



OPEN

Part 1: profiling extra cellular matrix core proteome of human fetal nucleus pulposus in search for regenerative targets

Shanmuganathan Rajasekaran¹✉, Chitraa Tangavel², Niek Djuric^{2,3}, Muthurajan Raveendran⁴, Dilip Chand Raja Soundararajan¹, Sharon Miracle Nayagam², Monica Steffi Matchado² & K. S. Sri Vijay Anand¹

Intervertebral disc degeneration is accompanied by a loss of Extra-cellular matrix (ECM) due to an imbalance in anabolic and catabolic pathways. Identifying ECM proteins with anabolic and/or regenerative potential could be the key to developing regenerative therapies. Since human fetal discs grow and develop rapidly, studying these discs may provide valuable insights on proteins with regenerative potential. This study compares core matrisome of 9 fetal and 7 healthy adult (age 22–79) nucleus pulposus (NP), using a proteomic and bioinformatic approach. Of the 33 upregulated proteins in fetus NP's, 20 of which were involved in ECM assembly pathways: fibromodulin, biglycan, heparan sulfate proteoglycan 2, chondroitin sulfate proteoglycan 4, procollagen C-endopeptidase enhancer and Collagen—type 1a1, 1a2, 6a1, 6a3, 11a1, 11a2, 12a1, 14a1 and 15a1. Moreover, 10 of the upregulated proteins were involved in growth pathways 'PI3L-Akt signaling' and 'regulation of insulin like growth factor transport and uptake.' Thrombospondin 1,3 and 4, tenascin C, matrilin-3, and collagen- type 1a1, 1a2, 6a1, 6a3 and 9a1. Additionally, matrilin-2 and 'Collagen triple helix repeat containing 1' were identified as possible regenerative proteins due to their involvement in 'Regeneration' and 'tissue development' respectively. In conclusion, the consistency of human fetal NP's differs greatly from that of healthy adults. In view of these outcomes, the core matrisome of human fetal discs contains an abundant number of proteins that could potentially show regenerative properties, and their potential should be explored in future machinal experiments.

A major supportive structure of the human adult spine is the intervertebral disc (IVD). The IVD consists of three major components: firstly, the nucleus pulposus (NP), which is a gel-like structure located at the inner part of the IVD; secondly, the annulus fibrosus (AF), a fibrous structure which is layered around the NP, and at last, the cartilage endplate (CEP), which is a thin cartilage sheet between the disc and the vertebral body¹. Together, these three structures create an elastic, hydrodynamic, weight bearing cushion for the spine, thereby allowing it to absorb compressive loads while remaining flexible^{2–4}.

During aging, the structure and components of the IVD change, especially in the NP, since it has a great variety in ECM proteins^{5,6}. The dry weight of a NP consists approximately 70% of ECM, in which collagen type 2 and proteoglycans are the two most abundant proteins types⁶. These proteoglycans integrate with the collagen fibrils and elastin to create a well-structured ECM that contributes to the strength and elasticity of the disc^{4,7}. Besides, proteoglycans are highly charged and they are also responsible for the hydraulic pressure of the disc⁴. This pressure, in turn, is essential for its weight bearing capacity⁵. During aging, the disc tends to slowly lose its strength and elasticity. While in many cases, this does not cause serious problems; some discs will start to degenerate⁴. During disc degeneration (DD), catabolic pathways are upregulated and the anabolic pathways are downregulated⁸. This imbalance leads to a decrease in ECM protein content, which results in a reduction in size

¹Department of Spine Surgery, Ganga Hospital, 313, Mettupalayam Road, Coimbatore 641043, India. ²Ganga Research Centre, No 91, Mettupalayam Road, Coimbatore 641030, India. ³Department of Neurosurgery, Leiden University Medical Center, Leiden, The Netherlands. ⁴Department of Plant Biotechnology, Tamil Nadu Agricultural University, Coimbatore 641003, India. ✉email: rajasekaran.orth@gmail.com

and strength and is often accompanied by inflammation^{8,9}. Due to these changes, patients with DD usually suffer from low back pain, which is currently one of the leading causes of physical disability worldwide¹⁰. As of today, the only effective treatment for DD is surgery which urges the need for non-surgical-regenerative therapies. Such therapies could lead to major improvements in terms of both disability and healthcare costs.

As mentioned, the pathophysiology of this highly disabling disease involves a dysregulation of the catabolic/anabolic balance⁸. Therefore, it is crucial to explore all possible anabolic mechanisms that can be exploited. Recent research has focused on exploring these pathways in both healthy and degenerated discs^{4,6,8}. Stem cells, proteins and genetic therapeutic options are being explored for regenerative potential in disc degeneration using animal models and in-vitro culture methods. However, their successful integration in to human clinical practice is challenged by the intricate biology and homeostatic mechanisms of nucleus pulposus. Recently, nucleus pulposus has been identified to have two types of cells: notochordal and chondrocyte type cells. Notochordal cells play an important role in synthesizing proteoglycans and other ECM molecules¹¹. They originate in the embryo, continue through fetal life, and have been found to substantially decrease during ageing. The age of disappearance differs greatly from species to species¹², in humans this is believed to occur around age 10¹³. Interestingly, most molecular studies focus solely on adults, leaving fetal discs, which have abundant notochordal cells relatively undiscovered. Since fetal discs undergo rapid growth, their ECM proteome contain many proteins involved in anabolic pathways. Identification of these proteins could give rise to new insights in the pathophysiology and regenerative treatment strategies which revolve around the restoration of ECM in both quantity and quality.

Unfortunately, very little is known regarding the ECM of the human fetal NP. The studies that have been conducted in this field, focused on developmental structures and changes of specific ECM proteins^{14–17}. In addition, Caldeira et al., (2017) have compared the proteome between bovine- fetal and adult discs¹⁸. Until now, no studies have been conducted that performed a qualitative and quantitative exploration of the human fetal matrisome (ECM proteins).

Hence, the aim of this explorative study is to identify proteins of interest for future therapies through profiling the core matrisomal proteome of human fetal NP's, and compare it with the core matrisome of healthy adults.

Materials and methods

Patient population. The main focus of this study is to compare two groups: Fetal and healthy adult NP's. Fetal NP tissue was harvested from 9 miscarriages from all lumbar and thoracic spinal segments. Healthy adult NP tissues were harvested from brain dead organ donor volunteers, of which 2 discs were harvested from L3 to L4 and 5 discs from L4 to L5. Moreover, 11 degenerated discs were included for an additional analysis, 1 originated from the level L2L3, 1 from L3L4, 7 from L4L5 and 2 from L5S1. Pfirrmann scores of all included NP's are listed in Supplementary Table S1. Institutional Ethical Committee (Ganga hospital Ganga Medical Centre and Hospital Pvt Ltd Tamil Nadu) and Institutional Review Board approval were obtained to conduct research on these specimens. Informed consent was obtained for each included sample and all methods were carried out according to the Declaration of Helsinki.

Sample collection. Fetal specimens were gathered following spontaneous as well as medical abortions (24 weeks of gestational age). The spine segment was harvested through a posterior approach within 12 h of death under sterile operating conditions after obtaining an informed consent from the parents of the deceased fetus. Nine Fetal samples with normal MRI and CT were included and those with congenital anomalies were excluded from the study. Intervertebral disc tissues were dissected and subsequently separated into nucleus pulposus and annulus fibrosus using a surgical microscope.

Regarding the healthy adult NP tissues, our institution has been licensed by the government for harvesting bone through a voluntary organ donor program. The spine was procured as 2/3 motion segments after retrieval of organs such as kidneys, heart and liver for transplant surgeries. MRI was obtained and NP's were harvested after segregating bone and annulus fibrosus. Both Fetus and adult NP's were transferred to sterile cryopreservation vials, and snap frozen in liquid nitrogen before transport to the research laboratory.

Sample processing. Around 200 mg tissue from the 7 adult, 9 fetus and 11 degenerated discs was subjected for extraction of total proteins and tryptic digested peptides were subjected to ESI-LC-MS/MS with conditions as described in our earlier report: (Rajasekaran et al.¹⁹).

Bioinformatics analysis. MS/MS raw data acquired from Orbitrap Velos Pro Mass Spectrometer were analysed by Proteome Discoverer v1.4 using Mascot (Matrix Science, London, UK; version 2.4.1.0) and inbuilt SequestHT search algorithm. The peptide spectrum matches (PSMs) from SequestHT and Mascot were post-processed using the Percolator algorithm. The peptides with rank one and having a q-value < 0.01 were considered for protein identification.

Relative quantification by spectral count. Spectral counts obtained by LC/MS-MS were further normalization by normalized spectral abundance factor (NSAF) method, as described by Zybaylov et al.²⁰. NSAF is a relative quantification method based on protein length and spectral count. 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.

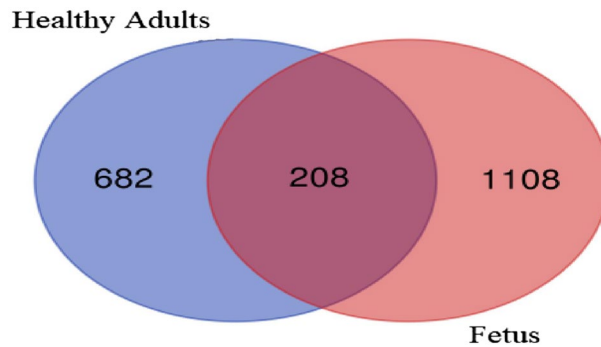


Figure 1. Venn diagram representing total number of identified proteins expressed in fetal and healthy adult NP's: Blue displays the unique number for healthy adults, Red the unique number for fetus, and Purple the number of share proteins.

Quantitative analysis. Out of the proteomic database, the core matrisome (ECM proteoglycans, glycoproteins and collagens) was selected for further analysis²¹. All selected proteins with > 2 unique peptide or 1 unique peptide with a PSM \geq 10 were included in the analysis²².

Statistical analysis. Data was analyzed using SPSS software version 25. Since the sample size was limited, differences in protein expression between groups were assessed using Mann–Whitney *U* tests. Groups were only compared if more than two samples expressed the respective protein in each group. Two-tailed alpha level was set at 0.05. Samples with missing values were excluded from the analysis.

