Neurospine 2020;17(2):426-442.
https://doi.org/10.14245/ns.2040056.028

Neurospine
pISSN 2586-6583 eISSN 2586-6591

Original Article

Corresponding Author
Shanmuganathan Rajasekaran
https://orcid.org/0000-0002-2966-6338

Department of Spine Surgery, Ganga Hospital, 313 Mettupalayam road, Coimbatore, India
E-mail: rajasekaran.orth@gmail.com

Received: January 31, 2020
Revised: March 1, 2020
Accepted: March 2, 2020

Proteomic Signatures of Healthy Intervertebral Discs From Organ Donors: A Comparison With Previous Studies on Discs From Scoliosis, Animals, and Trauma

Shanmuganathan Rajasekaran[1], Chitraa Tangavel[2], Dilip Chand Raja Soundararajan[1], Sharon Miracle Nayagam[2], Monica Steffi Matchado[2], Raveendran Muthurajan[3], K.S. Sri Vijay Anand[1], Sunmathi Rajendran[2], Ajoy Prasad Shetty[1], Rishi Mugesh Kanna[1], Dharmalingam Kuppamuthu[4]

[1]Department of Spine Surgery, Ganga Hospital, Coimbatore, India
[2]Ganga Research Centre, Coimbatore, India
[3]Department of Plant Biotechnology, Tamil Nadu Agricultural University, Coimbatore, India
[4]Aravind Medical Research Foundation, Madurai, India

Objective: To catalog and characterize the proteome of normal human intervertebral disc (IVD).
Methods: Nine magnetic resonance imaging (MRI) normal IVDs were harvested from 9 different brain dead yet alive voluntary organ donors and were subjected to electrospray ionization-liquid chromatography tandem mass spectrometry (ESI-LC-MS/MS) acquisition.
Results: A total of 1,116 proteins were identified. Functional enrichment analysis tool DAVID ver. 6.8 categorized: extracellular proteins (38%), intracellular (31%), protein-containing complex (13%), organelle (9%), membrane proteins (6%), supramolecular complex (2%), and 1% in the cell junction. Molecular function revealed: binding activity (42%), catalytic activity (31%), regulatory activity (14%), and structural activity (7%). Molecular transducer, transporter, and transcription regulator activity together contributed to 6%. A comparison of the proteins obtained from this study to others in the literature showed a wide variation in content with only 3% of bovine, 5% of murine, 54% of human scoliotic discs, and 10.2% of discs adjacent to lumbar burst fractures common to our study of organ donors. Between proteins reported in scoliosis and lumbar fracture patients, only 13.51% were common, further signifying the contrast amongst the various MRI normal IVD samples.
Conclusion: The proteome of "healthy" human IVDs has been defined, and our results show that proteomic data on IVDs obtained from scoliosis, fracture patients, and cadavers lack normal physiological conditions and should not be used as biological controls despite normal MRI findings. This questions the validity of previous studies that have used such discs as controls for analyzing the pathomechanisms of disc degeneration.

Keywords: Tandem mass spectrometry, Human intervertebral disc, Proteome, Degenerative disc disease, Tissue donors, Gene ontology

INTRODUCTION

Low back pain (LBP) is the most common musculoskeletal disorder affecting 540 million people globally.[1] Lumbar degenerative disc disease (DDD) is the leading cause of LBP and remains poorly understood. During disc degeneration, the intervertebral disc (IVD) undergoes physiological and morphological changes that lead to alterations in the extracellular matrix

(ECM). However, to understand the exact differences that occur in the disc proteome during this degenerative process, clear documentation of the proteins expressed in discs under normal physiological conditions is essential.

Characterization of the proteome of a normal IVD has been challenging due to the limited access to appropriate biological samples. To date, many comparative proteomic studies have used disc samples harvested from scoliosis, spinal trauma, and cadavers as controls to analyze diseased IVD as they appear normal in MRI. Control discs obtained from scoliosis subjects are reported to have an asymmetric loading pattern, leading to bone deformity,[2] and stress profiles similar to degenerated discs,[3] indicating pathomorphological changes that alter the cellular activity. The high-stress levels and asymmetrical loading in these discs might result in an altered biological composition, thus flawing the results of molecular comparative studies that have used these samples as "healthy" controls.

Lumbar spine burst fractures might also not represent actual controls, as 70% of the discs have been reported to have disc edema, bleeding/rupture, or displacement, which provides evidence of alteration in biochemical composition inside the IVD.[4] In an animal study of vertebral endplate damage induced changes in IVD, the elevation of cytotoxic lactate dehydrogenase enzyme, upregulated expression of pro-apoptotic proteins such as FasL and tumor necrosis factor-α, along with increase in catabolic gene expression (matrix metalloproteinase [MMP]-1 and MMP-13) demonstrates that discs procured from spaces adjacent to spinal trauma do not represent physiologically normal conditions.[5]

On the other hand, despite many studies stating that the integrity of proteins in autopsy specimens from 6 hours after death remains stable with only a minor degree of loss, hypoxia in these specimens is a critical agonal factor. Extended agonal state and hypoxia will lead to the increase in tissue lactate, thus lowering pH, which plays a significant role in alterations of RNA and protein integrity and may, therefore, fail to reflect homeostasis. Notably, it assumes more importance in a human IVD which is believed to be an avascular tissue and primarily depends on nutrition through solute transport occurring across end plates by the process of diffusion.

The majority of studies in LBP involve *in vivo* models of other animals such as bovine,[6] murine,[7] canine,[8] porcine, and caprine discs to understand the disease mechanisms in the human nucleus pulposus (NP). However, due to inter-species differences, their value for understanding disc pathology is limited, which emphasizes the need for proper human controls for molecular-level studies.

In this study, 9 MRI normal discs were harvested from brain dead yet alive voluntary asymptomatic organ donors, and we employed proteomic approach coupled with tandem mass

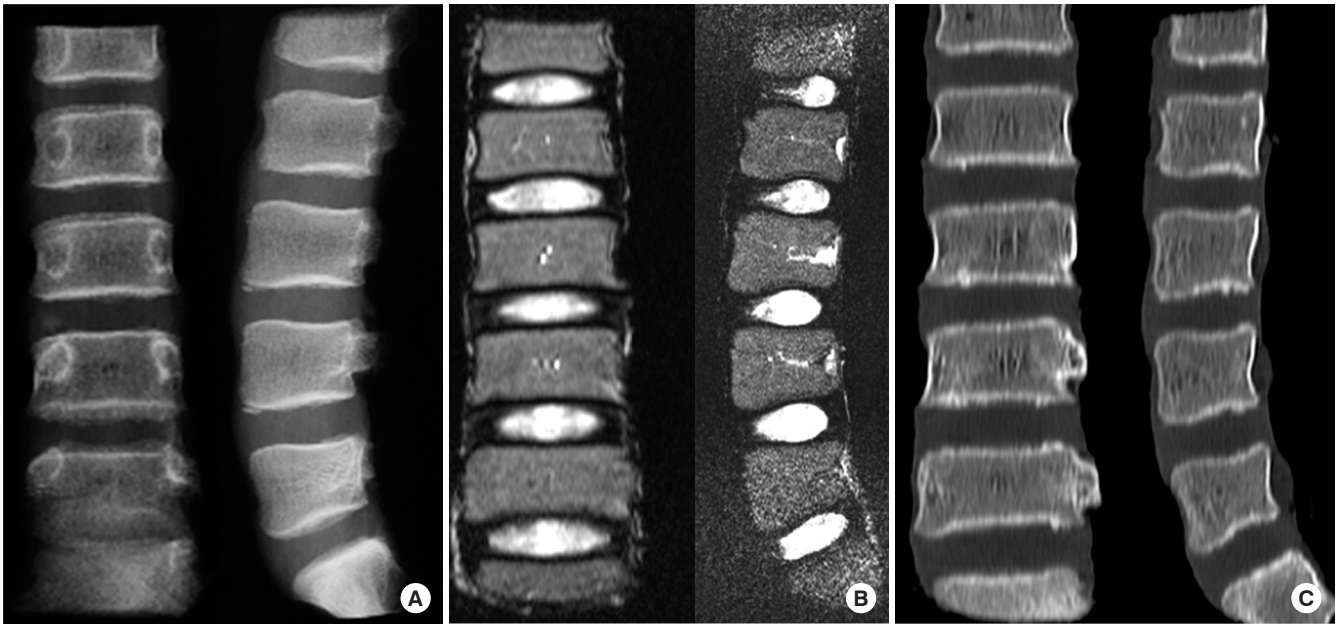


Fig. 1. Imaging of the acquired vertebral segments to choose radiologically normal disc specimens: (A) plain radiograph, (B) T2-weighted magnetic resonance image, and (C) computed tomography images demonstrating perfect normal intervertebral disc in the spine segment harvested from a 13-year-old voluntary organ donor, ideal to be labeled as a healthy normal disc.

spectrometry to derive the comprehensive list of proteins expressed in true biologically healthy control discs which would serve as the basis for identifying the biological processes (BPs) and pathways disrupted during aging and in DDD.

MATERIALS AND METHODS

The study was conducted after the approval of Institutional Review Board (IRB) of Ganga Medical Centre and Hospitals Pvt Ltd. (reference No. 2/2/2018) and samples were processed as per the ethical guidelines of the Indian Council of Medical research. Access to discs from organ donors was through licence obtained from regional organ transplant board for acquiring bone and musculoskeletal tissue.

1. Harvesting the Human IVD Tissues From Brain Dead Organ Donor

Lumbar spine segments were collected under sterile operating conditions as a source of bone allograft from 9 brain dead voluntary organ patients with no history of spinal trauma or spinal disorder after approval from the IRB. The removed spine segments were immediately investigated by plain radiography (Fig. 1A), 1.5 Tesla MRI (Fig. 1B), and computed tomography (Fig. 1C). The harvested segments which were completely normal in imaging (Fig. 2A) were segregated into vertebrae (Fig. 2B), annulus fibrosus, NP (Fig. 2C), and endplates (Fig. 2D) under strict aseptic precautions using microscope. To avoid cells from annulus fibrosus, we chose to dissect NP well short of the transitional zone. The bones were used as a source of allografts after processing them appropriately in a licensed bone

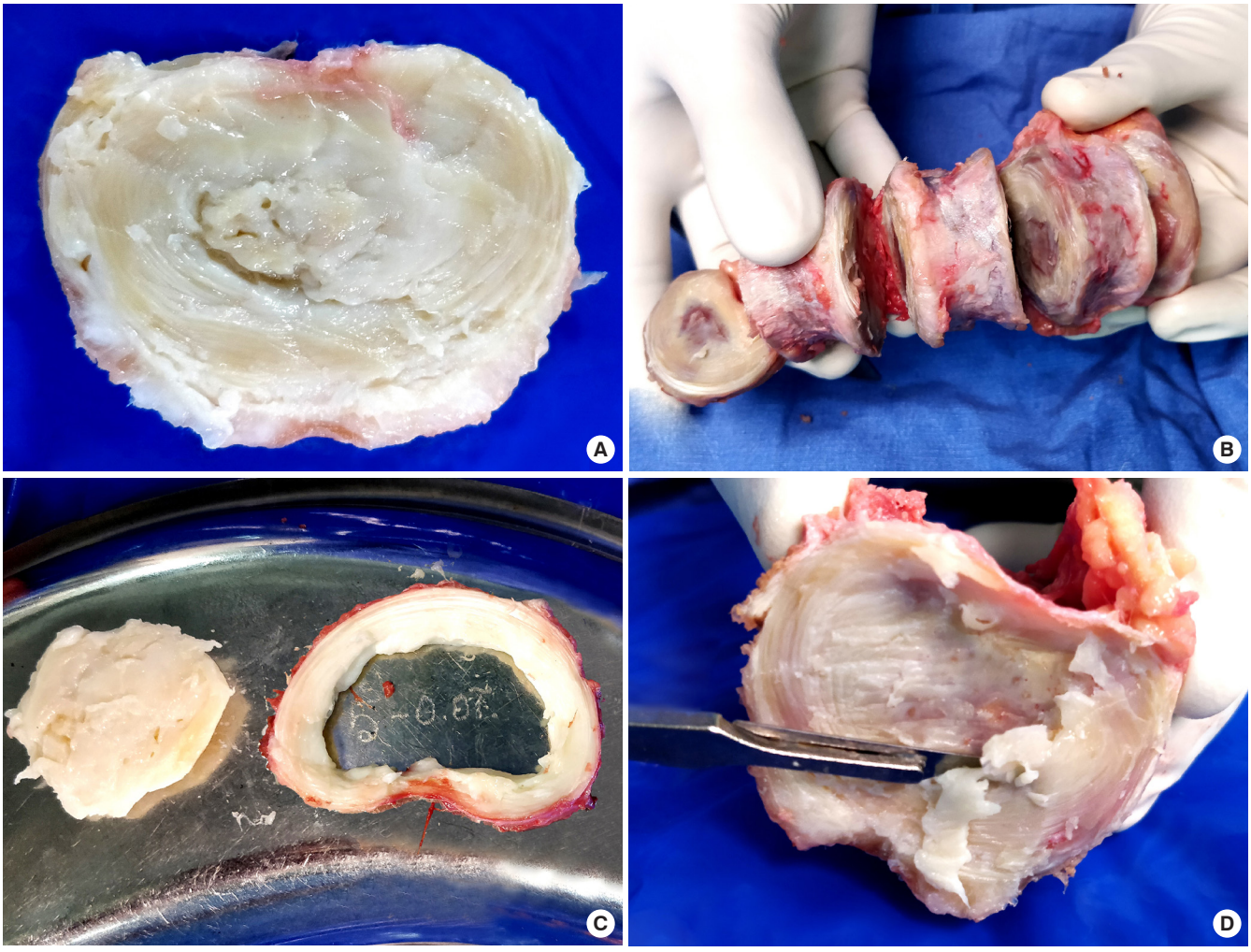


Fig. 2. Segregation of the intervertebral discs from the spine segment which was proven as radiologically normal: Intervertebral disc (A) isolated from bone (B) which is used as allograft after appropriate processing intervertebral disc dissected into nucleus pulposus and annulus fibrosus (C) and endplate (D) being scraped out from the bone for further research.

https://doi.org/10.14245/ns.2040056.028

bank (Fig. 2B). The remnant MRI normal NP of disc tissues were first washed with phosphate buffer solution (PBS) and stored in Cryovials and then snap-frozen immediately in liquid nitrogen containers. These tissues were carefully transported to the research center for further processing and analysis.

