



# Proteomic definitions of basement membrane composition in health and disease

**DOI:**

[10.1016/j.matbio.2016.08.006](https://doi.org/10.1016/j.matbio.2016.08.006)

**Document Version**

Accepted author manuscript

[Link to publication record in Manchester Research Explorer](#)

**Citation for published version (APA):**

Randles, M. J., Humphries, M. J., & Lennon, R. (2017). Proteomic definitions of basement membrane composition in health and disease. *Matrix biology : journal of the International Society for Matrix Biology*, 57-58, 12-28. <https://doi.org/10.1016/j.matbio.2016.08.006>

**Published in:**

Matrix biology : journal of the International Society for Matrix Biology

**Citing this paper**

Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

**General rights**

Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

**Takedown policy**

If you believe that this document breaches copyright please refer to the University of Manchester's Takedown Procedures [<http://man.ac.uk/04Y6Bo>] or contact [openresearch@manchester.ac.uk](mailto:openresearch@manchester.ac.uk) providing relevant details, so we can investigate your claim.



## **Proteomic definitions of basement membrane composition in health and disease**

Michael Randles<sup>1,2</sup> Martin J. Humphries<sup>1</sup> and Rachel Lennon<sup>1,2,3</sup>

*<sup>1</sup>Wellcome Trust Centre for Cell-Matrix Research, Faculty of Life Sciences University of Manchester, Manchester, UK; <sup>2</sup>Institute of Human Development, Faculty of Human Sciences, University of Manchester, Manchester, UK; <sup>3</sup>Department of Paediatric Nephrology, Central Manchester University Hospitals NHS Foundation Trust (CMFT), Manchester Academic Health Science Centre (MAHSC), Manchester, UK.*

Corresponding author: Dr Rachel Lennon, Wellcome Trust Centre for Cell-Matrix Research  
Michael Smith Building, University of Manchester, Manchester M13 9PT, UK.

Phone: 0044 (0) 161 2755498 Fax: 0044 (0) 161 2755082.

Email: [Rachel.Lennon@manchester.ac.uk](mailto:Rachel.Lennon@manchester.ac.uk)

## **Abstract**

Basement membranes are formed from condensed networks of extracellular matrix (ECM) proteins. These structures underlie all epithelial, mesothelial and endothelial sheets and provide an essential structural scaffold. Candidate-based investigations have established that predominant components of basement membranes are laminins, collagen type IV, nidogens and heparan sulfate proteoglycans. More recently, global proteomic approaches have been applied to investigate ECM and these analyses confirm tissue-specific ECM proteomes with a high degree of complexity. The proteomes consist of structural as well as regulatory ECM proteins such as proteases and growth factors. This review is focused on the proteomic analysis of basement membranes and illustrates how this approach can be used to build our understanding of ECM regulation in health and disease.

## **Key words**

Extracellular matrix, proteomics, glycoproteomics, degradomics, mass spectrometry, quantification, fractionation, enrichment, bioinformatics, systems biology.

## Introduction

Basement membranes (BMs) are essential extracellular matrix (ECM) structures within multicellular organisms. In most cases BMs form barriers that separate epithelial, mesothelial or endothelial cells from the underlying stroma. It is considered that laminin networks form the basis of all BMs [1, 2] and that type IV collagen networks provide structural strength and stability of BMs [3]. There are 16 different laminin  $\alpha\beta\gamma$  trimers [4] and three different type IV collagen  $\alpha\alpha$  trimers that are known to occur in nature [5] and these trimers form networks of sheet-like structures. Laminin and type IV collagen networks have differential expression throughout mammalian tissues [6, 7], most likely conferring different physical properties to BMs. Indeed BMs fulfil varied and sophisticated roles in normal tissues, in addition to forming barriers, they direct cell differentiation, morphology and survival [8, 9]. As a result, BMs comprise an extensive repertoire of ECM components. Other well-characterised BM components include: heparan sulphate proteoglycans (perlecan, agrin), nidogens, type XVIII collagen and type XV collagen. However, this list is not exhaustive and the tissue-, developmental stage- and disease-specific expression of all putative BM proteins has yet to be determined.

BMs must be carefully regulated in health and their disruption often results in disease. Altered BM homeostasis is seen with an accumulation of ectopic ECM within BMs in fibrosis [10-12], excessive BM glycosylation in diabetes mellitus [13, 14] and BM degradation associated with cancer metastasis [15-17]. Additionally, mutations in genes encoding BM components can cause hereditary diseases affecting multiple organs including nephropathy, sensorineural hearing loss, epidermolysis bullosa, ocular abnormalities and neuromuscular deficit [18-20]. It is therefore of clinical importance to characterise the spatio-temporal expression of BM proteins and the plethora of potential posttranslational modifications that occur in health and disease. This undertaking is a major challenge, however, with the development of ECM proteomics, new insights into the composition of ECM have been achieved [21] and the 'omic' approach complements more traditional candidate-based approaches.

## **Rationale for mass spectrometry-based proteomic investigation of BMs**

Gene expression analysis techniques, such as microarray and RNA sequencing, enable global quantification of BM components at the transcript level. These analyses allow the collection of vast amounts of information, and are often used as surrogates for protein abundance. However, although powerful, these techniques frequently misrepresent the abundance of proteins within biological systems due to post-transcriptional regulation of protein abundance [22]. This phenomenon is particularly evident for ECM proteins, which may have low turnover, and indeed many studies have shown poor correlation between ECM-transcript and ECM-protein level [23-25].

Mass spectrometry (MS) is a highly sensitive analytical technique that facilitates the detection and quantification of proteins, including those that are of low abundance, in a global manner. Moreover, unlike gene expression analysis, MS-based techniques enable researchers to probe the plethora of potential post-translational protein modifications. The analytical instrument, a mass spectrometer, generates charged gaseous ions, measures the mass-to-charge ratio ( $m/z$ ) of these ions and records the number of ions at each  $m/z$  value. MS analysis of whole proteins is known as "top down" proteomics. Since MS of whole proteins is less sensitive than MS of peptides [26] and biological samples are complex and frequently contain thousands of proteins, the  $m/z$  of an intact protein in a complex mixture by itself is insufficient for unique identification. In order to detect, and ideally quantify, all of the proteins present within a biological sample, proteins are digested to peptides referred to as "bottom up" proteomics. This is frequently performed using enzymatic trypsin digestion. The benefit of trypsin digestion is that predictable tryptic peptides are produced. Peptides have defined (either lysine or arginine) C-terminal protonated amino acids following trypsin digestion and this provides an advantage in peptide sequencing and subsequent database matching. However, the  $m/z$  of tryptic peptides in complex mixtures is also non-unique leading to ambiguity. To perform global shotgun proteomics experiments tandem MS (MS/MS) is

required, whereby peptide ions are selected, fragmented and detected. Information from both the precursor peptide (MS1) and the peptide fragments (MS2) enables the confident identification of peptides that in turn may be unique to, and therefore specific for, a given protein.

Peptide ions in the gaseous phase are most frequently generated by electrospray ionisation (ESI) [27]. These ions are then detected by mass analysers, which operate by measuring the trajectories of ions in an electric field or by trapping ions for further manipulation. Examples of mass analysers include: quadrupole mass filters, time of flight instruments, Fourier transform ion cyclotron analysers and the orbitrap. These analysers can be stand alone or assembled in series to take advantage of their various strengths.

Strategies to increase proteome coverage from a given sample, involve separating or fractionating the sample prior to analysis. This is achieved with high performance liquid chromatography (HPLC) [28], which is a frequently used modality, and prior to HPLC proteins can be separated by SDS-PAGE [29], off-gel electrophoresis (OGE) [30] and run as separate fractions. Although these techniques all increase total proteome coverage, they do not specifically increase ECM proteome coverage. For instance the introduction of OGE into a MS pipeline for ECM proteomics by Naba et al. led to a sixfold increase in proteins identified, but only a threefold increase in ECM proteins [31]. However, without this fractionation step growth factors were not identified within ECM enriched samples. In contrast, as discussed in this review approaches to isolate the ECM sub-proteome are essential for ECM/BM proteomic experiments.

To identify peptides, tandem MS (MS/MS), employs two analysis steps in series. Firstly, peptide ions within a narrow  $m/z$  range are selected by either data dependent or data independent acquisition. Data dependent acquisition (DDA) is the most commonly applied for global proteomics. Using DDA the most intense peptides are selected from an initial MS scan

(MS1). The exact mass of this precursor peptide ion is determined and it is then fragmented along its backbone, usually by collision induced dissociation (CID), electron transfer dissociation (ETD) or higher energy C-trap dissociation (HCD). These fragment ions are then detected, giving rise to characteristic spectra derived from a single precursor, which can be used to deduce the sequence of the precursor peptide. DDA is sequential, so the next most intense ion is selected and so forth. Limitations in acquisition speed mean that certain lower intensity peptides may not be selected and therefore undetected. In contrast, MS2 spectra are acquired from all precursor peptide ions in data independent acquisition (DIA) mode. One example of DIA is sequential window acquisition of all theoretical mass spectra (SWATH-MS). Theoretically, SWATH-MS allows all peptides to be analysed, and to identify peptides from the complete MS2 produced by mixed species of fragment ions. This technique negates the loss of low abundance peptide species theoretically increasing coverage of the proteome and the accuracy of quantification. The main challenges of DIA MS is the interpretation of multiplexed MS2 spectra originating from multiple precursors. Data interpretation is aided by spectral libraries containing MS of fragment ion spectra of peptides based on shotgun sequencing of natural, or synthetic peptides to obtain high-quality spectra for those proteins difficult to detect in the natural sample.

The next challenge is to determine the original proteins from which the identified peptides have arisen. Detected peptides can be compared to protein databases using algorithmic search engines. These databases have libraries of peptides predicted *in silico* from theoretical digests of the complete genome sequence of the organism of interest. By matching experimentally detected peptides to these databases, the identity of proteins in a given sample can be inferred. This is trivial in the case of peptides that are unique to a given protein, but pose a complex problem for peptides with sequences that are shared between proteins. As a result there is a wide range of software available for processing MS proteomic data and the most popular include; Mascot [32], SEQUEST [33], X!Tandem [34], MaxQuant [35], Andromeda [36] and Proteome Discoverer [37].

### ***Quantification***

Usually the aim of a MS experiment is to compare the relative abundance of proteins between different biological conditions. In addition, researchers would like to be able to discount co-purified contaminant proteins from their datasets by using 'subtractive proteomics' where proteins detected in similar quantities in control samples and experimental samples can be filtered from the data. It is not feasible to determine absolute quantification of all peptides in a global proteomic experiment, since this would require the use of peptide standards for each of the peptide identified and therefore relative quantification is performed.