Pathway analysis. In order to identify fetus specific proteins of interest for future regenerative treatment, a pathway analysis of all fetal specific proteins was performed using both STRING v10.5 and DAVID v.6.8 databases. Proteins were considered 'fetus specific' if they met one of the following three conditions: Firstly, they were uniquely described in fetal NP's in at least 2 fetal samples. Secondly, they were significantly upregulated in fetal NP's. Thirdly, in the case that only one sample was identified in the healthy adult group and no statistical analysis could be performed, an upregulation in fetal expression of at least a twofold.

Interaction analysis. In order to find out whether the identified pathways interact and share common initiators, a protein–protein interaction analysis was performed using Cytoscape v3.7²³.

Verifying in degenerated samples. In order to assess whether the protein that decrease from fetus to healthy adult NP's further decrease in severely degenerated discs, expression levels of all proteins of interest that were identified in regenerative pathways were used for an additional analysis. In this analysis, protein expression in 11 degenerated discs was compared to expression levels in fetus and healthy adult NP's.

Validation by immunohistochemistry. Immunohistochemistry (IHC) was used to validate proteomic data: expression differences between the groups were assessed for three proteins: Collagen 14A1, Fibromodulin and Biglycan. For Collagen 14A1 and Biglycan 4 fetus NP's (from one fetal spine) and 4 adult NP's from different organ donors were used. For Fibromodulin, two samples were used per group (all from different donors). Samples were fixed in formalin solution and subsequently embedded in paraffin. 5 μ m section from the paraffin blocks were used for IHC using a tree-step indirect method. For antigen retrieval, samples were heated in retrieval buffer until 15 lb per square inch pressure was reached, which was subsequently maintained for 2 min. Slices were rinsed in ethanol solutions and incubated in 0.1% hydrogen peroxide for blocking. Monoclonal antibodies for Fibromodulin (Santa Cruz Biotechnology, Inc USA) and Collagen14a1 (Santa Cruz Biotechnology, Inc USA), Biglycan (Santa Cruz Biotechnology, Inc USA), were used. Primary anti-bodies were incubated for 1 h and secondary antibodies (1:100 dilution) were developed with DAB and counterstained with Harris hematoxylin. Images were captured using Leica Application Suite Vs 4.5.0.418 software for a Leica DML light microscope. Evaluation of the slides was performed by an experienced pathologist. Slides were scored as negative, or as a mild/moderate/strong positive.

Results

Descriptive analysis of ECM. *Gene ontology.* In fetal NP's, a total of 1316 different proteins were identified, of which 1108 were unique to fetus and 208 were shared with adult NP's. In adult NP's, a total of 889 different proteins were identified, of which 682 unique to adult NP's (Fig. 1).

The quantitative analysis revealed that 48% of the fetal NP proteome consisted of ECM proteins (core matrisome + regulators + ECM affiliated + secreted factors), the remaining 52% was characterized as cellular. For healthy adults, the percentage of ECM proteins was 64%.

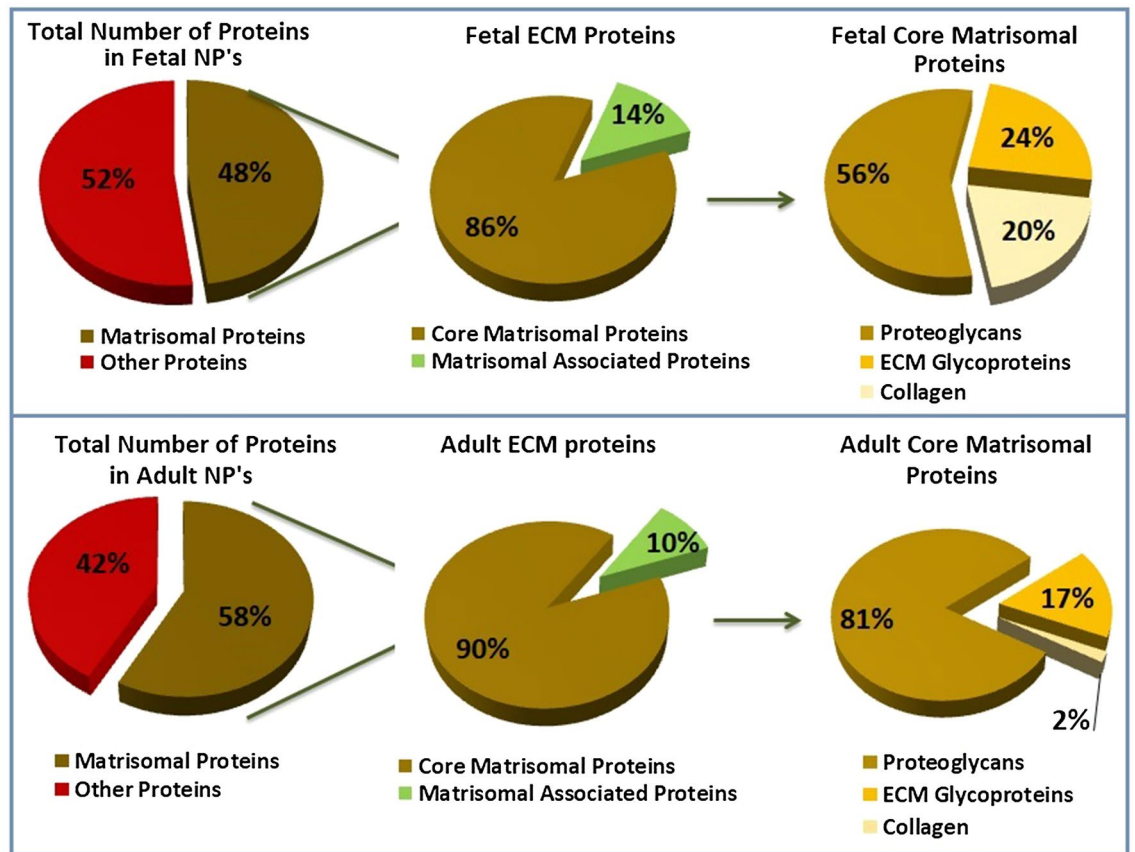


Figure 2. Relative protein composition fetus and healthy adult NP's: the pie chart exhibiting percentages of identified proteins distributed by matrisomal and other proteins.

The matrisome consisted of 86% of core matrisomal proteins in fetal NP's, similar results were seen in adult discs (90%). The core matrisome was subdivided in proteoglycans (56% in fetus and 82% in adult), glycoproteins (24% in fetus and 17% in adult) and collagens (20% in fetus and 2% in adult).

The composition of the discs is illustrated in Fig. 2.

Comparing fetal and adult core matrisome. After applying the exclusion criteria, a total of 64 core matrisomal proteins were subjected for further analysis. Out Of these 64, 25 proteins were specific to fetus, of which 2 proteoglycans, 15 glycoproteins, and 8 collagen types (Fig. 3, Supplementary Table 2). In adult NP's only 6 proteins were specific and belonged to glycoproteins (Supplementary Table 2).

Furthermore, 22 out of the 64 proteins were expressed in > 2 samples in each group and were selected for statistical analysis. The analysis for proteoglycans revealed a significantly lower expression of aggrecan ($p=0.001$), chondroadherin ($p=0.001$), lumican ($p=0.001$), osteoglycin ($p=0.004$), 'proline and arginine rich end leucine rich repeat protein' (PRELP) ($p=0.004$) and versican ($p=0.001$), and significantly higher expression of biglycan ($p=0.05$) and fibromodulin ($p=0.028$) in fetus (Fig. 4).

Regarding the glycoproteins, significant downregulations in fetal NP's as compared to adult NP's were found for 'ABI family, member 3 (NESH) binding protein' (ABI3BP) ($p=0.004$), 'cartilage oligomeric matrix protein' (COMP) ($p=0.001$) and fibronectin (FN1) ($p=0.002$). In contrast, 'procollagen C-endopeptidase enhancer (PCOLCE)' ($p=0.02$) and 'transforming growth factor-beta-induced protein ig-h3 (TGFB1)' ($p=0.003$) were both significantly upregulated in fetal NP's (Fig. 5).

At last, comparing fetal and adult collagen expression showed that collagen types 11a2 (COL11a2) ($p=0.012$), COL6a1 ($p=0.05$) and COL6a3 ($p<0.001$) were significantly upregulated in fetal NP's (Fig. 6).

Moreover, the remaining 11 proteins were present in both groups but were not eligible for statistical analysis (< 3 samples in one of the groups). Of those proteins, the proteoglycan 'Heparan sulfate proteoglycan 2' (HSPG2), the glycoproteins 'EGF-like repeats and discoidin I-like domains 3 (EDIL3)' and thrombospondin 1 (THBS1), and collagen type 15A1 (COL15A1) were upregulated > twofold in fetus samples. Hence these proteins were also integrated in the pathway analysis.

Identifying fetal proteins of interest for regenerative potential. Based on the previously described results, a total of 5 proteoglycans, 18 glycoproteins and 10 collagen types were selected (Table 1). STRING database provided a total of 56 pathways in which these proteins were involved. Out of these pathways, 7 were involved in synthesis of the extracellular matrix and 2 in growth related pathways (Table 2). Moreover the 33

