PORTLAND PRESS

BIOCHEMICAL JOURNAL
ACCEPTED MANUSCRIPT

Comparative proteomic assessment of matrisome enrichment methodologies

Lukas Krasny, Angela Paul, Patty Wai, Beatrice A Howard, Rachael C Natrajan and Paul H Huang

The matrisome is a complex and heterogeneous collection of extracellular matrix (ECM) and ECM-associated proteins that play important roles in tissue development and homeostasis. While several strategies for matrisome enrichment have been developed, it is currently unknown how the performance of these different methodologies compares in the proteomic identification of matrisome components across multiple tissue types. In this study, we perform a comparative proteomic assessment of two widely used decellularisation protocols and two extraction methods to characterise the matrisome in four murine organs (heart, mammary gland, lung and liver). We undertook a systematic evaluation of the performance of the individual methods on protein yield, matrisome enrichment capability and the ability to isolate core matrisome and matrisome-associated components. Our data finds that SDS decellularisation leads to the highest matrisome enrichment efficiency while the extraction protocol that comprises chemical and trypsin digestion of the ECM fraction consistently identifies the highest number of matrisomal proteins across all tissue examined. Matrisome enrichment had a clear benefit over non-enriched tissue for the comprehensive identification of matrisomal components in murine liver and heart. Strikingly, we find that all four matrisome enrichment methods led to significant losses in the soluble matrisome-associated proteins across all organs. Our findings highlight the multiple factors (including tissue type, matrisome class of interest and desired enrichment purity) that influence the choice of enrichment methodology and we anticipate that this data will serve as a useful guide for the design of future proteomic studies of the matrisome.

Cite as Biochemical Journal (2016) DOI: 10.1042/BCJ2016XXXX

Comparative proteomic assessment of matrisome enrichment methodologies

Lukas Krasny[1], Angela Paul[2], Patty Wai[3,4], Beatrice A Howard[3], Rachael C Natrajan[3,4] and Paul H Huang[1,5].

Affiliations:

[1] Division of Cancer Biology, The Institute of Cancer Research, London, SW3 6JB, UK.

[2] Proteomics Core Facility, The Institute of Cancer Research, London, SW3 6JB, UK.

[3] The Breast Cancer Now Toby Robins Research Centre, Division of Breast Cancer Research, The Institute of Cancer Research, London, SW3 6JB, UK.

[4] Division of Molecular Pathology, The Institute of Cancer Research, London, SW3 6JB, UK.

[5] Correspondence to:

Paul H Huang
Division of Cancer Biology
Institute of Cancer Research
237 Fulham Road
London SW3 6JB
United Kingdom
Email: paul.huang@icr.ac.uk

Short Title: Comparative proteomic assessment of matrisome enrichment methodologies.

Keywords: Matrisome, mass spectrometry, proteomics, extracellular matrix, tissue extraction

Total number of figures: 7 main and 1 supplemental
Total number of tables: 1

Abstract

The matrisome is a complex and heterogeneous collection of extracellular matrix (ECM) and ECM-associated proteins that play important roles in tissue development and homeostasis. While several strategies for matrisome enrichment have been developed, it is currently unknown how the performance of these different methodologies compares in the proteomic identification of matrisome components across multiple tissue types. In this study, we perform a comparative proteomic assessment of two widely used decellularisation protocols and two extraction methods to characterise the matrisome in four murine organs (heart, mammary gland, lung and liver). We undertook a systematic evaluation of the performance of the individual methods on protein yield, matrisome enrichment capability and the ability to isolate core matrisome and matrisome-associated components. Our data finds that SDS decellularisation leads to the highest matrisome enrichment efficiency while the extraction protocol that comprises chemical and trypsin digestion of the ECM fraction consistently identifies the highest number of matrisomal proteins across all tissue examined. Matrisome enrichment had a clear benefit over non-enriched tissue for the comprehensive identification of matrisomal components in murine liver and heart. Strikingly, we find that all four matrisome enrichment methods led to significant losses in the soluble matrisome-associated proteins across all organs. Our findings highlight the multiple factors (including tissue type, matrisome class of interest and desired enrichment purity) that influence the choice of enrichment methodology and we anticipate that this data will serve as a useful guide for the design of future proteomic studies of the matrisome.

Summary statement

In this study, we evaluate the performance of four common matrisome enrichment methods in murine organs. We show that different methods have distinct benefits and shortcomings and that choice of enrichment protocol is highly dependent on the desired proteomic outcome.

Abbreviation list

ABC	Ammonium bicarbonate
ANOVA	Analysis of Variance
BCA	Bicinchoninic acid
CHAPS	3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate
DNAse I	Deoxyribonuclease I
DTT	Dithiothreitol
ECM	Extracellular matrix
EDTA	Ethylenediamintetraacetic acid
FA	Formic acid
FDR	False discovery rate
IAA	Iodacetamide
iECM	Fraction of ECM insoluble in 8M urea
KIU	Kallikrein inhibitor unit
LC-MS/MS	Liquid chromatography-tandem mass spectrometry
MS/MS	Tandem mass spectrometry
PBS	Polybuffered Saline
RNAse A	Ribonuclease A
SCID	Severe combined immunodeficiency disease
SDS	Sodium dodecyl sulfate
sECM	Fraction of ECM soluble in 8M urea
TCEP	Triscarboxyethyl phosphine
TFA	Tifluoracetic acid

Introduction

The extracellular matrix (ECM) in the tissue microenvironment is a complex and heterogeneous collection of proteins that play important roles in tissue development and homeostasis [1]. In addition to functioning as scaffolds that confer structural integrity in mammalian tissues, ECM components provide biochemical and biophysical cues which are transmitted by cell surface receptors, for instance the integrins, to trigger intracellular signalling events that regulate fundamental cell decisions including proliferation, death and differentiation [2, 3]. It is therefore unsurprising that dysregulation in ECM biology is associated with a number of pathologies such as cancer, fibrosis, atherosclerosis, arthritis and a range of genetic disorders [3-8]. To provide a consensus for the classification of this class of proteins, there has been a recent effort to develop an *in silico* definition of the ECM and its associated proteins, collectively designated as the "matrisome" [9]. The matrisome is composed of the "core matrisome" which consists of ECM glycoproteins, collagens and proteoglycans and the "matrisome-associated proteins" that include ECM regulators, affiliated proteins and secreted factors [10]. The early draft of the matrisome was intended to be inclusive and the subcategories within matrisome-associated proteins as designated by Naba et al. contain multiple proteins that are predicted to interact with the core matrisome [9]. It should therefore be noted that the definition of these subcategories remains loose and the interaction of many of these matrisome-associated components with the ECM remains to be experimentally confirmed.

While the matrisome provides an *in silico* framework for classifying ECMs and their associated proteins, the exact composition of the ECM components *in vivo* varies between tissues and organisms [10]. Comprehensive experimental characterisation of the matrisome using proteomics has been challenging in part because many ECM proteins are large, highly glycosylated proteins that are frequently insoluble due to the presence of covalent crosslinks [11]. Furthermore, the matrisome is composed of a large number of proteins with a wide dynamic range, necessitating the use of enrichment strategies [11]. Several approaches

have been developed to enrich for core matrisome components and preserve matrisome-associated proteins while removing contaminating intracellular proteins. These matrisome enrichment strategies are broadly split into methods that employ decellularisation of intact tissue and those that extract ECM from native or crude tissue homogenate [11].

