Non-coding RNA Research 10 (2025) 177–191

Contents lists available at ScienceDirect

Non-coding RNA Research

journal homepage: www.keaipublishing.com/en/journals/non-coding-rna-research

Original Research Article

Study on molecular mechanism of intervertebral disc degeneration by single cell hdWGCNA combined with transcriptome sequencing

Xuan Zhao [a,b,1], Qijun Wang [a,b,1], Wei Wang [a,b,*], Xiaolong Chen [a,b,**], Shibao Lu [a,b,***]

[a] Department of Orthopedics, Xuanwu Hospital, Capital Medical University, Beijing, China
[b] National Clinical Research Center for Geriatric Diseases, Beijing, China

ARTICLE INFO

Keywords:
Intervertebral disc degeneration
Single cell analysis
hdWGCNA
Machine learning
Diagnostic markers

ABSTRACT

Background: Intervertebral disc degeneration (IVDD) is one of the important causes of lower back pain, seriously affecting people's health and quality of life. This research employs single-cell analysis to identify the specific cellular subtypes and key regulatory genes associated with IVDD.
Methods: We analyzed the single-cell data and screened cells that closely associated with the development of IVDD. The differential expression of feature genes between IVDD and control groups was analyzed. Additionally, drugs and regulatory transcription factors that interact with feature genes were predicted and clinically validated by reverse transcription quantitative real-time PCR (RT-qPCR), immunohistochemistry (IHC), and enzyme-linked immunosorbent assay (ELISA).
Results: Our study identified the Chond2 cell subtype associated with IVDD and selected four feature genes influencing the development of IVDD, namely IGFBP3, ACAN, VAPA and TMEM45A, through the high-dimensional weighted gene co-expression network analysis (hdWGCNA) analysis, least absolute shrinkage and selection operator (LASSO), and random forest (RF). Besides, compared to the MDD group, IGFBP3 and TMEM45A were significantly upregulated in the SDD group, while ACAN and VAPA showed no significant difference between the two groups. ELISA testing revealed a positive correlation between IGFBP3 concentration and the grading of IVDD. Furthermore, Celecoxib may be used to treat IVDD by inhibiting IGFBP3.
Conclusion: Our study identified the Chond2 cell subtype associated with IVDD and selected four feature genes influencing the development of IVDD, namely IGFBP3, ACAN, VAPA and TMEM45A. Our findings establish a robust theoretical foundation for the clinical diagnosis and treatment of IVDD patients.

1. Introduction

Intervertebral disc degeneration (IVDD) is a part of the body's degenerative process [1]. IVDD can lead to restricted lumbar spine mobility, making movements such as bending and turning difficult. It can also cause compression of the sciatic nerve, resulting in pain, numbness, or a tingling sensation in the lower limbs [2]. Severe IVDD can lead to muscle weakness, affecting daily activities. Currently, spinal pain is largely associated with IVDD. Treatments for IVDD include medication and surgical intervention, but almost all treatments aim to alleviate patient symptoms [3]. There is little exploration of suitable treatment methods from the perspective of altering the biological imbalance that has already occurred in intervertebral disc tissue. Research has shown that interventions at the genetic molecular level and cell or tissue transplantation hold great potential for the treatment of IVDD, providing significant room for choice. Therefore, it is crucial to uncover the key genes that influence the development of IVDD.

As transcriptome sequencing continues to advance, an escalating number of studies are uncovering genes closely linked to the progression of IVDD. For example, Xiang Q had identified ferroptosis-related genes such as HMOX1, MAPK1, and TXNRD1 as vital biomarkers for diagnosing IVDD [4]. Jiang X et al. identified that lncRNA GNAS-AS1 and

* Corresponding author. Department of Orthopedics, Xuanwu Hospital, Capital Medical University, Beijing, China.
** Corresponding author. Department of Orthopedics, Xuanwu Hospital, Capital Medical University, Beijing, China.
*** Corresponding author. Department of Orthopedics, Xuanwu Hospital, Capital Medical University, Beijing, China.
E-mail addresses: wangwei37@buaa.edu.cn (W. Wang), chensmalldragon@163.com (X. Chen), spinelu@163.com (S. Lu).
[1] These authors are co-first authors and contributed equally to this work.

https://doi.org/10.1016/j.ncrna.2024.09.003
Received 6 July 2024; Received in revised form 21 August 2024; Accepted 1 September 2024
Available online 2 September 2024

MIR100HG, which were associated with osteosarcoma, play a pivotal role in the diagnosis of patients with IVDD [5]. Xu S et al. identified 20 genes associated with aging with diagnostic value in patients with IVDD [6]. Based on the results of previous studies, we have found that transcriptomics can help identify biomarkers associated with specific diseases. These biomarkers can be used for early diagnosis, disease typing, and prognosis evaluation. Additionally, we can attain a deeper comprehension of the disease's pathogenesis by analyzing the disparities in gene expression between diseased and normal samples. Traditional method of gene expression analysis typically involves sequencing a mixture of a large number of cells, which can mask the differences between individual cells. Conversely, single-cell RNA sequencing (scRNA-seq) allows for comprehensive sequencing and analysis of RNA within individual cells, providing the full transcriptome of each cell. This technology helps us understand the differences in gene expression levels among cells, thereby revealing cellular diversity and functionality. Cherif H et al. revealed cell phenotypes and biomarkers associated with IVDD degeneration using single-cell sequencing analysis, including C2orf40, CD44, MT2A [7]. Zhang Y et al. discovered new subtypes of chondrocytes in the nucleus pulposus of IVDD through scRNA-Seq, providing a theoretical basis for a deeper understanding of the pathological changes in IVDD development [8]. In addition, research combining proteomic sequencing, RNA sequencing, and scRNA-seq identified SERPINA1 as a biomarker for predicting the progression of IVDD [9]. Wang D et al. used scRNA-seq to reveal interactions between major subgroups and changes in the microenvironment during IVDD. The above studies collectively demonstrate the key role of scRNA-seq in the diagnosis and treatment of IVDD. In addition, high-dimensional weighted gene co-expression network analysis (hdWGCNA) is an analysis method based on weighted gene co-expression networks. It can reveal the correlation between genes and discover changes in expression patterns under different biological conditions. By clustering genes into different modules and studying the correlation of these modules with specific biological features. Compared to WGCNA, hdWGCNA introduces hierarchical decomposition, enabling the unveiling of potential biological mechanisms and signaling pathways in complex biological systems. Therefore, fully utilizing this technology to study the mechanisms of IVDD development and potential therapeutic targets can enhance our understanding of IVDD.

In this study, we conducted in-depth analysis of single-cell expression profiles in IVDD, identifying cell subtypes and biomarkers closely associated with the development of IVDD. Our research findings provide important support for early diagnosis of IVDD patients, elucidating disease mechanisms, and advancing personalized medicine.

2. Materials and methods

2.1. Data collection

In this study, the GSE199866 dataset including the scRNA-seq data of four samples and GSE70362 dataset including the RNA-seq data of 16 control samples and 32 disease samples) were downloaded from Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/) database. The samples in the GSE199866 dataset were nucleus pulposus (NP) and annulus fibrosis (AF) cells isolated from non-degenerating (nD) and degenerating (D) discs of the same individual, named NPD, NPH, AFD, and AFH.

2.2. Acquisition of human NP tissues

This study involving human samples was supervised and approved by the Ethics Committee of Capital Medical University Xuanwu Hospital (Ethic number: 2024010-002). NP tissue samples were divided into mild (grades I-II) and severe (grades III-V) degenerative NP tissues according to the Pfirrmann grading system. A total of seven NP tissues from mild degenerative intervertebral disc (MDD) and fifteen NP tissues from severe degenerative intervertebral disc (SDD) were collected for analysis. The clinical information of patients was listed in additional file: Table S1. MDD tissues (Pfirrmann grade II) were obtained from patients undergoing surgery for spinal scoliosis, while the SDD NP tissues (Pfirrmann grade III-IV) were collected from IVDD patients with lumbar disc herniation, lumbar spinal stenosis or lumbar spondylolisthesis who underwent discectomy. The informed consent forms were duly signed by all participants, and none of the enrolled subjects had a medical history involving radiotherapy or chemotherapy, nor any prior surgical interventions.

2.3. Single-cell analysis

Firstly, the low-expression genes in expression profile were filtered out (nFeature_RNA >200 & nFeature_RNA <6000 & percent.mt < 5) by "Seurat" package. Subsequently, the data were sequentially subjected to normalization, standardization, principal component analysis (PCA), and uniform manifold approximation and projection (UMAP) processing. The Elbow Plot was employed to determine the optimal number of principal components (PCs). The location relationship between clusters was determined by UMAP analysis. Finally, cells related to IVDD were selected based on the annotation results using "celldex" package.

2.4. Cellchat analysis

We utilized standardized single-cell expression profile as input data, where the cell subtypes derived from single-cell analysis represent cellular information. The analysis focused on cell-related interactions and quantified closeness of interaction relationships based on strength and frequency of interactions between cells.

2.5. hdWGCNA analysis

SetupForWGCNA function was used to construct the co-expression network of cell expression genes, and soft threshold was set at seven. WGCNA dendrogram was visualized using PlotDendrogram, and then the level of module eigengenes (MEs) for each module was visualized through GetMEs function.

2.6. Identification of feature genes by machine learning algorithms

The least absolute shrinkage and selection operator (LASSO) and random forest (RF) algorithms were utilized to select feature genes for IVDD. "glmnet" package was employed to implement LASSO analysis. Feature importance was assessed using the RF algorithm, and the importance of features was evaluated based on the percent increase in MSE (%IncMSE). The top 10 features were selected for subsequent analysis.

2.7. Immunoinfiltration analysis

The differential expression of 29 immune cells between control and disease groups were analyzed. The expression profiles of immune cells were quantified using ssGSEA algorithm to analyze differential expression. The relationship between gene expression levels and immune cell abundance were investigated through Spearman correlation analysis.

2.8. Gene set enrichment analysis (GSEA) analysis

In order to further elucidate the relevant mechanism of the function of feature genes, we used the "GSEA" package in R software to conduct GSEA enrichment analysis. MsigDB database version 7.0 provided the background gene set, and annotated gene sets were used to annotate subtype pathways for difference analysis of pathways.