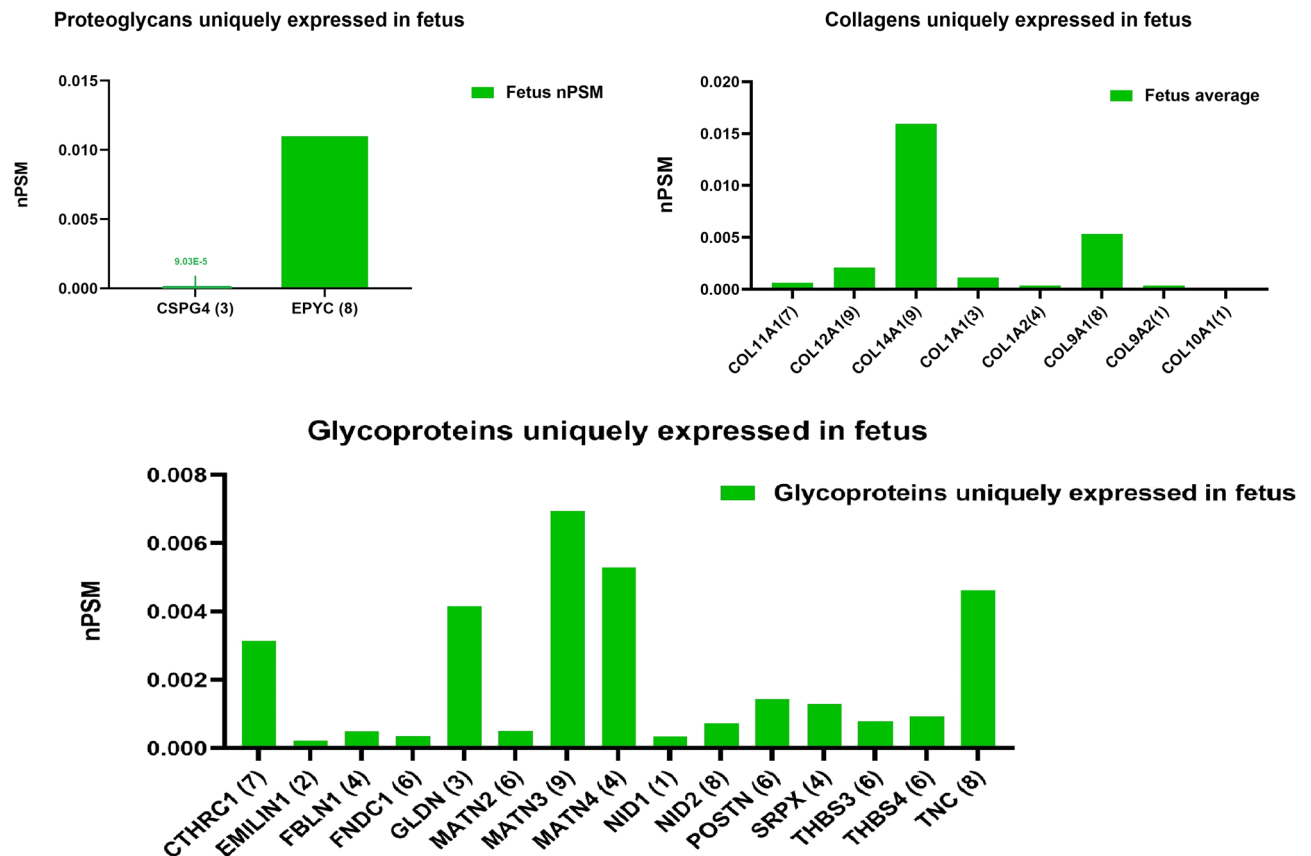


Figure 3. Proteins that were uniquely expressed in fetal NP's: the Y axis displays median normalized PSM's per group, X axis shows protein gene symbols accompanied by (number of samples). The upper left graph displays proteoglycans, the upper right collagens and the lower graph glycoproteins.

selected proteins were involved in 90 biological pathways, out of which 4 were growth related. All selected collagen types, biglycan, fibromodulin, HSPG2, 'PCOLCE' and 'Chondroitin sulfate proteoglycan 4 (CSPG4)' were involved in ECM assembly pathways. Regarding the growth-related pathways, collagen types 1A1, 1A2, 6A1, 6A1, 9A1, 11A1, and 11A2, thrombospondin 1, 3 and 4, matrilin 3 and tenascin C were identified (Table 2). Concerning the biological processes, collagen 1A1, 6A1, 11A1, 11A2 and 12A1, Collagen triple helix repeat containing 1 (CTHRC1), periostin and TGFBI, thrombospondin 3 and tenascin C were involved in 'tissue development'. Additionally, collagen 1A1, 1A2, fibromodulin, periostin and thrombospondin 1 are involved in 'cellular response to growth factor stimulus'. The last relevant pathway was 'regeneration', in which matrilin-2, periostin and tenascin C were present.

Interaction analysis. Protein–protein interaction analysis by Cytoscape revealed that biglycan has outgoing interactions to proteins in all the relevant ECM assembly pathways, and the P13K-Akt-pathway (Fig. S1). In addition, biglycan received incoming interactions from fibromodulin, TGFBI and HSPG1. Most importantly, it also received input from thrombospondin 4, matrilin-2, matrilin-3 and periostin. The interactions between these proteins connect the IGF-signaling pathway with some of tissue remodeling, cellular response to growth factors and regeneration processes through biglycan signaling (Fig. S2).

Verification in degenerated samples. The 24 proteins selected in the relevant regenerative pathways were submitted to an additional analyses, in which the expression levels in adult NP was compared to degenerated NP's. 14 out of the 24 proteins were uniquely expressed in fetal discs and absent in adult discs, in line with the decrease from fetus to adult, 5 proteins were also absent in degenerated discs: Collagen 9a1, CSPG4, Matrilin-3, thrombospondin 3 and CTHRC1. In addition, Matrilin-2 was expressed in only 1- and Collagen 11a1 in 2 degenerated samples. By contrast, 7 proteins that were absent in healthy adult discs were present in degenerated discs: Collagen 1a1, 1a2, 12A1, 14A1, tenascin, thrombospondin 4 and periostin (Fig. 7).

Of the 10 proteins that were expressed in both fetus and healthy adult NP's, Collagen 11A2 and Fibromodulin showed a significantly lower expression in degenerated compared to healthy adult NP's, which was in line with the decrease from fetus to healthy adult. In contrast, collagen 6a3 was upregulated in degenerated compared to healthy adult NP's. The other 8 proteins showed no significant differences (Fig. 8). In short, of the 24 relevant proteins that decreased in healthy adults compared to fetus, 8 proteins increased again upon degeneration.

Differences in proteoglycan expression between fetus and healthy adult NP's

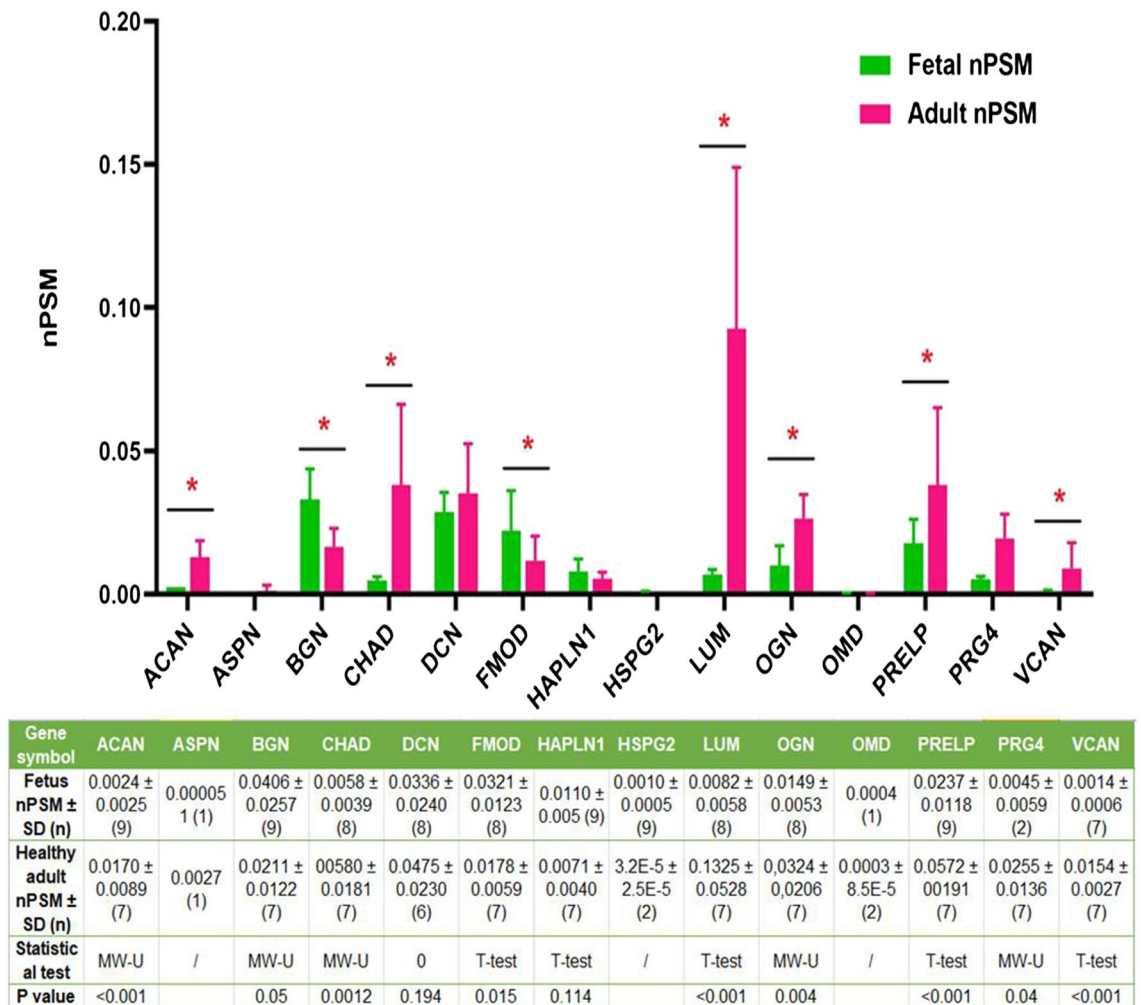


Figure 4. Differences in proteoglycan expression between fetus and healthy adults: The Y axis displays median normalized PSM's per group, X axis shows protein gene symbols, *indicates statistical significance and error bars are Interquartile ranges, which were only displayed if $n > 2$. The attached table contains the median (number of samples) for each group, and provides the according Mann Whitney U p-values if statistics could be performed ($n > 2$). '/' indicates that only 1 or 2 samples were present in one of the groups and no statistical test could be performed.

Validation by immunohistochemistry. 4 Fetus samples and 4 adult control discs underwent immunohistochemistry for validation. Protein expression of COL14a1, FMOD, and biglycan was validated by IHC. COL14A1 was present in all fetus NP's with mild or moderate intensity, and present in only one adult NP with mild intensity (Fig. 9A). Both fetus NP samples showed a strong positive FMOD staining, whereas in adult NP's only 1 sample was moderately positive whereas the other was negative. Together, these results validate the proteomic findings (Fig. 9B). Biglycan was strongly positive in 3 out of the 4 fetus samples, and mild to strongly positive in 4 adult samples. Thereby verifying their abundant presence in both, without any clear differences in quantity. The staining pattern in adults was completely diffuse, whereas in fetus it was centered around the cells (Fig. 9C).

Discussion

The fetal NP proteome is highly different from that of a healthy adult. Using a proteomic approach combined with a bioinformatics pathway analysis, this study identified significant upregulations or unique expressions in fetal NP's of 33 core matrisomal proteins, of which 24 involved ECM assembly or growth pathways. Of these 24, 2 proteins further decreased in degenerated discs, whereas 8 increased. Moreover, this study is the first to describe the core matrisomal proteome of human fetal NP's.