Decellularisation involves the use of weak detergents, trypsinization or low ionic strength buffers to disrupt cell membranes leading to the release of intracellular proteins into solution while preserving the insoluble ECM scaffold [12]. This approach is commonly used in the generation of scaffolds for tissue engineering and has been applied to different tissue types including rat lung and human and porcine myocardium [13-15]. Methods that are routinely employed to extract ECM from native tissue homogenate include the use of differential detergent extractions to sequentially fractionate distinct cellular compartments (cytosolic, nuclear, membrane and cytoskeletal), leaving an insoluble fraction that contains the ECM [9]. This approach has been deployed to characterise the matrisome in murine lung and mammary gland as well as human colon and liver [9, 16-18]. Another strategy for extraction of ECM from native tissue homogenate involves the use of high-salt buffers to remove intracellular components followed by solubilisation of the remaining pellet in a chaotrope (usually urea). This method has been successfully utilised in multiple studies to catalogue ECM proteins in rat mammary gland [19, 20]. In an elaboration of this method, Hill *et al*. performed a further chemical digestion step to aid in the solubilisation of the urea-insoluble fraction prior to trypsin digestion, improving the recovery of matrisome components identified in rat lungs [13]. Treatment by alkaline detergent has been used to enrich matrisomal components from murine glomeruli [21, 22]. The alkaline detergent improves the solubilisation of cellular components and disrupts cell-ECM interactions. Additionally, a combination of sequential extraction and decellularisation techniques has successfully been applied to enrich for cardiac ECM [23]. In this protocol, homogenized samples were first treated with high-salt buffer and the resulting pellet subjected to decellularisation in SDS

solution. Decellularised samples were then homogenised in guanidium-HCl buffer prior to proteomic analysis.

Given the diverse strategies available for matrisome enrichment, it is currently unknown how the performance of the different methodologies compares in the proteomic identification of matrisome components across multiple tissue types. This knowledge will be necessary for the development of the proposed “ECM atlas” that seeks to compile the ECM composition of different tissues as defined by mass spectrometry [10]. In this study, we perform a comparative proteomic assessment of four matrisome enrichment methodologies (two decellularisation and two native extraction approaches) in four murine organs (heart, mammary gland, lung and liver). We sought to address two important questions: 1. Do different enrichment methodologies introduce bias or sample loss in the identification of core matrisome and matrisome-associated proteins within the same organ? 2. Does tissue type influence the recovery of matrisome components across the different extraction methods? Our data finds that while SDS decellularisation leads to high matrisome enrichment efficiency across all studied tissue types, the chemical digestion method by Hill *et al*. consistently identified more matrisome proteins in the four organs examined [13]. Moreover, we show a clear benefit of matrisome enrichment over unenriched tissue for the identification of ECM components in murine liver and heart. We find that Triton X decellularisation is superior in enriching for proteoglycan components of the core matrisome compared to other enrichment methods. Our data also points to significant losses in matrisome-associated proteins across all matrisome enrichment methodologies evaluated in this study compared to non-enriched samples. Our results demonstrate that the choice of matrisome enrichment methodology is a key factor in determining both the enrichment efficiency and the number and type of matrisome proteins identified *in vivo*.

Experimental Section

Animal models and tissue collection

All animal work was carried out under UK Home Office project and personal licenses following local ethical approval from The Institute of Cancer Research Ethics Committee and in accordance with local and national guidelines. Hearts, livers, lungs and the number 3 mammary glands were dissected from 10-week-old to 14-week-old virgin female SCID Beige mice. Tissues were snap frozen in liquid nitrogen directly after the surgery and stored at -80 °C.

Tissue processing and ECM protein enrichment

Tissues were cut into small pieces, weighed and placed into precooled tubes for decellularisation or sequential extraction. Three biological and two technical replicates were performed for each matrisome enrichment method.

Triton X-100 decellularisation (based on Xu et al. [12]): Samples (10 – 100 mg) were placed into PBS solution with 10 KIU/ml aprotinin (Sigma) and washed for 48 h at 4 °C. Samples were then decellularised with a solution of Tris-HCl (10 mM, pH 8), 3% Triton X-100 (Sigma), 0.1% ethylenediamine tetraacetic acid (EDTA, Sigma) and 10 KIU/ml aprotinin for 72 h at 4 °C. In the next step, samples were incubated in a solution of 0.2 µg/ml of ribonuclease A (RNAse A, Sigma) and 0.2 mg/ml deoxyribonuclease I (DNAse I, Sigma) in 50mM Tris-HCl (pH 7.4), 10 mM $MgCl_2$ (Sigma) and 5mM $CaCl_2$ (Sigma) for 24 h at 37 °C. After 24 h of washing in PBS at 4 °C, decellularised samples were homogenized in 250 µl of 8M urea (Sigma), 100 mM ammonium bicarbonate (ABC, Sigma) using LabGEN700 homogenizer (Cole-Parmer, UK), protein concentration was measured by bicinchoninic acid (BCA) assay (Thermo Scientific) and homogenates were frozen and stored at -80 °C until protein digestion. During all washing and decellularisation steps, samples were slowly rotated and solutions were refreshed every 24 h.

SDS decellularisation (based on Xu et al. [12]): Samples (10 – 100 µg) were placed into PBS solution with 10 KIU/ml aprotinin and washed for 48 h at 4 °C. Then, samples were decellularised with the solution of Tris-HCl (10 mM, pH 8), 0.5% sodium dodecyl sulfate (SDS, Sigma), 0.1% EDTA and 10 KIU/ml aprotinin for 72 h at 4 °C. Subsequent steps are the same as for decellularisation by Triton X-100. During all washing and decellularisation steps, samples were slowly rotated and solutions were refreshed every 24 h.

Extraction method A (based on Hill *et al.* [13]): Samples were homogenized in 0.5 ml of Tris-HCl (50mM, pH 7.4), 0.25% 3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate hydrate (CHAPS, Sigma), 25mM EDTA, 3M NaCl (Sigma) and 10 KIU/ml aprotinin by LabGEN700 homogenizer. Homogenized tissue was spun at 15,000 rpm, 4 °C and the supernatant was removed. The pellet was resuspended in 0.5 ml of the same homogenization buffer and additionally washed two times. The pellet was then washed by 0.3 ml of 8M urea, 100mM ABC, 25mM tris(2-carboxyethyl)phosphine (TCEP, Sigma) and spun. The resulting pellet was resuspended in 86% TFA (Sigma), 100mM CNBr (Sigma) and incubated at 25 °C overnight in the dark. After digestion, samples were washed three times by H_2O and dried by SpeedVac (Thermo Scientific, UK). After resuspending in solution of 8M urea, 100mM ABC, protein concentration was measured by the BCA assay and samples were stored in -80 °C until digestion by trypsin.

Extraction method B (based on by Naba *et al*. [9]): Samples were extracted using the Compartmental Extraction Kit (Millipore) as per manufacturer's instructions. Briefly, tissues were homogenized in cold buffer C using the LabGEN700 homogenizer, rotated at 4 °C for 20 min and spun at 15,000 rpm, 4 °C for 20 min. The pellet was then washed by cold buffer W, rotated for 5 min at 4 °C and spun. Nuclear proteins were removed by resuspending the pellet in cold buffer N. After rotating for 20 min at 4 °C and centrifugation, supernatant was removed and the pellet was resuspended in buffer M, rotated for 20 min at 4 °C and spun. CS buffer pre-warmed to room temperature was added to the pellet, rotated 20 min at room

temperature and spun. The pellet was immediately washed, resuspended in cold buffer C, rotated for 5 min at 4 °C and spun. The final pellet was resuspended in 8M urea, 100mM ABC, protein concentration was measured by BCA assay and stored in -80 °C. Protease inhibitors were added to all extraction buffers.

