

RESEARCH

Open Access



# Machine learning algorithm predicts fibrosis-related blood diagnosis markers of intervertebral disc degeneration

Wei Zhao<sup>1,2†</sup>, Jinzheng Wei<sup>2†</sup>, Xinghua Ji<sup>3†</sup>, Erlong Jia<sup>1</sup>, Jinhu Li<sup>1\*</sup> and Jianzhong Huo<sup>4\*</sup>

## Abstract

**Background** Intervertebral disc cell fibrosis has been established as a contributing factor to intervertebral disc degeneration (IDD). This study aimed to identify fibrosis-related diagnostic genes for patients with IDD.

**Methods** RNA-sequencing data was downloaded from Gene Expression Omnibus (GEO) database. The diagnostic genes was identified using Random forest based on the differentially expressed fibrosis-related genes (DE-FIGs) between IDD and control samples. The immune infiltration states in IDD and the regulatory network as well as potential drugs targeted diagnostic genes were investigated. Quantitative Real-Time PCR was conducted for gene expression validation.

**Results** *CEP120* and *SPDL1* merged as diagnostic genes. Substantial variations were observed in the proportions of natural killer cells, neutrophils, and myeloid-derived suppressor cells between IDD and control samples. Further experiments indicated that *AC144548.1* could regulate the expressions of *SPDL1* and *CEP120* by combining *hsa-miR-5195-3p* and *hsa-miR-455-3p*, respectively. Additionally, transcription factors FOXM1, PPARG, and ATF3 were identified as regulators of *SPDL1* and *CEP120* transcription. Notably, 56 drugs were predicted to target these genes. The down-regulation of *SPDL1* and *CEP120* was also validated.

**Conclusion** This study identified two diagnostic genes associated with fibrosis in patients with IDD. Additionally, we elucidated their potential regulatory networks and identified target drugs, which offer a theoretical basis and reference for further study into fibrosis-related genes involved in IDD.

**Keywords** Intervertebral disc degeneration, Fibrosis, Diagnostic genes, GSEA, Immune Infiltration, Regulatory network

<sup>†</sup>Wei Zhao, Jinzheng Wei and Xinghua Ji contributed equally to this work.

\*Correspondence:

Jinhu Li  
xiaofuzigo@163.com  
Jianzhong Huo  
huojianzhong@126.com

<sup>1</sup>First Hospital of Shanxi Medical University, Taiyuan, Shanxi Province, PR China

<sup>2</sup>Shanxi Medical University, Taiyuan, Shanxi Province, PR China

<sup>3</sup>Tongji Shanxi Hospital, Shanxi Bethune Hospital, Shanxi Academy of Medical Sciences, Third Hospital of Shanxi Medical University, Taiyuan, Shanxi Province, PR China

<sup>4</sup>Taiyuan Central Hospital of Shanxi Medical University, Taiyuan, Shanxi Province, PR China



## Introduction

Intervertebral disc degeneration (IDD) is one of the leading causes of low back pain (LBP) affecting ~40% of adults worldwide [1]. This condition is frequently linked to cervical spondylosis, lumbar disc herniation, lumbar spinal stenosis, and lumbar spondylolisthesis [2]. Patients with IDD often experience high disability rates, which significantly affects their quality of life and imposes substantial economic burdens [3]. The intervertebral discs consist of the annulus fibrosus (AF), the nucleus pulposus (NP), and the hyaline cartilage plates of the upper and lower vertebral bodies. IDD mainly involves structural damage to the intervertebral disc and is characterized by an imbalance between catabolic and anabolic processes [4]. This encompasses myeloid nucleus senescence and apoptosis, progressive degradation of the extracellular matrix (ECM), fibro-annular fibrosis culminating in disc bulging, loss of NP and water content, and diminished disc height [5, 6]. Therefore, assessment frameworks like the Pfirrmann grading criteria are employed, relying on disc height, structure, and magnetic resonance imaging signal intensity to clinically classify degeneration levels as stages I–V [7, 8]. Besides, the etiology and pathogenesis of IDD are complex. On the one hand, factors such as age, genetics (e.g., polymorphisms in proteoglycan and collagen-encoding genes), and lifestyle (e.g., occupation, smoking, lack of physical activity, and night work) play contributory roles [9]. On the other hand, cellular processes such as apoptosis, autophagy, and the release of pro-inflammatory factors are involved. Moreover, the pathogenesis of IDD is also heritable, although the exact genetic and molecular mechanisms are still under investigation.

Fibrosis, a pathological process that occurs in various organs, is usually caused by dysregulation of tissue repair in response to chronic inflammation. This can lead to ECM remodeling and excessive accumulation of ECM components [10], consequently altering tissue biomechanical properties. It has been found that alterations in the concentration and morphology of types I and II collagen, and fibronectin during disc fibrosis are closely associated with the progression of disc degeneration [11, 12]. Moreover, myofibroblasts (MF) and macrophages are the main triggers for the progression of fibrosis [10, 13]. Within fibrotic tissues, the regulation of MFs allows for collagen deposition in the ECM, which in turn alters the structure and function of the tissue [14–16]. Also, it was found that MFs were involved in AF repair during IDD [17]. Macrophages can regulate cellular fibrosis through transglutaminases and matrix metalloproteinases (MMPs), which are present in herniated discs and have been widely reported; however, only MMP12 has been identified as a fibrosis marker in IDD [14, 18–22].

In this study, we downloaded two datasets (GSE150408 and GSE124272) from the Gene Expression Omnibus

(GEO) database and analyzed them separately to obtain 336 differentially expressed genes (DEGs). Concurrently, we obtained a total of 2,539 fibrosis-associated genes (FIGs) from the GeneCards database. By intersecting these with the 336 DEGs, we obtained 29 differentially expressed FIGs (DE-FIGs). Next, using the random forest (RF) model with a fold number of 3, we successfully screened out two diagnostic genes—*CEP120* and *SPDL1*. The expressions of these genes were validated within both the GSE150408 dataset (training set) and the GSE124272 dataset (validation set), and their predictive potential was confirmed through receiver operating characteristic (ROC) curve analysis. Further exploration included single-sample gene set enrichment analysis (ssGSEA), as well as the construction of miRNA-mRNA-transcription factor (TF) networks and competing endogenous RNA (ceRNA) networks. These networks shed light on the underlying mechanisms relevant to the diagnostic genes. Additionally, diagnostic genes-associated drug prediction was also performed. Finally, we verified the significance of the DEGs in blood samples of healthy subjects and patients with IDD using quantitative real-time polymerase chain reaction (qRT-PCR). These efforts culminated in the identification of two diagnostic genes, which provide a theoretical basis and reference value for the study of fibrosis-related genes in IDD.

## Materials and methods

### Data sources

The GSE150408 and GSE124272 data sets were downloaded from the GEO database (<http://www.ncbi.nlm.nih.gov/geo/>). GSE150408 contains whole blood samples from 17 patients with IDD and 17 controls, whereas GSE124272 includes 8 pairs of whole blood samples from patients with lumbar disc prolapse and healthy controls. Additionally, 8,621 FIGs were obtained from the GeneCards database (<http://www.genecards.org/>) (Additional file 1). After a screening process based on a relevance score greater than 1, the selection was narrowed down to 2,539 FIGs for subsequent analyses (Additional file 2).

### Identification of DE-FIGs

Limma package 3.44.3 [23] was employed to screen DEGs between IDD samples and controls in GSE124272 and GSE150408 respectively, which were defined as DEGs1 and DEGs2. Then, the common DEGs between DEGs1 and DEGs2 were obtained by overlap analysis and functional enrichment analyses by clusterProfiler v3.16.0 [24] was subsequently conducted with a threshold of  $p < 0.05$  [25–27]. Furthermore, the DEGs that intersected with the 2,539 FIGs were classified as DE-FIGs.

### Selection of diagnostic genes and nomogram construction

To obtain diagnostic genes for IDD, eight machine learning algorithms were used, including Logistic Regression (LR), RF by randomForest v4.6-14, Gradient Decision Tree (GBDT), eXtreme Gradient Boosting (XGB) by xgboost v1.4.1.1, Support Vector Machine (SVM) by e1071 v1.7-3 [28], Artificial Neural Network (ANN) by neuralnet v1.44.2 [29], Decision Tree (DT), Adaboost v4.2 [29], and MultinomialNB (MNB) by kalR. For detecting the most suitable model for DE-FIGs, all the samples in the GSE150408 dataset were randomly divided into training sets and validation sets according to an  $n$ -fold from 1 to 7. The regression analysis of the eight algorithms was conducted in the training set, and the verification was performed in the validation set. The most suitable algorithm was selected according to the comprehensive effects of the  $n$ -fold and each algorithm. Moreover, the diagnostic genes for IDD were identified based on the area under the ROC curve (AUC) values of the model under different numbers of variables. Finally, a nomogram consisting of the diagnostic genes was constructed, which was validated by the C-index, slope of the calibration curve, and decision curve analysis (DCA) curve.

### Diagnostic effectiveness of the diagnostic genes

To investigate whether the diagnostic genes as a whole could successfully distinguish IDD samples from healthy controls, principal component analysis (PCA) was performed on both the GSE150408 and GSE124272 datasets. Also, ROC curves were drawn for each diagnostic gene as well as collectively for all the diagnostic genes to differentiate between their diagnostic potential in IDD samples in the GSE 150,408 dataset. Simultaneously, similar ROC curves were plotted for the GSE124272 dataset to validate the results.