The incorporation of 'heavy or light' isotope labeled lysines and arginines into different biological samples can be used to aid quantification of proteins by MS. This approach, known as stable isotopic labeling using amino acids in cell culture (SILAC), enables peptides from different samples to be distinguished within a single MS analysis [38]. This technique is used mostly in cell culture and use in vivo is limited, however, this has been performed with the generation of the SILAC mouse [39]. Another labeling technique, termed iTRAQ, utilizes isobaric tags that have the same intact mass but have different masses when fragmented and detected at MS2 [40]. The fragment spectra from the isobaric tags are then used as a correlate of relative peptide abundance. The benefit of performing labelled proteomics is that technical variability can be negated by multiplexing and analysing samples simultaneously in a single combined MS analysis. The drawback of such experiments is that samples are increased a number of fold in complexity as well as individual samples being more dilute in the combined sample.

Label-free quantification strategies allow the comparison of many MS analyses. Spectral counting, which correlates the number of observed mass spectra matched to peptides from a protein to the abundance of that protein [41], peptide counting [42], peptide ion signal measurements [43] are examples of label-free quantification and a combination of these parameters has also been used [44]. Spectral counting is suitable for any MS data, but is less



sensitive than peptide ion signal measurements. In order to use peptide ion measurements chromatographic alignment is key, leading to potential issues pertaining to the properties of LC columns during use, resulting variability in retention times of identical peptides in different MS runs. Label-free proteomics is suitable for any sample type including human tissue and has the benefit of reduced cost as expensive labeling procedures are avoided. The continual development of instrumentation and software for MS analysis has increased the precision and accuracy of both labeled and label-free quantification.

### ***In silico* prediction of the matrisome**

In 2009 Ricard-Blum and colleagues published an article describing MatrixDB, an online resource cataloguing protein-protein and protein-carbohydrate interaction data [45]. MatrixDB focused on ECM proteins, originally reporting 259 extracellular proteins, but also included cellular proteins which interact with ECM proteins. Two years later Hynes and colleagues published a considerable body of work defining the genes within the mouse and human genomes that encode ECM proteins. The resultant database was the first comprehensive *in silico* prediction of the *in vivo* 'matrisome' [46]. The term matrisome was first coined in the context of BMs [47], referring to the structural components of ECM, but has since been expanded to include proteins that interact with, or modify, the ECM. Identifying protein domains that are definitive components of ECM protein structure led to the first drafts of the human and mouse matrisomes. These protein domains were based on published data and an iterative query of the UniProt protein database [48]. Initial lists of putative ECM proteins were refined by the removal of proteins that contain domains that preclude a protein from being an ECM protein. These dynamic lists are now manually curated and constantly updated with the growing understanding of the ECM [21]. The current mouse and human matrisomes contain 1110 and 1021 proteins respectively. These catalogues form immensely powerful resources for BM researchers, allowing the rapid identification of known ECM proteins from a complex dataset.

The number of matrisome proteins that are expressed in BMs is unclear. Using gene ontology definitions, 103 proteins from mouse and human are annotated as BM, but this list includes cell surface proteins that are involved in BM organisation, such as dystroglycan and CD151. In contrast, the *in silico* matrisome has a much stricter annotation with only 24 proteins being annotated as BM [21]. These are the most well characterised BM proteins, including: type IV collagens, laminins, collagen XV, collagen XVIII, perlecan and nidogens. Because ontology definitions and large databases may contain misleading or incorrect BM annotations, it is essential to manually check individual proteins against the literature or perform immunolocalisation experiments to confirm BM localisation. By utilising *in silico* predications and the plethora of large proteomic datasets that are available it is possible to map tissue expression of putative BM proteins, based on manual curation of GO and the human matrisome, showing the most highly expressed BM proteins across tissues (Figure 1). GO, Matrixdb and the matrisome are extremely useful resources for BM researchers, however, biochemical analysis with solubility profiling and immunolocalisation of putative BM components is required to unequivocally assign BM localisation to a given protein.

### **ECM enrichment strategies for MS-analyses**

Advances in proteomic pipelines have enabled the detection and quantification of thousands of peptides in very complex mixtures. However, because most MS experimental set-ups use data dependent acquisition therefore identifying only the most abundant peptide ions for given m/z, certain protein subsets are inevitably under-sampled. BM proteins are underrepresented in the vast majority of available whole tissue or whole cell datasets because the experimental approaches used for protein extraction are not optimised for identifying BM proteins. To analyse BM composition it is therefore important to consider enriching for this sub-proteome prior to MS analysis.

Analysis of BM proteins is limited by their physicochemical properties: they are frequently large in size, have copious post-translational modifications and because of their high degree of

crosslinking are poorly soluble. The majority of MS experiments analyse whole cell or tissue lysates generated by detergent-based lysis buffers. These lysates are saturated with cellular proteins that invariably mask the identification of BM proteins. Moreover, highly insoluble proteins, including BM proteins, are often cleared from tissue lysates through centrifugation and discarded in the pellet. Therefore the BM remains an under-sampled proteome despite the growing number of MS analyses that have been performed. What is more, the BM of interest is often embedded within larger structures and difficult to isolate or extract. For example, at the dermal-epidermal junction, the skin BM is linked via type VII collagen anchoring fibrils to the underlying dermis, which is rich in type I and III fibrillar collagens. A second example is the glomerular BM within the kidney, which is sandwiched between epithelial (podocyte) and endothelial cell layers, and interacts with mesangial cells and ECM. The close association of BMs to cells and other ECM structures poses difficulties in dissociating and isolating BMs. However, a number of research groups are starting to overcome the challenges associated with the proteomic analysis of BMs.

The low solubility of BM proteins can be utilised to isolate BM protein-enriched fractions. A number of methods have now been described [49] and these follow variations of a common theme with the following sequence. Tissue samples are first homogenised to increase the surface area to volume ratio enabling efficient diffusion of buffers into the tissue. The second step typically uses high concentration NaCl to release proteins that are loosely bound to the ECM. The third and fourth steps use, Triton- or sodium dodecyl sulfate (SDS)-based lysis buffers and mild mechanical disruption for several hours to achieve the removal of cellular components. In the final step, the proteins are solubilised and denatured using urea or guanidine and proteins precipitated with ethanol, methanol or acetone. The most insoluble fraction typically contains the vast majority of the collagen, laminin and other core BM proteins, which are discarded in most tissue lysis processes. The earlier fractions often contain more soluble BM-associated proteins, such as growth factors and matrix-remodelling proteins, which are also considered to be part of the matrisome. Therefore, analysis of all fractions

generated will produce the largest coverage of the BM proteome. Co-purified contaminant proteins can be filtered from BM proteome datasets using 'subtractive proteomics' to this end both gene ontology definitions and the matrisome resource are powerful tools.

## **Tissue proteomes**

### **Vascular ECM proteomics**

BMs are crucial for the structural integrity of blood vessel walls, and remodelling of these ECM structures has been implicated in a number of vascular pathologies. Mayr and colleagues have studied a number of vascular proteomes using ECM enrichment coupled to MS. Their ECM extraction process commences with 0.5 M NaCl to release loosely bound ECM proteins, SDS to remove cellular components and 4 M guanidine to solubilise the detergent resistant ECM fraction [50]. Normal human aorta was studied first and, by combining MS data from NaCl extracts and guanidine extracts, 103 ECM proteins were detected. ECM proteins that had not previously been associated with the ECM of blood vessels were revealed, including podocan, sclerostin and agrin. Podocan is an atypical proteoglycan, which had only been identified in the glomerular BM as a product of kidney podocytes [51]. Having characterised the ECM of normal human aorta, the MS analysis pipeline was applied to three additional studies of vascular ECM, with all datasets available via the website: <http://www.vascular-proteomics.com/>. Next, aortic disease with abdominal aortic aneurysms (AAA) [52], which are characterised by pathological elastolysis and collagenolysis of the aortic ECM was investigated. Collagen XII, tenascin, thrombospondin 2, fibronectin, ACLP, and periostin were found to increase in AAA samples relative to normal aorta. Furthermore, subsequent analysis revealed MMP-12 induced extensive fragmentation of fibronectin and cleaved collagen XII, tenascin, and periostin in AAA samples.

In a subsequent study, the same group analysed ECM derived from cardiovascular tissues from a porcine model of ischemia/reperfusion (I/R) injury [53]. Two distinct cardiac regions were analysed using both proteomics and gene expression: the focal injury in the left ventricle

and the border region close to the occluded coronary artery. As reported by others in subsequent studies [23-25], the regulation of mRNA levels did not correlate with regulation of ECM proteins. Although gene expression of ECM proteins was similar in both regions, the corresponding protein levels were much higher in the focal lesion. The analysis identified proteins that contribute to cardiac remodeling in this context including cartilage intermediate layer protein 1, matrilin-4, extracellular adipocyte enhancer binding protein 1, collagen -1(XIV), asporin and prolargin.

Venous hypertension was the focus of a subsequent proteomics study. This causes significant morbidity in a variety of clinical states, including varicose veins. Mayr and colleagues utilised their ECM enrichment-MS pipeline to study ECM remodeling in varicose saphenous veins [54]. Morphologically varicose veins demonstrated extensive neointima formation with subendothelial fibrosis, wall thickening and luminal dilation. MS analysis detected 48 novel venous ECM proteins, in addition to significantly altered ECM protein abundance in varicose veins compared to normal saphenous veins. These changes included loss of aggrecan and several small leucine-rich proteoglycans and a compensatory increase in collagen I and laminins  $\beta 1$ ,  $\beta 2$  and  $\gamma 1$ . In addition, network-forming type IV collagen  $\alpha 1$  and  $\alpha 2$  were significantly decreased. This loss of the normal structural BM component and increase in the fibril-forming type I collagen, suggests overall there may be stiffening of the veins. In contrast to the increase in type I collagen, most ECM proteins were decreased in varicose veins, without a concomitant reduction in gene expression, suggesting that the changes occur at the level of post-transcriptional regulation. To investigate the overall decrease in abundance of ECM proteins in varicose veins further, the data were searched for proteases in that could explain these changes. Two mast cell proteases were identified in varicose veins: chymase and tryptase  $\beta 1$ , both of which localised in the neointima and vessel wall of varicose veins. Further analysis revealed that (BM)-associated proteins, such as collagen VI, perlecan and tenascins, were predominantly affected by digestion with chymase. In contrast, no effect was

observed on type I collagen, suggesting that increased chymase levels cannot attenuate collagen I accumulation in varicose veins.

Overall, these analyses further our understanding of the composition of the vascular extracellular environment and provide insight into ECM remodeling and degradation during pathological processes.

### **The tumour niche**

The tumour niche is a fascinating ECM proteome, which has been studied using human mammary carcinoma xenografts [55, 56]. In these MS experiments the ECM produced by host mouse stromal cells could be readily distinguished from the ECM produced by the human carcinoma due to the differences in mouse and human protein sequence. Naba et al. (number of reference), used cancer cell lines of differing metastatic potential to investigate how the tumours differed in their extracellular niche. The number of proteins that can be compared in this elegant xenograph approach is limited due to the homology between mouse and human protein sequences, which results in limited numbers of peptides that can be used for relative quantification of tumour and stromal contributions to the ECM proteome. Methodologically, the insolubility of the ECM was utilised in order to produce ECM-enriched samples and in these experiments a CNMCS (Cytosol/Nucleus/Membrane/Cytoskeleton) Compartmental Protein Extraction kit was used to separate other protein compartments from the ECM.