2.9. Gene set variation analysis (GSVA) analysis

GSVA scored gene sets of interest comprehensively, transforming gene-level changes into pathway-level changes, thereby assessing biological functions of samples. The gene set downloaded from the Molecular Signatures Database was scored using "GSVA" package to analyze potential biological functional differences among the samples.

2.10. Regulatory network of feature genes

The transcriptional regulatory factors of key genes were predicted using "RcisTarget" package. All computations performed by RcisTarget were based on motifs. The normalized enrichment score (NES) of motifs depended on the total number of motifs in the database. In addition to the motifs annotated by the source data, this study inferred further annotation files based on motif similarity and gene sequences. The NES for each motif is calculated based on the area under the curve (AUC) distribution of all motifs in the gene set.

2.11. RNA isolation and reverse transcription quantitative real-time PCR (RT-qPCR)

NP tissues from patients were extracted and homogenized using a homogenizer and plastic pestle on ice for 10 to 15 strokes. Total RNA was isolated from NP tissues using Trizol reagent (Takara, Dalian, China) and then converted to cDNA. Next, RT-PCR was conducted using AceQ qPCR SYBR Green Master Mix (Q111-02, Vazyme, China) in a 7500 real-time PCR machine (Applied Biosystems, Inc., Carlsbad, CA, USA) in accordance with the manufacturer's instructions. The sequences of the qPCR primers are provided in additional file: Table S2. GAPDH is used as an internal reference gene, and the log(2^-(ΔΔCt)) method is employed as a statistical analysis approach.

2.12. Immunohistochemistry (IHC)

For IHC, NP sections were prepared as described previously. Briefly, sections were blocked by incubating with 10 % bovine serum albumin (BSA), followed by incubation with primary antibodies against IGFBP3 (1:100) (10189-2-AP, Proteintech, China). Subsequently, the sections were thoroughly washed and incubated with the corresponding secondary antibody conjugated to HRP at a dilution of 1:300. Moreover, protein visualization was achieved using a DAB staining kit (8059, Cell Signaling Technology, USA), followed by counterstaining with hematoxylin. Finally, image acquisition was performed utilizing a microscope (Axio Lab.A1, Zeiss, Heidenheim, Germany).

2.13. Enzyme-linked immunosorbent assay (ELISA)

IGFBP3 in NP tissues from patients was quantified using ELISA kit (Human IGFBP-3 ELISA Kit, MultiScience, China). A 1x PBS solution with 1 % phenylmethylsulfonyl fluoride (PMSF) was prepared, and then 5 times the weight of the sample was added to prechilled PBS. Homogenization was performed using an ultrasonic crusher. The tissue lysates were centrifuged at 12,000 rpm for 10 min. The supernatant was detected by ELISA analysis. Values from each sample were normalized to the total protein content determined using a BCA assay (23225, Thermo Fisher Scientific, USA).

2.14. Drug prediction of feature genes

The DGIdb database is a database for drug-gene interactions and genomic information. We utilized this database to predict potential drugs or small molecular compounds that interact with feature genes.

2.15. Statistical analysis

All statistical analyses were conducted using R language. The comparison between two groups was conducted using Wilcoxon.test. For clinical validation results, statistical significance was determined using an unpaired Student's t-test between the two groups. $P < 0.05$ was considered statistically significant.

3. Results

3.1. Identification of key cell subtypes in IVDD

The quality control of single-cell expression profile data was conducted and we retained 13,899 cells from four IVDD-related scRNA-seq samples for further analysis. Furthermore, batch effects between samples were identified through PCA dimensionality reduction analysis. Then, we further utilized Harmony analysis to reduce dimensionality and correct for batch effects, as shown in Figure S1 and Figure S2. Subsequently, 13,899 cells were clustered into 11 clusters using the "Seurat" package (Fig. 1A). According to the annotation of clusters based on the literature and CellMarker database [10–12], 11 clusters were annotated into seven cell types: Chond1, Chond2, Chond3, CPC, fibroNPC, HomC, and NPPC, as shown in Fig. 1B. Fig. 1C showed the 10 genes with the highest standard deviation, namely MIA, ACAN, COL3A1, COL1A1, BIRC5, CDK1, STMN1, JUN, BRD2, PDGFRA and PRRX1. Fig. S3A and B also displayed the normalized expression levels of representative marker genes for each subtype. The gene expression characteristics of each cell type were listed in Table S3. We also analyzed the proportion of each cell type in the disease samples (AFD, NPD) compared with the control samples (AFH, NPH) and found that Chond2 proportion in disease samples was notably higher than that in control samples (Fig. 1D). Through Gene Ontology (GO) analysis, we found that the Chond2 subtype was significantly enriched in related biological functions such as hypoxia, apoptosis, glycolysis and inflammatory response (Fig. S3C and Table S4).

3.2. Annotation of clustered subtypes and analysis of receptor-ligand pairs

Considering higher proportion of the Chond2 subtype in the disease samples, we further subset the Chond2 subtype and sequentially analyzed it through PCA, Harmony, and FindClusters to obtain 13 clusters, as shown in Fig. 2A. There were significant differences in type and number of cluster cells between disease and control samples. Among them, the proportion of clusters C0, C2 and C4 in disease samples was higher than that in control samples (Fig. 2B). Subsequently, to further analyze the communication network between different clustered cells, we employed CellChat analysis. The results revealed complex interactive relationships among these cell subtypes and our research results indicated an enhanced communication between CPC, Chond2, and NPCC (Fig. 2C–D). Subsequently, we analyzed the single-cell trajectory of Chond2. From the graph, we observed that Chond2 transitioned from cluster 5 in the early stage of differentiation, and then successively differentiated into cluster 2, cluster 6, and cluster 4 (Fig. 2E–G).

3.3. Identification of module genes in Chond2 subtype by hdWGCNA

To determine the gene co-expression network in Chond2 subtype and biomarkers in disease development, hdWGCNA analysis was performed. Firstly, the soft threshold value of seven was determined by the function of "TestSoftPowers" (Fig. 3A). A total of five gene modules were detected, including yellow, turquoise, brown, green, and blue modules (Fig. 3B). The dendrogram of WGCNA in Fig. 3C illustrated the hierarchical clustering of genes based on their expression patterns. Furthermore, through analysis of the relationship between modules and MEs levels, we found that MEs levels of the turquoise, blue, and green

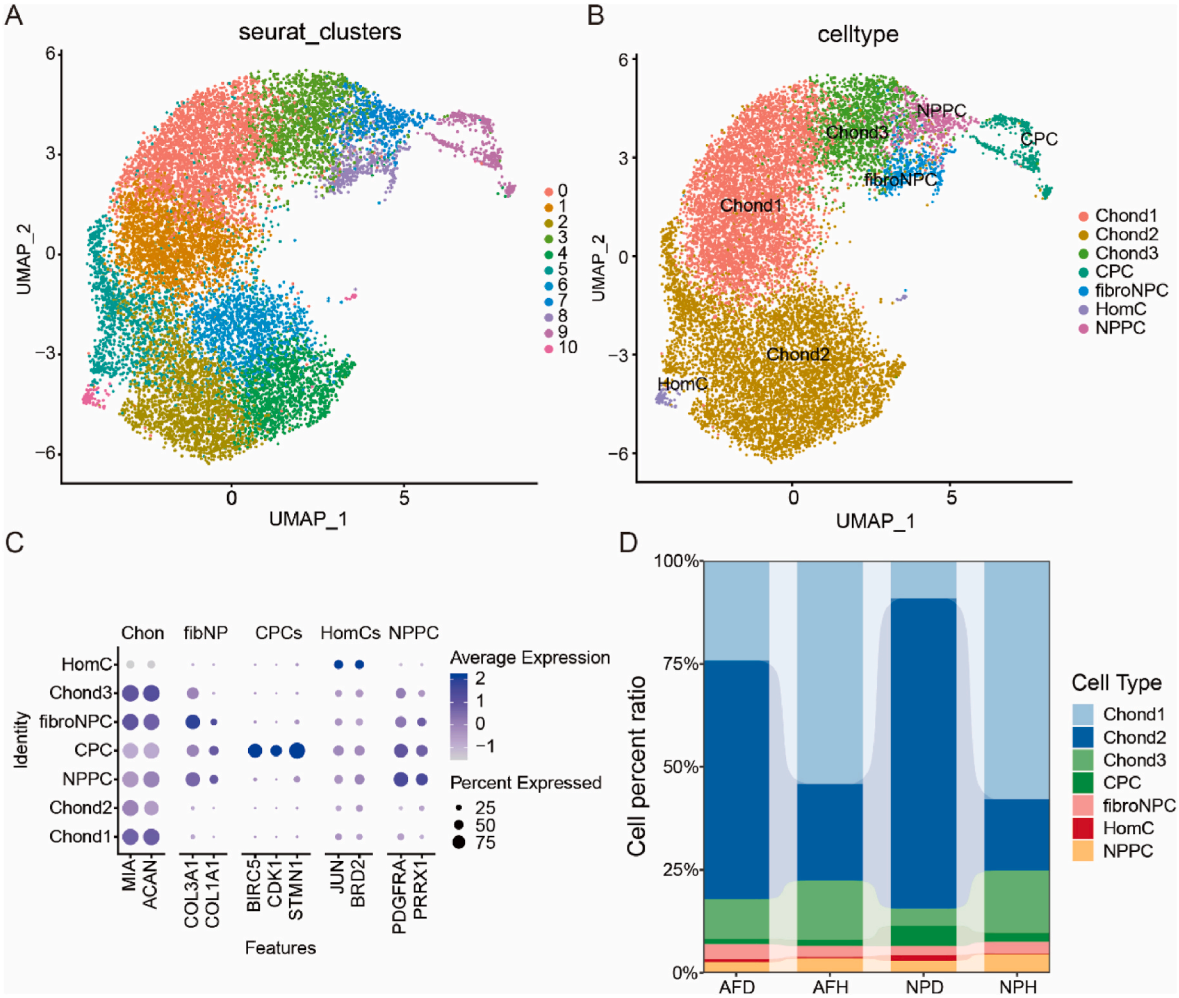

Fig. 1. Cell annotation processing and identification of key cell subtypes in IVDD.
(A) The cells were divided into 11 clusters by UMAP algorithm; (B) Cell annotation results for 11 clusters; (C) Expression of characteristic genes of different cell clusters in 7 cell types; (D) The difference of the proportion of 7 kinds of cells in different samples. AF: annulus fibrosus; NP: nucleus pulposus.

modules were elevated in the disease cell subtypes. Therefore, we selected the top 25 genes based on k-Module Membership (kME) values from these 3 module genes for further investigation (Fig. 3D–E).

