteristic.

proteases, inflammatory cytokines, and growth factors present in the degenerated IVD matrix facilitate this process (9). Abnormal lymphatic vessel formation induced by IVDD may resemble that observed in lymphedema, obesity, and cancer, closely correlating with the occurrence and progression of the disease (34). Modulating

lymphatic vessel formation could potentially become a future therapeutic target for IVDD. However, the mechanistic association between lymphatic vessels and IVDD remains unconfirmed, with a scarcity of relevant research. In this study, a diagnostic signature related to lymphatic vessels was generated by

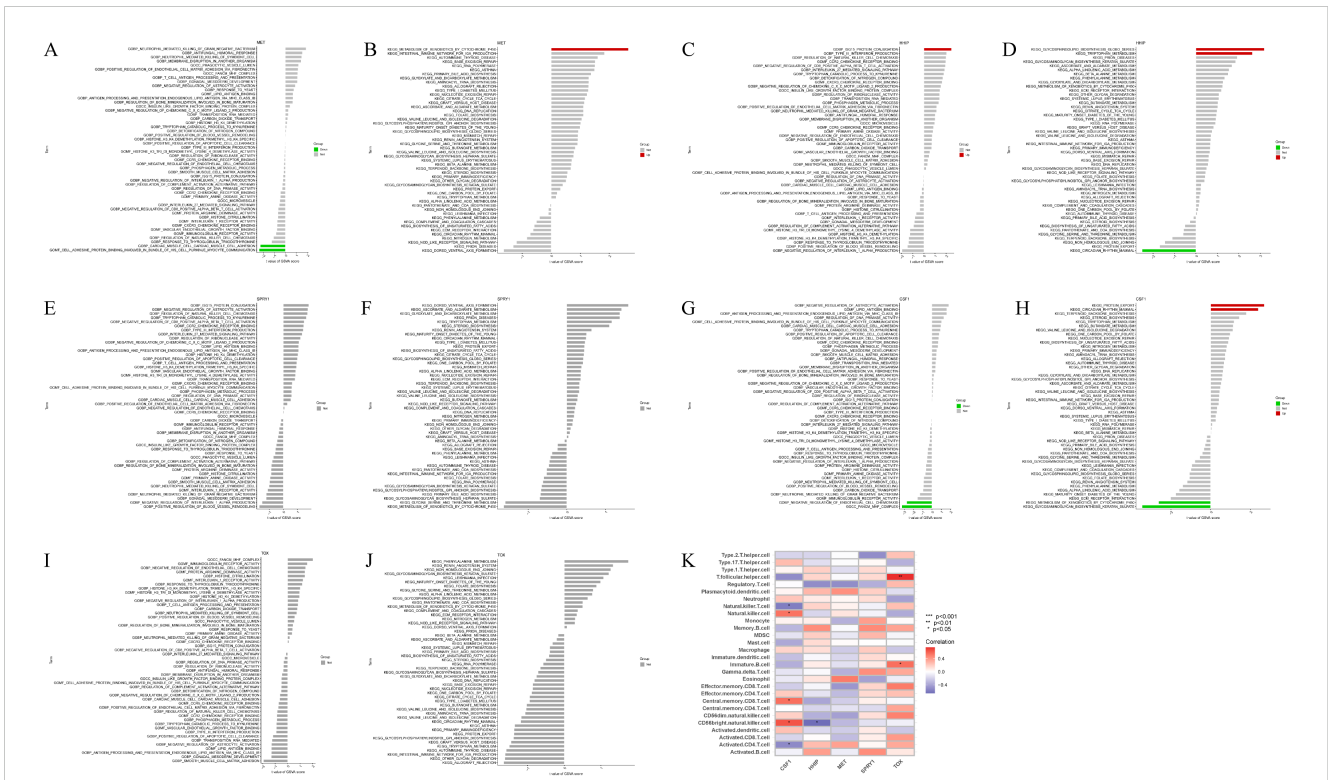


FIGURE 6 GSEA of model genes in the training set and their correlation with 28 immune cells. (A, B) GO and KEGG analysis in GSEA of MET; (C, D) GO and KEGG analysis in GSEA of HHIP; (E, F) GO and KEGG analysis in GSEA of SPRY1; (G, H) GO and KEGG analysis in GSEA of CSF1; (I, J) GO and KEGG analysis in GSEA of TOX; (K) Correlation between five model genes and 28 immune cells, red represents positive correlation, blue represents negative correlation, the darker the color the stronger the correlation. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

integrating gene expression profiles and LAGs using machine learning algorithms. The performance of the diagnostic model was evaluated using both training and external validation sets. In addition, GSEA was used to explore the potential mechanisms of the model genes and their correlation with immune cell infiltration and to predict potential drugs and ceRNA networks targeting the model genes. Our results may offer potential applications for IVDD patients, such as early diagnosis and the selection of new therapeutic targets.

In this study, three IVDD-related datasets were initially acquired from the GEO database and merged to form the training set, with GSE150408 serving as the validation set. Bioinformatic techniques were then used to examine the immune cell characteristics within the training set, revealing significant statistical discrepancies in CD8+ T cells and neutrophils. IVDD samples showed a significant increase in neutrophils and a marked decrease in CD8+ T cells compared to the control group. This is consistent with the findings of Kaneyama et al. (35) that nucleus pulposus (NP) tissue extruded from degenerated IVD induces an immune response, leading to infiltration of immune cells that disrupt the physiological barrier of IVD cells, resulting in NP cell apoptosis. We then screened DEGs and LAGs from the training set and four databases (GeneCards, MSigDB, Gene Ontology, KEGG), and obtained 15 DELAGs in the training set after taking the intersection.

