Peptide location fingerprinting identifies species- and tissue-conserved structural remodelling of proteins as a consequence of ageing and disease

Alexander Eckersleya, Matiss Ozolsa,c,d, Peikai Chen, Vivian Tam, Liam J. Ward, Judith A. Hoylanda, Andrew Traffordb, Xi-Ming Yuan, Herbert B. Schiller, Danny Chanf and Michael J. Sherratta

a - Division of Cell Matrix Biology & Regenerative Medicine, School of Biological Science, Faculty of Biology, Medicine and Health, The University of Manchester, Manchester, United Kingdom
b - Division of Cardiovascular Sciences, School of Biological Science, Faculty of Biology, Medicine and Health, The University of Manchester, Manchester, United Kingdom
c - Department of Human Genetics, Wellcome Sanger Institute, Genome Campus, Hinxton, United Kingdom
d - British Heart Foundation Centre of Research Excellence, University of Cambridge, Cambridge, United Kingdom
e - Department of Orthopaedics and Traumatology, The University of Hong Kong-Shenzhen Hospital (HKU-SZH), Shenzhen, Guangdong 518053, China
f - School of Biomedical Sciences, The University of Hong Kong, Hong Kong, China
g - Department of Clinical Science, Intervention and Technology, Karolinska Institutet, Stockholm, Sweden
h - Department of Forensic Genetics and Forensic Toxicology, National Board of Forensic Medicine, Linköping, Sweden
i - Department of Forensic Genetics and Forensic Toxicology, National Board of Forensic Medicine, Linköping, Sweden
j - Institute of Lung Health and Immunity and Comprehensive Pneumology Center, Helmholtz Munich, Member of the German Center for Lung Research (DZL), Munich, Germany

Corresponding authors. alexander.eckersley@manchester.ac.uk, michael.j.sherratt@manchester.ac.uk.
https://doi.org/10.1016/j.matbio.2022.05.007

Abstract

Extracellular matrices (ECMs) in the intervertebral disc (IVD), lung and artery are thought to undergo age-dependent accumulation of damage by chronic exposure to mechanisms such as reactive oxygen species, proteases and glycation. It is unknown whether this damage accumulation is species-dependent (via differing lifespans and hence cumulative exposures) or whether it can influence the progression of age-related diseases such as atherosclerosis. Peptide location fingerprinting (PLF) is a new proteomic analysis method, capable of the non-targeted identification of structure-associated changes within proteins. Here we applied PLF to publicly available ageing human IVD (outer annulus fibrosus), ageing mouse lung and human arterial atherosclerosis datasets and bioinformatically identified novel target proteins alongside common age-associated differences within protein structures which were conserved between three ECM-rich organs, two species, three IVD tissue regions, sexes and in an age-related disease. We identify peptide yield differences across protein structures which coincide with biological regions, potentially reflecting the functional consequences of ageing or atherosclerosis for macromolecular assemblies (collagen VI), enzyme/inhibitor activity (alpha-2 macroglobulin), activation states (complement C3) and interaction states (laminins, perlecan, fibronectin, filamin-A, collagen XIV and apolipoprotein-B). Furthermore, we show that alpha-2 macroglobulin and collagen XIV exhibit possible shared structural consequences in IVD ageing and arterial atherosclerosis, providing novel links between an age-related disease and intrinsic ageing. Crucially, we also demonstrate that fibronectin, laminin beta chains and filamin-A all exhibit conserved age-associated structural differences between mouse lung and human IVD, providing evidence that ECM, and their associating proteins, may be...
Introduction

In contrast to the dynamic intracellular environment where proteins are replaced hourly or daily, extracellular matrix (ECM)-rich tissues are susceptible to a unique form of ageing which is governed by the longevity of their constituent extracellular proteins. For instance, collagens in the intervertebral disc (IVD) [1] and elastic fibres in lung [2] remain present throughout the human lifespan. As a consequence of this extremely low turnover, many ECM components in tissues such as skin, tendon, lung, IVD and artery are thought to accumulate damage over time, via chronic exposure to a variety of mechanisms, leading to age-related remodelling and disease [3]. These include reactive oxygen species (ROS), markers of which are known to be upregulated in aged IVD [4,5], artery [6] and lung [7]. ROS can directly oxidise sensitive amino acid residues (Cys, Met, Trp, Tyr and His) leading to changes in higher order structure [8,9]. Damage to ECM components can also be incurred by proteolytic degradation and fragmentation through chronic exposure to proteases [10,11], which are often elevated in aged and degenerate tissues; as seen for matrix metalloproteinases (MMPs) 1 and 2 in both aged IVD [12] and artery [13]. Glycation of proteins and cross-linking of long-lived ECM components by advanced glycation end products (AGEs) in ageing connective tissues like IVD [14], lung [15] and artery [16] can also affect the stiffness, accessibility and interactivity of matrix proteins with age. Although the mechanical forces involved may differ between these tissues, repeated mechanical deformation can also lead to ECM fragmentation [10,17,18].

Despite evidence of these ECM-specific ageing mechanisms in the connective tissues of human and model organisms, it is unknown whether their effects are species-dependant, because of differing lifespans. For instance, with an approximate lifespan of only 2.5 years, mice have long been used as a biological model for human ageing [19]. Therefore, the determination of whether the matrisome ages comparably within these species is a relevant question in ageing research. Furthermore, it is also unknown whether the gradual accumulation of damage to long-lived ECM components can influence the progression of age-related diseases, some of which, such as arterial atherosclerosis and lung chronic obstructive pulmonary disease (COPD), have been attributed to an accelerated form of tissue ageing [20,21]. Therefore, the identification of age-susceptible proteins that are conserved between species, organs and with age related diseases, and the characterisation of shared mechanisms and functional consequences, is crucial for the understanding of connective tissue ageing and the discovery of new therapeutic targets.

The age-dependant degeneration of ECM architectures, such as for elastic and collagen fibres, is well documented in tissues such as the IVD, lung, aorta and skin [22–25]. However, time-dependant changes to higher order molecular structure, and their downstream consequences, are challenging to identify as they may be independent of changes in protein transcription, abundance, or architecture. Liquid chromatography tandem mass spectrometry (LC-MS/MS) has been used previously to investigate targeted post-translational modifications to protein structures [26]. However, due to the irregular and often non-specific nature of age-associated, variable modifications, accrued over time, this approach is ill-suited for the protein structure-related study of ageing and disease in ECM-rich connective tissues.

Peptide location fingerprinting (PLF) is a new and emerging proteomic analysis method, capable of the non-targeted and unbiased identification of structure-associated changes within proteins by mapping and quantifying LC-MS/MS-detected tryptic peptides within specific protein regions [27–29]. This approach takes advantage of differences in the regional digestibility of proteins, because of solubility, stability and enzyme susceptibility, to measure characteristic peptide patterns (fingerprints) across their higher order structures. Therefore, by comparing LC-MS/MS datasets, PLF can identify significant fluctuations in peptide yield patterns across protein structure as a consequence of ageing (from whole tissue protein extractions [27,28]), induced damage [29] and even tissue diversity [30] (from microfibril-enriched samples). These identified changes in structure can be reflective of damage modifications or changes in protein conformation, interaction states, activation or synthesis. Due to their insolubility (resultant from highly cross-linked networks), PLF is particularly effective at measuring structure-associated changes within ECM assemblies. For instance, we have previously used PLF to reliably identify ultraviolet radiation-induced, regional damage within ECM proteins as an in vitro model of skin photoageing [29], as well as tissue-specific differences between eye- (ciliary body) and skin-derived fibrillin-1 [30].
Recently, through the development of the Manchester PLF webtool (MPLF), we applied PLF as a proteomic biomarker discovery tool to both self-generated human LC-MS/MS datasets of skin photoageing [27] and to historical tendon [27] and IVD [28] ageing datasets, sourced from a public database (PRIDE). In these studies, we identified potential photoageing- and age-related proteins which were independent of changes in whole protein abundance, demonstrating that PLF identifies structure-associated differences that are unique to the methodology. Most crucially however, we also showed that regional peptide yield differences within these proteins could be correlated to specific biological mechanisms, such as increased collagen I, II and V synthesis in young IVD compared to aged (due to the higher presence of peptides from C-terminal propeptide regions), and increased collagen I and II degradation in aged IVD compared to young (due to the higher presence of peptides from regions downstream of their prominent MMP cleavage sites) [28].