3.4. Screening feature genes of IVDD

To further identify key genes influencing IVDD, we utilized the module genes with the top 25 kME values from the turquoise, blue, and green modules obtained in the previous step, and employed LASSO regression and RF algorithms to screen for feature genes associated with IVDD. The LASSO identified a total of 11 genes as feature genes for IVDD, as shown in Fig. 4A–B. Besides, we used RF screening to select the top 10 ranked genes as feature genes for IVDD (Fig. 4C). We took the intersection of these genes with the feature genes identified by LASSO, resulting in a total of 4 intersecting genes, containing IGFBP3, ACAN, VAPA, and TMEM45A (Fig. 4D). Then, the expression of these four key intersecting genes was examined in 7 cell subtypes. The results indicated that IGFBP3, VAPA, and TMEM45A exhibited higher expression levels in Chond2, while ACAN showed relatively higher expression in Chond3 (Fig. 4E–G). Together, our research findings suggested that the identified four genes might serve as potential predictive factors for IVDD and could potentially impact personalized treatment strategies.

3.5. Immune microenvironment analysis

This study analyzed differential expression of feature genes based on GSE70362 dataset. Fig. 5A showed that IGFBP3, ACAN, and TMEM45A exhibited significant differential expression between disease and control groups ($p < 0.001$). The microenvironment significantly influences the sensitivity of disease diagnosis. Therefore, we further explored potential molecular mechanisms through which feature genes influence the progression of IVDD by analyzing relationship between feature genes of IVDD and immune infiltration. Firstly, we analyzed the proportion of immune cell content in each patient and correlation between immune cells, as shown in Fig. 5B–C. Furthermore, the research results indicated that compared to control patients, patients in the disease group showed a significant increase in the proportions of APC co-stimulation, Mast cells, MHC class I, parainflammation, Th2 cells and Type I IFN Response, etc. (Fig. 5D). We explored relationship between feature genes and immune cells, and found that IGFBP3 was positively correlated with Type II IFN Response, MHC class I, and parainflammation, while significantly negatively correlated with B cells and Th1 cells. ACAN was significantly positively correlated with B cells and T cell co-inhibition, and significantly negatively correlated with Type II IFN Response, parainflammation, and MHC class I. VAPA was significantly positively correlated with Th1 cells and pDCs, and significantly negatively correlated with Type II IFN Response and parainflammation. TMEM45A was significantly positively correlated with Type I IFN Response,

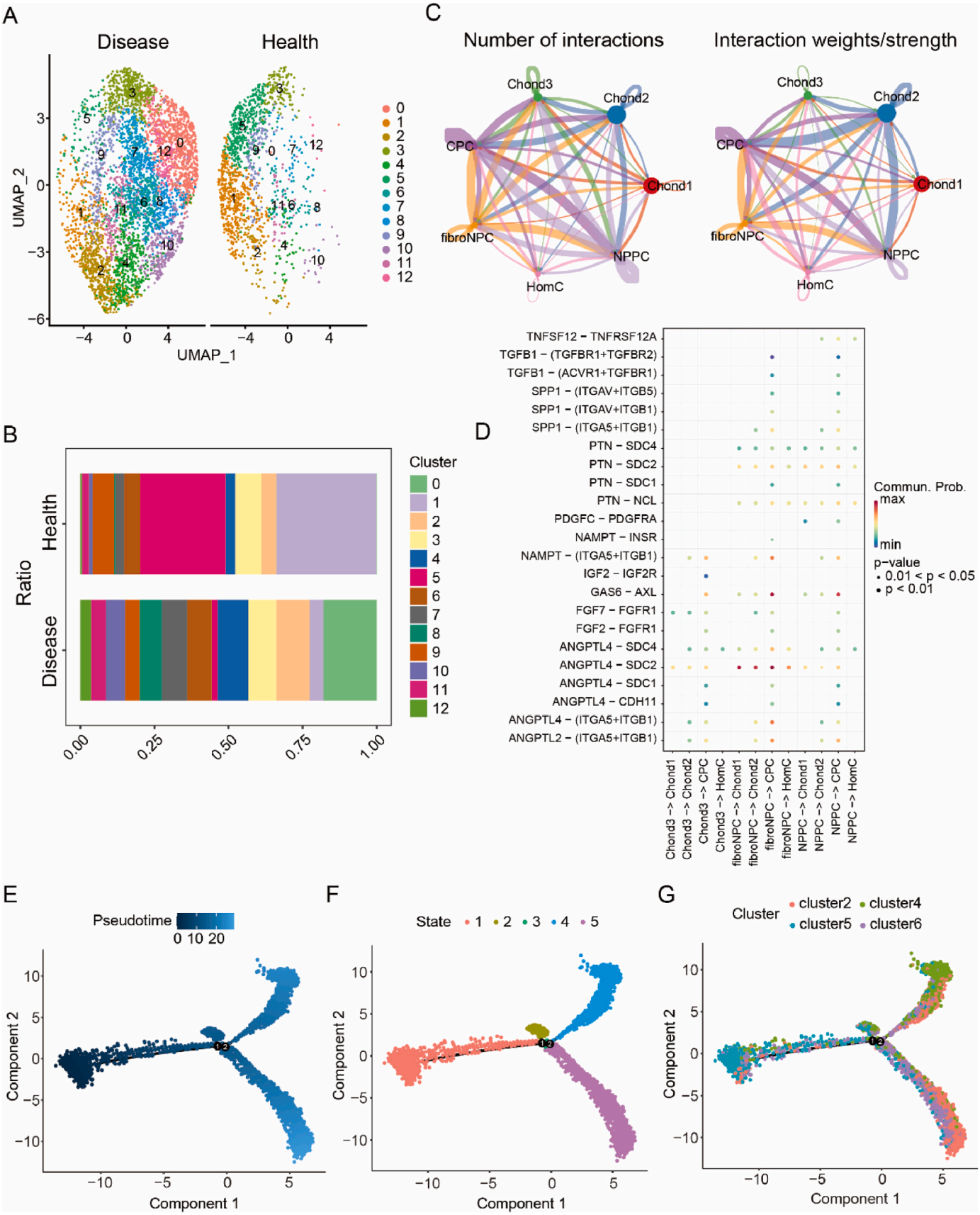

Fig. 2. Annotation of clustered subtypes and analysis of receptor-ligand pairs.
(A) Cell clustering results of disease and control samples; (B) The difference of the proportion of 12 kinds of cells in two groups of samples; (C) The cellular interaction network among 7 types of cells; (D) Bubble chart of receptor-ligand interactions between 12 types of cells; (E–G) Pseudotemporal analysis and developmental trajectory of Chond2 cells.

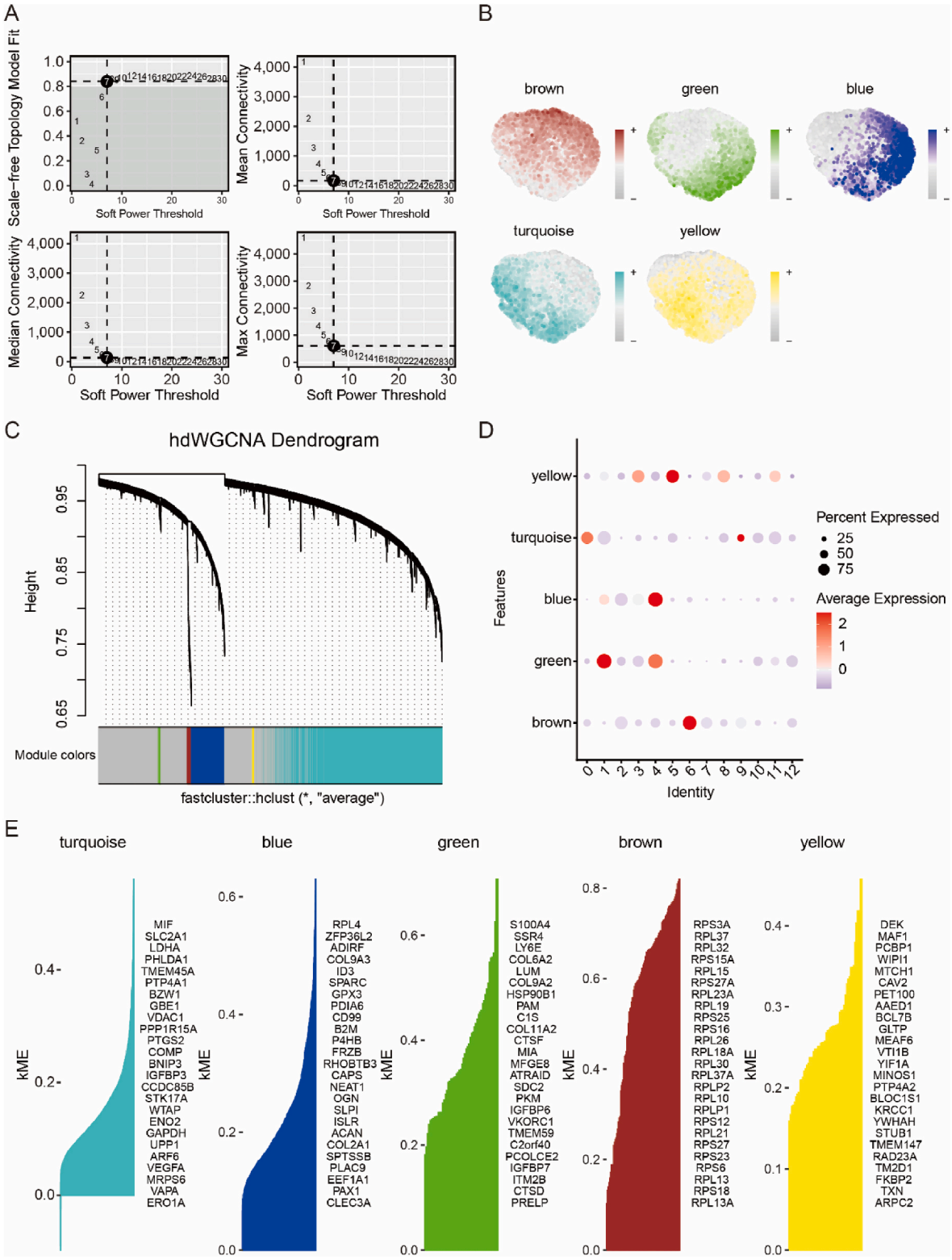

Fig. 3. Identification of module genes in Chond2 subtype.
(A) The scale-free exponent and average connectivity of various soft thresholds.; (B) Feature plot of module scores.; (C) Dendrogram of gene clustering; (D) Correlation among feature genes in the module; (E) Top feature genes in different modules.