MS analysis revealed that certain BM components were expressed only by the tumour cells alone (laminin chains  $\alpha 3$ ,  $\beta 3$ ,  $\gamma 2$ ), by both compartments (laminin chains  $\alpha 5$ ,  $\gamma 1$ , type VI collagen  $\alpha 3$ , perlecan) or by the stroma alone (nidogens 1 and 2). In addition, several BM components (laminin chains  $\beta 1$  and  $\beta 2$ , and the collagen chains IV- $\alpha 1$ , IV- $\alpha 5$  and XVIII- $\alpha 1$ ) were secreted in poorly metastatic tumours mostly by the stroma, whereas in highly metastatic tumours, they were secreted by both the tumour cells and the stromal cells. In addition, 36 ECM proteins were found within the tumour niche in both highly metastatic and poorly

metastatic tumour models. However, the origin differed as a function of the metastatic potential. In poorly metastatic tumours, fibronectin and periostin were secreted by the tumour cells, but in the presence of highly metastatic tumour cells stromal cells were induced to also secrete these proteins. Overall, these studies demonstrate the dynamic nature of the ECM, the potential for switch with regard to cellular origin and the association of certain ECM components with disease severity.

### **Proteomic analysis of purified basement membrane from the eye**

The mammalian eye is an ideal tissue system to discover the composition of BMs via MS-based proteomics. This is because the BMs in the eye are thick and relatively accessible. For example in humans the thickness of the retinal BM (or inner limiting membrane, ILM) is approximately 1-2  $\mu\text{m}$  and the lens capsule (LC) ranges from 10-15  $\mu\text{m}$ . Moreover, the ILM can be easily dissected away from vitreous body and lens capsule. Balasubramani and colleagues performed studies of BMs from the adult human eye [31]. The ILM, the retinal vascular BMs, and the LC, were isolated for analysis by MS. Peptide and protein identifications were performed using MaxQuant and this led to the identification of 1129 proteins [57]. According to definitions based on MatrixDB, the *in silico* matrisome and the UniProt database, 202 proteins were designated as ECM. The dominant type IV collagen trimer in lens capsule (LC) and retinal vascular BMs had a chain composition of  $\alpha 1\alpha 1\alpha 2$ , in contrast the dominant type IV collagen in the ILM was the  $\alpha 3\alpha 4\alpha 5$ , which has a more tissue restricted expression pattern (Figure 1).

### **Glomerular ECM proteomics**

Glomeruli are the individual filtration units of the kidney and are comprised of a bundle of specialised capillaries enclosed by Bowman's capsule. Glomerular capillary walls act as selective filters that are lined by endothelial cells, covered by epithelial cells known as

podocytes and between the two cell layers there is the glomerular BM (GBM). The ECM compartments in the glomerulus are the GBM of the capillary walls in addition to the mesangial matrix and the matrix associated with Bowman's capsule. We investigated the human glomerular ECM using proteomics and identified 144 ECM proteins [58]. The ECM signature we described includes all previously identified glomerular ECM components, in addition to many new and abundant components. We demonstrated that both type VI collagen and tubulointerstitial nephritis antigen-like 1 were abundant components of the normal human glomerular ECM. By generating and analysing a protein-protein interaction network for the glomerular ECM, we identified a core of highly connected structural components. Topological network analysis confirmed that BM proteins formed a highly connected "core" sub-network, whereas ECM-associated proteins were less clustered and connected in the network.

We subsequently narrowed our focus and analysed the specific cellular contributions to the GBM. To do this, we isolated ECMs derived from podocytes and glomerular endothelial cells *in vitro* [59]. With subsequent MS analysis, we identified 127 extracellular proteins in endothelial-derived matrices and 142 extracellular proteins in podocyte-derived matrices. These analyses indicated distinct contributions to the GBM by the two cell types. Furthermore, coculture of podocytes and endothelial cells resulted in organisation of their ECM into structures between the two cell types that, compared with the thin, underlying ECM produced in monoculture, more closely resembled the *in vivo* GBM. This suggested a role for cell-cell crosstalk in the production of the GBM. Importantly, MS analysis revealed that compared to monoculture, the coculture ECM proteome better resembled the tissue-derived glomerular ECM dataset, indicating its relevance to the GBM *in vivo*. One component of the *in vivo* glomerular matrix identified in coculture, but absent in monoculture, was hemicentin a protein which is key for connecting adjacent basement membranes [60] (Figure 2). During glomerular development the GBM forms by the fusion of two distinct BM one derived from glomerular endothelial cells and the other from podocytes, these data suggest that hemicentin may play an important role in this process. Interestingly, BM proteins detected in the *in vivo* glomerular



matrix, but absent from the *in vitro*-derived ECMs include laminin  $\alpha 1$  and  $\alpha 2$  and type IV collagen  $\alpha 6$ . Laminin  $\alpha 1$  and  $\alpha 2$  are usually expressed by mesangial cells and type IV collagen  $\alpha 6$  is present in Bowman's capsule, suggesting contributions of these compartments to the *in vivo* glomerular matrix.

To understand better how the glomerular ECM changes in disease, we used a mouse model of early glomerular dysfunction. Increased albumin excretion into the urine (albuminuria) is a marker of glomerular barrier dysfunction and predicts the development of chronic kidney disease [61, 62]. FVB/NHanHsd mice have fewer glomeruli than C57BL/6J OlaHsd mice, and within strains males have fewer glomeruli compared with females [63]. As a result, albuminuria develops earliest in FVB/NHanHsd male mice and latest in C57BL/6J OlaHsd female mice. Applying our ECM enrichment-MS pipeline, we discovered unique strain- and sex-dependent glomerular ECM signatures that correlated with variations in albuminuria [64]. The ECM proteome differences we discovered were concomitant with striking ultrastructural thickening and irregularity in the GBM. These changes in BM structure occurred in the absence of differences in abundance of known GBM components, such as type IV collagen  $\alpha 3\alpha 4\alpha 5$  and laminin  $\alpha 5\beta 2\gamma 1$ . Instead netrin 4, type I collagen and tenascin C were significantly altered.

### **Lung ECM proteomics**

Pulmonary fibrosis, the result of uncontrolled ECM deposition, is a major cause of mortality in the developed world [65], yet, this process remains poorly understood. However, the ECM proteome of the murine lung has now been studied in depth in order to better understand this pathology. In 2012 Hynes *et al* analysed the proteome of enriched ECM extracted from murine lung [46]. The MS analyses identified 143 ECM proteins. The vast majority of collagens, proteoglycans and expected BM components were identified. In 2015 a further two studies were published mapping the changes in lung ECM upon acute injury. Emson and colleagues used an innovative proteomics platform quantifying the fractional synthesis rates of proteins via stable isotope labelling and LC/MS-based mass isotopomer analysis [66]. Mice were

labelled with heavy water for up to 21 days following the induction of lung fibrosis with bleomycin. Lung samples were collected for sequential protein extraction to fractionate cellular, guanidine-soluble ECM proteins and residual, insoluble ECM proteins. This approach enabled the quantification of ECM protein kinetics during the onset of chemically-induced pulmonary fibrosis. The different fractions from the ECM enrichment protocol were analysed and found that the more insoluble the ECM compartment, the slower the fractional synthesis rate. The interpretation of these data is that more insoluble ECM components were older and more mature, which was supported by quantification of pyridinoline cross-links, which increased with insolubility. The synthesis rates of biglycan and decorin, two small leucine-rich proteoglycans associated with collagen fibril assembly and growth factor signalling, were noted to be particularly rapid, with > 60% labelled in both control and bleomycin-dosed lungs after just one week. Similarly label incorporation into fibronectin was >75% in both control and bleomycin-dosed lungs within one week. Bleomycin-induced injury led to a moderate increases in newly synthesised ECM proteins after one week, but large increases in the amount of newly synthesised ECM proteins after three weeks. The percentage of newly synthesised fibril-forming collagens type I, III, V and VI were particularly increased by bleomycin-induced injury, in addition to BM proteins type IV collagens, laminins and nidogens.

Schiller *et al* also utilised the bleomycin model to induce acute fibrotic lung injury and studied both the acute injury and resolution [23]. Total tissue and bronchoalveolar lavage fluid (BALF) was analysed over the course of 8 weeks [23] and a 4-step detergent extraction, was utilised to increase ECM proteome coverage, as described previously. Coupling this approach to a next-generation Quadrupole–Orbitrap mass spectrometer (Q Exactive) enabled accurate intensity-based label-free quantification. Overall 8,366 proteins, of which 435 were known matrisome proteins, were detected.

The authors also coupled this proteomic analysis to complementary RNA-sequencing data. Interestingly BM proteins were reduced by bleomycin injury even though their transcripts were

upregulated. In fact, the relationship between transcript data and proteomic data was found to be complex and not directly correlative. The  $\alpha$ 3- and  $\alpha$ 5-laminins were most abundant in adult mouse lung, followed by  $\alpha$ 4-laminins. The  $\alpha$ 1- and  $\alpha$ 2-laminins, which are mainly restricted to embryonic development [67], were at least ten-fold lower than all other laminins. Upon injury, the high-abundance laminin chains were down regulated at day 3 and day 14 together with the type IV collagen chains. In contrast, the laminin  $\alpha$ 1-,  $\alpha$ 2-, and  $\alpha$ 4-chains were upregulated early upon injury, suggesting a potential role in the early phase of tissue repair. Interestingly, these developmental laminins are implicated in the pathogenesis of Alport syndrome caused by type IV collagen mutations and associated with kidney, eye and hearing defects [68]. In addition, complement components and ECM proteases ADAMTS1, ADAMTS15, MMP2, MMP12 and MMP14 were increased in abundance with bleomycin injury. Moreover, BM proteins were found to become more soluble with acute lung injury due to proteolytic processing. By taking these data in the context of Emson and colleagues study, acute lung injury appears to promote the synthesis of BM proteins, but this is overcome by enhanced degradation resulting in overall loss of BM components. In contrast, other ECM proteins, including fibronectin and fibrillar collagens, have increased synthesis that is not balanced by degradation leading to accumulation of these ECM proteins in the damaged lung.

By analysing the soluble fractions in addition to the insoluble fractions generated during the ECM enrichment protocol, Schiller *et al* were able to detect altered association of morphogens to the insoluble matrix upon injury. The authors also assessed whether proteins changed in terms of solubility and indeed the BM protein netrin 1 became more soluble with injury without an overall change in abundance.