Subsequently, functional enrichment analysis was performed on DELAGs. According to GO analysis, DELAGs were enriched in both BP and MF. BP mainly includes branching epithelial cell and structure morphogenesis, regulation of animal organ morphogenesis, branching structure, and epithelial cell morphogenesis. MF mainly includes growth factor activity, transmembrane receptor protein tyrosine kinase activity, virus receptor activity, transmembrane receptor protein kinase activity, exogenous protein binding, growth factor binding, protein tyrosine kinase activity, and protein phosphatase binding. KEGG enrichment analysis revealed that DELAGs primarily influenced the Rap1 signaling pathway, MAPK signaling pathway, Ras signaling pathway, and PI3K-Akt signaling pathway in the IVDD group. The Ras signaling pathway is closely linked to inflammatory responses (36). Both the MAPK signaling pathway (23) and PI3K-Akt signaling pathway have been extensively linked to the occurrence and progression of IVDD. PI3K-Akt signaling pathway is involved in IVD cell proliferation, senescence, and apoptosis, and activation of this pathway can delay the progression of IVDD by upregulating SOX9 expression (37). Furthermore, studies indicate that cyclic mechanical stretching can ameliorate NP cell degeneration via the PI3K-Akt signaling pathway (38–40). GSEA unveiled that the genes within the IVDD group are predominantly associated with the Notch and Ras/ERK signaling pathways. Notably, a study delineated that the activation effects of Notch signaling exert cell type-specific influences on genes regulating IVD matrix synthesis and degradation metabolism (41).

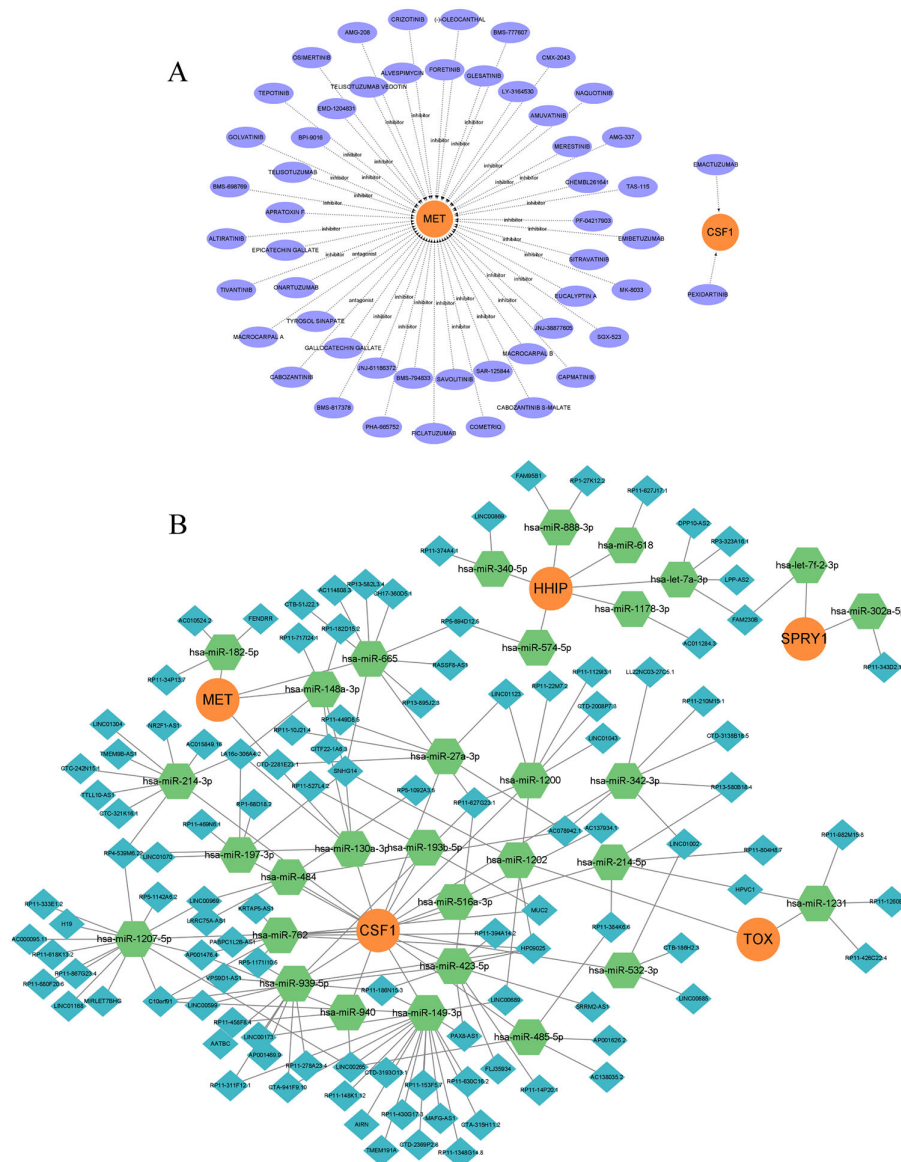


FIGURE 7
Drug regulatory network and ceRNA network. (A) Prediction of drug–gene interactions for model genes, orange represents model genes, purple represents predicted drugs; (B) ceRNA network, orange circle represents model gene, green hexagon represents miRNA, and blue diamond represents lncRNA.

Specifically, the activation of Notch signaling can promote the expression of matrix degradation genes in anulus fibrosus (AF) and ATDC5 cells and inhibit the expression of matrix anabolism genes. In NP cells, this effect inhibited the expression of matrix degradation genes (including MMP3, MMP13, Adamts4, and Adamts5), and attenuated the expression of MMP13 induced by TNF- α and macrophages. In addition, it has been demonstrated that the expression of Notch signaling molecules and their downstream target genes can be detected in AF and NP cells in IVD tissues (42), with a significant increase in Notch signaling activity noted in degenerated IVD tissue compared to healthy human IVD (43).

In comparison to conventional statistical models, machine learning can provide better predictive performance and capture the complex interactions among predictors and the non-linear

relationship between predictors and outcomes (44). Based on this, we employed four machine-learning algorithms and identified the top 5 genes (MET, HHIP, SPRY1, CSF1, and TOX) in the XGB model as model genes through comparative analysis. To validate the reliability of these model genes, we constructed a diagnostic model for predicting IVDD utilizing these five model genes. It is necessary to note that evaluation of model performance and external validation are imperative for the constructed model (45). The dataset utilized for external validation should also differ from the training set in terms of both temporality and geography, as external validation constitutes the model’s redevelopment process on a validation set (46, 47). Calibration and discrimination are some of the most basic and important indicators for model performance (48). Calibration refers to the degree of consistency between the