### ssGSEA

ssGSEA analysis was performed for the diagnostic genes by clusterProfiler v3.16.1, using GSE150408 as the input dataset. The correlation coefficient between a single diagnosis gene expression and each gene in the GSE150408 dataset was calculated and ranked. Then, according to the correlation coefficient sorting, GSEA was applied with the filter conditions:  $|NES| > 1$ ; adjusted  $p < 0.05$ ;  $q < 0.25$  for Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses.

### Immune infiltration analysis

The GSE150408 dataset was used as input in the MCP-counter algorithm (<http://github.com/ebecht/MCP-counter>) to calculate the content of eight immune cells (CD8<sup>+</sup> T cells, T cells, cytotoxic lymphocytes, B lineage, natural killer cells, monocytic lineage, myeloid dendritic

cells, and neutrophils), one fibroblast, and one epithelial cell in IDD samples and controls. In addition, single-sample gene set enrichment analysis (ssGSEA) analysis was performed to compute the proportions of 28 types of immune cells in IDD samples and controls in the GSE150408 dataset.

### Construction of miRNA-mRNA-TF and ceRNA regulatory networks

The level of mRNA is regulated by both miRNAs and TFs. Therefore, both miRNAs and TFs for diagnostic genes were predicted. First, the miRWalk website (<http://mirwalk.umm.uni-heidelberg.de/>) was used to predict miRNAs of diagnostic genes, and the miRNAs with energy  $< -25$  were selected. ChEA3 (<http://amp.pharm.mssm.edu/chea3/>) was subsequently utilized to predict TFs of the diagnostic genes, and the TFs with chip-seq were selected. The gene-TF relationship pairs and gene-miRNA relationship pairs were integrated into a miRNA-gene-TF network.

Meanwhile, the selected miRNAs of diagnostic genes acquired above were used to predict their target long non-coding RNAs (lncRNAs) from the Star base website (<http://starbase.sysu.edu.cn/>) with the selection criteria of clipExpNum  $> 5$ . After combining gene-miRNA and miRNA-lncRNA regulation pairs, the ceRNA network was established.

Afterwards, target drugs of diagnostic genes were predicted through the Comparative Toxicogenomics Database (CTD) (<http://ctdbase.org/>), and a drug-mRNA network was constructed. All the networks were visualized by Cytoscape v3.7.2.

### Validation by qRT-PCR

The qRT-PCR analysis was performed peripheral blood mononuclear cells (PBMCs) isolated from 24 whole blood samples (12 IDD samples vs. 12 Controls) to further investigate the gene expression pattern of signature genes. TRIzol reagent was first used to extract total RNA following the manufacturer's procedure (Thermo Fisher Scientific, Waltham, MA, USA). After detection of concentration and purity, 1  $\mu$ g total RNAs were reverse transcribed to synthesize cDNAs using the SureScript-First-strand-cDNA-synthesis-kit (Servicebio, Wuhan, China). A 20  $\mu$ L reaction system, including 3  $\mu$ L diluted cDNAs, 5  $\mu$ L of 2 $\times$ Universal Blue SYBR Green qPCR Master Mix, and 1  $\mu$ L each of forward and reverse primer were combined for qRT-PCR reaction with a BIO-RAD CFX96 Touch TM PCR detection system (Bio-Rad, Hercules, CA, USA). The primer sequences of signature genes are shown in Table 1. The amplification conditions included the following parts: 95  $^{\circ}$ C for 60 s, 40 cycles at 95  $^{\circ}$ C for 20 s, 55  $^{\circ}$ C for 20 s, and 72  $^{\circ}$ C for 30 s. For subsequent melting curve analysis, a single peak represented

**Table 1** Primers for qRT-PCR used in the current study

Primer	Sequence
CEP120 F	AGTTGGCTACTGATCCTGTGG
CEP120 R	GGAGTTTGATAGGAGTACGCTGT
SPDL1 F	TGATGACCATGACTGAGCAGA
SPDL1 R	ATAACTCTTACTGCTTCTGTGCC
internal reference-GAPDH F	CCCATCACCATCTTCCAGG
internal reference-GAPDH R	CATCACGCCACAGTTTCCC

F: forward; R: reverse

the reliability of the primer, and gene expression was computed with the  $2^{-\Delta\Delta Ct}$  method.

**Results**

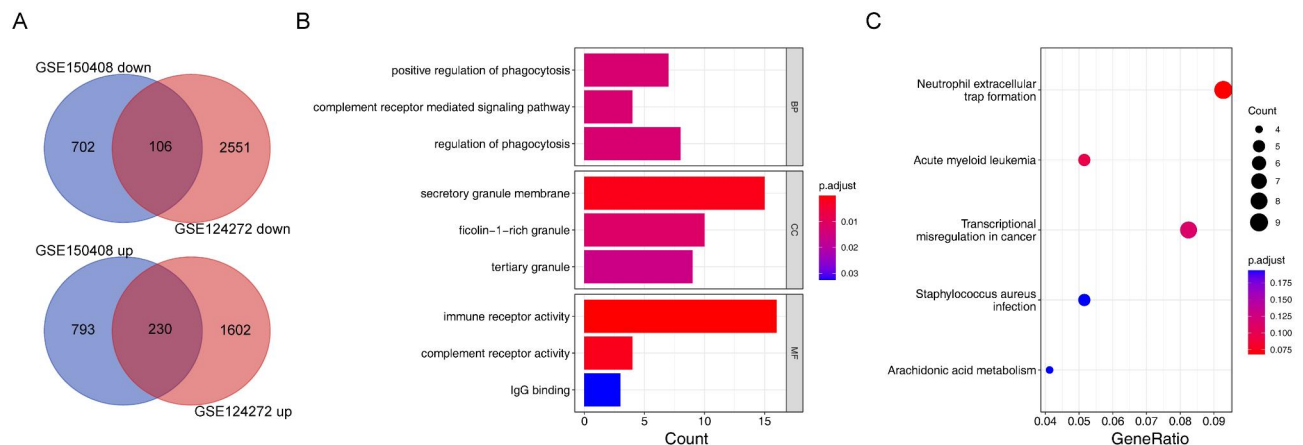
**A total of 336 common DEGs were identified**

According to the differential analysis between IDD and healthy samples, 4,489 DEGs1 and 1,832 DEGs2 were screened out from the GSE124272 and GSE150408 datasets, respectively (Additional files 3–5). After comparing these datasets, 336 common DEGs emerged, which included 230 up-regulated and 106 down-regulated genes (Fig. 1A). The functional enrichment analyses indicated that these 336 common DEGS were associated with 9

KEGG pathways and 9 GO terms (Fig. 1B C). Involvement in the KEGG pathways included, among others, neutrophil extracellular trap formation, acute myeloid leukemia, and transcriptional misregulation in cancer; while for the GO terms it included, amongst others, positive regulation of phagocytosis, complement receptor-mediated signaling pathway, regulation of phagocytosis, secretory granule membrane, and immune receptor activity.

**CEP120 and SPDL1 were selected as diagnostic genes for IDD**

A total of 29 DE-FIGs were obtained from the overlap analysis between 336 common DEGs and 2,539 FIGs (Additional file 6, Fig. 2A). Moreover, the results of eight machine learning algorithms showed that the optimal fold number was 3 and RF was the most suitable model. As a result, an *n*-fold of 3 for the RF model was selected to construct a diagnostic model (Table 2; Fig. 2B). Subsequently, the RF model generated importance values for each variable, denoted as %IncMSE. Based on these values, the 29 DE-FIGs were ranked from highest to lowest. Using this ranking, the predicted AUC values of the



**Fig. 1** Identification and evaluation of common DEGs between the GSE124272 and GSE150408 datasets. **(A)** Venn diagram of common DEGs. **(B)** The enriched GO terms for 336 common DEGs. The horizontal coordinates represent the gene ratio, and the vertical coordinates represent the GO term. Bubble plots are arranged from smallest to largest *p*-value. **(C)** KEGG bubble chart for 336 common DEGs. The horizontal coordinates represent the gene ratio, and the vertical coordinates represent the terms enriched by KEGG. Bubble plots are sorted by *p*-value from smallest to largest. DEGs, differentially expressed genes; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes

**Table 2** The *n*-fold of random forest model

<i>n</i> -fold	LR	RF	XBG	SVM	ANN	DT	ADABOOST	MNB	Average
1	0.75	0.50	0.75	0.50	1.00	0.75	0.75	0.75	0.72
2	0.75	1.00	0.75	1.00	0.75	0.75	0.75	1.00	0.84
3	0.67	1.00	1.00	1.00	1.00	0.83	1.00	1.00	0.94
4	1.00	1.00	1.00	0.75	0.75	0.50	1.00	0.75	0.84
5	1.00	1.00	1.00	1.00	0.50	0.50	0.75	0.50	0.78
6	0.50	1.00	0.50	1.00	0.50	0.50	0.50	0.50	0.63
7	0.50	1.00	0.50	1.00	0.50	0.50	0.50	0.50	0.63
Average	0.74	0.93	0.79	0.89	0.71	0.62	0.75	0.71	0.77

RF model under different variables were displayed in a line chart (Fig. 2C). Notably, it can be observed that when the number of variables reached 2, the AUC value reached 1 and remained at 1 as the number of variables increased. Therefore, these two genes, namely *CEP120* and *SPDL1*, were considered diagnostic genes.