2. Protein Extraction from Human IVD Tissues and Clean-up Prior to In-Gel Prefractionation

Around 200 mg of NP snap-frozen in liquid nitrogen was pulverized and resuspended in 2 mL of radio immunoprecipitation assay buffer (RIPA) buffer (50 mM Tris-Hcl, 150 mM NaCl, 0.5% sodium deoxycholate, 0.1% sodium dodecyl sulfate [SDS], 1% Triton X-100, 10% glycerol) and vortexed for 30 seconds. Further, the samples were sonicated thrice in pulse mode for 15 seconds on and off with an amplitude of 35%. The sample tubes were kept immersed in an ethanol ice bath throughout the process. The sonicated sample was incubated on ice for 30 minutes and centrifuged at 10,000 rpm for 20 minutes at 4°C. The supernatant contained hydrophilic, and mild hydrophobic proteins were labelled as S1 fraction. The pellet expected to contain membrane and hydrophobic proteins were extracted by adding 1 mL of 2% SDS and the homogenized sample was further boiled at 70°C for 10 minutes. The mixture was centrifuged at 12,000 rpm for 30 minutes at 23°C, and the resultant supernatant was labelled as S2 fraction.

3. Sample Clean-up using an Organic Solvent to Remove the Interfering Glycans From the Proteins before Proteome Analysis Using Tandem Mass Spectrometry

Fractions S1 and S2 obtained for each sample were cleaned up using organic solvent chloroform: methanol: water in the ratio 1:4:1:3 (v/v) (1 volume of the sample; 4 volumes of methanol: 1 volume of chloroform and 3 volumes of water) and vortexed for 30 seconds twice. The mixture was centrifuged at 10,000 rpm for 20 minutes at 4°C. The chloroform layer containing the protein was removed, and an equal volume of methanol was added and centrifuged at 10,000 rpm for 20 minutes at 4°C. The resulting pellet was rinsed twice with methanol and centrifuged with the same parameters as specified above. The resulting pellet is so fragile; hence the methanol should be discarded with caution without disturbing the pellet, which is further air-dried for 3 minutes and resuspended in 200 μL of 10 mM Tris-HCl 8.0, 0.1% SDS.

Fifty micrograms of the total proteins were then prefractionated in 10% SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and stained with colloidal coomassie brilliant blue G250.

4. Tryptic Digestion, Peptide Extraction, and Purification Using C18 Spin Columns

The total proteins of the human IVD were subjected to tryptic in-gel digestion, followed by the analysis of the tryptic peptides using tandem mass spectrometry.

The entire lane was sliced into 10 fractions and subjected to tryptic digestion at a concentration of 600 ng/fraction. The tryptic peptides were purified using C18 columns (Agilent Peptide Cleanup C18 Spin Column [#Cat No. 51882750]) before mass spectrometry as per the manufacturer's instruction. Each of the 3 fractions was pooled, dried, and resuspended in an appropriate volume of 0.1% formic acid with 5% acetonitrile. Around 1,000 fmol (femtomole) of the tryptic peptides from each fraction were loaded to nano-liquid chromatography (LC) and subjected to tandem mass acquisition. The resultant raw files from the electrospray ionization-liquid chromatography tandem mass spectrometry (ESI-LC-MS/MS) were processed in PD ver. 1.4, as described in the (Fig. 3).

5. Relative Quantification by Spectral Count

Spectral counts obtained by LC/MS-MS were further subjected to normalization by normalized spectral abundance factor (NSAF) method.[9]

NSAF is calculated as follows:

$$(NSAF)_K = \frac{\left(\frac{SpC}{L}\right)_K}{\sum_{i=1}^{N}\left(\frac{SpC}{L}\right)_i}$$

SpC: Number of spectral counts; L: Protein Length; k: individual protein

6. Gene Ontology and Pathway Enrichment Analysis

To interpret the functional enrichment of healthy IVD proteins, gene ontology (GO), including BP, molecular function (MF), and cellular component (CC), was performed using 'cluster profiler' in R vs. 3.6.1 (2019-07-05) (R Foundation, Vienna, Austria) program. Besides, pathway enrichment analysis was performed using the Reactome database.

7. Clustering and Protein-Protein Interaction Network

The identified proteins were mapped using the Search Tool for the Retrieval of Interacting Genes (STRING; http://string-db.org) with default settings, and the disconnected nodes were excluded from the interaction analysis. Protein Interaction Network (PPI) network with the interaction score >0.4 was extracted from STRING, and the network was further analyzed using Cytoscape *cytoHubba*.[10] Top 10 ranked proteins were

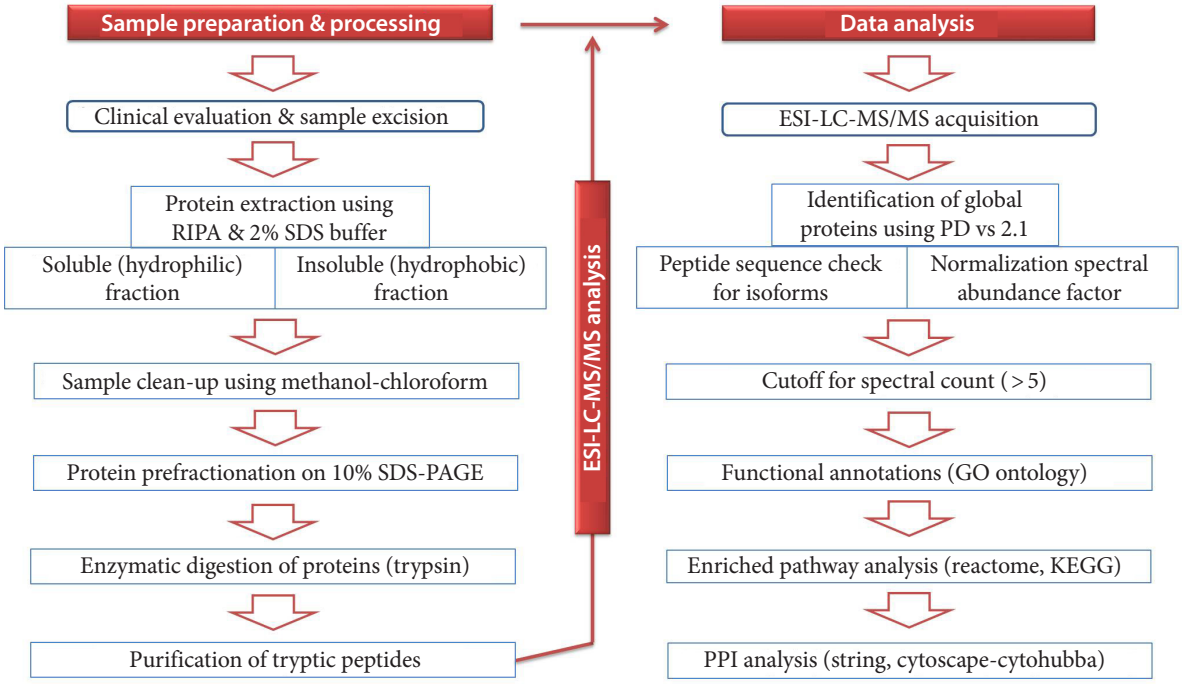


Fig. 3. Schematic illustration of the proteomics work flow adopted for this study. RIPA, radio immunoprecipitation assay buffer; SDS, sodium dodecyl sulfate; PAGE, polyacrylamide gel electrophoresis; ESI-LC-MS/MS, electrospray ionization-liquid chromatography tandem mass spectrometry; GO, gene ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; PPI, Protein Interaction Network.

identified using *cytoHubba* plugin based on one of the important network scoring methods: *Degree*. Proteins with higher degrees are considered to be more essential proteins.

8. Histology and Immunohistochemistry

Among 9 IVD NP tissues collected for the study, only 2 were subjected to Immunohistochemical (IHC) analysis. IHC validation of the proteomic findings was performed by indirect 3-step immunoperoxidase-technique with an avidin-biotin complex (ABC) detection system (paraffin). Lumbar spines were fixed in 4% paraformaldehyde overnight at 4°C, decalcified at room temperature, dehydrated in ethanol, washed with xylene, and embedded in paraffin. Tissues sections were cut at a thickness of 5–6 μm using a microtome (Leica Microsystems, Wetzlar, Germany) with overnight incubation at 37°C, deparaffinized in xylene, dehydrated in ethanol, and incubated with 0.3% hydrogen peroxide in absolute methanol for 20 minutes at room temperature to inhibit endogenous peroxidase activity. Then sections were re-hydrated with 96% alcohol , followed by 70% ethanol for 10 minutes and double distilled water for 3 times. To eliminate chemical modifications, antigen retrieval was performed and incubated overnight with a 1% primary antibody in bovine serum albumin in PBS. Followed PBS wash, sections were subjected to incubation with biotinylated secondary antibody for 60 minutes and washed with PBS for 3 times, 5 minutes each before incubating for 30 minutes with horseradish peroxidase labeled ABC complex. The colored reaction product was developed with 3, 3-diaminobenzidine tetrahydrochloride. Finally, the sections were counterstained lightly with hematoxylin for microscopic examination of the antigen-antibody complex.

9. Data Submission

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD016560 and 10.6019/PXD016560.

RESULTS

A total of 1,116 different proteins were identified, and the results are discussed as per the protocols followed to ensure the maximal retrieval of proteins along with the findings of the various methods employed before optimization.

https://doi.org/10.14245/ns.2040056.028

1. Enrichment of ECM Proteins From NP Tissues

The ECM proteins are large, diverse molecules with high interconnectivity, which makes them insoluble in low concentration of salts and detergents. In this study, we tried to extract the ECM proteins with various buffers (1X laemelli, 2% SDS, guanidium-HCl, Tris-acetate Nacl, and RIPA) with different composition and found sequential extraction with 22 different buffers: RIPA followed by 2% SDS was found to yield better prefractionation profile in 10% SDS-PAGE as observed in (Supplementary Fig. 1A).

Following sequential enrichment, the interference of the high concentration of glycosaminoglycans (GAGs) in the protein prefractionation on SDS-PAGE was a significant challenge. The interfering GAGs, salts, detergents were precipitated using: acetone; ethanol; tricholoroacetic acid (TCA)-acetone; 3kDa membrane cutoff device and methanol-chloroform. A broadening of lanes was observed with the use of ethanol and 3kDa cutoff device, showing the inefficiency of the methods to remove salts. Similarly, acetone precipitation failed to remove the interfering GAGs (Supplementary Fig. 1B). TCA-acetone precipitation shows the loss of proteins, especially around the 66-kDa region (circled in Fig. 4B).

Interestingly the prominent 67-kDa protein was completely TCA soluble. Out of these, methanol-chloroform was found to be efficient as shown (Supplementary Fig. 1B) in terms of removal of glycans, salts and detergents, and the recovery of the total protein (Table 1) was close to 62%. Supplementary Fig. 2A illustrates the efficiency of a sample clean-up by methanol-chloroform with that of acetone, and 3-kDa membrane cutoff device.

Following sample clean-up with methanol-chloroform and prefractionation on 10% SDS-PAGE, the functional annotation for BP in the samples revealed proteins involved in complementary activation, positive regulation of inflammatory response, and inflammatory response proteins (Supplementary Fig. 2B) which failed to show up before sample clean-up. This indicates the importance of sample clean-up in detecting proteins of low abundance, which might be of clinical significance.

2. Identification of NP Related Proteomic Signatures by 1D Gel Analysis

We deployed shotgun proteomics to determine the total proteins in healthy IVD tissues of 9 organ donor samples by using LC-MS/MS analysis. A total of 1,116 proteins were identified, and the identified proteins' accessions and their peptides were verified against the Uniprot database (2018). Uncharacterized or putative or fragment proteins were excluded, and they are documented in Supplementary Table 1.

The entire matrisome list of normal IVD was created by comparing it with the matrisome database (http://matrisome.org/). The quantitative analysis based on their peptide spectral matches (PSM) revealed that normal NP proteome consisted of 57% of matrisomal proteins (core matrisome, regulators, ECM affiliated, and secreted factors) and 43% of other proteins (cyto-

Table 1. Protein recovery and loss estimated for organic solvents used in this study

Solvents	Initial concentration	Final concentration	% Loss	% Gain
Methanol-chloroform	100	62	38	62
Ethanol	100	65	35	65
Acetone	100	83	17	83
TCA-acetone	100	43.7	56.3	43.7
3-kDa cutoff device	100	51.6	48.4	51.6

TCA, tricholoroacetic acid.

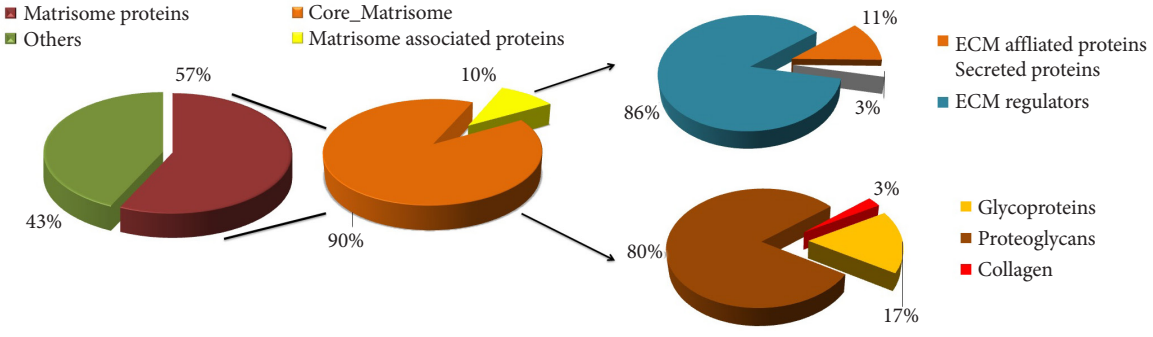


Fig. 4. Categorization of organ donor intervertebral disc nucleus pulposus matrisomal proteins from this study using www.matrisome.org database. The pie chart exhibits the distribution of the matrisomal proteins based on their abundance. ECM, extracellular matrix.

skeletal proteins, intracellular proteins, membrane proteins, and blood components). The matrisome further consisted of 90% of core matrisomal proteins and 10% of matrisome-associated proteins (Fig. 4). The core matrisome was further characterized as proteoglycans (80%), glycoproteins (17%), and collagens (3%). The matrisome-associated proteins were categorized as ECM regulators (86%), ECM-affiliated (11%), secreted proteins (3%). The quantitative distribution of the proteins is illustrated in Fig. 4.