predicted risk of predictors and the observed results, which can be evaluated by calibration curves (49). Within this study, the calibration curve of the diagnostic model closely matches the observed results curve, indicating that the diagnostic model we constructed has a high predictive ability for IVDD. Concurrently, as a calibration supplement, we introduced the DCA, a method to determine whether the information provided by the prediction model for decision-making in clinical practice is more beneficial than harmful (50). In this study, the DCA of the model gene in the training set consistently favors patients, demonstrating the model gene's stable diagnostic value for IVDD. Moreover, the model must also possess the discriminatory ability to differentiate between true positives and true negatives, typically assessed using the C-index, which is akin to the AUC in ROC; higher values indicate a stronger ability of the model to distinguish between true positives and true negatives (51). The C-indexes of the model genes in the training and validation sets were 0.973 and 0.772, respectively, and the accuracy was 0.933 and 0.735, respectively. These findings suggest that the diagnostic model constructed based on the model genes has a high degree of confidence in distinguishing between the normal group and the IVDD group. Notably, the overall efficacy of the model gene surpasses that of the five individual model genes, which is consistent with the multi-molecular driving properties of IVDD. Moreover, the expression levels of these five model genes in the original data display significant disparities between the normal and IVDD groups, with MET, SPRY1, and TOX demonstrating lower expression in IVDD samples compared to the normal group, potentially acting as protective genes for IVDD, while CSF1 and HHIP exhibit increased expression in IVDD samples, which could be used as the risk gene of IVDD. Thymocyte Selection Associated High Mobility Group Box (TOX) is a nuclear DNA-binding protein that plays an important role in the development of CD4+ T cells, NK cells, and intrinsic lymphocytes and is a key regulator of Exhausted T cell development (52). T-cell exhaustion is a pervasive phenomenon that serves to prevent immune overactivation and establish immune homeostasis in response to chronic inflammatory stimuli while limiting T-cell-mediated immunopathology (53). In this process, TOX induces CD4+ T cells to produce IL-10, thereby regulating inflammation (54). The exogenous addition of IL-10 has been demonstrated to alleviate IL-1 β -induced degeneration of NP cells. Furthermore, IL-10 treatment has been shown to significantly suppress mRNA expression of type X collagen, as well as degradation of aggrecan and type II collagen, through the inhibition of the p38/MAPK pathway. In addition, IL-10 up-regulates mRNA expression of SOX9, exerting a protective effect against IVD (55). Hedgehog interacting protein (HHIP) is a type I transmembrane glycoprotein that plays a role in a number of biological processes, including development and angiogenesis (56, 57). Cross-sectional studies have found that circulating HHIP levels are significantly elevated in overweight/obese women and positively correlate with blood glucose, insulin, and body mass index, while HHIP levels are regulated by blood glucose and insulin levels. Obesity is currently acknowledged as an independent risk factor for the onset of Intervertebral Disc Disease (IVDD) and is significantly correlated with elevated levels of IL-6 and systemic

pro-inflammatory cascades. Therefore, increased circulating levels of HHIP may be linked to the activation of inflammatory pathways associated with IVDD degeneration. Intervertebral disc degeneration (IVDD) is a multifactorial disease. Genetic susceptibility, mechanical stress, obesity, and smoking have all been identified as risk factors for IVDD. While there is limited research directly examining the association between the aforementioned model genes and IVDD, further investigation into the pathogenesis of IVDD and a more detailed analysis of the functions of the five model genes may gradually elucidate the interactions between them.

To further explore the biological functions of the model gene, we conducted Gene Set Variation Analysis (GSVA) and immune cell correlation analysis. The results revealed that CSF1 is primarily enriched in the biosynthesis process of glycosaminoglycans (GAG). It has been established that GAG plays a pivotal role in the human IVD. During IVDD, proteoglycans containing GAG are degraded and depleted, resulting in the loss of extracellular matrix integrity (58), and abnormal changes in GAG chain replacement during degeneration exacerbate this process (59). Hence, to maintain normal IVD function, high concentrations, and highly charged proteoglycans must be sustained, enabling functional proteoglycans to possess as many sulfated GAG chains as possible. However, enzymes involved in GAG biosynthesis (such as XT-1 and GlcAT-I) are downregulated with age and degeneration, resulting in weakened proteoglycan function and abnormal elongation of GAG chains (60, 61). Thus, we hypothesize that CSF1 mediates pathological changes in IVD by regulating GAG biosynthesis. Research has shown that during IVDD, immune cells, and their inflammatory factors can infiltrate the IVD through defects in the cartilaginous endplate and AF, accelerating catabolism and inducing inflammatory responses (62). Therefore, immunotherapy targeting immune cells is viewed as a novel strategy to alleviate IVDD (63). Meanwhile, genes associated with immune cells may also serve as biomarkers for IVDD (64). Based on this, we analyzed the inter-association between model genes and 28 immune cells. The findings revealed significant correlations between TOX, CSF1, and HHIP and immune cells, mainly focusing on Natural killer cells, CD8/CD4 T cells, and Immature B cells. It has been reported that activated natural killer cells are involved in the pathological process of IVDD (65). Further studies revealed that natural killer cells mainly have cytotoxic effects on NP cells (66). During IVDD, the protruding IVD tissue initiates an immune response, with the adaptive immune response being primarily manifested by the activation of T cell and B cell subsets (67). CD4 T cell subsets are involved in the regulation of inflammatory responses and offer potential targets for immunotherapy in IVDD (68). Upon activation, B cells produce antibodies, which are involved in the immune regulation and inflammatory response of IVDD (67). However, aberrant T-cell differentiation can result in the overexpression of inflammatory cytokines and the aberrant expression of B cells, both of which are critically linked to IVDD (67, 69).