parainflammation, and MHC class I (Fig. 5E). Finally, we analyzed correlation between four feature genes and different immune factors. As shown in Fig. 6, the results indicated a significant positive correlation between ACAN and immune factors such as CXCL14, LAG3, IL6R, HLA-C and CCR10. IGFBP3 was negatively correlated with CCL13, IL10, 1L6R, HLA-DPB1 and CCR10. We also found that TMEM45A was significantly negatively correlated with CCL28, BTLA, TNFRSF14, TAPBP, CXCR1, etc. VAPA was significantly positively correlated with TGFBR1, BTLA, MICB, CCR10 and other immune factors. These results indicated a close correlation between feature genes and the level of immune cell

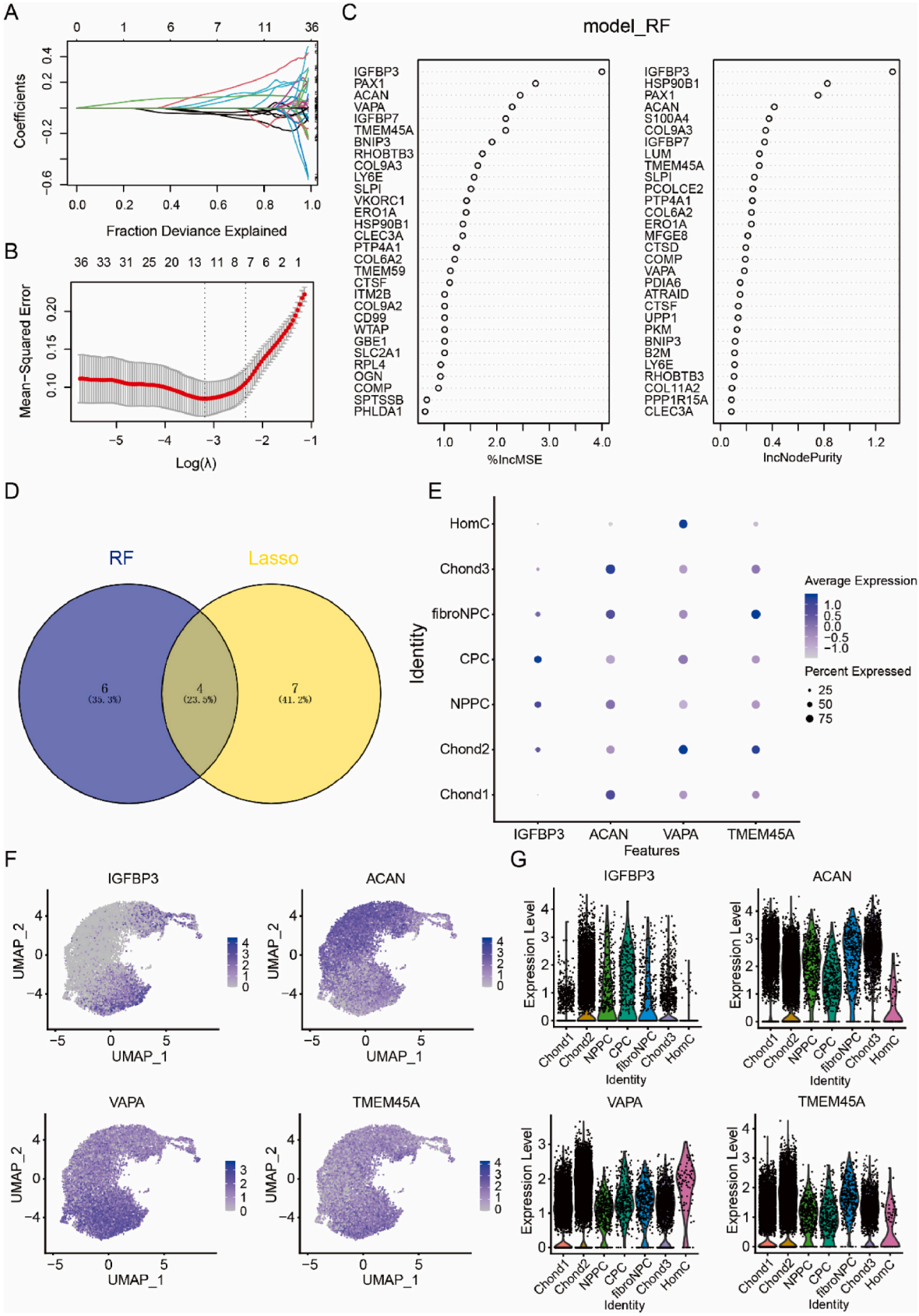

Fig. 4. Screening feature genes of IVDD.
(A) Distribution of LASSO coefficients and gene combinations at the minimum lambda value.; (B) Ten-fold cross-validation results for tuning parameter selection in the LASSO model; (C) Display of feature genes in RF model; (D) LASSO and RF gene intersection Venn diagram; (E–G) The expression of feature genes in different cellular subtypes.

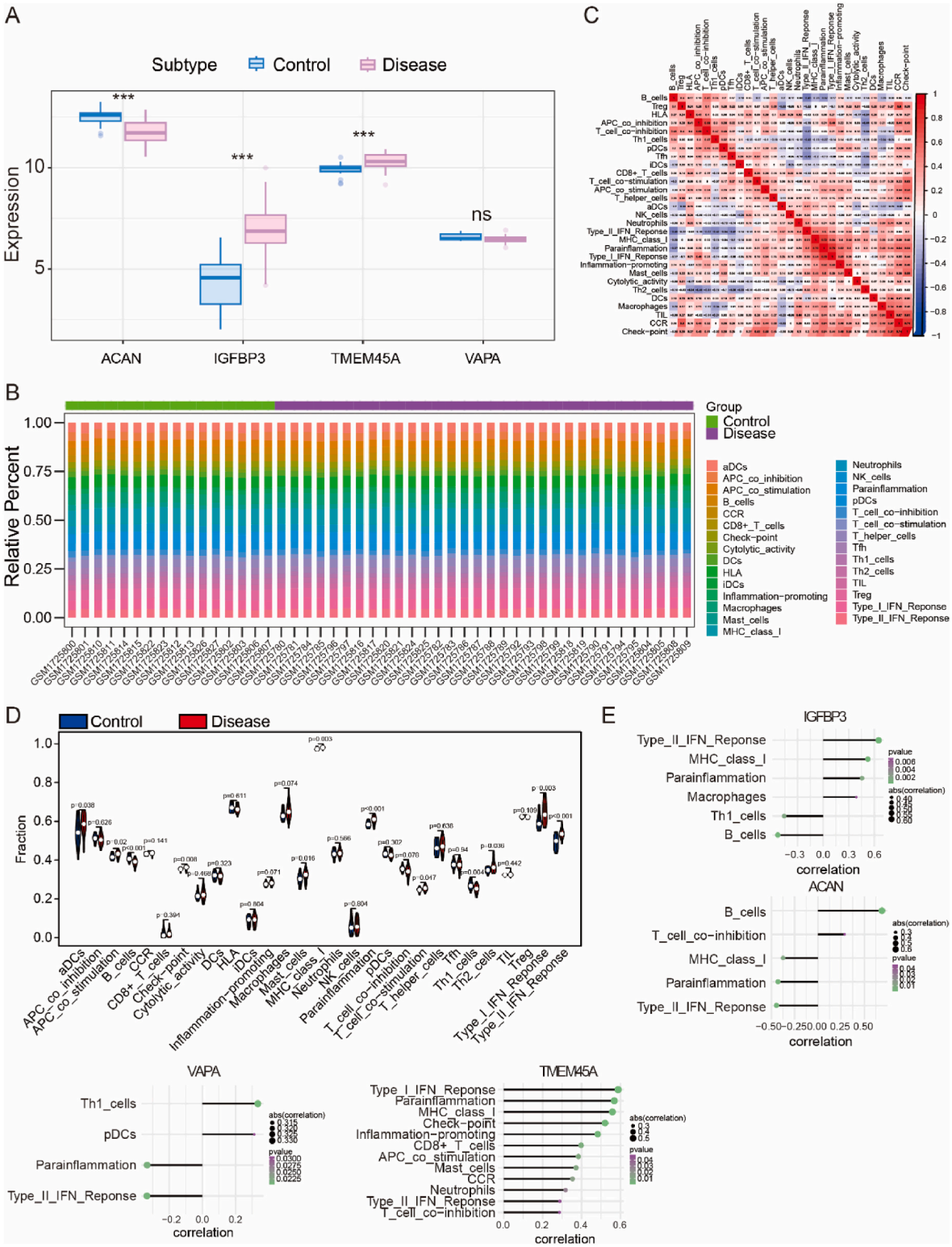

Fig. 5. Immunoinfiltration analysis revealed the key immune cells that play a role in IVDD.
(A) Differential expression of feature genes in control and disease groups; (B) The relative percentages of 29 immune cell subtypes in the control group and the disease group patients; (C) Pearson correlation among the 29 immune cell types; (D) Differences in immune cell composition between the control group and the disease group patients; (E) Correlation between feature genes and immune cell abundance.

infiltration in IVDD. This association may be of significant importance for understanding the pathogenic mechanisms of IVDD and immune regulation processes, providing new research directions for prevention of related diseases.