It is clear that PLF can be used as a bioinformatic, proteomic screening tool to interrogate structure-associated changes to ageing ECM proteins in multiple tissues, using historical label-free LC-MS/MS datasets [27,28]. In this study, we aim to apply this in silico PLF analysis to three previously published datasets representative of ECM-rich connective tissue ageing (mouse lung [31] and human IVD outer annulus fibrosus [32]) and age-related disease (human arterial atherosclerosis [33,34]) (Fig. 1). Datasets from IVD, lung and artery were chosen as these organs all undergo age-related remodelling [31,35–38] which can severely impact their function and in turn morbidity [39,40] or mortality [41,42] for the host organism. Although there are some differences in their respective ECM proteomes (matrisomes), many extracellular assemblies, including fibrillar collagens, elastic fibres and proteoglycans, are common not only to these three organs but to other ECM-rich tissues such as skin and tendon. As such, findings of this study will likely be relevant to ECM ageing in health and disease across multiple organs and connective tissues. Although these ageing and atherosclerosis datasets were previously used to investigate global differences in proteome composition and specific differences in relative protein abundance, modification-related differences to ECM protein structures were not interrogated. This provides us with a unique opportunity to apply PLF not only to bioinformatically identify novel protein targets, but also crucial evidence of age-associated differences to structure which are conserved between human and mouse, three ECM-rich organs and in an age-related disease.

### Results and discussion

#### PLF reveals age-affected proteins in IVD outer annulus fibrosus exhibiting tissue region-conserved and -specific fluctuations in peptide yield within protein structures

The human ageing IVD dataset was previously used in the development of the spatiotemporal proteomic analysis platform: DIPPER (http://www.sbms.hku.hk/dclab/DIPPER/). This includes the detailed analysis of 11 spatially resolved tissue regions (posterior, lateral, anterior and central portions of inner and outer AF, and nucleus pulposus) of three discs (L3/4, L4/5 and L5/S1) sourced from two cadaveric males (aged 16 and 59 years respectively) [32]. This formerly published study revealed a profound shift in proteomic composition and abundance along the anteroposterior and lateral axes of aged disc, with the inner and outer AF regions converging. The inner AF portions of this dataset were previously analysed by PLF [28]; by comparing proteomes separately derived from the posterior, anterior and lateral regions, we were able to show tissue region-conserved and -specific differences in peptide yield patterns within ageing ECM proteins such as aggrecan and collagen I, II, and V alpha chains [28]. Here, we chose to focus on the outer AF (OAF), which previously showed a clear shift of ECM composition between aged and young [32], and is therefore most likely to harbour structure-associated evidence of ageing in long-lived matrix proteins. Since IVD ageing and degeneration is profoundly regional, we stratified this comparison between anterior, posterior and left lateral regions as before, to reveal tissue location-specific similarities and differences. Young and aged discs were compared: three vs three discs for lateral and anterior and three vs two discs for posterior (young posterior L4/5 dataset omitted due to insufficient peptide identification for PLF).

Peptides corresponding to 681 proteins in anterior, 805 in left lateral and 437 in posterior were identified in both young and aged disc OAFs by MS/MS ion searches (Fig. S1; peptide lists: Tables S1 – S3). Principal component analyses (PCAs) of peptide spectral counts demonstrated good separation of data between young and aged groups for all three tissue regions, with young discs clustering in both anterior and left lateral comparisons (Fig. S2). PLF analysis led to the identification of 284 proteins in total (across all tissue regions) with protein regions exhibiting significantly different peptide yields between aged and young discs, 26 of which were shared between all three tissue regions (Fig. S3). When compared to a list of proteins which were significantly different in relative abundance between young and aged OAF, previously determined by
Fig. 1. Experimental workflow and methodology for the application of PLF to ageing and age-related disease in three distinct tissues. Raw LC-MS/MS datasets derived from mouse lung [31], human IVD outer annulus fibrosus (AF) [32], and human atherosclerotic artery [33,34] (A) were downloaded from the Proteomics Identification Database (PRIDE) repository. After peptide identification, PLF was used to compare peptide yield across protein structures between young and aged groups for mouse lung and human IVD and between atherosclerotic plaque and internal control artery to reveal proteins with age- or age-related disease-dependant structural differences. Additionally, these analyses were stratified by tissue location for IVD (posterior, left lateral and anterior) and by sex for artery to expose region- and sex-related similarities.
Peptide location fingerprinting identifies species- and tissue differences. Finally, peptide yield differences across affected proteins, commonly identified between datasets, were cross compared to reveal matching patterns of ageing and disease and to expose potential evidence of conserved mechanisms of change with age. Four ECM proteins in particular exhibited clear, profound tissue region-conserved differences in peptide yield between young and aged, with patterns that are indicative of unique consequences of ageing. These are the alpha-3 chain of the microfibrillar collagen VI (COL6A3), the alpha-2 chain of the fibrillar collagen I (COL1A2) and the ECM remodelling cartilage intermediate layer proteins (CILP)−1 and −2 (Fig. 3).

Collagen VI exists as a multifunctional microfibril composed of trimeric alpha chains (mainly alphas-1,−2 and −3) that regulates tissue homeostasis [44] through its extended network with other ECM components [45] and cell interactions [46]. Previous analysis of this dataset showed that alpha chains-1,−2 and −3 all had significantly lower relative abundances in aged IVD OAF compared to young, indicating a reduction of collagen VI with age [32]. Here, we observed two specific regions of COL6A3 which displayed significantly higher peptide yields in posterior, lateral and anterior tissue regions of young OAF than in aged. These corresponded directly to the VWFA domains 10 and 11 (Fig. 3A), which are located within the globular double beaded region of the collagen VI microfibril and directly flank the interbead triple helical region [44]. Since collagens are known to be long-lived [1], it is possible that collagen VI may have accrued structural damage (e.g. from ROS, proteases or glycation) over time which may be reflected in the more globular regions of its microfibrillar ultrastructure. Collagen VI is particularly sensitive to MMPs-2 and −9, of which MMP-2 is upregulated in aged and degenerated IVD [47,48]. Alternatively, the beaded region of the collagen VI microfibril is known to interact with numerous matrix proteins such as the proteoglycans biglycan and decorin and matrilins-1, −3 and −4, which in turn link to fibril forming collagen II and aggrecan to form an extensive ECM network [45]. Differences in peptide yield could reflect a change in the interaction state of aged microfibrils compared to young, which may affect the availability of interactive regions in the bead to trypsin digestion with functional consequences to known cell-collagen VI communication in the OAF [49].

The AF is comprised of hierarchical, concentric rings of collagen I fibrils (lamellae), which are thought to be remodelled in ageing [50]. Through PLF analysis of the inner AF component of this dataset, we had previously demonstrated that COL1A2 yields significantly more peptides within the C-terminal propeptide region of young compared to aged discs, and that this pattern was consistent between posterior, lateral and anterior tissue regions [28]. Here, we show that this same pattern exists within the OAF, with COL1A2 yielding significantly higher peptides within two segments, one of which spans this same C-terminal propeptide region (also consistent between posterior, lateral and anterior; Fig. 3B). Both the N- and C-terminal propeptide regions are cleaved from procollagen prior to fibril formation [51,52]. Therefore, the higher abundance of peptides from C-terminal propeptide region may be indicative of heightened collagen synthesis in young AF compared to aged. Previous studies of healthy tendon have shown that collagen is diurnally maintained via a circadian cycle of degradation and synthesis [53]. Should a similar maintenance process exist in IVD, we show evidence of its potential disruption with age both for outer (Fig. 3B) and inner AF [28].
Classification analysis of proteins with structure-associated modifications revealed ECM proteins as the major class affected by age in IVD OAF. PLF analysis identified 162 proteins in anterior, 119 in left lateral and 102 in posterior tissue regions with significant differences in peptide yield across structure (24%, 15% and 23% of total proteins identified in both young and aged, respectively per region; full PLF analysis results: Tables S4 – S6).
CILPs-1 and -2 are monomeric ECM glycoproteins which have been implicated both in osteoarthritis [54,55] and IVD degeneration [56,57]. Although CILP1 expression was shown to increase both in articular cartilage and IVD with age [58,59], we show that structure-associated differences in CILPs-1 and -2 also exist within the AF (Fig. 3C, D). Several segments along the structures of CILPs-1 and -2 displayed either significantly higher or lower peptide yields in aged compared to young. Due to these age-dependent fluctuations, which were markedly conserved between all tissue regions, this perhaps reflects a more global change in the higher order structure of these proteins than seen for COL6A3 and COL1A2 which may affect function.