Non-enriched sample: Tissue was homogenized in 8M urea, 50 mM ABC, 75 mM NaCl with protease inhibitors by LabGEN700 homogenizer. Proteins were directly precipitated by 1.6 ml of ice cold acetone (VWR Chemicals) overnight at -20 °C. Precipitate was centrifuged at 15,000 rpm, 4 °C for 20 min and resuspended in 8M urea, 100mM ABC. Protein concentration was measured by BCA assay and samples were stored at -80 °C.

Protein digestion and sample preparation

20 µg of protein for each sample was reduced by 20mM dithiothreitol (DTT, Sigma), at 56 °C for 40 min and alkylated by 30mM iodoacetamide (IAA, Sigma) at 25 °C for 25 min in dark. After dilution to 2M Urea, 100mM ABC, all samples except those extracted by the Extraction method A protocol were deglycosylated by 0.2 µg of protein-N-glycosidase F (PNGase F, New England Biolabs). Samples extracted by Extraction method B protocol were digested by 0.2 µg of Endoproteinase Lys-C (Sigma) at 37 °C for 2 h and subsequently by 0.4 µg of trypsin (Life Technologies Ltd.) at 37 °C overnight. Other samples were directly digested by 0.4 µg of trypsin at 37 °C overnight. Digestion was stopped by acidification of the solution using 10% trifluoracetic acid (Sigma).

After digestion, samples were desalted by solid phase extraction using C18 OMIX tips (Agilent) or OPTI trap (Optimize Technologies), dried by SpeedVac and dissolved in 2% acetonitrile (ACN, VWR Chemicals)/0.1% formic acid (FA, Sigma). Concentration of peptides after digestion and desalting was measured by BCA assay.

Mass Spectrometry

Samples were analysed by liquid chromatography-tandem mass spectrometry (LC-MS/MS) using the LTQ Orbitrap Velos (Thermo Scientific) mass spectrometer coupled on-line with the Eksigent nanoLC automated system. 1 µg of sample was loaded onto Thermo Acclaim PepMap 100 C18 (100 µm x 2 cm, 5 µm) guard column and separated on in-house packed column (75 µm x 28 cm, Reprosil-Pur C18-AQ 3 µm particles) using the linear gradient from 5 to 40% ACN/0.1% FA in 88 min or 240 min for analysis of non-enriched samples with long gradient. The constant flow rate of 300 nl/min was used for all samples.

Mass spectra were acquired in data dependent mode using Thermo Xcalibur (ver. 2.2.42) software in mass range from m/z 375 to 2000. The twenty most intense precursor ions were selected for fragmentation by collision-induced dissociation with normalized collision energy 40 and dynamic exclusion for 40 sec within a mass range of ± 10 ppm and repeat count 1.

Data processing and analysis

MS/MS mass spectra of precursors with charge state of 2^+, 3^+ and 4^+ were extracted by Proteome Discoverer (ver. 1.4, Thermo Scientific). All MS/MS spectra were analysed using an in-house Mascot server (ver. 2.3.02, Matrix Science) which searched the Uniprot database selected for Mus musculus (16769 entries, downloaded 29/04/2015), with the enzyme specificity set as C-terminal lysine and arginine or CNBr+C-terminal lysine and arginine and 2 possible missed cleavages. Mascot was searched with a fragment ion mass tolerance of 0.6 Da and a parent ion tolerance of 10 ppm. Carbamidomethylation of cysteine was specified as a fixed modification while oxidation of lysine, methionine and proline and acetylation of lysine and the N-terminus were specified as variable modifications.

Scaffold (ver. 4.4.6, Proteome Software Inc.) was used to validate MS/MS based peptide and protein identifications. Peptide identifications were accepted if they could achieve an FDR less than 1.0% by the Scaffold Local FDR algorithm. Protein identifications were accepted if they could be established at greater than 95.0% probability and contained at

least 2 identified peptides. Protein probabilities were assigned by the Protein Prophet algorithm [24]. Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony.

Proteins identified by Mascot search and validated by Scaffold were then annotated using the mouse matrisome database MatrisomeDB from http://matrisomeproject.mit.edu/ (downloaded 21/10/2015).

Statistical analysis

"Percentage protein extraction yield" is defined as (amount of protein extracted from tissue)*100/(wet weight of the tissue). Statistical evaluation of percentage extraction yields across the different matrisome enrichment methods was performed by ANOVA with the Tukey corrected multiple comparisons and plots were generated using GraphPad Prism6 software (ver. 6.07). The value of $p < 0.05$ was considered statistically significant.

Datasets from both technical replicates were combined and only proteins identified in at least 2 of 3 biological replicates were considered. Proteins annotated using the MatrisomeDB were assigned as matrisomal and the "percentage matrisome enrichment" is defined as the (number matrisomal proteins)*100/(total number of proteins identified). The on-line tool Venny 2.1 (http://bioinfogp.cnb.csic.es/tools/venny/index.html) was used for comparison of identified matrisomal proteins and for generation of Venn diagrams.

Results and Discussion

The aim of this study was to evaluate four published protocols for matrisome enrichment in terms of 1) protein extraction yield, 2) the effectiveness of matrisome enrichment and 3) the number of unique matrisomal proteins identified by LC-MS/MS analysis. We utilised two decellularisation methods and two extraction methods to characterise four murine organs (heart, mammary gland, lung and liver). Identified matrisomal proteins were compared

across the four methods to determine the relative ability of each method to enrich for matrisomal components. All enrichment experiments were performed in 3 biological replicates with 2 technical repeats.

Matrisome enrichment workflows

Decellularisation by anionic (SDS) or non-ionic (Triton X-100) detergents are the most frequently used approaches in tissue engineering for the preparation of collagenous scaffolds [25]. After immersing the tissue in buffer with detergent to disrupt and remove cellular components, the decellularised scaffolds can then be homogenized in 8M urea and digested by proteases for proteomic analysis of matrisomal components (Fig. 1). However decellularisation techniques were originally developed for the purposes of tissue engineering where the complete removal of cellular proteins is required to avoid potential undesirable immune responses [26]. Such complete decellularisation requires at least several hours to days of extensive physical and chemical treatment which may lead to protein degradation.

In contrast to decellularisation, two recently developed protocols for extraction of matrisomal proteins from homogenised tissue offers a more rapid approach [11] (Fig. 1). The principle is based on sequential removal of soluble cellular components after tissue homogenization with the remaining insoluble fraction containing enriched matrisomal components. In this study, we have employed two extraction methods, Extraction method A (EMA) is based on the protocol by Hill *et al*. [13] and Extraction method B (EMB) by Naba *et al.* [9]. EMA employs a buffer containing CHAPS and high ionic strength (3M NaCl) to lyse cells and extract soluble cellular proteins while minimising matrisomal loss (Fig. 1). Matrisomal proteins are then denatured and solubilized by 8M urea resulting in a fraction of soluble ECM and pellet of insoluble ECM [13]. The insoluble pellet is then treated with CNBr to decrease the complexity of the protein meshwork and improve accessibility for trypsin prior to mass spectrometry analysis (Fig. 1). EMB employs a commercial cellular compartment enrichment kit to sequentially remove cellular components (cytoplasmic, nuclear, membrane and

cytoskeletal proteins) from tissue (Fig. 1) [9]. The resultant pellet (which contains the matrisomal components) is solubilised in 8M urea, deglycosylated and digested with LysC to decrease the complexity of the ECM meshwork prior to digestion with trypsin and mass spectrometry analysis (Fig. 1).