Further, a nomogram comprising *CEP120* and *SPDL1* was constructed (Fig. 2D) with a C-index of 0.7681661 and a corrected C-index of 0.7326992, indicating that the prediction of the nomogram was accurate (Fig. 2E). Moreover, within the high-risk threshold range from 0 to 1, the nomogram for *CEP120* and *SPDL1* could provide benefit, with the net benefit higher than the curve of each single-gene. This indicated that the overall effectiveness of the nomogram exceeded that of a single gene, affirming the selection of these diagnostic genes as reasonable and effective (Fig. 2F).

**The diagnostic effectiveness of the diagnostic genes**

The PCA results for both the GSE150408 and GSE124272 datasets showed that these two diagnostic genes, when considered together, could effectively distinguish IDD samples from healthy controls (Fig. 3A B).

Furthermore, the ROC curve for each diagnostic gene in both datasets showed AUC values greater than 0.75, which illustrated the diagnostic efficacy of each gene (Fig. 3C D). Simultaneously, the ROC curves for the combination of *CEP120* and *SPDL1* also had AUC values

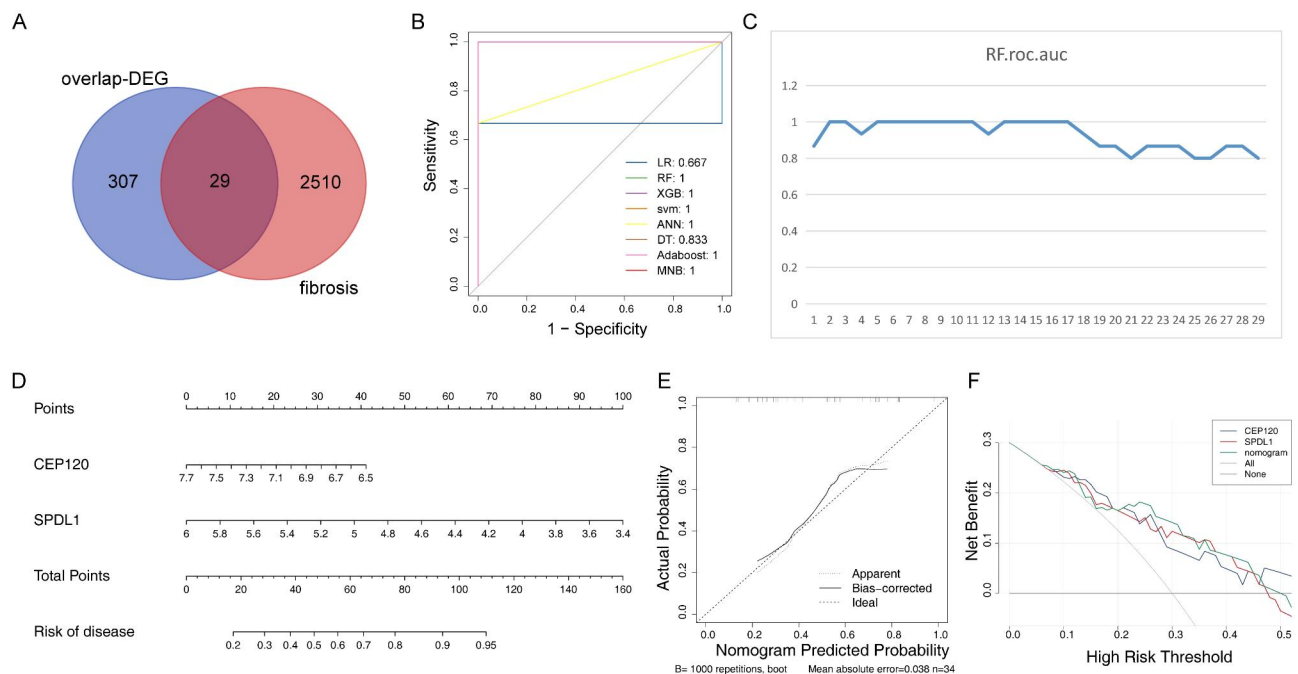
greater than 0.75 in both datasets, which also suggests the overall diagnostic effectiveness of the diagnostic genes was similar to their individual effects, confirming their substantial diagnostic power (Fig. 3E F).

**ssGSEA**

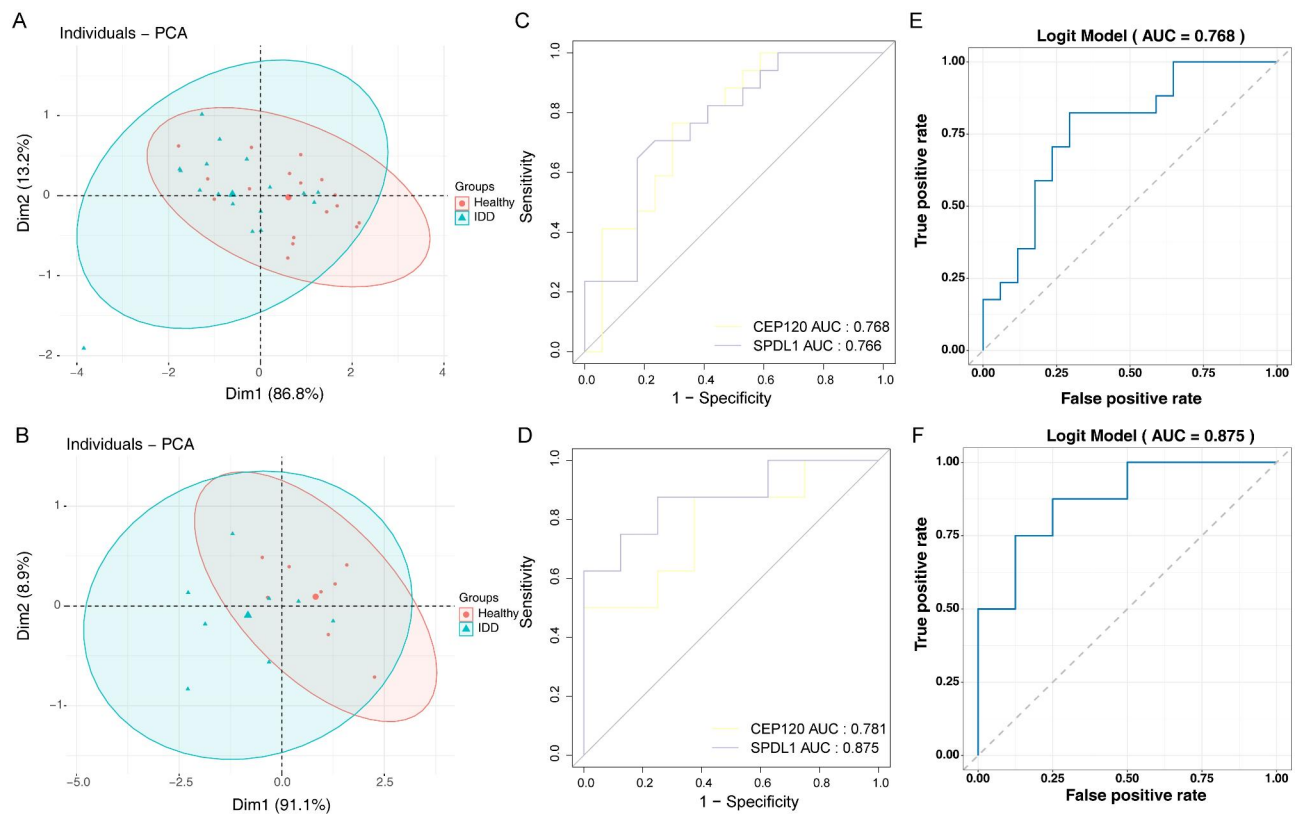
The top 10 enriched GO terms and KEGG pathways of *CEP120* and *SPDL1* are displayed in Fig. 4A D. Both diagnostic genes were found to participate in significant biological processes, including cellular nitrogen compound catabolic processes, corporation organization, cytoplasmic translation, DNA conformation change, and DNA-dependent DNA replication. Additionally, common KEGG pathways of both diagnostic genes included cell cycle, neuroactive ligand-receptor interaction, olfactory transduction, peroxisome, ribosome, spliceosome, and ubiquitin-mediated proteolysis.

**Immune infiltration between IDD and control samples**

The MCP-counter algorithm results of the GSE150408 dataset for both IDD and control samples showed that CD8 T cells were most abundant, while fibroblasts and epithelial cells were the least prevalent (Fig. 5A). Among the eight cell types analyzed, the proportion of natural killer cell was significantly lower in IDD samples, while neutrophils represented a larger percentage in IDD samples (Fig. 5B).