3. GO Analysis

A cutoff PSM ≥5 was applied to the total proteins identified

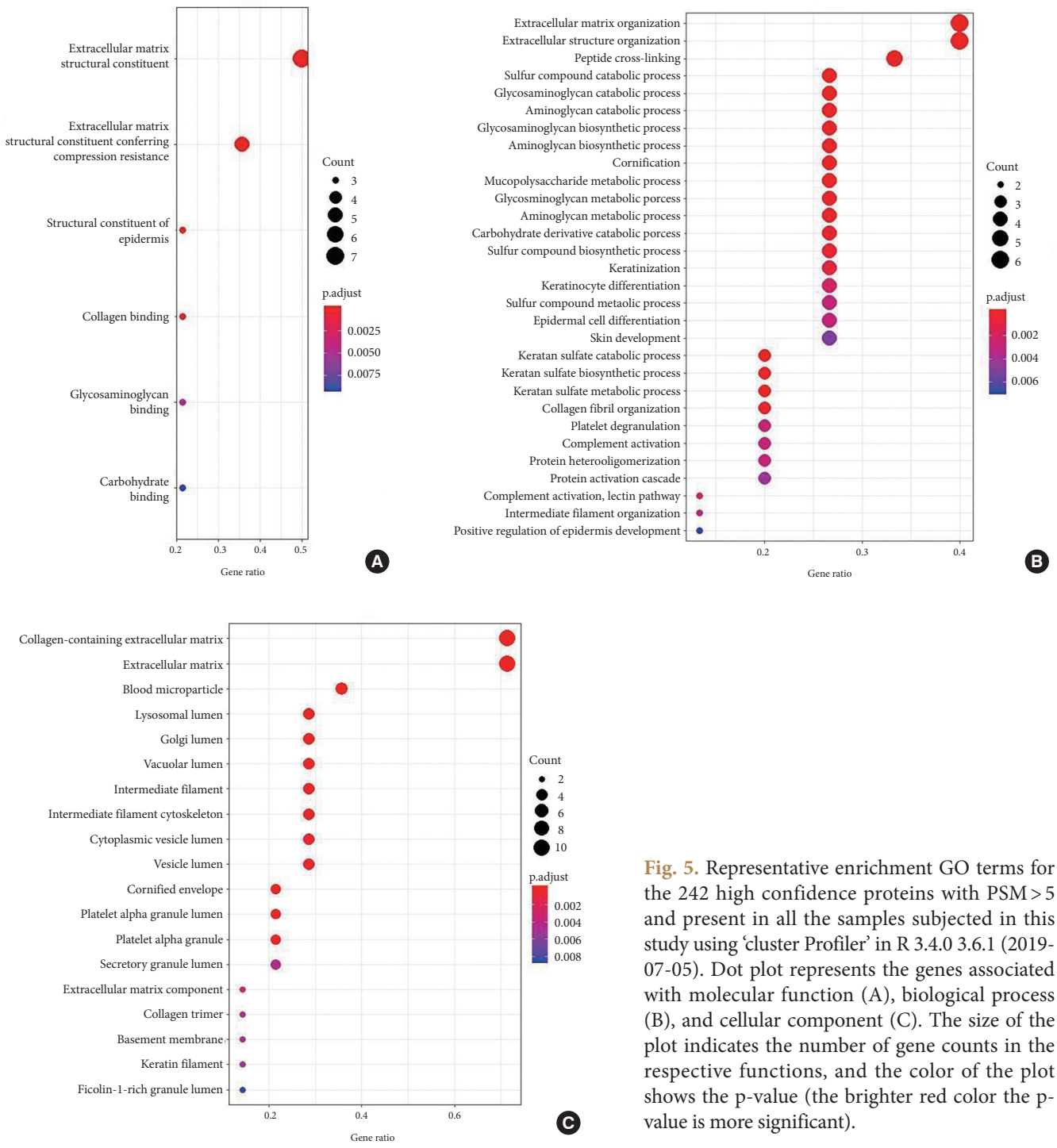


Fig. 5. Representative enrichment GO terms for the 242 high confidence proteins with PSM > 5 and present in all the samples subjected in this study using 'cluster Profiler' in R 3.4.0 3.6.1 (2019-07-05). Dot plot represents the genes associated with molecular function (A), biological process (B), and cellular component (C). The size of the plot indicates the number of gene counts in the respective functions, and the color of the plot shows the p-value (the brighter red color the p-value is more significant).

https://doi.org/10.14245/ns.2040056.028

Table 2. List of top 10 enriched pathways from Search Tool for the Retrieval of Interacting Genes database

No.	Enriched pathway	Proteins involved	p-value	Gene symbol
Based on p-value				
1	Formation of the cornified envelope	29	1.66E-24	*CDSN, DSG1, DSP, JUP, KRT1, KRT10, KRT14, KRT15, KRT16, KRT17, KRT19, KRT2, KRT33A, KRT4, KRT5, KRT6A, KRT6B, KRT6C, KRT72, KRT73, KRT76, KRT77, KRT78, KRT79, KRT8, KRT80, KRT84, KRT9, TGM5*
2	Extracellular matrix organization	37	8.90E-24	*A2M, ACAN, ASPN, BGN, CD47, COL10A1, COL11A2, COL14A1, COL15A1, COL1A1, COL1A2, COL2A1, COL6A1, COL6A2, COL6A3, COMP, DCN, FGA, FGB, FGG, FMOD, GDF5, HAPLN1, HSPG2, HTRA1, LOX, LUM, MMP10, PCOLCE, PCOLCE2, PPIB, THBS1, TIMP1, TNC, TTR, VCAN, VTN*
3	Platelet degranulation	22	1.04E-16	*A1BG, A2M, ALB, ANXA5, APOA1, CD109, CLEC3B, CLU, F13A1, FGA, FGB, FGG, HRG, ITIH4, QSOX1, SERPINA1, SERPING1, TF, THBS1, TIMP1, TIMP3, TTN*
4	ECM proteoglycans	18	9.26E-16	*ACAN, ASPN, BGN, COL1A1, COL1A2, COL2A1, COL6A1, COL6A2, COL6A3, COMP, DCN, FMOD, HAPLN1, HSPG2, LUM, TNC, VCAN, VTN*
5	Integrin cell surface interactions	17	6.49E-14	*CD47, COL10A1, COL1A1, COL1A2, COL2A1, COL6A1, COL6A2, COL6A3, COMP, FGA, FGB, FGG, HSPG2, LUM, THBS1, TNC, VTN*
6	Platelet activation, signaling and aggregation	25	6.50E-14	*A1BG, A2M, ALB, ANXA5, APOA1, CD109, CLEC3B, CLU, COL1A1, COL1A2, F13A1, FGA, FGB, FGG, HRG, ITIH4, QSOX1, RASGRP2, SERPINA1, SERPING1, TF, THBS1, TIMP1, TIMP3, TTN*
7	Innate immune system	45	2.71E-13	*A1BG, ACTG1, ANXA2, ARG1, C1S, C3, C9, CALML5, CASP8, CD47, CFB, CFI, CHI3L1, CLU, DCD, DSG1, DSP, FABP5, FGA, FGB, FGG, FLG2, FTL, GSN, HBB, HP, HRNR, JUP, KRT1, LTF, LYZ, MIF, PKM, PLA2G2A, QSOX1, RASGRP2, S100A1, S100A8, S100A9, SERPINA1, SERPINB1, SERPING1, TNFAIP6, TTR, VTN*
8	Hemostasis	34	1.31E-12	*A1BG, A2M, ALB, ANXA2, ANXA5, APOA1, CD109, CD47, CLEC3B, CLU, COL1A1, COL1A2, F13A1, FGA, FGB, FGG, HBB, HBD, HRG, ITIH4, KIF1B, MIF, QSOX1, RASGRP2, SERPINA1, SERPINA5, SERPINC1, SERPINE2, SERPING1, TF, THBS1, TIMP1, TIMP3, TTN*
9	Neutrophil degranulation	29	1.53E-11	*A1BG, ANXA2, ARG1, C3, CALML5, CD47, CHI3L1, DSG1, DSP, FABP5, FLG2, FTL, GSN, HBB, HP, HRNR, JUP, KRT1, LTF, LYZ, MIF, PKM, QSOX1, S100A8, S100A9, SERPINA1, SERPINB1, TNFAIP6, TTR*
10	Degradation of the extracellular matrix	17	7.17E-11	*A2M, ACAN, COL10A1, COL11A2, COL14A1, COL15A1,COL1A1, COL1A2, COL2A1, COL6A1, COL6A2, COL6A3, DCN, HSPG2, HTRA1, MMP10, TIMP1*
Based on number of proteins involved				
1	Immune system	57	3.58E-10	*A1BG, ACTG1, ANXA1, ANXA2, ARG1, BLMH, C1S, C3, C9, CA1, CALML5, CASP8, CD47, CFB, CFI, CHI3L1, CLU, COL1A1, COL1A2, COL2A1, CSF3R, DCD, DSG1, DSP, F13A1, FABP5, FGA, FGB, FGG, FLG2, FTL, GSN, HBB, HP, HRNR, JUP, KRT1, LTF, LYZ, MIF, MSN, PKM, PLA2G2A, QSOX1, RASGRP2, S100A1, S100A8, S100A9, SERPINA1, SERPINB1, SERPING1, TIMP1, TNFAIP6, TNFRSF11B, TTR, VIM, VTN*
2	Innate immune system	45	2.71E-13	*A1BG, ACTG1, ANXA2, ARG1, C1S, C3, C9, CALML5, CASP8, CD47, CFB, CFI, CHI3L1, CLU, DCD, DSG1, DSP, FABP5, FGA, FGB, FGG, FLG2, FTL, GSN, HBB, HP, HRNR, JUP, KRT1, LTF, LYZ, MIF, PKM, PLA2G2A, QSOX1, RASGRP2, S100A1, S100A8, S100A9, SERPINA1, SERPINB1, SERPING1, TNFAIP6, TTR, VTN*
3	Extracellular matrix organization	37	8.90E-24	*A2M, ACAN, ASPN, BGN, CD47, COL10A1, COL11A2, COL14A1, COL15A1, COL1A1, COL1A2, COL2A1, COL6A1, COL6A2, COL6A3, COMP, DCN, FGA, FGB, FGG, FMOD, GDF5, HAPLN1, HSPG2, HTRA1, LOX, LUM, MMP10, PCOLCE, PCOLCE2, PPIB, THBS1, TIMP1, TNC, TTR, VCAN, VTN*

(Continued to the next page)

https://doi.org/10.14245/ns.2040056.028

Table 2. Continued

No.	Enriched pathway	Proteins involved	p-value	Gene symbol
Based on number of proteins involved				
4	Developmental biology	37	1.28E-08	*ACTG1, ANK2, CDSN, COL2A1, COL6A1, COL6A2, COL6A3, DSG1, DSP, H2AFV, JUP, KRT1, KRT10, KRT14, KRT15, KRT16, KRT17, KRT19, KRT2, KRT33A, KRT4, KRT5, KRT6A, KRT6B, KRT6C, KRT72, KRT73, KRT76, KRT77, KRT78, KRT79, KRT8, KRT80, KRT84, KRT9, MSN, TGM5*
5	Hemostasis	34	1.31E-12	*A1BG, A2M, ALB, ANXA2, ANXA5, APOA1, CD109, CD47, CLEC3B, CLU, COL1A1, COL1A2, F13A1, FGA, FGB, FGG, HBB, HBD, HRG, ITIH4, KIF1B, MIF, QSOX1, RASGRP2, SERPINA1, SERPINA5, SERPINC1, SERPINE2, SERPING1, TF, THBS1, TIMP1, TIMP3, TTN*
6	Formation of the cornified envelope	29	1.66E-24	*CDSN, DSG1, DSP, JUP, KRT1, KRT10, KRT14, KRT15, KRT16, KRT17, KRT19, KRT2, KRT33A, KRT4, KRT5, KRT6A, KRT6B, KRT6C, KRT72, KRT73, KRT76, KRT77, KRT78, KRT79, KRT8, KRT80, KRT84, KRT9, TGM5*
7	Neutrophil degranulation	29	1.53E-11	*A1BG, ANXA2, ARG1, C3, CALML5, CD47, CHI3L1, DSG1, DSP, FABP5, FLG2, FTL, GSN, HBB, HP, HRNR, JUP, KRT1, LTF, LYZ, MIF, PKM, QSOX1, S100A8, S100A9, SERPINA1, SERPINB1, TNFAIP6, TTR*
8	Platelet activation, signaling and aggregation	25	6.50E-14	*A1BG, A2M, ALB, ANXA5, APOA1, CD109, CLEC3B, CLU, COL1A1, COL1A2, F13A1, FGA, FGB, FGG, HRG, ITIH4, QSOX1, RASGRP2, SERPINA1, SERPING1, TF, THBS1, TIMP1, TIMP3, TTN*
9	Disease	24	0.0039	*ACAN, ACTG1, ALB, BGN, CP, DCN, FGA, FGB, FGG, FMOD, HSPG2, IPO5, LTF, LUM, OGN, OMD, PRDX1, PRDX2, PRELP, RBP4, THBS1, TTR, VCAN, YWHAE*
10	Platelet degranulation	22	1.04E-16	*A1BG, A2M, ALB, ANXA5, APOA1, CD109, CLEC3B, CLU, F13A1, FGA, FGB, FGG, HRG, ITIH4, QSOX1, SERPINA1, SERPING1, TF, THBS1, TIMP1, TIMP3, TTN*

to increase the confidentiality of the GO process. With this filter, 242 of 1,116 proteins were considered for GO enrichment analysis under BPs, MF, and CC using cluster profiler in R ver. 3.6.1 (2019-07-05). The overall spread of GO that was all significantly enriched is depicted in Fig. 5.