In this study we applied a shotgun proteomic approach using nanoflow reversed-phase liquid chromatography coupled on-line to the Orbitrap Velos mass spectrometer. Of the 200-300 proteins which were identified in a typical sample after enrichment using this set-up, 30-70 were assigned as matrisomal proteins. It should be noted that the number of identified proteins may vary on LC-MS system used as well as the stringency employed during data processing (see methods for details). No additional fractionation (SDS-PAGE, off-gel electrophoresis, etc.) was used in this study.

Heart

The heart is composed of myocardial muscle and collagenous heart valves and because of the periodic contractions of the myocardium, there are high demands for tissue flexibility and contractility [27]. The ECM in the heart is necessary for the distribution of mechanical forces throughout the myocardium and transmission of electromechanical signals maintaining systolic and diastolic function [28, 29]. Dysregulation of ECM such as the increased accumulation of collagen is associated with the development of cardiovascular disease [29, 30].

Comparative assessment of the protein extraction yield (defined as the percentage of extracted protein as a function of organ wet weight) from heart shows an average yield of 0.66 – 2.47% with no statistical difference across the four methods (Fig. 2A). The EMA protocol identified the highest number of matrisomal proteins in the heart with 33 proteins found in at least 2 of 3 biological replicates (Fig. 2B). However, when percentage matrisome enrichment (defined as the percentage of identified proteins which are catalogued in the

Matrisome Project Database) is considered, decellularisation by SDS performed best with 33% of total identified proteins being assigned as matrisome proteins (Fig 2C). As depicted in the Venn diagram in Fig. 2D, 22 proteins were commonly identified by all enrichment techniques showing good overlap in their enrichment capability, with only few proteins uniquely enriched by just one of the methods.

The proteomic data was mapped onto the mouse matrisome database (MatrisomeDB, http://matrisomeproject.mit.edu/), where matrisomal proteins are divided into the "core matrisome" (encompassing glycoproteins, collagens and proteoglycans) and the "matrisome-associated proteins" (encompassing affiliated proteins, regulators and secreted factors) [9]. Glycoproteins and collagens were the most represented matrisomal classes in heart in our analysis (Fig. 2E and Table 1). The proteins identified in these two classes have previously been documented in human and porcine hearts. For example, Didangelos et al. identified the glycoproteins fibronectin (FINC) and fibulins (FBLN), the proteoglycan perlecan (PGBM) and multiple collagens in human aorta while Barallobre-Barreiro et al., isolated the laminin class of glycoproteins from porcine heart [15, 31, 32]. The method of enrichment in these studies is based on a combination of extraction and decellularisation as the tissue was diced into smaller pieces (approximately 15-20 mg/piece) but not homogenized or minced. Interestingly, we did not observe much enrichment of matrisome-associated proteins, in particular there were no matrisome-affiliated proteins and secreted factors found in our analysis. This data is consistent with a previous study by de Castro Bras et al. who enriched and analysed matrisomal proteins from the mouse left ventricle but was unable to identify any matrisome-associated proteins in the insoluble fraction [23]. One potential reason for the lack of matrisome-associated proteins could be the result of multiple steps of solubilisation and washes to remove intracellular components in the four matrisome enrichment strategies during which soluble matrisome-associated proteins are likely to be washed away.

To test this hypothesis, we performed an LC-MS/MS analysis of murine heart lysates that were not subjected to enrichment. While the non-enriched samples showed poorer recovery of glycoproteins and collagens, we identified 10 and 14 matrisome-associated proteins using a short (88min) and long (4 hour) gradient respectively in the non-enriched samples (Fig 2E). This finding demonstrates that matrisome enrichment leads to a significant loss of matrisome-associated proteins in the heart and that these components can be directly identified in non-enriched samples. Although the enrichment methodologies (in particular the EMA protocol) were capable of enriching several ECM regulators, direct LC-MS/MS of non-enriched samples was superior in the number of proteins identified in this class.

Mammary gland

The mammary gland is composed of several lobules encapsulated and separated by basement membrane and connective tissue which is surrounded by a fatty layer that needs to be effectively removed before LC-MS/MS analysis [33]. Many matrisomal proteins have been reported as important players during the mammary gland development and branching of ducts, including PGBM, matrix metalloproteinases (MMPs) and their inhibitors and transforming growth factor beta (TGFβ) [33]. Moreover, the mammary gland stroma undergoes extensive changes and remodelling during pubertal development, pregnancy and after postpartum involution [34].

Analysis of the protein extraction yields revealed that Triton X decellularisation achieved the best recovery with an average of 1.5% wet weight (Fig. 3A). The other three methods showed low extraction yield with SDS decellularisation yielding the lowest average of 0.11%. The highest number of 41 matrisomal proteins was identified in the samples extracted by the EMA protocol (Fig. 3B). Conversely, SDS decellularisation resulted in the identification of the lowest number of 23 matrisomal proteins. Similar to murine heart, SDS decellularisation also showed the highest matrisomal enrichment in the mammary gland of 38% (Fig. 3C). 18 matrisome proteins were commonly identified by all methods while 9 matrisomal proteins

were exclusively identified by Hill's method (Fig. 3D and Table 1). These 9 proteins include the collagens XI (COBA1) and XV (COFA1), the proteoglycan prolargin (PRELP), the ECM affiliated protein annexin 6 (ANXA6) and four ECM regulators: three forms of alpha-1-antitrypsin (A1AT1, A1AT2, A1AT4) and serpin H (SERPH). PRELP has previously been identified by O'Brien *et al*. in rat mammary glands [34] while COBA1, COFA1, ANXA6, A1AT and SERPH have been detected in mouse mammary fat pad xenograft models of breast carcinoma [18].

Proteins of the core matrisome were the dominant matrisomal proteins identified in the enriched mammary gland samples (Fig. 3E). For example, 15 collagens were identified by the EMA protocol which is consistent with the previously reported study by O'Brien *et al*., where 19 collagens were identified in rat mammary gland using the same enrichment method [34]. In the glycoprotein class of core matrisome proteins, 8 different laminin chains were identified by O'Brien *et al*. compared to 4 in our analysis [34]. Overall, 60 matrisomal proteins were identified in the study by O'Brien *et al*. versus 41 identified in our dataset. However it should be noted that O'Brien *et al*. performed extensive fractionation of the sample into 17 fractions by SDS-PAGE prior to LC-MS/MS analysis while our study has no fractionation steps after matrisome enrichment. This extensive fractionation may account for the higher number of matrisomal proteins identified in the O'Brien *et al.* study.

Unexpectedly, the four enrichment methods did not dramatically increase the number of identified core matrisomal proteins when compared to the non-enriched samples (Fig. 3E). 30 core matrisome components were identified in the non-enriched sample while 29 proteins were identified after Triton X decellularisation, 23 after SDS decellularisation, 34 after the EMA protocol extraction and 30 after extraction by the EMB protocol. Increasing the liquid chromatography gradient from 88 minutes to 4 hours increased the number of identified core matrisomal proteins in the non-enriched sample to 39. Of the four enrichment strategies, the EMA protocol yielded the highest number of 7 matrisome-associated proteins. However

similar to the heart, matrisome-associated proteins were poorly enriched by the four methods compared to non-enriched samples in the mammary gland (Fig. 3E).

Lung

The lung is a complex organ composed of airways and alveoli intertwined with veins. The lung ECM is primarily composed of collagen and elastin fibres. Because of the requirements for very high flexibility, elastin is placed among individual alveoli and capillaries where it contributes to lung elasticity during inhalation and plays a major role in the intrinsic recoil of the lung [35]. The more rigid ECM molecules of collagen I, II and III are found in the bronchus, bronchioles and veins while collagen IV and V are present in the basement membrane of capillaries and alveoli [35]. Lung-specific ECM components include the surfactants which consist of lipoproteins that are present in the alveoli where they decrease the air-liquid tension and thus protect alveoli from collapse [36].