Furthermore, we have identified potential therapeutic candidates for IVDD. DGIdb employs a data-mining approach to

identify potential therapeutic targets or priority drug development based on specific genes (70). A total of 52 potential drugs for the treatment of IVDD were identified based on DGIdb. The majority of these drugs target the MET gene, followed by the CSF1 gene. Mesenchymal to epithelial transition factor (MET), also known as Cellular-mesenchymal to epithelial transition factor (c-Met) or Hepatocyte growth factor receptor (HGFR), is a member of the Protein tyrosine kinases family (71). c-MET is a specific receptor for hepatocyte growth factor (HGF). It has been reported that c-MET is expressed in NP cells, and activation of signaling by HGF binding to c-MET promotes cell proliferation and exerts anti-inflammatory effects (72). During IVDD, inflammatory stimuli enhance c-MET expression, while HGF treatment leads to decreased c-MET expression. Meanwhile, HGF treatment enhances hypoxia-inducible factor-1 α (HIF-1 α) expression to promote NP cell proliferation (73). In addition, activation of HGF/c-MET signaling inhibits the elevation of cyclooxygenase-2, MMP-3, and MMP-9 in NP cells after TNF- α stimulation (72). It has been demonstrated that apoptosis and ECM degradation play key roles in the progression of IVDD (74, 75). Thus, it is evident that HGF targeting c-MET has a protective effect on NP cells and can delay the process of IVDD by promoting cell proliferation and inhibiting the degradation of ECM. Colony-stimulating factor-1 (CSF1), also known as macrophage colony-stimulating factor (M-CSF), controls macrophage production, differentiation, and function (76, 77). Patients with IVDD are infiltrated with a large number of inflammatory cells, of which macrophages are the main inflammatory cells capable of infiltrating into the NP, and the number of macrophages is positively correlated with the grade of IVDD (66, 78). Macrophages that migrate to degenerating intervertebral discs (IVDs) differentiate into distinct cellular phenotypes (M1 or M2 type) in response to the local microenvironment. This differentiation process is regulated by a number of factors. It was demonstrated that p38 activation in NP cells could induce macrophage differentiation towards the M1 phenotype by modulating the local pro-inflammatory microenvironment. Conversely, blocking p38 activation in NP cells could inhibit M1 phenotypic differentiation and reduce CSF1 and IFN- γ secretion by NP cells, while simultaneously neutralizing CSF1 and IFN- γ -induced macrophage differentiation towards the M1 phenotype (79). In addition, CSF1 is also associated with microglial activation and proliferation. When NP is exposed to the dorsal root ganglia of the spinal cord, there is an upregulation of CSF1 expression in the dorsal root ganglia of the spinal cord and CSF1 receptor (CSF1R) in spinal cord microglia, which are activated at this time, and this behavior contributes to the pathogenesis of discogenic back pain through central sensitization (76). In summary, MET and CSF1 may be potential therapeutic targets for IVDD, but their specific mechanisms in IVDD development and progression require further experimental elucidation and assessment. Additionally, the drugs targeting these two genes based on theoretical prediction also need to be verified by more in-depth animal experiments and clinical trials after elucidating the mechanism of action of the two genes.

Recently, non-coding RNAs have been identified as pivotal regulators of gene expression (80). According to reports, non-coding RNAs can modulate the activation of autophagy in immune cells, thereby participating in the mediation of IVDD by directly or indirectly targeting autophagy-related genes and associated signaling pathways (81–83). Furthermore, lncRNA, miRNA, and mRNA can form ceRNA networks to regulate the occurrence and progression of IVDD (84, 85). We constructed potential ceRNA regulatory networks targeting these five model genes, providing new insights into the regulatory mechanisms and targeted therapies for IVDD. Within the constructed ceRNA network, multiple nodes have been reported to participate in the regulation of intervertebral disc degeneration (IVDD). LINC00969 is highly expressed in NP tissues and cells of IVDD patients, promoting IVD degeneration by sponging miR-335-3p and modulating NLRP3 inflammasome activity (86). In NP cells, overexpression of LINC00689 mediates autophagy via the miR-3127-5p/ATG7 axis, facilitating NP cell proliferation and suppressing apoptosis (87). NR2F1-AS1 is upregulated in IVD tissues of IVDD patients or NP cells treated with IL-1 β or TNF- α , exacerbating IL-1 β -induced extracellular matrix degradation and NP cell apoptosis, and regulates IVDD progression through miR-145-5p-mediated FOXO1 signaling pathway (88).

Although our results show accurate diagnostic efficacy of the model genes and have been validated in external datasets, there remain some limitations that require clarification. First, the sample size of gene expression profiles obtained from public databases is slightly insufficient, and individual differences between samples may affect the generalizability of the results. Furthermore, although we used cross-validation in model training, the relatively small sample size may still lead to overfitting of the model, thus affecting the robustness and generalizability of the results. Therefore, we will increase the sample size in future studies to avoid these problems and thus improve the accuracy of model prediction. Secondly, the lack of detailed clinical characteristics in the acquired data, such as age, degree of IVDD graded by Pfirrmann, duration of the disease, affected IVD level, and IVDD type (spinal stenosis, degenerative spondylolisthesis, intervertebral disc disease), which limits us to further reveal the potential association between model genes and certain traits. In addition, the targeted drugs and ceRNA regulatory networks derived from the model genes are at the analysis and hypothesis stage, requiring further validation through *in vitro* and *in vivo* experiments. Future studies will involve the design of prospective research to gather more comprehensive and multidimensional data for the verification of our findings. Meanwhile, various molecular biology techniques will be employed to conduct *in vitro* and *in vivo* experiments, aiming to fully understand the role of model genes and their potential regulatory mechanisms in IVDD.

5 Conclusions

In this research, four machine-learning methods were used to screen five model genes associated with lymphatic vessels, namely

MET, HHIP, SPRY1, CSF1, and TOX. A diagnostic model with high predictive value was constructed based on these genes, effectively identifying patients with IVDD. Furthermore, potential therapeutic drugs for these model genes were predicted, and a lncRNA-miRNA-mRNA regulatory network was developed. Through comprehensive analysis, the potential association between lymphatic vessels and the occurrence and progression of IVDD was explored. Further research is urgently needed to validate and elucidate the regulatory mechanisms between them. Ultimately, our research provides novel insights into the pathogenesis of IVDD and contributes to the discovery of new therapeutic targets.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/[Supplementary Material](#).

Author contributions

ML: Conceptualization, Formal analysis, Methodology, Visualization, Writing – original draft, Writing – review & editing. SL: Data curation, Formal analysis, Investigation, Visualization, Writing – original draft. YW: Formal analysis, Investigation, Methodology, Visualization, Writing – original draft. GZ: Formal analysis, Investigation, Methodology, Visualization, Writing – original draft. FH: Formal analysis, Investigation, Methodology, Visualization, Writing – original draft. QZ: Formal analysis, Investigation, Methodology, Visualization, Writing – original draft. PS: Conceptualization, Data curation, Funding acquisition, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing. HZ: Conceptualization, Funding acquisition, Investigation, Methodology, Resources, Supervision, Writing – original draft, Writing – review & editing.