**Fig. 2** Identification and evaluation of diagnostic genes for IDD. (A) Venn diagram of differentially expressed fibrosis genes. (B) ROC curves for the eight models at  $n$ -fold=3. (C) Predicted area under the curve values of RF models with different numbers of variables. (D) Nomogram to predict the survival rate of patients with IDD. (E) Calibration curve of the nomogram. (F) RF modeling for decision. IDD, intervertebral disc degeneration; RF, random forest; ROC, receiver operating characteristic



**Fig. 3** The diagnostic accuracy of the diagnostic genes. **(A)** PCA results of GSE150408. **(B)** PCA results of GSE124272. **(C)** ROC curves for diagnostic genes of GSE150408. **(D)** ROC curves for diagnostic genes of GSE124272. **(E)** Logistic regression ROC curve of GSE150408. **(F)** Logistic regression ROC curve of GSE124272. PCA, principal component analysis; ROC, receiver operating characteristic

In addition, it can be seen from the ssGSEA results that the proportions of myeloid-derived suppressor cells (MDSCs) and neutrophils varied between IDD and healthy samples, with both cell types being more abundant in IDD samples (Fig. 5C-D).

#### The miRNA-mRNA-TF and ceRNA regulatory networks and drug prediction

The miRNA prediction result from the miRwalk website indicated that *CEP120* and *SPDL1* are predicted to target 17 miRNAs. Further, 120 TFs were predicted to interact with these two diagnostic genes. Consequently, a complex miRNA-mRNA-TF regulatory network was constructed, consisting of 139 nodes (comprising the two diagnostic genes, 17 miRNAs, and 120 TFs) and 156 edges (Fig. 6A). For instance, miRNAs such as *hsa-miR-5195-3p*, *hsa-miR-6801-5p*, *hsa-miR-671-5p*, *hsa-miR-4707-5p*, *hsa-miR-3651*, along with TFs including FOXM1, PPARG, ATF3, ZNF217, and ZBTB33, were identified as potential regulators of *SPDL1*. On the other hands, the transcription of *CEP120* was regulated by *hsa-miR-150-3p*, *hsa-miR-4709-3p*, *hsa-miR-455-3p*, *hsa-miR-6836-5p*, *hsa-miR-6742-5p*, *hsa-miR-4776-3p*, *hsa-miR-1538*, *hsa-miR-7112-5p*, *hsa-miR-12,115*, *hsa-miR-663a*, *hsa-miR-4469*, *hsa-miR-6803-5p*, along with

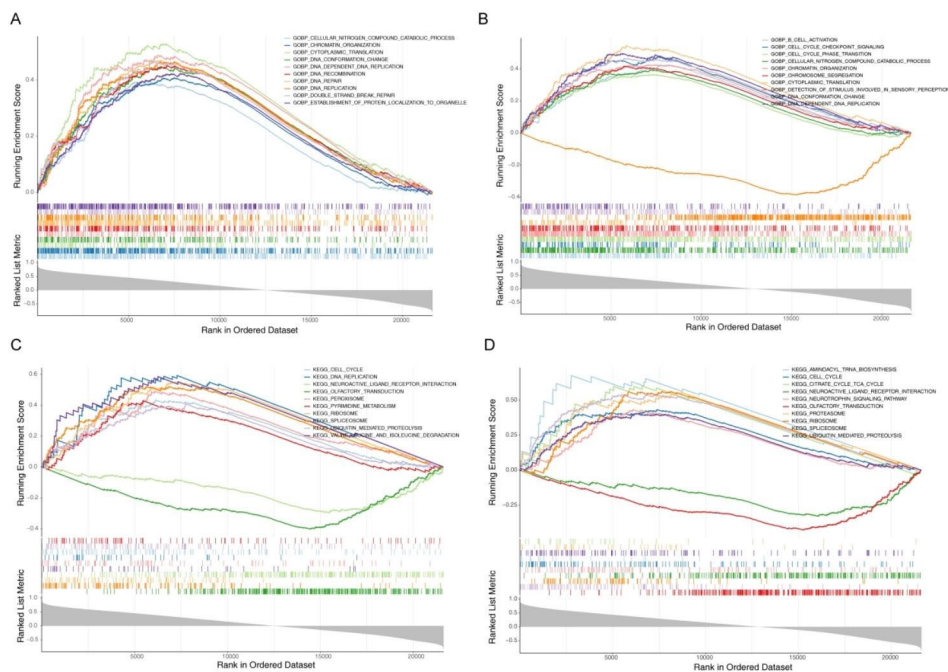
TFs including GATA2, ELF1, PPARD, GATA1, ELK4, SP4, and TET1.

Besides, based on the 17 target miRNAs of the two diagnostic genes, 31 lncRNAs were predicted to interact. However, these lncRNAs were only predicted by four miRNAs (*hsa-miR-455-3p*, *hsa-miR-663a*, *hsa-miR-671-5p*, and *hsa-miR-5195-3p*). Therefore, a ceRNA network was constructed consisting of 37 nodes (31 lncRNAs, 4 miRNAs, and the 2 diagnostic genes) and 42 edges. For example, *AC144548.1* was identified as a regulator of *SPDL1* and *CEP120* through *hsa-miR-5195-3p* and *hsa-miR-455-3p*, respectively (Fig. 6B).

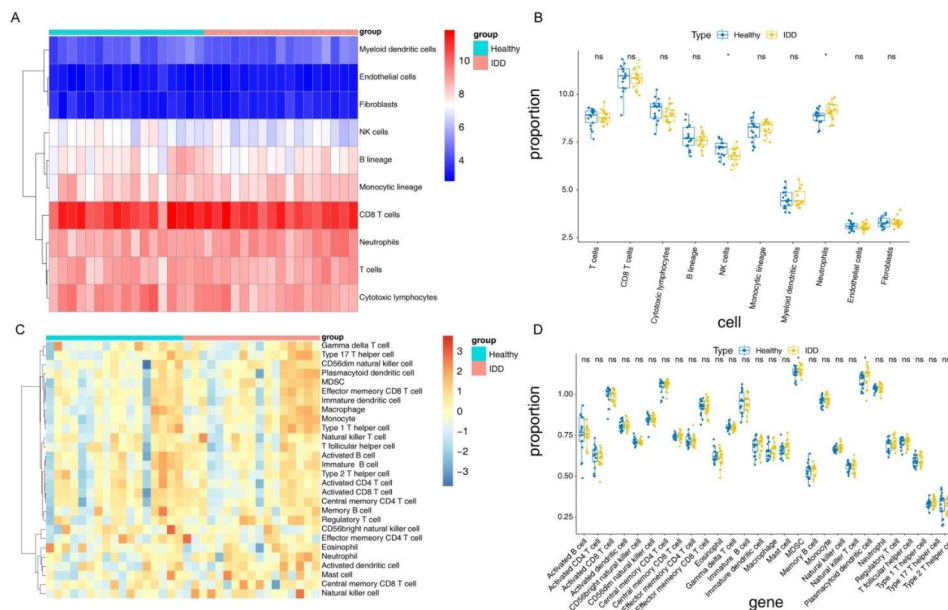
Finally, a total of 56 drugs were predicted to be associated with the two diagnostic genes. D014028, D000077210, D000077185, D004317, D001564, and D014212 were identified as common drugs for both diagnostic genes (Fig. 6C). These findings provide insights into potential therapeutic interventions and the regulatory mechanisms of these diagnostic genes in the context of the disease.

#### Validation of two diagnostic genes

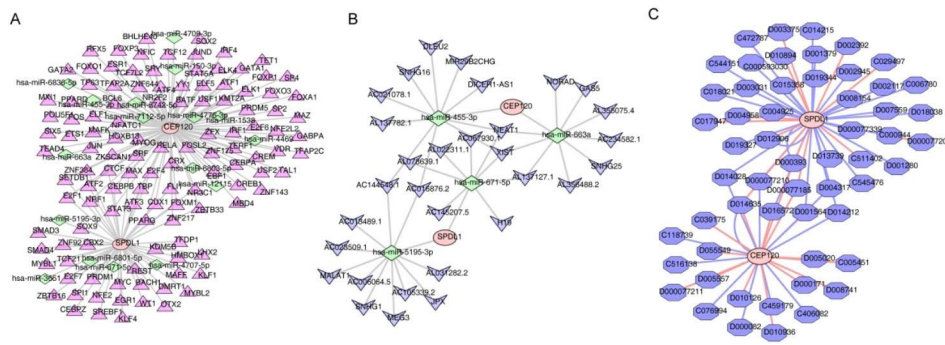
To confirm the reliability of the two diagnostic genes obtained, the expressions of *CEP120* and *SPDL1* were verified by qRT-PCR. It can be seen from the results that