The average protein extraction yield in murine lung after matrisome enrichment was 0.5 – 1% wet weight with no statistically significant differences between the four methods (Fig. 4A). Enrichment by the EMA protocol identified the largest number of 58 matrisomal proteins (Fig. 4B) while the highest level of matrisome enrichment of 61% was achieved after SDS decellularisation (Fig. 4C). The number of proteins identified using the EMB protocol in our study (between 37 to 70 proteins in three independent biological replicates) is comparable with the published study by Naba *et al*., where 55 matrisomal proteins were identified in murine lungs in one biological replicate in the absence of fractionation by off-gel electrophoresis [9]. Collectively, 42 matrisomal proteins were identified in at least 2 of 3 biological replicates using the EMB protocol in our study.

Comparative analysis revealed 34 matrisomal proteins which were commonly identified by all four enrichment methods (Fig. 4D). 12 proteins were identified exclusively after extraction by the EMA protocol. Annexin 2 (ANXA2), A1AT, Galectin-3 (LEG3) and pulmonary

surfactant-associated proteins A and B (SFTPA, PSPB) were among the proteins identified exclusively using this method (Table 1). These proteins have previously been identified in published murine lung matrisome studies [9, 13]. Additionally, the ECM regulator plasminogen (PLMN) was detected exclusively in Triton X decellularisation and the glycoprotein Latent-transforming growth factor beta-binding protein 4 (LTBP4) was only found in the EMB protocol. Similar to the heart and mammary gland, collagens and glycoproteins were the most prominent classes of matrisomal proteins that were recovered (Fig. 4E). Interestingly, Triton X decellularisation and EMA extraction were superior in the enrichment of proteoglycans in the lung (Fig. 4E). Consistent with our analyses in the other organs, the four matrisome enrichment methods led to significant losses in matrisome-associated proteins compared to the non-enriched samples (Fig. 4E).

Liver

The ECM of the liver comprises less than 3% of the normal liver and is limited to the Glisson's capsule, portal tracts and central veins [37]. Unlike other organs, there is no physical basement membrane surrounding hepatocytes, although the individual proteins of the basement membrane are present [37, 38]. Instead, hepatocytes are surrounded directly by matrix consisting of FINC, collagens and basement membrane proteins which facilitates rapid diffusion between the blood and cells [39]. The low ECM levels in the liver may be a possible explanation for the low protein extraction yields of between 0.17-0.5% average wet weight with no statistically significant differences between the four methods (Fig. 5A). The highest number of 40 matrisomal proteins was identified by the EMA method (Fig. 5B) and the highest percentage of matrisome enrichment was obtained by SDS decellularisation, where 48% of total proteins were assigned as matrisomal (Fig. 5C). 23 proteins were common across the four employed methods (Fig. 5D and Table 1). A number of proteins were uniquely identified by specific methods. For instance, SDS decellularisation identified 6 unique matrisomal proteins including laminins β3 (LAMB3) and γ2 (LAMC2), LTBP4, and three types of collagens (CO4A1, CO7A1 and CODA1), The EMA protocol found 5 unique

proteins such as laminins α4 (LAMA4) and β1 (LAMB1), collagens XVI (COGA1) and XVIII (COIA1) and α-1-antitrypsin (A1AT2). Triton X decellularisation identified 2 proteoglycans lumican (LUM) and decorin (PGS2) while the EMB protocol identified 2 matrisome associated proteins ERGIC-53 (LMAN1) and serine-protease inhibitor A3K (SPA3K).

The benefit of matrisome enrichment in the liver is pronounced compared to non-enriched samples where a short gradient identified less than half of the proteins in the core matrisome recovered by the matrisome enrichment methods (Fig. 5E). For example, no proteoglycans were identified in non-enriched samples, compared to 3 proteoglycans - PGBM, byglican (PGS1), PRELP - identified after EMB protocol extraction and 5 proteoglycans (PGBM, LUM, PGS1, PGS2, PRELP) identified after Triton X decellularisation (Table I). Importantly, our findings demonstrate that enrichment is necessary for comprehensive analysis of the liver matrisome, which may be the result of the low ECM content in this organ.

Comparative assessment across organs

Across the four organs examined, the highest number of matrisomal proteins was detected in the lung tissue where 64 proteins were identified, 12 of which were unique to lung (Fig. 6A and B). In particular, glycoproteins were found in high abundance in lung compared to other organs (Fig. 6A) with the following lung-specific glycoproteins identified in our study: agrin (AGRN), fibrillin-2 (FBN2), fibulin-3 (FBLN3), TGF-beta-induced protein ig-h3 (BGH3), LTBP4 and von Willebrand factor (VWF) (Fig. 6B). In contrast, matrisomal enrichment of the heart showed the lowest number of 38 matrisomal proteins identified with only few proteoglycans and no affiliated proteins detected (Fig. 6A). Furthermore, unlike the other three organs, no heart specific proteins were identified in our study (Fig 6B). In addition, glycoprotein emilin-2 (EMIL2), collagens XXIV (COOA1) and XXVIII (COSA1), A1AT1 and ANXA6 were uniquely identified in mammary glands while alpha 5 chain of type VI collagen (CO6A5) and alpha 1 chain of type XIV collagen (CODA1), affiliated proteins galectin-9 (LEG9) and LMAN and regulator SPA3K were unique to liver tissue.

Our analysis finds 29 matrisomal proteins that are common across all four organs (Fig. 6B). These common matrisome components comprise primarily proteins of the collagen and glycoprotein classes (Table 1). The identification of common matrisome proteins in this study will contribute to the efforts by several groups to comprehensively define the "ECM atlas" across multiple tissue types [10]. It should be noted that our study is qualitative in nature and does not provide any quantitative information of the relative levels of the 29 matrisomal proteins in each of the four organs examined. This list of common matrisomal components will be useful for the development of Selected Reaction Monitoring (SRM) approaches for quantitative profiling of the matrisome in multiple tissue types; where the identity of the proteins and peptides of interest are required *a priori* for assay development [40]. A recent study describes the development of an SRM method based on QconCat technology, which utilises 83 stable isotope labelled peptides representing 48 different ECM proteins to measure decellularised human hearts [14]. 60% of our 29 common matrisomal proteins are found in the QconCat library indicating that this quantitative SRM method can be readily extended to quantitative matrisomal measurements in lung, mammary gland and liver.

Comparative assessment across methods

To evaluate performance of individual methods across the four organs, a scatter plot displaying percentage matrisome enrichment versus number of identified unique matrisomal protein was generated (Fig. 7A). Our study finds that while SDS decellularisation provides the highest percentage matrisome enrichment across all four organs, the number of identified matrisomal proteins is consistently the lowest. This result may be due to the negative charge of SDS which causes protein denaturation and tissue disintegration. While this property of SDS facilitates the effective removal of the cellular proteins during washing steps, it also leads to losses in matrisomal proteins leading to our observed high percentage matrisome enrichment accompanied by low numbers of matrisome protein identifications. Consistent with this idea, the loss of matrisomal glucosoaminoglycans and reduced collagen

integrity after SDS decellularisation has been described in aortic valves , annulus fibrosus and rat tail [12, 25, 41, 42].