The recovery after bleomycin-induced injury MS was used to investigate the resolution of fibrogenesis [23]. 154 ECM proteins had changes in abundance during injury/fibrogenesis compared to the resolution phase. This analysis revealed emilin-2 and collagen XXVIII as putative constituents of the repair ECM. Furthermore, in order to correlate temporal MS protein

levels with the degree of fibrosis/stiffness, the authors used a lung compliance test as a surrogate for lung fibrosis/stiffness. Decreased compliance is indicative of increased lung fibrosis/stiffening. In this analysis 54 matrixome proteins had a significant correlation with lung compliance. For example, BM proteins such as type IV collagen  $\alpha 5$  correlated positively with compliance, whereas tenascin C inversely correlated with compliance. These studies highlight the importance of BM proteins for lung function, lung compliance and for the resolution of injury.

### **Expression and interactions of basement membrane proteins**

By combining whole normal human tissue proteomic datasets, it is possible to map the expression of BM proteins. From these open access proteomic datasets we selected BM proteins and identified the expression of 71 BM proteins across 39 tissues (Figure 1). Some BM proteins were detected in every tissue: COL18A1, FBLN1, FN1 and LAMB2 and many proteins were present in only one or two tissues: AGRN, COL4A2, EFEMP1, HSPG2, LAMB1, LAMC1, NID1 and TGFBI. However, because these datasets were not generated with a focus on the ECM, a number of BM components that one would expect to be detected in the same tissue (given known chain combinations), were not co-detected. For example, although COL4A2 was detected in 37/39 tissues, COL4A1 was only detected in 30/39 tissues. This is in contrast to datasets that have used ECM enrichment strategies. Using datasets that used ECM enrichment from normal tissues, the expression of 62 BM components could be mapped (Figure 3 and 4). Here, there were fewer (11) available datasets. COL15A1, COL18A1, COL4A1, COL4A2, FN1, HSPG2, LAMA2, LAMA5, LAMB2, LAMC1, NID1, NID2, TGFBI and TNC were detected in every dataset and AGRN, COL4A4, COL4A6, COL5A1, FBN1, LAMA2, LAMA3, LAMA4, LAMB1, TIMP3 and VWA1 were detected in 10/11 datasets. Thus highlighting the core BM components.

In whole tissue proteomic datasets, the network-forming COL4A3, COL4A4, COL4A5 and COL4A6 proteins were detected in between 13-33% of tissues; thus, these subunits appear

to have a highly tissue restricted expression pattern. In contrast, ECM-enriched datasets reveal that these proteins are detected in 80-90% tissues. The same is true for LAMA2, LAMA3 and LAMA4, which appear to have restricted expression in whole proteomic datasets, but are detected in 90% of ECM enriched datasets. This could reflect the lower abundance of these laminin and collagen isoforms in comparison to the more abundant and ubiquitous COL4A1, COL4A2, LAMA5, LAMB1 and LAMB2 isoforms. Also of note, the number of tissues where a BM protein is expressed correlates with the number of known protein-protein interactions that the protein forms (Figure 4).

## **Future directions**

### **Glycoproteomics**

MS-based studies have started to reveal the true extent of protein post-translational modifications (PTMs). Thousands of modifications can now be characterised and localised to specific protein sequences. The quantification of PTM levels across different biological samples is also possible with label-free methods [69]. Furthermore, it is possible to record the occupancy or stoichiometry of PTM sites on a large scale. However, a complete inventory of sites has not been established for any PTM. Glycosylation is particularly prevalent in BM proteins and these modifications are essential for the normal function of BMs. Although it is now possible to define modification sites of proteins, determining the structure of glycosylation is a complex problem. N-linked glycosylation modifications can be extremely heterogeneous, and glycosylated tryptic peptides can be too large and hydrophilic to be captured by conventional reversed-phase liquid chromatography. One option to avoid this problem is to use proteases with different cleavage specificities such as chymotrypsin, Lys-N, or endoproteinase Glu-C in addition to trypsin and these approaches may yield further insight into the extent and nature of BM protein glycosylation.

Glycocapture is a technique for enriching glycoproteins for analysis. These techniques often involve labelling glycan groups such as using metabolic labelling with chemoselective ligation,

such as click chemistry. In the first description of click chemistry cycloaddition of an azide and alkyne for metabolic labelling of glycans required the use of copper (I) ions, which are cytotoxic at high concentration [70]. In further studies, copper-free ligation of azide and alkyne, in addition to Staudinger ligation between azide and phosphine, have been developed for labelling glycans in live cells [71, 72]. Other approaches to enrich for glycoproteins include lectin chromatography [73], titanium dioxide enrichment [74, 75], solid-phase extraction of N-linked glycopeptides [76], and the use of hydrazide chemistry to modify glycans for subsequent enrichment [77, 78].

### **Alternative digestion strategies**

Collagen IV is the most abundant structural protein in BMs. Tissue specific roles for type IV collagen  $\alpha3\alpha4\alpha5$  and broader expression of  $\alpha1\alpha1\alpha2$  are reflected in the phenotypes of human diseases caused by genetic mutations on type IV collagen [79, 80]. Type IV collagen has a large number of PTMs, including hydroxylation of proline and lysine and glycosylation of lysine residues, which contribute to the structural features required for its different biological functions. Both glycosylation and hydroxylation of type IV collagen was investigated by Vanacore and colleagues [81]. In order to characterise PTM sites in type IV collagen comprehensively by MS, the authors needed to overcome a number of limiting factors. Firstly, glycosylated lysine residues do not undergo tryptic cleavage. Secondly, *in silico* analyses suggested that trypsin, the most commonly used proteolytic enzyme for MS analyses, would generate very long and heavily modified collagen IV peptides, which are not easily detectable through conventional MS-based identification. This problem is evident in many proteomic analyses of BM proteins, in which collagen IV is reported with low abundance and low sequence coverage. Through *in silico* prediction the endoproteinase Glu-C produced the best results in terms of shorter peptide length and number of PTMs per peptide compared with trypsin.

A multilevel strategy for the analysis of mouse Col4a1, which included either trypsin or Glu-C digested Col4a1, and three different fragmentation approaches, either collision induced dissociation (CID), electron transfer dissociation (ETD) or higher energy C-trap dissociation (HCD). All three fragmentation strategies produced complementary information. The same parent ion analysed with the different fragmentation approaches enabled primary structure information to be inferred by one method where the other method did not provide informative data. By using these complementary methods, 82% of sequence coverage was achieved for the 7S domain and 81.5% for the full-length Col4a1, mapping 39 glycosylated hydroxylysine residues, 148 4-hydroxyproline residues, and seven 3-hydroxyproline residues. Additional deglycosylation experiments with PNGase F, revealed a single N-glycosylation site in the 7S domain. These findings revealed the extent of collagen IV modification and lay the foundation for dissecting the key role of these modifications in health and disease.

### **Degradomics**

Within BMs proteolytic activity is crucial because new BM components are synthesised at a relatively rapid rate even in normal tissues [66] and therefore old components need to be degraded to prevent pathological accumulation of BM components. Proteolytic activity within the ECM is dysregulated in disease, for instance loss of TIMP3 is associated with the rupture of atherosclerotic plaques [82, 83]. Within tissues, proteomics can be used to detect the presence of peptide fragments, which in turn can be used to infer the activity of specific proteolytic enzymes [52, 54]. Proteomics can also be utilised for global analysis of protease cleavage site analysis, and this approach is also known as degradomics. With this technique information about both the protein targets and sites of cleavage for a specific protease can be revealed. These approaches rely on either positive or negative selective enrichment of N-termini generated by modifying free  $\alpha$ -amines and the neo-N-termini of protease cleavage products [84-88].

Terminal amine isotopic labelling of substrates (TAILS) is a three-step quantitative proteomics approach for labelling and isolating N-terminal peptides before and after exposure to a protease of interest. TAILS has been applied to matrix metalloproteinase (MMP)-2 cleavage sites in mouse fibroblast secretomes, revealing 288 sites [89]. Using TAILS more than 3000 cleavage sites were identified for five different astacins. Meprin  $\alpha$  and meprin  $\beta$  proteases exhibited a strong preference for aspartate in the peptide (P)1 position because of a conserved positively charged residue in the active cleft subsite (S)1. This unparalleled specificity has not been found for other families of extracellular proteases. Interestingly, cleavage specificity is also strongly influenced by proline in P2 or P3 leading to a rare example of subsite cooperativity. A further example is the cleavage site profiling of the proteases ADAM10 and ADAM17, which have been shown to play important roles in inflammation, cancer, and neurodegenerative disorders [90]. The cleavage site specificities of ADAM10 and ADAM17 were shown to vary at the P1' site: while both enzymes cleave at leucine, only ADAM10 expresses preference toward aromatic amino acids, whereas ADAM17 exhibits the highest preference for valine, in good agreement with ADAM10/17 cleavage sites previously identified [91].

A combined approach of iTRAQ-based terminal amine isotopic labeling of substrates (iTRAQ-TAILS) with SILAC was utilised by Keller and colleagues to assign proteins to a specific cell type by MS1- and their cleavage by MS2-based quantification in the same experiment. By using this hybrid technique they investigated the matrix metalloproteinase MMP10-dependent processing of extracellular proteins derived from either keratinocyte or fibroblast secretomes to give insight into processing of proteins such as collagens and laminins at the dermal epidermal interface [92]. In a subsequent study the same group applied the same methodology and identified integrin  $\alpha 6$ , cysteine-rich angiogenic inducer 61 and dermokine as novel MMP10 targets. Further experiments validating this data using transgenic mice overexpressing MMP10 suggested that MMP10 modulates local adhesion and migration during wound healing and cancer development [93]. A greater understanding about BM degradation in disease



processes, through the identification of cleavage sites and peptide fragments also has the potential for identifying new disease biomarkers [94].

## **Conclusion**

The field of MS-based proteomics has enormous potential for systems-level analyses of biological systems, including the complex extracellular niche of BMs. Recent developments mean that not only can proteins be detected and quantified, but their PTMs mapped and degradation products measured. Many of the studies described above have applied these techniques generating data-rich information about their respective tissue systems. In turn, these data have lead researchers to important discoveries in BM biology. From these complex datasets emerge several key questions, what is a bona fide BM proteins, and how is the BM regulated and turned over? These same proteomics approaches and their future refinements will provide the tools to systematically examine this complexity. Proteomic datasets also represent valuable resources for the wider research communities and increasingly investigators are depositing data in open access formats. Key to the interpretation of MS data is the bioinformatic analyses and here the interface between biologists and informaticians will be vital for the ongoing development of the field.

## **Acknowledgements**

MJR was supported by a project grant awarded to RL from Kidney Research UK (RP52/2014). RL is supported by a Wellcome Trust Intermediate Clinical Fellowship (090006). We also acknowledge core funding from the Wellcome Trust (088785/Z/09/Z) awarded to the Wellcome Trust Centre for Cell-Matrix Research, University of Manchester, UK.

## **Author contributions**

MJR designed the original figures, MJR, MJH and RL prepared the manuscript. All authors have approved the final article.

## Statement of competing financial interests

The authors declare no conflict of interest.

