Article

Peptide location fingerprinting reveals tissue region-specific differences in protein structures in an ageing human organ

Alexander Eckersley[1]*, Matiss Ozols[1,4], Peikai Chen[5,6], Vivian Tam[5], Judith A Hoyland[1,3], Andrew Trafford[2], Danny Chan[5] and Michael J Sherratt[1]*

[1] Division of Cell Matrix Biology & Regenerative Medicine, School of Biological Science, Faculty of Biology, Medicine and Health, The University of Manchester, Manchester, UK; matiss.ozols@manchester.ac.uk (M.O.)

[2] Division of Cardiovascular Sciences, School of Medical Sciences, Faculty of Biology, Medicine and Health, The University of Manchester, Manchester UK; Andrew.W.Trafford@manchester.ac.uk (A.T.)

[3] NIHR Manchester Biomedical Research Centre, Central Manchester Foundation Trust, Manchester Academic Health Science Centre, Manchester, UK; judith.a.hoyland@manchester.ac.uk (J.A.H.)

[4] Department of Human Genetics, Wellcome Sanger Institute, Genome Campus, Hinxton, UK; mo11@sanger.ac.uk (M.O.)

[5] School of Biomedical Sciences, The University of Hong Kong, Hong Kong, China; vivtam@hku.hk / viv.tam@gmail.com (V.T.) chand@hku.hk (D.C.)

[6] The University of Hong Kong - Shenzhen Hospital (HKU-SZH), Shenzhen, China; pkchen@hku-szh.org (P. C.)

* Correspondence: alexander.eckersley@manchester.ac.uk (A.E.); michael.j.sherratt@manchester.ac.uk (M.J.S)

Abstract: In ageing tissues, long-lived extracellular matrix (ECM) proteins are susceptible to the accumulation of structural damage due to diverse mechanisms including glycation, oxidation and protease cleavage. Peptide location fingerprinting (PLF) is a new mass spectrometry (MS) analysis technique capable of identifying proteins exhibiting structural differences in complex proteomes. PLF applied to published young and aged intervertebral disc (IVD) MS datasets (posterior, lateral and anterior regions of the annulus fibrosus), identified 268 proteins with age-related structural differences. For several ECM assemblies (collagens I, II and V and aggrecan), these differences were markedly conserved between degeneration-prone (posterior and lateral) and resistant (anterior) regions. Significant differences in peptide yields, observed within collagen I, II and V α-chains (COL1A2, COL2A1, COL5A1), were located within their triple helical regions and/or cleaved C-terminal propeptides, indicating potential accumulation of damage and impaired maintenance in ageing. Several proteins (COL5A1, COL2A1 and aggrecan) also exhibited tissue region (lateral)-specific differences in structure between aged and young, suggesting that some ageing mechanisms may act locally within tissues. This study not only provides evidence of age-related changes in ECM protein structures which are tissue-region specific, but also highlights the ability of PLF to identify potential protein biomarkers of localised tissue remodelling.

Keywords: proteomics; peptide location fingerprinting; extracellular matrix; biomarkers; ageing; intervertebral disc; spine; mass spectrometry

1. Introduction

Intervertebral discs (IVDs) are fibrocartilaginous cushions residing between vertebrae, which confer flexibility and stability to the spinal column, and function as effective shock absorbers over an individual's lifetime [1]. Age-associated degeneration of the IVD in the lower lumbar spine is the leading cause of lower back pain within an ageing population [2], imposing major socioeconomic and healthcare burdens.

IVDs (located between endplates of vertebrae) consist of two extracellular matrix (ECM)-rich tissue compartments: 1) a central gel-like nucleus pulposus (NP), composed predominantly of the chondroitin sulphate proteoglycans (CSPGs) aggrecan and versican and a relatively disordered network of collagen II; and 2) an encircling annulus fibrosus (AF) consisting of highly ordered, concentric rings (lamellae) of parallel collagen I and elastic fibres arranged in a crisscross pattern [3,4]. These compartments are large, avascular and subject to mechanical loads, creating a challenging environment for tissue repair and regeneration (mediated mainly by a small population of chondrocyte-like cells in the NP and fibroblast-like cells in the AF [5]). Crucially, regional differences in ECM composition exist within the AF. ECM components within the inner AF reflect the interface between the NP and outer AF, being comprised of not only collagen I fibrils but also CSPGs and collagen II fibrils [4,6].

Clinically, the onset and progression of IVD ageing and degeneration is characterised typically by a gradual reduction of hydration, pressure and height of the NP, and the formation of tears in AF lamellae making the disc more prone to bulging, protrusion or herniation [7]. This structural damage to the AF, and hence susceptibility to failure, is profoundly regional, with degeneration mainly localised in the posterior-lateral portion of the IVD [8,9]. The dynamic alteration in ECM composition and abundance likely underpins this regional progression of IVD ageing and degeneration with decreasing levels of CSPGs in both the NP and AF and increasing levels of collagen I fibres, leading to a fibrotic, dehydrated phenotype [4,10–13]. Additionally, we previously showed that the protein composition of the inner AF becomes gradually more similar to the outer AF with ageing [13]. It is clear therefore that understanding the mechanisms and consequences of ECM remodelling is key to defining and potentially reversing IVD degeneration and ageing.
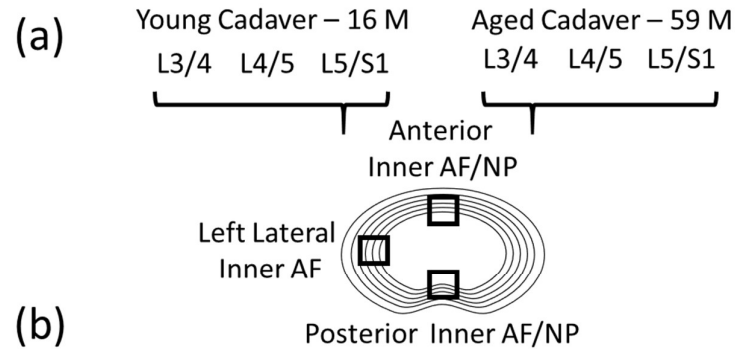
ECM assemblies and components can be long-lived, with biological half-lives measured in years or decades. For example, lung elastic fibres and collagens in cartilaginous tissues remain present throughout the life of an individual [14,15]. Consequently, ECM assemblies residing in the IVD (such as collagen I in the AF [16] and aggrecan in the NP [17]) are susceptible to the progressive accumulation of damage to both their molecular structures and networks over time, leading to loss of functions of many diverse ECM components. The potential causative mechanisms of this remodelling include the chronic exposure of ECM proteins to elevated protease mediated cleavage [18], cross-linking by sugars leading to advanced glycation end products (AGEs) [19], mechanical fragmentation [20] and oxidation by reactive oxygen species (ROS) [21]. These processes may lead to a spectrum of modifications to protein higher order structures that are challenging to detect. Although several proteomics studies have characterised IVD ageing and degeneration through changes in protein presence and abundance alone [12,13,22–24], the identification of ECM proteins susceptible to damage modifications to structure is also crucial both for the discovery of novel ageing biomarker candidates and for deciphering the underlying mechanisms involved.

Peptide location fingerprinting (PLF) is a new computational technique capable of identifying proteins exhibiting structural differences in mass spectrometry (MS) datasets derived from complex biological samples regardless of the causal mechanisms [25]. In standard proteomics, proteins are exposed to enzymes (e.g. trypsin), generating peptides whose sequence identities and relative abundances are then measured by liquid chromatography tandem MS (LC-MS/MS). PLF relies on the fact that, for a given protein, the pattern of peptides detected by LC-MS/MS following protease digestion is determined by its higher order structure and reflects its solubility, stability and enzyme susceptibility. Therefore, differences in protein structure, due to either the accumulation of damage (via ageing and disease), changes in interaction states with other protein or differential synthesis, may be detected by PLF [25–27]. This is particularly true for the highly insoluble ECM proteins which exist within tightly bound networks. In standard LC-MS/MS, peptides are mapped to proteins enabling quantification of protein abundance. PLF maps and
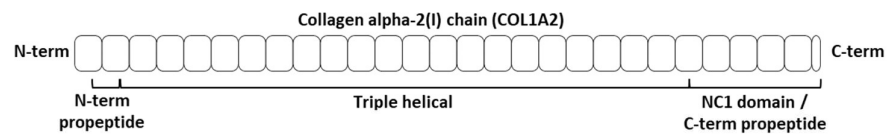
quantifies peptides (by spectral counting) within specific protein regions instead, enabling the detection of statistical differences along the primary structure (peptide yield patterns) [25,26]. Previously, we have used PLF to reproducibly detect UVR-induced damage within the structures of photosensitive skin ECM proteins, *in vitro* [26] and tissue-specific differences in eye (ciliary zonule) and skin fibrillin-1 [27]. Crucially, and most recently, we have developed and employed an online webtool (MPLF) to identify potential bi-omarker proteins with structural differences between photoaged and photo-protected skin and young and aged tendon [25]. Comparisons with conventional protein quantification also revealed that PLF identifies significant differences which are unique to the methodology, particularly in ECM proteins, demonstrating that modifications to protein structure may not correlate with protein abundance in damage-susceptible long-lived proteins [25].

