*current issues in molecular biology*

MDPI

*Article*

# Reproducibility of Transcriptomic Signatures and Candidate Genes in Human Intervertebral Disc Degeneration: A Discovery–Validation Re-Analysis of Public Datasets

**Sridhar Muthusami [1], Kaviyaprabha Rangaraj [2], Sathish Muthu [3,4,*], Girinivasan Chellamuthu [4,5], Yogavalli Madhavan [6], Khan Sharun [7], Bharathi Muruganantham [2], Ramya Lakshmi Rajendran [8,9,10], Prakash Gangadaran [8,9] and Byeong-Cheol Ahn [8,9,11,*]**

1 Centre for Cancer Research, Department of Biochemistry, Karpagam Academy of Higher Education, Eachanari, Coimbatore 641021, Tamil Nadu, India; sridhar.m@kahedu.edu.in
2 Centre for Bioinformatics, Department of Biochemistry, Karpagam Academy of Higher Education, Eachanari, Coimbatore 641021, Tamil Nadu, India; kaviya97900@gmail.com (K.R.); bharathi.m@kahedu.edu.in (B.M.)
3 Central Research Laboratory, Aarupadai Veedu Medical College and Hospital, Vinayaka Mission's Research Foundation (Deemed to Be University), Puducherry 607402, India
4 Department of Translational Research, Orthopaedic Research Group, Coimbatore 641045, Tamil Nadu, India; giri.c.nivasan@gmail.com
5 Department of Orthopaedics, Kauvery Hospitals, Trichy 620001, Tamil Nadu, India
6 Department of Microbiology, Karpagam Academy of Higher Education, Eachanari, Coimbatore 641021, Tamil Nadu, India; kalyanimadhu2709@gmail.com
7 Graduate Institute of Medicine, Yuan Ze University, Taoyuan 32003, Taiwan; sharunkhansk@gmail.com
8 Department of Nuclear Medicine, School of Medicine, Kyungpook National University, Daegu 41944, Republic of Korea; ramyag@knu.ac.kr (R.L.R.); prakashg@knu.ac.kr (P.G.)
9 Cardiovascular Research Institute, Kyungpook National University, Daegu 41944, Republic of Korea
10 BK21 FOUR KNU Convergence Educational Program of Biomedical Sciences for Creative Future Talents, Department of Biomedical Sciences, School of Medicine, Kyungpook National University, Daegu 41944, Republic of Korea
11 Department of Nuclear Medicine, Kyungpook National University Hospital, Daegu 41944, Republic of Korea
* Correspondence: drsathishmuthu@gmail.com (S.M.); abc2000@knu.ac.kr (B.-C.A.)

## Abstract

***Purpose:*** Integrative analyses of public transcriptomic datasets are widely used to nominate hub genes and therapeutic targets in intervertebral disc degeneration (IVDD), but whether such candidates replicate across datasets is rarely tested. We assessed the reproducibility of gene- and pathway-level signatures of IVDD using a discovery–validation design. ***Methods:*** Six Gene Expression Omnibus (GEO) series used in an overlap-and-network pipeline (four degeneration-versus-control comparisons and two cell-type or intervention contrasts) were re-analysed dataset by dataset (limma or DESeq2, with design adjustment where needed). Thirty-three candidate genes from that pipeline were audited, and a weighted Stouffer meta-analysis of the four discovery datasets was tested for replication in three independent validation cohorts (GSE70362, 48 samples from 18 donors; the non-overlapping samples of GSE23130; GSE34095). We also quantified genome-wide agreement between datasets, checked sample uniqueness, performed pre-ranked gene set enrichment analysis (Hallmark and Reactome) and analysed a combined protein–protein interaction network. ***Results:*** Samples 1–15 of GSE23130 were the specimens of GSE15227, and 23 PubMed Central articles cite both accessions. Genome-wide agreement between degeneration datasets was negligible (Spearman ρ −0.13 to 0.13). Of 1252 genes significant in the discovery meta-analysis, 49% changed in the same direction in validation (binomial $p = 0.78$); 15 replicated and two remained robust to leave-one-dataset-out analysis. None of the *33 candidates* replicated robustly; ACAN, COL9A3 and COL11A2 changed significantly in the opposite direction in validation, and the apparent downregulation of MMP3 and MMP9 arose only from non-degeneration contrasts. A reproducible overlap step selected 14 mainly interferon-related

Check for updates

Academic Editors: Koji Akeda, Nobuyuki Fujita and Akira Shinohara

Received: 27 October 2026
Revised: 6 November 2026
Accepted: 11 November 2026
Published: 16 November 2026

**Copyright:** © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

genes driven by these contrasts. At the pathway level, E2F target and G2-M checkpoint (proliferation) gene sets were enriched in both meta-analyses and had positive enrichment scores in all nine datasets, whereas inflammatory and TNF-α/NF-κB signatures did not replicate. ***Conclusions:*** Gene-level IVDD signatures derived from small public datasets are not reproducible, and overlap-based hub genes should not be presented as therapeutic targets without independent validation. Within the datasets examined, a cell proliferation programme was the most consistent pathway-level signal; whether it is a feature of disc degeneration requires experimental validation. We provide a checklist for future secondary analyses of IVDD transcriptomes.

**Keywords:** intervertebral disc degeneration; transcriptomics; reproducibility; meta-analysis; Gene Expression Omnibus; validation; cell proliferation

## 1. Introduction

Intervertebral disc degeneration (IVDD) is a major cause of chronic low back pain, affecting millions worldwide and contributing substantially to disability and socioeconomic burden [1,2]. In a population-based Japanese cohort (the Wakayama Spine Study; $n = 975$), MRI-defined IVDD was present in 71% of men and 77% of women younger than 50 years and in more than 90% of both sexes older than 50 years [3], and degeneration grade increases with age at every lumbar level in both sexes [4]. Despite its prevalence, IVDD remains underrecognised as a global health issue, with limited curative treatment options available [5].

The intervertebral disc (IVD) is a complex fibrocartilaginous structure composed of the nucleus pulposus (NP), annulus fibrosus (AF), transitional zone (TZ), and cartilaginous endplates (CEPs). These components collectively maintain spinal flexibility, load distribution, and mechanical stability [6,7]. Degeneration disrupts this balance, leading to extracellular matrix (ECM) breakdown, altered biomechanics, and inflammatory cascades [7]. Risk factors such as ageing, genetic predisposition, obesity, smoking, occupational stress, and low-grade infection accelerate this process [8,9].

At the molecular level, IVDD is characterised by an imbalance between anabolic and catabolic activities within disc cells, with altered ECM synthesis, matrix metalloproteinase (MMP) and ADAMTS activity, inflammatory cytokine signalling, cellular senescence and apoptosis [10–12]. Genetic polymorphisms in collagen genes, aggrecan (ACAN), the vitamin D receptor and cartilage intermediate-layer protein have been linked to susceptibility [13–15].

Public repositories such as the Gene Expression Omnibus (GEO) [16] have made disc transcriptomes widely available, and many secondary analyses combine several GEO series to nominate hub genes, biomarkers and therapeutic targets for IVDD [17]. Such analyses face well-known threats to reproducibility: very small groups, heterogeneous tissues and grading systems, cross-platform batch effects, confounding by age, and comparisons that do not address degeneration at all [18,19]. A common approach selects genes that are differentially expressed in several datasets and ranks them in protein–protein interaction (PPI) networks, but the resulting candidates are seldom tested in independent cohorts.

We previously applied such an overlap-and-network pipeline to six GEO series and prioritised 15 upregulated and 18 downregulated candidate genes, dominated by collagens and matrix-modifying enzymes. Here, prompted by peer review, we examined whether these candidates and the underlying signatures are reproducible. We re-analysed each dataset with explicit designs, distinguished degeneration comparisons from cell-type and intervention contrasts, checked sample uniqueness across series, quantified genome-wide

agreement between datasets, and tested a discovery meta-analysis in three independent validation cohorts at the gene and pathway levels. Our aim was to establish which, if any, transcriptomic features of IVDD are robust enough to guide experimental work.

## 2. Methods

### *2.1. Dataset Identification and Selection*

The six original series (GSE15227, GSE17782, GSE27494, GSE132958, GSE147383 and GSE167199) were selected in January 2026 from GEO (https://www.ncbi.nlm.nih.gov/geo/, accessed on 4 November 2026) because they contained human disc tissue or disc cells with an annotated two-group comparison. To identify validation cohorts and to document the search, GEO DataSets was searched on 4 November 2026 with the query ("intervertebral disc degeneration" OR "intervertebral disc" OR "nucleus pulposus" OR "annulus fibrosus") AND "Homo sapiens" [Organism] AND "gse" [Entry Type], restricted to expression profiling by array or by high-throughput sequencing. The search returned 77 series (Supplementary Table S1). Validation cohorts had to be bulk human nucleus pulposus (NP) or annulus fibrosus (AF) tissue, compare degenerated with non- or less-degenerated discs within one series, provide gene-level data, and not overlap the discovery samples. Three series met these criteria: GSE70362, GSE23130 and GSE34095 (Figure 1). GSE56081 (lncRNA array without gene-level annotation), GSE186542 (NCBI-processed counts available for only one control), single-cell, cell culture, blood and non-disc series were excluded; the reasons for each series are listed in Supplementary Table S1.