The EMA extraction protocol identified the highest number of matrisomal proteins in all analysed tissues, despite having similar percentage matrisomal enrichment as Triton X decellularisation and the EMB protocol (Fig. 7A and B). The capability of the EMA protocol to identify more matrisomal proteins is also reflected by the number of proteins identified exclusively by this method, particularly in lung and mammary gland (Fig. 7B and Table 1). For example, A1AT2 was identified exclusively by the EMA protocol in all analysed tissues. In addition, SFTPA, PSPB and ANXA6 were identified only in samples after extraction with this method. The percentage matrisome enrichment and number of matrisomal proteins identified by Triton X decellularisation and the EMB protocol were very similar across all tissue (Fig 7A). The only exception is lung tissue, where the EMB extraction protocol provided superior matrisomal enrichment compared to Triton X decellularisation.

There is a significant overlap of identified matrisomal proteins between methods across all analysed tissues, with 38 out of 79 matrisomal proteins commonly identified by all four methods (Fig. 7B). These proteins include important core matrisomal glycoproteins (such as fibrillin 1 (FBN1), fibrinogens (FIBA, FIBB, FIBG), FINC and laminins) and collagens (Table 1). Furthermore in agreement with previous studies showing that Triton X improves the recovery of proteoglycans [43], our experiments also demonstrate that Triton X decellularisation provided superior enrichment of the proteoglycan class of matrisomal proteins across all four organs (Figures 2E, 3E, 4E and 5E).

In terms of number of unique matrisome proteins identified, our study shows that lung and mammary gland do not appear to benefit from matrisome enrichment as direct LC-MS/MS analysis of the non-enriched samples leads to the identification of similar or higher numbers of core matrisome components (Fig. 3E, 4E). One potential reason for this unexpected

finding is that these two tissue types have higher levels of overall ECM content compared to heart and liver. Taking collagen content as an indicator for overall ECM content in an organ, the content of collagen in rat lung is 11.3% of dry fat-free weight, while rat ventricle contains 2.96% and rat liver 0.64% [44]. In rat mammary gland, collagen forms 9.9% of wet fat-free weight while the content of non-collagenous protein is ten times lower [45]. The high (lung and mammary gland) and low (heart and liver) collagen content - and thus an approximation of ECM content in the tissue - is in good agreement with the requirement for matrisome enrichment in our study. For the latter two organs, direct LC-MS/MS analysis of unenriched samples is sufficient to identify the abundant core matrisomal proteins.

It has been suggested that matrisome enrichment is required to increase the depth of protein sequence coverage due to the large number of ECM proteins displaying a wide dynamic range [11]. We sought to determine if non-enriched samples have reduced sequence coverage by taking a selection of 4 proteins from distinct matrisomal classes and plotting the percentage protein sequence coverage across all methods and organs (Figure S1). We find that there is no clear relationship between sequence coverage and type of enrichment. While the non-enriched samples show reduced protein sequence coverage across all organs in some proteins (FINC and PGBM), this is not always the case. For instance, in the case of CO1A1 (Figure S1C), the non-enriched samples display comparable sequence coverage to the other four enrichment methods in all organs examined except the heart. In another example of TGM2 (Figure S1D), the non-enriched samples have higher or similar sequence coverage compared to EMA in the lung and liver but lower sequence coverage in the heart and mammary gland. This data demonstrates that percentage sequence coverage is not always reduced in non-enriched samples and is likely dependent on the type of organ, method of choice and protein of interest.

Our comparative analysis also revealed that a large proportion of matrisome-associated proteins are lost during enrichment using any of the four methods. These results are perhaps expected given that the majority of proteins in the matrisome-associated class are soluble and readily lost during enrichment. In the deep proteomic study of mouse lung ECM by Schiller et al., only 53 of a total 264 matrisome-associated proteins were found to be enriched in the insoluble fraction after quantitative detergent solubility profiling [16]. Similarly, in a study of the human ventricle ECM performed by Barallobre-Barreiro [46], a higher abundance of matrisome-associated proteins such as S100 proteins or ECM peptidases was found in the soluble fraction. Our data highlights that in-depth proteomic analysis of non-enriched samples is currently the best available tool for the analysis of this class of proteins, however further development of new enrichment strategies to better isolate matrisome-associated proteins with improved sequence coverage is necessary.

Conclusion

In summary, we have performed a comparative proteomic analysis of four matrisome enrichment strategies in four different organs. Our data emphasizes that choice of enrichment strategy is dependent on tissue type, matrisome class of interest and desired matrisome enrichment purity. Based on our results, SDS decellularisation is a good option for attaining high purity of matrisome enriched samples, while the EMA extraction protocol can be universally used to achieve high levels of matrisome protein identification. Triton X decellularisation provides good enrichment of proteoglycans while in-depth proteomic analysis of non-enriched samples is necessary to identify matrisome-associated proteins. We further show that in some organs such as mammary glands and lungs, matrisome enrichment is not superior to proteomic analysis of non-enriched samples based on the number of matrisomal proteins identified and suggest that this may be due to the overall higher content of ECM in these tissues. We anticipate that future work comparing matrisome enrichment versus non-enriched samples in additional tissue types of differing ECM content

will provide more clarity of the necessity for prior sample enrichment in the proteomic characterisation of the matrisome.

Acknowledgments

We acknowledge access to mass spectrometry instrumentation in the ICR proteomics core facility.

Declarations of interest

The authors declare no competing financial interest.

Funding information

This work was supported by grants from the Institute of Cancer Research (ICR) and Breast Cancer Now (2014NovPR360) to PHH, a Breast Cancer Now Scientific Fellowship (SF01May2011) to RCN, and Breast Cancer Now programmatic funding (BAH and RCN).

Author contribution

L.K., A.P., P.W., B.A.H. and R.C.N. performed the research; L.K., A.P. and P.H.H. analysed the data; L.K. and P.H.H. wrote the paper with input from all the authors.

References

1 Mouw, J. K., Ou, G. and Weaver, V. M. (2014) Extracellular matrix assembly: a multiscale deconstruction. Nat Rev Mol Cell Biol. **15**, 771-785

2 Pickup, M. W., Mouw, J. K. and Weaver, V. M. (2014) The extracellular matrix modulates the hallmarks of cancer. EMBO Rep. **15**, 1243-1253

3 Payne, L. S. and Huang, P. H. (2013) The pathobiology of collagens in glioma. Mol Cancer Res. **11**, 1129-1140

4 Lu, P., Weaver, V. M. and Werb, Z. (2012) The extracellular matrix: a dynamic niche in cancer progression. J Cell Biol. **196**, 395-406

5 Bateman, J. F., Boot-Handford, R. P. and Lamande, S. R. (2009) Genetic diseases of connective tissues: cellular and extracellular effects of ECM mutations. Nat Rev Genet. **10**, 173-183

6 Wallace, K., Burt, A. D. and Wright, M. C. (2008) Liver fibrosis. Biochem J. **411**, 1-18

7 Roach, H. I., Aigner, T., Soder, S., Haag, J. and Welkerling, H. (2007) Pathobiology of osteoarthritis: pathomechanisms and potential therapeutic targets. Curr Drug Targets. **8**, 271-282

8 Chistiakov, D. A., Sobenin, I. A. and Orekhov, A. N. (2013) Vascular extracellular matrix in atherosclerosis. Cardiol Rev. **21**, 270-288

9 Naba, A., Clauser, K. R., Hoersch, S., Liu, H., Carr, S. A. and Hynes, R. O. (2012) The matrisome: in silico definition and in vivo characterization by proteomics of normal and tumor extracellular matrices. Mol Cell Proteomics. **11**, M111 014647

10 Naba, A., Clauser, K. R., Ding, H., Whittaker, C. A., Carr, S. A. and Hynes, R. O. (2016) The extracellular matrix: Tools and insights for the "omics" era. Matrix Biol. **49**, 10-24