The observation that these four exemplar matrix proteins (Fig. 3) all displayed remarkable consistency between posterior, lateral and anterior regions of the OAF in their peptide yield patterns is important, as it provides clear evidence of tissue-wide mechanisms and consequences of ECM ageing. Despite this, a subset of proteins displayed peptide yield patterns that were tissue region-unique in parts of their structures but conserved in others (Fig. 4). This is interesting as it may indicate a superimposition of both tissue-wide and region-specific mechanisms or consequences of ageing which are acting upon the same proteins and may be reflected in the peptide yield patterns. These proteins include the alpha-1 chain of the collagen VI (COL6A1), the proteoglycan versican and the immune response-regulating complement C3.

The tissue-region conserved, significantly higher peptide yields observed in the VWFA2 domain of young COL6A1 compared to aged OAF (Fig. 4A) is analogous to that seen in the VWFA11 domain of COL6A3 (Fig. 3A) since both are located in similar positions along their chains, just downstream of the triple helical region. Aged COL6A1 also exhibited a significantly higher peptide yield within the N-terminal VWFA1 domain compared to young, which was specific to the posterior region of the OAF (Fig. 4A). As with COL6A3, both these globular regions of COL6A1 also correspond to the double beaded ultrastructural region of its constituent collagen VI microfibril. The observation that two major alpha chains of collagen VI had structure-dependant differences closely associated to the microfibril bead further corroborates the hypothesis that damage to this long-lived ECM assembly may accrue with age or that the interaction state of aged microfibrils may differ compared to young. Although this appears to be predominantly tissue-wide (manifesting as conserved peptide patterns in all three tissue regions), posterior-specific mechanisms or consequences of ageing may also exist within the collagen VI microfibril.

Versican is a highly interactive proteoglycan, which together with collagen VI, exists as an integral component of translaminar cross bridges in the AF [60]. Previous analysis of this dataset demonstrated a significantly lower relative abundance for versican in aged OAF compared to young [32]. We reveal that two specific 50 aa segments of versican yielded more peptides in young OAF than in aged across all tissue regions, located at the interface between the GAGα and GAGβ regions (Fig. 4B). Versican exists as four different isoforms in tissues with the V0 isoform containing both GAGα and β regions and the V1 and V2 isoforms containing only GAGα and GAGβ regions respectively [61]. The distribution of these isoforms was previously characterised in foetal IVD, where V0 was the predominating isof orm [62], and in aged adult IVD, where the V1 isoform was reported as most prevalent (although the data was not shown) [63]. The higher peptide yields located at the interface between these GAG regions in young may indicate changing distributions of these versican isoforms with age, which may have functional implications. In addition, a lateral-specific difference in peptide yield was seen within the L2 module of versican, which yielded more peptides in aged than in young OAF (Fig. 4B). All isoforms of versican are capable of binding hyaluronan through the L1 and L2 modules [64,65], creating a network capable of regulating cell adhesion [66]. Structure-associate differences within these modules may be indicative of a lateral tissue region-specific disruption between versican and hyaluronan.

Age-associated differences in protein structure, were not limited to ECM proteins as exemplified by complement C3, which has been described as the Swiss-army knife of both innate and adaptive immunity. Its native form goes through a series of activation stages with each product and fragment released playing a unique functional role, including the opsonisation of foreign antigens, amplification of the complement response and immune cell recruitment [67]. PLF revealed significantly higher peptide yields within three segments of the C3 structure in young OAF compared to aged (Fig. 4C), one which was both anterior- and lateral-specific and two unique to
posterior. These segments all fall within the region of the protein’s C3dg degradation product, produced when C3b (the active form of C3) binds to complement receptors, to form C3c which is devoid of the C3dg region [68]. While C3c is released from cell, the C3dg fragment remains bound and exerts a variety of effector functions including immune modulation [67]. The significant differences in peptide

Fig. 3. ECM proteins displaying tissue region-conserved differences in peptide yield across their modular structures between young and aged IVD OAF. Protein primary sequences were segmented into 50 aa-sized segments and LC-MS/MS-detected peptide sequences were quantified within each segment (bar graphs = average, normalised PSMs; missing values [mv] per segment are indicated at y = 0; error bars = standard deviation [SD]). Modular fluctuations in peptide yields across protein structures were assessed by subtracting the average, normalised peptide counts per segment in young from aged (Line graphs = aged-young PSMs/segment aa size [50]; below zero/mv line = higher in young, above zero/mv line = higher in aged; composite line graphs are normalised between all three tissue regions to show maximum and minimum differences in aged compared to young [Δ], see Fig. S5 for non-normalised) and statistically compared between groups (unpaired Bonferroni-corrected, repeated measures ANOVAs: *, p ≤ 0.05; **, p ≤ 0.01; ***, p ≤ 0.001; composite line graphs: stars = significant in all three tissue regions; aa ranges of Uniprot-sourced functional domains and regions are indicated). COL6A3 (A) contained multiple segments within the C-terminal half of the protein which displayed significantly higher peptide yields in young than in aged. These were consistent between posterior, lateral and anterior regions of the IVD OAF and coincided specifically with von Willebrand Factor type A (VWFA) domains 10 and 11. COL1A2 (B) also contained two segments which exhibited higher peptide yields in young than in aged for all three tissue regions and span across both the triple helical and the C-terminal propeptide regions (downstream of a prominent MMP cleavage site). The N-terminal half of CILP1 (C) yielded higher peptides in aged than in young (one significantly different segment coincided with the Ig-like C2-type domain) whereas the C-terminal half yielded more peptide in young than aged. These peptide yield difference patterns were remarkably conserved between all three tissue regions, across the entire modular structure of CILP1. Similarly, this pattern was also tissue-region conserved for CILP2 (D), which contained a segment at the protein centre which yielded significantly higher peptides in aged than in young and two near the C-terminal end which yielded significantly more in young than in aged. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
yields corresponding this fragment may reflect di-
s
tinct activation states of complement C3 between
young and aged OAF, potentially revealing clues to
a changing immune response in ageing. The possi-
bility that PLF could be used to assess the activation
state of proteins in standard label-free LC-MS/MS
datasets opens a potential new avenue for future
study.

In addition to demonstrating differences in struc-
ture which were either conserved or unique to all
three OAF regions tested, the specific peptide yield
patterns displayed by these proteins may serve as
fingerprints of particular mechanisms or consequen-
ces of ageing, as previously shown [28]. Crucially,
this was not limited to ECM proteins alone, as these
fingerprints were also used to investigate a member

![Fig. 4.](image_url)

Proteins displaying both tissue region-conserved and -specific differences in peptide yield across their modular structures between young and aged IVD OAF. LC-MS/MS-detected peptide sequences were quantified within each 50 aa segment (bar graphs = average, normalised PSMs; mv indicated at y = 0; error bars = SD). Differences in peptide yields across protein structures were assessed by subtracting average, normalised PSMs per segment in young from aged (Line graphs = aged-young PSMs/segment length; below zero/mv line = higher in young, above zero/mv line = higher in aged; composite line graphs are normalised between all three tissue regions, see Fig. S6 for non-normalised) and statisti-
cally compared (unpaired Bonferroni-corrected, repeated measures ANOVAs: *, p ≤ 0.05; **, p ≤ 0.01; ***, p ≤ 0.001; composite line graphs: stars = significant in all tissue regions; coloured arrows = significant, tissue region-specific differen-
ties; aa ranges of Uniprot-sourced domains indicated). Two segments coinciding with the VWFA2 domain of COL6A1 (A) displayed higher peptide yields in young than in aged for all three OAF tissue regions. However, two segments near the N-terminal end of the protein (corresponding to VWFA1) exhibited significantly higher peptide yields in aged than in young that were posterior-specific. Similarly, versican (B) contained two segments near the protein centre which also exhibited higher peptide yields in young than in aged and were consistent between all three tissue regions. These corresponded to the interface between glycosaminoglycan (GAG) α and GAGβ regions of the protein. However, one segment near the N-terminal end (corresponding to the link-2 [L2] module) yielded significantly more peptides in aged than in young, exclu-
sively in the left lateral OAF. Three segments within complement C3 (C) had significantly higher peptide yields in young than in aged OAF; one which was both anterior- and left lateral-specific, coinciding with the N-terminal half of the C3dg degradation product, and two which were posterior-specific, at C-terminal half of the same fragment. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
of the complement pathway. Here, we demonstrate that patterns within these proteins, which coincided with functional regions, could provide clues as to their state within a macromolecular assembly (collagen VI), isoform composition (versican) or activation (C3) state within OAF tissue, and how this may change with ageing.