3.6. Clarify the signaling pathways of feature genes based on GSEA and GSVA analysis

We investigated potential molecular mechanisms underlying progression of IVDD by analyzing specific signaling pathways enriched by

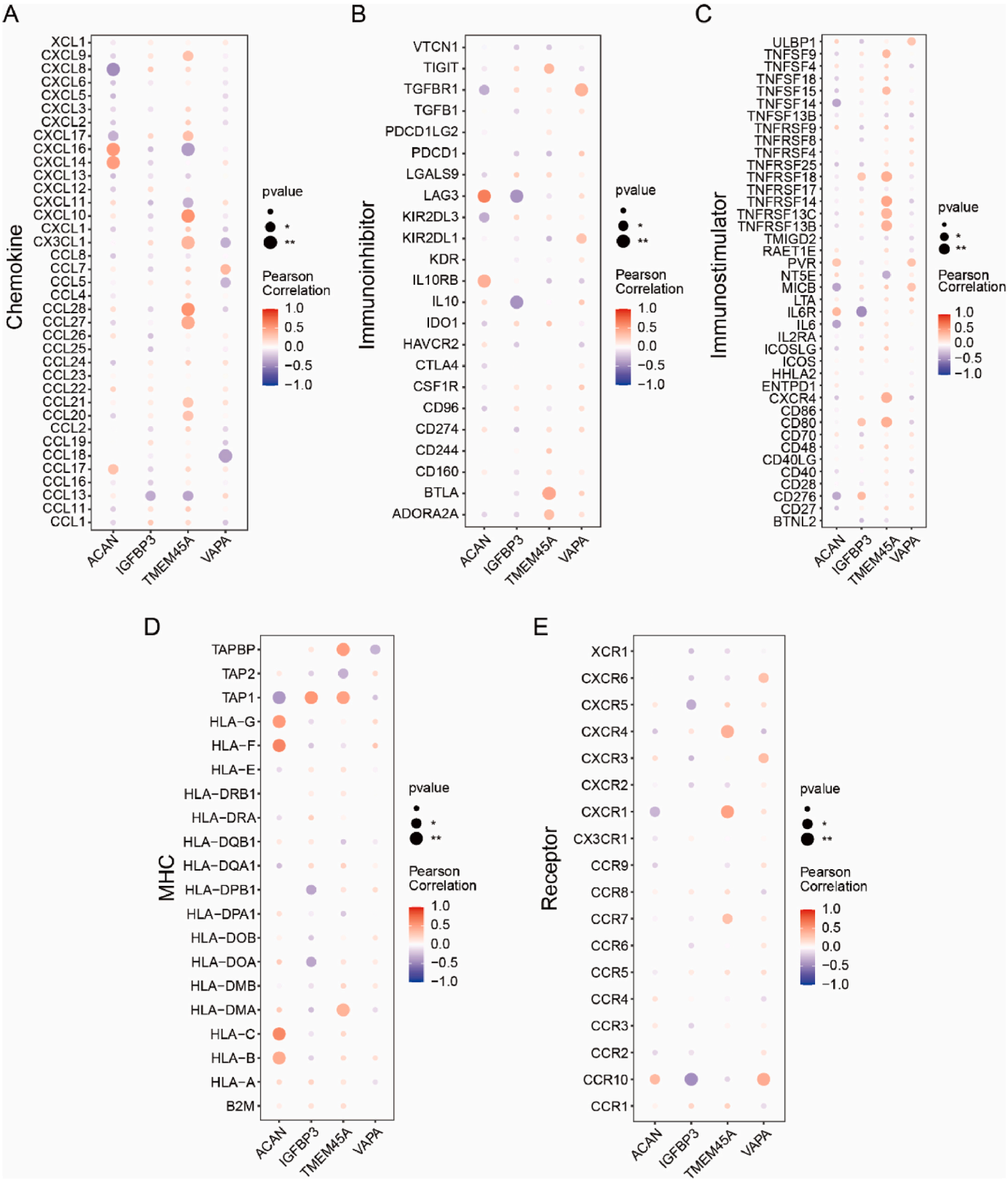

Fig. 6. Correlation between feature genes and immune factors.
(A) Correlation between feature genes and chemokines; (B) Correlation between feature genes and immunoinhibitory molecules; (C) Correlation between feature genes and immunostimulatory molecules; (D)Correlation between feature genes and MHCs; (E) Correlation between feature genes and receptors.

these feature genes. GSEA results showed that the pathways enriched by IGFBP3 including gap junction, p53 signaling pathway, TGF-β signaling pathway, etc. ACAN-enriched pathways included butanoate metabolism, FoxO signaling pathway and relaxin signaling pathway. The pathways enriched from VAPA included mRNA surveillance pathway, proteasome, spliceosome, etc. The pathways enriched by TMEM45A included IL-17 signaling pathway, p53 signaling pathway, TNF signaling pathway, etc. (Fig. 7). In addition, GSVA results showed that high expression of IGFBP3 was enriched in IL6_JAK_STAT3_SIGNALING, IL2_STAT5_SIGNALING, and NOTCH_SIGNALING, high expression of ACAN was enriched in PEROXISOME, UV_RESPONSE_DN, and TGF_BETA_SIGNALING, high expression of VAPA was enriched in

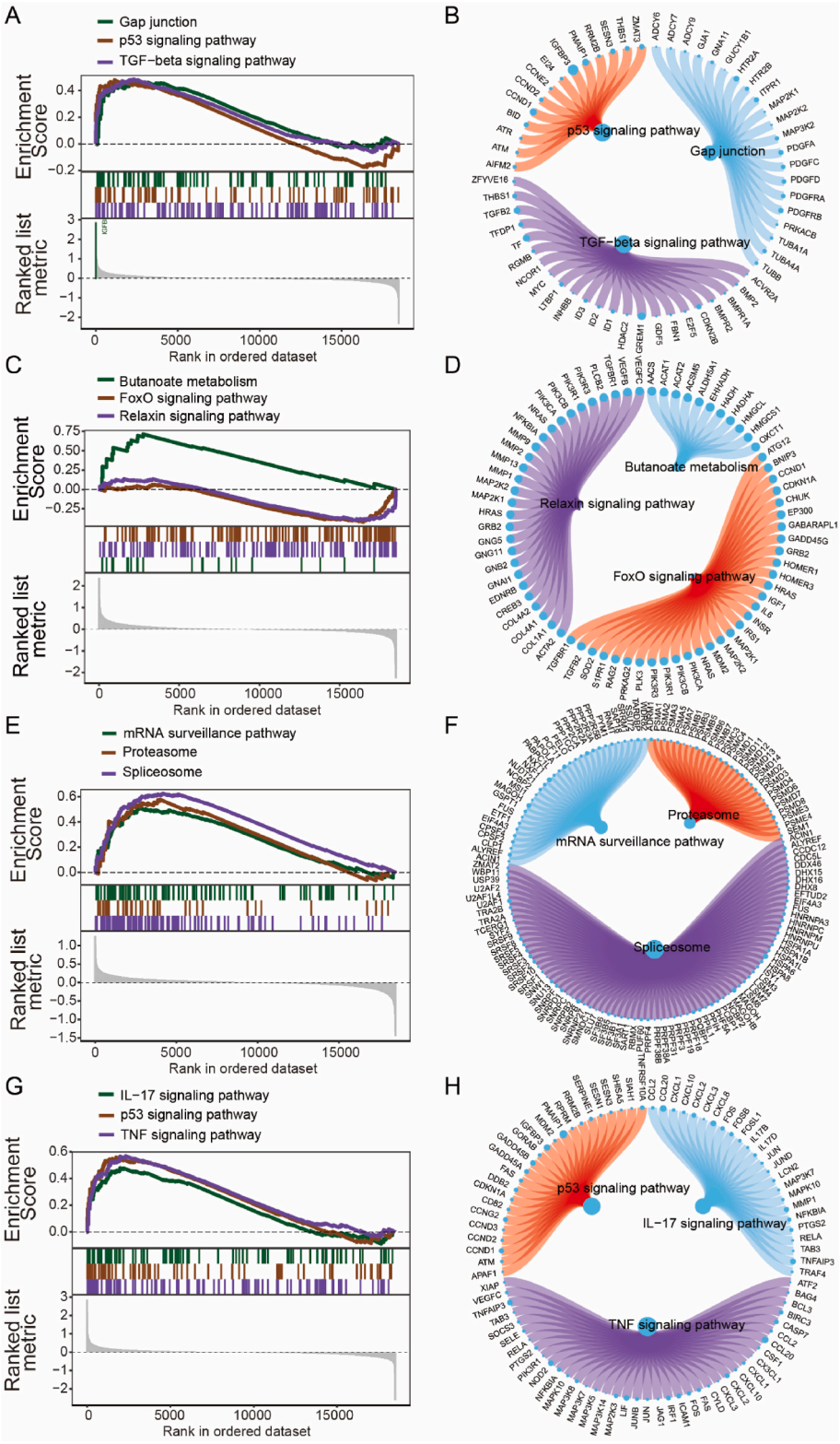

Fig. 7. GSEA analysis revealed the signaling pathways through which feature genes exert their effects.
(A–B) IGFBP3 GSEA enrichment result; (C–D) ACAN GSEA enrichment result; (E–F) VAPA GSEA enrichment result; (G–H) TMEM45A GSEA enrichment result.

pathways such as MTORC1, NOTCH_SIGNALING, and WNT_BETA_CATENIN_SIGNALING, and high expression of TMEM45A was enriched in pathways such as TNFA_SIGNALING_VIA_NFKB, and P53_PATHWAY (Fig. 8). This implies that these feature genes may exert an influence on progression of IVDD via these signaling pathways.

3.7. Correlation analysis, clinical validation and drug prediction of feature genes

To further investigate the inter-group variations in gene expression related to diseases, we conducted an analysis of the expression levels of the top 20 genes ranked by relevance score. We found significant differences in expression of ACAN, COL9A3, COL9A2, CILP, ASPN, HTRA1, VCAN, and IL1B between control and disease patients (Fig. 9A). Additionally, this study revealed a significant correlation between the expression levels of four feature genes and multiple disease-related genes. Specifically, TMEM45A showed a significant positive correlation with HTRA1 (r = 0.729) and a significant negative correlation with PRPF31 (r = −0.621), while IGFBP3 was positively correlated with PRPF31 (r = 0.374) and negatively correlated with COL9A2 (r = −0.384) (Fig. 9B). The above analysis results further indicated that these four feature genes were closely related to the development of IVDD.