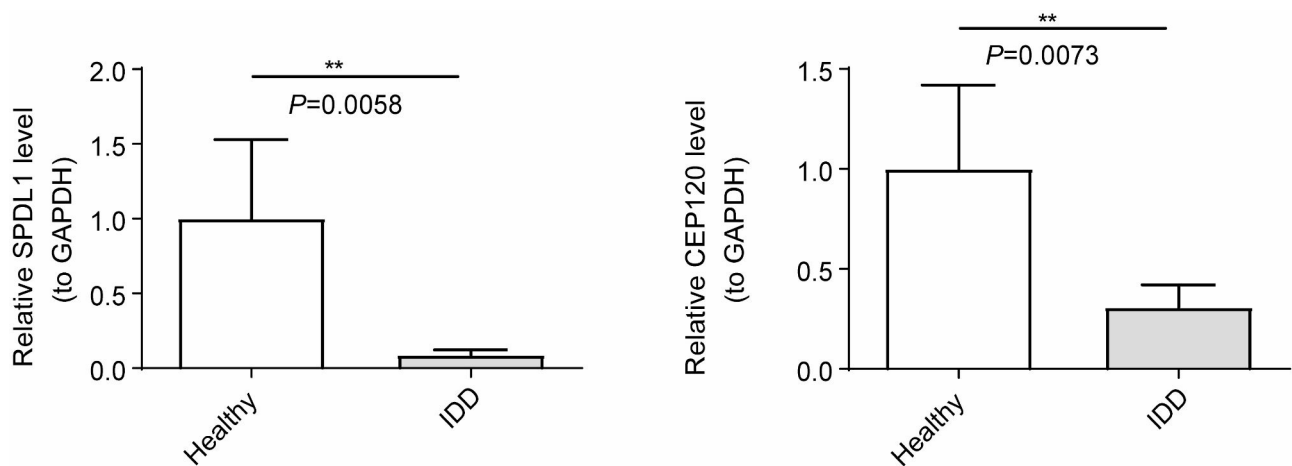
**Fig. 4** Single-sample GSEA analysis. **(A)** The top 10 results for GO enrichment of *CEP120*. The vertical coordinates represent the ES; positive ES indicates that the functional gene set is enriched in the front of the sequence; negative ES indicates that the functional gene set is enriched at the back of the sequence. The horizontal coordinates represent genes, and each small vertical line represents a gene. Overall, all genes in the top 10 are up-regulated. **(B)** The top 10 results for GO enrichment of *SPDL1*. Overall, 9 genes are up-regulated, and 1 gene is down-regulated. **(C)** The top 10 results for KEGG enrichment of *CEP120*. Overall, 8 genes are up-regulated and 2 are down-regulated. **(D)** The top 10 results for KEGG enrichment of *SPDL1*. Overall, 8 genes are up-regulated and 2 are down-regulated. ES, enrichment score; GO, Gene Ontology; GSEA, gene set enrichment analysis; KEGG, Kyoto Encyclopedia of Genes and Genomes



**Fig. 5** Immune-related analysis of diagnostic genes. **(A)** Heat map of different cell contents derived by MCP-counter analysis. The first row indicates the sample grouping, with blue for healthy samples and red for disease samples. Each row indicates the quantity of each cell in different samples, while each column indicates the total cell counts in each sample. The tree on the left side represents the results of cluster analysis, grouping different cell types from various samples. **(B)** Box plot of cell content between groups as derived by MCP-counter analysis. **(C)** Heat map of the different cell contents derived from ssGSEA analysis. The first row indicates the sample grouping, with blue for healthy samples and red for disease samples. Each row indicates the quantity of each of the 28 cell types in different samples, while each column indicates the cell counts in each sample. The tree on the left side represents the results of the cluster analysis of different cell types from different samples. **(D)** Box plot of cell content between groups as derived by ssGSEA analysis



**Fig. 6** Construction of diagnostic genes-related regulatory networks. **(A)** The miRNA-mRNA-TF network diagram. The red circles indicate mRNAs; the pink triangles indicate TFs; the green diamonds indicate miRNAs. **(B)** The mRNA-miRNA-lncRNA network diagram. The red circles indicate mRNAs; the blue arrow shapes indicate lncRNAs; the green diamond indicates miRNAs. **(C)** Drug-diagnostic gene network diagram. The red circles indicate mRNAs; the blue hexagons indicate drugs



**Fig. 7** Comparison of differences in expression of diagnostic genes between healthy (n=12) and IDD (n=12) samples by qRT-PCR ( $p < 0.01$ )

*CEP120* expressions were lower in IDD samples compared to controls ( $p < 0.0058$ ); similarly, *SPDL1* expression was down-regulated in IDD samples ( $p < 0.0073$ ) (Additional file 7, Fig. 7). These findings provide experimental validation of the gene expression differences observed between healthy and IDD samples, supporting the diagnostic significance of *CEP120* and *SPDL1*.

**Discussion**

The intervertebral disc, a complex cartilaginous tissue, connects adjacent vertebral bodies and maintains the mechanical load that allows the spine to move. In healthy intervertebral discs, homeostasis between assimilation and catabolic processes maintains the state of the ECM. Unfortunately, aging and constant mechanical stress can disrupt the metabolic environment of disc cells. This leads to the dysregulation of various factors in the surrounding environment, potentially resulting in the breakdown of macromolecules and the subsequent development of IDD [30].

Current treatments for IDD are either surgical interventions or the relief and management of individual symptoms. Nevertheless, various innovative approaches have been developed, including therapeutic protein or stem cell injections, gene therapy, molecular therapy, and tissue engineering. These approaches have shown promising results in animal models and hold promise for clinical applications [31, 32]. To expedite their clinical application, it is necessary to conduct in-depth research into the underlying mechanisms of IDD, examining key biomarkers, signaling pathways, and immune infiltration. In the study, through various bioinformatics analyses, we preliminarily identified two diagnostic genes and the associated regulatory networks, along with potential drugs that could be relevant to the progression of IDD.

The integration of bioinformatics has significantly enriched our study. We performed differential analysis on two datasets, GSE150408 and GSE124272, sourced from the GEO database, to obtain 336 DEGs, 230 up-regulated and 106 down-regulated. In particular, we focused on two promising diagnostic candidate genes, *CEP120* and



*SPDL1*, which were selected through the RF model with a fold of 3. These genes were evaluated in the GSE150408 dataset and validated in the GSE124272 dataset.

CEP120 is a sub-centromere enrichment protein involved in centromere elongation, including centromere replication, assembly, elongation, and maturation [33–38]. During the cell cycle, CEP120 levels gradually increase from the early S phase to the M phase and decrease significantly at the end of mitosis. Its overexpression not only leads to centriole overgrowth but also generates atypical redundant centrioles [36]. On the other hand, the lack of CEP120 results in delayed or stalled cell cycles in vivo. It is worth noting that disc fibrosis has been associated with telomere shortening and DNA damage response (DDR) [39–41]. In patients with IDD, the ends of DNA strands are incompletely replicated during continuous cellular replication, causing gradual shortening, decreased telomerase activity, and activation of aging-related pathways [42]. In this context, intracellular DDR is activated in response to changes in telomere length, suggesting a potential correlation between the DDR process and down-regulation of CEP120 [43]. In addition, the activation of cellular senescence in NP cells can arrest the cell cycle and elevate reactive oxygen species (ROS) concentrations. DDR triggered by oxidative activation further induces senescence, degeneration, and fibrosis of intervertebral disc cells [44]. Additionally, various inflammatory factors such as TNF- $\alpha$ , IL-1 $\alpha$ , IL-1 $\beta$ , IL-6, IL-17, and various chemokines can enhance the catabolism of disc ECM and promote the inflammatory response, exacerbating the disease [45–48].

SPDL1 is a functional glycosylated protein that has an inhibitory effect on activated T cells [49, 50]. It is generated through selective splicing and translation of PD-L1 precursor mRNA [51–53]. SPDL1 is also an active circulating protein that induces apoptosis of CD8<sup>+</sup> T cells and impairs the ability of these effector cells to kill tumor cells [54, 55]. Patients with high levels of SPDL1 tend to benefit more from anti-PD-L1 treatment. However, it is noteworthy that *SPDL1* expression does not always correlate with immunosensitivity, especially in lung cancer [56, 57]. Whether PD-L1 positively or negatively regulates *SPDL1* remains controversial. Besides, the SPDL1 level is elevated in many inflammatory diseases and is thus considered to be an inflammatory marker reflecting the widespread expression of mPD-L1 in an inflammatory environment. Based on the findings of our current research, we believe that *CEP120* can slow down mitosis and bring tissue development to a near-standstill, resulting in qualitative changes such as cell apoptosis. Meanwhile, the high *SPDL1* expression could alter the immune microenvironment, inhibiting immune responses and making the disc tissue less resistant to external stimuli or tolerant thereof. This cascade of events eventually leads

to fibrosis, which may serve to stabilize the disc. However, the association between IDD and *CEP120* as well as *SPDL1* is relatively unexplored in existing literature. Therefore, our findings may indeed provide a new avenue for the diagnosis and treatment of IDD.

For the immune microenvironment of patients with IDD, the MCP-counter algorithm was used to quantify immune cells, fibroblasts, and epithelial cells in the GSE150408 dataset. This analysis revealed that the content of natural killer cells and neutrophils differed significantly between groups. The observed increase in neutrophils aligns with previous literature, where stimulation of IL-1 $\beta$  led to activated disc cells and the secretion of neutrophil-initiated cytokines, which are associated with inflammatory response in IDD [58]. This is consistent with the involvement of neutrophils in the cellular NLRP3/caspase-1 pathway and IL-1 $\beta$  secretion, as reported in earlier studies [59]. Simultaneously, NP cell damage caused by disc degeneration leads to a massive proliferation of natural killer cells, which induces hypersensitivity reactions [60, 61]. Furthermore, the ssGSEA results show higher infiltration of MDSCs and neutrophils in the IDD samples. Persistent inflammation during tissue aging is associated with a compensatory anti-inflammatory response that prevents excessive tissue damage. This response includes the activation of an associated immunosuppressive network that involves an increase in the number of MDSCs, regulatory T cells, and macrophages [62]. MDSCs, in particular, have been implicated in the development of cellular fibrosis by producing immunomodulatory mediators that stimulate the disposition of fibroblasts and stromal proteins a process aimed at limiting harmful inflammation [63].