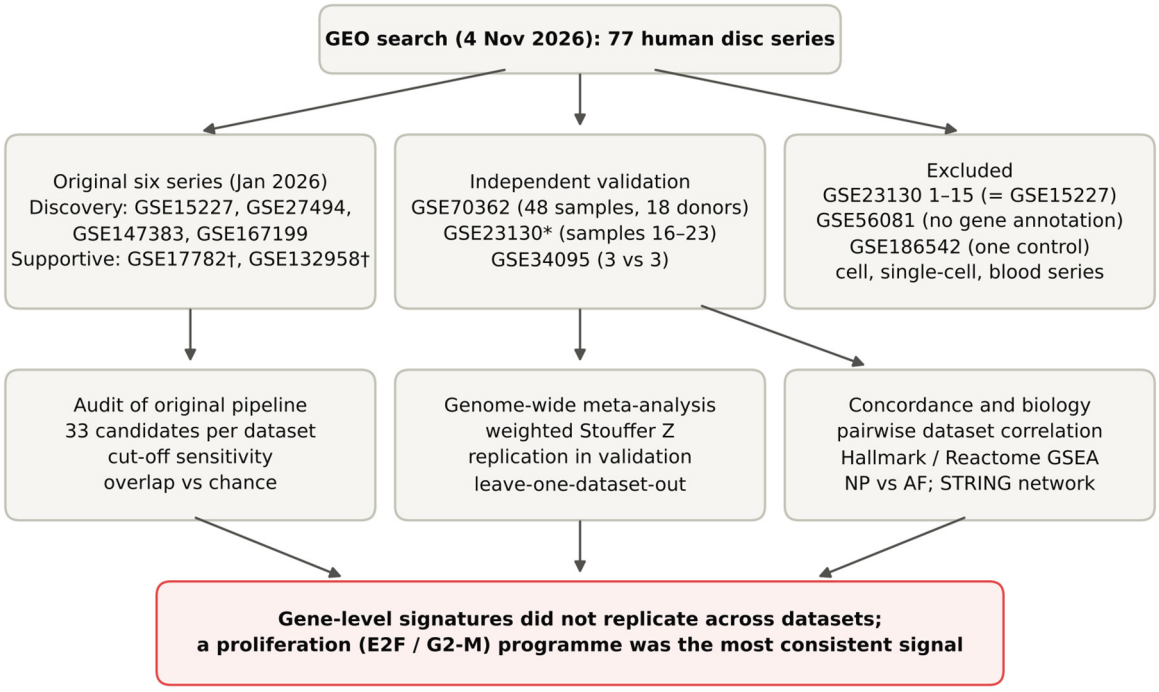


**Figure 1.** Study design. The six original series were re-analysed dataset by dataset; four degeneration comparisons formed the discovery set, and two non-degeneration contrasts (†) were analysed separately. Three independent cohorts formed the validation set (* GSE23130 samples 16–23; samples 1–15 duplicate GSE15227).

The degeneration datasets differ in their design (Table 1). In GSE27494, all samples were annulus cells cultured in three dimensions from four donors, each with a control and an IL-1β-treated culture, so the grade IV versus grade III comparison contrasts three donors with one. In GSE147383, AF and NP samples were obtained from the same two

degenerated and two non-degenerated donors. GSE17782 contrasts annulus disc cells with adipose-derived mesenchymal stem cells (a cell-type comparison), and GSE132958 compares TNFR1-knockdown with non-targeting control NP cells from a single donor, each with and without TNF-α (an intervention comparison); these two series were analysed separately as supportive datasets and excluded from all meta-analyses. GSE70362 contains AF and NP from 24 discs from 18 donors (Thompson grades I–V) processed in three batches.

**Table 1.** Datasets analysed in this study and the comparison used for each.

| S. No. | GEO Accession | Description | Comparison (Test vs. Reference) | Platform (Type) |
|---|---|---|---|---|
| 1. | GSE15227 | Discovery. Asporin is expressed at higher levels in more-degenerate human intervertebral disc tissue (annulus). | Grade IV (n = 3) vs. grade II–III (n = 12) | GPL1352—Affymetrix Human X3P (microarray) |
| 2. | GSE17782 | Supportive. Human mesenchymal stem cells with and without TGF-β vs. human annulus disc cells. | Annulus disc cells (n = 7) vs. adipose-derived MSCs (n = 6); cell-type contrast | GPL1352—Affymetrix Human X3P (microarray) |
| 3. | GSE27494 | Discovery. Annulus cells cultured in 3D from four donors, each with a control and an IL-1β-treated culture. | Grade IV (3 donors; n = 6) vs. grade III (1 donor; n = 2); IL-1β as covariate | GPL1352—Affymetrix Human X3P (microarray) |
| 4. | GSE132958 | Supportive. Nucleus pulposus cells from one donor: TNFR1 knockdown and non-targeting control, each with and without TNF-α. | TNFR1 knockdown (n = 4) vs. non-targeting control (n = 4); intervention contrast | GPL16791—Illumina HiSeq 2500 (RNA-seq) |
| 5. | GSE147383 | Discovery. Annulus fibrosus and nucleus pulposus tissue from degenerated and non-degenerated discs. | Degenerated (2 donors; AF + NP, n = 4) vs. non-degenerated (2 donors; n = 4); tissue as covariate | GPL570—Affymetrix U133 Plus 2.0 (microarray) |
| 6. | GSE167199 | Discovery. The miRNA, lncRNA and mRNA profiles in intervertebral disc degeneration (nucleus pulposus; mRNA analysed). | IVDD (n = 3; age 57–65) vs. control (n = 3; age 29–35) | GPL24676—Illumina NovaSeq 6000 (RNA-seq) |
| 7. | GSE70362 | Validation. Annulus fibrosus and nucleus pulposus from 24 discs from 18 donors (Thompson I–V). | Grade IV–V (n = 20) vs. I–III (n = 28); tissue and batch as covariates | GPL17810—Affymetrix HuGene 2.0 ST (microarray) |
| 8. | GSE23130 | Validation. Annulus tissue; samples 16–23 only (samples 1–15 are the GSE15227 specimens). | Grade IV–V (n = 5) vs. II–III (n = 3) | GPL1352—Affymetrix Human X3P (microarray) |
| 9. | GSE34095 | Validation. Disc tissue from patients with degenerative disc disease and adolescents with scoliosis. | Degenerative (n = 3; age 49–54) vs. scoliosis (n = 3; age 11–16) | GPL96—Affymetrix HG-U133A (microarray) |

### 2.2. Sample Uniqueness

Because GSE15227 and GSE23130 were generated by the same laboratory on the same platform, we tested for shared samples by computing Pearson correlations between all pairs of samples across the two series, using probes with mean log2 expression above 8 in both, and compared each sample's best cross-series match with the strongest within-series correlation. To gauge how often the two accessions are used together, PubMed Central was searched for articles mentioning both accession numbers.

### 2.3. Differential Expression

Each dataset was analysed separately. For microarrays, a limma-equivalent moderated t-statistic (linear model with empirical Bayes variance moderation) [20] was applied to the submitter-normalised series-matrix values, log2-transformed where necessary, as in GEO2R; RMA was not re-applied, and for each gene the probe with the smallest *p*-value was retained. For RNA-seq, DESeq2 (Wald test) [21] was applied to the NCBI-generated raw counts (GRCh38.p13). Where the design required it, covariates were added: tissue (AF/NP) for GSE147383; IL-1β treatment for GSE27494; tissue and batch for GSE70362. Degeneration was defined as Thompson grade IV–V versus I–III in GSE70362 and GSE23130, consistent with grade IV versus II–III in GSE15227. For GSE23130, only samples 16–23 were used (Section 3.1). In GSE70362, NP and AF were also analysed separately, together with a degeneration-by-tissue interaction and a linear trend across grades. *p*-Values were adjusted by the Benjamini–Hochberg false discovery rate (FDR). Group assignments, covariates *and* library sizes are listed in Supplementary Table S2, with complete per-dataset results in Supplementary Table S3.

### 2.4. Audit of the Original Candidate Genes

The original pipeline defined DEGs by |log2 fold change (log2FC)| ≥ 2 and adjusted $p \leq 0.05$, retained genes dysregulated in at least three of the six datasets, and ranked them by degree centrality in STRING networks visualised in Cytoscape v3.10.1, using cytoHubba [22] and a degree threshold of 10 [23]. For each of the 33 resulting candidates, we tabulated the per-dataset log2FC and FDR and counted the datasets meeting the primary threshold and two relaxed thresholds (|log2FC| ≥ 1 with FDR ≤ 0.05 or ≤0.10) in the reported direction. We then re-ran the overlap step genome-wide on the six datasets and compared the number of genes passing it with the number expected by chance, obtained by permuting each dataset's DEG calls across genes 500 times.