## References

- [1] J.H. Miner, C. Li, J.L. Mudd, G. Go, A.E. Sutherland, Compositional and structural requirements for laminin and basement membranes during mouse embryo implantation and gastrulation, *Development* 131(10) (2004) 2247-2256.
- [2] N. Smyth, H.S. Vatansever, P. Murray, M. Meyer, C. Frie, M. Paulsson, D. Edgar, Absence of basement membranes after targeting the LAMC1 gene results in embryonic lethality due to failure of endoderm differentiation, *The Journal of cell biology* 144(1) (1999) 151-60.
- [3] E. Poschl, U. Schlotzer-Schrehardt, B. Brachvogel, K. Saito, Y. Ninomiya, U. Mayer, Collagen IV is essential for basement membrane stability but dispensable for initiation of its assembly during early development, *Development* 131(7) (2004) 1619-28.
- [4] E. Hohenester, P.D. Yurchenco, Laminins in basement membrane assembly, *Cell adhesion & migration* 7(1) (2013) 56-63.
- [5] J. Khoshnoodi, V. Pedchenko, B.G. Hudson, Mammalian collagen IV, *Microscopy research and technique* 71(5) (2008) 357-70.
- [6] J.H. Miner, B.L. Patton, S.I. Lentz, D.J. Gilbert, W.D. Snider, N.A. Jenkins, N.G. Copeland, J.R. Sanes, The laminin alpha chains: expression, developmental transitions, and chromosomal locations of alpha1-5, identification of heterotrimeric laminins 8-11, and cloning of a novel alpha3 isoform, *The Journal of cell biology* 137(3) (1997) 685-701.
- [7] J.H. Miner, J.R. Sanes, Collagen IV alpha 3, alpha 4, and alpha 5 chains in rodent basal laminae: sequence, distribution, association with laminins, and developmental switches, *The Journal of cell biology* 127(3) (1994) 879-91.
- [8] R.O. Hynes, The extracellular matrix: not just pretty fibrils, *Science* 326(5957) (2009) 1216-9.
- [9] S. Li, D. Harrison, S. Carbonetto, R. Fassler, N. Smyth, D. Edgar, P.D. Yurchenco, Matrix assembly, regulation, and survival functions of laminin and its receptors in embryonic stem cell differentiation, *The Journal of cell biology* 157(7) (2002) 1279-90.
- [10] G. Raghu, L.J. Striker, L.D. Hudson, G.E. Striker, Extracellular matrix in normal and fibrotic human lungs, *Am Rev Respir Dis* 131(2) (1985) 281-9.
- [11] M. Zeisberg, Y. Maeshima, B. Mosterman, R. Kalluri, Renal fibrosis. Extracellular matrix microenvironment regulates migratory behavior of activated tubular epithelial cells, *Am J Pathol* 160(6) (2002) 2001-8.
- [12] C. Yang, M. Zeisberg, B. Mosterman, A. Sudhakar, U. Yerramalla, K. Holthaus, L. Xu, F. Eng, N. Afdhal, R. Kalluri, Liver fibrosis: insights into migration of hepatic stellate cells in response to extracellular matrix and growth factors, *Gastroenterology* 124(1) (2003) 147-59.
- [13] V.M. Monnier, D.R. Sell, R.H. Nagaraj, S. Miyata, S. Grandhee, P. Odetti, S.A. Ibrahim, Maillard reaction-mediated molecular damage to extracellular matrix and other tissue proteins in diabetes, aging, and uremia, *Diabetes* 41 Suppl 2 (1992) 36-41.
- [14] M. Brownlee, A. Cerami, H. Vlassara, Advanced glycosylation end products in tissue and the biochemical basis of diabetic complications, *N Engl J Med* 318(20) (1988) 1315-21.
- [15] G.P. Siegal, S.H. Barsky, V.P. Terranova, L.A. Liotta, Stages of neoplastic transformation of human breast tissue as monitored by dissolution of basement membrane components. An immunoperoxidase study, *Invasion Metastasis* 1(1) (1981) 54-70.
- [16] L.A. Liotta, J. Kleinerman, P. Catanzaro, D. Rynbrandt, Degradation of basement membrane by murine tumor cells, *J Natl Cancer Inst* 58(5) (1977) 1427-31.
- [17] M. Egeblad, M.G. Rasch, V.M. Weaver, Dynamic interplay between the collagen scaffold and tumor evolution, *Curr Opin Cell Biol* 22(5) (2010) 697-706.

- [18] L. Bruckner-Tuderman, Collagen VII and bullous disorders of the skin, *Dermatology* 189 Suppl 2 (1994) 16-20.
- [19] D.F. Barker, S.L. Hostikka, J. Zhou, L.T. Chow, A.R. Oliphant, S.C. Gerken, M.C. Gregory, M.H. Skolnick, C.L. Atkin, K. Tryggvason, Identification of mutations in the COL4A5 collagen gene in Alport syndrome, *Science* 248(4960) (1990) 1224-7.
- [20] M. Zenker, T. Aigner, O. Wendler, T. Tralau, H. Muntefering, R. Fenski, S. Pitz, V. Schumacher, B. Royer-Pokora, E. Wuhl, P. Cochat, R. Bouvier, C. Kraus, K. Mark, H. Madlon, J. Dotsch, W. Rascher, I. Maruniak-Chudek, T. Lennert, L.M. Neumann, A. Reis, Human laminin beta2 deficiency causes congenital nephrosis with mesangial sclerosis and distinct eye abnormalities, *Hum Mol Genet* 13(21) (2004) 2625-32.
- [21] A. Naba, K.R. Clauser, H. Ding, C.A. Whittaker, S.A. Carr, R.O. Hynes, The extracellular matrix: Tools and insights for the "omics" era, *Matrix Biol* 49 (2016) 10-24.
- [22] T.C. Walther, M. Mann, Mass spectrometry-based proteomics in cell biology, *Journal of Cell Biology* 190(4) (2010) 491-500.
- [23] H.B. Schiller, I.E. Fernandez, G. Burgstaller, C. Schaab, R.A. Scheltema, T. Schwarzmayr, T.M. Strom, O. Eickelberg, M. Mann, Time- and compartment-resolved proteome profiling of the extracellular niche in lung injury and repair, *Molecular systems biology* 11(7) (2015) 819.
- [24] V. Kuttner, C. Mack, K.T. Rigbolt, J.S. Kern, O. Schilling, H. Busch, L. Bruckner-Tuderman, J. Dengjel, Global remodelling of cellular microenvironment due to loss of collagen VII, *Molecular systems biology* 9 (2013) 657.
- [25] M.J. Randles, A.S. Woolf, J.L. Huang, A. Byron, J.D. Humphries, K.L. Price, M. Kolatsi-Joannou, S. Collinson, T. Denny, D. Knight, A. Mironov, T. Starborg, R. Korstanje, M.J. Humphries, D.A. Long, R. Lennon, Genetic Background is a Key Determinant of Glomerular Extracellular Matrix Composition and Organization, *J Am Soc Nephrol* 26(12) (2015) 3021-3034.
- [26] R. Aebersold, M. Mann, Mass spectrometry-based proteomics, *Nature* 422(6928) (2003) 198-207.
- [27] J.B. Fenn, M. Mann, C.K. Meng, S.F. Wong, C.M. Whitehouse, Electrospray ionization for mass spectrometry of large biomolecules, *Science* 246(4926) (1989) 64-71.
- [28] D.F. Hunt, J.R. Yates, 3rd, J. Shabanowitz, S. Winston, C.R. Hauer, Protein sequencing by tandem mass spectrometry, *Proc Natl Acad Sci U S A* 83(17) (1986) 6233-7.
- [29] A. Shevchenko, M. Wilm, O. Vorm, M. Mann, Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels, *Anal Chem* 68(5) (1996) 850-8.
- [30] M. Heller, P.E. Michel, P. Morier, D. Crettaz, C. Wenz, J.D. Tissot, F. Reymond, J.S. Rossier, Two-stage Off-Gel isoelectric focusing: protein followed by peptide fractionation and application to proteome analysis of human plasma, *Electrophoresis* 26(6) (2005) 1174-88.
- [31] A. Naba, K.R. Clauser, S. Hoersch, H. Liu, S.A. Carr, R.O. Hynes, The matrisome: in silico definition and in vivo characterization by proteomics of normal and tumor extracellular matrices, *Mol Cell Proteomics* 11(4) (2012) M111 014647.
- [32] D.N. Perkins, D.J. Pappin, D.M. Creasy, J.S. Cottrell, Probability-based protein identification by searching sequence databases using mass spectrometry data, *Electrophoresis* 20(18) (1999) 3551-67.
- [33] J.K. Eng, A.L. McCormack, J.R. Yates, An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database, *Journal of the American Society for Mass Spectrometry* 5(11) (1994) 976-89.
- [34] R. Craig, R.C. Beavis, TANDEM: matching proteins with tandem mass spectra, *Bioinformatics* 20(9) (2004) 1466-1467.
- [35] J. Cox, M. Mann, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification, *Nat Biotechnol* 26(12) (2008) 1367-72.
- [36] J. Cox, N. Neuhauser, A. Michalski, R.A. Scheltema, J.V. Olsen, M. Mann, Andromeda: a peptide search engine integrated into the MaxQuant environment, *J Proteome Res* 10(4) (2011) 1794-805.