The ssGSEA enrichment results for the two diagnostic genes repeatedly highlighted specific biological processes and signaling pathways. These include DNA conformation change, DNA-dependent DNA replication, as well as the KEGG signaling pathways like the spliceosome, ubiquitin-mediated proteolysis, and cell cycle. These repeated enrichments indicate a strong relationship between these diagnostic genes and IDD progression, particularly in processes related to cell cycle, DNA damage, repair, replication, and degradation. Furthermore, upon literature review, we found that several molecular biomarkers are widely used to detect cellular senescence and degeneration. These include oncogene p53, cell cycle kinase-dependent (CDK) inhibitors p16 and p21, cell cycle regulators (retinoblastoma protein Rb), p38, and telomere length [64]. There is evidence that the central signaling pathways mediating disc cell fibrosis are the p53-p21-retinoblastoma protein (Rb) pathway and the p16-Rb pathway, but the exact role of each signaling molecule has not been fully elucidated [65].

In the p53-p21-Rb pathway, p53 plays an irreplaceable role in inducing cell senescence by responding to telomerase shortening and DDR (i.e., it initiates the irreversible process that brings the cell cycle to a standstill) [64]. At the site of DNA damage, following the direct phosphorylation of p53, p21 (a crucial downstream regulator of p53 signaling) inhibits CDK2 activity. This, in turn, prevents Rb phosphorylation, thus blocking the activation of E2F factors and delaying the cell cycle transition from the G1 to S phase ultimately leading to cell cycle arrest [66, 67]. This mechanism may be one of the potential reasons why *CEP120* expression is not elevated in this context. Notably, in NP and AF specimens of IDD, the expression of *p53*, *p21*, and *Rb* genes were significantly up-regulated, while telomere shortening and decreased telomerase activity could be detected [68, 69], confirming the oxidative activation of this pathway in NP cells.

In the p16-Rb signaling pathways, oxidative activation of p16 can inhibit the expression of *CDK4* and *CDK6*, thereby inhibiting the phosphorylation of Rb and delaying or halting cell cycle progression [69, 70]. The number of p16-positive cells in intervertebral disc tissue was positively correlated with Pfirrmann grading as well [68]. This correlation is attributed to that the p16-Rb pathway mediates mitochondrial damage with high glucose concentration, exacerbating the production of ROS in the intervertebral disc [71].

Given that the level of mRNA is regulated by both miRNAs and TFs, we predicted 17 target miRNAs and 120 TFs relevant to the two diagnostic genes. This prediction allowed us to construct a miRNA-mRNA-TF regulatory network. According to the existing studies, we found that p53 can regulate the expression of *hsa-miR-663a*, which in turn has a positive effect on cell proliferation, colony formation, and targeted invasion during the cell cycle, with E2F2 as the key TF target [72]. By acting on *CCND1* and *CDC34*, *miR-671-5p* not only inhibits cell proliferation, cell cycle progression, as well as cloning in vitro and in vivo, but also promotes cell apoptosis [73]. Overexpression of *hsa-miR-5195-3p*, a potential regulator of the TGF $\beta$  signaling pathway, can significantly down-regulate c-MYC and cyclin D1, concomitantly affecting p21 levels by increasing its expression [74]. Besides, the deletion of p16 suppresses the down-regulation of TF E2F1/2 levels and regulates oxidative stress, thereby slowing down disc degeneration [75]. According to the results of the miRNA-mRNA-TF regulatory network, *CEP120* was considered to be involved in this process as well.

It was also found that NRF1 induces autophagy and inhibits apoptotic responses by promoting *Atg7* expression in compressed NP cells, delaying disc degeneration [76]. Phosphorylation of STAT3 accelerates NP apoptosis and IDD by degrading ECM [77]. Our forecast results for TF also confirm this view.

The prediction of potential drugs for *CEP120* and *SPDL1*, as target genes, revealed 56 potential drugs or molecular compounds. After review, we found that D014028 (tobacco smoke pollution) and D001564 (benzo(a)pyrene), both components of cigarette smoking, accelerated the degradation of collagen and proteoglycan in rat intervertebral discs thereby causing apoptosis of NP cells and eventually IDD. The therapeutic effect of D000077185 (resveratrol) on IDD is related to its antioxidant and anti-inflammatory activity, while resveratrol also affects NP cell apoptosis, autophagy, and ECM biosynthesis through various signaling pathways [78–80]. However, the specific mechanism of action of the two genes in IDD requires further experimental studies. These findings suggest that once the specific mechanisms of *CEP120* and *SPDL1* in IDD are elucidated, these drugs and molecular compounds could offer new therapeutic avenues for the treatment of IDD.

In summary, this study has elucidated the diagnostic genes-related biological processes and pathways associated with the development of IDD through comprehensive bioinformatics analysis. These findings contribute to our understanding of the etiology of IDD and shed light on potential diagnostic genes (i.e., *CEP120* and *SPDL1*) and therapeutic agents. Notably, the study predicted various therapeutic agents, including tobacco smoke pollution, benzo(a)pyrene, and resveratrol, which may play roles in the development and progression of IDD. However, research on these drugs is currently in the animal stage, and further clinical studies are needed to investigate their effects in patients with IDD.

Our study is not without limitations. First, all results were derived by bioinformatics algorithms without raw data from clinical samples to assess the quality of sequencing samples, such as Pfirrmann grading. Second, fibrosis-related genes in IDD still need to be explored in depth by various in vivo or in vitro experiments. We subsequently need more studies to validate and reveal further mechanisms, which could lead to the discovery of new potential therapeutic targets for IDD.

## Conclusion

In summary, *CEP120* and *SPDL1* were identified as key fibrosis-related diagnostic genes for patients with IDD. Additionally, a potential regulatory network and therapeutic agents targeting these two key genes were preliminarily predicted. These findings provide a reference for further study of IDD in fibrosis-related genes.

## Abbreviations

IDD	Intervertebral disc degeneration
LBP	Low back pain
AF	Annulus fibrosus
NP	Nucleus pulposus
ECM	Extracellular matrix
FN	Fibronectin

MF	Myofibroblasts
MMPs	Matrix metalloproteinases
DEGs	Differentially expressed genes
FIGs	Fibrosis-associated genes
DE-FIGs	Differentially expressed FIGs
RF	Random forest
ROC	Receiver operating characteristic
GSEA	Gene set enrichment analysis
TFs	Transcription factors
qRT-PCR	Quantitative real-time polymerase chain reaction
LR	Logistic Regression
GBDT	Gradient boosting Decision Tree
SVM	Support Vector Machine
ANN	Artificial Neural Network
DT	Decision Tree
MNB	MultinomialNB
AUC	The area under the ROC curve
DCA	Decision curve analysis
ssGSEA	Single-sample gene set enrichment analysis
ceRNA	Competing endogenous RNA
CTD	Comparative Toxicogenomics Database

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12920-023-01705-6>.

Supplementary Material 1  
Supplementary Material 2  
Supplementary Material 3

## Acknowledgements

Not applicable.

## Author contributions

W.Z. and J.H. designed this study. J.W. and X.J. conducted the experiments. E.J. helped with data analyses. W.Z. and J.W. wrote the manuscript, and J.L. revised the manuscript. All authors have read and approved the manuscript.

## Funding

This work was supported by Shanxi Province Science and Technology Research Project (No.20150313012-1), Key Re-search and Development (R&D) Projects of Shanxi Province (No.201803D31134), Applied Basic Research Program of Shanxi Province (201901D111409), Taiyuan Science and Technology Programme (No.202201), and Subject of Shanxi Provincial Health Commission (2020011).

## Data Availability

The data presented in this study are openly available in the GEO database (<https://www.ncbi.nlm.nih.gov/geo/>), reference number [GSE150408 and GSE124272].

## Declarations

### Competing interests

The authors declare no competing interests.

### Ethics approval and consent to participate

Research carried out here were in compliance with the Helsinki Declaration. The protocol of this study was approved by the Ethics Committee of First Hospital of Shanxi Medical University. Written informed consent forms were obtained from patients and healthy controls before blood taking.

### Consent for publication

Not applicable.