### 2.5. Meta-Analysis, Replication and Robustness

Per-dataset two-sided *p*-values were converted to signed Z-scores and combined with a weighted Stouffer method [24], with weights equal to the square root of the number of independent donors. The discovery meta-analysis combined GSE15227, GSE27494, GSE147383 and GSE167199 (genes measured in all four); the validation meta-analysis combined GSE70362, GSE23130 (samples 16–23) and GSE34095 (genes measured in at least two). Combined *p*-values were FDR-adjusted genome-wide. A discovery gene (FDR ≤ 0.05) was considered replicated if its one-sided validation *p*-value in the discovery direction remained significant after FDR adjustment across all testable discovery genes, and robust if it also remained significant in each leave-one-dataset-out discovery meta-analysis. Overall sign concordance was tested with a binomial test. A signature score (mean standardised expression of the 50 most strongly upregulated minus the 50 most strongly downregulated discovery genes) was evaluated in each validation cohort by the area under the receiver operating characteristic curve (AUC), with a permutation test against 1000 random gene sets. Genome-wide agreement between datasets was measured as the Spearman correlation

of signed Z-scores. Statistical power was summarised as the minimum log2FC detectable with 80% power at a two-sided α of 0.001, using each gene's moderated standard deviation.

### *2.6. Pathway and Network Analyses*

Pre-ranked gene set enrichment analysis (GSEA) [25] was performed with GSEApy [26] on the discovery and validation meta-analysis Z-scores and on each dataset separately, using the MSigDB Hallmark [27] and Reactome collections (gene sets of 15–500 genes; 1000 permutations). A pathway was considered replicated if it reached FDR ≤ 0.05 in the same direction in both meta-analyses; robustness was assessed by leave-one-dataset-out discovery analyses and by sign consistency across datasets. Functional enrichment of the two candidate lists was computed with the STRING v12.5 enrichment module against the whole genome [28]. The 33 candidates were combined into one STRING network (all evidence channels; confidence ≥ 0.7, with ≥0.4 as a sensitivity analysis); densely connected cores were identified by k-core decomposition, the principle underlying MCODE [29], and modules by Louvain community detection [30]. Analyses were performed in Python 3 (NumPy, SciPy, pandas, PyDESeq2, GSEApy and NetworkX); the scripts are provided as Supplementary Materials.

## 3. Results

### *3.1. Duplicate Samples Across GEO Series*

Samples 1–15 of GSE23130 were the specimens of GSE15227. Each GSE15227 sample's best match in GSE23130 was the identically titled sample ("Disc tissue 1–15"), in the same order (r = 0.89–0.96, against a median between-sample correlation of 0.64 within GSE15227), and the Thompson grade agreed for 14 of the 15 pairs (Supplementary Table S4). Only samples 16–23 of GSE23130 (five grade IV–V and three grade II–III) were therefore used for validation. Twenty-three full-text articles in PubMed Central mention both accessions (Supplementary Table S5); we did not assess whether each treated them as independent, but the overlap creates a clear risk of double counting.

### *3.2. The Original Candidate Genes Were Not Consistently Dysregulated*

The number of DEGs at the primary threshold varied widely, from 15 and 19 in GSE147383 and GSE27494 to 691 in GSE167199 and 1371 in the cell-type contrast GSE17782 (Supplementary Table S6). None of the 33 candidates met the primary threshold in its reported direction in three or more datasets; the maximum was two (COL9A3, COL5A3 and CTSS), each time including a supportive dataset, and no candidate did so in more than one degeneration dataset (Supplementary Table S7). Relaxed thresholds raised the maximum to three of six datasets (COL5A3) and two of four degeneration datasets. Directions frequently disagreed: COL1A2 and COL3A1, reported as upregulated, were higher in GSE15227 but lower in GSE147383, GSE167199 and GSE17782, and MMP9, reported as downregulated, was strongly upregulated in GSE167199 (log2FC 9.11). The downregulation of MMP3 and MMP9 came from GSE17782 and GSE132958, where it is an expected consequence of the contrast (in GSE132958, TNF-α exposure increased MMP3, log2FC 5.16), and from no degeneration dataset.

When the overlap step was re-run genome-wide, 14 genes passed it in the same direction at the primary threshold, against 0.2 expected by chance (permutation $p = 0.002$; Figure 2A; Supplementary Table S8). None was among the 33 candidates. Thirteen of the 14 were downregulated in GSE167199, GSE17782 and GSE132958, and most were interferon-stimulated or inflammatory genes (ISG15, MX2, OAS2, RSAD2, IFI27, IL1B, IL32, MMP1 and KYNU; Figure 2B). The procedure therefore selected genes driven by two non-degeneration contrasts and one small tissue dataset, rather than a degeneration signature.

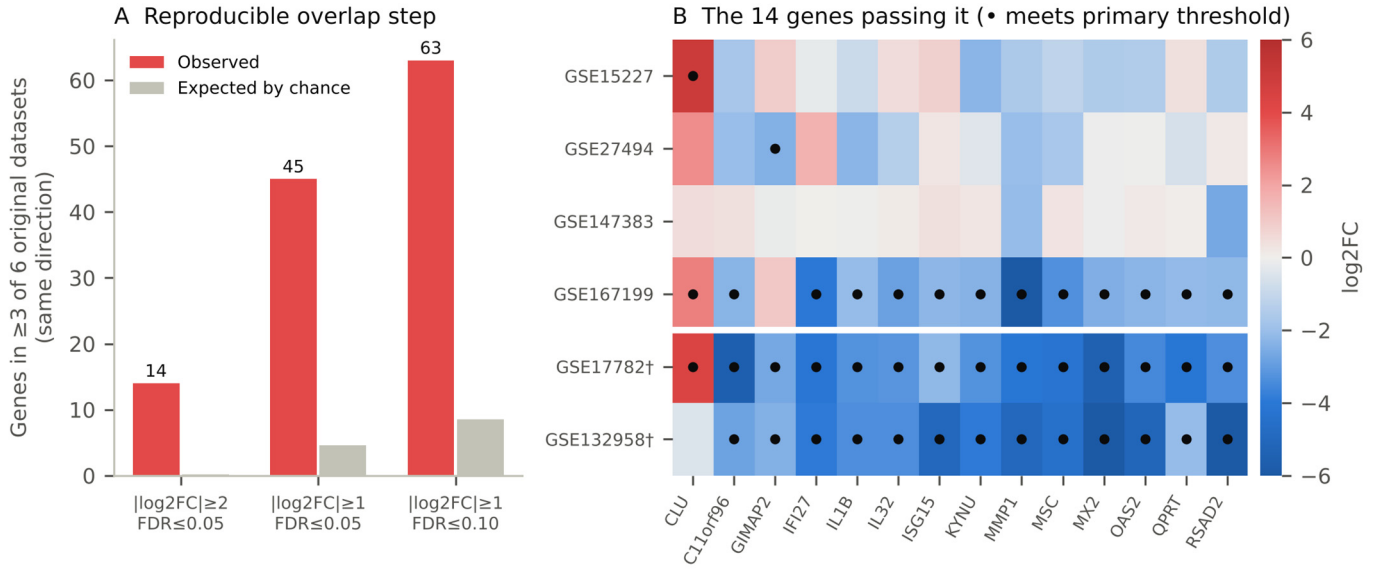


**Figure 2.** Audit of the overlap step. (**A**) Genes passing the "dysregulated in at least three of six datasets" rule in the same direction, observed versus expected by chance (500 permutations), at three cut-offs. (**B**) log2FC of the 14 genes passing the rule at the primary cut-off in the six original datasets; dots mark the primary threshold ($|\log2FC| \geq 2$, FDR $\leq 0.05$). † Supportive non-degeneration contrasts.

### *3.3. Genome-Wide Agreement Between Datasets Was Negligible*

Signed gene-level Z-scores were barely correlated between datasets (Figure 3; Supplementary Table S9). Across the 21 pairs of degeneration datasets, Spearman's ρ ranged from −0.13 to 0.13 (median 0.03); GSE167199 was negatively correlated with GSE70362 and GSE147383, and GSE15227 was negatively correlated with the non-overlapping GSE23130 samples from the same laboratory. By contrast, within GSE70362 the degeneration effects estimated separately in NP and AF were moderately correlated ($\rho = 0.56$), and no gene showed a significant degeneration-by-tissue interaction (FDR $\leq 0.05$), indicating that within-study effects are largely shared between compartments while between-study agreement is near zero. Limited power did not obviously explain this: in the five datasets assessed, the median minimum detectable log2FC ranged from 0.26 to 0.78, and at least 94% of genes had a minimum detectable log2FC of 2 or less in all but the eight-sample GSE23130 subset (78%; Supplementary Table S6).

### *3.4. Discovery Signatures Did Not Replicate*

The discovery meta-analysis identified 1252 genes at FDR $\leq 0.05$ (639 up, 613 down; 1192 without design adjustment). Of the 1232 that could be tested in validation, 603 (49%) changed in the same direction, no more than expected by chance (binomial $p = 0.78$; Figure 4A). Fifteen genes replicated at a replication FDR $\leq 0.05$, and only two, ABHD2 and the cell-cycle inhibitor CDKN2C, remained significant in every leave-one-dataset-out discovery analysis (Supplementary Table S10). A score built from the 50 most up- and downregulated discovery genes did not discriminate degenerated discs in validation: in GSE70362 it was, if anything, inversely related to degeneration (AUC 0.33, $p = 0.044$; ρ with grade −0.38; permutation $p$ versus random gene sets 0.11), and the AUCs were 0.40 and 0.33 in GSE23130 and GSE34095.