- [37] N. Colaert, H. Barsnes, M. Vaudel, K. Helsens, E. Timmerman, A. Sickmann, K. Gevaert, L. Martens, Thermo-msf-parser: an open source Java library to parse and visualize Thermo Proteome Discoverer msf files, *J Proteome Res* 10(8) (2011) 3840-3.
- [38] S.E. Ong, B. Blagoev, I. Kratchmarova, D.B. Kristensen, H. Steen, A. Pandey, M. Mann, Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics, *Molecular & Cellular Proteomics* 1(5) (2002) 376-386.
- [39] S. Zanivan, M. Krueger, M. Mann, In vivo quantitative proteomics: the SILAC mouse, *Methods in molecular biology* 757 (2012) 435-50.
- [40] P.L. Ross, Y.L.N. Huang, J.N. Marchese, B. Williamson, K. Parker, S. Hattan, N. Khainovski, S. Pillai, S. Dey, S. Daniels, S. Purkayastha, P. Juhasz, S. Martin, M. Bartlett-Jones, F. He, A. Jacobson, D.J. Pappin, Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents, *Molecular & Cellular Proteomics* 3(12) (2004) 1154-1169.
- [41] H.B. Liu, R.G. Sadygov, J.R. Yates, A model for random sampling and estimation of relative protein abundance in shotgun proteomics, *Analytical chemistry* 76(14) (2004) 4193-4201.
- [42] Y. Ishihama, Y. Oda, T. Tabata, T. Sato, T. Nagasu, J. Rappsilber, M. Mann, Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein, *Molecular & cellular proteomics : MCP* 4(9) (2005) 1265-72.
- [43] J.C. Silva, R. Denny, C.A. Dorschel, M. Gorenstein, I.J. Kass, G.Z. Li, T. McKenna, M.J. Nold, K. Richardson, P. Young, S. Geromanos, Quantitative proteomic analysis by accurate mass retention time pairs, *Analytical chemistry* 77(7) (2005) 2187-200.
- [44] N.M. Griffin, J. Yu, F. Long, P. Oh, S. Shore, Y. Li, J.A. Koziol, J.E. Schnitzer, Label-free, normalized quantification of complex mass spectrometry data for proteomic analysis, *Nature biotechnology* 28(1) (2010) 83-9.
- [45] E. Chautard, L. Ballut, N. Thierry-Mieg, S. Ricard-Blum, MatrixDB, a database focused on extracellular protein-protein and protein-carbohydrate interactions, *Bioinformatics* 25(5) (2009) 690-1.
- [46] A. Naba, K.R. Clauser, S. Hoersch, H. Liu, S.A. Carr, R.O. Hynes, The matrisome: in silico definition and in vivo characterization by proteomics of normal and tumor extracellular matrices, *Mol Cell Proteomics* 11(4) (2012) M111 014647.
- [47] G.R. Martin, H.K. Kleinman, V.P. Terranova, S. Ledbetter, J.R. Hassell, The regulation of basement membrane formation and cell-matrix interactions by defined supramolecular complexes, *Ciba Found Symp* 108 (1984) 197-212.
- [48] C. UniProt, The Universal Protein Resource (UniProt) in 2010, *Nucleic Acids Res* 38(Database issue) (2010) D142-8.
- [49] A. Naba, K.R. Clauser, R.O. Hynes, Enrichment of Extracellular Matrix Proteins from Tissues and Digestion into Peptides for Mass Spectrometry Analysis, *J Vis Exp* (101) (2015) e53057.
- [50] A. Didangelos, X. Yin, K. Mandal, M. Baumert, M. Jahangiri, M. Mayr, Proteomics characterization of extracellular space components in the human aorta, *Mol Cell Proteomics* 9(9) (2010) 2048-62.
- [51] M.D. Ross, L.A. Bruggeman, B. Hanss, M. Sunamoto, D. Marras, M.E. Klotman, P.E. Klotman, Podocan, a novel small leucine-rich repeat protein expressed in the sclerotic glomerular lesion of experimental HIV-associated nephropathy, *Journal of Biological Chemistry* 278(35) (2003) 33248-33255.
- [52] A. Didangelos, X. Yin, K. Mandal, A. Saje, A. Smith, Q. Xu, M. Jahangiri, M. Mayr, Extracellular matrix composition and remodeling in human abdominal aortic aneurysms: a proteomics approach, *Mol Cell Proteomics* 10(8) (2011) M111 008128.
- [53] J. Barallobre-Barreiro, A. Didangelos, F.A. Schoendube, I. Drozdov, X. Yin, M. Fernandez-Caggiano, P. Willeit, V.O. Puntmann, G. Aldama-Lopez, A.M. Shah, N. Domenech, M. Mayr, Proteomics analysis of cardiac extracellular matrix remodeling in a porcine model of ischemia/reperfusion injury, *Circulation* 125(6) (2012) 789-802.

- [54] J. Barallobre-Barreiro, R. Oklu, M. Lynch, M. Fava, F. Baig, X. Yin, T. Barwari, D.N. Potier, H. Albadawi, M. Jahangiri, K.E. Porter, M.T. Watkins, S. Misra, J. Stoughton, M. Mayr, Extracellular Matrix Remodeling in Response to Venous Hypertension: Proteomics of Human Varicose Veins, *Cardiovasc Res* (2016).
- [55] A. Naba, K.R. Clauser, J.M. Lamar, S.A. Carr, R.O. Hynes, Extracellular matrix signatures of human mammary carcinoma identify novel metastasis promoters, *Elife* 3 (2014) e01308.
- [56] A. Naba, K.R. Clauser, C.A. Whittaker, S.A. Carr, K.K. Tanabe, R.O. Hynes, Extracellular matrix signatures of human primary metastatic colon cancers and their metastases to liver, *Bmc Cancer* 14 (2014).
- [57] G. Uechi, Z. Sun, E.M. Schreiber, W. Halfter, M. Balasubramani, Proteomic View of Basement Membranes from Human Retinal Blood Vessels, Inner Limiting Membranes, and Lens Capsules, *J Proteome Res* (2014).
- [58] R. Lennon, A. Byron, J.D. Humphries, M.J. Randles, A. Carisey, S. Murphy, D. Knight, P.E. Brenchley, R. Zent, M.J. Humphries, Global Analysis Reveals the Complexity of the Human Glomerular Extracellular Matrix, *J Am Soc Nephrol* (2014).
- [59] A. Byron, M.J. Randles, J.D. Humphries, A. Mironov, H. Hamidi, S. Harris, P.W. Mathieson, M.A. Saleem, S.S. Satchell, R. Zent, M.J. Humphries, R. Lennon, Glomerular Cell Cross-Talk Influences Composition and Assembly of Extracellular Matrix, *J Am Soc Nephrol* (2014).
- [60] M.A. Morrissey, D.P. Keeley, E.J. Hagedorn, S.T. McClatchey, Q. Chi, D.H. Hall, D.R. Sherwood, B-LINK: a hemicentin, plakin, and integrin-dependent adhesion system that links tissues by connecting adjacent basement membranes, *Dev Cell* 31(3) (2014) 319-31.
- [61] G.C. Viberti, R.D. Hill, R.J. Jarrett, A. Argyropoulos, U. Mahmud, H. Keen, Microalbuminuria as a predictor of clinical nephropathy in insulin-dependent diabetes mellitus, *Lancet* 1(8287) (1982) 1430-2.
- [62] C.E. Mogensen, C.K. Christensen, Predicting diabetic nephropathy in insulin-dependent patients, *The New England journal of medicine* 311(2) (1984) 89-93.
- [63] D.A. Long, M. Kolatsi-Joannou, K.L. Price, C. Dessapt-Baradez, J.L. Huang, E. Papakrivopoulou, M. Hubank, R. Korstanje, L. Gnudi, A.S. Woolf, Albuminuria is associated with too few glomeruli and too much testosterone, *Kidney Int* 83(6) (2013) 1118-29.
- [64] M.J. Randles, A.S. Woolf, J.L. Huang, A. Byron, J.D. Humphries, K.L. Price, M. Kolatsi-Joannou, S. Collinson, T. Denny, D. Knight, A. Mironov, T. Starborg, R. Korstanje, M.J. Humphries, D.A. Long, R. Lennon, Genetic Background is a Key Determinant of Glomerular Extracellular Matrix Composition and Organization, *J Am Soc Nephrol* 26(12) (2015) 3021-34.
- [65] T.R. Cox, J.T. Erler, Remodeling and homeostasis of the extracellular matrix: implications for fibrotic diseases and cancer, *Dis Model Mech* 4(2) (2011) 165-78.
- [66] M.L. Decaris, M. Gatmaitan, S. FlorCruz, F. Luo, K. Li, W.E. Holmes, M.K. Hellerstein, S.M. Turner, C.L. Emson, Proteomic analysis of altered extracellular matrix turnover in bleomycin-induced pulmonary fibrosis, *Mol Cell Proteomics* 13(7) (2014) 1741-52.
- [67] N.M. Nguyen, R.M. Senior, Laminin isoforms and lung development: all isoforms are not equal, *Dev Biol* 294(2) (2006) 271-9.
- [68] D.R. Abrahamson, K. Isom, E. Roach, L. Stroganova, A. Zelenchuk, J.H. Miner, P.L. St John, Laminin compensation in collagen alpha3(IV) knockout (Alport) glomeruli contributes to permeability defects, *J Am Soc Nephrol* 18(9) (2007) 2465-72.
- [69] J.V. Olsen, M. Mann, Status of Large-scale Analysis of Post-translational Modifications by Mass Spectrometry, *Molecular & Cellular Proteomics* 12(12) (2013) 3444-3452.
- [70] S.T. Laughlin, N.J. Agard, J.M. Baskin, I.S. Carrico, P.V. Chang, A.S. Ganguli, M.J. Hangauer, A. Lo, J.A. Prescher, C.R. Bertozzi, Metabolic labeling of glycans with azido sugars for visualization and glycoproteomics, *Methods Enzymol* 415 (2006) 230-50.
- [71] J.M. Baskin, J.A. Prescher, S.T. Laughlin, N.J. Agard, P.V. Chang, I.A. Miller, A. Lo, J.A. Codelli, C.R. Bertozzi, Copper-free click chemistry for dynamic in vivo imaging, *Proc Natl Acad Sci U S A* 104(43) (2007) 16793-7.