Recently, we have also published DIPPER, a detailed spatiotemporal proteomic analysis of the ageing and degenerating human IVD, which included high resolution label-free LC-MS/MS performed on 11 separate regions of discs sourced from a young (16 yr) and an aged (59 yr) male cadaveric lumbar spines [13]. This study revealed novel, region-specific differences in ECM proteome (matrisome) composition and abundance, which demonstrated a dynamic shift along the lateral and anteroposterior axes, with a rapid decline of NP proteins and increase in AF proteins in the aged inner AF. Although this proteomic study was highly detailed, modification-related differences to ECM protein structures were not interrogated. Since degeneration in the ageing disc is profoundly regional [8,9], and PLF can be applied post hoc to historical LC-MS/MS data (as shown for aging tendon [25]), further analysis of the DIPPER dataset presents us with a timely opportunity to test: i) the hypothesis that protein structures (as indicated by peptide yield pattern) are spatially- and potentially age-dependent in the IVD and ii) the ability of PLF to detect protein structural differences within localised tissue regions.
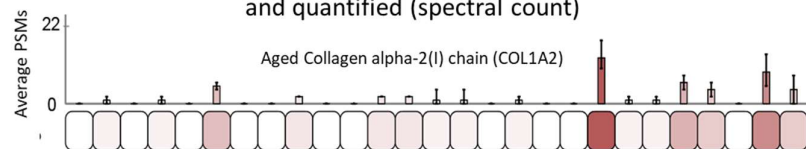
As the inner AF appeared particularly susceptible to matrisomal changes [13], we applied PLF to LC-MS/MS datasets corresponding to the inner AFs of three distinct IVD regions (posterior, left lateral and anterior; Figure 1a). This enabled the detection of ageing structural differences in ECM proteins between the degeneration prone (posterior and lateral) and resistant (anterior) regions of the IVD (Figure 1b) and revealed biomarker candidates with unique and common age-related peptide yield patterns.
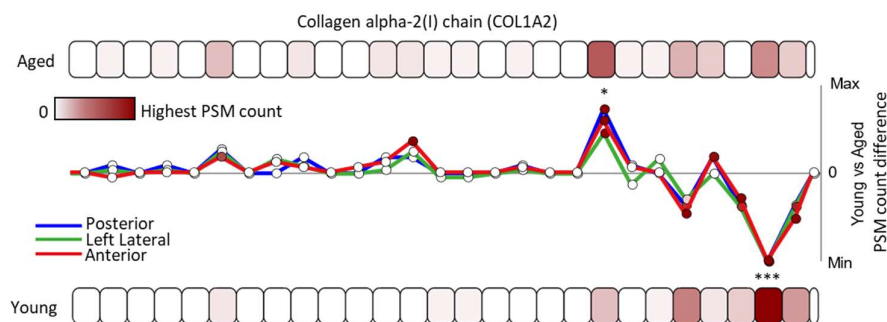
Figure 1. Experimental design for the application of PLF to three distinct regions of young and aged IVDs. (**a**) Raw LC-MS/MS datasets (from two males: young [16 yrs] and aged [59 yrs]), corresponding to the posterior inner AF/NP, left lateral inner AF and anterior inner AF/NP regions of the L3/4, L4/5 and L5/S1 lower lumbar IVDs were downloaded from the PRIDE repository (PXD017740) [13]. (**b**) After MS/MS peptide identification, the MPLF webtool [25] was used to perform PLF on identified peptides and compare peptide yields across protein structures between young and aged discs – 3 vs. 3 discs for anterior and left lateral datasets and 3 vs 2 discs for posterior datasets (young anterior L4/5 disc was omitted from analysis due to insufficient identified peptides for PLF analysis). As previously described [25], (**b, i**) proteins were divided into 50 amino acid (aa)-sized segments; the collagen I α2 chain (COL1A2) is shown here as an example. (**b, ii**) Peptides were mapped to their respective segments, summed, normalised between aged and young groups based on the median whole protein spectrum count and averaged for each group (average peptide spectrum match [PSM] count per segment). (**b, iii**) Peptide yields in each segment were statistically compared between young and aged using an unpaired, Bonferroni-corrected, repeated measures ANOVA. Proteins containing segments with significant differences in peptide yield between young and aged were

considered biomarker candidates of ageing. To visualise these differences along the protein structure, average PSM counts per segment in the young group were subtracted from those in aged and compared between posterior, left lateral and anterior IVD regions.

2. Results and Discussion

2.1. Peptide Location Fingerprinting Reveals IVD Region-specific Biomarker Candidates of Ageing

MS/MS peptide searches identified 260 posterior, 391 left lateral and 303 anterior inner AF-specific proteins in both young and aged groups (Figure S1; peptide reports: Tables S1 – S3). Principal component analyses of peptide spectral counts demonstrated clear separations of data between young and aged discs for all three IVD regions tested (Figure S2), reflecting a global compositional difference in ageing similar to that showcased previously in the DIPPER study [13]. PLF analysis identified a combined total of 268 proteins (across tissue regions) with significant differences in peptide yield patterns along their primary structures (Figure 2a; full PLF analysis reports: Tables S4 – S6). When compared to biomarker candidates identified in the DIPPER study by label-free quantification of relative protein abundance [13], PLF identified 119 proteins that were unique to this approach, with age-associated significant differences in protein structure but not abundance (Figure S3).

Of the 268 potential ageing biomarkers, 38 were shared between posterior, lateral and anterior inner AF regions (Figure 2a). Classification of these shared biomarker candidates into ECM-specific protein classes highlighted several matrix-associated proteins, including the major structural collagens of the AF (collagen I and V) and NP (collagen II and VI), elastic fibre associated proteins (fibrillin-1 and LTBP2) and aggrecan as key affected components (Figure 2b). Some of the identified proteins have been linked with disease including osteoarthritis-associated [28,29] cartilage intermediate layer proteins (CILP) 1 and 2, levels of which were shown to differ significantly between degenerated and non-degenerated AFs and NPs [12], and the cell-matrix adhesion-mediating thrombospondins (THBS) 2 [30] and 4 [31]. Levels of THBS4 were shown previously in the DIPPER study to be significantly higher in young inner AF than in aged [13]. However, to our knowledge, age-dependent differences in disc THSB1 have not been previously shown, although two single nucleotide polymorphisms were found to be associated with IVD degeneration [32].
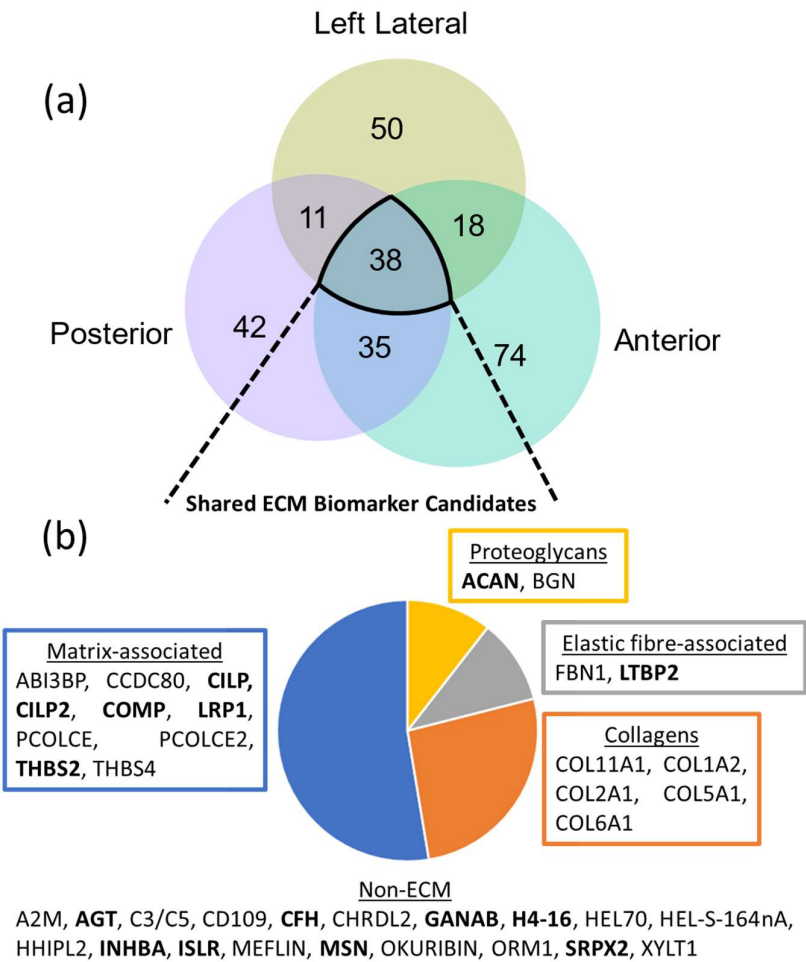
Figure 2. ECM-associated proteins make up half of the biomarker candidates identified with significant, age-dependant structural differences in all three IVD regions. PLF revealed 126 proteins in posterior (Table S4), 117 in left lateral (Table Scheme 5. and 165 in anterior (Table S6) IVD regions with significant differences in peptide yield across their structures. (**a**) Of these, 38 were shared between the three regions. (**b**) Classification of these shared biomarker candidates revealed ECM proteins as the major class, which included several ECM-associated proteins, collagens, two elastic fibre-specific proteins and two proteoglycans. Protein biomarker candidate identities which were uniquely discovered by PLF and not by label-free quantification of abundance in the DIPPER study [13] (Figure S3) are highlighted in bold.