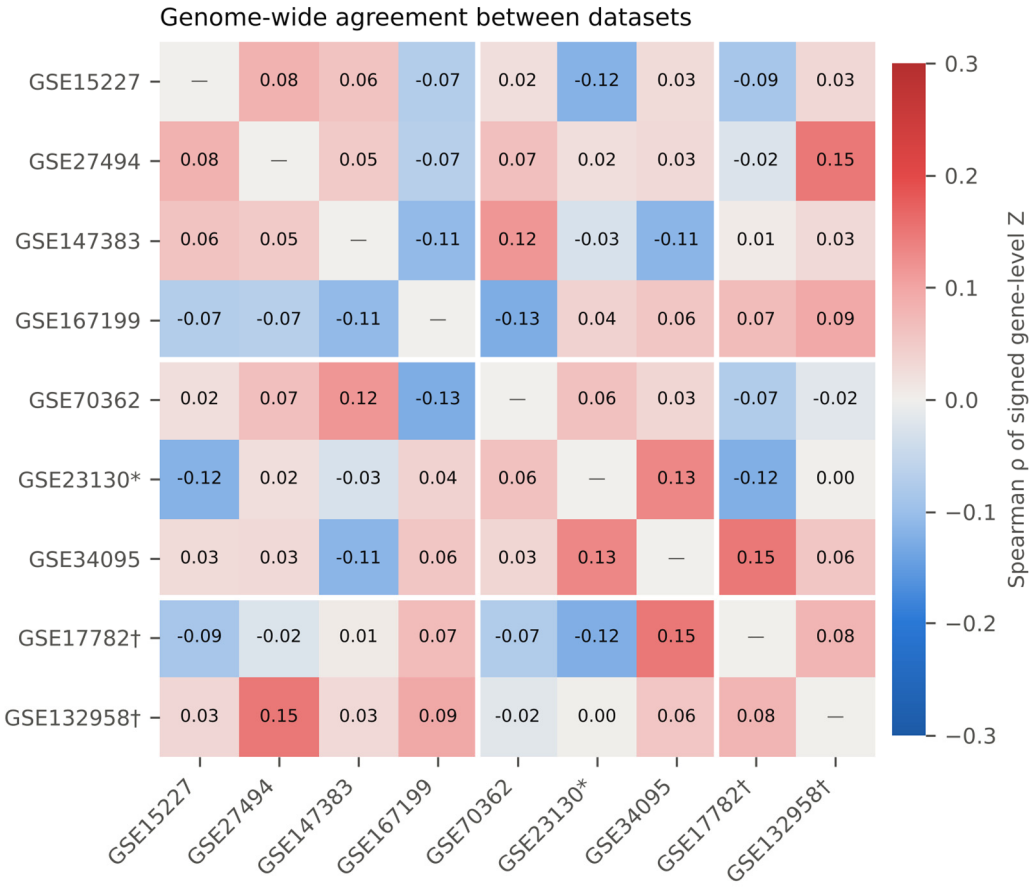


**Figure 3.** Genome-wide agreement between datasets. Spearman correlation of signed gene-level Z-scores for each pair of datasets: discovery (left block), validation (middle) and supportive non-degeneration contrasts (right; †). * GSE23130 samples 16–23 only.

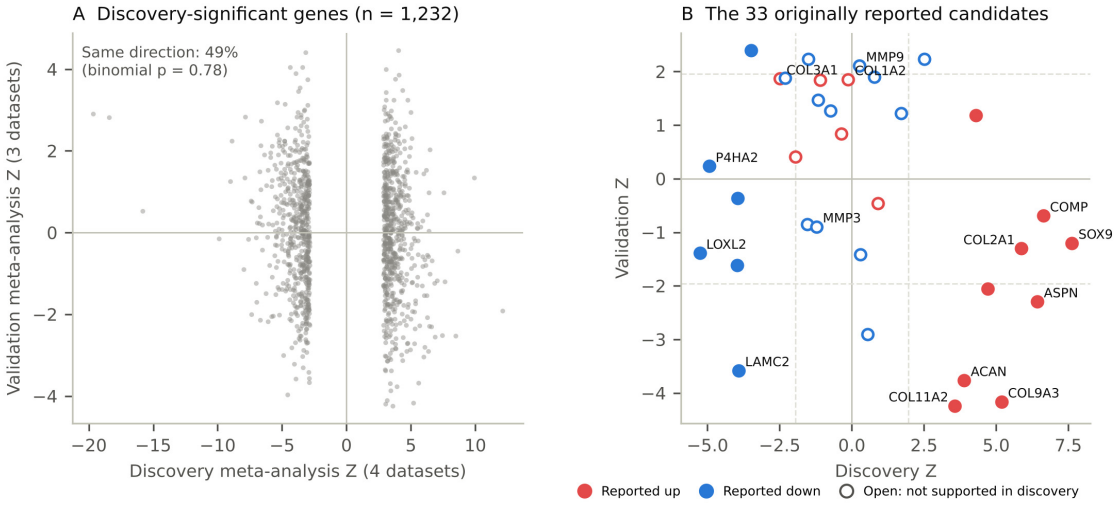


**Figure 4.** Replication of discovery signals. (**A**) Discovery versus validation meta-analysis Z-scores for the 1232 discovery-significant genes that could be tested in validation; points are spread evenly across concordant and discordant quadrants. (**B**) The 33 originally reported candidates (red, reported upregulated; blue, reported downregulated; filled, supported in the discovery meta-analysis; dashed lines, $|Z| = 1.96$).

Of the 33 candidates, 15 were significant in the discovery meta-analysis in their reported direction (COL2A1, COL9A3, COL11A1, COL11A2, P3H2, COMP, ASPN, SOX9 and ACAN up; COL4A5, P4HA2, COL4A1, COL6A6, LOXL2 and LAMC2 down), but only LAMC2

replicated, and it was not robust to leave-one-out analysis (Table 2; Figure 4B). Three upregulated candidates changed significantly in the opposite direction in validation (ACAN, COL9A3 and COL11A2; validation FDR 0.024–0.040), with consistent downregulation in the largest cohort, GSE70362 (log2FC −0.81, −0.82 and −1.21). MMP3 and MMP9 were not significant in either meta-analysis.

**Table 2.** The 33 originally reported candidate genes in the discovery and validation meta-analyses.

| Gene | Reported | Discovery Z | Discovery FDR | Supported | Validation Z | Validation FDR | GSE70362 log2FC | Replicated |
|---|---|---|---|---|---|---|---|---|
| COL1A2 | Up | −0.12 | 0.961 | No | 1.85 | 0.411 | 0.22 | — |
| COL2A1 | Up | 5.87 | $8.7 \times 10^{-7}$ | Yes | −1.30 | 0.622 | −0.61 | No |
| COL3A1 | Up | −2.49 | 0.098 | No | 1.87 | 0.407 | 0.04 | — |
| COL9A3 | Up | 5.18 | $2.4 \times 10^{-5}$ | Yes | −4.16 | 0.026 | −0.82 | No |
| COL11A1 | Up | 4.31 | $8.3 \times 10^{-4}$ | Yes | 1.19 | 0.665 | 0.15 | No |
| COL11A2 | Up | 3.57 | 0.009 | Yes | −4.24 | 0.024 | −1.21 | No |
| COL8A1 | Up | −0.37 | 0.867 | No | 0.84 | 0.787 | — | — |
| COL14A1 | Up | −1.95 | 0.222 | No | 0.41 | 0.909 | 0.05 | — |
| PLOD2 | Up | −1.10 | 0.548 | No | 1.84 | 0.417 | 0.17 | — |
| COL25A1 | Up | 0.90 | 0.638 | No | −0.46 | 0.899 | 0.00 | — |
| P3H2 | Up | 4.71 | $1.8 \times 10^{-4}$ | Yes | −2.05 | 0.347 | −0.29 | No |
| COMP | Up | 6.64 | $1.3 \times 10^{-8}$ | Yes | −0.69 | 0.836 | −0.00 | No |
| ASPN | Up | 6.43 | $4.5 \times 10^{-8}$ | Yes | −2.29 | 0.272 | −1.79 | No |
| SOX9 | Up | 7.63 | $2.4 \times 10^{-11}$ | Yes | −1.20 | 0.659 | −0.06 | No |
| ACAN | Up | 3.89 | 0.003 | Yes | −3.76 | 0.040 | −0.81 | No |
| CTSS | Down | 0.78 | 0.696 | No | 1.90 | 0.397 | 0.20 | — |
| COL15A1 | Down | 1.70 | 0.304 | No | 1.22 | 0.651 | 0.06 | — |
| BMP1 | Down | −2.32 | 0.130 | No | 1.88 | 0.402 | 0.05 | — |
| COL4A2 | Down | −1.16 | 0.521 | No | 1.47 | 0.561 | 0.53 | — |
| COL4A5 | Down | −3.96 | 0.003 | Yes | −0.37 | 0.924 | −0.01 | No |
| P4HA2 | Down | −4.94 | $7.0 \times 10^{-5}$ | Yes | 0.23 | 0.955 | −0.06 | No |
| COL1A1 | Down | 2.52 | 0.094 | No | 2.23 | 0.289 | 0.50 | — |
| CTSB | Down | −0.73 | 0.715 | No | 1.27 | 0.633 | 0.13 | — |
| COL18A1 | Down | 0.30 | 0.895 | No | −1.41 | 0.583 | −0.10 | — |
| COL4A4 | Down | −1.54 | 0.368 | No | −0.85 | 0.786 | 0.01 | — |
| COL6A2 | Down | 0.54 | 0.798 | No | −2.91 | 0.131 | −0.59 | — |
| COL4A1 | Down | −3.49 | 0.011 | Yes | 2.39 | 0.245 | 0.78 | No |
| COL6A6 | Down | −3.98 | 0.003 | Yes | −1.62 | 0.503 | −0.09 | No |
| LOXL2 | Down | −5.27 | $1.7 \times 10^{-5}$ | Yes | −1.39 | 0.592 | −0.33 | No |
| LAMC2 | Down | −3.92 | 0.003 | Yes | −3.58 | 0.056 | −0.52 | Yes |
| MMP9 | Down | 0.27 | 0.907 | No | 2.11 | 0.327 | 0.24 | — |
| COL5A3 | Down | −1.50 | 0.383 | No | 2.23 | 0.289 | 0.17 | — |
| MMP3 | Down | −1.23 | 0.495 | No | −0.90 | 0.770 | −0.03 | — |