Our findings are in line with those of Smith et al.¹⁷, who performed an immunohistochemical analysis of multiple proteoglycan and collagen types in the human fetal disc. They also found a low expression of aggrecan,

Differences Glycoprotein expression between fetus and healthy adult NP's

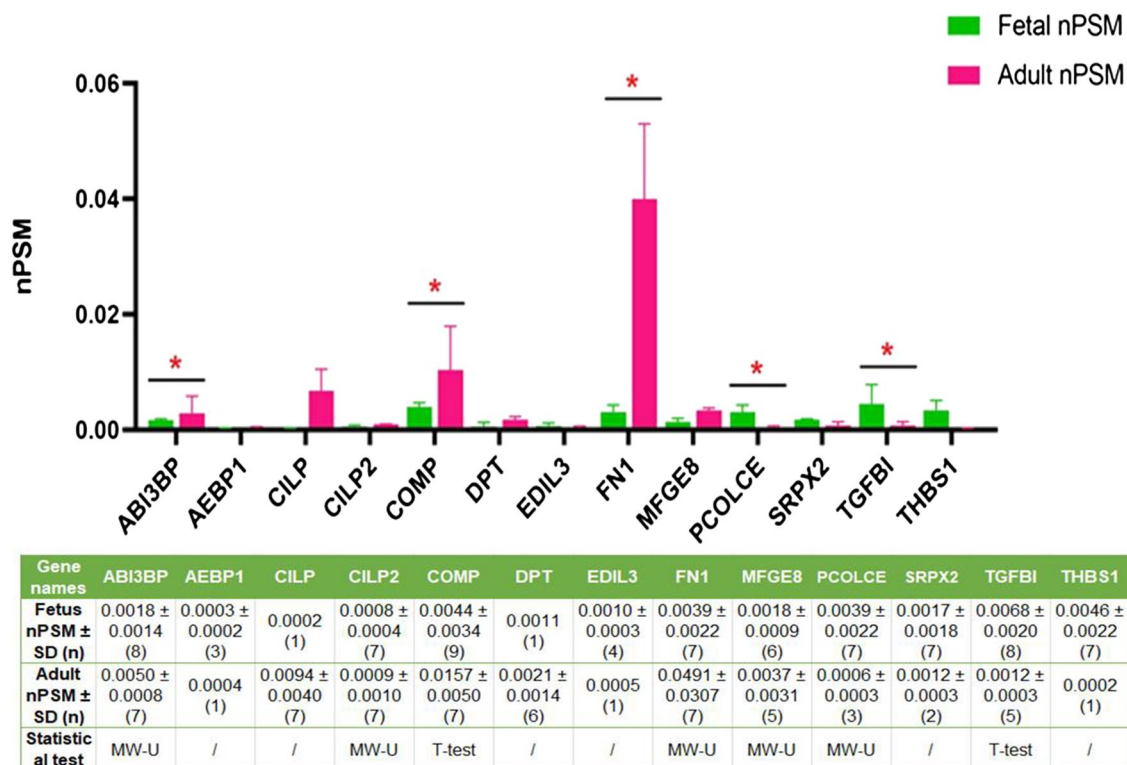


Figure 5. Differences in glycoprotein expression between fetus and healthy adults: The Y axis displays median normalized PSM's per group, X axis shows protein gene symbols, error bars are Interquartile ranges and * indicates statistical significance. The attached table contains the median (number of samples) for each group, and provides the according Mann Whitney U p-values if statistics could be performed ($n > 2$). '/' indicates that only 1 or 2 samples were present in one of the groups and no statistical test could be performed.

versican and perlecan (a synonym for HSPG2) in the intervertebral disc. In addition, they reported on the presence of both collagen type 1 and 2 in the nucleus pulposus, which was confirmed by our results. Collagen type 1 is known for its presence in AF but not in NP tissue. Their presence in fetus NP's could indicate that the distinguishment between AF and NP regions is not yet fully developed at this stage.

When comparing our results with proteomic data from bovine discs as described by Caldeira et al.¹⁸, we found the overall distribution of the matrisome to be quite comparable. When comparing fetus and healthy adult for significant differences, similar results in terms of expression were found for chondroadherin, lumican, PRELP, cartilage oligomeric matrix protein and Fibronectin. However, some results were different; in bovine adult discs, the expression of fibromodulin and 'Hyaluronan and proteoglycan link protein 1' was found higher as compared to fetal discs, which contradicted the differences in expression levels of human discs. Nevertheless, it should be noted that the higher expression in human fetus was only significant for fibromodulin. In order to verify whether these differences can be attributed to interspecies variation or to unsystematic error due to small sample sizes, more proteomic studies on fetal discs are needed. In addition, Bach et al. performed a qualitative proteomic study using notochordal cell conditioned medium from humans, canine and porcine²⁴. Their results were comparable with expression profile of fetus NP's: apart from heparan sulfate proteoglycan, periostin and Collagen triple helix repeat containing 1, all identified proteins of interest were expressed in their medium.

Differences between healthy and degenerated adult discs. The additional analysis with degenerated discs revealed that of the 24 identified relevant proteins, only fibromodulin and collagen11a2 further decreased in degenerated discs compared to healthy adults. This places additional interest on these 2 proteins as potential targets for regenerative therapy. By contrast, even though in low concentrations and in mostly limited amount of samples, 9 proteins that were absent in adult discs were re-expressed in the degenerated state and one showed significantly increased expression compared to healthy adult discs (Collagen 1a1, 1a2, 6a3, 11a1, 12A1, 14A1, Matrilin-2, tenascin, thrombospondin 4 and periostin). This could mean that these proteins are involved in inducing the degenerated state, for example through functioning as an inhibitor or regulator in the regenerative processes identified. By contrast, these proteins could also be upregulated as a compensation mechanism. If this is the case, further increasing their concentrations may aid regeneration and slow further decay. In order to distinguish between harmful and compensatory, mechanical experiments are required, which is beyond the scope of this study. However, in order to identify the best candidates for future mechanical experiments, we will elaborate on the specific functions on all 24 identified proteins below.

Differences in collagen expression between fetus and healthy adult NP's

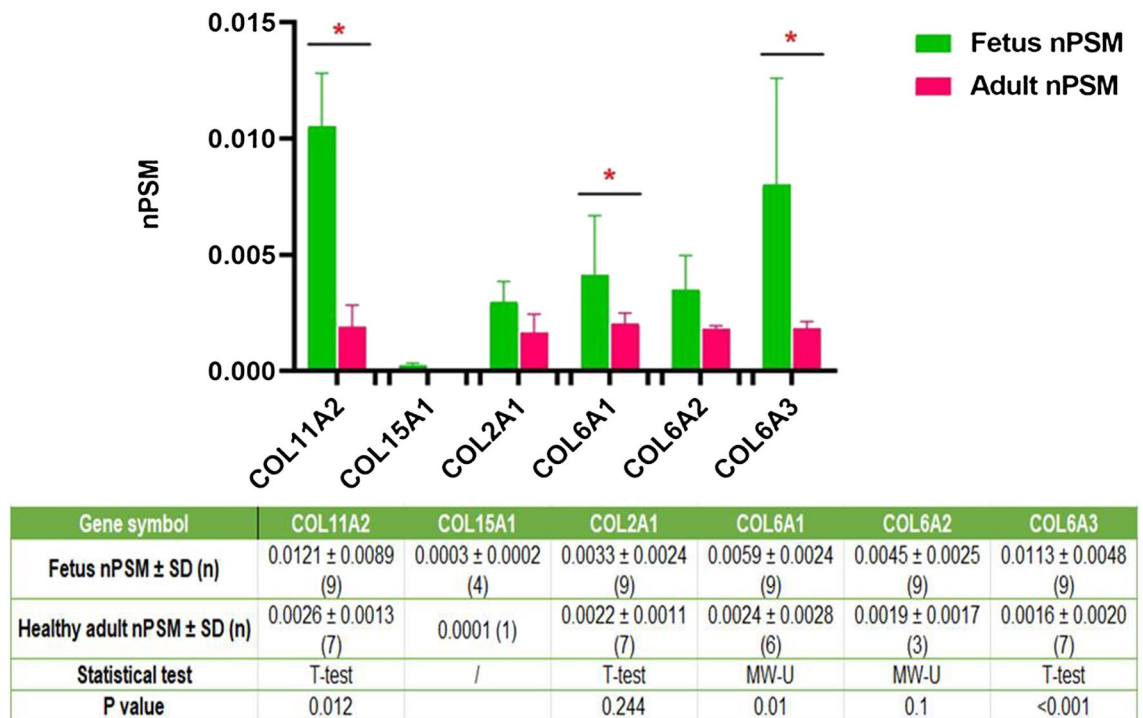


Figure 6. Differences in collagen expression between fetus and healthy adults: The Y axis displays median normalized PSM's per group, X axis shows protein gene symbols, error bars are Interquartile ranges and * indicates statistical significance. The attached table contains the median (number of samples) for each group, and provides the according Mann Whitney U p-values if statistics could be performed ($n > 2$). '/' indicates that only 1 or 2 samples were present in one of the groups and no statistical test could be performed.

ECM assembly pathways: proteoglycans. Regarding the proteoglycans, only Fibromodulin, CSPG4, HSPG2 and biglycan showed up in relevant ECM growth related pathways. All four were involved in 'glycosaminoglycan metabolism', which support fibrils and control hydrostatic pressure²⁵. CSPG4 is a glycosaminoglycan with anti-inflammatory, anti-catabolic to anti-apoptotic potential, which is currently used as a treatment for osteoarthritis of the knee joint²⁶. Furthermore, recent literature revealed that biglycan and fibromodulin are involved in both the assembly and regulation of collagen fibrillogenesis⁹. Interestingly, biglycan was also identified as an initiator for multiple pathways in the interaction analysis. It is a cell-associated class- I SLRPs and is secreted as probiglycan which is cleaved by the BMP-1 to form mature biglycan²⁷. Secreted biglycan interacts with collagen type I, II, III, and VI and elastin and becomes sequestered in the ECM of most of the organs²⁸. In the developing human fetal discs, biglycan is widely distributed and its presence separates the interspace of the developing intervertebral disc from that of the cartilage end plate. Moreover, biglycan also has various interactions with growth factors and receptors, and through these interactions it acts as a signaling molecule that regulates cell growth and morphogenesis²⁹.