The decline of the proteoglycan aggrecan in IVD ageing due to proteolysis by MMPs and aggrecanases is well established [33]. Here, we show evidence of structural differences within the aggrecan core protein between aged and young IVD which may relate to these mechanisms. Similarly, the degeneration of elastic fibre and collagen architectures are a hallmark of ageing in many connective tissues such as IVD, skin and aorta [34–38]. Our data shows that, in the IVD, this process may extend specifically to modification-associated changes in the protein structures of elastic fibre proteins fibrillin-1 and latent TGFβ-binding protein (LTBP) 2 and in multiple α-chains from collagens I, II, V, VI and XI.

To investigate whether the age-dependent differences in peptide yield patterns along the primary structure of ECM proteins are consistent between regions, members of these 38 shared biomarker candidates were further examined.

2.2. Age-associated Fluctuations in PeptideYield Along the Primary Structures of COL1A2 and PCOLCE are Markedly Conserved Between Posterior, Lateral and Anterior IVD Regions

Both the collagen I α2 chain (COL1A2) and procollagen C-endopeptidase enhancer 1 (PCOLCE; PCPE1) were identified with age-associated modifications to their structures in all three IVD regions (Figure 2b). The COL1A2 chain is a major structural component of collagen I fibrils, forming a tightly bound triple helix alongside two α1 chains (COL1A1) within a monomer. The extracellular glycoprotein PCOLCE is an enhancer that binds to the C-terminal propeptide of procollagens and drives the activity of tolloid proteinases, their cleavage of procollagens and their subsequent assembly into fibres [39]. PCOLCE therefore plays a crucial role in collagen deposition and associated morbidities such as fibrosis [40] .

Comparisons between posterior, lateral and anterior regions revealed notable, consistent age-dependant differences in peptide yield across the protein structures of COL1A2 and PCOLCE. On the C-terminal side of the COL1A2 protein, a characteristic significant increase and decrease in peptide yield pattern was observed in aged compared to young, along the primary structure for all three IVD regions (Figure 3a). Similarly, a characteristic increase in peptide yield was seen on the N-terminal side of aged PCOLCE compared to young, followed by a significant decrease in peptide yield on the C-terminal side (Figure. 3b). This pattern was also consistent for posterior, lateral and anterior inner AF regions. While there were age-dependant differences observed along the structures of both proteins, it is important to consider that peptide yields were also stable between aged and young in many protein regions (such as on the N-terminal side of COL1A2). This lack of difference between young and old was also consistent between IVD regions.
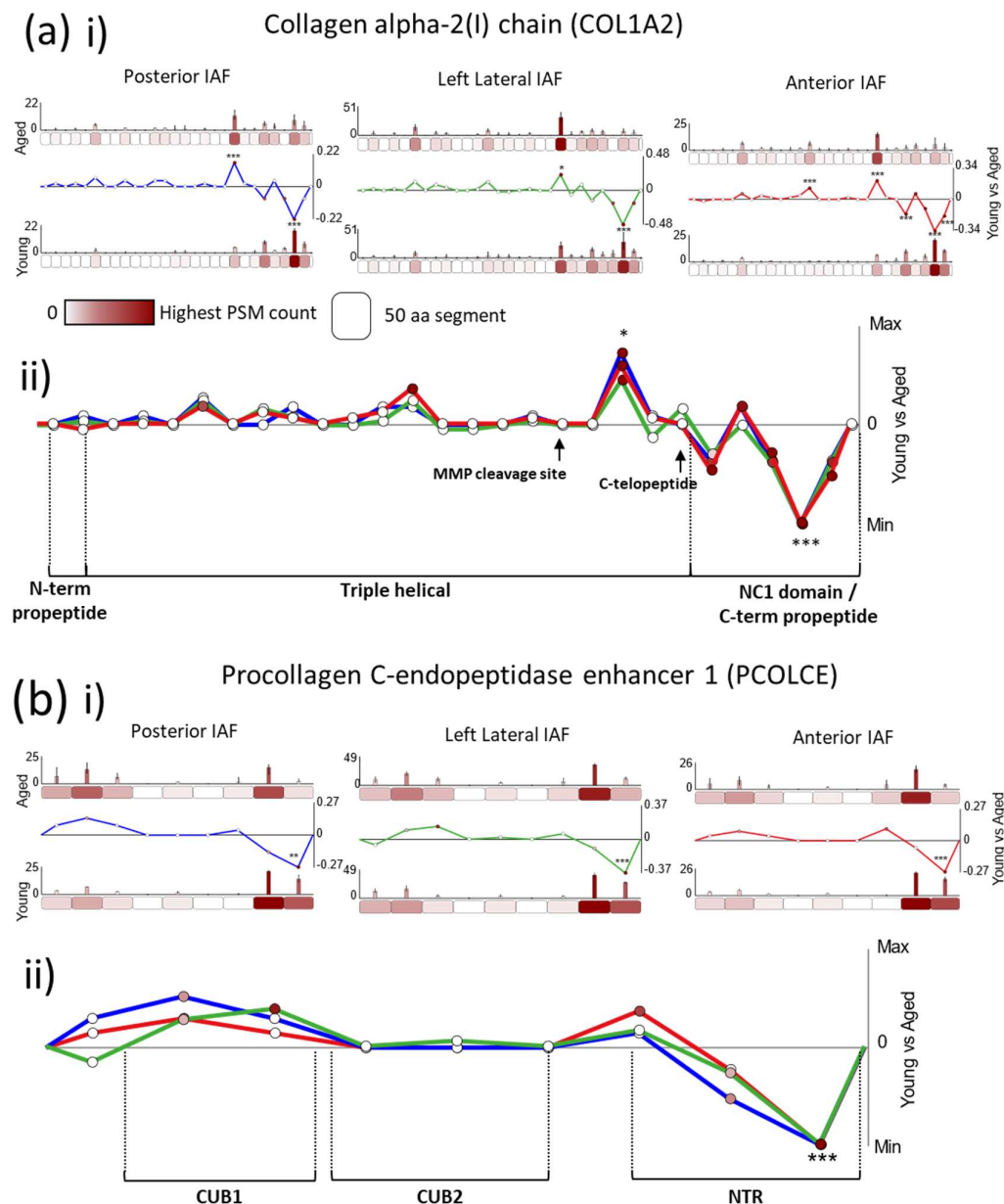
Figure 3. COL1A2 and PCOLCE protein structures exhibit age-dependant differences in peptide yield patterns which were conserved between posterior, left lateral and anterior IVD regions. PSMs were summed in each 50 aa protein segment, normalised and averaged for young and aged groups (bar graphs = average, normalised PSMs, error bars = standard deviation). Average PSMs per segment in young were subtracted by those in aged and divided by segment aa length (50) to reveal fluctuations in peptide yield along protein structures (line graph y axes = aged – young PSM counts / segment length, normalised between tissue regions in composite graphs; *, $p \leq 0.05$; ** $p \leq 0.01$; ***, $p \leq 0.001$, Bonferroni-corrected, repeated measures ANOVA; black * in composite line graph = significant in all three IVD regions; functional domains, sourced from UniProt, are indicated). (**a, i**) Several COL1A2 segments exhibited significant differences in peptide yield between young and aged within the C-terminal half of the protein (blue line: posterior, green: left lateral, red: anterior). (**a, ii**) Crucially, these differences in peptide yield along the COL1A2 structure are markedly consistent between all the three regions with a similarly higher peptide yield displayed in segment 20 of aged compared to young in and lower peptide yield in the last three segments, nearest to the C-terminus. (**b, i**) The last 50 aa segment on the C-terminal end of PCOLCE also exhibited a significantly lower peptide yield in aged compared to young for the inner AFs of all three IVD regions. (**b, ii**) Peptide yield difference patterns along the primary structure were also similar between all the three regions, with a consistently higher peptide yield in aged compared to young measured on the N-terminal side of the protein fol-lowed by a lower yield in the last two segments near the C-terminal end.

The significantly higher peptide yield, observed on the C-terminal end of young COL1A2 compared to aged, corresponds directly to its non-collagenous (NC1) domain (Figure 3a, ii). These domains are cleaved prior to fibril assembly [41] by C-proteinases such as bone morphogenic protein-1/tolloid proteinase (BMP1) from fully folded procollagen, releasing a C-terminal propeptide. Since the release of this domain is critical for the formation and assembly of the collagen triple helix protomer [42], its higher detection in young IVD suggests heightened collagen synthesis compared to aged. Also in support of this, the significantly lower peptide yield in aged PCOLCE compared to young, seen on its C-terminal end for all IVD regions, coincides directly with its netrin (NTR) domain (Figure. 3b, ii); critical for the super-stimulation of BMP1 and its cleavage of the collagen NC1 domain prior to fibril assembly [43]. This lower peptide yield in the netrin domain of aged PCOLCE may impact on its interaction with BMP1. Age-related changes to PCOLCE structure may have resulted in masking of this protein region, which would explain its lower availability to trypsin and potentially BMP1, also impacting on collagen synthesis. Recent studies of collagen turnover in healthy tendon have shown that collagen is not repaired [44,45] but instead maintained daily, through a cycle of synthesis and degradation that is controlled by the circadian rhythm [46]. It is likely that a similar maintenance program exists in young, healthy AF (compositionally similar to tendon) and here, we show evidence this may deteriorate with age.