- [72] E. Saxon, C.R. Bertozzi, Cell surface engineering by a modified Staudinger reaction, *Science* 287(5460) (2000) 2007-10.
- [73] T. Plavina, E. Wakshull, W.S. Hancock, M. Hincapie, Combination of abundant protein depletion and multi-lectin affinity chromatography (M-LAC) for plasma protein biomarker discovery, *J Proteome Res* 6(2) (2007) 662-71.
- [74] M.R. Larsen, S.S. Jensen, L.A. Jakobsen, N.H. Heegaard, Exploring the sialome using titanium dioxide chromatography and mass spectrometry, *Mol Cell Proteomics* 6(10) (2007) 1778-87.
- [75] B.L. Parker, P. Gupta, S.J. Cordwell, M.R. Larsen, G. Palmisano, Purification and identification of O-GlcNAc-modified peptides using phosphate-based alkyne CLICK chemistry in combination with titanium dioxide chromatography and mass spectrometry, *J Proteome Res* 10(4) (2011) 1449-58.
- [76] Y. Tian, K.S. Kelly-Spratt, C.J. Kemp, H. Zhang, Mapping tissue-specific expression of extracellular proteins using systematic glycoproteomic analysis of different mouse tissues, *J Proteome Res* 9(11) (2010) 5837-47.
- [77] H. Zhang, E.C. Yi, X.J. Li, P. Mallick, K.S. Kelly-Spratt, C.D. Masselon, D.G. Camp, 2nd, R.D. Smith, C.J. Kemp, R. Aebersold, High throughput quantitative analysis of serum proteins using glycopeptide capture and liquid chromatography mass spectrometry, *Mol Cell Proteomics* 4(2) (2005) 144-55.
- [78] B. Wollscheid, D. Bausch-Fluck, C. Henderson, R. O'Brien, M. Bibel, R. Schiess, R. Aebersold, J.D. Watts, Mass-spectrometric identification and relative quantification of N-linked cell surface glycoproteins, *Nat Biotechnol* 27(4) (2009) 378-86.
- [79] B.G. Hudson, The molecular basis of Goodpasture and Alport syndromes: beacons for the discovery of the collagen IV family, *J Am Soc Nephrol* 15(10) (2004) 2514-27.
- [80] E. Plaisier, O. Gribouval, S. Alamowitch, B. Mougnot, C. Prost, M.C. Verpont, B. Marro, T. Desmetre, S.Y. Cohen, E. Roullet, M. Dracon, M. Fardeau, T. Van Agtmael, D. Kerjaschki, C. Antignac, P. Ronco, COL4A1 mutations and hereditary angiopathy, nephropathy, aneurysms, and muscle cramps, *N Engl J Med* 357(26) (2007) 2687-95.
- [81] T. Basak, L. Vega-Montoto, L.J. Zimmerman, D.L. Tabb, B.G. Hudson, R.M. Vanacore, Comprehensive Characterization of Glycosylation and Hydroxylation of Basement Membrane Collagen IV by High-Resolution Mass Spectrometry, *J Proteome Res* 15(1) (2016) 245-58.
- [82] M. Cardellini, R. Menghini, E. Martelli, V. Casagrande, A. Marino, S. Rizza, O. Porzio, A. Mauriello, A. Solini, A. Ippoliti, R. Lauro, F. Folli, M. Federici, TIMP3 is reduced in atherosclerotic plaques from subjects with type 2 diabetes and increased by SirT1, *Diabetes* 58(10) (2009) 2396-401.
- [83] R. Stohr, M. Cavalera, S. Menini, M. Mavilio, V. Casagrande, C. Rossi, A. Urbani, M. Cardellini, G. Pugliese, R. Menghini, M. Federici, Loss of TIMP3 exacerbates atherosclerosis in ApoE null mice, *Atherosclerosis* 235(2) (2014) 438-43.
- [84] K. Gevaert, M. Goethals, L. Martens, J. Van Damme, A. Staes, G.R. Thomas, J. Vandekerckhove, Exploring proteomes and analyzing protein processing by mass spectrometric identification of sorted N-terminal peptides, *Nature biotechnology* 21(5) (2003) 566-9.
- [85] L. McDonald, D.H. Robertson, J.L. Hurst, R.J. Beynon, Positional proteomics: selective recovery and analysis of N-terminal proteolytic peptides, *Nature methods* 2(12) (2005) 955-7.
- [86] C. Ji, N. Guo, L. Li, Differential dimethyl labeling of N-termini of peptides after guanidination for proteome analysis, *Journal of proteome research* 4(6) (2005) 2099-108.
- [87] J.C. Timmer, M. Enoksson, E. Wildfang, W. Zhu, Y. Igarashi, J.B. Denault, Y. Ma, B. Dummitt, Y.H. Chang, A.E. Mast, A. Eroshkin, J.W. Smith, W.A. Tao, G.S. Salvesen, Profiling constitutive proteolytic events in vivo, *The Biochemical journal* 407(1) (2007) 41-8.
- [88] S. Mahrus, J.C. Trinidad, D.T. Barkan, A. Sali, A.L. Burlingame, J.A. Wells, Global sequencing of proteolytic cleavage sites in apoptosis by specific labeling of protein N termini, *Cell* 134(5) (2008) 866-76.

- [89] O. Kleifeld, A. Doucet, U. auf dem Keller, A. Prudova, O. Schilling, R.K. Kainthan, A.E. Starr, L.J. Foster, J.N. Kizhakkedathu, C.M. Overall, Isotopic labeling of terminal amines in complex samples identifies protein N-termini and protease cleavage products, *Nature biotechnology* 28(3) (2010) 281-8.
- [90] P. Saftig, K. Reiss, The "A Disintegrin And Metalloproteases" ADAM10 and ADAM17: Novel drug targets with therapeutic potential?, *Eur J Cell Biol* 90(6-7) (2011) 527-535.
- [91] J. Tucher, D. Linke, T. Koudelka, L. Cassidy, C. Tredup, R. Wichert, C. Pietrzik, C. Becker-Pauly, A. Tholey, LC-MS based cleavage site profiling of the proteases ADAM10 and ADAM17 using proteome-derived peptide libraries, *J Proteome Res* 13(4) (2014) 2205-14.
- [92] P. Schlage, T. Kockmann, J.N. Kizhakkedathu, U. auf dem Keller, Monitoring matrix metalloproteinase activity at the epidermal-dermal interface by SILAC-iTRAQ-TAILS, *Proteomics* 15(14) (2015) 2491-502.
- [93] P. Schlage, T. Kockmann, F. Sabino, J.N. Kizhakkedathu, U. Auf dem Keller, Matrix Metalloproteinase 10 Degradomics in Keratinocytes and Epidermal Tissue Identifies Bioactive Substrates With Pleiotropic Functions, *Mol Cell Proteomics* 14(12) (2015) 3234-46.
- [94] F. Genovese, A.A. Manresa, D.J. Leeming, M.A. Karsdal, P. Boor, The extracellular matrix in the kidney: a source of novel non-invasive biomarkers of kidney fibrosis?, *Fibrogenesis Tissue Repair* 7(1) (2014) 4.
- [95] R. Lennon, A. Byron, J.D. Humphries, M.J. Randles, A. Carisey, S. Murphy, D. Knight, P.E. Brenchley, R. Zent, M.J. Humphries, Global analysis reveals the complexity of the human glomerular extracellular matrix, *J Am Soc Nephrol* 25(5) (2014) 939-51.
- [96] A. Naba, K.R. Clauser, C.A. Whittaker, S.A. Carr, K.K. Tanabe, R.O. Hynes, Extracellular matrix signatures of human primary metastatic colon cancers and their metastases to liver, *BMC Cancer* 14 (2014) 518.

## Figure Legends

**Figure 1: Basement membrane protein expression across different tissue systems.** 39 normal tissues, known to contain basement membranes, were selected from proteomicsDB (March 2016) and interrogated with proteins that are annotated as basement membrane by Gene Ontology (GO:0005604), with the following exceptions based on cross-reference with the human matrisome (March 2016) and UniProt database: ACHE, ALB, ANXA2, ANXA2P2, APLP1, CCDC80, CD151, DAG1, DLG1, DST, ENTPD2, GSTO1, RELL2, RPSA, RUNX1, SMC3SPN and USH2A. The heat map is clustered and based on Euclidean distance and displays approximate relative expression based on MS1 iBAQ abundance. Gene names are used for clarity.

**Figure 2: Overlap of *in vivo* and *in vitro* glomerular extracellular matrix (ECM) proteome.** Glomerular cells in culture produce basement membrane (BM) proteins, but both the composition and organization of BM proteins is more similar to *in vivo* when glomerular cells are cocultured. GEnCs, glomerular endothelial cells; GBM, glomerular basement membrane. Gene names are used for clarity.

**Figure 3: Analysis of ECM-enriched proteomic datasets.** ECM proteomic datasets from normal tissue [31, 50, 54, 57, 64, 95, 96] were screened for proteins that are annotated as basement membrane by Gene Ontology (GO:0005604), with the following exceptions based on cross-reference with the human matrisome (March 2016) and UniProt database: ACHE, ALB, ANXA2, ANXA2P2, APLP1, CCDC80, CD151, DAG1, DLG1, DST, ENTPD2, GSTO1, RELL2, RPSA, RUNX1, SMC3SPN and USH2A. Gene names are used for clarity.

**Figure 4: Interactions between basement membrane proteins.** Protein-protein interaction networks display basement membrane (BM) proteins detected in datasets from whole tissue (top) or ECM enriched samples (bottom). The colour and size of the nodes (circles) indicate



the number of tissue datasets a BM protein is detected in. The edges (lines) between nodes represent known protein-protein interactions. Box and whisker plots show the degree (number of protein-protein interactions) on the y-axis and number of tissues on the x-axis. Box plots indicate 25<sup>th</sup> and 75<sup>th</sup> percentiles (lower and upper bounds, respectively), 1.5× interquartile range (whiskers), individual data points (coloured diamonds), the median (black line) and mean (diamond within a black box). Gene names are used for clarity.