Received: 14 May 2023 / Accepted: 15 October 2023

Published online: 01 November 2023

## References

1. Fan H, Chen Z, Tang H, Shan L, Chen Z, Liu S, Zhang Y, Guo X, Yang H, Hao D. Necroptosis of nucleus pulposus cells involved in intervertebral disc degeneration through MyD88 signaling. *Front Endocrinol.* 2022;13:994307.
2. Oichi T, Taniguchi Y, Oshima Y, Tanaka S, Saito T. Pathomechanism of intervertebral disc degeneration. *JOR Spine.* 2020;3(1):e1076.
3. Hoy D, March L, Brooks P, Blyth F, Woolf A, Bain C, Williams G, Smith E, Vos T, Barendregt J, et al. The global burden of low back pain: estimates from the global burden of Disease 2010 study. *Ann Rheum Dis.* 2014;73(6):968–74.
4. Kos N, Gradisnik L, Velnar T. A brief review of the degenerative intervertebral disc Disease. *Med Archives (Sarajevo Bosnia Herzegovina).* 2019;73(6):421–4.
5. Kirnaz S, Capadona C, Wong T, Goldberg J, Medary B, Sommer F, McGrath L, Härtl R. Fundamentals of intervertebral disc degeneration. *World Neurosurg.* 2022;157:264–73.
6. Xin J, Wang Y, Zheng Z, Wang S, Na S, Zhang S. Treatment of intervertebral disc degeneration. *Orthop Surg.* 2022;14(7):1271–80.
7. Jiang Y, Yu L, Luo X, Lin Y, He B, Wu B, Qu J, Wu T, Pu-Yeh W, Zhang C, et al. Quantitative synthetic MRI for evaluation of the lumbar intervertebral disk degeneration in patients with chronic low back pain. *Eur J Radiol.* 2020;124:108858.
8. Pfirrmann C, Metzendorf A, Zanetti M, Hodler J, Boos N. Magnetic resonance classification of lumbar intervertebral disc degeneration. *Spine.* 2001;26(17):1873–8.
9. Feng Y, Egan B, Wang J. Genetic factors in intervertebral disc degeneration. *Genes & Diseases.* 2016;3(3):178–85.
10. Baues M, Dasgupta A, Ehling J, Prakash J, Boor P, Tacke F, Kiessling F, Lammers T. Fibrosis imaging: current concepts and future directions. *Adv Drug Deliv Rev.* 2017;121:9–26.
11. Yee A, Lam M, Tam V, Chan W, Chu I, Cheah K, Cheung K, Chan D. Fibrotic-like changes in degenerate human intervertebral discs revealed by quantitative proteomic analysis. *Osteoarthritis Cartil.* 2016;24(3):503–13.
12. Oegema T, Johnson S, Aguiar D, Ogilvie J. Fibronectin and its fragments increase with degeneration in the human intervertebral disc. *Spine.* 2000;25(21):2742–7.
13. Feng G, Zhang Z, Dang M, Zhang X, Doleyres Y, Song Y, Chen D, Ma P. Injectable nanofibrous spongy microspheres for NR4A1 plasmid DNA transfection to reverse fibrotic degeneration and support disc regeneration. *Biomaterials.* 2017;131:86–97.
14. Lupher M, Gallatin W. Regulation of fibrosis by the immune system. *Adv Immunol.* 2006;89:245–88.
15. Prakash J, Pinzani M. Fibroblasts and extracellular matrix: Targeting and therapeutic tools in fibrosis and cancer. *Adv Drug Deliv Rev.* 2017;121:1–2.
16. Rajendran R, Sukumaran A. Editorial. Oral submucous fibrosis: revised hypotheses as to its cause. *J Contemp Dent Pract.* 2013;14(5):i–iii.
17. Au T, Lam T, Peng Y, Wynn S, Cheung K, Cheah K, Leung V. Transformation of resident notochord-descendent nucleus pulposus cells in mouse injury-induced fibrotic intervertebral discs. *Aging Cell.* 2020;19(11):e13254.
18. Franco C, Patricia H, Timo S, Claudia B, Marcela H. Matrix metalloproteinases as regulators of Periodontal inflammation. *Int J Mol Sci* 2017, 18(2).
19. Kanerva A, Kommonen B, Grönblad M, Tolonen J, Habtemariam A, Virri J, Karaharju E. Inflammatory cells in experimental intervertebral disc injury. *Spine.* 1997;22(23):2711–5.
20. Nakazawa K, Walter B, Laudier D, Krishnamoorthy D, Mosley G, Spiller K, latridis J. Accumulation and localization of macrophage phenotypes with human intervertebral disc degeneration. *The Spine Journal: Official Journal of the North American Spine Society.* 2018;18(2):343–56.
21. Lv F, Peng Y, Lim F, Sun Y, Lv M, Zhou L, Wang H, Zheng Z, Cheung K, Leung V. Matrix metalloproteinase 12 is an indicator of intervertebral disc degeneration co-expressed with fibrotic markers. *Osteoarthritis Cartil.* 2016;24(10):1826–36.
22. Vasiliadis E, Kaspiris A, Grivas T, Khaldi L, Lamprou M, Pneumaticos S, Nikolopoulos K, Korres D, Papadimitriou E. Expression of macrophage elastase (MMP12) in rat tail intervertebral disc and growth plate after asymmetric loading. *Bone & Joint Research.* 2014;3(9):273–9.
23. Smyth G. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol.* 2004;3:Article3.