A second COL1A2 protein region, located near the C-terminal end of the triple helical domain (Figure 3a, ii), also differed significantly between aged and young (higher peptide yields in aged) for all three IVD regions. As this protein region is located well within the triple helix of collagen I, peptides derived from this domain are likely to be from mature collagen fibrils. In osteoporosis, fragments of the C-terminal telopeptide of collagen I α1 are frequently used as markers of collagen degradation in bone [47,48]. Although the COL1A2 protein region seen here is located upstream of its C-terminal telopeptide (Figure 3a, ii, black arrow), a higher peptide yield in aged IVD may also be indicative of collagen fibril degradation as a consequence of damage accumulation to structure. COL1A2 is a substrate for matrix metalloproteinases (MMPs) 1, 2, 3, 8 and 14 [49] (see interaction network Figure S4), all of which were shown to be upregulated in degenerated IVD [50]. All of these (except MMP3) preferentially cleave the α-chains of collagens I, II and III approximately ¾ from the N-terminus along the mature chain [51,52]. The COL1A2 region, which had higher peptide yields in aged than in young, is located downstream of this prominent MMP cleavage site (indicated in Figure 3a, ii, black arrow) and corresponds to this ¾ degradation fragment. This suggests that chronic exposure to elevated endogenous MMPs, thought to be one of several ageing mechanisms driving the accumulation of damage to long-lived ECM assemblies, is likely to play role in the higher peptide yield observed in the triple helical domain of aged COL1A2 compared to young.
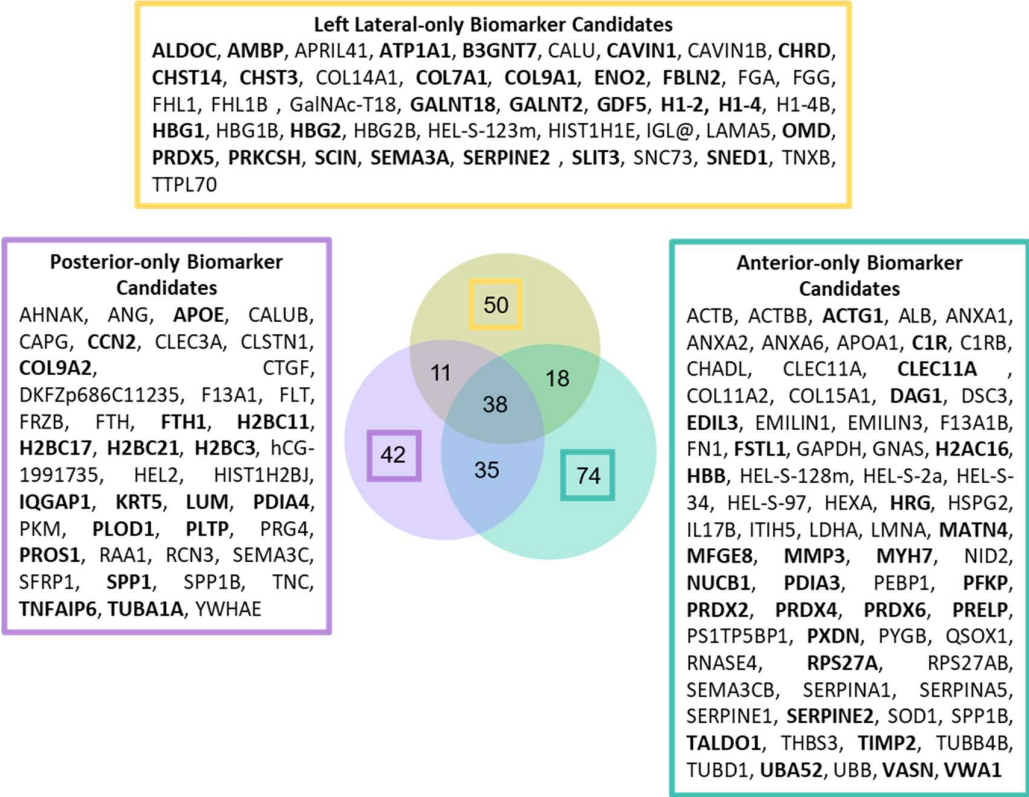
In addition to BMP1, PCOLCE also binds fibronectin [53], LTBP1, THSB1, aggrecan and biglycan [54] (and their associated glycosaminoglycans) and COL1A2 binds COL5A1 [55] within the wider ECM network (Figure S4). If the structural differences seen in aged PCOLCE and COL1A2 compared to young (Figure 3) result in the impediment of these interactions, it could drive the deterioration of the wider proteostasis in ageing IVD [13].

The remarkable consistency in the age-dependant peptide yield patterns of COL1A2 and PCOLCE structures between the posterior, lateral and anterior regions of the AF (Figure 3) is an interesting observation as it suggests that similar ageing mechanisms may be at play tissue wide. However, some proteins exhibited significant structural differences in peptide yield which were entirely tissue region-specific.

2.3. Age-related Peptide Yield Patterns Along the Primary Structures of the α1 Chains of Collagens II and V and Aggrecan Exhibit Regional Tissue Specificity Within the IVD

A sub-population of IVD proteins showed significant differences in structure which were region-specific (42 for posterior, 50 for lateral and 74 for anterior) as a consequence

of ageing (Figure 4). These proteins (potentially biomarkers) included several prominent ECM components such as tenascin C and proteoglycan 4 (exclusive to posterior), collagen XIV α1 and osteomodulin (unique to left lateral), and EMILIN-1 and nidogen 2 (exclusive to anterior) (Figure S5).

Left Lateral-only Biomarker Candidates
ALDOC, AMBP, APRIL41, **ATP1A1**, **B3GNT7**, CALU, **CAVIN1**, CAVIN1B, **CHRD**, **CHST14**, **CHST3**, COL14A1, **COL7A1**, **COL9A1**, **ENO2**, **FBLN2**, FGA, FGG, FHL1, FHL1B , GalNAc-T18, **GALNT18**, **GALNT2**, **GDF5**, **H1-2**, **H1-4**, H1-4B, **HBG1**, HBG1B, **HBG2**, HBG2B, HEL-S-123m, HIST1H1E, IGL@, LAMA5, **OMD**, **PRDX5**, **PRKCSH**, **SCIN**, **SEMA3A**, **SERPINE2** , **SLIT3**, SNC73, **SNED1**, TNXB, TTPL70

Posterior-only Biomarker Candidates
AHNAK, ANG, **APOE**, CALUB, CAPG, **CCN2**, CLEC3A, CLSTN1, **COL9A2**, CTGF, DKFZp686C11235, F13A1, FLT, FRZB, FTH, **FTH1**, **H2BC11**, **H2BC17**, **H2BC21**, **H2BC3**, hCG-1991735, HEL2, HIST1H2BJ, **IQGAP1**, **KRT5**, **LUM**, **PDIA4**, PKM, **PLOD1**, **PLTP**, PRG4, **PROS1**, RAA1, RCN3, SEMA3C, SFRP1, **SPP1**, SPP1B, TNC, **TNFAIP6**, **TUBA1A**, YWHAE

Anterior-only Biomarker Candidates
ACTB, ACTBB, **ACTG1**, ALB, ANXA1, ANXA2, ANXA6, APOA1, **C1R**, C1RB, CHADL, CLEC11A, **CLEC11A** , COL11A2, COL15A1, **DAG1**, DSC3, **EDIL3**, EMILIN1, EMILIN3, F13A1B, FN1, **FSTL1**, GAPDH, GNAS, **H2AC16**, **HBB**, HEL-S-128m, HEL-S-2a, HEL-S-34, HEL-S-97, HEXA, **HRG**, HSPG2, IL17B, ITIH5, LDHA, LMNA, **MATN4**, **MFGE8**, **MMP3**, **MYH7**, NID2, **NUCB1**, **PDIA3**, PEBP1, **PFKP**, **PRDX2**, **PRDX4**, **PRDX6**, **PRELP**, PS1TP5BP1, **PXDN**, PYGB, QSOX1, RNASE4, **RPS27A**, RPS27AB, SEMA3CB, SERPINA1, SERPINA5, SERPINE1, **SERPINE2**, SOD1, SPP1B, **TALDO1**, THBS3, **TIMP2**, TUBB4B, TUBD1, **UBA52**, UBB, **VASN**, **VWA1**

Venn diagram values: 50, 11, 18, 38, 42, 35, 74

Figure 4. A total of 166 proteins were identified with significant, age-dependant structural differences that were entirely IVD-region specific. Protein biomarker candidate identities which were uniquely discovered by PLF and not by label-free quantification of abundance in the DIPPER study (Figure S3) [13] are highlighted in bold.