Figure 1

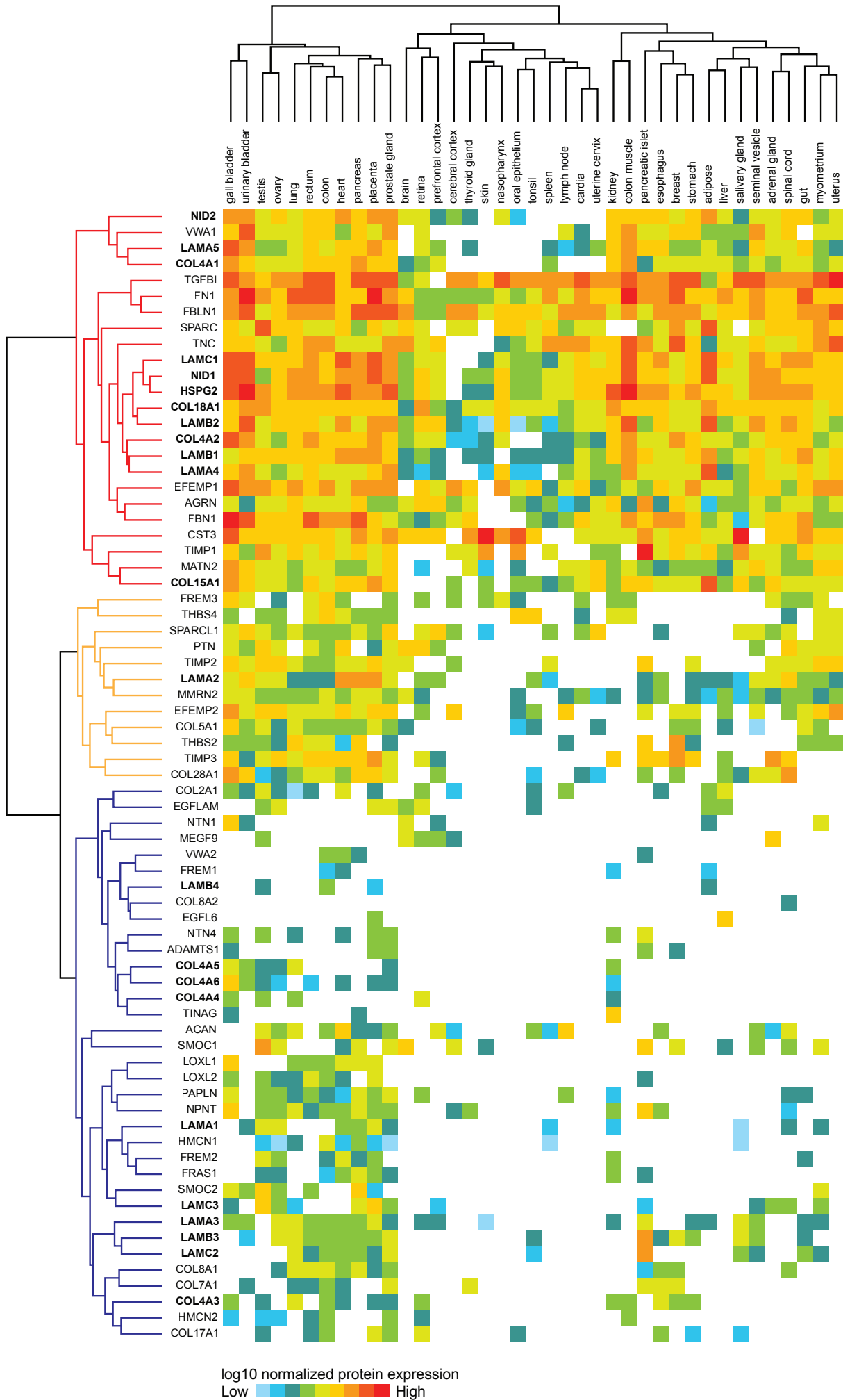


Figure 2

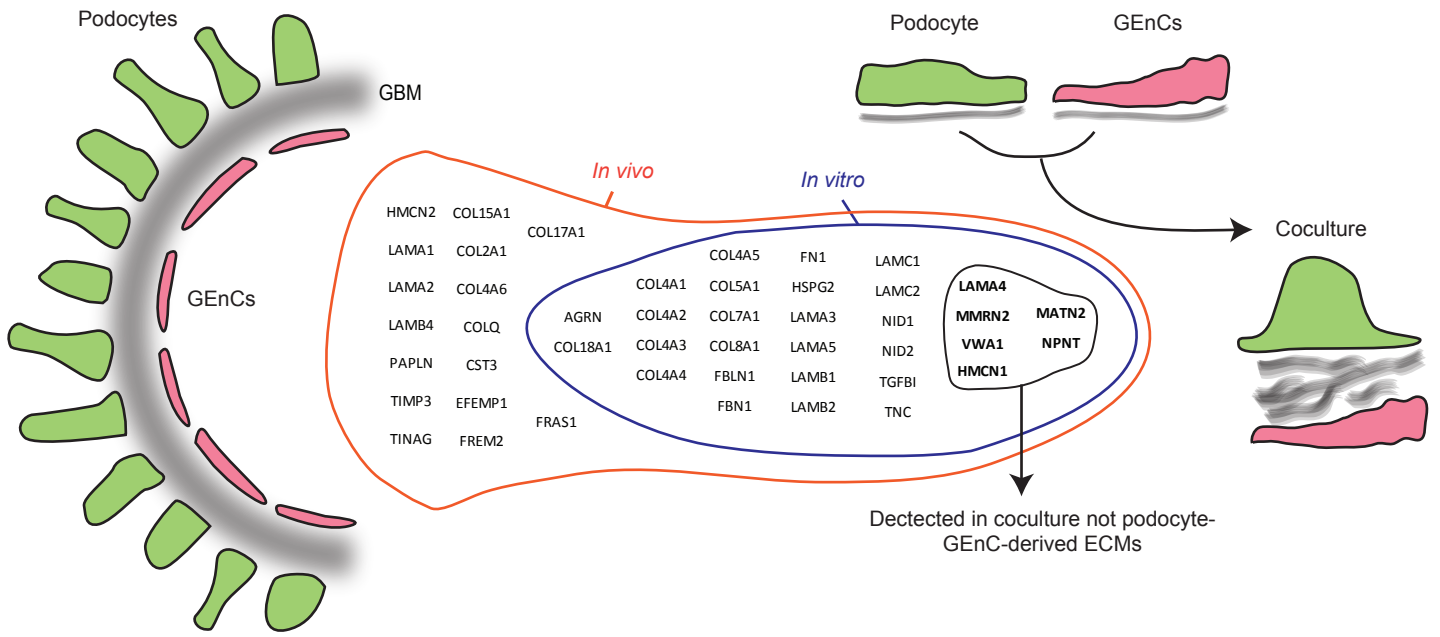
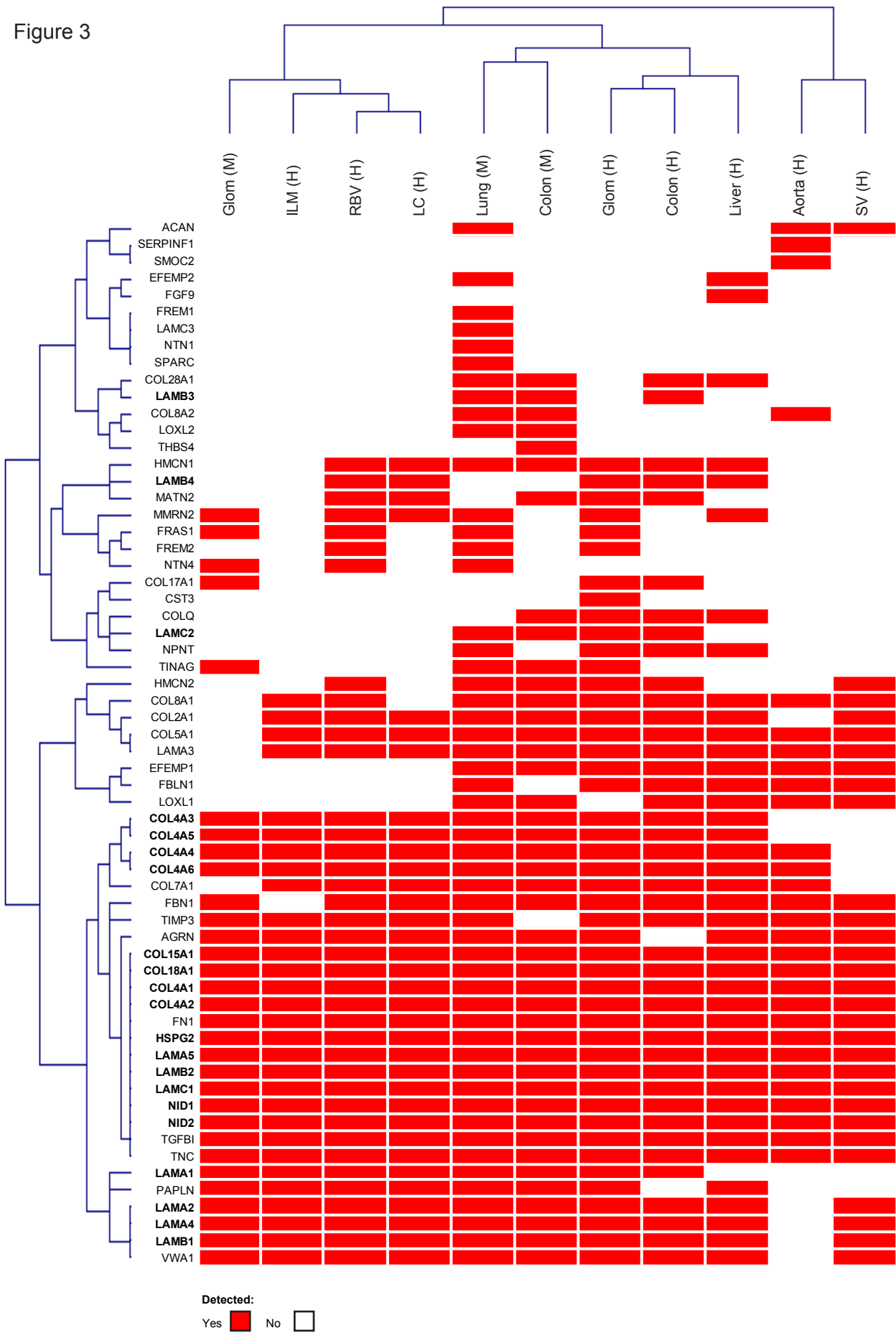


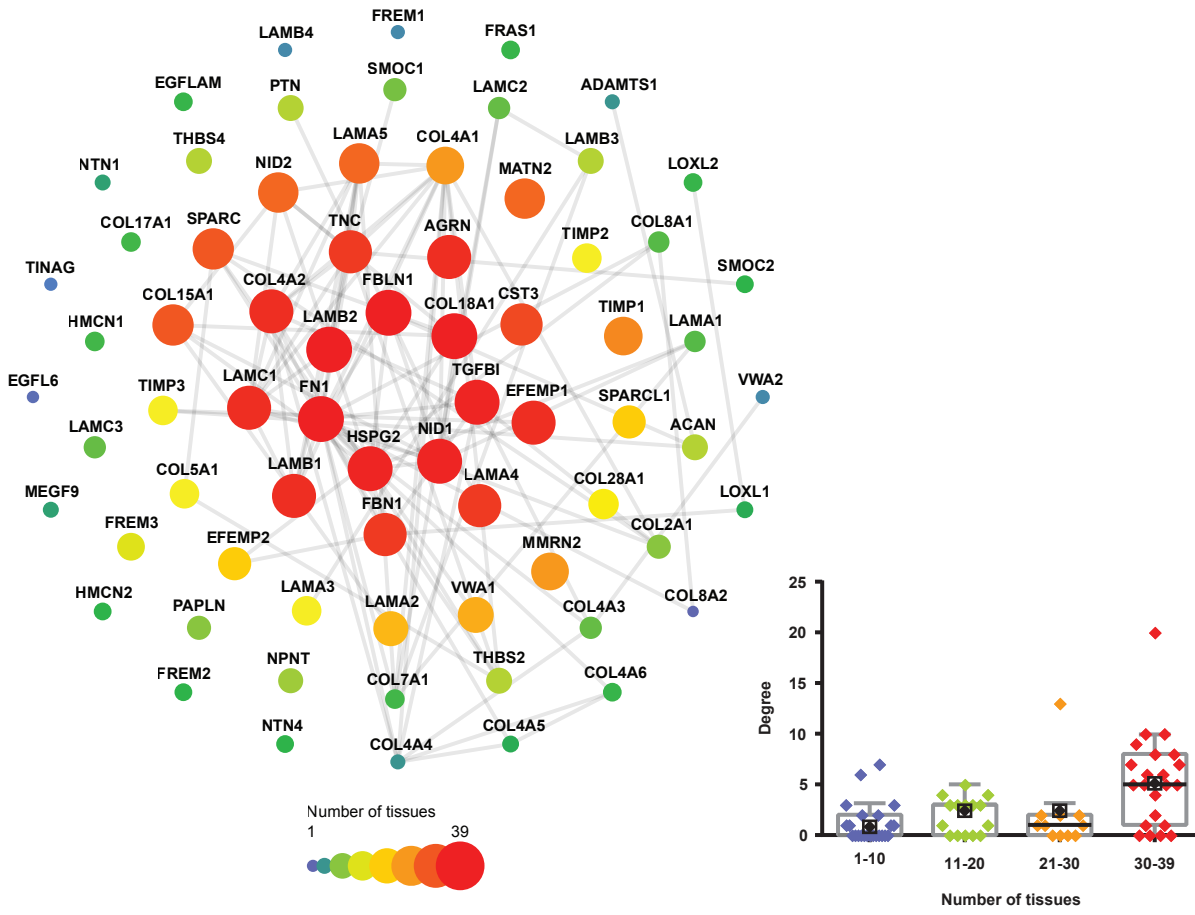
Figure 3



Extracellular matrix signatures of human primary metastatic colon cancers and their metastases to liver

Figure 4

Whole tissue proteomics



ECM enrichment proteomics