Collagens and basement membrane proteins amongst ECM components most affected by modifications to structure in ageing mouse lung

The mouse lung dataset was previously used for the development of an ageing lung atlas resource, comprised of both single cell transcriptomic and tissue proteomic analyses (https://theislab.github.io/LungAgingAtlas/) [31]. The proteomic analysis consisted of bulk whole lungs sourced from four young and four aged mice (3 and 24 months old). In addition to a global deregulation of protein expression with age, this formerly published study revealed extensive ECM remodelling between young and aged mice including alterations in the relative abundance of fibril-bridging (XIV, XVI), basement membrane (IV) and microfibrillar (VI) collagen alpha chains [31]. Although this dataset consisted of fractionated samples of soluble and insoluble proteins, we chose to focus on the ECM-rich fractions, which previously exhibited the largest separation of aged and young sample clusters by PCA [31] and are therefore most likely to harbour structure-associated evidence of ageing in long-lived matrix proteins.

MS/MS ion searches identified peptides corresponding to 668 proteins in both young and aged mouse lung groups (Fig. S7 A; peptide list: Table S7). PCA of peptides and their associated spectral counts demonstrated good separation of data between young and aged groups, with aged samples forming a more distinctive cluster compared to young samples (Fig. S7 B). PLF analysis identified a total of 140 proteins displaying significant differences in peptide yields across their structures between aged and young groups (full PLF analysis results: Table S8). When compared to a list of proteins which were significant different in relative abundance between young and aged mouse lung, previously determined by label-free quantification [31], PLF identified 134 targets of ageing which were potentially unique to the methodology, particularly for collagens and basement membrane laminins (Fig. S8). Classification analysis of the 140 shortlisted proteins revealed numerous collagens, elastic fibre-associated proteins and basement membrane laminins as the main ECM components structurally affected in ageing mouse lung (Fig. 5). Architectural remodelling of ECM components in aged human lung has long been documented [35], including decreases in elastic fibre presence and an increased basement membrane thickness within alveolar walls. Here, we show for the first time that many proteins pertinent to these assemblies exhibit structure-associated modifications on a molecular level, providing clues as to specific targets of ageing mechanisms.

As performed for ageing human IVD (Figs. 3, 4), specific age-dependant fluctuations in peptide yields across the structures of key lung ECM proteins were related to biological domains or regions in order to interrogate the functional consequences of their unique structural changes in ageing (Fig. 6). These were the alpha 1 chain of collagen I (Col1a1) and four basement membrane component proteins including the alpha 2 chain collagen IV (Col4a2), laminin alpha chains −3 (Lama3) and −5 (Lama5) and heparan sulphate proteoglycan (HSPG) 2 core protein (perlecan).

As the main, long-lived structural ECM scaffold component present in almost all connective tissues, the age-dependant remodelling of collagen I is well documented [69]. In healthy lung however, most ageing studies failed to identify changes to collagen presence (see reviews: [22]) and although the relative proportions of collagen I to III was shown to decrease in aged rats [70], this was not reproduced in mice [71]. In support of this, previous quantification analysis of the dataset analysed here found that the abundances of most lung collagens I and III chains did not change between aged and young mice [31,72]. Despite this, we identified a segment within Col1a1 displaying a significantly higher peptide yield in young than in aged (Fig. 6A), indicating that PLF may provide a more sensitive means of characterising age-associated changes to lung collagen I, which may occur on a more molecular level. This difference was seen in the triple helical domain of mature collagen I, suggesting that these modifications may be as a result of damage accumulation to structure. MMP abundance was previously shown not to change in this dataset [31], and activity of MMP-2 and −9 were shown to decrease with age in rat lungs [73]. Hence, this suggests that the structure-associated differences observed here for Col1a1 in mouse lung may not be due to changes in the expressions of secreted MMPs with age, in contrast to IVD. In support of this, the differences in Col1a1 peptide yield observed did not appear to coincide with its most prominent MMP cleavage site (Fig. 6A, black arrow). The presence of AGEs and related crosslinks in rat lung collagen however was previously shown to increase with age [74], with decreases in collagen solubility reported as a result [75]. As solubility is thought to play central role in the regional digestibility of ECM proteins [27,30], the differences observed in the peptide yield patterns of the Col1a1 structure may instead be as a result of AGE accumulation with ageing.

Previous quantification analysis of this dataset showed higher levels of Col4a3 aged mice
compared to young [31], indicating potential changes to collagen IV presence in ageing lung. Here we reveal that Col4a2 yielded significantly more peptides in young than in aged (Fig. 6B) within a protein region corresponding to canstatin (NC1). Interestingly, this domain is released as a functional fragment (matrikine) during the proteolytic degradation of collagen IV by membrane-anchored (MT) MMPs-14 and /C0 [76,77]. Although previous analysis showed that no change in the levels of secreted MMPs [31,72], the MT-MMPs [78] were not identified. It is possible that changes in MMPs 14 and 15 exist because of ageing which may lead to differences in the levels of this fragment, reflected here. Canstatin has been shown to actively inhibit cell proliferation, apoptosis and angiogenesis is therefore presumed to play a role in tissue homeostasis, with important implications for cancer therapy [77]. As such, decreasing levels of this fragment with age may be an important driver of ageing and age-related diseases in lung.

Levels of Lama3 and Lama4 were recently shown to be significantly decreased in decellularized aged mouse lung compared to young, including a reduction in the expression of both laminin proteins for resident cells which were subsequently cultured on these aged scaffolds [79]. Although previous analysis indicated no significant changes in laminin chain abundance in this lung dataset [31], PLF analysis revealed multiple segments within the coiled-coil region and G domains of Lama3 and Lama5 which exhibited significant differences in peptide yield between aged and young mice (Fig. 6C,D). The laminins exist as trimeric complexes of alpha, beta and gamma chains. These chains wrap around each other in their coiled-coil regions, which then split into three N-terminal finger-like structures capable of interacting with other matrix components. At the C-terminal base of this complex is a region made of five globular (G) domains formed by the alpha chain alone which is capable of binding cell integrins and syndecans [80]. Significant differences in the availability of epitopes within the G domains in particular could be indicative of impaired cell-mediated interactions as a consequence of ageing. Alternatively, several MMPs are capable of cleaving laminin (including the calstatin-producing MMP14 discussed earlier) and some have even been shown to release functional matrikines [81]. As with collagen IV, differences in the structures of Lama3 and Lama5 could...
also be as a result of changes in MT-MMP activity with age.

To our knowledge, changes in perlecan have yet to be reported in healthy lung ageing however, due to its interactions with a multitude of ECM components, cell receptors and even growth factors, this proteoglycan has been implicated in the orchestration of age-related fibrosis in various connective tissues [82]. These many interactions are protein region-specific, with different domains of perlecan exerting unique functional roles. Significant differences in the peptide yield along the structure of this
protein were observed within two domains in particular; L4A (higher in aged) and LG3 (higher in young) (Fig. 6E). The L4-containing region is capable of binding growth factors (e.g. PDGF [83] and FGF7 [84]) to modulate signalling in tissue, whereas the LG-containing region is a more prolific binder of ECM components (e.g. nidogens, fibulins, elastin, collagen VI, matriglycan [85]). Differences in digestibility of domains within these regions between young and aged mice could reflect a change in these interactions because of ageing. Interestingly the LG3 domain can be cleaved and released by BMP1 or cathepsin L, as a fragment known as endorepellin which is capable of modulating cell signalling [86,87]. Significantly more peptides were observed at the perlecan LG3 domain of young lung compared to aged, potentially corresponding to an increased presence of this fragment. Since this matrikine has been implicated as a mediator of fibrosis, age-associated changes in its presence may be a key driver of disease phenotypes in the ageing lung.

Basement membranes are composed of a meshwork of networked collagen IV, laminins and perlecan [88] all of which displayed structure-associated differences as a consequence of ageing (Fig. 6), providing further evidence of potential disruption at the alveolar barrier, the crucial interface between capillaries and alveoli. In addition to potential disruptions in protein interactions (e.g. perlecan), the unique peptide fingerprints displayed by these proteins may enable the identification of specific mechanisms or consequences of ageing, such as changes in specific protease activity (such as MT-MMP cleavage of collagen IV and laminins) and accumulations of AGEs (on collagen I).

Conserved differences within protein regions observed between human IVD and mouse lung ageing

Despite drastically shorter lifespans (~2.5 vs. ~70 years), mouse models are commonly used to characterise progression and treatment of human ageing and disease [19]. A recent study showed that aged mice exhibit typical features of human IVD ageing and degeneration, including loss of disc height and bulging alongside molecular changes and behaviours linked to increased pain compared to young mice [89]. Similarly, another study found that alterations in lung function and micromechanics in ageing mice were similar to that in humans [90]. Despite this, due to the longevity of ECM components which dominate tissues like lung and IVD, it is unclear whether mice can successfully model structure-associated changes to matrix proteins in humans, over many years. To examine whether certain mechanisms or consequences of ageing may be shared between mouse and human proteins, we compared those identified with structure-associated alterations in human IVD ageing with those identified in mouse lung.