Importantly, some proteins even displayed age-dependant differences in peptide yield which were shared between all AF regions (Figure. 2b) on one side of their protein structures but region-specific on the other side, as seen for collagen V α1 (COL5A1), collagen II α1 (COL2A1) and aggrecan core protein (Figure 5). For COL5A1, a characteristic significant decrease in peptide yield in aged compared to young, followed by an increase can be seen in the C-terminal end, along the primary structure, which remains consistent for all three IVD regions (Figure 5a). This is similar also for COL2A1 which exhibited the same significant rise and fall in peptide yield along its chain in all three IVD regions, on its C-terminal side (Figure 5b). However, both α1 chains of collagens V and II displayed additional N-terminal, significant increases in peptide yield in aged compared to young that were lateral-specific (green brackets in Figure 5a, ii and b, ii). Contrastingly, a significantly higher peptide yield was observed near the N-terminal end of aged aggrecan than in young, for both the posterior and anterior inner AF regions (Figure 5c). However, this pattern was reversed for left lateral, with the same protein region yielding significant lower peptide counts in aged aggrecan than in young (green bracket in Figure 5c, ii).
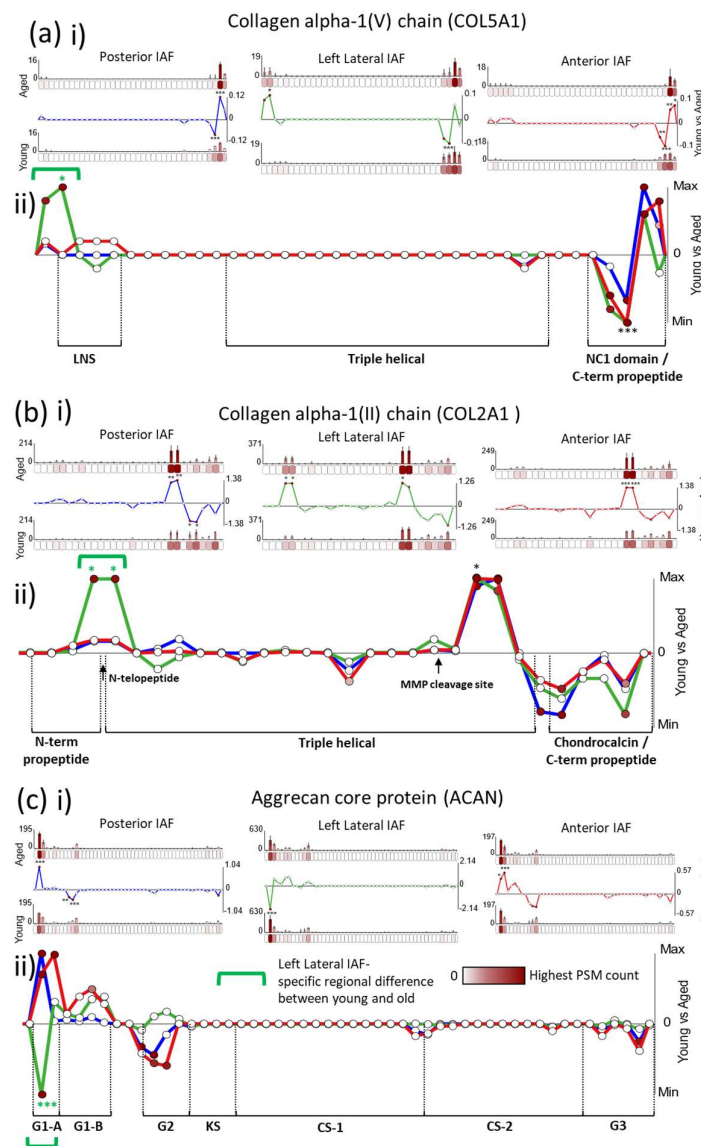
Figure 5. COL5A1, COL2A1 and aggrecan all exhibit lateral, inner AF-specific differences in peptide yield not seen in posterior or anterior IVD regions. PSMs were summed in each 50 aa protein segment, normalised and averaged for young and aged groups (bar graphs = average, normalised PSMs, error bars = standard deviation). Average PSMs per segment in young were subtracted by those in aged and divided by segment aa length (50) to reveal fluctuations in peptide yield along protein structures (line graph y axes = aged – young PSM counts / segment length, normalised between tissue regions in composite graphs; *, $p \leq 0.05$; ** $p \leq 0.01$; ***, $p \leq 0.001$, Bonferroni-corrected, repeated measures ANOVA; black * in composite line graph = significant in all three IVD regions; functional domains, sourced from UniProt, are indicated). (**a, i**) Several segments on the C-terminal end of COL5A1 exhibited significant differences in peptide yields between young and aged for all three IVD regions. However, a left lateral-specific significant difference in yield was also observed in its second segment, on the N-terminal end of the protein. (**a, ii**) Fluctuations in COL5A1 peptide yield on the C-terminal end follow a similar pattern between all three regions, with a consistent fall and rise of peptide yield observed in aged compared to young for the last four segments of its primary structure. However, a clear left lateral-specific increase in peptide yield can also be seen on the N-terminal end. (**b, i**) COL2A1 also had several protein segments on the C-terminal side which exhibited significant differences in peptide yield. (**b, ii**) These displayed a characteristic rise and fall peptide yield pattern in aged compared to young which were markedly consistent between the three IVD regions. However, once again, two COL2A1 segments on the N-terminal side of the protein exhibited significantly higher yields for aged than for young that were unique to the left lateral region of the disc. (**c, i**) Segments two and three on the N-terminal side of the aggrecan protein also exhibited significant differences in peptide yield between young and aged. (**c, ii**) These differences were significantly higher in aged aggrecan within the posterior and anterior inner AF regions, compared to young, but significantly lower in left lateral.

The tissue region-conserved, higher peptide yields seen on the C-terminal ends of young COL5A1 and COL2A1 compared to aged (Figure 5a, ii and b, ii) also corresponded to their cleaved C-terminal propeptides, as seen for COL1A2 (Figure 3a, ii). This once again reflects a possible tissue-wide perturbation in collagen maintenance [46] as a consequence of ageing, for collagen II and V fibrils in addition to collagen I. Interestingly, we noticed that the C-terminal propeptides of COL1A2 and COL5A1 are enriched in ROS-sensitive residues [56] (Figure. S6), indicating a potential role for these fragments as anti-oxidants, whose presence is reduced in aged IVD (Figure 3a, ii and Figure 5a, ii). Again in common with COL1A2 (Figure 3a, ii), aged COL2A1 also exhibited an AF region-wide significant increase in peptide yield compared to young, downstream to a prominent MMP cleavage site (Figure 5b, ii, black arrow) [51,52], suggesting that chronic degradation of collagen II by MMPs may be an ageing mechanism at play.

The lateral-specific, significant rise in peptide yield seen on the N-terminal end of aged COL5A1 compared to young (green bracket in Figure 5a, ii) corresponded partly to its laminin G-like (LNS) domain [57]. Although capable of matriglycan interactions [58], the function of this domain specifically within collagen V fibrils remains elusive. Here, we show clear evidence that the binding availability of the LNS domain is significantly higher exclusively in left lateral aged COL5A1 than for all other groups (young, aged and posterior, anterior). For COL2A1, these aged, lateral-specific increases in peptide yield (Figure 5b) corresponded directly to the interface between the N-terminal propeptide and triple helical region, where the telopeptide is located. Collectively these age-dependant observations in COL5A1 and COL2A1 may indicate the degradation of their respective collagen V and II fibrils that is exclusive to the lateral regions of the AF.

The heightened peptide yield observed near the N-terminal end of aged posterior and anterior aggrecan compared to young, which was reversed (lower) for aged left lateral, coincides with the immunoglobulin fold of its globular G1-A domain (Figure 5c, ii), [33]. The A subdomain in particular is critical for stabilising and enhancing the binding of its adjacent B subdomains to hyaluronan [59]. This interaction is crucial for the anchorage of aggrecan to collagen fibrils in articular cartilage [60] (Figure S4). Several ageing mechanisms have been suggested for age-related changes previously observed in aggrecan structure. Interestingly, Sivan *et al*. proposed that, as a consequence of elevated proteases in ageing discs, proteolysis on the N-terminal end of aggrecan results in only the G1 domain remaining bound to the hyaluronan, which leads to the accumulation of unbound aggrecan fragments in the NP [33]. Additionally, aged aggrecan is known to accumulate AGEs on its lysine residues *in vivo* [37] and that lysine residues within the G1 domain are critical for its hyaluronan interaction [33]. Collectively, these insights highlight the functional importance of this N-terminal protein region in aggrecan and may provide clues for the mechanisms (e.g. proteolysis or glycation) behind the tissue region-specific differences in peptide yield observed within this protein region between aged and young.

It is important to note that these displays of tissue region-specific modifications to protein structures were not limited to ECM proteins. For instance, similar peptide pattern differences were observed within the immune regulator complement factor H (CFH) (Figure S7), which also exhibited anterior inner AF-specific regional difference between young and aged, on the C-terminal side of its primary structure, and also tissue region-wide differences near its N-terminus. CFH plays a key role in regulating the progression of inflammation in age-related diseases such as vascular atherosclerosis [61] and macular degeneration [62]. The structural modifications observed here suggest that similar CFH-driven inflammation pathways may be involved in regions of the inner AF.

Collectively these observations are indicative not only of regional downstream consequences of ageing to major ECM components, on a molecular level, but also of ageing mechanisms acting exclusively within specific IVD regions.