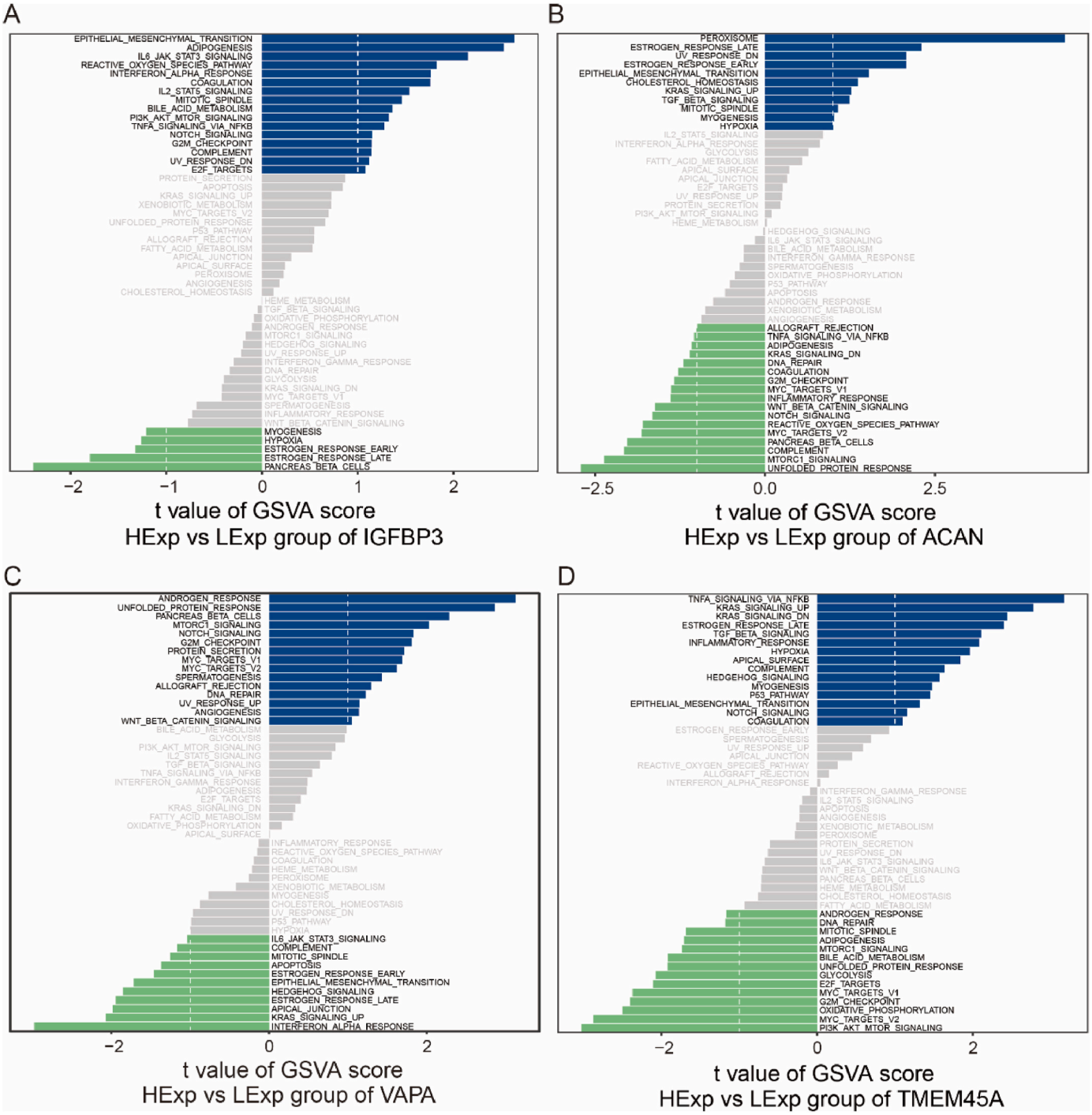

Fig. 8. GSVA analysis elucidated the enriched signaling pathways in the group of feature genes with differential expression.
(A) IGFBP3 GSVA enrichment result; (B) ACAN GSVA enrichment result; (C) VAPA GSVA enrichment result; (D) TMEM45A GSVA enrichment result.

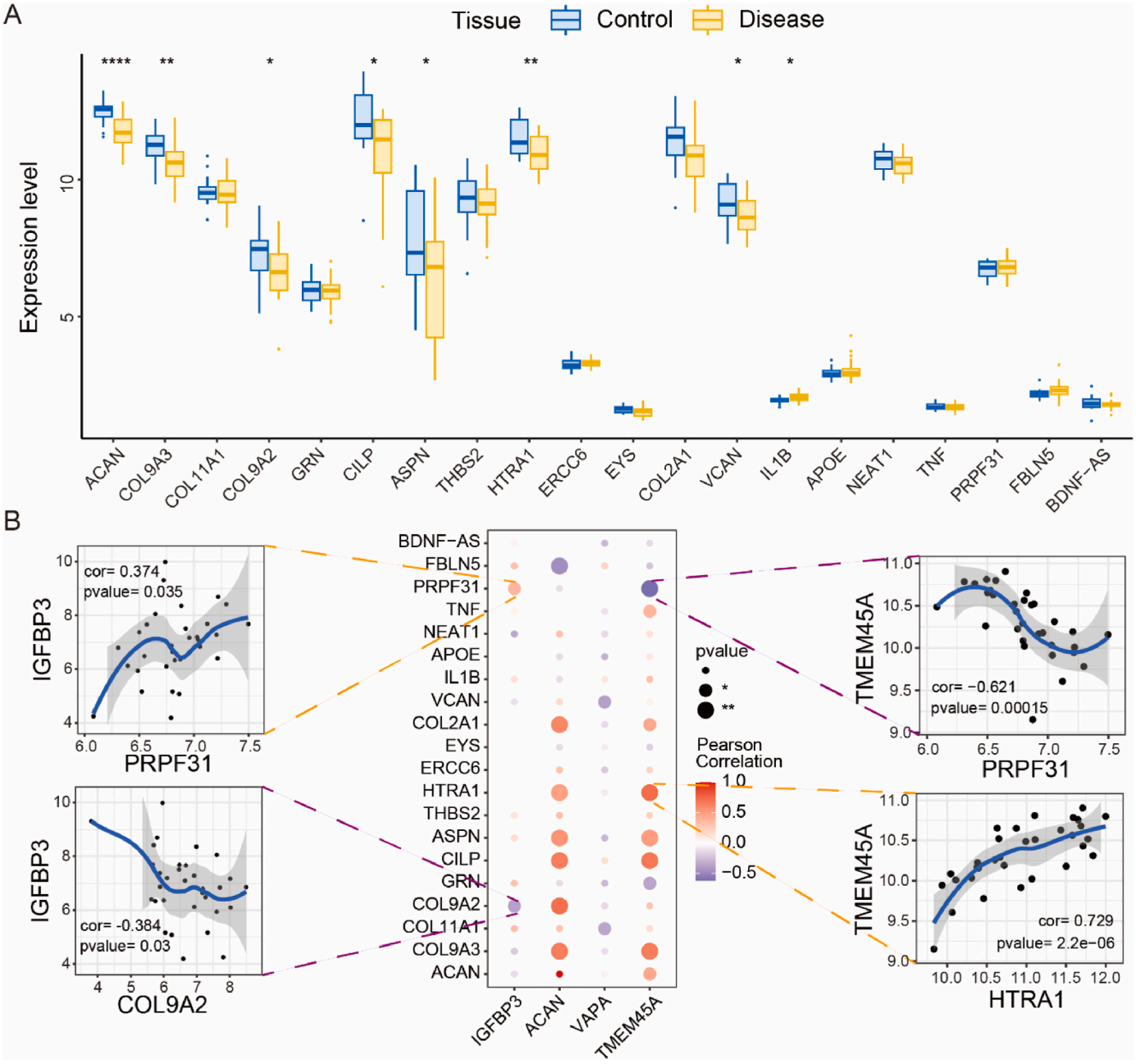

Fig. 9. Correlation of feature genes with disease progression genes.
(A) Differential expression of disease-regulated genes; (B) Pearson correlation analysis of feature genes with disease-related genes.

Therefore, clinical samples were collected and clinically verified by RT-qPCR, IHC and ELISA. The RT-qPCR results showed that compared to the MDD group, IGFBP3 and TMEM45A were upregulated in the SDD group, while ACAN and VAPA showed no significant differences between two groups (Fig. 10A–E). The IHC results indicated that the proportion of IGFBP3^{+} NPCs in SDD group was higher than that in MDD group (Fig. 10F–G). Additionally, the ELISA results indicated that IGFBP3 expression in Pfirrmann grade III and IV samples were higher than grade II, and the concentration of IGFBP3 was positively correlated with Pfirrmann grade (r = 0.6556, P = 0.0009) (Fig. 10H). Furthermore, we utilized DGIdb to analyze drugs that might interact with feature genes. The result indicated that the feature gene IGFBP3 interacted with two drugs (CELECOXIB and FLUOROURACIL) (Fig. 10I). CELECOXIB, a key medication for alleviating osteoarthritis, has been shown to inhibit the expression of IGFBP3. Therefore, we hypothesized that CELECOXIB might exert its therapeutic effects on IVDD by inhibiting the expression of IGFBP3. The genes screened by us are likely to play a significant role in the identification of novel therapeutic targets.