Even though these proteins display regenerative potential, it should be noted that the p-values of the differences in expressions of HSPG2, biglycan and Fibromodulin were > 0.01 , and that this study has performed multiple tests without corrections for multiple testing, hence the findings should be interpreted with caution. Similar caution in interpretation is required regarding the uniquely expressed CSPG4, since it was only present in 3 out of the 8 samples. Nevertheless, the difference in expression of FMOD was validated by IHC. Besides, based on their functions, all three seem to be potential candidates for regenerative therapies. Interestingly, the IHC data revealed that in fetus NP's fibromodulin was solely located in the nuclear membrane, whereas in adult NP's it was spread over the cell cytoplasm and interstitium. This illustrates that at 24 weeks; FMOD is not yet part of the extracellular matrix and still has to be excreted to fulfill its role in collagen fibrillogenesis.

ECM assembly pathways: glycoproteins. Out of the 18 relevant glycoproteins, PCOLCE was involved in 4 relevant ECM growth related pathways: 'Assembly of collagen fibrils and other multimeric structures', 'Collagen biosynthesis and modifying enzymes' and 'Collagen formation' and 'Crosslinking of collagen fibrils'. This enzyme is necessary for the conversion of procollagen to collagen (type I-III) and is thereby crucial in the creating of a fibrillary matrix³⁰.

ECM assembly pathways: collagens. Interestingly, all upregulated or unique fetus collagen types were involved in ECM assembly pathways: Assembly, anchoring and crosslinking of collagen fibrils, Collagen biosyn-

Protein name	Gene symbol	Cluster
Collagen triple helix repeat containing 1	CTHRC1	Glycoprotein
EGF-like repeats and discoidin I-like domains 3	EDIL3	Glycoprotein
Elastin microfibril interfacier 1	EMILIN1	Glycoprotein
Fibulins	FBLN1	Glycoprotein
Fibronectin type III domain containing 1	FNDC1	Glycoprotein
Gliomedin	GLDN	Glycoprotein
Isoform 3 of matrilin-2	MATN2	Glycoprotein
Isoform 2 of matrilin-3	MATN3	Glycoprotein
Matrilin-4	MATN4	Glycoprotein
Nidogen 2	NID2	Glycoprotein
Procollagen C-endopeptidase enhancer	PCOLCE	Glycoprotein
Periostin, osteoblast specific factor	POSTN	Glycoprotein
Sushi-repeat-containing protein, X-linked	SRPX	Glycoprotein
Transforming growth factor-beta-induced protein ig-h3	TGF-BI	Glycoprotein
Thrombospondin 1	THBS1	Glycoprotein
Thrombospondin 3	THBS3	Glycoprotein
Thrombospondin 4	THBS4	Glycoprotein
Tenascins	TNC	Glycoprotein
Biglycan	BGN	Proteoglycan
Chondroitin sulfate proteoglycan 4	CSPG4	Proteoglycan
Epiphycan	EPYC	Proteoglycan
Fibromodulin	FMOD	Proteoglycan
Heparan sulfate proteoglycan	HSPG2	Proteoglycan
Collagen, type I, alpha 1	COL1A1	Collagen
Collagen, type I, alpha 2	COL1A2	Collagen
Collagen, type VI, alpha 1	COL6A1	Collagen
Collagen, type VI, alpha 3	COL6A3	Collagen
Collagen, type IX, alpha 1	COL9A1	Collagen
Collagen, type XI, alpha 1	COL11A1	Collagen
Collagen, type XI, alpha 2	COL11A2	Collagen
Collagen, type XII, alpha 1	COL12A1	Collagen
Collagen, type XIV, alpha 1	COL14A1	Collagen
Collagen, type XV, alpha 1	COL15A1	Collagen

Table 1. Protein list of interest. Table lists the proteins of interest (uniquely expressed in fetus in at least 2 samples, significantly upregulated in fetus or upregulated > twofold if only 1 sample expressed the respective protein in the healthy adult group).

thesis and formation. Out of the 9 collagen types, collagen- 1A1, 1A2, 6A1, 6A3, 9A1, 11A1 and 11A2 were all characterized as structural types that were part of the matrix⁹.

In contrast, collagen 12A1 and 14A1 are not part of the major structure. Collagen type 12A1 is thought to play a role in regeneration of tail and spinal cord in *Notophthalmus viridescens*³¹, and is involved in the regulation and organization of collagen fibril bundles³². Collagen type 14A1 has a regulatory role in early stage fibrillogenesis³³, and plays a key role in regeneration in other tissues³⁴. In addition, both Collagen 12a1 and 14a1, are involved in hydration and thickness, at least in other tissues³⁵. At last, Collagen 15A1 has been characterized as a basement membrane protein that functions as a chondroitin sulfate proteoglycan³⁶. This type of protein has multiple beneficial functions as described in the paragraphs on proteoglycans. Nevertheless, the evidence on all three collagen types is limited and more research on the functions of collagen 12A1, 14A1 and 15A1 in intervertebral discs is needed.

Growth pathways. No proteoglycans were involved in any of the growth pathways. Out of the 5 selected glycoproteins, Thrombospondin 1, 2, 3, and 4 and Tenascin C were all involved in the ‘PI3K-Akt signaling pathway’. This pathway is involved in protein synthesis, inhibition of apoptosis and cell proliferation³⁷, these three have shown to play major roles in disc degeneration⁸. Thrombospondins have been identified as a downstream effector protein in this pathway that regulates angiogenesis^{38,39}. Interestingly, recent findings have illustrated that maintaining an anaerobe environment is crucial for the functioning of NP cells⁴⁰, thereby highlighting the importance of regulating angiogenesis. In addition, both Tenascin C and thrombospondins activate TGF- β , which induces the PI3K-Akt signaling^{39,41}, and thus makes their specific roles in the nucleus pulposus worth exploring.

	Proteins involved	P-value	Gene symbol
ECM assembly pathways			
Assembly of collagen fibrils and other multimeric structures	11	3.72e-17	COL11A1, COL11A2, COL12A1, COL14A1, COL15A1, COL1A1, COL1A2, COL6A1, COL6A3, COL9A1, PCOLCE
Collagen biosynthesis and modifying enzymes	11	7.47e-17	COL11A1, COL11A2, COL12A1, COL14A1, COL15A1, COL1A1, COL1A2, COL6A1, COL6A3, COL9A1, PCOLCE
Crosslinking of collagen fibrils	3	5.19e-05	COL1A1, COL1A2, PCOLCE
Glycosaminoglycan metabolism	4	0.00043	BGN, CSPG4, FMOD, HSPG2
Anchoring fibril formation	2	0.0015	COL1A1, COL1A2
Chondroitin sulfate biosynthesis	2	0.0021	BGN, CSPG4
Growth pathways			
PI3K-Akt signalling pathway	9	1.89e-08	COL1A1, COL1A2, COL6A1, COL6A3, COL9A1, THBS1, THBS3, THBS4, TNC
Regulation of Insulin-like Growth Factor (IGF) transport and uptake by Insulin-like Growth Factor Binding Proteins (IGFBPs)	2	0.0046	MATN3, TNC
Biological processes			
Tissue development	10	0.0043	COL11A1, COL11A2, COL12A1, COL1A1, COL6A1, CTHRC1, POSTN, TGFB1, THBS3, TNC
Cellular response to growth factor stimulus	5	0.0128	COL1A1, COL1A2, FMOD, POSTN, THBS1
Regeneration	3	0.0197	MATN2, POSTN, TNC

Table 2. Pathway analysis. Table displays the results of the STRING and DAVID pathway analysis. The first (left) column shows the identified pathways and biological processes that are involved in ECM assembly, the second column shows the number of proteins of interest involved in that pathway/process, the third list the p-value of the upregulation and the fourth displays the gene symbols of the proteins involved. In total, 24 proteins were identified in 8 pathways and 3 biological processes.

Differences in protein expression between fetus and degenerated NP's: 14 proteins from relevant pathways that were absent in healthy adult NP's

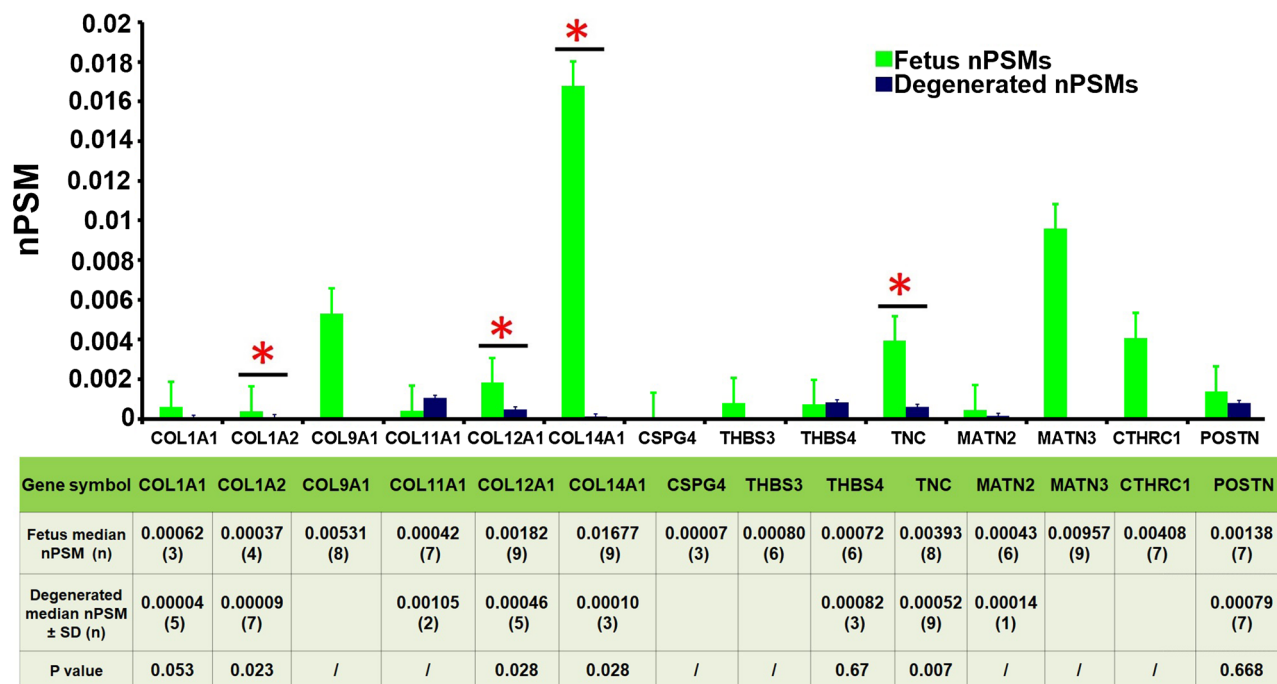


Figure 7. Differences in protein expression between fetus and degenerated NP's: 10 proteins from relevant pathways that were present in healthy adult NP's. The Y axis displays median normalized PSM's per group, X axis shows protein gene symbols, error bars are Interquartile ranges and *indicates statistical significance. The attached table contains the median (number of samples) for each group, and provides the according Mann Whitney U p-values if statistics could be performed (n > 2). '/' indicates that only 1 or 2 samples were present in one of the groups and no statistical test could be performed.