3. Materials and Methods

3.1. Dataset Collection

Young and aged IVD LC-MS/MS datasets were sourced from the Proteomics Identification Database (PRIDE) – Project ID PXD017740. These were originally generated for an earlier study which led to the spatiotemporal IVD proteomic resource, DIPPER - http://www.sbms.hku.hk/dclab/DIPPER/. For detailed methods of human tissue sourcing, sample preparation and MS please refer to this original publication [13]. In brief, the datasets were generated by LC-MS/MS (data-dependant acquisition on an Orbitrap Fusion Lumos Tribrid Mass Spectrometer) from trypsin/LysC-digested whole protein samples extracted from 11 separate regions of IVDs - three discs (L3/4, L4/5 and L5/S1), sourced from two male cadaveric human lumbar spines, one young (16-year-old) and one aged (59-year-old). Prior to LC-MS/MS, peptides were fractionated into four samples per IVD region by high pH reversed phase and separately analysed. For this study, only three out of the eleven IVD regions available for analysis were evaluated by PLF: the anterior, left lateral and posterior inner AF/NPs (Figure 1a).

It is important to highlight that the acquisition of non-degenerated human IVD tissue is challenging and such samples from young and aged individuals are rare. Although we acknowledge that this detailed analysis was performed on one young and one aged individual only, this limitation is outweighed by the merits of the study, since our findings are based on human *in vivo* data, and therefore of direct clinical relevance.

3.2. Peptide Identification

Raw MS spectrum files were converted to Mascot MGF files using RawConverter (Scripps Research Institute; CA, USA) [63] and peak spectra searched against the Uniprot human database (Swiss-Prot and TreEMBL; 2018) [64] using Mascot (Matrix Science, MA, USA). MS/MS ion searches were performed using the following parameters – instrument: ESI-TRAP; database: Swissprot_TreEMBL_2018_01; species: homo sapiens (161,629 protein isoforms); fragment tolerance: 20 ppm; parent tolerance: 6 ppm (monoisotopic); fixed modifications: +58 Da on C (carboxymethyl); variable modifications: +16 Da on M (oxidation); digestion enzyme: trypsin; max missed cleavages: 2; peptide charge: 2+, 3+ and 4+. Mascot search results were exported as DAT files.

Scaffold 5 (Proteome Software; OR, USA) was then used to generate high confidence PSMs by standard LFDR scoring. Mascot DAT files were imported into Scaffold 5 and individual fractionated peptide sample datasets were combined to form a single IVD region sample dataset using MudPIT (Stowers Institute; MO, USA) [65]. Only peptides exclusive to their matched proteins, and with a peptide probability of 95% minimum, were used for PLF analysis, giving a low false discovery rate (FDR) of 0.6% per group (posterior, left lateral and anterior). Peptide FDR was calculated by the PeptideProphet algorithm (WA, USA) [66] within Scaffold 5 using peptide probabilities designated by the Trans-Proteomic Pipeline. CSV files containing peptide lists were exported from Scaffold 5 for PLF analysis (Tables S1 – S3).

3.3. Peptide Location Fingerprinting

Peptide lists were imported into our MPLF webtool (accessible at https://www.manchesterproteome.manchester.ac.uk/#/MPLF) and three separate PLF analyses were performed on young vs. aged for each IVD region (posterior, left lateral and anterior). As previously described [25] (Figure 1b), protein primary sequences were computationally divided into 50 aa-sized segments and peptide sequences, and their spectrum counts, were mapped and summed within each segment. Sequences spanning two adjoining segments were counted in both. Peptide counts per segment were then normalised between aged and young groups based on the median whole protein spectrum count and averaged

per group. Average, normalised peptide counts in each segment for the young group were subtracted from those in the aged group and divided by the segment aa length (50) to reveal regional fluctuations in peptide yield within the primary structures of proteins, as a consequence of age. Average, normalised peptide counts per protein segment were statistically compared between young and aged groups using unpaired, Bonferroni-corrected, repeated measures ANOVAs (Tables S4 – S6). Proteins which were only present in one group (young or aged) were excluded from analysis. For more information on PLF and the MPLF webtool please refer to our earlier publication [25].

3.4. Interaction Network Analysis

Exemplar proteins COL1A2, PCOLCE, COL5A1, COL2A1 and aggrecan were used to construct an experimentally identified interaction network (Figure S4). ECM protein interactions were derived using Matrix DB [67] which were constructed and visualised using Cytoscape 3.4.0 [68].

4. Conclusions

Although ageing and degeneration of IVDs are the result of genetics, environments and their interplays [5], this study focused predominantly on the identification of proteins which may have accumulated modifications to structure over time. This was successfully enabled through the application of PLF to historic LC-MS/MS datasets generated from the inner AF of posterior, lateral and anterior regions of young and aged human IVDs. We demonstrated that, although many of these affected proteins were region-specific, several were identified in all three regions tested. Some of these shared candidates, such as COL1A2 and PCOLCE, exhibited remarkably consistent age-related differences in peptide yields along their protein structures between all tissue regions. This indicates that, in some instances, the same mechanisms of ageing may act on proteins across the entire tissue, leading to similar molecular-level consequences. However, several potential biomarkers identified, such as COL5A1, COL2A1 and aggrecan, also had lateral-specific differences in structure between aged and young, which suggests that some mechanisms acting on proteins may also be exclusive to certain tissue regions. This provides evidence that protein structures are potentially spatially- and age-dependent in the IVD and that PLF is capable of detecting differences within localised tissue regions. This study highlights the merits of PLF as an interrogator of structural changes within proteins and potentially as a powerful tool for biomarker discovery [25] which, in tandem with relative protein quantification (achieved in the DIPPER study [13]), allows for a more complete assessment of tissue proteostasis as a consequence of ageing and the identification of potential new therapeutic targets.

Author Contributions: Conceptualization, A.E., M.J.S. A.T.; methodology, A.E.; software, M.O.; validation, A.E.; formal analysis, A.E.; investigation, A.E.; resources, M.J.S.; data curation, A.E., M.O.;

Institutional Review Board Statement: Ethical review and approval were not required for this study as it was conducted using publicly available datasets from the PRIDE database, already published [13] and ethically approved.

Informed Consent Statement: This study was conducted using publicly available datasets from the PRIDE database already published [13]. Informed consent was obtained as part of that already published study.

Data Availability Statement: The LC-MS/MS raw data was downloaded from the ProteomeXchange Consortium via the PRIDE repository; dataset identifier PXD017740.

Interactive protein schematics showcasing the PLF analysis of these IVD datasets can be viewed via the MPLF webtool - https://www.manchesterproteome.manchester.ac.uk/#/MPLF by selecting the "Location Fingerprinter" tab. Prior to publication, this is accessible to reviewers using the following account – **Login: MICRA_Reviewer, Password : Proteome2021!**

The DIPPER resource [13] is accessible at: http://www.sbms.hku.hk/dclab/DIPPER/.

Conflicts of Interest: The authors have no conflicts of interest. Walgreens Boots Alliance approved the submission of this manuscript but exerted no editorial control.

References

1. Alonso, F.; Hart, D.J. Intervertebral Disk. In; Aminoff, M.J., Daroff, R.B.B.T.-E. of the N.S. (Second E., Eds.; Academic Press: Oxford, 2014; pp. 724–729 ISBN 978-0-12-385158-1.
2. Hoy, D.; March, L.; Brooks, P.; Blyth, F.; Woolf, A.; Bain, C.; Williams, G.; Smith, E.; Vos, T.; Barendregt, J. The global burden of low back pain: estimates from the Global Burden of Disease 2010 study. *Ann. Rheum. Dis.* **2014**, *73*, 968.
3. Mirza, S.K. Intervertebral Disk. In; Aminoff, M.J., Daroff, R.B.B.T.-E. of the N.S., Eds.; Academic Press: New York, 2003; pp. 683–689 ISBN 978-0-12-226870-0.
4. Ohnishi, T.; Novais, E.J.; Risbud, M. V Alterations in ECM signature underscore multiple sub-phenotypes of intervertebral disc degeneration. *Matrix Biol. Plus* **2020**, *6–7*, 100036, doi:https://doi.org/10.1016/j.mbplus.2020.100036.
5. Kepler, C.K.; Ponnappan, R.K.; Tannoury, C.A.; Risbud, M. V; Anderson, D.G. The molecular basis of intervertebral disc degeneration. *Spine J.* **2013**, *13*, 318–330, doi:https://doi.org/10.1016/j.spinee.2012.12.003.
6. Eyre, D.R.; Muir, H. Types I and II collagens in intervertebral disc. Interchanging radial distributions in annulus fibrosus. *Biochem. J.* **1976**, *157*, 267.
7. Vergroesen, P.-P.A.; Kingma, I.; Emanuel, K.S.; Hoogendoorn, R.J.W.; Welting, T.J.; van Royen, B.J.; van Dieën, J.H.; Smit, T.H. Mechanics and biology in intervertebral disc degeneration: a vicious circle. *Osteoarthr. Cartil.* **2015**, *23*, 1057–1070, doi:https://doi.org/10.1016/j.joca.2015.03.028.
8. Wilke, H.-J.; Kienle, A.; Maile, S.; Rasche, V.; Berger-Roscher, N. A new dynamic six degrees of freedom disc-loading simulator allows to provoke disc damage and herniation. *Eur. Spine J.* **2016**, *25*, 1363–1372.
9. Cheung, K.M.C.; Karppinen, J.; Chan, D.; Ho, D.W.H.; Song, Y.-Q.; Sham, P.; Cheah, K.S.E.; Leong, J.C.Y.; Luk, K.D.K. Prevalence and pattern of lumbar magnetic resonance imaging changes in a population study of one thousand forty-three individuals. *Spine (Phila. Pa. 1976).* **2009**, *34*, 934–940.
10. Gruber, H.E.; Hoelscher, G.L.; Ingram, J.A.; Bethea, S.; Zinchenko, N.; Hanley, E.N. Variations in aggrecan localization and gene expression patterns characterize increasing stages of human intervertebral disk degeneration. *Exp. Mol. Pathol.* **2011**, *91*, 534–539, doi:https://doi.org/10.1016/j.yexmp.2011.06.001.
11. Collin, E.C.; Carroll, O.; Kilcoyne, M.; Peroglio, M.; See, E.; Hendig, D.; Alini, M.; Grad, S.; Pandit, A. Ageing affects chondroitin sulfates and their synthetic enzymes in the intervertebral disc. *Signal Transduct. Target. Ther.* **2017**, *2*, 1–8.