Note: Z, weighted Stouffer Z-score (positive, higher in degenerated discs); FDR, genome-wide Benjamini–Hochberg adjusted *p*. Discovery: GSE15227, GSE27494, GSE147383, GSE167199. Validation: GSE70362, GSE23130 (samples 16–23), GSE34095. Supported: discovery FDR ≤ 0.05 in the reported direction. Replicated: replication FDR ≤ 0.05 in the discovery direction. — Not applicable or not measured. Per-dataset values are given in Supplementary Table S7.

### *3.5. A Proliferation Programme Was the Most Consistent Pathway-Level Signal*

At the pathway level, 18 Hallmark gene sets were enriched in the discovery and 25 in the validation meta-analysis (FDR ≤ 0.05), of which five agreed in direction: E2F targets, G2-M checkpoint, Myc targets V1 and oxidative phosphorylation (positive) and hypoxia (negative) (Table 3; Figure 5; Supplementary Table S11). Only the two proliferation sets were consistent across datasets: E2F targets and G2-M checkpoint had positive enrichment scores in all nine datasets, including the supportive contrasts, and remained significant in three of four leave-one-out analyses (FDR 0.054 and 0.17 when GSE167199 was omitted). Their shared leading-edge genes included BIRC5, AURKB, BUB1B, CDC25A, CDKN2A, CDKN2C, MCM3, MCM7, CENPE and NUSAP1. Hypoxia and oxidative phosphorylation reversed

direction when GSE167199 was omitted, and Myc targets lost significance without GSE15227. Ten Reactome pathways replicated, six of which describe mitosis (M phase, mitotic anaphase, separation of sister chromatids). At the sample level, however, a proliferation score did not discriminate grade IV–V from grade I–III discs in GSE70362 (AUC 0.58, $p = 0.36$; Supplementary Table S11).

**Table 3.** Hallmark gene sets significant in both the discovery and validation meta-analyses.

| Hallmark Gene Set | Discovery NES | Discovery FDR | Validation NES | Validation FDR | Datasets with NES > 0 (of 9) | Leave-One-Out (of 4) |
|---|---|---|---|---|---|---|
| E2F Targets | 2.10 | <0.001 | 2.19 | <0.001 | 9 | 3 |
| G2-M Checkpoint | 1.80 | 0.003 | 2.29 | <0.001 | 9 | 3 |
| Myc Targets V1 | 1.86 | 0.006 | 2.12 | <0.001 | 6 | 3 |
| Oxidative Phosphorylation | 1.71 | 0.006 | 1.42 | 0.035 | 7 | 2 |
| Hypoxia | −2.07 | $3.4 \times 10^{-4}$ | −1.50 | 0.028 | 3 | 3 |

Note: NES, normalised enrichment score from pre-ranked GSEA of meta-analysis Z-scores; FDR, GSEA false discovery rate. Datasets with positive NES, out of nine (seven degeneration datasets plus two supportive contrasts). Leave-one-out: number of the four leave-one-dataset-out discovery analyses with FDR ≤ 0.05 in the same direction.

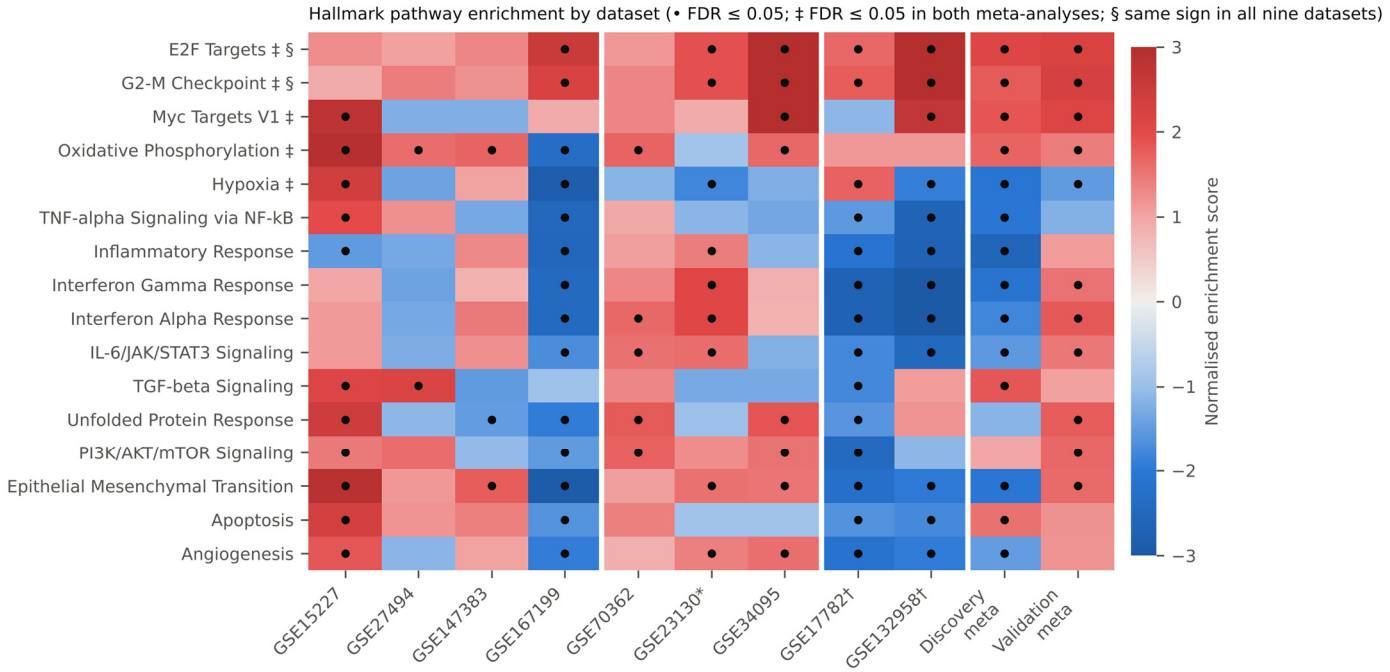


**Figure 5.** Hallmark pathway enrichment by dataset and in the two meta-analyses. Colour shows the normalised enrichment score; dots mark FDR ≤ 0.05. ‡ FDR ≤ 0.05 in the same direction in both meta-analyses; § same sign in all nine datasets. † Supportive non-degeneration contrasts; * GSE23130 samples 16–23.

Pathways often invoked in IVDD did not replicate. Inflammatory response, TNF-α signalling via NF-κB and interferon-γ response were depleted in the discovery meta-analysis, driven by GSE167199, but were not depleted in validation, where the interferon-γ response was enriched instead. TGF-β signalling was enriched in discovery only (NES 1.85), and the unfolded protein response and PI3K/AKT/mTOR signalling were enriched in validation only.

### *3.6. Functional Context and Interaction of the Candidate Genes*

The upregulated candidates were enriched for collagen fibril organisation, cartilage development and ECM structural constituents, and the downregulated candidates for ECM organisation, collagen catabolism, basement membrane and collagen type IV components; both lists were enriched for ECM–receptor interaction and PI3K/Akt signalling, largely reflecting their collagen-rich composition (Supplementary Table S12). In a combined high-confidence STRING network, 31 of the 33 candidates formed 88 edges, 36 (41%) of which linked an up- with a downregulated candidate, around a fibrillar collagen six-core (COL1A1, COL1A2, COL2A1, COL3A1, COL5A3, COL11A1 and COL11A2); MMP9 had the second-highest betweenness, bridging a cartilage matrix–protease module (ACAN, COMP, COL9A3, COL14A1, MMP3, MMP9, CTSB and CTSS) to the collagen modules (Figure 6; Supplementary Table S13). The network shows that the candidates describe one interconnected ECM system; because the individual genes did not replicate, it describes known biology rather than a disease signature.