24. Yu G, Wang L, Han Y, He Q. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012;16(5):284–7.
25. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000;28(1):27–30.
26. Kanehisa M. Toward understanding the origin and evolution of cellular organisms. *Protein Science: A Publication of the Protein Society*. 2019;28(11):1947–51.
27. Kanehisa M, Furumichi M, Sato Y, Kawashima M, Ishiguro-Watanabe M. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res*. 2023;51:D587–92.
28. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized Linear models via Coordinate Descent. *J Stat Softw*. 2010;33(1):1–22.
29. Gu W, Ming T, Xie Z. Developing a Genetic Biomarker-based Diagnostic Model for Major Depressive Disorder using Random Forests and Artificial Neural Networks. *Combinatorial chemistry & high throughput screening* 2022.
30. Saberi M, Zhang X, Mobasheri A. Targeting mitochondrial dysfunction with small molecules in intervertebral disc aging and degeneration. *Geroscience*. 2021;43(2):517–37.
31. Wu P, Kim H, Jang I. Intervertebral disc Diseases PART 2: a review of the current diagnostic and treatment strategies for intervertebral disc Disease. *Int J Mol Sci* 2020, 21(6).
32. Dowdell J, Erwin M, Choma T, Vaccaro A, Iatridis J, Cho S. Intervertebral disk degeneration and repair. *Neurosurgery*. 2017;80:46–554.
33. Mahjoub M, Xie Z, Stearns T. Cep120 is asymmetrically localized to the daughter centriole and is essential for centriole assembly. *J Cell Biol*. 2010;191(2):331–46.
34. Wu C, Yang M, Li J, Wang C, Cao T, Tao K, Wang B. Talpid3-binding centrosomal protein Cep120 is required for centriole duplication and proliferation of cerebellar granule neuron progenitors. *PLoS ONE*. 2014;9(9):e107943.
35. Comartin D, Gupta G, Fussner E, Coyaud É, Hasegan M, Archinti M, Cheung S, Pinchev D, Lawo S, Raught B, et al. CEP120 and SPICE1 cooperate with CPAP in centriole elongation. *Curr Biology: CB*. 2013;23(14):1360–6.
36. Lin Y, Wu C, Lin Y, Hsu W, Tang C, Chang C, Tang T. CEP120 interacts with CPAP and positively regulates centriole elongation. *J Cell Biol*. 2013;202(2):211–9.
37. Xie Z, Moy L, Sanada K, Zhou Y, Buchman J, Tsai L. Cep120 and TACCs control interkinetic nuclear migration and the neural progenitor pool. *Neuron*. 2007;56(1):79–93.
38. Sharma A, Gerard S, Olieric N, Steinmetz M. Cep120 promotes microtubule formation through a unique tubulin binding C2 domain. *J Struct Biol*. 2018;203(1):62–70.
39. Webley K, Bond J, Jones C, Blaydes J, Craig A, Hupp T, Wynford-Thomas D. Posttranslational modifications of p53 in replicative senescence overlapping but distinct from those induced by DNA damage. *Mol Cell Biol*. 2000;20(8):2803–8.
40. Bibby S, Urban J. Effect of nutrient deprivation on the viability of intervertebral disc cells. *Eur Spine Journal: Official Publication Eur Spine Soc Eur Spinal Deformity Soc Eur Sect Cerv Spine Res Soc*. 2004;13(8):695–701.
41. Wang Y, Che M, Xin J, Zheng Z, Li J, Zhang S. The role of IL-1 $\beta$  and TNF- $\alpha$  in intervertebral disc degeneration. *Biomed Pharmacotherapy = Biomedecine Pharmacotherapie*. 2020;131:110660.
42. Jeong S, Lee J, Kim K. In vitro lifespan and senescence mechanisms of human nucleus pulposus chondrocytes. *The Spine Journal: Official Journal of the North American Spine Society*. 2014;14(3):499–504.
43. di d'Adda F, Teo S, Jackson S. Functional links between telomeres and proteins of the DNA-damage response. *Genes Dev*. 2004;18(15):1781–99.
44. Dimozi A, Mavrogonatou E, Sklirou A, Kletsas D. Oxidative stress inhibits the proliferation, induces premature senescence and promotes a catabolic phenotype in human nucleus pulposus intervertebral disc cells. *Eur Cells Mater*. 2015;30:89–102. discussion 103.
45. Le Maitre C, Hoyland J, Freemont A. Catabolic cytokine expression in degenerate and herniated human intervertebral discs: IL-1 $\beta$  and TNF $\alpha$  expression profile. *Arthritis Res Therapy*. 2007;9(4):R77.
46. Shamji M, Setton L, Jarvis W, So S, Chen J, Jing L, Bullock R, Isaacs R, Brown C, Richardson W. Proinflammatory cytokine expression profile in degenerate and herniated human intervertebral disc tissues. *Arthritis Rheum*. 2010;62(7):1974–82.
47. Le Maitre C, Freemont A, Hoyland J. The role of interleukin-1 in the pathogenesis of human intervertebral disc degeneration. *Arthritis Res Therapy*. 2005;7(4):R732–745.
48. Séguin C, Pilliar R, Roughley P, Kandel R. Tumor necrosis factor-alpha modulates matrix production and catabolism in nucleus pulposus tissue. *Spine*. 2005;30(17):1940–8.
49. Takeuchi M, Doi T, Obayashi K, Hirai A, Yoneda K, Tanaka F, Iwai Y. Soluble PD-L1 with PD-1-binding capacity exists in the plasma of patients with non-small cell Lung cancer. *Immunol Lett*. 2018;196:155–60.
50. Liang Z, Tian Y, Cai W, Weng Z, Li Y, Zhang H, Bao Y, Li Y. High-affinity human PD-L1 variants attenuate the suppression of T cell activation. *Oncotarget*. 2017;8(51):88360–75.
51. Dezutter-Dambuyant C, Durand I, Alberti L, Bendriss-Vermare N, Valladeau-Guilemond J, Duc A, Magron A, Morel A, Sisirak V, Rodriguez C, et al. A novel regulation of PD-1 ligands on mesenchymal stromal cells through MMP-mediated proteolytic cleavage. *Oncoimmunology*. 2016;5(3):e1091146.
52. Ng K, Attig J, Young G, Ottina E, Papamichos S, Kotsianidis I, Kassiotis G. Soluble PD-L1 generated by endogenous retroelement exaptation is a receptor antagonist. *eLife* 2019, 8.
53. Mahoney K, Shukla S, Patsoukis N, Chaudhri A, Browne E, Arazi A, Eisenhaure T, Pendergraft W, Hua P, Pham H, et al. A secreted PD-L1 splice variant that covalently dimerizes and mediates immunosuppression. *Cancer Immunol Immunotherapy: CII*. 2019;68(3):421–32.
54. Tannig P, Peter A, Lapuente D, Klessing S, Damm D, Tenbusch M, Überla K, Temchura V. Modulation of Vaccine-Induced HIV-1-Specific Immune responses by Co-electroporation of PD-L1 Encoding DNA. *Vaccines* 2020, 8(1).
55. Orme J, Jazieh K, Xie T, Harrington S, Liu X, Ball M, Madden B, Charlesworth M, Azam T, Lucien F, et al. ADAM10 and ADAM17 cleave PD-L1 to mediate PD-(L)1 inhibitor resistance. *Oncoimmunology*. 2020;9(1):1744980.
56. Larrinaga G, Solano-Iturri J, Errarte P, Unda M, Loizaga-Iriarte A, Pérez-Fernández A, Echevarría E, Asumendi A, Manini C, Angulo J et al. Soluble PD-L1 is an Independent prognostic factor in Clear Cell Renal Cell Carcinoma. *Cancers* 2021, 13(4).
57. Murakami S, Shibaki R, Matsumoto Y, Yoshida T, Goto Y, Kanda S, Horinouchi H, Fujiwara Y, Yamamoto N, Ohe Y. Association between serum level soluble programmed cell death ligand 1 and prognosis in patients with non-small cell Lung cancer treated with anti-PD-1 antibody. *Thorac cancer*. 2020;11(12):3585–95.
58. Zhang Y, Chee A, Shi P, Adams S, Markova D, Anderson D, Smith H, Deng Y, Plastaras C, An H. Intervertebral disc cells produce interleukins found in patients with Back Pain. *Am J Phys Med Rehabil*. 2016;95(6):407–15.
59. He D, Zhou M, Bai Z, Wen Y, Shen J, Hu Z. Propionibacterium acnes induces intervertebral disc degeneration by promoting nucleus pulposus cell pyroptosis via NLRP3-dependent pathway. *Biochem Biophys Res Commun*. 2020;526(3):772–9.
60. Murai K, Sakai D, Nakamura Y, Nakai T, Igarashi T, Seo N, Murakami T, Kobayashi E, Mochida J. Primary immune system responders to nucleus pulposus cells: evidence for immune response in disc herniation. *Eur Cells Mater*. 2010;19:13–21.
61. Beaulieu A. Memory responses by natural killer cells. *J Leukoc Biol*. 2018;104(6):1087–96.
62. Salminen A. Increased immunosuppression impairs tissue homeostasis with aging and age-related Diseases. *J Mol Med*. 2021;99(1):1–20.
63. Huaux F. Interpreting Immunoregulation in Lung Fibrosis: a New Branch of the Immune Model. *Front Immunol*. 2021;12:690375.
64. Feng C, Liu H, Yang M, Zhang Y, Huang B, Zhou Y. Disc cell senescence in intervertebral disc degeneration: causes and molecular pathways. *Cell Cycle (Georgetown Tex)*. 2016;15(13):1674–84.
65. Gire V, Roux P, Wynford-Thomas D, Brondello JM, Dulic V. DNA damage checkpoint kinase Chk2 triggers replicative senescence. *EMBO J*. 2004;23(13):2554–63.
66. Gire V, Roux P, Wynford-Thomas D, Brondello J, Dulic V. DNA damage checkpoint kinase Chk2 triggers replicative senescence. *EMBO J*. 2004;23(13):2554–63.
67. Herbig U, Jobling W, Chen B, Chen D, Sedivy J. Telomere shortening triggers senescence of human cells through a pathway involving ATM, p53, and p21(CIP1), but not p16(INK4a). *Mol Cell*. 2004;14(4):501–13.
68. Kim K, Chung H, Ha K, Lee J, Kim Y. Senescence mechanisms of nucleus pulposus chondrocytes in human intervertebral discs. *The Spine Journal: Official Journal of the North American Spine Society*. 2009;9(8):658–66.
69. Gruber H, Watts J, Hoelscher G, Bethea S, Ingram J, Zinchenko N, Hanley E. Mitochondrial gene expression in the human annulus: in vivo data from annulus cells and selectively harvested senescent annulus cells. *The Spine Journal: Official Journal of the North American Spine Society*. 2011;11(8):782–91.

70. Ressler S, Bartkova J, Niederegger H, Bartek J, Scharffetter-Kochanek K, Jansen-Durr P, Wlaschek M. p16INK4A is a robust in vivo biomarker of cellular aging in human skin. *Aging Cell*. 2006;5(5):379–89.
71. Park J-S, Park J-B, Park I-J, Park E-Y. Accelerated premature stress-induced senescence of young annulus fibrosus cells of rats by high glucose-induced oxidative stress. *Int Orthop*. 2014;38(6):1311–20.
72. Wang S, Yang P, Feng X, Zhu Y, Qiu F, Hu X, Zhang S. Apigenin inhibits the growth of Hepatocellular Carcinoma Cells by affecting the expression of microRNA transcriptome. *Front Oncol*. 2021;11:657665.
73. Xin C, Lu S, Li Y, Zhang Y, Tian J, Zhang S, Yang S, Gao T, Xu J. Mir-671-5p inhibits Tumor Proliferation by blocking cell cycle in Osteosarcoma. *DNA Cell Biol*. 2019;38(9):996–1004.
74. Jahangiri Moez M, Bjeije H, Soltani B. Hsa-Mir-5195-3P induces downregulation of TGF $\beta$ R1, TGF $\beta$ R2, SMAD3 and SMAD4 supporting its Tumor suppressive activity in HCT116 cells. *Int J Biochem Cell Biol*. 2019;109:1–7.
75. Che H, Li J, Li Y, Ma C, Liu H, Qin J, Dong J, Zhang Z, Xian C, Miao D et al. p16 deficiency attenuates intervertebral disc degeneration by adjusting oxidative stress and nucleus pulposus cell cycle. *eLife* 2020, 9.
76. Li S, Hua W, Wang K, Gao Y, Chen S, Liu W, Song Y, Wu X, Tu J, Kang L, et al. Autophagy attenuates compression-induced apoptosis of human nucleus pulposus cells via MEK/ERK/NRF1/Atg7 signaling pathways during intervertebral disc degeneration. *Exp Cell Res*. 2018;370(1):87–97.
77. Bai X, Jiang M, Wang J, Yang S, Liu Z, Zhang H, Zhu X. In vitro Cyanidin attenuates the apoptosis of rat nucleus pulposus cells and the degeneration of intervertebral disc via the JAK2/STAT3 signal pathway and. *Pharm Biol*. 2022;60(1):427–36.
78. Huo Y, Yang D, Lai K, Tu J, Zhu Y, Ding W, Yang S. Antioxidant effects of Resveratrol in Intervertebral Disk. *J Invest Surgery: Official J Acad Surg Res*. 2022;35(5):1135–44.
79. Hu H, Li L, Liu Y, Wang S, Xie S, Sun J. [Effect of resveratrol on high mobility group box-1 protein signaling pathway in cartilage endplate degeneration caused by inflammation]. *Zhongguo Xiu Fu Chong Jian Wai Ke Za Zhi = Zhongguo Xiu fu chongjian waikexue zazhi = Chinese. J Reparative Reconstr Surg*. 2022;36(4):461–9.
80. Liu M, Zhang L, Zang W, Zhang K, Li H, Gao Y. Pharmacological effects of Resveratrol in Intervertebral Disc Degeneration: A literature review. *Orthop Surg*. 2022;14(12):3141–9.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.