Out of a total of 284 age-affected proteins in human IVD OAF (across posterior, lateral and anterior regions) and 140 proteins in mouse lung, 39 were shared between both species (Fig. S9). Classification analysis of these shared potential biomarkers of both human and mouse ageing revealed ECM proteins, mainly collagens and laminins, as the class predominantly affected in both tissues, accounting for ~40% of proteins shared between species (Fig. 7). This reflects an age-associated remodelling of the ECM which exists within the connective tissues of both human and mouse, regardless of differences in lifespan.

Although these 39 shared proteins were identified as targets of ageing in both human IVD and mouse lung, the mechanisms or consequences leading to those structure-associated changes may be unique to each. To determine which proteins may have evidence of the same ageing mechanisms or consequences that are irrespective of species and organs, peptide patterns across structures were compared to reveal protein regions of coinciding peptide yield differences. Of these 39 proteins, only four had regions exhibiting the same age-dependant differences in peptide yields for both human IVD and mouse lung: the adhesive ECM protein fibronectin, the basement membrane laminins beta-1 (Lamb1) and beta-2 (Lamb2) and the actin crosslinker, filamin-A (Fig. 8).

Fibronectin is a fibril-forming ECM glycoprotein, capable of a variety of cell and matrix interaction, which regulate cell adhesion, proliferation and activity [91] and elastic fibre assembly [92]. Aged mouse lungs demonstrate a decreased expression of fibronectin compared to young [79] and the ability of fibroblasts to secrete fibronectin was also shown to wane with age [93]. In contrast, higher levels of fibronectin were previously observed in aged bovine nucleus pulposus compared to young [36] and fragments of fibronectin have been detected in AFs from moderately degenerated IVDs [94]. The soluble form of fibronectin exists as a compact dimer connected at the C-terminus by a disulphide bond. Elongation of the dimer into fibrils is induced by integrin-binding of two central regions of the monomers after which clustering of fibronectin-bound integrins promotes the crosslinking of these fibrils into a stable matrix. Crucially, the significantly lower peptide yields seen consistently within a central region of fibronectin in both aged mouse lung and human IVD, compared to young, coincides specifically with this cell attachment region (Fig. 8A). This indicates that connective tissue ageing both in humans and mice may have consequences to either fibronectin fibrillogenesis or general cell adhesion. Since these cell interactions are crucial to connective tissue homeostasis, these regional changes to fibronectin...
may play a prevalent role in its age-dependant disruption regardless of species.

In addition to two laminin alpha chains (Fig. 6C, D), structure-associated differences were also identified in two beta chain (Fig. 8B, C) from aged mouse lung when compared to young, further evidencing a wider disruption of this basement membrane assembly. In addition, these higher peptide yields, within the N-terminal domain of Lamb1 (Fig. 8B) and within the coiled coil (I) domain of Lamb2 (Fig. 8C), were consistently seen in both aged human anterior IVD and aged mouse lung when compared to young. Although the role of laminins as basement membrane components is well characterised in lung [88], their function in the AF of the IVD is more elusive. They are widely present in the pericellular matrix of the IVD nucleus pulposus, however their distribution in the AF is more limited, especially in adult tissues [95]. Regardless of the unique roles these complexes may play in each tissue, the same protein regions were demonstrably affected in both species, potentially indicating a conserved mechanism or consequence of ageing. The N-terminal domain is critical for the self-assembly of laminins, their interactions with other ECM assemblies and their incorporation into basement membranes [96]. It is possible that the differences seen within the structures of Lamb1 and Lamb2 reflect a shift in their interaction states with similar consequences to both tissues.

Filamin-A is an important crosslinker of orthogonal actin fibrils (F-actin), involved in the dynamic remodelling of the actin cytoskeleton which governs cell motility, locomotion and mechanical resistance [97]. Filamin-A exists as an L-shaped homodimer (linked at the C-terminal end) and contains two actin-binding regions, one near the N-terminus (ABD) another further along the protein within the rod-1 domain [98]. As such, the dimer binds two F-actins at a 90° angle with its rod-1 domains resting flat along the filaments [97]. Similar regions within the rod 2 domain of filamin-A yielded significantly more peptides in young than in aged, for both mouse lung and human anterior IVD (Fig. 8D). Interestingly, this domain

---

**Fig. 7.** Classification analysis of common proteins in mouse lung and human IVD, identified with significant structure-associated modifications as a consequence of ageing. PANTHER classification of the 39 shared proteins identified ECM proteins (e.g. collagen and laminin chains, fibronectin and fibrillin-1) and cytoskeletal proteins (e.g. dynein and spectrin) as the top two classes structurally affected by ageing in both human IVD and mouse lung (classes and their protein identities are displayed; minimum 2 proteins per class; total number of proteins per class indicated in brackets). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
functions to bind integrins, linking the actin cytoskeleton to the cell membrane and in turn to the ECM matrix [99]. As such, differences within this region may reflect a changed state of resident cell interactions with the ECM, which remains consistent between both species and organs. Interestingly, as components of the pericellular matrix, fibronectins and laminins are both integral to cell adhesion while filamin-A plays a role in anchoring the actin cytoskeleton to the outer ECM [99]. This suggests perhaps that the conserved structure-associated differences, seen in both mouse lung and human IVD, are connected to a changing cellular microenvironment during connective tissue aging. To our knowledge, this study is the first to provide evidence that ECM proteins are subjected to similar structure-associated differences in mouse and human ageing, irrespective of their vastly different lifespans and tissue functions. It indicates that universal markers of connective tissue ageing may exist within the ECM and provides clues into mechanisms or consequences which may be conserved and therefore crucial to our understanding of ageing.

Sex-conserved differences identified within protein structures from human atherosclerotic plaque

The human atherosclerotic dataset was previously used in proteomic analyses of distinct arterial lesions and includes label-free nLC-MS/MS of plaque core, fatty streak and internal control tissue.
sourced from ten aged males and females [33,34]. Former study of this dataset revealed sex-conserved differences in the relative abundance of ECM components, such as proteoglycans mimecan and biglycan, which are indicative of ECM remodelling in atherosclerosis [33]. For this study, we chose to investigate plaque core tissue specifically (in comparison to internal control) as it exhibited the largest differences in proteomic composition [33], and therefore disease progression. Since sex-specific differences in abundance were observed previously for this dataset in key players of disease progression (e.g. fibrinogens and apolipoproteins) [33,34], we too stratified this comparison to males and females in order to further investigate similarities between sexes.

Peptides corresponding to 507 proteins in males and 458 proteins in females were identified in both atherosclerotic plaque and control artery by MS/MS peptide ion searches (Fig. S10; peptide lists: Tables S9, S10). PCA of peptides and their associated spectral counts revealed good separation of data between plaque and control artery for both male and female groups (Fig. S11). Male plaque samples in particular formed a distinct cluster in comparison to other groups, suggesting sex-specific homogeneity. PLF analysis led to the identification of 242 proteins across both sexes exhibiting significant differences in peptide yields across structures between plaque and control artery, 63 of which were shared between males and females (Fig. S12). When compared to a list of proteins which were significantly different in relative abundance between control and plaque, previously determined by label-free quantification [34], PLF identified 152 targets of atherosclerosis which were potentially unique to the methodology, particularly for collagens (Fig. S13). Classification analysis of the 242 shortlisted proteins once again identified ECM proteins as the major class structurally modified in arterial atherosclerosis, comprising ~20% of affected proteins shared between male and females (Fig. 9). This included chains from two basement membrane collagens IV and XVIII and the fibrillar collagen I and its bridging collagen XII as well as proteoglycans perlecan and mimecan.

Peptide yield differences across the structures of PLF-identified proteins were further compared between males and females to investigate whether fluctuations between plaque and control arteries were sex-dependent. Additionally, significant differences within key proteins were related to biological domains or regions to reveal potential functional consequences to arterial atherosclerosis. Several key proteins had regions which yielded significantly different peptides in plaque compared to control consistently between males and females. This included proteoglycans perlecan and mimecan, the actin crosslinking alpha actinin-4 (ACTN4) and the LDL-associated apolipoprotein-B (APOB) (Fig. 10).