References

1. GBD 2019 Diseases and Injuries Collaborators. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet*. (2020) 396:1204–22. doi: 10.1016/S0140-6736(20)30925-9
2. Peng BG. Pathophysiology, diagnosis, and treatment of discogenic low back pain. *World J Orthop*. (2013) 4:42–52. doi: 10.5312/wjo.v4.i2.42
3. Oliver G. Lymphatic vasculature development. *Nat Rev Immunol*. (2004) 4:35–45. doi: 10.1038/nri1258
4. Randolph GJ, Ivanov S, Zinselmeyer BH, Scallan JP. The lymphatic system: integral roles in immunity. *Annu Rev Immunol*. (2017) 35:31–52. doi: 10.1146/annurev-immunol-041015-055354
5. Padera TP, Meijer EF, Munn LL. The lymphatic system in disease processes and cancer progression. *Annu Rev BioMed Eng*. (2016) 18:125–58. doi: 10.1146/annurev-bioeng-112315-031200
6. Hou T, Liu Y, Wang X, et al. Ginsenoside Rg1 promotes lymphatic drainage and improves chronic inflammatory arthritis. *J Musculoskelet Neuronal Interact*. (2020) 20:526–34.
7. Chang JE, Turley SJ. Stromal infrastructure of the lymph node and coordination of immunity. *Trends Immunol*. (2015) 36:30–9. doi: 10.1016/j.it.2014.11.003
8. Kliskey K, Williams K, Yu J, Jackson D, Urban J, Athanasou N. The presence and absence of lymphatic vessels in the adult human intervertebral disc: relation to disc pathology. *Skeletal Radiol*. (2009) 38:1169–73. doi: 10.1007/s00256-009-0770-2
9. Kashima TG, Dongre A, Athanasou NA. Lymphatic involvement in vertebral and disc pathology. *Spine (Phila Pa 1976)*. (2011) 36:899–904. doi: 10.1097/BRS.0b013e3182050284
10. Kang JD, Georgescu HI, Mcintyre-Larkin L, Stefanovic-Racic M, Donaldson WF 3rd, Evans CH. Herniated lumbar intervertebral discs spontaneously produce matrix metalloproteinases, nitric oxide, interleukin-6, and prostaglandin E2. *Spine (Phila Pa 1976)*. (1996) 21:271–7. doi: 10.1097/00007632-199602010-00003
11. Le Maitre CL, Freemont AJ, Hoyland JA. The role of interleukin-1 in the pathogenesis of human intervertebral disc degeneration. *Arthritis Res Ther*. (2005) 7:R732–745. doi: 10.1186/ar1732
12. Özcan A, Collado-Diaz V, Egholm C, Tomura M, Gunzer M, Halin C, et al. CCR7-guided neutrophil redirection to skin-draining lymph nodes regulates cutaneous

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the Natural Science Foundation of Gansu Province (No. 23JRR0994).

Acknowledgments

We would like to acknowledge the GEO (GSE153761, GSE147383, GSE124272, and GSE150408) network for providing data.

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fimmu.2024.1441028/full#supplementary-material>