11 Byron, A., Humphries, J. D. and Humphries, M. J. (2013) Defining the extracellular matrix using proteomics. Int J Exp Pathol. **94**, 75-92

12 Xu, H., Xu, B., Yang, Q., Li, X., Ma, X., Xia, Q., Zhang, Y., Zhang, C. and Wu, Y. (2014) Comparison of decellularization protocols for preparing a decellularized porcine annulus fibrosus scaffold. PLoS One. **9**, e86723

13 Hill, R. C., Calle, E. A., Dzieciatkowska, M., Niklason, L. E. and Hansen, K. C. (2015) Quantification of extracellular matrix proteins from a rat lung scaffold to provide a molecular readout for tissue engineering. Mol Cell Proteomics. **14**, 961-973

14 Johnson, T. D., Hill, R. C., Dzieciatkowska, M., Nigam, V., Behfar, A., Christman, K. L. and Hansen, K. C. (2016) Quantification of decellularized human myocardial matrix: A comparison of six patients. Proteomics Clin Appl. **10**, 75-83

15 Barallobre-Barreiro, J., Didangelos, A., Schoendube, F. A., Drozdov, I., Yin, X., Fernandez-Caggiano, M., Willeit, P., Puntmann, V. O., Aldama-Lopez, G., Shah, A. M., Domenech, N. and Mayr, M. (2012) Proteomics analysis of cardiac extracellular matrix remodeling in a porcine model of ischemia/reperfusion injury. Circulation. **125**, 789-802

16 Schiller, H. B., Fernandez, I. E., Burgstaller, G., Schaab, C., Scheltema, R. A., Schwarzmayr, T., Strom, T. M., Eickelberg, O. and Mann, M. (2015) Time- and compartment-resolved proteome profiling of the extracellular niche in lung injury and repair. Mol Syst Biol. **11**, 819

17 Naba, A., Clauser, K. R., Whittaker, C. A., Carr, S. A., Tanabe, K. K. and Hynes, R. O. (2014) Extracellular matrix signatures of human primary metastatic colon cancers and their metastases to liver. BMC Cancer. **14**, 518

18 Naba, A., Clauser, K. R., Lamar, J. M., Carr, S. A. and Hynes, R. O. (2014) Extracellular matrix signatures of human mammary carcinoma identify novel metastasis promoters. Elife. **3**, e01308

19 O'Brien, J., Hansen, K., Barkan, D., Green, J. and Schedin, P. (2011) Non-steroidal anti-inflammatory drugs target the pro-tumorigenic extracellular matrix of the postpartum mammary gland. Int J Dev Biol. **55**, 745-755

20 Maller, O., Hansen, K. C., Lyons, T. R., Acerbi, I., Weaver, V. M., Prekeris, R., Tan, A. C. and Schedin, P. (2013) Collagen architecture in pregnancy-induced protection from breast cancer. J Cell Sci. **126**, 4108-4110

21 Randles, M. J., Woolf, A. S., Huang, J. L., Byron, A., Humphries, J. D., Price, K. L., Kolatsi-Joannou, M., Collinson, S., Denny, T., Knight, D., Mironov, A., Starborg, T., Korstanje, R., Humphries, M. J., Long, D. A. and Lennon, R. (2015) Genetic Background is a Key Determinant of Glomerular Extracellular Matrix Composition and Organization. J Am Soc Nephrol. **26**, 3021-3034

22 Lennon, R., Byron, A., Humphries, J. D., Randles, M. J., Carisey, A., Murphy, S., Knight, D., Brenchley, P. E., Zent, R. and Humphries, M. J. (2014) Global analysis reveals the complexity of the human glomerular extracellular matrix. J Am Soc Nephrol. **25**, 939-951

23 de Castro Bras, L. E., Ramirez, T. A., DeLeon-Pennell, K. Y., Chiao, Y. A., Ma, Y., Dai, Q., Halade, G. V., Hakala, K., Weintraub, S. T. and Lindsey, M. L. (2013) Texas 3-step decellularization protocol: looking at the cardiac extracellular matrix. J Proteomics. **86**, 43-52

24 Nesvizhskii, A. I., Keller, A., Kolker, E. and Aebersold, R. (2003) A statistical model for identifying proteins by tandem mass spectrometry. Anal Chem. **75**, 4646-4658

25 Gilbert, T. W., Sellaro, T. L. and Badylak, S. F. (2006) Decellularization of tissues and organs. Biomaterials. **27**, 3675-3683

26 Keane, T. J., Londono, R., Turner, N. J. and Badylak, S. F. (2012) Consequences of ineffective decellularization of biologic scaffolds on the host response. Biomaterials. **33**, 1771-1781

27 Curtis, M. W. and Russell, B. (2011) Micromechanical regulation in cardiac myocytes and fibroblasts: implications for tissue remodeling. Pflugers Arch. **462**, 105-117

28 Kandalam, V., Basu, R., Abraham, T., Wang, X., Awad, A., Wang, W., Lopaschuk, G. D., Maeda, N., Oudit, G. Y. and Kassiri, Z. (2010) Early activation of matrix metalloproteinases underlies the exacerbated systolic and diastolic dysfunction in mice lacking TIMP3 following myocardial infarction. Am J Physiol Heart Circ Physiol. **299**, H1012-1023

29 Baudino, T. A., Carver, W., Giles, W. and Borg, T. K. (2006) Cardiac fibroblasts: friend or foe? Am J Physiol Heart Circ Physiol. **291**, H1015-1026

30 Chapman, D., Weber, K. T. and Eghbali, M. (1990) Regulation of fibrillar collagen types I and III and basement membrane type IV collagen gene expression in pressure overloaded rat myocardium. Circ Res. **67**, 787-794

31 Didangelos, A., Yin, X., Mandal, K., Baumert, M., Jahangiri, M. and Mayr, M. (2010) Proteomics characterization of extracellular space components in the human aorta. Mol Cell Proteomics. **9**, 2048-2062

32 Didangelos, A., Yin, X., Mandal, K., Saje, A., Smith, A., Xu, Q., Jahangiri, M. and Mayr, M. (2011) Extracellular matrix composition and remodeling in human abdominal aortic aneurysms: a proteomics approach. Mol Cell Proteomics. **10**, M111 008128

33 Wiseman, B. S. and Werb, Z. (2002) Stromal effects on mammary gland development and breast cancer. Science. **296**, 1046-1049

34 O'Brien, J. H., Vanderlinden, L. A., Schedin, P. J. and Hansen, K. C. (2012) Rat mammary extracellular matrix composition and response to ibuprofen treatment during postpartum involution by differential GeLC-MS/MS analysis. J Proteome Res. **11**, 4894-4905

35 Balestrini, J. L. and Niklason, L. E. (2015) Extracellular matrix as a driver for lung regeneration. Ann Biomed Eng. **43**, 568-576

36 Han, S. and Mallampalli, R. K. (2015) The Role of Surfactant in Lung Disease and Host Defense against Pulmonary Infections. Ann Am Thorac Soc. **12**, 765-774

37 Bedossa, P. and Paradis, V. (2003) Liver extracellular matrix in health and disease. J Pathol. **200**, 504-515

38 Martinez-Hernandez, A. and Amenta, P. S. (1993) The hepatic extracellular matrix. I. Components and distribution in normal liver. Virchows Arch A Pathol Anat Histopathol. **423**, 1-11

39 Martinez-Hernandez, A. (1985) The hepatic extracellular matrix. II. Electron immunohistochemical studies in rats with CCl4-induced cirrhosis. Lab Invest. **53**, 166-186