Differences in protein expression between adult and degenerated NP's: 10 proteins from relevant pathways that were present in healthy adult NP's

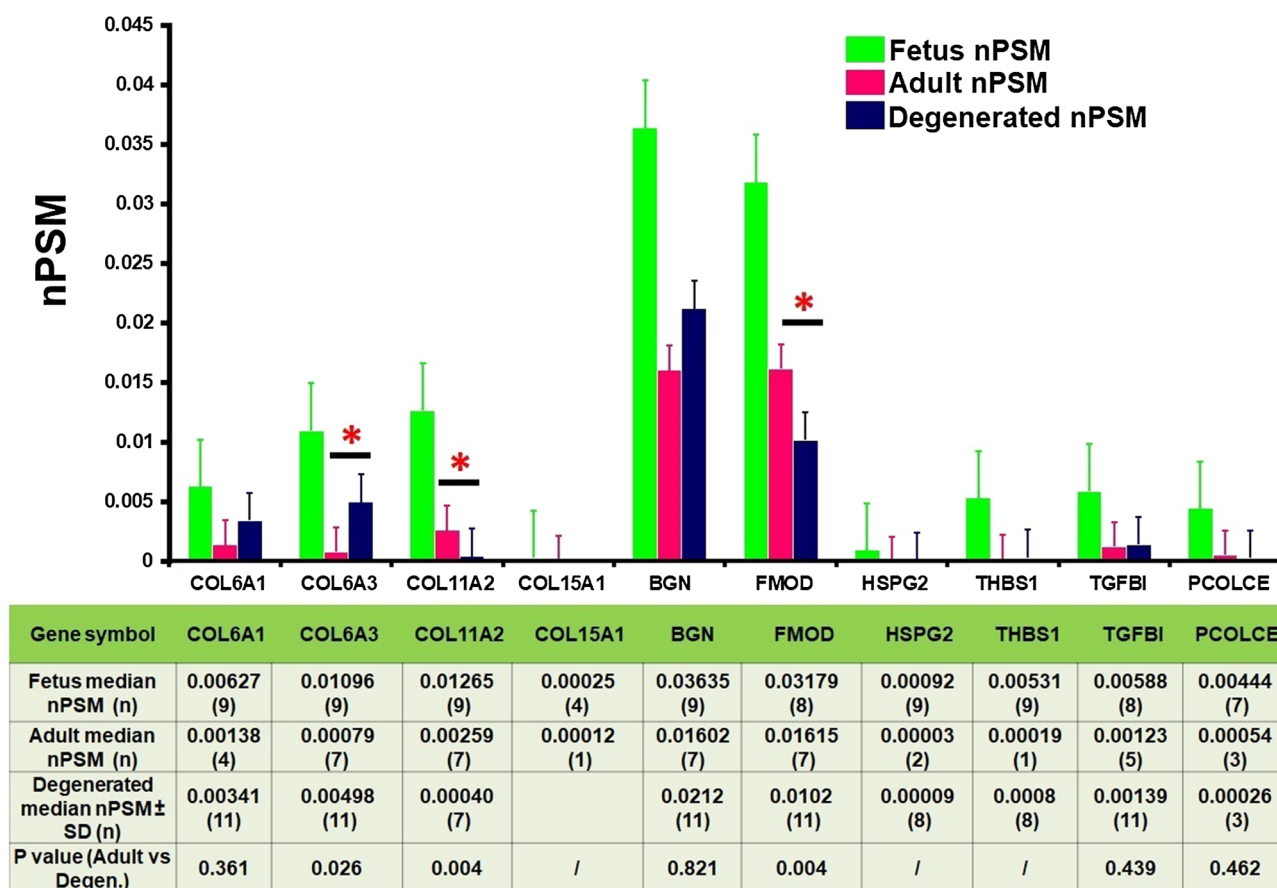


Figure 8. Differences in protein expression between healthy adult and degenerated NP's: 10 proteins from relevant pathways that were present in healthy adult NP's. The Y axis displays median normalized PSM's per group, X axis shows protein gene symbols, error bars are Interquartile ranges and *indicates statistical significance between healthy and degenerated Adults. The attached table contains the median (number of samples) for each group, and provides the according Mann Whitney U p-values if statistics could be performed ($n > 2$). '/' indicates that only 1 or 2 samples were present in one of the groups and no statistical test could be performed.

Moreover, both tenascin C and Matrilin-3 were involved in the regulation of (IGF). This growth factor has shown to stimulate IVD cell proliferation and ECM synthesis⁴². Its expression rate is carefully regulated by Matrilin-3, thereby preventing overexpression⁴³. At last, all collagen types characterized in the growth pathways were identified as structural collagens, and were targets in the pathways instead of initiators or regulators.

Biological processes. With ten proteins involved, the most significantly upregulated biological process was 'tissue development'. This process featured some proteins, which tissue development properties have been discussed earlier under the ECM assembly pathways: Collagen 1A1, 6A1, 11A1, 11A2 and 12A1. The remaining five proteins were involved in the development of a variety of tissues, some more relevant than others. Firstly, Tenascin C was involved in wound healing, during which it stimulates fibroblasts to increase the ECM protein synthesis⁴⁴, thereby making it an attractive target for regenerative treatment. Interestingly, Thrombospondin 3 has proven to be important for osteogenesis, but its development functions in the NP remain unknown⁴⁵. Moreover, TGFBI is involved in lumen formation during angiogenesis⁴⁶. Because the nucleus pulposus homeostasis is maintained best in an anaerobic environment⁴⁰, TGFBI might not be a preferential choice for regenerative therapies. The fourth protein involved was CTHRC1, which positively regulates bone formation⁴⁷. Nonetheless, its overexpression is also associated with increased angiogenesis⁴⁸, and a decrease in collagen type 1⁴⁹, thereby questioning its practicability. The last protein engaged in tissue development was periostin, which is a protein participating in osteogenesis and post-injury tissue regeneration processes such as fibrosis and wound healing⁵⁰. During these processes, periostin expression leads to degradation of ECM through upregulation of matrix metalloproteinases⁵⁰, which is reflected by an upregulation of periostin in degenerated discs⁵¹. Taken together,

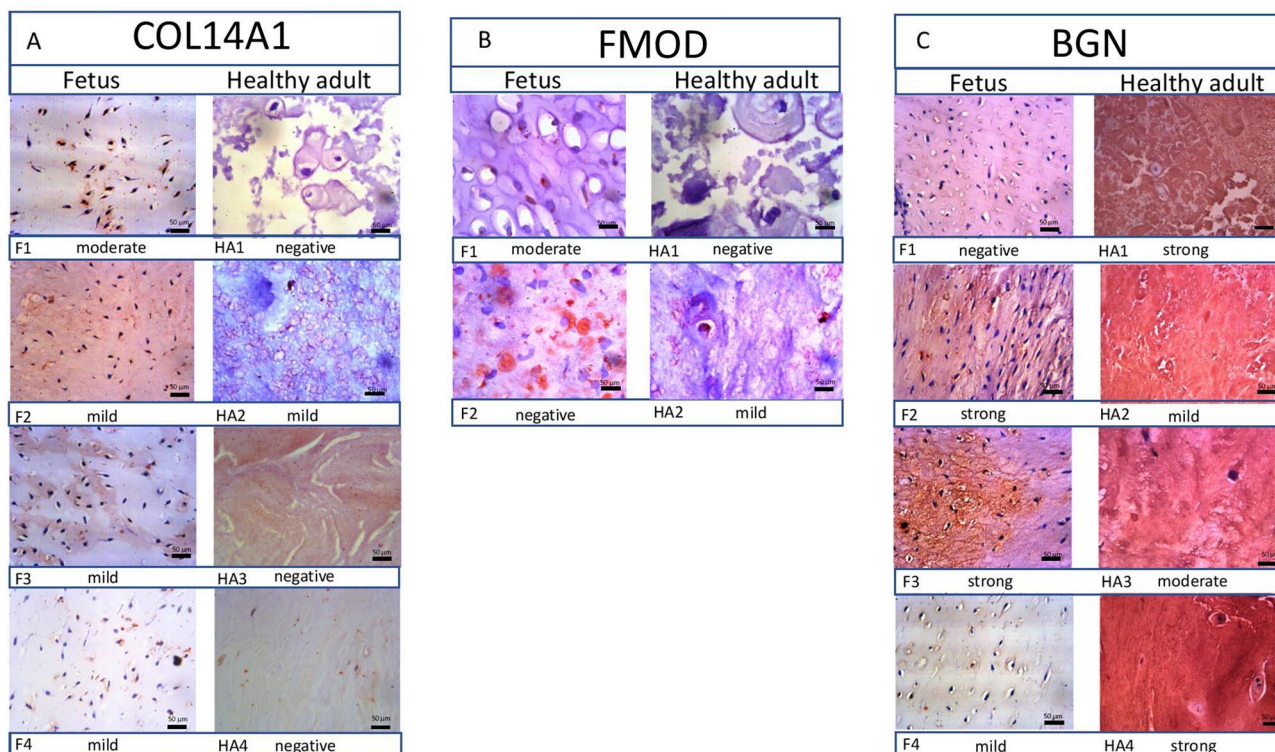


Figure 9. Validation by immunohistochemistry: protein expression of Col14a1, FMOD, and BGN was validated by immunohistochemistry (magnification 400 ×, scale bar = 50 μm). (A) COL14a1 was strongly positive in the interstitium and cell margins of fetus NP's (F), Healthy adult (HA) samples, faint staining was seen in the interstitium of HA2. (B) FMOD was moderately positive in the cell margins and interstitium of F1, but negatively stained in the rest of the NP's. In HA, only sample HA2 was stained mildly positive. (C) BGN was stained positively in all HA and all but one F NP with varying intensity. In HA, the staining was diffuse in the interstitium but not in the NP cells, whereas in F NP's the staining was positive in cytoplasm and interstitium.

periostin seems very unsuitable for a regenerative therapy, and future treatments should focus on inhibiting this protein instead.