In vitro, the removal of HSPG was shown to increase binding of low-density lipoproteins (LDLs) within a representative subendothelial matrix [100]. As such, the reduction of perlecan on arterial walls is thought to contribute to the disease progression of atherosclerosis [101]. Here, we show that these changes may not be limited to differences in protein abundance but also structure-associated. The higher peptide yields observed within both male and female plaque-sourced perlecan compared to control were very specific to the L4 B domain of the proteoglycan (Fig. 10A). This pattern is similar to that seen for mouse lung ageing, where higher peptide yields were also seen within the L4 A domain of perlecan in aged compared to young (Fig. 6E). This potentially indicates that the L4-containing regions of this protein may be particularly sensitive to structure-associated changes. This region interacts with the basement membrane ECM protein WARP and collagen VI [85] and is capable of sequestering growth factors [83,84], as mentioned previously. Perhaps structure-associated differences within these L4 regions reflect disturbances in these interactions in both atherosclerosis and ageing, drawing parallels between the two processes.

Mimecan is a keratan sulphate proteoglycan that has been described as an emerging biomarker of atherosclerosis [102]. In human carotid plaques, its presence was recently shown to be downregulated in vascular smooth muscle cells but upregulated in certain plaque regions (close to collagen fibres and neutral lipids) [103]. The same study also showed a positive correlation between plaque levels of mimecan and age, with higher levels predicting future cardiovascular-related death. Previous analysis of mimecan within this proteomics dataset indicated significantly reduced levels in plaque core compared to control [33,34]. Here, we show that mimecan derived from both male and female plaque yielded consistently higher peptides within LRR1 when compared to control (Fig. 10B). This structure-associated difference may reflect changes in mimecan activity and its protein interactions with potential consequences to collagen fibrillogenesis [104] which may provide clues to the mechanism of plaque formation.

Alpha-actinin exists as an antiparallel homodimer with its monomers associating along their spectrin repeats to form a cylindrical rod domain [105]. The N-terminal ends of this rod are flanked by two actin-binding CH1 and CH2 domains, enabling the formation of crosslinks between F-actin. As such, ACTN4 plays a crucial role in the regulation of the actin cytoskeleton, and therefore cell motility. Higher peptide yields were observed within the spectrin 4 domain of plaque-derived ACTN4 of both males and females, compared to control artery (Fig. 10C) indicating a potential change in this complex’s higher order structure. To our knowledge, the effect of
Fig. 9. Classification analysis of proteins with structure-associated differences reveals ECM proteins as the main class affected in atherosclerosis. PLF analysis identified 151 proteins in females and 154 proteins in males with significant differences in peptide yield across structure (33% and 30% of protein identities shared between control and plaque, ...
atherosclerosis on ACTN4 has yet to be demonstrated, although exposure of oxidised LDLs (oxLDLs) to podocytes was recently shown to significantly enhance podocyte migration by increasing ACTN4 expression [106]. As a major crosslinker of actin, changes in ACTN4 are associated to alterations in cell motility and migration [107]. As such, structure-related differences within this protein may reflect an altered morphological state of resident plaque cells compared those in control artery.

APOB is the primary lipoprotein component of LDLs (and very low-density lipoproteins, VLDLs) and through its tight association, makes up ~50% of the surface of these spherical particles. APOB is highly implicated in atherosclerosis, through its interaction with cell surface LDL-receptors which drive the accumulation of oxLDLs in the intima, under the vascular epithelium [108]. Six heparin-binding (HB) domains exist along the structure of this lipoprotein which are thought to interact with proteoglycans within the subendothelium during early atherosclerosis.

Fig. 10. Proteins displaying sex-conserved differences in peptide yield across their modular structures between human atherosclerotic plaque and internal control artery. nLC-MS/MS-detected peptide sequences were quantified within each 50 aa segment (bar graphs = average, normalised PSMs; mv indicated at y = 0; error bars = SD). Differences in peptide yields across protein structures were assessed by subtracting average, normalised PSMs per segment in control from plaque groups (Line graphs = plaque-control PSMs/segment length; below zero/mv line = higher in control, above zero/mv line = higher in plaque; composite line graphs are normalised between sexes, see Fig. S14 for non-normalised) and statistically compared (paired Bonferroni-corrected, repeated measures ANOVAs: *, p ≤ 0.05; **, p ≤ 0.01; ***, p ≤ 0.001; composite line graphs: stars = significant in both females and males). Two segments within perlecan (A) and another within mimecan (B) displayed significantly higher peptide yields in plaque than in control that were consistent between females and males. These corresponded to the L4 B domain of perlecan and the first leucine-rich repeat (LRR) of mimecan. ACTN4 (C) also had one segment within its fourth spectrin repeat, near the C-terminal end of the protein which yielded more peptides in plaque than in control for both male and female groups. Apolipoprotein B (D) contained two segments in the C-terminal half of the protein which exhibited significantly higher peptide yields in control compared to plaque consistently between males and females (lying within the fifth heparin-binding domain [HB]). A segment on the N-terminal side of APOB however, did yield significantly more peptides in control than in plaque which were female-specific (blue arrow). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

respectively; full PLF analysis results: Tables S11, S12.) PANTHER classification analysis [43] indicated ECM proteins as the main class affected between male and female groups (classes and their protein identities are displayed with coloured frames matching their respective slice within the pie charts; minimum 3 proteins per class; total number of proteins per class indicated in brackets). Cytoskeletal proteins (e.g. actins and myosins), metabolite interconversion enzymes (e.g. superoxide dismutases), protein modifying enzymes (e.g. MMPs and cathepsin B), and protease inhibitors (e.g. serpins) were also some of the top classes affected. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
atherosclerosis [109]. Interestingly, raised peptide yields were observed in segments within the HB5 domain of APOB from both control male and female artery compared to plaque (Fig. 10D). Disruptions or changes associated to this domain could have implications on the ability of LDLs to interact with proteoglycans, potentially providing evidence of shared consequences in atherosclerosis between males and females. Perlecain, mimecan, ACTN4 and APOB all displayed regional significant differences in peptide yield, between plaque and control artery, that were consistent between males and females, providing clear evidence of sex-wide consequences of atherosclerosis (Fig. 10). PLF analysis of plaque and control artery proteomes revealed that structure-associated changes are not limited to just ageing proteins, but also exist as a consequence of age-related diseases like atherosclerosis. These differences were not only observed in ECM proteins (perlecain, mimecan), but also cytoskeletal cross-linkers (ACTN4) and in the primary lipoprotein APOB, reflecting a global shift in matrix remodelling, cell migration and LDL retention.

**Conserved differences within protein regions of alpha-2 macroglobulin and collagen XIV observed between IVD ageing and arterial atherosclerosis**

Atherosclerosis has been characterised as premature biological ageing, where mechanisms such as cellular senescence and its associated promotion of inflammation and perturbed maintenance of the ECM drive the progression of disease [20]. Links between atherosclerosis and IVD ageing and degeneration in particular have also been shown; with aortic plaque lesions, occluded lumbar arteries and high serum cholesterol levels all associated [110]. Concordantly, levels of oxidised (ox-) LDLs

---

**Fig. 11.** Classification analysis of proteins identified with significant structure-associated differences in both arterial atherosclerosis and IVD ageing. PANTHER classification analysis [43] identified several protein classes affected as a consequence of both age-related disease and ageing (classes and their protein identities are displayed with coloured frames matching their respective slice within the pie chart; minimum 3 proteins per class; total number of proteins per class indicated in brackets). As well as ECM proteins, this included cytoskeletal proteins (e.g. actin and keratin), protein-binding activity modulators (e.g. protease inhibitors A2M, serpins and TIMP1), metabolite interconversion enzymes (e.g. superoxide dismutase 3) and transfer/carrier proteins (e.g. apolipoproteins) and as the top five. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
(accumulation of which in the vascular intima is a hallmark of atherosclerosis [108]) were also positive correlated with the progression degeneration in both the nucleus pulposus and OAF [111].

Despite these links, it remains unknown whether proteins are similarly affected by ageing in IVD and atherosclerosis. To examine whether certain mechanisms or consequences of ageing are shared in age-related diseases like atherosclerosis, we next compared the proteins identified with structure-associated alterations in IVD OAF ageing with those in identified in arterial atherosclerosis. Of the 242 proteins identified for atherosclerosis (in both male and female) and 284 proteins identified for IVD ageing (across posterior, lateral and anterior tissue regions) with structure-associated differences, 81 were shared (Fig. S15). Classification analysis of these shared potential targets (Fig. 11) identified ECM protein as the major class affected in both IVD ageing and arterial atherosclerosis, including multiple alpha chains of collagens (I, IV, V, VI, XIV), elastic fibre-associated proteins (fibulins, LTBP2) tenascins (-C, -X) and proteoglycans (perlecan and versican). ECM proteins accounted for ~25% of proteins identified with significant structure-associated differences in both ageing and atherosclerosis, indicating a shared, matrisomal remodelling between IVD and artery.