12.   Yee, A.; Lam, M.P.Y.; Tam, V.; Chan, W.C.W.; Chu, I.K.; Cheah, K.S.E.; Cheung, K.M.C.; Chan, D. Fibrotic-like changes in degenerate human intervertebral discs revealed by quantitative proteomic analysis. *Osteoarthr. Cartil.* **2016**, *24*, 503–513, doi:https://doi.org/10.1016/j.joca.2015.09.020.

13.   Tam, V.; Chen, P.; Yee, A.; Solis, N.; Klein, T.; Kudelko, M.; Sharma, R.; Chan, W.C.W.; Overall, C.M.; Haglund, L. DIPPER, a spatiotemporal proteomics atlas of human intervertebral discs for exploring ageing and degeneration dynamics. *Elife* **2020**, *9*, e64940.

14.   Shapiro, S.D.; Endicott, S.K.; Province, M. a.; Pierce, J. a.; Campbell, E.J. Marked longevity of human lung parenchymal elastic fibers deduced from prevalence of D-aspartate and nuclear weapons-related radiocarbon. *J. Clin. Invest.* **1991**, *87*, 1828–1834, doi:10.1172/JCI115204.

15.   Sivan, S.-S.; Wachtel, E.; Tsitron, E.; Sakkee, N.; van der Ham, F.; DeGroot, J.; Roberts, S.; Maroudas, A. Collagen turnover in normal and degenerate human intervertebral discs as determined by the racemization of aspartic acid. *J. Biol. Chem.* **2008**, *283*, 8796.

16.   Gruber, H.E.; Hanley, E.N. Observations on morphologic changes in the aging and degenerating human disc: secondary collagen alterations. *BMC Musculoskelet. Disord.* **2002**, *3*, 9.

17.   Craddock, R.J.; Hodson, N.W.; Ozols, M.; Shearer, T.; Hoyland, J.A.; Sherratt, M.J. Extracellular matrix fragmentation in young, healthy cartilaginous tissues. *Eur. Cells Mater.* **2018**, *35*, 34.

18.   Xu, H.; Mei, Q.; Xu, B.; Liu, G.; Zhao, J. Expression of matrix metalloproteinases is positively related to the severity of disc degeneration and growing age in the East Asian lumbar disc herniation patients. *Cell Biochem. Biophys.* **2014**, *70*, 1219–1225.

19.   Hoy, R.C.; D'Erminio, D.N.; Krishnamoorthy, D.; Natelson, D.M.; Laudier, D.M.; Illien-Jünger, S.; Iatridis, J.C. Advanced glycation end products cause RAGE-dependent annulus fibrosus collagen disruption and loss identified using in situ second harmonic generation imaging in mice intervertebral disk in vivo and in organ culture models. *JOR spine* **2020**, e1126.

20.   Alkhatib, B.; Gawri, R.; önnerfjord, P.; Ouellet, J.; Roughley, P.; Steffen, T.; Haglund, L.A. Chondroadherin Fragmentation as a Biochemical Marker for Early Stage Disk Degeneration. *Glob. Spine J.* **2012**, *2*, s-0032.

21.   Feng, C.; Yang, M.; Lan, M.; Liu, C.; Zhang, Y.; Huang, B.; Liu, H.; Zhou, Y. ROS: crucial intermediators in the pathogenesis of intervertebral disc degeneration. *Oxid. Med. Cell. Longev.* **2017**, *2017*.

22.   Babu, N.S.; Krishnan, S.; Swamy, C.V.B.; Subbaiah, G.P.V.; Reddy, A.V.G.; Idris, M.M. Quantitative proteomic analysis of normal and degenerated human intervertebral disc. *Spine J.* **2016**, *16*, 989.

23.   Rajasekaran, S.; Tangavel, C.; Soundararajan, D.C.R.; Nayagam, S.M.; Matchado, M.S.; Muthurajan, R.; Anand, K.S.S.V.; Rajendran, S.; Shetty, A.P.; Kanna, R.M. Proteomic Signatures of Healthy Intervertebral Discs From Organ Donors: A Comparison With Previous Studies on Discs From Scoliosis, Animals, and Trauma. *Neurospine* **2020**, *17*, 426.

24.   Ye, D.; Liang, W.; Dai, L.; Zhou, L.; Yao, Y.; Zhong, X.; Chen, H.; Xu, J. Comparative and quantitative proteomic analysis of normal and degenerated human annulus fibrosus cells. *Clin. Exp. Pharmacol. Physiol.* **2015**, *42*, 530–536.

25.   Ozols, M.; Eckersley, A.; Mellody, K.T.; Mallikarjun, V.; Warwood, S.; O'Cualain, R.; Knight, D.; Watson, R.E.B.; Griffiths, C.E.M.; Swift, J.; et al. Peptide location fingerprinting reveals modification-associated biomarker candidates   of ageing in human tissue proteomes. *Aging Cell* **2021**, e13355, doi:10.1111/acel.13355.

26.   Eckersley, A.; Ozols, M.; O'Cualain, R.; Keevill, E.J.; Foster, A.; Pilkington, S.; Knight, D.; Griffiths, C.E.M.; Watson, R.E.B.; Sherratt, M.J. Proteomic fingerprints of damage in extracellular matrix assemblies. *Matrix Biol. Plus* **2020**, *5*, 100027, doi:10.1016/j.mbplus.2020.100027.

27.   Eckersley, A.; Mellody, K.T.; Pilkington, S.; Griffiths, C.E.M.; Watson, R.E.B.; O'Cualain, R.; Baldock, C.; Knight, D.; Sherratt, M.J. Structural and compositional diversity of fibrillin microfibrils in human tissues. *J. Biol. Chem.* **2018**, *293*, 5117, doi:10.1074/jbc.RA117.001483.

28.   Tsuruha, J.; Masuko-Hongo, K.; Kato, T.; Sakata, M.; Nakamura, H.; Nishioka, K. Implication of cartilage intermediate layer protein in cartilage destruction in subsets of patients with osteoarthritis and rheumatoid arthritis. *Arthritis Rheum.* **2001**, *44*, 838–845.

29.   Bernardo, B.C.; Belluoccio, D.; Rowley, L.; Little, C.B.; Hansen, U.; Bateman, J.F. Cartilage intermediate layer protein 2 (CILP-2) is expressed in articular and meniscal cartilage and down-regulated in experimental osteoarthritis. J. Biol. Chem. 20. *J. Biol. Chem.* **2011**, *286*, 37758–37767.

30.   Bornstein, P.; Agah, A.; Kyriakides, T.R. The role of thrombospondins 1 and 2 in the regulation of cell–matrix interactions, collagen fibril formation, and the response to injury. *Int. J. Biochem. Cell Biol.* **2004**, *36*, 1115–1125.

31.   Subramanian, A.; Schilling, T.F. Thrombospondin-4 controls matrix assembly during development and repair of myotendinous junctions. *Elife* **2014**, *3*, e02372.

32.   Yuan, B.; Ji, W.; Fan, B.; Zhang, B.; Zhao, Y.; Li, J. Association analysis between thrombospondin-2 gene polymorphisms and intervertebral disc degeneration in a Chinese Han population. *Medicine (Baltimore).* **2018**, *97*.

33.   Sivan, S.S.; Wachtel, E.; Roughley, P. Structure, function, aging and turnover of aggrecan in the intervertebral disc. *Biochim. Biophys. Acta - Gen. Subj.* **2014**, *1840*, 3181–3189, doi:https://doi.org/10.1016/j.bbagen.2014.07.013.

34.   Sherratt, M.J. Tissue elasticity and the ageing elastic fibre. *Age* **2009**, *31*, 305–325, doi:10.1007/s11357-009-9103-6.

35.   Watson, R.E.B.; Griffiths, C.E.M.; Craven, N.M.; Shuttleworth, C.A.; Kielty, C.M. Fibrillin-rich microfibrils are reduced in photoaged skin. Distribution at the dermal-epidermal junction. *J. Invest. Dermatol.* **1999**, *112*, 782–787, doi:10.1046/j.1523-1747.1999.00562.x.

36. Lovell, C.R.; Smolenski, K.A.; Duance, V.C.; Light, N.D.; Young, S.; Dyson, M. Type I and III collagen content and fibre distribution in normal human skin during ageing. *Br. J. Dermatol.* **1987**, *117*, 419–428.