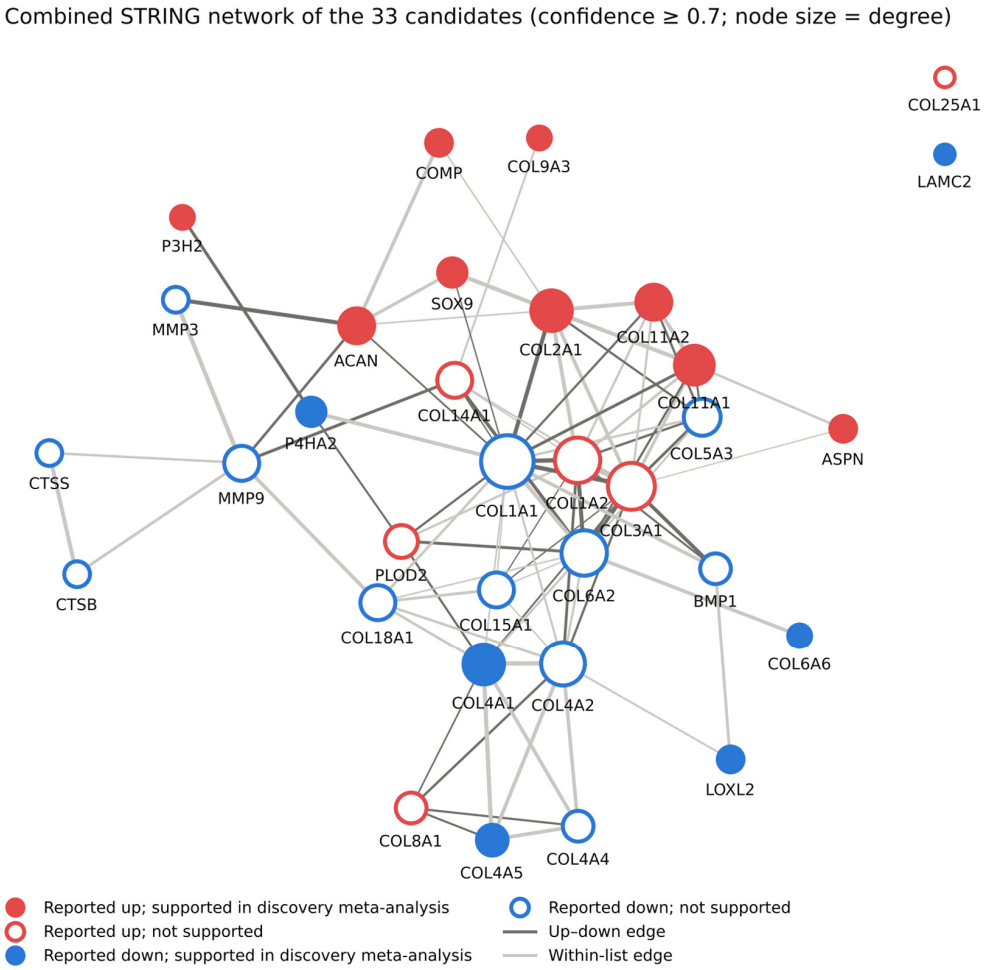


**Figure 6.** The combined STRING network of the 33 originally reported candidates (confidence ≥ 0.7). Node size is proportional to degree; red, reported upregulated; blue, reported downregulated; filled, supported in the discovery meta-analysis; dark edges link up- and downregulated candidates. No candidate was robust in validation.

## 4. Discussion

### *4.1. Principal Findings*

Using a discovery–validation design, we found that gene-level transcriptomic signatures of IVDD derived from public datasets were not reproduced. Genome-wide agreement

between degeneration datasets was close to zero, a 1252-gene discovery signature changed in the validation cohorts no more consistently than chance, and none of the 33 candidates prioritised by an overlap-and-network pipeline replicated robustly; three changed significantly in the opposite direction. We also identified two avoidable sources of error that are probably common: the same specimens deposited under two GEO accessions (GSE15227 and GSE23130), and the inclusion of cell-type and intervention contrasts as if they were degeneration comparisons. The latter produced an apparent downregulation of MMPs and an interferon-dominated overlap signature. The most consistent pathway-level signal across the datasets examined was a proliferation programme, detected in every dataset.

### *4.2. Why Public IVDD Transcriptomes Disagree*

Several factors probably contribute. The comparisons differ in what they contrast: grade IV versus II–III within surgical AF tissue, degenerated surgical NP versus NP from younger control or scoliosis patients, and cultured cells. Age is confounded with degeneration in several series (for example, controls aged 29–35 versus patients aged 57–65 in GSE167199, and adolescents versus adults in GSE34095), so "degeneration" signatures partly reflect age, diagnosis and surgical indication. Groups are very small, and samples within a series are not always independent (AF and NP from the same donors in GSE147383 and GSE70362; one grade III donor in GSE27494). Platforms, normalisation and batches differ, and batch effects can dominate biological signals in high-throughput data [19]. Similar problems have been documented for microarray studies in general, where many published analyses could not be reproduced from the deposited data [18]. The moderate agreement between NP and AF within GSE70362, set against near-zero agreement between studies, suggests that study-level factors outweigh compartment-level biology.

### *4.3. Implications for Hub Gene and Biomarker Studies*

Our audit shows how an overlap-and-network pipeline can produce apparently coherent but non-reproducible candidates. Requiring a gene to be dysregulated in several datasets does not ensure consistency if direction is not required, if non-degeneration contrasts are included, or if datasets share samples; in our data, the reproducible version of the step selected genes driven by two non-degeneration contrasts. Network centrality then adds coherence rather than evidence: collagens are densely interconnected in STRING whatever their direction of change, and in a small, dense network every node can exceed a degree threshold. Translating such candidates into therapeutic targets, as our original manuscript did, is not justified without independent validation. We therefore did not perform drug repurposing or connectivity map analyses, because querying a non-reproducible signature would compound the problem. The previously described upregulation of asporin in more-degenerate disc tissue [31] and interferon signalling in degenerative annuli [32] were reported from GSE15227 and GSE70362 themselves; our results indicate that such single-series findings require replication in independent cohorts.

### *4.4. A Proliferation Programme in Degenerated Discs*

The E2F target and G2-M checkpoint gene sets were the only signals that were enriched in both meta-analyses and positive in every dataset. This agrees with histological evidence that the cell clusters characteristic of degenerated discs arise from increased disc cell proliferation [33], and with the coexistence of proliferation and senescence in degenerate discs: the replicated leading-edge genes included the cyclin-dependent kinase inhibitors CDKN2A (p16INK4a), a marker of the accelerated senescence described in degenerated discs [12], and CDKN2C, one of the two robustly replicated genes. Because the supportive cell-culture contrasts were also positive, part of this signal may reflect proliferative states common

to disc cells rather than degeneration specifically, and a sample-level proliferation score did not discriminate grade in GSE70362. The signal should therefore be interpreted as a reproducible direction of change, not as a biomarker. The reduced hypoxia signature was not robust, which is notable given the central role of hypoxia-inducible factor signalling in NP cell metabolism [34] and the possibility that annular fissures and vascular ingrowth alter disc oxygenation; it warrants testing in better-matched cohorts.

### *4.5. Inflammation, ECM Genes and Genetic Susceptibility*

Inflammatory cytokines, NF-κB signalling and ECM catabolism are well established in IVDD [35–38], and obesity-related adipokines, low-grade infection and macrophages have been proposed as contributors [39–43]. Yet inflammatory and TNF-α/NF-κB gene sets did not replicate in our analysis; in discovery they were depleted, driven by one dataset. Inflammatory activity in IVDD is probably focal, time-dependent and partly protein-level, and bulk transcriptomes of end-stage surgical tissue may not capture it reproducibly. Similarly, collagen and aggrecan genes are structurally central [44,45], and polymorphisms in collagen genes and the aggrecan VNTR are associated with susceptibility [13,46,47], but their expression changes were inconsistent across datasets. Genetic association and expression change are distinct lines of evidence and should not be conflated.

### *4.6. Recommendations for Secondary Analyses of IVDD Transcriptomes*

On the basis of these findings, *we suggest that secondary analyses of public IVDD datasets: (i) verify that samples are unique across series before combining them; (ii) restrict degeneration signatures to comparisons of degenerated discs with control discs* and analyse cell-type or intervention contrasts separately; (iii) model known covariates such as tissue compartment, batch, treatment and donor, and report age for both groups; (iv) report the direction and magnitude of change for each candidate in each dataset; (v) combine evidence with direction-aware meta-analysis rather than counting overlaps; (vi) test candidates in independent cohorts before claiming biomarkers or targets; and (vii) deposit code and complete results.

### *4.7. Limitations*

This study has limitations. The validation evidence rests mainly on GSE70362; the other two validation cohorts are small, and their non-significance per dataset is uninformative on its own. GSE70362 contains two tissues and up to two discs per donor; we adjusted for tissue and batch but could not fully model donor correlation, and only three donors had discs of discordant degeneration status, too few for a within-donor analysis. Grading systems and degeneration cut-offs differ between series. Our re-analysis used a re-implementation of the GEO2R procedures and may differ in detail from the original analyses, which may explain why the original selection of the 33 candidates could not be reproduced. Bulk transcriptomes cannot resolve cell-type composition, and single-cell datasets were not analysed. Finally, no experimental validation was performed; the proliferation finding is a hypothesis for direct testing, for example by Ki-67 and p16 immunohistochemistry across Thompson grades in independent tissue.