To determine whether any of these 81 common protein targets had evidence of structure-associated mechanisms or consequences that may be shared between ageing and age-related atherosclerosis, peptide patterns across their structure were compared, revealing regions with coinciding peptide yield differences. Since elements of atherosclerosis have been viewed previously as “accelerated forms of vascular ageing” [20], we compared the peptide yield differences between aged and young IVD to the same differences between plaque and control. Two proteins were identified with regions yielding very similar peptide yield differences between aged vs. young IVD and plaque vs control artery: the protease inhibitor A2M and the alpha-1 chain of the fibrillar collagen-bridging collagen XIV (COL14A1) (Fig. 12).

A2M is a large, tetrameric glycoprotein capable of non-specifically inhibiting all four classes of proteases [112], including ECM-remodelling MMPs [113] and ADAMTSs [114,115]. This inhibition is triggered via the cleavage of a “bait” region within A2M.

**Fig. 12.** A2M and COL14A1 displayed conserved differences in peptide yields within regions of their modular structures between ageing in IVD OAF and age-related atherosclerosis in artery. LC-MS/MS-detected peptide sequences were quantified within each 50 aa segment (bar graphs = average, normalised PSMs; mv indicated at y = 0; error bars = SD). Differences in peptide yields across protein structures were assessed by subtracting average, normalised PSMs per segment in young from aged for IVD and in control for artery (IVD line graph = aged-young PSMs/segment length; artery line graph = plaque-control PSMs/segment length; below zero/mv line = higher in young/control, above zero/mv line = higher in aged/plaque; composite line graphs are normalised between all IVD and artery groups) and statistically compared (Bonferroni-corrected, repeated measures ANOVAs, unpaired for IVD, paired for artery; *, p ≤ 0.05; **, p ≤ 0.01; ***, p ≤ 0.001; composite line graphs: stars = significant in all comparison; black arrows = regions displaying ageing and atherosclerosis-conserved differences; aa ranges of Uniprot-sourced domains indicated). The MG2 domain of A2M (A) contained segments which exhibited the same higher yield of peptides in young than in aged, for two tissue regions of the IVD (posterior and anterior), and in control compared to plaque for female arteries. These differences were significantly different for all except anterior IVD. One segment corresponding to the fibronectin type-III (Fn) 8 domain of COL14A1 (B) yielded the same significantly higher peptides in aged anterior IVD than in young as seen for male plaque when compared to control artery. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
Fig. 13. PLF identifies targets of ageing and atherosclerosis that are potentially unique to the methodology, alongside possible affected mechanisms and functional consequences. For instance, PLF revealed several basement membrane components in the IVD that were sensitive to structure-associated differences, but did not change in relative abundance.
by the target protease which initiates a conformational change within the A2M complex, leading to the entrapment of the protease in a dimeric cage [116]. Higher peptide yields for both young IVD compared to aged and control artery compared to plaque were observed in the MG2 domain (Fig. 12A), which is buried within the centre of the A2M monomer [117] and frames an entrance to the central cavity of the dimer [117]. Since the tetrameric complex of A2M exists in two conformational states (protease bound or unbound), and since the MG2 domains frames the entrance to its entrapping cavity, the structure-associated differences observed may reflect changes in these states between both young and aged IVD and control and plaque artery. This could represent a dynamic shift in protease inhibition within both tissues because of ageing and age-related atherosclerosis.

Collagen XIV belongs to the subfamily of so-called fibril-associated collagens with interrupted triple-helices (FACITs). It exists as a three-pronged homotrimer of COL14A1 chains which function to bridge adjacent collagen I fibrils [118]. Collagen XIV was found to prevent the lateral fusion of collagen I fibrils by limiting fibril diameter [119] and is therefore highly present in tissue regions bearing high mechanical stress [120]. Previous analysis of the IVD ageing dataset showed a lower relative abundance of COL14A1 in full aged IVD compared to young [32] which may contribute to the remodelling of fibrillar collagen. Here, we show that COL14A1 from aged anterior OAF yields more peptides than young within the Fn8 domain, near the centre of the structure (Fig. 12B). Interestingly this same domain also yielded more peptides in male atherosclerotic plaque than in control artery, perhaps indicating a similar structure-associated consequence as observed in aged IVD. Both the IVD AF and artery are comprised of hierarchical, concentric rings of collagen I fibrils. Collagen fibrils from atherosclerotic plaque are synthesised by smooth muscle cells during disease progression, and their degradation is thought to affect plaque stability during disease [121]. It is possible that structure-dependant changes in collagen XIV may contribute to the degradation or remodelling of collagen in both tissues, through the disruption of the interfibrillar bridges they form.

A2M and collagen XIV both exhibited regional peptide yield differences with possible shared structural consequences between ageing in IVD and atherosclerosis in artery. As such, these identified, common markers may provide new links between ageing and atherosclerosis, albeit in two disparate tissues. These potentially corroborate the notion that age-related diseases harbour accelerated forms of tissue ageing, with mechanisms and consequences that may be comparable in both.

Conclusion

The identification of age-susceptible protein structures that are conserved between species, organs and in age-related disease, and the characterisation of shared mechanisms and functional consequences, is crucial for the understanding of connective tissue ageing and the future development of new therapeutic interventions. The application of PLF as a proteomic discovery tool to publically available ageing human IVD, ageing mouse lung and human atherosclerosis datasets provided potentially crucial evidence of common age-associated differences to protein structures, some of which were conserved between human and mouse, three different ECM-rich organs and in an age-related disease. Crucially, PLF identifies structure-associated differences in ECM proteins which remained undetected by conventional whole protein quantification approaches previously performed [31,32,34] (Figs 13A, S4, S8 and S13), therefore enabling a more complete assessment of tissue proteostasis in ageing and disease than previously achieved. Furthermore, the potential of PLF to be used for the interrogation of functional consequences and mechanisms of ageing and disease across species is particularly promising; with possible changes in

(as reported previously [32]) between aged and young (A; Fig. S4). Multiple underlying mechanisms or functional consequences of ageing/atherosclerosis were potentially revealed through regional peptide fingerprints (significant differences) within protein structures (B). PLF identified possible changes in: (i) macromolecular structure, such as for the collagen VI microfibril, where globular regions within its double bead, closest to the triple helices, exhibited differences in peptide yield in IVD ageing (for α1 and α3); (ii) enzyme inhibition activity, such as for A2M where the MG2 domains that frame the entrance to the proteaseinhibiting cavity exhibited peptide yield differences which may be linked to distinct conformational states between aged and young IVD and atherosclerotic and control artery; (iii) protein activation, such as for complement C3 where regions corresponding to both ends of the C3dg fragment exhibited significant differences in peptide yield, indicating the heightened potential transition of C3b to C3c in young compared to aged IVD; (iv) MMP degradation, such as for collagen IV where the NC1 domain of the α2 chain yielded significantly more peptides in young mouse lung than in aged, which may be linked to the known degradation and release of the canstatin matrinke by MT-MMPs; (v) ECM-cell communication, such as a perturbed integrin-mediated link between the rod-2 domain of the cytoskeletal filament-A and the cell-binding domain of fibronectin or G2 and G3 domains of laminin, all of which exhibited significant differences in peptide yield in both human IVD and mouse lung ageing. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
macromolecular structure, enzyme/inhibitor activity, protein activation and ECM-cell communication being revealed (Fig. 13B). The MPLF webtool is publicly available and can be used to analyse any data-dependant acquisition (DDA), label-free LC-MS/MS datasets. PLF is therefore likely to be applicable many more public LC-MS/MS datasets, enabling the exploration of these biological processes in ageing and disease (e.g. COPD [122] and tumorigenesis [123]), the appropriability of animal models, and perhaps the testing of treatment efficacies. Finally, the prospect that ECM components and their associated proteins may be subjected to potentially similar mechanisms or consequences of ageing in both mouse and human and in age-related atherosclerosis is interesting, as it suggests the potential for universal targets that may be present irrespective of differences in life-spans, tissue functions and disease progressions.

Materials and methods

Dataset sourcing and summaries of sample preparation and mass spectrometry methods

All proteomic LC-MS/MS datasets used in this study were sourced from the Proteomics Identification Database (PRIDE) repository. The human IVD, artery, and mouse lung datasets were selected for PLF analysis because 1) they enabled ageing and age-related disease comparisons (with a focus on ECM-ageing) across two species and three organs; 2) common ECM proteins are well represented, with high peptide coverages and yields within their respective proteomes (crucial for maximising the success of PLF analysis and the comparison between datasets); and 3) they were generated using DDA, label-free LC-MS/MS, all of which are currently necessary for the PLF analysis approach.