The second biological process identified was 'cellular response to growth factor stimulus', which included 5 proteins. As with the pathways, collagen 1A1 and 1A2 were substrates in this process⁵². Of the three remaining proteins, both fibromodulin and thrombospondin reflected their role in this process through binding to transforming growth factor-beta (TGF-β)⁵³, thereby modulating its signal transduction. The third protein was periostin, which was regarded as unsuitable for regenerative therapies as mentioned above, it will not be discussed further.

The last upregulated biological process in fetal specific proteins was 'regeneration', which encountered 3 proteins. Out of which periostin's and tenascin's effects in wound healing have both been discussed above. At last, matrilin-2 it is known to promote muscle, neuron, liver and skeletal regeneration, which is mediated through TGF-β, bone morphogenetic protein 7 and Smad signaling⁵⁴. Matrilins are generally multidomain extracellular adaptor proteins widely involved in the formation of both collagen-dependent & collagen-independent filamentous network⁵⁵. Although functions of matrilin types are still unknown, matrilin-2 is the largest because of its two VWA domains that are connected by ten epidermal growth factor-like (EGF) modules⁵⁶. They play a huge role in remodeling of tissue architecture, and they are widely expressed during differentiation and regeneration process in myogenic, neural and other tissues. It also plays a major role in tumorigenesis⁵⁴.

Matrilin-2 can be detected in almost every other tissue type; they get secreted and deposited. They can be seen in loose and dense connective tissue, skeletal and smooth muscle, skin, subepithelial basement membrane, uterus, peripheral and central nervous system⁵⁵. They function as interacting with other ECM molecules through their VWA domains to assemble into a filamentous network connected to collagen fibrils and proteoglycans⁵⁶.

Study by Zhang et al., 2014 proposes matrilin-2 as a potent biomarker of Osteoarthritis in Articular Cartilage during its early stage⁵⁷. For investigating the distribution patterns of matrilin-2, scientists have raised polyclonal antibodies specific to matrilin-2. With these antibodies they were able to detect two isoforms that are very specific to matrilin-2. Through immunohistochemistry, they were able to detect wide expression of matrilin-2 in loose & dense connective tissue, some specific epithelial cells and much inherently in some cartilage and bone⁵⁷. By using a developmental animal model of Duncan-Hartley guinea pigs, the team also found increased expression of matrilin-2 in a damaged articular cartilage⁵⁷. Role of matrilin-2 as an important ECM protein has been established in mice for proper and timely myogenic differentiation and skeletal muscle regeneration⁵⁸.

Unfortunately, its regenerative potential in disc disease is still unknown, as is the case with most discussed proteins. Nevertheless, Matrilin-2 seems to be a regenerative therapy candidate of high potential.

To summarize the relevance of the biological processes in addition to the pathways: The regenerative potential of tenascin C and fibromodulin was added upon with ECM assembly and interacting with TGF- β respectively. Most importantly, matrilin-2 clearly stood out due to its involvement in regeneration of many other tissues.

Limitations. This study has several strongpoints: It was the first to compare the proteome of the human fetal and adult core matrisome. In addition, this study combined a qualitative, quantitative, statistical and pathway analysis in order to provide an overview of the findings and to put them in perspective in terms of future clinical applications. This study also has some limitations, even though multiple statistical tests were performed, no correction for multiple testing was implemented. This decision was made due to the explorative nature of the study and the low number of samples available. Nevertheless, some of the significant differences should therefore be interpreted with caution. Furthermore, the detection threshold for proteins identification of this study was limited by the sensitivity of mass spectrometer (q -value < 0.01). This limited sensitivity means that potentially interesting proteins with lower expression values may have been overlooked, and that proteins which are identified as unique in fetus could also be expressed in adult NP's in very low quantities. This was illustrated by the validation through IHC, which is considered as a highly sensitive technique: Adult NP 2 showed a moderately positive staining for COL14a1 whereas none of the proteomic samples expressed COL14a1. Nevertheless, as very large differences in protein expression are of similar interest as uniquely expressed proteins, this limitation of proteomics should not be a major concern.

In short, this study provided an overview of the human fetal and healthy adult proteome and provided a list of fetal proteins of interest with potential for regenerative therapies for disc degeneration. In order to assess whether these potential candidates indeed have regenerative potential, future studies should focus on testing their regenerative capacities in mechanical experiments, in which an abundance of these proteins is added to adult discs to see if they can prevent degeneration, or added to degenerated discs to identify if they can reverse degeneration.

Received: 29 November 2019; Accepted: 19 August 2020

Published online: 24 September 2020

References

- Humzah, M. D. & Soames, R. W. Human intervertebral disc: structure and function. *Anat. Rec.* **220**, 337–356. <https://doi.org/10.1002/ar.1092200402> (1988).
- Roberts, S. Disc morphology in health and disease. *Biochem. Soc. Trans.* **30**, 864–869 (2002).
- Shapiro, I. M., Vresilovic, E. J. & Risbud, M. V. Is the spinal motion segment a diarthrodial polyaxial joint: what a nice nucleus like you doing in a joint like this?. *Bone* **50**, 771–776. <https://doi.org/10.1016/j.bone.2011.12.004> (2012).
- Vo, N. V. *et al.* Molecular mechanisms of biological aging in intervertebral discs. *J. Orthop. Res.* **34**, 1289–1306. <https://doi.org/10.1002/jor.23195> (2016).
- Roughley, P. J., Alini, M. & Antoniou, J. The role of proteoglycans in aging, degeneration and repair of the intervertebral disc. *Biochem. Soc. Trans.* **30**, 869–874 (2002).
- Yee, A. *et al.* Fibrotic-like changes in degenerate human intervertebral discs revealed by quantitative proteomic analysis. *Osteoarthritis Cartil.* **24**, 503–513. <https://doi.org/10.1016/j.joca.2015.09.020> (2016).
- Roughley, P. J. Biology of intervertebral disc aging and degeneration: involvement of the extracellular matrix. *Spine* **29**, 2691–2699 (2004).
- Ding, F., Shao, Z. W. & Xiong, L. M. Cell death in intervertebral disc degeneration. *Apoptosis* **18**, 777–785. <https://doi.org/10.1007/s10495-013-0839-1> (2013).
- Feng, H., Danfelter, M., Stromqvist, B. & Heinegard, D. Extracellular matrix in disc degeneration. *J. Bone Joint Surg. Am.* **88**(Suppl 2), 25–29. <https://doi.org/10.2106/JBJS.E.01341> (2006).
- GDALAP Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet* **388**, 1545–1602. [https://doi.org/10.1016/s0140-6736\(16\)31678-6](https://doi.org/10.1016/s0140-6736(16)31678-6) (2016).
- Cappello, R., Bird, J. L., Pfeiffer, D., Bayliss, M. T. & Dudhia, J. Notochordal cell produce and assemble extracellular matrix in a distinct manner, which may be responsible for the maintenance of healthy nucleus pulposus. *Spine* **31**, 873–882. <https://doi.org/10.1097/01.brs.0000209302.00820.fd> (2006) ((Discussion 883)).
- Hunter, C. J., Matyas, J. R. & Duncan, N. A. Cytomorphology of notochordal and chondrocytic cells from the nucleus pulposus: a species comparison. *J. Anat.* **205**, 357–362. <https://doi.org/10.1111/j.0021-8782.2004.00352.x> (2004).
- Trout, J. J., Buckwalter, J. A., Moore, K. C. & Landas, S. K. Ultrastructure of the human intervertebral disc. I. Changes in notochordal cells with age. *Tissue Cell* **14**, 359–369. [https://doi.org/10.1016/0040-8166\(82\)90033-7](https://doi.org/10.1016/0040-8166(82)90033-7) (1982).
- Hayes, A. J., Smith, S. M., Gibson, M. A. & Melrose, J. Comparative immunolocalization of the elastin fiber-associated proteins fibrillin-1, LTBP-2, and MAGP-1 with components of the collagenous and proteoglycan matrix of the fetal human intervertebral disc. *Spine* **36**, E1365–1372. <https://doi.org/10.1097/BRS.0b013e31821fd23e> (2011).
- Buckwalter, J. A., Kuettner, K. E. & Thonar, E. J. Age-related changes in articular cartilage proteoglycans: electron microscopic studies. *J. Orthop. Res.* **3**, 251–257. <https://doi.org/10.1002/jor.1100030301> (1985).
- Buckwalter, J. A., Pedrini-Mille, A., Pedrini, V. & Tudisco, C. Proteoglycans of human infant intervertebral disc Electron microscopic and biochemical studies. *J. Bone Joint Surg. Am.* **67**, 284–294 (1985).
- Smith, S. M., Whitelock, J. M., Iozzo, R. V., Little, C. B. & Melrose, J. Topographical variation in the distributions of versican, aggrecan and perlecan in the foetal human spine reflects their diverse functional roles in spinal development. *Histochem Cell Biol.* **132**, 491–503. <https://doi.org/10.1007/s00418-009-0623-z> (2009).
- Caldeira, J. *et al.* Matrisome profiling during intervertebral disc development and ageing. *Sci. Rep.* **7**, 11629. <https://doi.org/10.1038/s41598-017-11960-0> (2017).
- Rajasekaran, S. *et al.* ISSLS PRIZE IN CLINICAL SCIENCE 2017: Is infection the possible initiator of disc disease? An insight from proteomic analysis. *Eur. Spine J.* **26**, 1384–1400. <https://doi.org/10.1007/s00586-017-4972-3> (2017).
- Zybilov, B. *et al.* Statistical analysis of membrane proteome expression changes in *Saccharomyces cerevisiae*. *J. Proteome Res.* **5**, 2339–2347. <https://doi.org/10.1021/pr060161n> (2006).