## 5. Conclusions

Gene-level transcriptomic signatures and overlap-based hub genes in IVDD were not reproducible across public datasets, and part of their apparent consistency arose from duplicated samples and non-degeneration contrasts. The 33 collagen-dominated candidates from our original analysis did not replicate and should not be considered therapeutic targets. Across the datasets examined, the most consistent pathway-level signal was a cell proliferation programme; given the limitations of the available datasets, it should be

regarded as a hypothesis for experimental validation rather than an established feature of disc degeneration.

Future secondary analyses should verify sample uniqueness, separate degeneration from other contrasts, report per-dataset effects and test candidates in independent cohorts. Larger, age-matched, well-annotated cohorts with compartment-resolved and single-cell data, together with experimental validation of the proliferation programme, are needed before transcriptomic findings can guide therapy for IVDD.

**Supplementary Materials:** The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/cimb48111236/s1.

**Author Contributions:** S.M. (Sridhar Muthusami): Project administration, formal analysis, data curation, supervision, investigation, and original draft preparation. K.R.: Software, methodology, formal analysis, writing—original draft, and editing. S.M. (Sathish Muthu): Conceptualisation, methodology, validation, software, formal analysis, investigation, resources, data curation, supervision, project administration, and original draft preparation. G.C.: Software, formal analysis, visualisation, writing—review and editing. Y.M.: Data curation and literature collection. K.S.: Visualisation, writing—review and editing. B.M.: Resources, investigation, project administration, and supervision. R.L.R.: Validation, data interpretation, visualisation, writing—review and editing. P.G.: Conceptualisation, supervision, data interpretation, funding acquisition, writing—review and editing. B.-C.A.: Conceptualisation, supervision, project administration, funding acquisition, and writing—review and editing. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** As all human data were obtained from publicly accessible databases and fully deidentified, in accordance with the National Ethical Guidelines for Biomedical and Health Research Involving Human Participants (Indian Council of Medical Research, 2017), this study was exempted from ethics committee review.

**Informed Consent Statement:** Informed consent was waived for this study as all participant data were retrieved from publicly available databases and completely anonymised, in accordance with the National Ethical Guidelines for Biomedical and Health Research Involving Human Participants (Indian Council of Medical Research, 2017).

**Data Availability Statement:** All datasets analysed are publicly available in the NCBI Gene Expression Omnibus under accession numbers GSE15227, GSE17782, GSE27494, GSE132958, GSE147383, GSE167199, GSE70362, GSE23130 and GSE34095. Sample assignments; complete per-dataset differential expression results; meta-analysis, pathway and network results; the duplicate-sample map; and the literature search are provided in Supplementary Tables S1–S13.

**Acknowledgments:** The authors gratefully acknowledge the Karpagam Academy of Higher Education, Coimbatore, India and the Department of Science and Technology, India, for their research support. During preparation of the original manuscript, Wordvice AI and ChatGPT (GPT-5.5, OpenAI) were used under author supervision to improve English expression, clarify structure and adjust the word count. The authors checked the edited version and take full responsibility for the final manuscript.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. GBD 2021 Low Back Pain Collaborators. Global, regional, and national burden of low back pain, 1990–2020, its attributable risk factors, and projections to 2050: A systematic analysis of the Global Burden of Disease Study 2021. *Lancet Rheumatol.* **2023**, *5*, e316–e329. [CrossRef] [PubMed]
2. Ravindra, V.M.; Senglaub, S.S.; Rattani, A.; Dewan, M.C.; Härtl, R.; Bisson, E.; Park, K.B.; Shrime, M.G. Degenerative Lumbar Spine Disease: Estimating Global Incidence and Worldwide Volume. *Glob. Spine J.* **2018**, *8*, 784–794. [CrossRef] [PubMed]
3. Teraguchi, M.; Yoshimura, N.; Hashizume, H.; Muraki, S.; Yamada, H.; Minamide, A.; Oka, H.; Ishimoto, Y.; Nagata, K.; Kagotani, R.; et al. Prevalence and distribution of intervertebral disc degeneration over the entire spine in a population-based cohort: The Wakayama Spine Study. *Osteoarthr. Cartil.* **2014**, *22*, 104–110. [CrossRef] [PubMed]

4. Machino, M.; Nakashima, H.; Ito, K.; Tsushima, M.; Ando, K.; Kobayashi, K.; Imagama, S. Influence of Age and Gender on Intervertebral Disk Degeneration and Height in the Thoracolumbar Spine. *Spine Surg. Relat. Res.* **2021**, *6*, 379–387. [CrossRef] [PubMed]
5. Samanta, A.; Lufkin, T.; Kraus, P. Intervertebral disc degeneration—Current therapeutic options and challenges. *Front. Public Health* **2023**, *11*, 1156749. [CrossRef] [PubMed]
6. Izzo, R.; Guarnieri, G.; Guglielmi, G.; Muto, M. Biomechanics of the spine. Part I: Spinal stability. *Eur. J. Radiol.* **2013**, *82*, 118–126. [CrossRef] [PubMed]
7. Muthu, S.; Visawanathan, V.K.; Chellamuthu, G. Mechanical Basis of Lumbar Intervertebral Disk Degeneration. *Indian Spine J.* **2024**, *7*, 121. [CrossRef] [PubMed]
8. Elfering, A.; Semmer, N.; Birkhofer, D.; Boos, N. Risk factors for lumbar disc degeneration: A 5-year prospective MRI study in asymptomatic individuals. *Spine* **2002**, *27*, 125–134. [CrossRef] [PubMed]
9. Zielinska, N.; Podgórski, M.; Haładaj, R.; Polguj, M.; Olewnik, Ł. Risk Factors of Intervertebral Disc Pathology—A Point of View Formerly and Today—A Review. *J. Clin. Med.* **2021**, *10*, 409. [CrossRef] [PubMed]
10. Ząbek, Z.; Wyczałkowska-Tomasik, A.; Poboży, K.; Adamus, J.P.; Turek, G.; Ząbek, M.; Pączek, L. Understanding the Microenvironment of Intervertebral Disc Degeneration: A Comprehensive Review of Pathophysiological Insights and Therapeutic Implications. *Int. J. Mol. Sci.* **2025**, *26*, 9938. [CrossRef] [PubMed]
11. Dowdell, J.; Erwin, M.; Choma, T.; Vaccaro, A.; Iatridis, J.; Cho, S.K. Intervertebral Disk Degeneration and Repair. *Neurosurgery* **2017**, *80*, S46–S54. Erratum in *Neurosurgery* **2018**, *83*, 1084. https://doi.org/10.1093/neuros/nyw078. [CrossRef] [PubMed]
12. Le Maitre, C.L.; Freemont, A.J.; Hoyland, J.A. Accelerated cellular senescence in degenerate intervertebral discs: A possible role in the pathogenesis of intervertebral disc degeneration. *Arthritis Res. Ther.* **2007**, *9*, R45. [CrossRef] [PubMed]
13. Mayer, J.E.; Iatridis, J.C.; Chan, D.; Qureshi, S.A.; Gottesman, O.; Hecht, A.C. Review: Genetic Polymorphisms Associated with Intervertebral Disc Degeneration. *Spine J. Off. J. N. Am. Spine Soc.* **2013**, *13*, 299–317. [CrossRef] [PubMed]
14. S.M., H.; M.K., S.; G.K., C.; Bhat, D.I. Association of *CILP*, *COL9A2* and *MMP3* Gene Polymorphisms with Lumbar Disc Degeneration in an Indian Population. *J. Mol. Neurosci.* **2018**, *66*, 378–382. [CrossRef] [PubMed]
15. Kelempisioti, A.; Eskola, P.J.; Okuloff, A.; Karjalainen, U.; Takatalo, J.; Daavittila, I.; Niinimäki, J.; Sequeiros, R.B.; Tervonen, O.; Solovieva, S.; et al. Genetic susceptibility of intervertebral disc degeneration among young Finnish adults. *BMC Med. Genet.* **2011**, *12*, 153. [CrossRef] [PubMed]
16. Barrett, T.; Wilhite, S.E.; Ledoux, P.; Evangelista, C.; Kim, I.F.; Tomashevsky, M.; Marshall, K.A.; Phillippy, K.H.; Sherman, P.M.; Holko, M.; et al. NCBI GEO: Archive for functional genomics data sets—Update. *Nucleic Acids Res.* **2013**, *41*, D991–D995. [CrossRef] [PubMed]
17. Tan, J.; Shi, M.; Li, B.; Liu, Y.; Luo, S.; Cheng, X. Role of arachidonic acid metabolism in intervertebral disc degeneration: Identification of potential biomarkers and therapeutic targets via multi-omics analysis and artificial intelligence strategies. *Lipids Health Dis.* **2023**, *22*, 204. [CrossRef] [PubMed]
18. Ioannidis, J.P.A.; Allison, D.B.; Ball, C.A.; Coulibaly, I.; Cui, X.; Culhane, A.C.; Falchi, M.; Furlanello, C.; Game, L.; Jurman, G.; et al. Repeatability of published microarray gene expression analyses. *Nat. Genet.* **2009**, *41*, 149–155. [CrossRef] [PubMed]
19. Leek, J.T.; Scharpf, R.B.; Bravo, H.C.; Simcha, D.; Langmead, B.; Johnson, W.E.; Geman, D.; Baggerly, K.; Irizarry, R.A. Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat. Rev. Genet.* **2010**, *11*, 733–739. [CrossRef] [PubMed]
20. Ritchie, M.E.; Phipson, B.; Wu, D.; Hu, Y.; Law, C.W.; Shi, W.; Smyth, G.K. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **2015**, *43*, e47. [CrossRef] [PubMed]
21. Love, M.I.; Huber, W.; Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **2014**, *15*, 550. [CrossRef] [PubMed]
22. Chin, C.H.; Chen, S.H.; Wu, H.H.; Ho, C.-W.; Ko, M.-T.; Lin, C.-Y. CytoHubba: Identifying hub objects and sub-networks from complex interactome. *BMC Syst. Biol.* **2014**, *8*, S11. [CrossRef] [PubMed]
23. Barabási, A.L.; Oltvai, Z.N. Network biology: Understanding the cell's functional organization. *Nat. Rev. Genet.* **2004**, *5*, 101–113. [CrossRef] [PubMed]
24. Whitlock, M.C. Combining probability from independent tests: The weighted Z-method is superior to Fisher's approach. *J. Evol. Biol.* **2005**, *18*, 1368–1373. [CrossRef] [PubMed]
25. Subramanian, A.; Tamayo, P.; Mootha, V.K.; Mukherjee, S.; Ebert, B.L.; Gillette, M.A.; Paulovich, A.; Pomeroy, S.L.; Golub, T.R.; Lander, E.S.; et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 15545–15550. [CrossRef] [PubMed]
26. Fang, Z.; Liu, X.; Peltz, G. GSEApy: A comprehensive package for performing gene set enrichment analysis in Python. *Bioinformatics* **2023**, *39*, btac757. [CrossRef] [PubMed]
27. Liberzon, A.; Birger, C.; Thorvaldsdóttir, H.; Ghandi, M.; Mesirov, J.P.; Tamayo, P. The Molecular Signatures Database Hallmark gene set collection. *Cell Syst.* **2015**, *1*, 417–425. [CrossRef] [PubMed]