The young and aged human IVD datasets were generated by Tam and Chen et al. and originally used in the development of the spatiotemporal IVD proteomic resource - DIPPER (http://www.sbms.hku.hk/dclab/DIPPER), already published [32]. As detailed in their original publication, 11 separate IVD regions were carefully dissected from three discs (L3/4, L4/5 and L5/S1), acquired from one aged (59 yr old) and one young (16 yr old) male. Only three of these regions were analysed for this study (posterior, left lateral and anterior portions of the OAF). To summarise their procedures pertinent to this study, after pulverisation from frozen with a freeze mill, proteins were mechanically and chemically extracted by ten freeze-thaw cycles and sonication prior to agitation for two days within an extraction buffer comprised of guanidine hydrochloride, sodium acetate, 6-aminocaproic acid and a protease inhibitor cocktail. Samples were then ultracentrifuged and supernatants were ethanol-precipitated prior to re-centrifugation to leave protein pellets. These were re-solubilised by sonication in urea and ammonium bicarbonate buffer prior to reduction with TCEP and alkylation with iodoacetic acid. Samples were then digested with trypsin/LysC and peptides then split into four fractions by high pH reversed phase fractionation. LC-MS/MS datasets were generated by DDA using an Orbitrap Fusion Lumos Tribrid Mass Spectrometer. Raw MS datasets were downloaded from PRIDE - Project ID PXD017740 for PLF analysis. For more detailed methods of tissue sourcing, sample preparation and MS, please refer to the original publication [32].

The young and aged mouse lung datasets were originally generated by Angelidis and Simon et al. and used in the investigation of ageing by deep tissue quantitative proteomics, already published [31]. To summarise their procedures pertinent to this study, whole bulk lung was taken from young and aged mice (3 and 24 months old; N = 4) and homogenised in PBS containing protease inhibitors. After centrifugation, soluble proteins were separated from insoluble via three incubation steps with three separate buffers of increasing chaotropic activity. All three buffers contained NaCl, Tris–HCl, glycerol, protease inhibitors, Benzonase (Merck) and IGE-PAL-CA-630 (Sigma). However, buffer 2 contained added 0.5% sodium deoxycholate and 0.1% SDS while buffer 3 had higher concentrations of salt (500 mM NaCl compared to 150 mM), sodium deoxycholate (2%) and SDS (1%). Treatment with these buffers resulted in an insoluble protein pellet, enriched in ECM proteins, which was further heated, reduced in TCEP and alkylated in chloroacetamide within a guanidinium hydrochloride buffer. This was then mechanically disrupted using a micro-dounce prior to ultrasonication and overnight digestion with trypsin/LysC to create an ECM-rich peptide fraction. Peptides were purified and LC-MS/MS datasets generated by DDA using a Q-Exactive Hybrid Quadrupole-Orbitrap Mass Spectrometer (Thermo). Raw MS datasets was downloaded from PRIDE - Project ID PXD012307 for PLF analysis. For more detailed methods of tissue sourcing, sample preparation and MS, please refer to the original publication [31].

The human atherosclerotic artery datasets were originally generated by Liang and Ward et al. and used in the proteomic investigation of carotid plaques in men and women, already published [33,34]. To summarise their procedures pertinent to this study, biopsies (4 mm) were taken from atherosclerotic plaque centre regions and internal control regions of carotid arteries obtained from patients (N = 10 male and 10 female; age range 60 to 83 yrs) undergoing endarterectomy. These were flash frozen and crushed prior to sonication in TriZol LS (Life Technologies). Phase separation with chloroform
was performed to remove nucleic acids and proteins precipitated by isopropyl alcohol. The centrifuged protein pellet was then resuspended in a urea/thio-urea buffer containing protease inhibitors (PelaBloc; Sigma) prior to reduction with dithiothreitol and alkylation with iodoacetamide. Samples were filtered (3 kDa cut-off) and digested with trypsin. Peptides were purified and nLC-MS/MS datasets generated using an LTQ Orbitrap Velos Pro-Mass Spectrometer (Thermo). Raw MS datasets was downloaded from PRIDE - Project ID PXD003930 for PLF analysis. For more detailed methods of tissue sourcing, sample preparation and MS, please refer to the original publications [33,34].

**Peptide location fingerprinting**

Peptide location fingerprinting was applied using our predeveloped MPLF webtool as previously described (https://www.manchesterproteome.manchester.ac.uk/#/MPLF) [27,28]. Analyses were performed separately on each OAF region (posterior, lateral and anterior) for the human IVD young vs. aged datasets and also separately on males and females for the human arterial atherosclerotic plaque vs. control datasets.

Peptide list CSV files were imported into the webtool, and the primary sequences of matched proteins were bioinformatically divided into 50 aa-sized segments. Identified peptide sequences were then mapped and quantified within each segment by spectral count. Peptide sequences which spanned two connecting segments were counted in both. Total spectral counts were then summed per segment. To detect protein region-specific differences in peptide yield that were unskewed by differences in whole protein abundance, summed peptide counts per segment were median normalised based on the experiment-wide total spectrum counts of their corresponding whole proteins. Additionally, proteins which were exclusively present in one experimental group (young or control) but not the other (aged or plaque) were excluded from comparisons. These steps minimised the effect of whole protein relative abundance on regional comparisons of peptide yield across protein structure and ensured that the identification of proteins with structure-related differences were independent of protein presence or quantity. Modular fluctuations in peptide yield across protein structures were revealed by subtracting the average, normalised peptide counts per segment in one experimental group from the other and dividing them by the segment length (50 aa). Average, normalised peptide counts in each segment were statistically compared between experimental groups using Bonferroni-corrected, repeated measures ANOVAs (unpaired for young vs aged; paired for plaque vs control). For further details on PLF and its application, please refer to our MPLF webtool development publication [27].

**Author contributions**

AE conceived and designed the study, performed all PLF analysis, interpreted the data, prepared the figures, and wrote the paper. AE and MJS
contributed to the acquisition of funding. MO contributed to the running of PLF and to the maintenance, support, and continued development of the MPLF webtool. MJS and AT contributed to study conception. MJS contributed to experimental design and preparation of figures. PC, VT, LJW, JAH, XMY, HBS, DC and MJS contributed to the interpretation of results. All authors contributed to editing of the paper.

Funding disclosure

This study was jointly funded by the Manchester Institute for Collaborative Research on Ageing (MICRA) (AE [PI], AT, MO and MJS [Co-I]) and Walgreens Boots Alliance (WBA) (MJS), UK. In line with WBA’s Corporate Social Responsibility Policies, no animal testing has been conducted as part of this study. All mouse mass spectrometry datasets have been acquired (downloaded) from a public repository - PRIDE. MO would like to acknowledge funding from the British Heart Foundation Centre of Research Excellence (BHF CRE) and Wellcome Trust grant (108413/A/15/D), DC is supported by RGC European Union - Hong Kong Research and Innovation Cooperation Co-funding Mechanism (E-HKU703/18), and European Union’s Horizon 2020 research and innovation program (iPSpine; grant agreement number 825925) and PC by the Shenzhen “Key Medical Discipline Construction Fund” (SZXK077). JAH and MJS would like to acknowledge funding from the UK Engineering and Physical Sciences Research Council (EPSRC; EP/V011065/1). The work on the human atherosclerotic artery datasets was supported by grants to XMY from the Swedish Heart Lung Foundation, Stroke Foundation, Olle Engkvist Foundation, Swedish Gamla Tjänarinnor Foundation and Linköping University Hospital Research Fund.

Declaration of Competing Interest

The authors declare no conflicts of interest. WBA approved manuscript submission but exerted no editorial control.

Acknowledgements

We would like to thank the University of Manchester BioMS facility for providing the resources needed for MS analysis and the research IT services for their help in hosting and running the MPLF webtool.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.matbio.2022.05.007.

Data availability

The LC-MS/MS datasets analysed by PLF in this paper were downloaded from the PRIDE repository: human IVD ageing - PXD017740; mouse lung ageing - PXD012307; human arterial atherosclerosis - PXD003930. PLF analyses of all datasets, and associated protein schematics can be interactively viewed and downloaded via the MPLF webtool at https://www.manchesterproteome.manchester.ac.uk/#/MPLF under the “Location Fingerprinter” tab [27,28].

References


Peptide location fingerprinting identifies species- and tissue