21. Naba, A. *et al.* The extracellular matrix: tools and insights for the “omics” era. *Matrix Biol.* **49**, 10–24. <https://doi.org/10.1016/j.matbio.2015.06.003> (2016).
22. Higdon, R. & Kolker, E. A predictive model for identifying proteins by a single peptide match. *Bioinformatics* **23**, 277–280. <https://doi.org/10.1093/bioinformatics/btl595> (2007).
23. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504. <https://doi.org/10.1101/gr.1239303> (2003).
24. Bach, F. C. *et al.* Soluble and pelletable factors in porcine, canine and human notochordal cell-conditioned medium: implications for IVD regeneration. *Eur. Cell Mater.* **32**, 163–180. <https://doi.org/10.22203/ecm.v032a11> (2016).
25. Viola, M., Douglas, T. E., Alaniz, L. & Bartolini, B. Glycosaminoglycans metabolism. *Biochem. Res. Int.* **2012**, 245792. <https://doi.org/10.1155/2012/245792> (2012).
26. Henrotin, Y., Mathy, M., Sanchez, C. & Lambert, C. Chondroitin sulfate in the treatment of osteoarthritis: from in vitro studies to clinical recommendations. *Ther. Adv. Musculoskelet. Dis.* **2**, 335–348. <https://doi.org/10.1177/1759720X10383076> (2010).
27. Scott, I. C. *et al.* Bone morphogenetic protein-1 processes procollagen. *J. Biol. Chem.* **275**, 30504–30511. <https://doi.org/10.1074/jbc.M004846200> (2000).
28. Douglas, T., Heinemann, S., Bierbaum, S., Scharnweber, D. & Worch, H. Fibrillogenesis of collagen types I, II, and III with small leucine-rich proteoglycans decorin and biglycan. *Biomacromol* **7**, 2388–2393. <https://doi.org/10.1021/bm0603746> (2006).
29. Iozzo, R. V. & Schaefer, L. Proteoglycans in health and disease: novel regulatory signaling mechanisms evoked by the small leucine-rich proteoglycans. *FEBS J.* **277**, 3864–3875. <https://doi.org/10.1111/j.1742-4658.2010.07797.x> (2010).
30. Bourhis, J. M. *et al.* Procollagen C-proteinase enhancer grasps the stalk of the C-propeptide trimer to boost collagen precursor maturation. *Proc. Natl. Acad. Sci. USA* **110**, 6394–6399. <https://doi.org/10.1073/pnas.1300480110> (2013).
31. Wei, Y. & Tassava, R. A. Expression of type XII collagen by wound epithelial, mesenchymal, and endodermal cells during blastema formation in regenerating newt (*Notophthalmus viridescens*) tails. *J. Morphol.* **230**, 177–186. [https://doi.org/10.1002/\(SICI\)1097-4687\(199611\)230:2%3c177::AID-JMOR5%3e3.0.CO;2-E](https://doi.org/10.1002/(SICI)1097-4687(199611)230:2%3c177::AID-JMOR5%3e3.0.CO;2-E) (1996).
32. Chiquet, M., Birk, D. E., Bonnemant, C. G. & Koch, M. Collagen XII: Protecting bone and muscle integrity by organizing collagen fibrils. *Int. J. Biochem. Cell Biol.* **53**, 51–54. <https://doi.org/10.1016/j.biocel.2014.04.020> (2014).
33. Ansoorge, H. L. *et al.* Type XIV collagen regulates fibrillogenesis: premature collagen fibril growth and tissue dysfunction in null mice. *J. Biol. Chem.* **284**, 8427–8438. <https://doi.org/10.1074/jbc.M805582200> (2009).
34. Tono-Oka, S., Tanase, S., Miike, T. & Tanaka, H. Transient expression of collagen type XIV during muscle development and its reappearance after denervation and degeneration. *J. Histochem. Cytochem.* **44**, 907–918 (1996).
35. Hemmavanh, C., Koch, M., Birk, D. E. & Espana, E. M. Abnormal corneal endothelial maturation in collagen XII and XIV null mice. *Invest. Ophthalmol. Vis. Sci.* **54**, 3297–3308. <https://doi.org/10.1167/iops.12-11456> (2013).
36. Li, D., Clark, C. C. & Myers, J. C. Basement membrane zone type XV collagen is a disulfide-bonded chondroitin sulfate proteoglycan in human tissues and cultured cells. *J. Biol. Chem.* **275**, 22339–22347. <https://doi.org/10.1074/jbc.M000519200> (2000).
37. Hemmings, B. A. & Restuccia, D. F. PI3K-PKB/Akt pathway. *Cold Spring. Harb. Perspect. Biol.* **4**, a011189. <https://doi.org/10.1101/cshperspect.a011189> (2012).
38. Chen, J. *et al.* Akt1 regulates pathological angiogenesis, vascular maturation and permeability in vivo. *Nat. Med.* **11**, 1188–1196. <https://doi.org/10.1038/nm1307> (2005).
39. Chistiakov, D. A., Melnichenko, A. A., Myasoedova, V. A., Grechko, A. V. & Orekhov, A. N. Thrombospondins: a role in cardiovascular disease. *Int. J. Mol. Sci.* <https://doi.org/10.3390/ijms18071540> (2017).
40. Feng, G. *et al.* Hypoxia promotes nucleus pulposus phenotype in 3D scaffolds in vitro and in vivo: laboratory investigation. *J. Neurosurg. Spine* **21**, 303–309. <https://doi.org/10.3171/2014.4.Spine13870> (2014).
41. Kasprzycka, M., Hammarstrom, C. & Haraldsen, G. Tenascins in fibrotic disorders—from bench to bedside. *Cell. Adher. Migr.* **9**, 83–89. <https://doi.org/10.4161/19336918.2014.994901> (2015).
42. Osada, R. *et al.* Autocrine/paracrine mechanism of insulin-like growth factor-1 secretion, and the effect of insulin-like growth factor-1 on proteoglycan synthesis in bovine intervertebral discs. *J. Orthop. Res.* **14**, 690–699. <https://doi.org/10.1002/jor.1100140503> (1996).
43. Trehan, S. C. Q. Matrilin-3, a regulator of chondrocyte hypertrophy, bone mineral density, and osteoarthritis, binds bone morphogenetic protein-2, vascular endothelial growth factor165 and insulin-like growth factor. *55th Annual Meeting of the Orthopaedic Research Society* Post No. 427 (2009).
44. Li, J., Liu, Y., Wang, Y. & Xu, W. Expression of tenascin-C in a rat vocal fold injury model and its regulation of fibroblasts. *Laryngoscope* **128**, E316–e322. <https://doi.org/10.1002/lary.27164> (2018).
45. Hankenson, K. D., Hormuzdi, S. G., Meganck, J. A. & Bornstein, P. Mice with a disruption of the thrombospondin 3 gene differ in geometric and biomechanical properties of bone and have accelerated development of the femoral head. *Mol. Cell. Biol.* **25**, 5599–5606. <https://doi.org/10.1128/mcb.25.13.5599-5606.2005> (2005).
46. Newman, A. C., Nakatsu, M. N., Chou, W., Gershon, P. D. & Hughes, C. C. The requirement for fibroblasts in angiogenesis: fibroblast-derived matrix proteins are essential for endothelial cell lumen formation. *Mol. Biol. Cell* **22**, 3791–3800. <https://doi.org/10.1091/mbc.E11-05-0393> (2011).
47. Kimura, H. *et al.* Cthrc1 is a positive regulator of osteoblastic bone formation. *PLoS ONE* **3**, e3174. <https://doi.org/10.1371/journal.pone.0003174> (2008).
48. Lee, J. *et al.* CTHRC1 promotes angiogenesis by recruiting Tie2-expressing monocytes to pancreatic tumors. *Exp. Mol. Med.* **48**, e261. <https://doi.org/10.1038/emm.2016.87> (2016).
49. Pyagay, P. *et al.* Collagen triple helix repeat containing 1, a novel secreted protein in injured and diseased arteries, inhibits collagen expression and promotes cell migration. *Circ. Res.* **96**, 261–268. <https://doi.org/10.1161/01.Res.0000154262.07264.12> (2005).
50. Conway, S. J. *et al.* The role of periostin in tissue remodeling across health and disease. *Cell Mol Life Sci* **71**, 1279–1288. <https://doi.org/10.1007/s00018-013-1494-y> (2014).
51. Tsai, T. T. *et al.* Increased periostin gene expression in degenerative intervertebral disc cells. *Spine J.* **13**, 289–298. <https://doi.org/10.1016/j.spinee.2013.01.040> (2013).
52. Romanelli, R. G. *et al.* Effect of pentoxifylline on the degradation of procollagen type I produced by human hepatic stellate cells in response to transforming growth factor-beta 1. *Br. J. Pharmacol.* **122**, 1047–1054. <https://doi.org/10.1038/sj.bjp.0701484> (1997).
53. Hildebrand, A. *et al.* Interaction of the small interstitial proteoglycans biglycan, decorin and fibromodulin with transforming growth factor beta. *Biochem. J.* **302**(Pt 2), 527–534 (1994).
54. Korpos, E., Deak, F. & Kiss, I. Matrilin-2, an extracellular adaptor protein, is needed for the regeneration of muscle, nerve and other tissues. *Neural. Regen. Res.* **10**, 866–869. <https://doi.org/10.4103/1673-5374.158332> (2015).
55. Malin, D. *et al.* The extracellular-matrix protein matrilin 2 participates in peripheral nerve regeneration. *J. Cell. Sci.* **122**, 995–1004. <https://doi.org/10.1242/jcs.040378> (2009).
56. The matrilins: a novel family of oligomeric extracellular matrix proteins. <https://www.ncbi.nlm.nih.gov/pubmed/10367731> (2019).
57. Zhang, S. *et al.* Matrilin-2 is a widely distributed extracellular matrix protein and a potential biomarker in the early stage of osteoarthritis in articular cartilage. *Biomed. Res. Int.* **2014**, 986127. <https://doi.org/10.1155/2014/986127> (2014).
58. Deak, F. *et al.* Extracellular deposition of matrilin-2 controls the timing of the myogenic program during muscle regeneration. *J. Cell Sci.* **127**, 3240–3256. <https://doi.org/10.1242/jcs.141556> (2014).

Acknowledgements

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.

Author contributions

S.R., C.T. and N.D. wrote the main manuscript text, C.T., S.M.N. and M.S.M. performed the laboratory work, and S.R., M.R., D.C.R.S., K.S. supervised.

Funding

This study was funded by Ganga Orthopaedic Research & Education Foundation.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-72859-x>.

Correspondence and requests for materials should be addressed to S.R.

Reprints and permissions information is available at www.nature.com/reprints.

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020