28. Szklarczyk, D.; Kirsch, R.; Koutrouli, M.; Nastou, K.; Mehryary, F.; Hachilif, R.; Gable, A.L.; Fang, T.; Doncheva, N.T.; Pyysalo, S.; et al. The STRING database in 2023: Protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res.* **2023**, *51*, D638–D646. [CrossRef] [PubMed]
29. Bader, G.D.; Hogue, C.W. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinform.* **2003**, *4*, 2. [CrossRef] [PubMed]
30. Blondel, V.D.; Guillaume, J.L.; Lambiotte, R.; Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech.* **2008**, *2008*, 10008. [CrossRef]
31. Gruber, H.E.; Ingram, J.A.; Hoelscher, G.L.; Zinchenko, N.; HanleyJr, E.N.; Sun, Y. Asporin, a susceptibility gene in osteoarthritis, is expressed at higher levels in the more degenerate human intervertebral disc. *Arthritis Res. Ther.* **2009**, *11*, R47. [CrossRef] [PubMed]
32. Kazezian, Z.; Gawri, R.; Haglund, L.; Ouellet, J.; Mwale, F.; Tarrant, F.; O'gAora, P.; Pandit, A.; Alini, M.; Grad, S. Gene expression profiling identifies interferon signalling molecules and IGFBP3 in human degenerative annulus fibrosus. *Sci. Rep.* **2015**, *5*, 15662. [CrossRef] [PubMed]
33. Johnson, W.E.B.; Eisenstein, S.M.; Roberts, S. Cell cluster formation in degenerate lumbar intervertebral discs is associated with increased disc cell proliferation. *Connect. Tissue Res.* **2001**, *42*, 197–207. [CrossRef] [PubMed]
34. Risbud, M.V.; Guttapalli, A.; Stokes, D.G.; Hawkins, D.; Danielson, K.G.; Schaer, T.P.; Albert, T.J.; Shapiro, I.M. Nucleus pulposus cells express HIF-1α under normoxic culture conditions: A metabolic adaptation to the intervertebral disc microenvironment. *J. Cell. Biochem.* **2006**, *98*, 152–159. [CrossRef] [PubMed]
35. Zhang, G.; Liu, M.; Chen, H.; Wu, Z.; Gao, Y.; Ma, Z.; He, X.; Kang, X. NF-κB signalling pathways in nucleus pulposus cell function and intervertebral disc degeneration. *Cell Prolif.* **2021**, *54*, e13057. [CrossRef] [PubMed]
36. Guo, Q.; Jin, Y.; Chen, X.; Ye, X.; Shen, X.; Lin, M.; Zeng, C.; Zhou, T.; Zhang, J. NF-κB in biology and targeted therapy: New insights and translational implications. *Signal Transduct. Target Ther.* **2024**, *9*, 53. [CrossRef] [PubMed]
37. Tisherman, R.; Coelho, P.; Phillibert, D.; Wang, D.; Dong, Q.; Vo, N.; Kang, J.; Sowa, G. NF-κB Signaling Pathway in Controlling Intervertebral Disk Cell Response to Inflammatory and Mechanical Stressors. *Phys. Ther.* **2016**, *96*, 704–711. [CrossRef] [PubMed]
38. Rajasekaran, S.; Tangavel, C.; KS, S.V.A.; Soundararajan, D.C.R.; Nayagam, S.M.; Matchado, M.S.; Raveendran, M.; Shetty, A.P.; Kanna, R.M.; Dharmalingam, K. Inflammaging determines health and disease in lumbar discs-evidence from differing proteomic signatures of healthy, aging, and degenerating discs. *Spine J. Off. J. N. Am. Spine Soc.* **2020**, *20*, 48–59. [CrossRef] [PubMed]
39. Shin, J.H.; Park, S.; Cho, H.; Kim, J.H.; Choi, H. Adipokine human Resistin promotes obesity-associated inflammatory intervertebral disc degeneration via pro-inflammatory cytokine cascade activation. *Sci. Rep.* **2022**, *12*, 8936. [CrossRef] [PubMed]
40. Ruiz-Fernández, C.; Francisco, V.; Pino, J.; Mera, A.; González-Gay, M.A.; Gómez, R.; Lago, F.; Gualillo, O. Molecular Relationships among Obesity, Inflammation and Intervertebral Disc Degeneration: Are Adipokines the Common Link? *Int. J. Mol. Sci.* **2019**, *20*, 2030. [CrossRef] [PubMed]
41. Zhang, Y.; Wang, Y.; Yuan, Y.; Lin, Y.; Lin, B.; Zhou, H. Propionibacterium acnes induces cartilaginous endplate degeneration by promoting MIF expression via the NF-κB pathway. *J. Orthop. Surg.* **2020**, *15*, 213. [CrossRef] [PubMed]
42. Zheng, Y.; Lin, Y.; Chen, Z.; Jiao, Y.; Yuan, Y.; Li, C.; Xu, X.; Cao, P. Propionibacterium acnes induces intervertebral discs degeneration by increasing MMP-1 and inhibiting TIMP-1 expression via the NF-κB pathway. *Int. J. Clin. Exp. Pathol.* **2018**, *11*, 3445–3453. [PubMed]
43. Dou, Y.; Zhang, Y.; Liu, Y.; Sun, X.; Liu, X.; Li, B.; Yang, Q. Role of macrophage in intervertebral disc degeneration. *Bone Res.* **2025**, *13*, 15. [CrossRef] [PubMed]
44. Empere, M.; Wang, X.; Prein, C.; Aspberg, A.; Moser, M.; Oohashi, T.; Clausen-Schaumann, H.; Aszodi, A.; Alberton, P. Aggrecan governs intervertebral discs development by providing critical mechanical cues of the extracellular matrix. *Front. Bioeng. Biotechnol.* **2023**, *11*, 1128587. [CrossRef] [PubMed]
45. Liu, Y.; Chao, T.; Li, Y.; Peng, P.; Ge, G.; Geng, D.; Dong, L.; Wang, C. An enzyme-proof glycan glue for extracellular matrix to ameliorate intervertebral disc degeneration. *Nat. Commun.* **2025**, *16*, 3629. [CrossRef] [PubMed]
46. Xie, G.; Liang, C.; Yu, H.; Zhang, Q. Association between polymorphisms of collagen genes and susceptibility to intervertebral disc degeneration: A meta-analysis. *J. Orthop. Surg.* **2021**, *16*, 616. [CrossRef] [PubMed]
47. Gu, J.; Guan, F.; Guan, G.; Xu, G.; Wang, X.; Zhao, W.; Ji, Y.; Yan, J. Aggrecan variable number of tandem repeat polymorphism and lumbar disc degeneration: A meta-analysis. *Spine* **2013**, *38*, E1600–E1607. [CrossRef] [PubMed]


**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.