In BP, the majority of the proteins were associated with the ECM and structural organization along with other metabolic processes that involve keratinization, cell differentiation, complement activation, collagen fibril organization, and cornification whereas in MF proteins were associated with structural constitution and resistance of matrix compression. As expected, in the CC category, proteins expressed were related to the ECM followed by the membrane and cytoplasmic regions.

4. Pathway Enrichment Analysis

Pathway analysis was performed with 242 proteins using the STRING database, and a total of 105 significant pathways were found to be enriched (Supplementary Table 2). Out of these, pathways required for IVD homeostasis were found to be highly significant ($p < 0.05$). Pathway enrichment revealed that most of the glycoproteins and proteoglycans were involved in ECM organization pathways. Collagens and glycoproteins identified were involved in activation of the Collagen biosynthesis, collagen chain trimerization, and integrin cell surface interaction pathways. ECM glycoproteins, collagens, and proteoglycans were a part of nonintegrin membrane-ECM interactions and carbohydrate metabolism. Cytoskeletal proteins were observed in the formation of the cornified envelope pathways, developmental biology, and Keratan Sulfate Biosynthesis. These pathways along with their genes, are enlisted in (Table 2). While the most significantly enriched pathway was the formation of a cornified envelope ECM organization, it was interesting to observe that the maximum number of proteins was involved in the immune system pathways followed by ECM organization.

5. Cluster Analysis and PPI Networks

The filtered proteins were clustered into 5 main categories with high interactions: collagens, cytoskeletons, proteoglycans, glycoproteins, and serine protease inhibitors (SERPINs) (Fig. 6). The total PPI network obtained from STRING comprised 221 connected nodes and 325 edges with an enrichment p-value of < 1.0e-16.

Top 10 ranked hub proteins were predicted using *cytohubba* based on 3 crucial network topological parameters: degree, maximum neighborhood component (MNC), and closeness (Fig. 7A, B). From both degree and MNC parameters, the top 10

https://doi.org/10.14245/ns.2040056.028

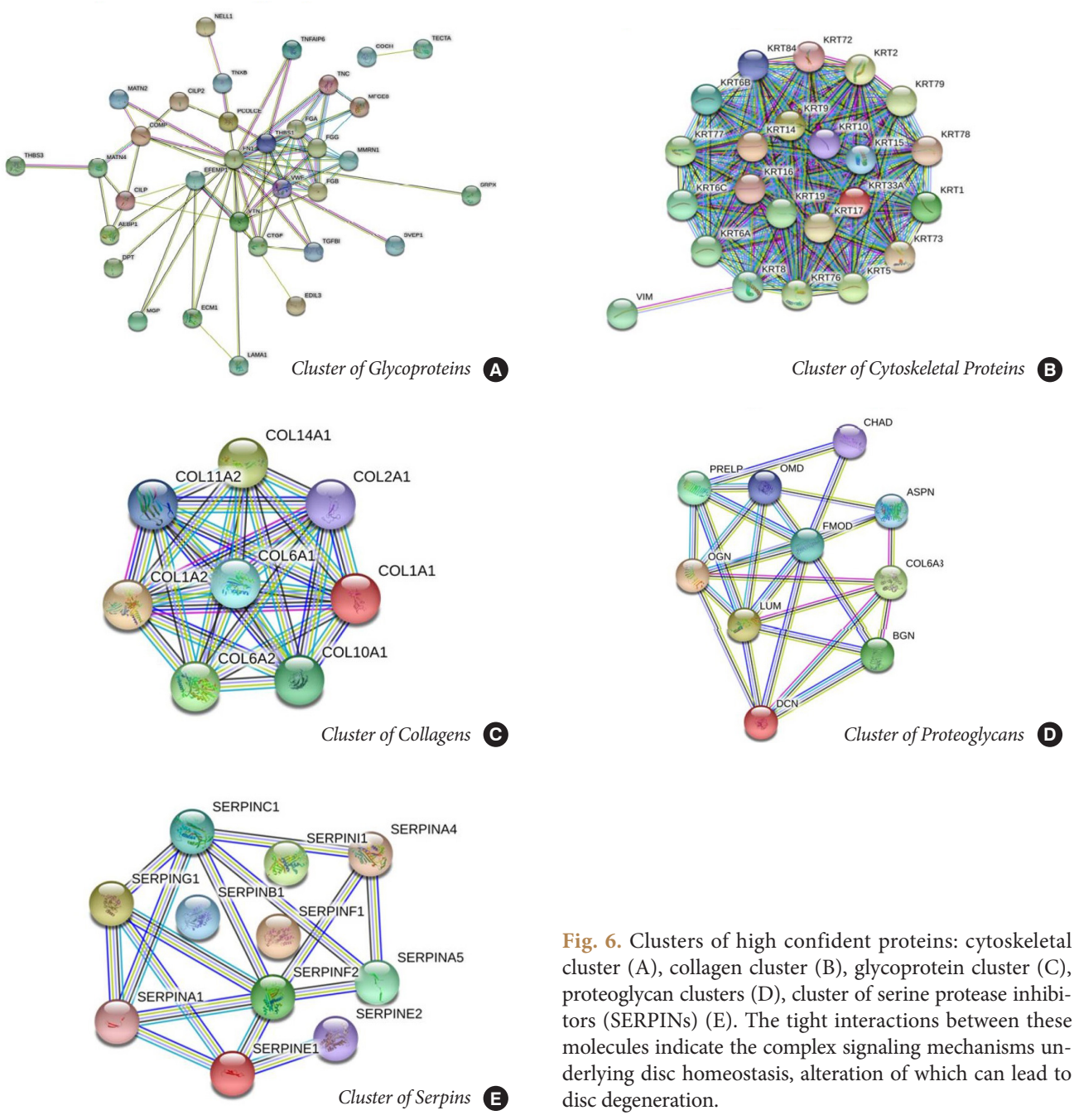


Fig. 6. Clusters of high confident proteins: cytoskeletal cluster (A), collagen cluster (B), glycoprotein cluster (C), proteoglycan clusters (D), cluster of serine protease inhibitors (SERPINs) (E). The tight interactions between these molecules indicate the complex signaling mechanisms underlying disc homeostasis, alteration of which can lead to disc degeneration.

proteins ranked in the following order: ALB (albumin), FN1 (glycoprotein), GAPDH (glycolytic pathway enzyme), TIMP1 (ECM regulators), HP (haptoglobin), C3 (complement), COL1A1, APOA1 (lipoprotein), VTN (vitronectin, glycoprotein), and CLU (clusterin). Based on closeness, we observe the same order of proteins as above except SERPINA1 (ECM regulators) and ceruloplasmin instead of APOA1 and CLU.

Comparative analysis between organ donor IVD proteins obtained in this study with reported proteome literature: (1) other species (murine, ovine, and bovine); (2) scoliosis; and (3) lumbar spines burst fracture patients

We performed a literature survey to compare our results with existing candidate proteins associated with IVD reported by transcriptional profiling and protein localization studies in bovine and murine (Supplementary Table 1).[11-19] The comparison showed 24 proteins common to human and bovine and 57 proteins common to human and murine (Fig. 8A). The percentage/number of total proteins may be varied since the compari-

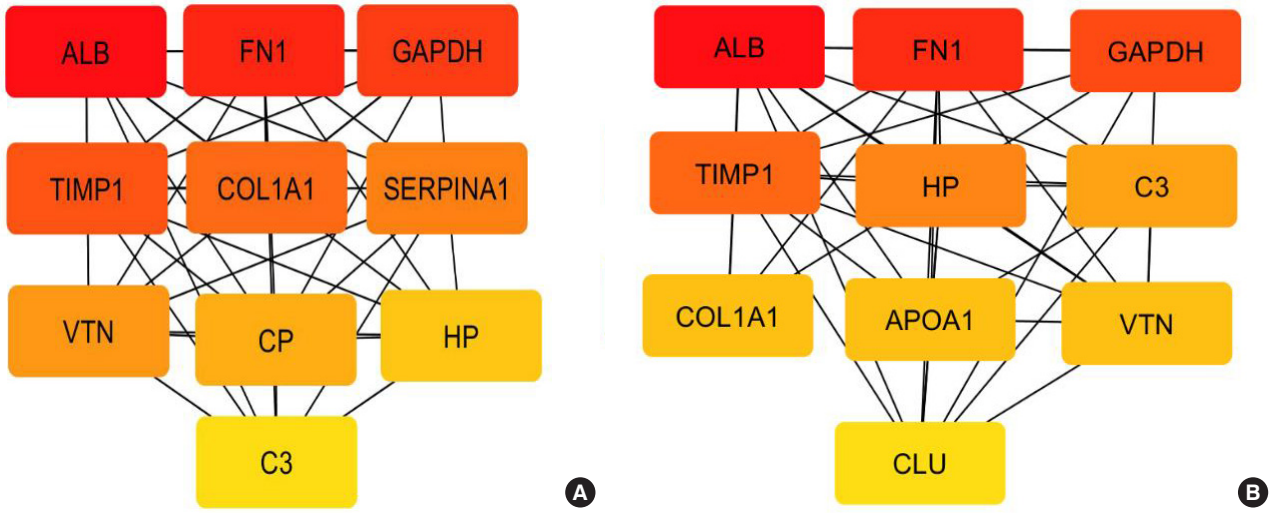


Fig. 7. (A) Top 10 ranked (HUB) proteins based on closeness were; ALB (albumin), FN1 (glycoprotein), GAPDH (glycolytic pathway enzyme), TIMP1 (ECM regulators), COL1A1 (collagen), SERPINA1 (ECM regulators), VTN (vitronectin - a glycoprotein), CP (ceruloplasmin), HP (haptoglobin), and C3 (complement). (B) Among the top 10 based on both degree and MNC parameters, we observed a similar order of proteins except APOA1 (lipoprotein), and CLU (Clusterin) instead of CP and SERPINA1.

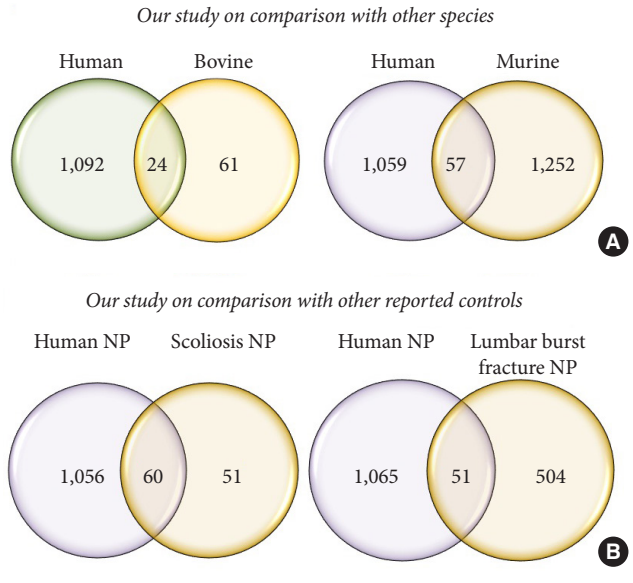


Fig. 8. (A) Venn diagrams comparing the proteins extracted in this with other literature only 24 proteins were common to bovine data versus 57 common to murine study. This shows the vast variation in the proteomic constitution of animals and humans, therefore questioning the suitability of animal models for translational research in disc degeneration. (B) All human discs appearing normal in magnetic resonance imaging cannot be considered as pure controls as only 60 proteins were common to our data and discs obtained from scoliosis and 57 common to discs harvested from burst fractures. NP, nucleus pulposus.

son was made irrespective of met hodology.

Similarly, we found 60 proteins from our study to be in common to scoliosis and 51 proteins common to lumbar spine burst fracture discs, as reported in the literature (Fig. 8B). Another important finding is that between the reported 111 proteins from scoliosis control discs[20] and 692 proteins lumbar burst spine fracture control discs,[21] only 15 proteins were common (Supplementary Table 3).

6. IHC Evaluation

IHC analysis was performed to evaluate the localization and expression of a subset of IVD proteins found in our study. Both lumbar spinal NP tissues subjected to IHC analysis showed positive staining against collagen type II, fibromodulin (FMOD), serine protease inhibitor (SERPINA1), and SERPINE2 (Fig. 9). Type II collagen was found throughout the matrix reconfirming the normalcy of adult discs. We found more pronounced staining of FMOD in cytoplasmic regions, whereas SERPINA1 and SERPINE2 along with cytoplasm as well as nuclear areas. As expected, all the proteins were found much localized to the matrix regions.