- inflammation and infection. *Sci Immunol.* (2022) 7:eabi9126. doi: 10.1126/sciimmunol.abi9126
13. Cao S, Liu H, Fan J, Yang K, Yang B, Wang J, et al. An oxidative stress-related gene pair (CCNB1/PKD1), competitive endogenous RNAs, and immune-infiltration patterns potentially regulate intervertebral disc degeneration development. *Front Immunol.* (2021) 12:765382. doi: 10.3389/fimmu.2021.765382
 14. Ye L, Mei R, Liu WT, Ren H, Zhang XX. Machine learning-aided analyses of thousands of draft genomes reveal specific features of activated sludge processes. *Microbiome.* (2020) 8:16. doi: 10.1186/s40168-020-0794-3
 15. Toh TS, Dondelinger F, Wang D. Looking beyond the hype: Applied AI and machine learning in translational medicine. *EBioMedicine.* (2019) 47:607–15. doi: 10.1016/j.ebiom.2019.08.027
 16. Huang L, Wang L, Hu X, Chen S, Tao Y, Su H, et al. Machine learning of serum metabolic patterns encodes early-stage lung adenocarcinoma. *Nat Commun.* (2020) 11:3556. doi: 10.1038/s41467-020-17347-6
 17. Shen B, Yi X, Sun Y, Bi X, Du J, Zhang C, et al. Proteomic and metabolomic characterization of COVID-19 patient sera. *Cell.* (2020) 182:59–72.e15. doi: 10.1016/j.cell.2020.05.032
 18. Xia J, Xie Z, Niu G, Lu Z, Wang Z, Xing Y, et al. Single-cell landscape and clinical outcomes of infiltrating B cells in colorectal cancer. *Immunology.* (2023) 168:135–51. doi: 10.1111/imm.v168.1
 19. Su W, Wei Y, Huang B, Ji J. Identification of hub genes and immune infiltration in psoriasis by bioinformatics method. *Front Genet.* (2021) 12:606065. doi: 10.3389/fgene.2021.606065
 20. Furió-Tarí P, Tarazona S, Gabaldón T, Enright AJ, Conesa A. spongeScan: A web for detecting microRNA binding elements in lncRNA sequences. *Nucleic Acids Res.* (2016) 44:W176–180. doi: 10.1093/nar/gkw443
 21. Wu F, Huang X, Zhang Z, Shao Z. A meta-analysis assessing the association between COL11A1 and GDF5 genetic variants and intervertebral disc degeneration susceptibility. *Spine (Phila Pa 1976).* (2020) 45:E616–e623. doi: 10.1097/BRS.00000000000003371
 22. Zhao R, Liu W, Xia T, Yang L. Disordered mechanical stress and tissue engineering therapies in intervertebral disc degeneration. *Polymers (Basel).* (2019) 11(7):1151. doi: 10.3390/polym11071151
 23. Zhang HJ, Liao HY, Bai DY, Wang ZQ, Xie XW. MAPK /ERK signaling pathway: A potential target for the treatment of intervertebral disc degeneration. *BioMed Pharmacother.* (2021) 143:112170. doi: 10.1016/j.biopha.2021.112170
 24. Kiraz M, Demir E. Relationship of lumbar disc degeneration with hemoglobin value and smoking. *Neurochirurgie.* (2020) 66:373–7. doi: 10.1016/j.neuchi.2020.06.133
 25. Cannata F, Vadalà G, Ambrosio L, Fallucca S, Napoli N, Papalia R, et al. Intervertebral disc degeneration: A focus on obesity and type 2 diabetes. *Diabetes Metab Res Rev.* (2020) 36:e3224. doi: 10.1002/dmrr.v36.1
 26. Francisco V, Pino J, González-Gay M, Lago F, Karppinen J, Tervonen O, et al. A new immunometabolic perspective of intervertebral disc degeneration. *Nat Rev Rheumatol.* (2022) 18:47–60. doi: 10.1038/s41584-021-00713-z
 27. Hiyama A, Mochida J, Iwashina T, Omi H, Watanabe T, Serigano K, et al. Transplantation of mesenchymal stem cells in a canine disc degeneration model. *J Orthop Res.* (2008) 26:589–600. doi: 10.1002/jor.20584
 28. Erario M, Croce E, Moviglia Brandolino MT, Moviglia G, Grangeat AM. Ozone as Modulator of resorption and inflammatory response in extruded nucleus pulposus herniation. Revising concepts. *Int J Mol Sci.* (2021) 22(18):9946. doi: 10.3390/ijms22189946
 29. Habtemariam A, Virri J, Grönblad M, Holm S, Kaigle A, Karaharju E. Inflammatory cells in full-thickness annulus injury in pigs. An experimental disc herniation animal model. *Spine (Phila Pa 1976).* (1998) 23:524–9. doi: 10.1097/00007632-199803010-00002
 30. Virri J, Grönblad M, Seitsalo S, Habtemariam A, Kääpä E, Karaharju E. Comparison of the prevalence of inflammatory cells in subtypes of disc herniations and associations with straight leg raising. *Spine (Phila Pa 1976).* (2001) 26:2311–5. doi: 10.1097/00007632-200111010-00004
 31. Geiss A, Sobottke R, Delank KS, Eysel P. Plasmacytoid dendritic cells and memory T cells infiltrate true sequestrations stronger than subligamentous sequestrations: evidence from flow cytometric analysis of disc infiltrates. *Eur Spine J.* (2016) 25:1417–27. doi: 10.1007/s00586-015-4325-z
 32. Amin RM, Andrade NS, Neuman BJ. Lumbar disc herniation. *Curr Rev Musculoskelet Med.* (2017) 10:507–16. doi: 10.1007/s12178-017-9441-4
 33. Tammela T, Alitalo K. Lymphangiogenesis: Molecular mechanisms and future promise. *Cell.* (2010) 140:460–76. doi: 10.1016/j.cell.2010.01.045
 34. Oliver G, Kipnis J, Randolph GJ, Harvey NL. The lymphatic vasculature in the 21 (st) century: novel functional roles in homeostasis and disease. *Cell.* (2020) 182:270–96. doi: 10.1016/j.cell.2020.06.039
 35. Kaneyama S, Nishida K, Takada T, Suzuki T, Shimomura T, Maeno K, et al. Fas ligand expression on human nucleus pulposus cells decreases with disc degeneration processes. *J Orthop Sci.* (2008) 13:130–5. doi: 10.1007/s00776-007-1204-4