37. Sivan, S.S.; Tsitron, E.; Wachtel, E.; Roughley, P.; Sakkee, N.; Van Der Ham, F.; Degroot, J.; Maroudas, A. Age-related accumulation of pentosidine in aggrecan and collagen from normal and degenerate human intervertebral discs. *Biochem. J.* **2006**, *399*, 29–35.

38. Urabe, G.; Hoshina, K.; Shimanuki, T.; Nishimori, Y.; Miyata, T.; Deguchi, J. Structural analysis of adventitial collagen to feature aging and aneurysm formation in human aorta. *J. Vasc. Surg.* **2016**, *63*, 1341–1350.

39. Vadon-Le Goff, S.; Kronenberg, D.; Bourhis, J.-M.; Bijakowski, C.; Raynal, N.; Ruggiero, F.; Farndale, R.W.; Stöcker, W.; Hulmes, D.J.S.; Moali, C. Procollagen C-proteinase Enhancer Stimulates Procollagen Processing by Binding to the C-propeptide Region Only*. *J. Biol. Chem.* **2011**, *286*, 38932–38938, doi:https://doi.org/10.1074/jbc.M111.274944.

40. Lagoutte, P.; Bettler, E.; Vadon-Le Goff, S.; Moali, C. Procollagen C-proteinase enhancer-1 (PCPE-1), a potential biomarker and therapeutic target for fibrosis. *Matrix Biol. Plus* **2021**, 100062, doi:https://doi.org/10.1016/j.mbplus.2021.100062.

41. Canty, E.G.; Kadler, K.E. Procollagen trafficking, processing and fibrillogenesis. *J. Cell Sci.* **2005**, *118*, 1341–1353.

42. Khoshnoodi, J.; Cartailler, J.-P.; Alvares, K.; Veis, A.; Hudson, B.G. Molecular Recognition in the Assembly of Collagens: Terminal Noncollagenous Domains Are Key Recognition Modules in the Formation of Triple Helical Protomers*. *J. Biol. Chem.* **2006**, *281*, 38117–38121, doi:https://doi.org/10.1074/jbc.R600025200.

43. Bekhouche, M.; Kronenberg, D.; Vadon-Le Goff, S.; Bijakowski, C.; Lim, N.H.; Font, B.; Kessler, E.; Colige, A.; Nagase, H.; Murphy, G.; et al. Role of the Netrin-like Domain of Procollagen C-Proteinase Enhancer-1 in the Control of Metalloproteinase Activity*. *J. Biol. Chem.* **2010**, *285*, 15950–15959, doi:https://doi.org/10.1074/jbc.M109.086447.

44. Heinemeier, K.M.; Schjerling, P.; Øhlenschlæger, T.F.; Eismark, C.; Olsen, J.; Kjær, M. Carbon-14 bomb pulse dating shows that tendinopathy is preceded by years of abnormally high collagen turnover. *FASEB J.* **2018**, *32*, 4763–4775.

45. Heinemeier, K.M.; Schjerling, P.; Heinemeier, J.; Magnusson, S.P.; Kjaer, M. Lack of tissue renewal in human adult Achilles tendon is revealed by nuclear bomb 14C. *FASEB J.* **2013**, *27*, 2074–2079.

46. Chang, J.; Garva, R.; Pickard, A.; Yeung, C.Y.C.; Mallikarjun, V.; Swift, J.; Holmes, D.F.; Calverley, B.; Lu, Y.; Adamson, A.; et al. Circadian control of the secretory pathway maintains collagen homeostasis. *Nat. Cell Biol.* **2020**, *22*, 74–86, doi:10.1038/s41556-019-0441-z.

47. Song, L. Chapter One - Calcium and Bone Metabolism Indices. In; Makowski, G.S.B.T.-A. in C.C., Ed.; Elsevier, **2017**; Vol. 82, pp. 1–46 ISBN 0065-2423.

48. Christgau, S.; Rosenquist, C.; Alexandersen, P.; Bjarnason, N.H.; Ravn, P.; Fledelius, C.; Herling, C.; Qvist, P.; Christiansen, C. Clinical evaluation of the Serum CrossLaps One Step ELISA, a new assay measuring the serum concentration of bone-derived degradation products of type I collagen C-telopeptides. *Clin. Chem.* **1998**, *44*, 2290–2300.

49. Packard, B.Z.; Artym, V. V; Komoriya, A.; Yamada, K.M. Direct visualization of protease activity on cells migrating in three-dimensions. *Matrix Biol.* **2009**, *28*, 3–10, doi:https://doi.org/10.1016/j.matbio.2008.10.001.

50. Wang, W.-J.; Yu, X.-H.; Wang, C.; Yang, W.; He, W.-S.; Zhang, S.-J.; Yan, Y.-G.; Zhang, J. MMPs and ADAMTSs in intervertebral disc degeneration. *Clin. Chim. Acta* **2015**, *448*, 238–246, doi:https://doi.org/10.1016/j.cca.2015.06.023.

51. Chung, L.; Dinakarpandian, D.; Yoshida, N.; Lauer-Fields, J.L.; Fields, G.B.; Visse, R.; Nagase, H. Collagenase unwinds triple-helical collagen prior to peptide bond hydrolysis. *EMBO J.* **2004**, *23*, 3020–3030.

52. Laronha, H.; Caldeira, J. Structure and function of human matrix metalloproteinases. *Cells* **2020**, *9*, 1076.

53. Weiss, T.; Brusel, M.; Rousselle, P.; Kessler, E. The NTR domain of procollagen C-proteinase enhancer-1 (PCPE-1) mediates PCPE-1 binding to syndecans-1,-2 and-4 as well as fibronectin. *Int. J. Biochem. Cell Biol.* **2014**, *57*, 45–53.

54. Salza, R.; Peysselon, F.; Chautard, E.; Faye, C.; Moschcovich, L.; Weiss, T.; Perrin-Cocon, L.; Lotteau, V.; Kessler, E.; Ricard-Blum, S. Extended interaction network of procollagen C-proteinase enhancer-1 in the extracellular matrix. *Biochem. J.* **2014**, *457*, 137–149.

55. Symoens, S.; Renard, M.; Bonod-Bidaud, C.; Syx, D.; Vaganay, E.; Malfait, F.; Ricard-Blum, S.; Kessler, E.; Van Laer, L.; Coucke, P. Identification of binding partners interacting with the α1-N-propeptide of type V collagen. *Biochem. J.* **2011**, *433*, 371–381.

56. Ozols, M.; Eckersley, A.; Platt, C.I.; Stewart-McGuinness, C.; Hibbert, S.A.; Revote, J.; Li, F.; Griffiths, C.E.M.; Watson, R.E.B.; Song, J. Predicting Proteolysis in Complex Proteomes Using Deep Learning. *Int. J. Mol. Sci.* **2021**, *22*, 3071.

57. Fallahi, A.; Kroll, B.; Warner, L.R.; Oxford, R.J.; Irwin, K.M.; Mercer, L.M.; Shadle, S.E.; Oxford, J.T. Structural model of the amino propeptide of collagen XI α1 chain with similarity to the LNS domains. *Protein Sci.* **2005**, *14*, 1526–1537.

58. Hohenester, E. Laminin G-like domains: dystroglycan-specific lectins. *Curr. Opin. Struct. Biol.* **2019**, *56*, 56–63.

59. Watanabe, H.; Cheung, S.C.; Itano, N.; Kimata, K.; Yamada, Y. Identification of hyaluronan-binding domains of aggrecan. *J. Biol. Chem.* **1997**, *272*, 28057–28065.

60. Kiani, C.; Liwen, C.; Wu, Y.J.; Albert, J.Y.E.E.; Burton, B.Y. Structure and function of aggrecan. *Cell Res.* **2002**, *12*, 19–32.

61. Calippe, B.; Augustin, S.; Beguier, F.; Charles-Messance, H.; Poupel, L.; Conart, J.-B.; Hu, S.J.; Lavalette, S.; Fauvet, A.; Rayes, J. Complement factor H inhibits CD47-mediated resolution of inflammation. *Immunity* **2017**, *46*, 261–272.

62. Klein, R.; Knudtson, M.D.; Klein, B.E.K.; Wong, T.Y.; Cotch, M.F.; Liu, K.; Cheng, C.Y.; Burke, G.L.; Saad, M.F.; Jacobs Jr, D.R. Inflammation, complement factor h, and age-related macular degeneration: the Multi-ethnic Study of Atherosclerosis. *Ophthalmology* **2008**, *115*, 1742–1749.

63. He, L.; Diedrich, J.; Chu, Y.-Y.; Yates III, J.R. Extracting accurate precursor information for tandem mass spectra by RawConverter. *Anal. Chem.* **2015**, *87*, 11361–11367.

64.   Consortium, U. UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* **2016**, *45*, D158–D169.

65.   Washburn, M.P.; Wolters, D.; Yates, J.R. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* **2001**, *19*, 242–247.

66.   Keller, A.; Nesvizhskii, A.I.; Kolker, E.; Aebersold, R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* **2002**, *74*, 5383–5392.

67.   Launay, G.; Salza, R.; Multedo, D.; Thierry-Mieg, N.; Ricard-Blum, S. MatrixDB, the extracellular matrix interaction database: updated content, a new navigator and expanded functionalities. *Nucleic Acids Res.* **2015**, *43*, D321–D327.

68.   Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **2003**, *13*, 2498–2504.