DISCUSSION

This study is the first to document the exact proteome profiling of adult NP of MRI normal IVDs, harvested from brain

https://doi.org/10.14245/ns.2040056.028

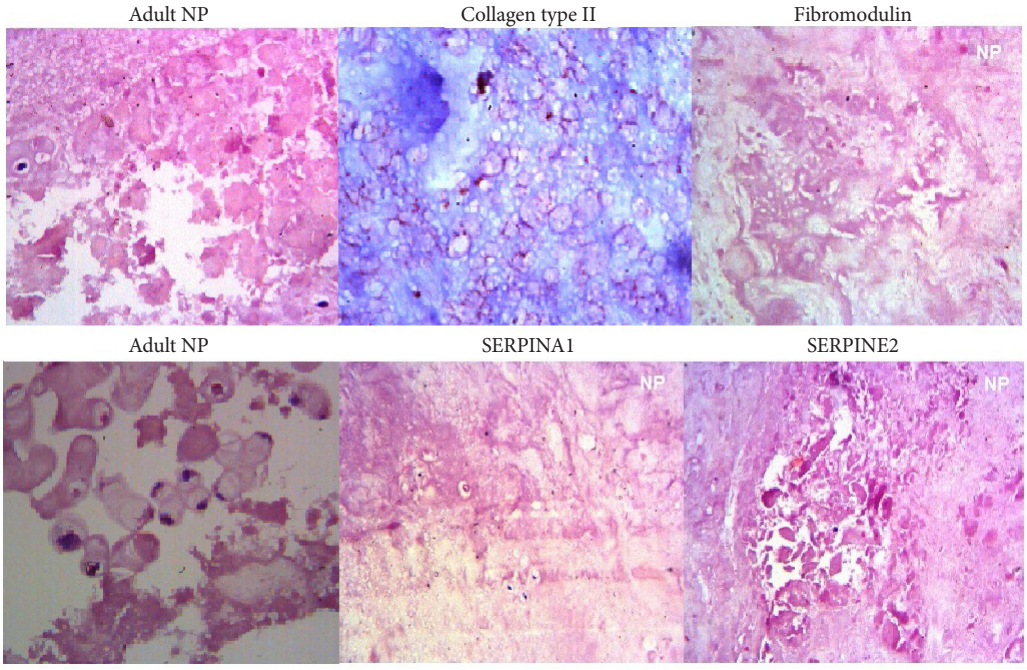


Fig. 9. Immunohistochemical staining in the nucleus pulposus of disc samples from normal discs against a subset of intervertebral disc proteins (collgen type II, FMOD, SERPINA1, and SERPINE2). Immunohistochemistry specimens were counter stained with hematoxylin. The magnification was ×400 for all the specimens. FMOD, fibromodulin; SERPIN, serine protease inhibitor (SERPINA1-Alpha-1-Antitrypsin; SERPINE2-Protease Nexin 1).

dead yet alive organ donors without history LBP. We employed an in-gel-based proteomic approach coupled with tandem mass spectrometry, and identified a total of 1,116 different proteins and have characterized the proteome of adult NP.

Two significant challenges that exist in the application of proteomics in human IVD include (1) identification of the normal IVD proteome and (2) optimization of appropriate methodology for maximal extraction of proteins from the complex cartilage tissues.

1. Normal IVD Proteome

While blood samples have been accessed from healthy volunteers to define the normal plasma proteome, the harvest of normal IVDs from asymptomatic individuals is unethical and impossible, leaving behind a vast lacuna in the literature on what constitutes a normal disc proteome. Previously, many authors have tried to address this problem by harvesting IVDs from patients undergoing surgery for scoliosis and trauma, and some have done studies on cadaveric discs also. Though they appear normal in MRI, at the molecular level, they might not be accurate biological controls due to altered mechanical stress levels and remodeling secondary to uneven/asymmetric loading in scoliosis and endplate damage or inflammation in discs adjacent to fractures. In this study, we have characterized the proteome of adult human NP by analysing 9 IVDs harvested from brain dead yet alive organ donors which showed each sample id positively correlated with each other and also compared it with the reported proteome of discs obtained from patients with scoliosis and lumbar fractures. A vast variability of the proteome of scoliotic and lumbar fractured samples when compared to the normal proteome of our data, again stresses the fact that they might not be appropriate biological controls. All the samples analysed in our study showed positive correlation with each other (Fig. 10).

2. Optimization of Appropriate Methodology

Standardization of proteomic experiment methodology is essential for reproducibility results in experiments and requires the selection of proper buffer solutions and clean-up agents. Optimal buffer conditions are necessary for removing interfering substances and reducing the sample complexity of IVD. We compared 5 different buffers such as 1X laemelli, 2% SDS, guanidium-HCl, Tris-acetate NaCl, and RIPA and found that sequential extraction using radio RIPA buffer followed by 2% SDS resulted in the maximum yield of proteins. Further, minimal protein loss during sample clean-up is of paramount im-

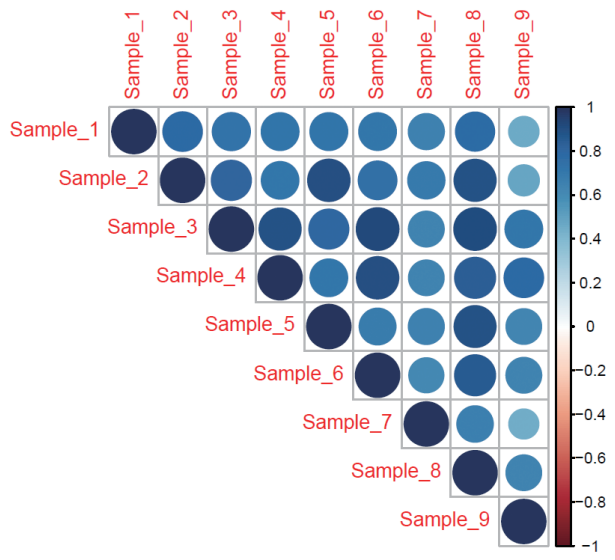


Fig. 10. Pair-wise correlation analysis for 9 intervertebral disc samples using 'corrplot' package ver. 0.84 in R (ver. 3.6.1). Positive correlations were displayed in blue color. The color intensity and the size of the circle are propotional to correlation coefficents. In the right side of the correlogram, the legend color shows the correlation coefficents and the corresponding colors. All samples show positive correlation with each other.

portance to enhance the sensitivity and scalability of protein identification. We used the following agents' acetone; ethanol; TCA-acetone; 3-kDa membrane cutoff device and methanol-chloroform; and out of these, methanol-chloroform was found to be efficient in terms of removal of glycans, salts and detergents, and recovery of total proteins. Another interesting observation was the contribution of many uncharacterized proteins and cDNA fragments to nearly 6% of the total proteins identified in the samples before the clean-up, signifying the necessity for sample clean-up process to remove such proteins.

3. Validation of the NP Phenotype

To validate the proteomic characterization of the normal NP, we employed a candidate approach for NP associated genes previously identified by transcriptome, microarray, immunolocalization, proteomics analysis of bovine,[6] murine,[7] canine,[8] and human disc tissues.[14,17,20,21] Out of 31 reported candidate genes, 15 were found in our study. Similar to other molecular studies on human NP using various platforms, our proteomics analysis also identified well-characterized core matrisome proteins such as ACAN (aggrecan), COL12A1 (collagen 12A1), DCN (decorin), LUM (lumican), ANXA3 (annexin), A2M (alpha 2 macroglobulin), and intracellular proteins such as Keratins-KRT19, KRT18, KRT8, VIM (vimentin), and DSC2 (desmocollin-2). These findings validate our study on control NP and demonstrate its normalcy.

4. Proteomic Signature of Adult Human IVD

The proteome of NP was characterized by the presence of 1,116 proteins with 57% of matrisomal proteins (core matrisome, regulators, ECM affiliated, and secreted factors) and 43% of other proteins (cytoskeletal proteins, intracellular proteins, membrane proteins, and blood components). Among the 1,116 proteins, 242 proteins with a PSM > 5 were subjected to GO analysis, which revealed that the majority of these proteins were involved in ECM organization. These proteins were mainly concentrated in the ECM, followed by the membrane and cytoplasmic regions. Further, 105 significant pathways representing mechanisms of IVD homeostasis was identified. Among these pathways, it was surprising to see that the maximum number of proteins were involved in the immune system (57 proteins) and innate immune system (45 proteins) followed by 37 proteins involved in ECM organization and Development Biology. Based on enrichment (p-values), the most significant pathway was the formation of cornified envelope followed by ECM organization and platelet degranulation. Further, the core matrisome groups of glycoproteins, proteoglycans, and collagens were involved in integrin cell surface interaction-pathway, which is believed to play a significant role in maintaining cellular homeostasis by mediating the cross-talk mechanisms between intracellular and extracellular region. Another interesting finding was that among the clusters of proteins that form tight interactions with each other, apart from collagens, glycoproteins and proteoglycans, SERPINs, and keratins were also found.

5. Collagens

Collagen 1 and 2 are the most discussed collagens in IVD, and in our study, 14 Collagen types-1, -2, -5, -6, -7, -8, -9, -10, -11, -14, -15, -18, -20, and -24 were identified in the NP. Collagen 2 is believed to be the main constituent in NP, which forms a fibrillary network and helps in aggregating proteoglycan to provide tensile strength.[22] Collagen types-3, -6, -9, and -11 have also been reported to contribute to disc homeostasis.[23] Apart from cell adhesion and cross bridging, collagen 3 are believed to be associated with collagen 6. Collagen 6 forms chondrons and fibrous capsules that can encapsulate cells and have got cytoprotective effect by counteracting apoptosis and oxidative damage. While collagen 9 cross-links to collagen 2 and regulates its size, collagen 11 plays a vital role in maintaining the

https://doi.org/10.14245/ns.2040056.028

spacing and width of type 2 collagen. However, the role of collagen types-18, -20, and -24 in IVD remains unknown and needs further research. The presence of 14 types of collagen in the NP of human IVD has never been reported before and it could be that though collagen 2 constitutes the most of NP, the regulatory and supporting roles of other collagens might be very crucial for the survival and functioning of other molecules and has been neglected by most researchers.

6. Glycoproteins

Among the highly ranked proteins, which formed a cluster, some of the notable glycoproteins which formed tight interaction with other proteins include: FN1 (fibronectin), THBS1 (thrombospondin-1), EFEMP1 (epidermal growth factor-containing fibulin-like extracellular matrix protein-1), VTN, MTN2 (matrilin-2), MTN4 (matrilin-4), MMRN1 (multimerin-1), TGFB1 (transforming growth factor beta-1), and TNC (tenascin-C). FN1 is a normal constituent of ECM known to bind collagen molecules and has an immense role in cell adhesion and cell motility. The upregulation of FN1 has been associated with aging and DDD in previous studies.[24] On the other hand, downregulation of THBS1 is associated with DDD. Matrilins are essential adhesive glycoproteins for maintaining the ECM volume and also are believed to have regenerative potential.[25] Multimerin is another adhesive glycoprotein that is soluble and has been reported to be present in endothelial cells, with its role in IVD unknown.[26] EFEMP1 also known as fibulin-3 is an essential modulator of ECM as it upregulates the expression of tissue inhibitor of metalloproteinase (TIMP)-1 & 3, thereby inhibiting the expression of MMP-2, MMP-3, and MMP-9.[27] Similarly, TGFB1 stimulates the ECM precursors and has excellent potential in tissue regeneration. TNC has immense anti-inflammatory, and antiangiogenic properties and its downregulation have been associated with aging, and DDD.[28] Vitronectin is an essential ECM glycoprotein, which binds to integrin and thus promotes cell adhesion.[29] A quantitative proteomic study estimating the expression of these molecules is critical to understand their role in aging and DDD. Apart from their role in the maintenance of ECM, a majority of glycoproteins also have metabolic and functional properties.

7. Proteoglycans

OMD (osteomodulin), FMOD, ASPN (asporin), BGN (biglycan), DCN, PRELP (prolargin), OGN (osteoglycin), LUM, and chondroadherin (CHAD) formed significant interactions among the proteoglycans. These molecules belong to a family of proteins called small leucine-rich proteoglycans (SLRPs), which have been implicated in IVD homeostasis, and DDD.[30] FMOD and LUM are close homologs, and while the accumulation of FMOD is related to aging and DDD, an abundance of LUM indicates healthy IVD. DCN and BGN are closely associated with each other and are believed to have a significant role in fetal development and its regulation. Apart from interacting with collagens, DCN, BGN, FMOD, and LUM also regulate growth factors such as TGFB and fragmentation of these SLRPs, induce an inflammatory state, and initiate the degradative process. The upregulation of ASPN negatively regulates the expression of collagen 2 and Aggrecan by inhibiting TGFB and is associated with IVD degeneration.[31]

8. Serine Protease Inhibitors

Twelve different types of SERPINs identified were found to interact with each other in our study. Some of the SERPINs such as SERPINA1, A3, and E2 have been studied *in vivo* and *in vitro* in disc disease. Neutrophil elastase is a critical regulatory molecule in degenerative and inflammatory diseases through its active involvement in proteolysis of elastin and collagen-IV of ECM. SERPINA1, which is one of the top-ranked proteins in our PPI analysis, plays a vital role in inhibiting neutrophil elastase and thereby is essential to inhibit a catabolic cascade of events.[32] While SERPINE2 inhibits the extracellular proteases via the internalization process,[33] SERPINA3 inhibits a pro-inflammatory enzyme called neutrophil cathepsin G and helps in limiting ECM remodeling, coagulation, apoptosis, and inflammatory process.[34] The expression and function of other SERPINs and their correlation in normal IVD need to be elaborated with more experimental evidence to understand its role in the maintenance of normal homeostasis.

9. Cytoskeletal Proteins

Besides SERPINs, tight interactions were noted between the intermediate filament keratins, which are one of the 3 major cytoskeletal assemblies in higher organisms.[35] Intermediate filament network play a crucial role in the spatial organization of microtubules and microfilaments or both. Keratins are responsible for cellular homeostasis, resistance to stress response exerted by mechanical and nonmechanical factors, cellular integrity, growth, and survival. Expression of KRT 8 and KRT 18, a member of intermediate filament assembly, which exerts resistance towards multiple mechanical or toxic mediated stress responses, in our study serves as a proof of concept for disc homeostasis. VMN is a cytoskeletal protein, and it was found to

be interacting with KRT8 in our study. VIM is known for several cellular functions and significant physiological changes during development and its loss is associated with DDD.[36] KRT 19, which is considered as an NP marker, was also found in our study, which assures that the samples we have used are representative of normal IVD.

10. Can Discs Harvested From Scoliosis and Lumbar Burst Fractures Be Considered Controls?

We compared our proteomic data with that of studies which have used discs from lumbar burst fractures[21] and scoliosis.[20] This analysis showed only 60 proteins common to the scoliosis group and our research compared to 51 proteins common between the lumbar burst fracture study group and our research, indicating the vast difference in both their physiological states despite appearing normal in MRI.