 36. Deng Y, Gao X, Feng T, Wang Z, Xiao W, Xiong Z, et al. Systematically characterized mechanism of treatment for lumbar disc herniation based on Yaobitong capsule ingredient analysis in rat plasma and its network pharmacology strategy by UPLC-MS/MS. *J Ethnopharmacol.* (2020) 260:113097. doi: 10.1016/j.jep.2020.113097
 37. Li J, Wang C, Xue L, Zhang F, Liu J. Melatonin Suppresses Apoptosis of Nucleus Pulposus Cells through Inhibiting Autophagy via the PI3K/Akt Pathway in a High-Glucose Culture. *BioMed Res Int.* (2021) 2021:4604258. doi: 10.1155/2021/4604258
 38. Zhang X, Hu Z, Hao J, Shen J. Low intensity pulsed ultrasound promotes the extracellular matrix synthesis of degenerative human nucleus pulposus cells through FAK/PI3K/akt pathway. *Spine (Phila Pa 1976).* (2016) 41:E248–254. doi: 10.1097/BRS.0000000000001220
 39. Tan Y, Yao X, Dai Z, Wang Y, Lv G. Bone morphogenetic protein 2 alleviated intervertebral disc degeneration through mediating the degradation of ECM and apoptosis of nucleus pulposus cells via the PI3K/Akt pathway. *Int J Mol Med.* (2019) 43:583–92. doi: 10.3892/ijmm.2018.3972
 40. Cheng CC, Uchiyama Y, Hiyama A, Gajghate S, Shapiro IM, Risbud MV. PI3K/AKT regulates aggrecan gene expression by modulating Sox9 expression and activity in nucleus pulposus cells of the intervertebral disc. *J Cell Physiol.* (2009) 221:668–76. doi: 10.1002/jcp.v221:3
 41. Zheng Y, Liu C, Ni L, Liu Z, Mirando AJ, Lin J, et al. Cell type-specific effects of Notch signaling activation on intervertebral discs: Implications for intervertebral disc degeneration. *J Cell Physiol.* (2018) 233:5431–40. doi: 10.1002/jcp.v233:7
 42. Hiyama A, Skubutyte R, Markova D, Anderson DG, Yadla S, Sakai D, et al. Hypoxia activates the notch signaling pathway in cells of the intervertebral disc: implications in degenerative disc disease. *Arthritis Rheum.* (2011) 63:1355–64. doi: 10.1002/art.30246
 43. Wang H, Tian Y, Wang J, Phillips KLE, Binch ALA, Dunn S, et al. Inflammatory cytokines induce NOTCH signaling in nucleus pulposus cells: implications in intervertebral disc degeneration. *J Biol Chem.* (2013) 288:16761–74. doi: 10.1074/jbc.M112.446633
 44. Wang J, Kang Z, Liu Y, Li Z, Liu Y, Liu J. Identification of immune cell infiltration and diagnostic biomarkers in unstable atherosclerotic plaques by integrated bioinformatics analysis and machine learning. *Front Immunol.* (2022) 13:956078. doi: 10.3389/fimmu.2022.956078
 45. Harrell FE Jr., Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med.* (1996) 15:361–87. doi: 10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4
 46. Moons KG, Kengne AP, Grobbee DE, Royston P, Vergouwe Y, Altman DG, et al. Risk prediction models: II. External validation, model updating, and impact assessment. *Heart.* (2012) 98:691–8. doi: 10.1136/heartjnl-2011-301247
 47. Wang S, Tian S, Li Y, Zhan N, Guo Y, Liu Y, et al. Development and validation of a novel scoring system developed from a nomogram to identify Malignant pleural effusion. *EBioMedicine.* (2020) 58:102924. doi: 10.1016/j.ebiom.2020.102924
 48. Moons KG, De Groot JA, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. *PLoS Med.* (2014) 11:e1001744. doi: 10.1371/journal.pmed.1001744
 49. Royston P, Altman DG. External validation of a Cox prognostic model: principles and methods. *BMC Med Res Methodol.* (2013) 13:33. doi: 10.1186/1471-2288-13-33
 50. Vickers AJ, Holland F. Decision curve analysis to evaluate the clinical benefit of prediction models. *Spine J.* (2021) 21:1643–8. doi: 10.1016/j.spinee.2021.02.024
 51. Moons KG, Kengne AP, Woodward M, Royston P, Vergouwe Y, Altman DG, et al. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. *Heart.* (2012) 98:683–90. doi: 10.1136/heartjnl-2011-301246
 52. Aliahmad P, de la Torre B, Kaye J. Shared dependence on the DNA-binding factor TOX for the development of lymphoid tissue-inducer cell and NK cell lineages. *Nat Immunol.* (2010) 11:945–52. doi: 10.1038/ni.1930
 53. Zeng Z, Wei F, Ren X. Exhausted T cells and epigenetic status. *Cancer Biol Med.* (2020) 17:923–36. doi: 10.20892/j.issn.2095-3941.2020.0338
 54. Canaria DA, Rodriguez JA, Wang L, Ye FJ, Yan B, Wang M, et al. Tox induces T cell IL-10 production in a BATF-dependent manner. *Front Immunol.* (2023) 14:1275423. doi: 10.3389/fimmu.2023.1275423
 55. Ge J, Yan Q, Wang Y, Cheng X, Song D, Wu C, et al. IL-10 delays the degeneration of intervertebral discs by suppressing the p38 MAPK signaling pathway. *Free Radic Biol Med.* (2020) 147:262–70. doi: 10.1016/j.freeradbiomed.2019.12.040
 56. Olsen CL, Hsu PP, Glienke J, Rubanyi GM, Brooks AR. Hedgehog-interacting protein is highly expressed in endothelial cells but down-regulated during angiogenesis and in several human tumors. *BMC Cancer.* (2004) 4:43. doi: 10.1186/1471-2407-4-43
 57. Agrawal V, Kim DY, Kwon YG. Hhip regulates tumor-stroma-mediated upregulation of tumor angiogenesis. *Exp Mol Med.* (2017) 49:e289. doi: 10.1038/emmm.2016.139
 58. Liu X, Krishnamoorthy D, Lin L, Xue P, Zhang F, Chi L, et al. A method for characterising human intervertebral disc glycosaminoglycan disaccharides using liquid chromatography-mass spectrometry with multiple reaction monitoring. *Eur Cell Mater.* (2018) 35:117–31. doi: 10.22203/eCM.v035a09