3.8. *Transcriptional regulation analysis of feature genes*

The preliminary analysis revealed that four feature genes are regulated by multiple transcription factors. Therefore, we performed enrichment analysis using cumulative recovery curves for transcription factors. Motif-TF annotation and selection analysis of important genes indicated that motif cisbp_M0872 had highest standardized enrichment score (NES: 4.84). In addition, it was found that the transcription factor upstream of ACAN and IGFBP3 was ZNF32, with their binding site being

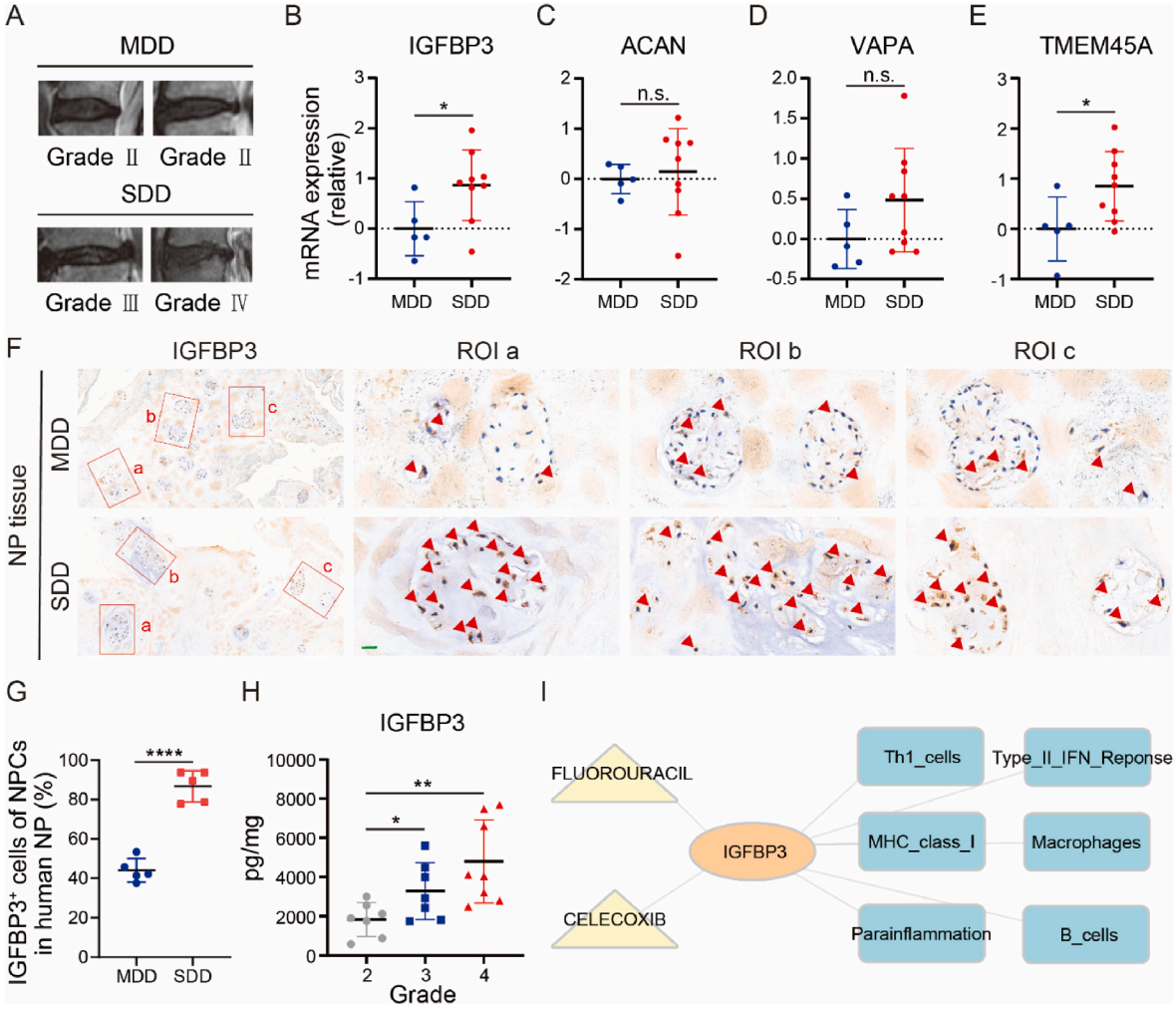

Fig. 10. Clinical validation and drug prediction.
(A–E) RT-qPCR was used to detect the differential expression of feature genes in MDD and SDD groups; (F–G) IHC was used to detect the proportion of IGFBP3^{+} NPCs in human NP in MDD and SDD groups. IGFBP3^{+} cells per field were determined from three non-overlapping fields per section and the average was taken from duplicates for each condition (n = 5). Scale bar = 20 μm ****P < 0.0001. ROI = region of interest; (H) The expression of IGFBP3 in Pfirrmann grade II - IV samples was determined by ELISA; (I) Prediction of drugs related to feature genes and their association with immune cells.

cisbp_M0437. The transcription factor upstream of ACAN, IGFBP3, and TMEM45A was FOXG1, with their binding site being cisbp_M0723 (Fig. S4).

4. Discussion

IVDD is a common and widely impacting spinal disease that affects the quality of life and health of individuals to a certain extent [13]. With the influence of aging populations and modern lifestyles, the incidence of IVDD is gradually increasing. Timely diagnosis of IVDD can lead to appropriate treatment measures before the condition worsens, thus avoiding serious consequences such as aggravated pain and nerve compression. scRNA-seq analysis holds important significance in the diagnosis of IVDD, as it allows for a more comprehensive understanding of the characteristics and changes in cellular composition within intervertebral disc tissue. This, in turn, facilitates a deeper exploration of the pathogenesis of IVDD, diagnostic biomarkers, and potential therapeutic targets. Based on scRNA-seq analysis, it was found that Chond2 cell subtypes were closely related to IVDD disease. Furthermore, the CellChat results indicated enhanced communication between CPC, Chond2, and NPCC. LASSO and RF algorithms identified four feature genes that influence the development of IVDD, namely IGFBP3, ACAN, VAPA and TMEM45A. Additionally, immune infiltration analysis revealed the differences in immune cell populations between control group and IVDD patients. Results from GSEA and GSVA indicated that these feature genes impact the progression of IVDD through different signaling pathways. Drug prediction results revealed an interaction between the feature gene IGFBP3 and two drugs, CELECOXIB and FLUOROURACIL. The ELISA testing revealed a positive correlation between IGFBP3 concentration and severity of IVDD. These findings provide new insights into understanding pathogenesis and biological significance for IVDD in depth.

The IGFBP3 gene encodes the Insulin-like Growth Factor Binding Protein 3 and is located on human chromosome 7 [14]. It contains multiple exons, and through transcription and translation, ultimately produces the IGFBP-3 protein [15]. IGFBP3 binds to insulin-like growth factor (IGF) and regulates different biological processes [16]. IGFBP-3 affects cell metabolism, and proliferation functions. Research by Chan YX et al. found that high levels of IGFBP3 may predict an increased incidence of colorectal cancer in elderly men [17]. Additionally, scholars such as Yoshino K discovered that IGFBP3 may be a key marker for enhancing the radiation sensitivity of squamous cell esophageal cancer [18]. Furthermore, Kazezian Z found that hyaluronic acid could downregulate the expression of IGFBP3 in bovine intervertebral discs [19]. Above studies suggest a strong association between IGFBP3 and the pathogenesis of tumors and IVDD. In this study, it is found that IGFBP3 plays a pivotal role as a regulatory gene in the development of

IVDD and ELISA tests revealed a positive correlation between the concentration of IGFBP3 and degree of IVDD. This further underscores the pivotal role of IGFBP3 in the pathogenesis of IVDD. Therefore, studying the IGFBP-3 gene and the protein it encodes can help us better understand the pathogenesis of IVDD disease. The ACAN gene encodes a protein called Aggrecan, which is one of the most abundant components in the cartilage matrix [20]. Aggrecan forms a complex mesh-like structure in the extracellular matrix of cartilage, interacting with other proteins such as collagen to maintain the structural integrity and function of cartilage tissue. Mutations or abnormal expression of the ACAN gene may lead to conditions such as defective cartilage development or degenerative joint diseases. The VAPA gene encodes a membrane protein that is involved in regulating intracellular lipid metabolism and transport, affecting normal cellular functions and survival. The TMEM45A gene is located on human chromosome 3, encoding transmembrane protein 45A, which participates in various biological processes within cells. Over time, an increasing body of research has demonstrated the roles of ACAN, VAPA, and TMEM45A in various diseases, including IVDD. Keskin et al. demonstrated the association of ACAN gene VNTR polymorphism with the development of Alzheimer's disease [21]. Uchida et al. revealed a relationship between ACAN gene mutations and an autosomal dominant short stature [22]. Researchers such as Chen X found that U2AF2 weakened IVDD by affecting ACAN [23]. Regarding VAPA, Zh et al. discovered that VAPA promoted the metastasis of liver cancer to bone [24]. The impact of TMEM45A on the proliferation and apoptosis of HPV-positive cervical cancer cells has been demonstrated [25]. Flamant et al. demonstrated that TMEM45A was crucial for chemoresistance in liver cancer cells [26]. Currently, no studies have revealed the relationship between VAPA, TMEM45A, and IVDD. However, our research has discovered that VAPA and TMEM45A are important regulatory genes in the occurrence of IVDD, which is the innovation of this study. In general, the four feature genes we have identified are of significant importance for understanding the pathogenesis of IVDD and developing treatment strategies. IVDD is a complex process involving multiple cellular and molecular mechanisms. Immune cells, including macrophages, lymphocytes, dendritic cells, etc., play crucial roles in the inflammatory response within the microenvironment of the IVDD. It is important that we observed differences in Mast_cells, MHC_class_I, Th2_cells, Type_I_IFN_Response, etc. between IVDD and control patients. MHC class I molecules are an important group of immune-related proteins that play a vital role in antigen presentation and immune recognition processes. These molecules assist the body in identifying and clearing infected or abnormal cells, maintaining immune homeostasis, and resisting pathogen invasion. Additionally, research has revealed the involvement of MHC class I in the pathogenesis and progression of various diseases. For example, Bao Y et al. found that PIKfyve up-regulated MHC Class I surface expression to enhance cancer immunotherapy [27]. Meanwhile, Chew GL and colleagues demonstrated that DUX4 inhibition of MHC class I promoted cancer immune evasion [28]. Our study has revealed the relevance of MHC class I to IVDD occurrence, which is crucial for further diagnosis and treatment of IVDD. The interaction between the immune microenvironment and the IVDD is dynamic and plays a pivotal and indispensable role in IVDD. Therefore, a comprehensive investigation into the role of the immune microenvironment in the pathogenesis of IVDD is imperative.