On analyzing the unique proteins of lumbar burst fracture disc group, it was interesting to note the presence of proteins such as HECTD1 (HECT domain E3 ubiquitin-protein ligase 1), ANKRD12 (ankyrin repeat domain 12), UBR4 (ubiquitin-protein ligase E3 component n-recognin 4), and LTN1 (listerin E3 ubiquitin-protein ligase 1) participating in either one of the processes such as class I major histocompatibility complex mediated antigen processing, antigen-presenting, and ubiquitination, which are catabolic degradative pathways mediated by the ubiquitin-proteasome complex. The upregulation of degenerative pathways in these MRI normal discs may be secondary to the traumatic molecular changes such as edema, bleeding, and endplate disruption occurring adjacent to the burst fractures. Besides, ubiquitin-mediated degradation has been reported as a very typical observation in our previous study indicating the rapid deterioration of the IVD intracellular proteins. Taken together, this illustrates that discs adjacent to LBF cannot be considered as true biological controls and studies treating them as controls will fail in biomarker discovery in DDD.

Similarly, the unique proteins from scoliotic discs such as AHSG (alpha 2-HS glycoprotein), DERA (deoxyribose phosphate aldolase), ORM1 (alpha-1-acid glycoprotein 1 precursor), PPBP (platelet basic protein precursor), VNN1 (pantetheinase precursor, vanin), PKM2 (pyruvate kinase), POTEKP (putative beta-actin-like protein 3) participate in platelet degranulation & neutrophil degranulation. Besides, the presence of proteins such as; ORM1, which is known to play a significant role in acute phase inflammation; DERA, a protein highly responsible for stress-induced damage by producing energy through deoxy nucleoside degradation and VNN1, an inducer of oxidative-stress response, clearly indicates the inflammatory response to the abnormal stress induced by asymmetric physical loads in these scoliotic discs. Therefore, these scoliotic discs are far from being considered normal discs and thus being an abysmal control; they cannot be used for comparative studies for defining the pathological states of degenerate human IVD.

Further, it is not surprising that most of the basic science research is being conducted in animal models, and regenerative therapies for DDD are under trial for their possible implications in humans. However, the basic requirement of this translational research is the need for animal tissues to resemble that of humans. However, from our analysis, it is evident that the proteome of the normal IVD of humans differs immensely from that of animal models such as bovine, murine, and canine discs. The extrapolation of results from molecular studies done in animals to solve human pathologies should be done with extreme caution considering the above facts.

11. Implications

To summarise, this study has defined the proteomic signature of human IVD for the first time by analyzing the MRI normal discs harvested from brain dead yet alive voluntary organ donors. These discs represent true biological controls, unlike discs harvested from patients with scoliosis and trauma, and therefore are ideal for comparative proteomics studies and biomarker discovery. Two essential steps for maximal identification of proteins from IVD, viz. sequential extraction using RIPA and 2% SDS followed by sample clean-up have been discussed. The various proteins, their BPs, along with pathways that play a significant role in disc homeostasis, have been enlisted. Through immunohistochemistry, we have validated the localization & expression of a subset of IVD proteins. Further, apart from collagens, proteoglycans, and glycoproteins, the importance of SERPINs and Keratins in maintaining disc homeostasis has been highlighted which needs further research to understand their regulatory mechanisms and involvement in disc degeneration.

12. Limitations

Only 9 IVDs were available for analysis at the time of this study. However, it is acceptable considering the rarity of MRI normal discs from voluntary organ donors. A comparison of proteins derived from our research with earlier reports might not be ideal as they vary in methodology. However, the total number of proteins identified in this study is much higher than the previously reported ones and it is highly unlikely that our

 https://doi.org/10.14245/ns.2040056.028

data can miss out on the proteins considered unique to the other groups. A detailed description of all the proteins and pathways is beyond the scope of this study, and each of these proteins has to analyzed individually for their role in health and disease.

CONCLUSION

We identified a total of 1,116 proteins in our study, which accounts to be the highest number of proteins reported in human IVD so far. The samples obtained from brain dead organ donors can be described as a biologically normal NP, whose cellular mechanism reflects the ideal scenario of a healthy disc. In comparison, the upregulations of inflammatory and degradative pathways in scoliosis and lumbar burst fracture discs, respectively, imply that these previously used controls are far from ideal controls. The methodology used in the analysis of the proteome also captures proteins involved in ECM homeostasis, which has not been reported until now in healthy NP.

CONFLICT OF INTEREST

The authors have nothing to disclose.

ACKNOWLEDGMENTS

We received funding from Ganga Orthopedic Research and Education Foundation (GOREF-2-2017). We thank Ms. Alishya Maria Jose for her help in sample collection and data maintenance and Ms. Sujitha M of Aravind Medical Research Foundation for LC-MS/MS acquisitions.

SUPPLEMENTARY MATERIALS

Supplementary materials can be found via https://doi.org/10.14245/ns.2040056.028.

REFERENCES

1. Hartvigsen J, Hancock MJ, Kongsted A, et al. What low back pain is and why we need to pay attention. Lancet 2018;391:2356-67.
2. Meir AR, Fairbank JC, Jones DA, et al. High pressures and asymmetrical stresses in the scoliotic disc in the absence of muscle loading. Scoliosis 2007;2:4. https://doi.org/10.1186/1748-7161-2-4.
3. Stokes IA, Burwell RG, Dangerfield PH, et al. Biomechanical spinal growth modulation and progressive adolescent scoliosis-a test of the 'vicious cycle' pathogenetic hypothesis: summary of an electronic focus group debate of the IBSE. Scoliosis 2006;1:16.
4. Sander AL, Laurer H, Lehnert T, et al. A clinically useful classification of traumatic intervertebral disk lesions. AJR Am J Roentgenol 2013;200:618-23.
5. Haschtmann D, Stoyanov JV, Gédet P, et al. Vertebral endplate trauma induces disc cell apoptosis and promotes organ degeneration in vitro. Eur Spine J 2008;17:289-99.
6. Caldeira J, Santa C, Osório H, et al. Matrisome profiling during intervertebral disc development and ageing. Sci Rep 2017;7:11629. https://doi.org/10.1038/s41598-017-11960-0.
7. McCann MR, Patel P, Frimpong A, et al. Proteomic signature of the murine intervertebral disc. PLoS One 2015;10:e0117807.
8. Erwin WM, DeSouza L, Funabashi M, et al. The biological basis of degenerative disc disease: proteomic and biomechanical analysis of the canine intervertebral disc. Arthritis Res Ther 2015;17:240. https://doi.org/10.1186/s13075-015-0733-z.
9. Zybailov B, Mosley AL, Sardiu ME, et al. Statistical analysis of membrane proteome expression changes in Saccharomyces cerevisiae. J Proteome Res 2006;5:2339-47.
10. Jeong H, Mason SP, Barabási AL, et al. Lethality and centrality in protein networks. Nature 2001;411:41-2.
11. Sive JI, Baird P, Jeziorsk M, et al. Expression of chondrocyte markers by cells of normal and degenerate intervertebral discs. Mol Pathol 2002;55:91-7.
12. Pattappa G, Li Z, Peroglio M, et al. Diversity of intervertebral disc cells: phenotype and function. J Anat 2012;221:480-96.
13. Lee CR, Sakai D, Nakai T, et al. A phenotypic comparison of intervertebral disc and articular cartilage cells in the rat. Eur Spine J 2007;16:2174-85.
14. Minogue BM, Richardson SM, Zeef LA, et al. Transcriptional profiling of bovine intervertebral disc cells: implications for identification of normal and degenerate human intervertebral disc cell phenotypes. Arthritis Res Ther 2010;12:R22.
15. Sakai D, Nakai T, Mochida J, et al. Differential phenotype of intervertebral disc cells: microarray and immunohistochemical analysis of canine nucleus pulposus and anulus fibrosus. Spine (Phila Pa 1976) 2009;34:1448-56.
16. Rodrigues-Pinto R, Richardson SM, Hoyland JA. Identification of novel nucleus pulposus markers: interspecies varia-

tions and implications for cell-based therapiesfor intervertebral disc degeneration. Bone Joint Res 2013;2:169-78.
17. Richardson SM, Ludwinski FE, Gnanalingham KK, et al. Notochordal and nucleus pulposus marker expression is maintained by sub-populations of adult human nucleus pulposus cells through aging and degeneration. Sci Rep 2017;7:1501.
18. Fujita N, Miyamoto T, Imai J, et al. CD24 is expressed specifically in the nucleus pulposus of intervertebral discs. Biochem Biophys Res Commun 2005;338:1890-6.
19. Tang X, Jing L, Chen J. Changes in the molecular phenotype of nucleus pulposus cells with intervertebral disc aging. PLoS One 2012;7:e52020.
20. Sarath Babu N, Krishnan S, Brahmendra Swamy CV, et al. Quantitative proteomic analysis of normal and degenerated human intervertebral disc. Spine J 2016;16:989-1000.
21. Yee A, Lam MP, Tam V, et al. Fibrotic-like changes in degenerate human intervertebral discs revealed by quantitative proteomic analysis. Osteoarthritis Cartilage 2016;24:503-13.
22. Lian C, Gao B, Wu Z, et al. Collagen type II is downregulated in the degenerative nucleus pulposus and contributes to the degeneration and apoptosis of human nucleus pulposus cells. Mol Med Rep 2017;16:4730-6.
23. Galbusera F, Wilke HJ. The mechanical role of collagen fibers in the intervertebral disc. In: Sharabi M, Wade K, Haj-Ali R. Biomechanics of the spine: basic concepts, spinal disorders and treatments. London: Academic Press; 2018. p. 105-23.
24. Greg Anderson D, Li X, Tannoury T, et al. A fibronectin fragment stimulates intervertebral disc degeneration in vivo. Spine (Phila Pa 1976) 2003;28:2338-45.
25. Malin D, Sonnenberg-Riethmacher E, Guseva D, et al. The extracellular-matrix protein matrilin 2 participates in peripheral nerve regeneration. J Cell Sci 2009;122(Pt 7):995-1004.
26. Hayward CP, Cramer EM, Song Z, et al. Studies of multimerin in human endothelial cells. Blood 1998;91:1304-17.
27. Lin Z, Wang Z, Li G, et al. Fibulin-3 may improve vascular health through inhibition of MMP-2/9 and oxidative stress in spontaneously hypertensive rats. Mol Med Rep 2016;13:3805-12.
28. Gruber HE, Ingram JA, Hanley EN Jr. Tenascin in the human intervertebral disc: alterations with aging and disc degeneration. Biotech Histochem 2002;77:37-41.
29. Wikipedia. Vitronectin [Internet]. 2019 [cited 2020 Jan 31]. Available from: https://en.wikipedia.org/w/index.php?title=Vitronectin&oldid=887410367.
30. Brown S, Melrose J, Caterson B, et al. A comparative evaluation of the small leucine-rich proteoglycans of pathological human intervertebral discs. Eur Spine J 2012;21 Suppl 2:S154-9.
31. Wang S, Liu C, Sun Z, et al. IL-1β increases asporin expression via the NF-κB p65 pathway in nucleus pulposus cells during intervertebral disc degeneration. Sci Rep 2017;7:4112. https://doi.org/10.1038/s41598-017-04384-3.
32. Morris CA, Pani AM, Mervis CB, et al. Alpha 1 antitrypsin deficiency alleles are associated with joint dislocation and scoliosis in Williams syndrome. Am J Med Genet C Semin Med Genet 2010;154C:299-306.
33. Low DA, Baker JB, Koonce WC, et al. Released protease-nexin regulates cellular binding, internalization, and degradation of serine proteases. Proc Natl Acad Sci U S A 1981;78:2340-4.
34. Chelbi ST, Wilson ML, Veillard AC, et al. Genetic and epigenetic mechanisms collaborate to control SERPINA3 expression and its association with placental diseases. Hum Mol Genet 2012;21:1968-78.
35. Jacob JT, Coulombe PA, Kwan R, et al. Types I and II Keratin Intermediate Filaments. Cold Spring Harb Perspect Biol 2018 Apr 2;10(4). pii: a018275. https://doi.org/10.1101/cshperspect.a018275.
36. Ivaska J, Pallari HM, Nevo J, et al. Novel functions of vimentin in cell adhesion, migration, and signaling. Exp Cell Res 2007;313:2050-62.

https://doi.org/10.14245/ns.2040056.028

Supplementary Table 1. Proteins reported in literature versus this study

Gene symbol	Name	Organisms	Method	Presence/ absence	Reference
Sox9	Transcription factor Sox-9-A	Bovine, Human, canine	In-situ hybridization	-	Sive et al., 2002[11]
ACAN	Aggrecan	Bovine, Human, canine	Microarray	+	Pattappa et al., 2012[12]
COL2A1	Collagen Type II A	Bovine, Human, canine	Microarray	+	Lee et al., 2007[13]
KRT19	Keratin 19	Bovine, Human, canine	Microarray	+	Minogue et al., 2010[14]
KRT18	Keratin 18	Bovine, Human, canine	Microarray	+	Minogue et al., 2010[14], Sakai et al., 2009[15]
NCAM1	Neural cell adhesion molecule 1	Bovine, canine	Microarray	-	Minogue et al., 2010[14],Sakai et al, 2009[15]
KRT8	Keratin 8	Bovine, Human, canine, Murine	Transcriptomic profiling	+	Minogue et al., 2010[14]
SNAP25	Synaptosomal-associated protein 25	Bovine	Transcriptomic profiling	-	Minogue et al., 2010[14]
CDH2	Cadherin-2	Bovine	Transcriptomic profiling	-	Minogue et al., 2010[14]
SOSTDC1	Sclerostin domain-containing protein 1	Bovine	Transcriptomic profiling	-	Minogue et al., 2010[14]
GPC3	Glypican-3	Murine	Transcriptomic profiling, proteome analysis	-	Lee et al., 2007[13]
ANXA3	Annexin A3	Murine	Transcriptomic profiling	+	Lee et al., 2007[13]
PTN	Pleiotrophin	Murine	Transcriptomic profiling	-	Lee et al., 2007[13]
A2M	alpha-2-macroglobulin	Human	Microarray	+	Rodrigues-Pinto et al, 2013[16]
DSC2	Desmocollin-2, cell surface markers	Human, Canine	Microarray	+	Sakai et al., 2009[15]
VIM	Vimentin	Murine	Transcriptomic profiling	+	Lee et al., 2007[13]
Pax-1	Paired box protein Pax-1	Human	Transcriptomic profiling	-	Richardson et al., 2017[17]
CA12	carbonic anhydrase-12	Human	Transcriptomic profiling	-	Richardson et al., 2017[17], Minogue et al., 2010[14]
CD24	cluster differentiation 24 (cell surface marker)	Murine	Transcriptome	-	Fujita et al., 2005[18], Tang et al., 2012[19]
TNFAIP6	TNF alpha-induced protein 6	Bovine	Microarray analysis	+	Minogue et al., 2010[14]
FOXF1	N-cad, forkhead box F1	Bovine, Human	Microarray analysis	-	Richardson et al., 2017[17], Minogue et al., 2010[14]
FOXF2	Forkhead box protein F2	Bovine	Microarray analysis	-	Minogue et al., 2010[14]

Candidate IVD-associated proteins/genes previously reported based on the proteome, transcriptional, immunolocation analysis of human, murine, bovine and canine IVD.
+/- indicates the presence/absence in our study.