59. Silagi ES, Shapiro IM, Risbud MV. Glycosaminoglycan synthesis in the nucleus pulposus: Dysregulation and the pathogenesis of disc degeneration. *Matrix Biol.* (2018) 71-72:368–79. doi: 10.1016/j.matbio.2018.02.025
60. Ye W, Zhou J, Markova DZ, Tian Y, Li J, Anderson DG, et al. Xylosyltransferase-1 expression is refractory to inhibition by the inflammatory cytokines tumor necrosis factor α and IL-1 β in nucleus pulposus cells: novel regulation by AP-1, Sp1, and Sp3. *Am J Pathol.* (2015) 185:485–95. doi: 10.1016/j.ajpath.2014.09.021
61. Collin EC, Carroll O, Kilcoyne M, Peroglio M, See E, Hendig D, et al. Ageing affects chondroitin sulfates and their synthetic enzymes in the intervertebral disc. *Signal Transduct Target Ther.* (2017) 2:17049. doi: 10.1038/sigtrans.2017.49
62. Bermudez-Lekerika P, Crump KB, Tseranidou S, Nüesch A, Kanelis E, Alminnawi A, et al. Immuno-modulatory effects of intervertebral disc cells. *Front Cell Dev Biol.* (2022) 10:924692. doi: 10.3389/fcell.2022.924692
63. Xu H, Li J, Fei Q, Jiang L. Contribution of immune cells to intervertebral disc degeneration and the potential of immunotherapy. *Connect Tissue Res.* (2023) 64:413–27. doi: 10.1080/03008207.2023.2212051
64. Zhang F, Cui D, Wang K, Cheng H, Zhai Y, Jiao W, et al. Identification and validation of ferroptosis signatures and immune infiltration characteristics associated with intervertebral disc degeneration. *Front Genet.* (2023) 14:1133615. doi: 10.3389/fgene.2023.1133615
65. Guo W, Mu K, Li WS, Gao SX, Wang LF, Li XM, et al. Identification of mitochondria-related key gene and association with immune cells infiltration in intervertebral disc degeneration. *Front Genet.* (2023) 14:1135767. doi: 10.3389/fgene.2023.1135767
66. Murai K, Sakai D, Nakamura Y, Nakai T, Igarashi T, Seo N, et al. Primary immune system responders to nucleus pulposus cells: evidence for immune response in disc herniation. *Eur Cell Mater.* (2010) 19:13–21. doi: 10.22203/eCM.v019a02
67. Song C, Zhou Y, Cheng K, Liu F, Cai W, Zhou D, et al. Cellular senescence - Molecular mechanisms of intervertebral disc degeneration from an immune perspective. *BioMed Pharmacother.* (2023) 162:114711. doi: 10.1016/j.biopha.2023.114711
68. Cunha C, GQT, Ribeiro-Machado C, CLP, Ferreira JR, Molinos M, et al. Modulation of the *in vivo* inflammatory response by pro- versus anti-inflammatory intervertebral disc treatments. *Int J Mol Sci.* (2020) 21(5):1730. doi: 10.3390/ijms21051730
69. Kämpe A, Knebel A, Carlson R, Rohn K, Tipold A. Evaluation of the involvement of Th17-cells in the pathogenesis of canine spinal cord injury. *PLoS One.* (2021) 16:e0257442. doi: 10.1371/journal.pone.0257442
70. Liu H, Chen D, Liu P, Xu S, Lin X, Zeng R. Secondary analysis of existing microarray data reveals potential gene drivers of cutaneous squamous cell carcinoma. *J Cell Physiol.* (2019) 234:15270–8. doi: 10.1002/jcp.v234.9
71. Yamashita J, Ogawa M, Yamashita S, Nomura K, Kuramoto M, Saishoji T, et al. Immunoreactive hepatocyte growth factor is a strong and independent predictor of recurrence and survival in human breast cancer. *Cancer Res.* (1994) 54:1630–3.
72. Ishibashi H, Tonomura H, Ikeda T, Nagae M, Sakata M, Fujiwara H, et al. Hepatocyte growth factor/c-met promotes proliferation, suppresses apoptosis, and improves matrix metabolism in rabbit nucleus pulposus cells *in vitro*. *J Orthop Res.* (2016) 34:709–16. doi: 10.1002/jor.23063
73. Tonomura H, Nagae M, Takatori R, Ishibashi H, Itsuji T, Takahashi K. The potential role of hepatocyte growth factor in degenerative disorders of the synovial joint and spine. *Int J Mol Sci.* (2020) 21(22):8717. doi: 10.3390/ijms21228717
74. Shamji MF, Setton LA, Jarvis W, So S, Chen J, Jing L, et al. Proinflammatory cytokine expression profile in degenerated and herniated human intervertebral disc tissues. *Arthritis Rheum.* (2010) 62:1974–82. doi: 10.1002/art.27444
75. Zhang F, Zhao X, Shen H, Zhang C. Molecular mechanisms of cell death in intervertebral disc degeneration (Review). *Int J Mol Med.* (2016) 37:1439–48. doi: 10.3892/ijmm.2016.2573
76. Egeland NG, Moen A, Pedersen LM, Brisby H, Gjerstad J. Spinal nociceptive hyperexcitability induced by experimental disc herniation is associated with enhanced local expression of Csf1 and FasL. *Pain.* (2013) 154:1743–8. doi: 10.1016/j.pain.2013.05.034
77. Hamilton JA. Colony-stimulating factors in inflammation and autoimmunity. *Nat Rev Immunol.* (2008) 8:533–44. doi: 10.1038/nri2356
78. Kawaguchi S, Yamashita T, Yokogushi K, Murakami T, Ohwada O, Sato N. Immunophenotypic analysis of the inflammatory infiltrates in herniated intervertebral discs. *Spine (Phila Pa 1976).* (2001) 26:1209–14. doi: 10.1097/00007632-200106010-00008
79. Yang C, Cao P, Gao Y, Wu M, Lin Y, Tian Y, et al. Differential expression of p38 MAPK α , β , γ , δ isoforms in nucleus pulposus modulates macrophage polarization in intervertebral disc degeneration. *Sci Rep.* (2016) 6:22182. doi: 10.1038/srep22182
80. Liu Y, Lu T, Liu Z, Ning W, Li S, Chen Y, et al. Six macrophage-associated genes in synovium constitute a novel diagnostic signature for osteoarthritis. *Front Immunol.* (2022) 13:936606. doi: 10.3389/fimmu.2022.936606
81. Mei L, Zheng Y, Gao X, Ma T, Xia B, Hao Y, et al. Hsa-let-7f-1-3p targeting the circadian gene Bmal1 mediates intervertebral disc degeneration by regulating autophagy. *Pharmacol Res.* (2022) 186:106537. doi: 10.1016/j.phrs.2022.106537
82. Wu T, Jia X, Zhu Z, Guo K, Wang Q, Gao Z, et al. Inhibition of miR-130b-3p restores autophagy and attenuates intervertebral disc degeneration through mediating ATG14 and PRKAA1. *Apoptosis.* (2022) 27:409–25. doi: 10.1007/s10495-022-01725-0
83. Wang XB, Wang H, Long HQ, Li DY, Zheng X. LINC00641 regulates autophagy and intervertebral disc degeneration by acting as a competitive endogenous RNA of miR-153-3p under nutrition deprivation stress. *J Cell Physiol.* (2019) 234:7115–27. doi: 10.1002/jcp.v234.5
84. Song J, Wang HL, Song KH, Ding ZW, Wang HL, Ma XS, et al. CircularRNA_104670 plays a critical role in intervertebral disc degeneration by functioning as a ceRNA. *Exp Mol Med.* (2018) 50:1–12. doi: 10.1038/s12276-018-0125-y
85. Xiang Q, Kang L, Wang J, Liao Z, Song Y, Zhao K, et al. CircRNA-CIDN mitigated compression loading-induced damage in human nucleus pulposus cells via miR-34a-5p/SIRT1 axis. *EBioMedicine.* (2020) 53:102679. doi: 10.1016/j.ebiom.2020.102679
86. Yu L, Hao Y, Xu C, Zhu G, Cai Y. LINC00969 promotes the degeneration of intervertebral disk by sponging miR-335-3p and regulating NLRP3 inflammasome activation. *IUBMB Life.* (2019) 71:611–8. doi: 10.1002/iub.v71.5
87. Wang C, Chen R, Zhu X, Zhang X. Long non-coding RNAs LINC00689 inhibits the apoptosis of human nucleus pulposus cells via miR-3127-5p/ATG7 axis-mediated autophagy. *Open Med (Wars).* (2022) 17:1821–32. doi: 10.1515/med-2022-0544
88. Du L, Li X, Gao Q, Yuan P, Sun Y, Chen Y, et al. LncRNA nuclear receptor subfamily 2 group F member 1 antisense RNA 1 (NR2F1-AS1) aggravates nucleus pulposus cell apoptosis and extracellular matrix degradation. *Bioengineered.* (2022) 13:2746–62. doi: 10.1080/21655979.2021.2016087