In conclusion, our research findings have revealed the significant roles of the key cell subtype Chond2 and four feature genes (IGFBP3, ACAN, VAPA, and TMEM45A) in the development of IVDD. This provides potential therapeutic targets for future IVDD research and treatments, holding certain clinical significance. Nevertheless, it is imperative to recognize limitations of this study. The limited number of samples in scRNA-seq may affect the reliability of the analysis results. Additionally, the complete elucidation of the molecular mechanisms underlying the functions of the four feature genes identified have not been fully elucidated and require further investigation. Despite the shortcomings of this article, our analysis results have also pointed to new directions for future research.

Funding

This study was supported by grants from Post-subsidy funds for National Clinical Research Center, Ministry of Science and Technology of China (No.303-01-001-0272-05), the National Key Research and Development Program of China (No.2020YFC2004900), R&D Program of Beijing Municipal Education Commission (22JG0059), the key Science and Technology Project of Beijing Municipal Education Commission (KZ20231002537), and the National Natural Youth Cultivation Project of Xuanwu Hospital of Capital Medical University (QNPY202316).

Ethics approval and consent to participate

The study was conducted in accordance with the Declaration of Helsinki, and approved by the Ethics Committee of Capital Medical University Xuanwu Hospital. (Ethic number: 2024010-002).

Consent for publication

Not applicable.

Availability of data and materials

The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

CRediT authorship contribution statement

Xuan Zhao: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation. **Qijun Wang:** Software, Formal analysis. **Wei Wang:** Conceptualization. **Xiaolong Chen:** Supervision, Conceptualization. **Shibao Lu:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors have declared no conflict of interest.

Acknowledgements

Not applicable.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ncrna.2024.09.003.

References

[1] S. Kirnaz, C. Capadona, T. Wong, J.L. Goldberg, B. Medary, F. Sommer, L. B. McGrath Jr., R. Härtl, Fundamentals of intervertebral disc degeneration, World Neurosurg 157 (2022) 264–273.
[2] R.D. Fraser, O.L. Osti, B. Vernon-Roberts, Intervertebral disc degeneration, Eur. Spine J. 1 (4) (1993) 205–213.
[3] P. Thavaneswaran, M. Vandepeer, Lumbar artificial intervertebral disc replacement: a systematic review, ANZ J. Surg. 84 (3) (2014) 121–127.
[4] Q. Xiang, Y. Zhao, W. Li, Identification and validation of ferroptosis-related gene signature in intervertebral disc degeneration, Front. Endocrinol. 14 (2023) 1089796.
[5] X. Jiang, J. Wu, C. Guo, W. Song, Key LncRNAs associated with oxidative stress were identified by GEO database data and whole blood analysis of intervertebral disc degeneration patients, Front. Genet. 13 (2022) 929843.

[6] S. Xu, H. Fu, S. Weng, X. Gu, J. Li, Derivation and comprehensive analysis of ageing-related genes in intervertebral disc degeneration for prediction and immunology, Mech. Ageing Dev. 211 (2023) 111794.
[7] H. Cherif, M. Mannarino, A.S. Pacis, J. Ragoussis, O. Rabau, J.A. Ouellet, L. Haglund, Single-cell RNA-seq analysis of cells from degenerating and non-degenerating intervertebral discs from the same individual reveals new biomarkers for intervertebral disc degeneration, Int. J. Mol. Sci. 23 (7) (2022).
[8] Y. Zhang, S. Han, M. Kong, Q. Tu, L. Zhang, X. Ma, Single-cell RNA-seq analysis identifies unique chondrocyte subsets and reveals involvement of ferroptosis in human intervertebral disc degeneration, Osteoarthritis Cartilage 29 (9) (2021) 1324–1334.
[9] X. Yang, Y. Lu, H. Zhou, H.T. Jiang, L. Chu, Integrated proteome sequencing, bulk RNA sequencing and single-cell RNA sequencing to identify potential biomarkers in different grades of intervertebral disc degeneration, Front. Cell Dev. Biol. 11 (2023) 1136777.
[10] H. Sun, X. Wen, H. Li, P. Wu, M. Gu, X. Zhao, Z. Zhang, S. Hu, G. Mao, R. Ma, W. Liao, Z. Zhang, Single-cell RNA-seq analysis identifies meniscus progenitors and reveals the progression of meniscus degeneration, Ann. Rheum. Dis. 79 (3) (2020) 408–417.
[11] Q. Ji, Y. Zheng, G. Zhang, Y. Hu, X. Fan, Y. Hou, L. Wen, L. Li, Y. Xu, Y. Wang, F. Tang, Single-cell RNA-seq analysis reveals the progression of human osteoarthritis, Ann. Rheum. Dis. 78 (1) (2019) 100–110.
[12] Y. Gan, J. He, J. Zhu, Z. Xu, Z. Wang, J. Yan, O. Hu, Z. Bai, L. Chen, Y. Xie, M. Jin, S. Huang, B. Liu, P. Liu, Spatially defined single-cell transcriptional profiling characterizes diverse chondrocyte subtypes and nucleus pulposus progenitors in human intervertebral discs, Bone Res 9 (1) (2021) 37.
[13] Y. Wang, J. Kang, X. Guo, D. Zhu, M. Liu, L. Yang, G. Zhang, X. Kang, Intervertebral disc degeneration models for pathophysiology and regenerative therapy -benefits and limitations, J. Invest. Surg. 35 (4) (2022) 935–952.
[14] M.B. Ranke, Insulin-like growth factor binding-protein-3 (IGFBP-3), Best Pract. Res. Clin. Endocrinol. Metabol. 29 (5) (2015) 701–711.
[15] A. Bhardwaj, K.A. Pathak, A. Shrivastav, S. Varma Shrivastav, Insulin-like growth factor binding protein-3 binds to histone 3, Int. J. Mol. Sci. 22 (1) (2021).
[16] S. Jogie-Brahim, D. Feldman, Y. Oh, Unraveling insulin-like growth factor binding protein-3 actions in human disease, Endocr. Rev. 30 (5) (2009) 417–437.
[17] Y.X. Chan, H. Alfonso, Chubb S.A. Paul, K.K.Y. Ho, P. Gerard Fegan, G.J. Hankey, J. Golledge, L. Flicker, B.B. Yeap, Higher IGFBP3 is associated with increased incidence of colorectal cancer in older men independently of IGF1, Clin. Endocrinol. 88 (2) (2018) 333–340.
[18] K. Yoshino, S. Motoyama, S. Koyota, K. Shibuya, S. Usami, K. Maruyama, H. Saito, Y. Minamiya, T. Sugiyama, J. Ogawa, IGFBP3 and BAG1 enhance radiation-induced apoptosis in squamous esophageal cancer cells, Biochem. Biophys. Res. Commun. 404 (4) (2011) 1070–1075.
[19] Z. Kazezian, Z. Li, M. Alini, S. Grad, A. Pandit, Injectable hyaluronic acid down-regulates interferon signaling molecules, IGFBP3 and IFIT3 in the bovine intervertebral disc, Acta Biomater. 52 (2017) 118–129.
[20] H. Watanabe, Aggrecan and versican: two brothers close or apart, Am. J. Physiol. Cell Physiol. 322 (5) (2022) C967–c976.
[21] T. Keskin, O. Avsar, S. Eliacik, F. Uysal Tan, Investigation of the relationship between ACAN gene VNTR polymorphism and Alzheimer's disease in Turkish population, Nucleos Nucleot. Nucleic Acids (2024) 1–10.
[22] N. Uchida, H. Shibata, G. Nishimura, T. Hasegawa, A novel mutation in the ACAN gene in a family with autosomal dominant short stature and intervertebral disc disease, Hum Genome Var 7 (1) (2020) 44.
[23] X. Chen, D. Cai, H. Li, Q. Wei, X. Li, Z. Han, J. Liang, J. Xie, J. Ruan, J. Liu, Z. Xiang, W. Dong, W. Guo, Exosomal U2AF2 derived from human bone marrow mesenchymal stem cells attenuates the intervertebral disc degeneration through circ_0036763/miR-583/ACAN axis, Regen Ther 25 (2024) 344–354.
[24] S. Zhang, X. Liao, S. Chen, W. Qian, M. Li, Y. Xu, M. Yang, X. Li, S. Mo, M. Tang, X. Wu, Y. Hu, Z. Li, R. Yu, A. Abudourousuli, L. Song, J. Li, Large oncosome-loaded VAPA promotes bone-tropic metastasis of hepatocellular carcinoma via formation of osteoclastic pre-metastatic niche, Adv. Sci. 9 (31) (2022) e2201974.
[25] Y. Liu, L. Liu, Z.X. Mou, TMEM45A affects proliferation, apoptosis, epithelial-mesenchymal transition, migration, invasion and cisplatin resistance of HPV-positive cervical cancer cell lines, Biochem. Genet. 60 (1) (2022) 173–190.
[26] L. Flamant, E. Roegiers, M. Pierre, A. Hayez, C. Sterpin, O. De Backer, T. Arnould, Y. Poumay, C. Michiels, TMEM45A is essential for hypoxia-induced chemoresistance in breast and liver cancer cells, BMC Cancer 12 (2012) 391.
[27] Y. Bao, Y. Qiao, J.E. Choi, Y. Zhang, R. Mannan, C. Cheng, T. He, Y. Zheng, J. Yu, M. Gondal, G. Cruz, S. Grove, X. Cao, F. Su, R. Wang, Y. Chang, I. Kryczek, M. Cieslik, M.D. Green, W. Zou, A.M. Chinnaiyan, Targeting the lipid kinase PIKfyve upregulates surface expression of MHC class I to augment cancer immunotherapy, Proc. Natl. Acad. Sci. U. S. A. 120 (49) (2023) e2314416120.
[28] G.L. Chew, A.E. Campbell, E. De Neef, N.A. Sutliff, S.C. Shadle, S.J. Tapscott, R. K. Bradley, DUX4 suppresses MHC class I to promote cancer immune evasion and resistance to checkpoint blockade, Dev. Cell 50 (5) (2019) 658–671.e7.