Supplementary Table 2. Detailed list of 105 enriched pathways from STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) database

No.	Enriched pathways	Total number of proteins involved	p-value	Gene symbols
1	Extracellular matrix organization	37	8.90E-24	*A2M, ACAN, ASPN, BGN, CD47, COL10A1, COL11A2, COL14A1, COL15A1, COL1A1, COL1A2, COL2A1, COL6A1, COL6A2, COL6A3, COMP, DCN, FGA, FGB, FGG, FMOD, GDF5, HAPLN1, HSPG2, HTRA1, LOX, LUM, MMP10, PCOLCE, PCOLCE2, PPIB, THBS1, TIMP1, TNC, TTR, VCAN, VTN*
2	ECM proteoglycans	18	9.26E-16	*ACAN, ASPN, BGN, COL1A1, COL1A2, COL2A1, COL6A1, COL6A2, COL6A3, COMP, DCN, FMOD, HAPLN1, HSPG2, LUM, TNC, VCAN, VTN*
3	Integrin cell surface interactions	17	6.49E-14	*CD47, COL10A1, COL1A1, COL1A2, COL2A1, COL6A1, COL6A2, COL6A3, COMP, FGA, FGB, FGG, HSPG2, LUM, THBS1, TNC, VTN*
4	Collagen biosynthesis and modifying enzymes	13	1.63E-10	*COL10A1, COL11A2, COL14A1, COL15A1, COL1A1, COL1A2, COL2A1, COL6A1, COL6A2, COL6A3, PCOLCE, PCOLCE2, PPIB*
5	Collagen formation	14	2.61E-10	*COL10A1, COL11A2, COL14A1, COL15A1, COL1A1, COL1A2, COL2A1, COL6A1, COL6A2, COL6A3, LOX, PCOLCE, PCOLCE2, PPIB*
6	Glycosaminoglycan metabolism	12	4.88E-07	*ACAN, B3GNT7, BGN, DCN, FMOD, HSPG2, LUM, OGN, OMD, PRELP, VCAN, XYLT1*
7	Keratan sulfate degradation	6	5.80E-07	*ACAN, FMOD, LUM, OGN, OMD, PRELP*
8	Gluconeogenesis	5	0.00054	*ENO1, GAPDH, PGAM2, PGK1, TPI1*
9	Formation of the cornified envelope	29	1.66E-24	*CDSN, DSG1, DSP, JUP, KRT1, KRT10, KRT14, KRT15, KRT16, KRT17, KRT19, KRT2, KRT33A, KRT4, KRT5, KRT6A, KRT6B, KRT6C, KRT72, KRT73, KRT76, KRT77, KRT78, KRT79, KRT8, KRT80, KRT84, KRT9, TGM5*
10	Platelet degranulation	22	1.04E-16	*A1BG, A2M, ALB, ANXA5, APOA1, CD109, CLEC3B, CLU, F13A1, FGA, FGB, FGG, HRG, ITIH4, QSOX1, SERPINA1, SERPING1, TF, THBS1, TIMP1, TIMP3, TTN*
11	Platelet activation, signalling and aggregation	25	6.50E-14	*A1BG, A2M, ALB, ANXA5, APOA1, CD109, CLEC3B, CLU, COL1A1, COL1A2, F13A1, FGA, FGB, FGG, HRG, ITIH4, QSOX1, RASGRP2, SERPINA1, SERPING1, TF, THBS1, TIMP1, TIMP3, TTN*
12	Innate immune system	45	2.71E-13	*A1BG, ACTG1, ANXA2, ARG1, C1S, C3, C9, CALML5, CASP8, CD47, CFB, CFI, CHI3L1, CLU, DCD, DSG1, DSP, FABP5, FGA, FGB, FGG, FLG2, FTL, GSN, HBB, HP, HRNR, JUP, KRT1, LTF, LYZ, MIF, PKM, PLA2G2A, QSOX1, RASGRP2, S100A1, S100A8, S100A9, SERPINA1, SERPINB1, SERPING1, TNFAIP6, TTR, VTN*
13	Hemostasis	34	1.31E-12	*A1BG, A2M, ALB, ANXA2, ANXA5, APOA1, CD109, CD47, CLEC3B, CLU, COL1A1, COL1A2, F13A1, FGA, FGB, FGG, HBB, HBD, HRG, ITIH4, KIF1B, MIF, QSOX1, RASGRP2, SERPINA1, SERPINA5, SERPINC1, SERPINE2, SERPING1, TF, THBS1, TIMP1, TIMP3, TTN*
14	Neutrophil degranulation	29	1.53E-11	*A1BG, ANXA2, ARG1, C3, CALML5, CD47, CHI3L1, DSG1, DSP, FABP5, FLG2, FTL, GSN, HBB, HP, HRNR, JUP, KRT1, LTF, LYZ, MIF, PKM, QSOX1, S100A8, S100A9, SERPINA1, SERPINB1, TNFAIP6, TTR*
15	Degradation of the extracellular matrix	17	7.17E-11	*A2M, ACAN, COL10A1, COL11A2, COL14A1, COL15A1, COL1A1, COL1A2, COL2A1, COL6A1, COL6A2, COL6A3, DCN, HSPG2, HTRA1, MMP10, TIMP1*
16	Post-translational protein phosphorylation	15	1.94E-10	*ALB, APOA1, C3, CP, FGA, FGG, MFGE8, MFI2, QSOX1, SERPINA1, SERPINC1, TF, TIMP1, TNC, VCAN*
17	Immune system	57	3.58E-10	*A1BG, ACTG1, ANXA1, ANXA2, ARG1, BLMH, C1S, C3, C9, CA1, CALML5, CASP8, CD47, CFB, CFI, CHI3L1, CLU, COL1A1, COL1A2, COL2A1, CSF3R, DCD, DSG1, DSP, F13A1, FABP5, FGA, FGB, FGG, FLG2, FTL, GSN, HBB, HP, HRNR, JUP, KRT1, LTF, LYZ, MIF, MSN, PKM, PLA2G2A, QSOX1, RASGRP2, S100A1, S100A8, S100A9, SERPINA1, SERPINB1, SERPING1, TIMP1, TNFAIP6, TNFRSF11B, TTR, VIM, VTN*

(*Continued to the next page*)

https://doi.org/10.14245/ns.2040056.028

https://doi.org/10.14245/ns.2040056.028

www.e-neurospine.org

Supplementary Table 2. Continued

No.	Enriched pathways	Total number of proteins involved	p-value	Gene symbols
18	Assembly of collagen fibrils and other multimeric structures	12	6.15E-10	*COL10A1, COL11A2, COL14A1, COL15A1, COL1A1, COL1A2, COL2A1, COL6A1, COL6A2, COL6A3, LOX, PCOLCE*
19	regulation of insulin-like growth factor transport and uptake by insulin-like growth factor binding proteins	15	1.03E-09	*ALB, APOA1, C3, CP, FGA, FGG, MFGE8, MFI2, QSOX1, SERPINA1, SERPINC1, TF, TIMP1, TNC, VCAN*
20	Diseases associated with glycosaminoglycan metabolism	10	3.96E-09	*ACAN, BGN, DCN, FMOD, HSPG2, LUM, OGN, OMD, PRELP, VCAN*
21	Collagen chain trimerization	10	8.43E-09	*COL10A1, COL11A2, COL14A1, COL15A1, COL1A1, COL1A2, COL2A1, COL6A1, COL6A2, COL6A3*
22	Developmental biology	37	1.28E-08	*ACTG1, ANK2, CDSN, COL2A1, COL6A1, COL6A2, COL6A3, DSG1, DSP, H2AFV, JUP, KRT1, KRT10, KRT14, KRT15, KRT16, KRT17, KRT19, KRT2, KRT33A, KRT4, KRT5, KRT6A, KRT6B, KRT6C, KRT72, KRT73, KRT76, KRT77, KRT78, KRT79, KRT8, KRT80, KRT84, KRT9, MSN, TGM5*
23	Collagen degradation	11	1.37E-08	*COL10A1, COL11A2, COL14A1, COL15A1, COL1A1, COL1A2, COL2A1, COL6A1, COL6A2, COL6A3, MMP10*
24	Formation of fibrin clot (clotting cascade)	9	5.04E-08	*A2M, F13A1, FGA, FGB, FGG, SERPINA5, SERPINC1, SERPINE2, SERPING1*
25	Defective ST3GAL3 causes MCT12 and EIEE15	6	5.33E-08	*ACAN, FMOD, LUM, OGN, OMD, PRELP*
26	Binding and uptake of ligands by scavenger receptors	9	5.63E-08	*ALB, APOA1, COL1A1, COL1A2, FTL, HBA1, HBB, HP, HPX*
27	Nonintegrin membrane-ECM interactions	10	6.87E-08	*COL10A1, COL11A2, COL1A1, COL1A2, COL2A1, HSPG2, THBS1, TNC, TTR, VTN*
28	Amyloid fiber formation	11	7.28E-08	*APCS, APOA1, APOA4, FGA, GSN, HSPG2, LTF, LYZ, MFGE8, TGFBI, TTR*
29	Defective CHST6 causes MCDC1	6	7.93E-08	*ACAN, FMOD, LUM, OGN, OMD, PRELP*
30	Defective B4GALT1 causes B4GALT1-CDG (CDG-2d)	6	7.93E-08	*ACAN, FMOD, LUM, OGN, OMD, PRELP*
31	Metabolism of carbohydrates	17	3.11E-07	*ACAN, B3GNT7, BGN, DCN, ENO1, FMOD, GAPDH, HSPG2, LUM, OGN, OMD, PGAM2, PGK1, PRELP, TPI1, VCAN, XYLT1*
32	Common pathway of fibrin clot formation	7	3.54E-07	*F13A1, FGA, FGB, FGG, SERPINA5, SERPINC1, SERPINE2*
33	Scavenging of heme from plasma	6	5.80E-07	*ALB, APOA1, HBA1, HBB, HP, HPX*
34	Keratan sulfate biosynthesis	7	1.03E-06	*ACAN, B3GNT7, FMOD, LUM, OGN, OMD, PRELP*
35	Regulation of complement cascade	8	2.02E-06	*C1S, C3, C9, CFB, CFI, CLU, SERPING1, VTN*
36	Regulation of TLR by endogenous ligand	6	3.19E-06	*FGA, FGB, FGG, S100A1, S100A8, S100A9*
37	Diseases of glycosylation	11	9.04E-06	*ACAN, BGN, DCN, FMOD, HSPG2, LUM, OGN, OMD, PRELP, THBS1, VCAN*

(*Continued to the next page*)

Supplementary Table 2. Continued

No.	Enriched pathways	Total number of proteins involved	p-value	Gene symbols
38	Erythrocytes take up oxygen and release carbon dioxide	4	7.48E-05	*CA1, CA2, HBA1, HBB*
39	Platelet aggregation (plug formation)	6	8.18E-05	*COL1A1, COL1A2, FGA, FGB, FGG, RASGRP2*
40	Intrinsic pathway of fibrin clot formation	5	0.00011	*A2M, SERPINA5, SERPINC1, SERPINE2, SERPING1*
41	A tetrasaccharide linker sequence is required for GAG synthesis	5	0.00022	*BGN, DCN, HSPG2, VCAN, XYLT1*
42	Syndecan interactions	5	0.00022	*COL1A1, COL1A2, THBS1, TNC, VTN*
43	Erythrocytes take up carbon dioxide and release oxygen	4	0.00024	*CA1, CA2, HBA1, HBB*
44	Apoptotic execution phase	6	0.00037	*CASP8, DSG1, DSP, GSN, HIST1H1C, VIM*
45	Signaling by PDGF	6	0.00054	*COL2A1, COL6A1, COL6A2, COL6A3, THBS1, THBS3*
46	Antimicrobial peptides	7	0.00076	*CLU, DCD, LTF, LYZ, PLA2G2A, S100A8, S100A9*
47	Crosslinking of collagen fibrils	4	0.0008	*COL1A1, COL1A2, LOX, PCOLCE*
48	Scavenging by class A receptors	4	0.0008	*APOA1, COL1A1, COL1A2, FTL*
49	Apoptotic cleavage of cellular proteins	5	0.00082	*CASP8, DSG1, DSP, GSN, VIM*
50	Metal sequestration by antimicrobial proteins	3	0.00088	*LTF, S100A8, S100A9*
51	Defective B4GALT7 causes EDS, progeroid type	4	0.0011	*BGN, DCN, HSPG2, VCAN*
52	Defective B3GAT3 causes JDSSDHD	4	0.0011	*BGN, DCN, HSPG2, VCAN*
53	Defective B3GALT6 causes EDSP2 and SEMDJL1	4	0.0011	*BGN, DCN, HSPG2, VCAN*
54	Signaling by receptor tyrosine kinases	15	0.0013	*ACTG1, CILP, COL11A2, COL1A1, COL1A2, COL2A1, COL6A1, COL6A2, COL6A3, FGFBP2, HSPB1, JUP, MST1, THBS1, THBS3*
55	Defective CHST3 causes SEDCJD	3	0.0016	*BGN, DCN, VCAN*
56	Defective CHST14 causes EDS, musculocontractural type	3	0.0016	*BGN, DCN, VCAN*
57	Defective CHSY1 causes TPBS	3	0.0016	*BGN, DCN, VCAN*
58	Retinoid metabolism and transport	5	0.0016	*APOA1, APOA4, HSPG2, RBP4, TTR*
59	Integrin alphaIIb beta3 signaling	4	0.002	*FGA, FGB, FGG, RASGRP2*

(*Continued to the next page*)

https://doi.org/10.14245/ns.2040056.028

Supplementary Table 2. Continued

No.	Enriched pathways	Total number of proteins involved	p-value	Gene symbols
60	Dermatan sulfate biosynthesis	3	0.0029	*BGN, DCN, VCAN*
61	MET activates PTK2 signaling	4	0.0031	*COL11A2, COL1A1, COL1A2, COL2A1*
62	Caspase-mediated cleavage of cytoskeletal proteins	3	0.0035	*CASP8, GSN, VIM*
63	Disease	24	0.0039	*ACAN, ACTG1, ALB, BGN, CP, DCN, FGA, FGB, FGG, FMOD, HSPG2, IPO5, LTF, LUM, OGN, OMD, PRDX1, PRDX2, PRELP, RBP4, THBS1, TTR, VCAN, YWHAE*
64	GRB2:SOS provides linkage to MAPK signaling for Integrins	3	0.0042	*FGA, FGB, FGG*
65	Dissolution of fibrin clot	3	0.0042	*ANXA2, HRG, SERPINE2*
66	Cell-Cell communication	7	0.0045	*ACTG1, ANG, CD47, FLNC, JUP, KRT14, KRT5*
67	Cell junction organization	6	0.0045	*ACTG1, ANG, FLNC, JUP, KRT14, KRT5*
68	CS/DS degradation	3	0.0047	*BGN, DCN, VCAN*
69	p130Cas linkage to MAPK signaling for integrins	3	0.0047	*FGA, FGB, FGG*
70	Signaling by high-kinase activity BRAF mutants	4	0.0051	*ACTG1, FGA, FGB, FGG*
71	Visual phototransduction	6	0.0055	*APOA1, APOA4, HSPG2, RBP4, RDH8, TTR*
72	Glycolysis	5	0.0064	*ENO1, GAPDH, PGAM2, PGK1, TPI1*
73	Gene and protein expression by JAK-STAT signaling after Interleukin-12 stimulation	4	0.0065	*ANXA2, CA1, MIF, MSN*
74	MAP2K and MAPK activation	4	0.007	*ACTG1, FGA, FGB, FGG*
75	Signaling by moderate kinase activity BRAF mutants	4	0.007	*ACTG1, FGA, FGB, FGG*
76	Paradoxical activation of RAF signaling by kinase inactive BRAF	4	0.007	*ACTG1, FGA, FGB, FGG*
77	Plasma lipoprotein assembly	3	0.008	*A2M, APOA1, APOA4*
78	NCAM1 interactions	4	0.0085	*COL2A1, COL6A1, COL6A2, COL6A3*
79	Chondroitin sulfate biosynthesis	3	0.0089	*BGN, DCN, VCAN*
80	TLR cascades	7	0.0099	*CASP8, FGA, FGB, FGG, S100A1, S100A8, S100A9*
81	Initial triggering of complement	3	0.0113	*C1S, C3, CFB*
82	Alternative complement activation	2	0.012	*C3, CFB*

(*Continued to the next page*)

https://doi.org/10.14245/ns.2040056.028

Supplementary Table 2. Continued

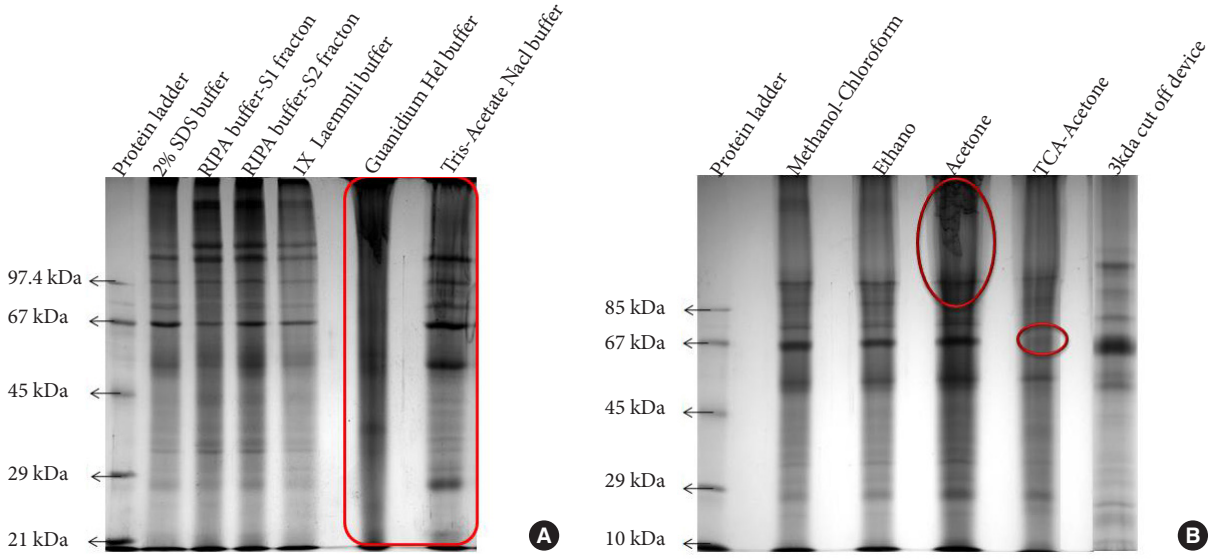
No.	Enriched pathways	Total number of proteins involved	p-value	Gene symbols
83	Deregulated CDK5 triggers multiple neurodegenerative pathways in Alzheimer disease models	3	0.0123	*PRDX1, PRDX2, YWHAE*
84	Apoptosis	7	0.0135	*CASP8, DSG1, DSP, GSN, HIST1H1C, VIM, YWHAE*
85	The canonical retinoid cycle in rods (twilight vision)	3	0.0135	*RBP4, RDH8, TTR*
86	Signaling by RAS mutants	4	0.0155	*ACTG1, FGA, FGB, FGG*
87	Activation of C3 and C5	2	0.0185	*C3, CFB*
88	HDL assembly	2	0.0185	*A2M, APOA1*
89	Terminal pathway of complement	2	0.0222	*C9, CLU*
90	Signaling by BRAF and RAF fusions	4	0.0225	*ACTG1, FGA, FGB, FGG*
91	Plasma lipoprotein remodeling	3	0.024	*ALB, APOA1, APOA4*
92	HDL remodeling	2	0.026	*ALB, APOA1*
93	Adherens junctions interactions	3	0.0278	*ACTG1, ANG, JUP*
94	Chylomicron assembly	2	0.0304	*APOA1, APOA4*
95	Chylomicron remodeling	2	0.0304	*APOA1, APOA4*
96	Vesicle-mediated transport	15	0.0308	*ACTG1, ALB, ANK2, APOA1, COL1A1, COL1A2, FTL, HBA1, HBB, HP, HPX, KIF1B, SERPINA1, TF, YWHAE*
97	Plasma lipoprotein assembly, remodeling, and clearance	4	0.0318	*A2M, ALB, APOA1, APOA4*
98	Interleukin-4 and Interleukin-13 signaling	5	0.0318	*ANXA1, COL1A2, F13A1, TIMP1, VIM*
99	Detoxification of reactive oxygen species	3	0.0332	*PRDX1, PRDX2, SOD3*
100	Apoptotic cleavage of cell adhesion proteins	2	0.0338	*DSG1, DSP*
101	GP1b-IX-V activation signalling	2	0.0338	*COL1A1, COL1A2*
102	Type I hemidesmosome assembly	2	0.0338	*KRT14, KRT5*
103	Reversible hydration of carbon dioxide	2	0.0382	*CA1, CA2*
104	Retinoid cycle disease events	2	0.0382	*RBP4, TTR*
105	G alpha (i) signalling events	10	0.0499	*AGT, ANXA1, APOA1, APOA4, C3, HSPG2, INSL5, RBP4, RDH8, TTR*

ECM, extracellular matrix; TLR, Toll-like receptor; GAG, glycosaminoglycans; PDGF, platelet-derived growth factor; EDS, Ehlers–Danlos syndrome; TPBS, temtamy preaxial brachydactyly syndrome; PTK2, protein tyrosine kinase 2; MAPK, mitogen-activated protein kinase; HDL, high-density lipoprotein.

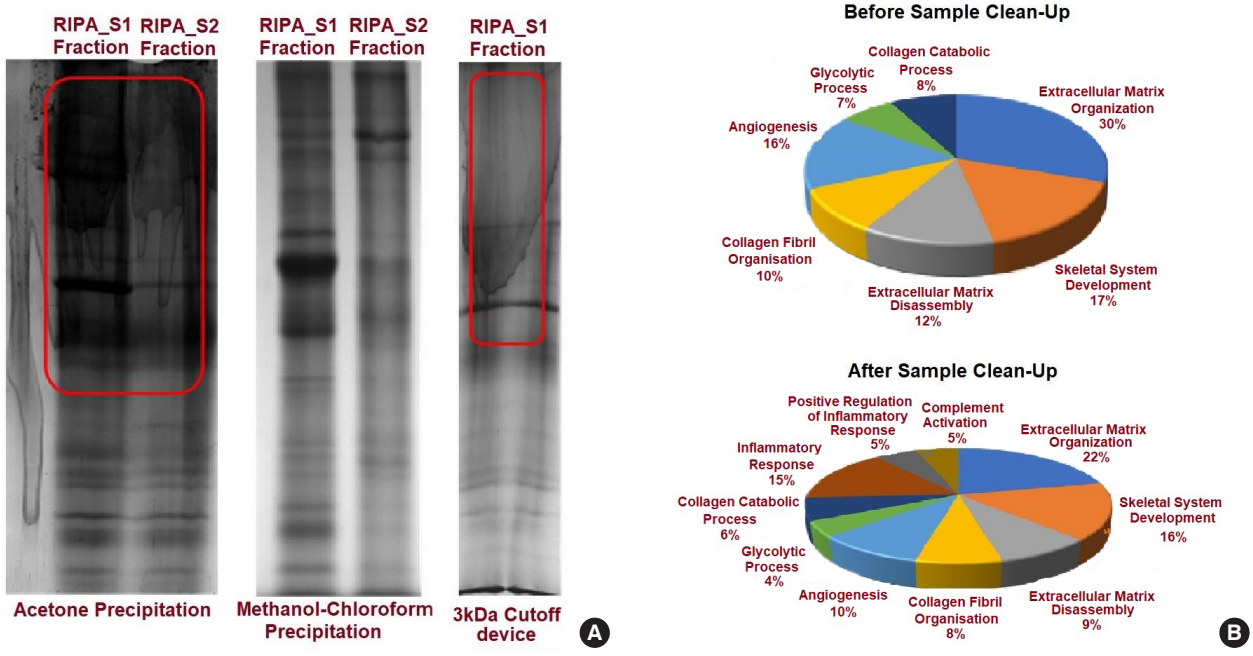
Supplementary Table 3. Catalog of proteins that were found in our study and other literatures

Study population of MRI normal discs	Collagens	Keratins	SERPIN super family of proteins
Discs harvested from brain dead organ donors in living conditions-Our study	COL1A2	KRT1	SERPINA1
	COL1A1	KRT2	SERPINA3
	COL2A1	KRT4	SERPINA4
	COL5A3	KRT5	SERPINA5
	COL6A1	KRT6A	SERPINB1
	COL6A2	KRT6C	SERPINC1
	COL6A6	KRT15	SERPINE1
	COL7A1	KRT16	SERPINE2
	COL8A1	KRT17	SERPINF1
	COL9A2	KRT19	SERPINF2
	COL10A1	KRT33A	SERPING1
	COL11A2	KRT72	SERPINI1
	COL14A1	KRT73	
	COL15A1	KRT76	
	COL18A1	KRT79	
	COL20A1	KRT80	
	COL24A1	KRT84	
		KRT85	
Samples harvested from discs during scoliosis surgery- Yee et al., 2016[19]	COL1A2	Nil	Nil
	COL14A1		
	COL16A1		
	COL20A1		
Samples harvested from discs during scoliosis surgery- Yee et al., 2016[19]	COL1A1	KRT1	SERPINA1
	COL1A2	KRT2	SERPINC1
	COL2A1		SERPINE2
	COL3A1		
	COL6A3		

MRI, magnetic resonance imaging; SERPIN, serine protease inhibitors.



Supplementary Fig. 1. Comparative proteome profiling of nucleus pulposus observed during protein extraction and protein clean-up followed by prefractionation on 10% sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis. (A) Total protein extraction using various buffers: 2% SDS; radio immunoprecipitation assay buffer (RIPA) buffer (S1 fraction) + 2% SDS (S2 fraction); 1X laemmli buffer; guanidium-HCl buffer; tris-acetate buffer, red-colored circle indicates spillage pattern due to interfering glycans. (B) Sample clean-up using various organic solvents after extraction of proteins to remove the interfering glycans before proteome analysis by ESI-LC-MS/MS. The red circle indicates the interfering glycans which were not removed after precipitation with acetone. With tricholoroacetic acid (TCA)-acetone, the loss of proteins is shown with a circle in the respective lane. ESI-LC-MS/MS, electrospray ionization-liquid chromatography tandem mass spectrometry.



Supplementary Fig. 2. Comparative profile of the intervertebral disc nucleus pulposus tissue on a 10% sodium dodecyl sulfate-polyacrylamide gel electrophoresis after sample clean-up using: acetone; methanol-chloroform; 3-kDa membrane cutoff device. (A) Presence of interfering glycan's during prefractionation (circled in red), in acetone precipitation and 3 kDa, cutoff device compared to methanol-chloroform. (B) Biological processes of proteins identified before and after sample clean-up with methanol-chloroform organic solvent. RIPA, radio immunoprecipitation assay buffer.

https://doi.org/10.14245/ns.2040056.028