40 Lange, V., Picotti, P., Domon, B. and Aebersold, R. (2008) Selected reaction monitoring for quantitative proteomics: a tutorial. Mol Syst Biol. **4**, 222

41 Kasimir, M. T., Rieder, E., Seebacher, G., Silberhumer, G., Wolner, E., Weigel, G. and Simon, P. (2003) Comparison of different decellularization procedures of porcine heart valves. Int J Artif Organs. **26**, 421-427

42 Cartmell, J. S. and Dunn, M. G. (2000) Effect of chemical treatments on tendon cellularity and mechanical properties. J Biomed Mater Res. **49**, 134-140

43 Yanagishita, M., Podyma-Inoue, K. A. and Yokoyama, M. (2009) Extraction and separation of proteoglycans. Glycoconj J. **26**, 953-959

44 Neuman, R. E. and Logan, M. A. (1950) The determination of collagen and elastin in tissues. J Biol Chem. **186**, 549-556

45 Harkness, M. L. and Harkness, R. D. (1956) The effect of pregnancy and lactation on the collagen content of the mammary gland of the rat. J Physiol. **132**, 476-481

46 Barallobre-Barreiro, J., Oklu, R., Lynch, M., Fava, M., Baig, F., Yin, X., Barwari, T., Potier, D. N., Albadawi, H., Jahangiri, M., Porter, K. E., Watkins, M. T., Misra, S., Stoughton, J. and Mayr, M. (2016) Extracellular matrix remodelling in response to venous hypertension: proteomics of human varicose veins. Cardiovasc Res. **110**, 419-430

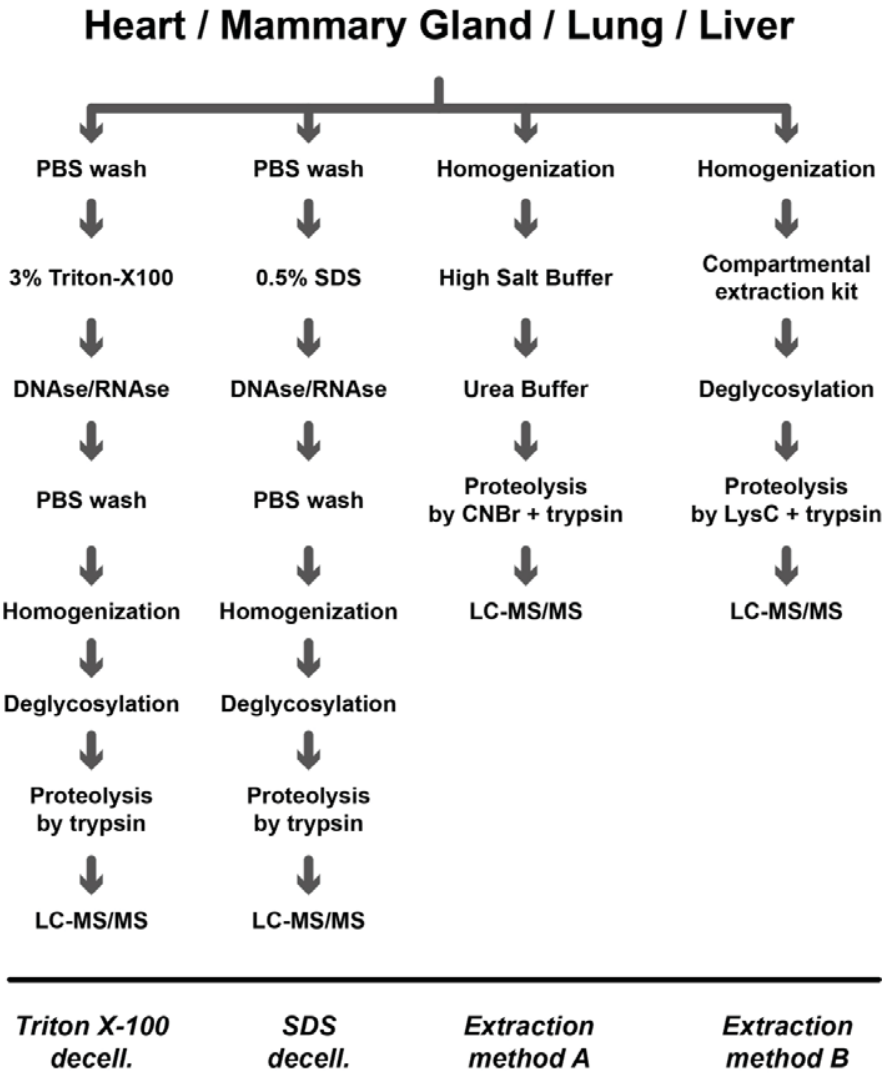

Figure 1: Matrisome enrichment workflows. Schematic of the workflow depicting key sample processing steps during matrisome enrichment by four individual methods and subsequent LC-MS/MS analysis.

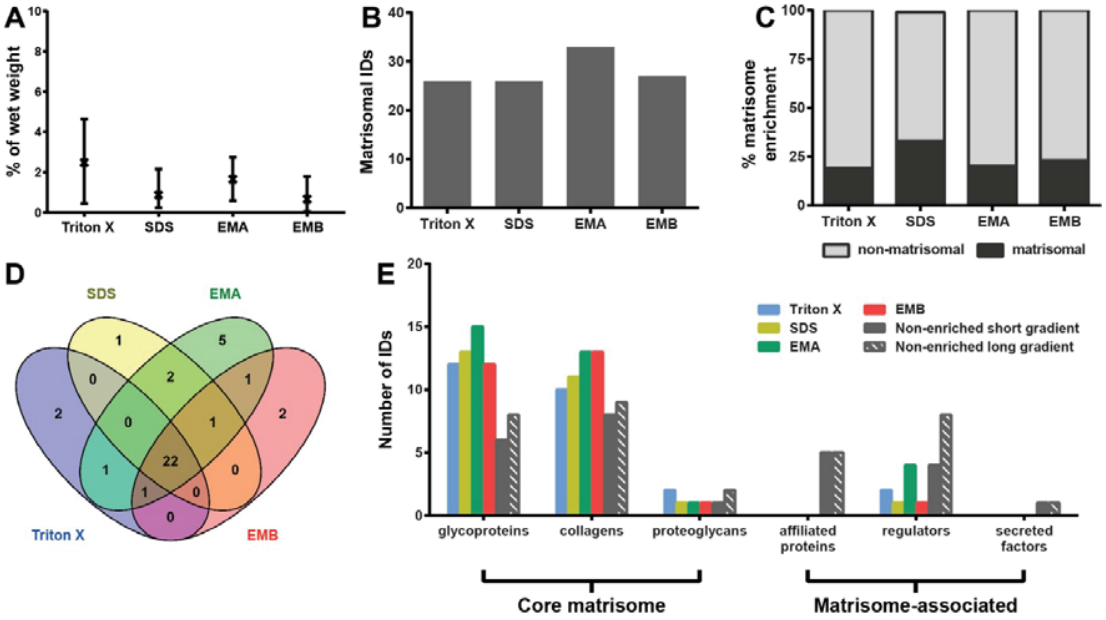

Figure 2: Comparison of methods for enrichment of the heart matrisome. A) Percentage protein extraction yield after enrichment by individual methods. Results from 3 biological replicates are shown as mean and range (n=3). B) Number of identified matrisomal proteins after enrichment. C) Proportional representation of identified matrisomal and non-matrisomal proteins after enrichment. D) Venn diagram depicting overlap of identified matrisomal proteins across the individual enrichment methods. E) Distribution of identified proteins in matrisomal classes for enriched and non-enriched samples. Only proteins detected in at least 2 biological replicates were considered.

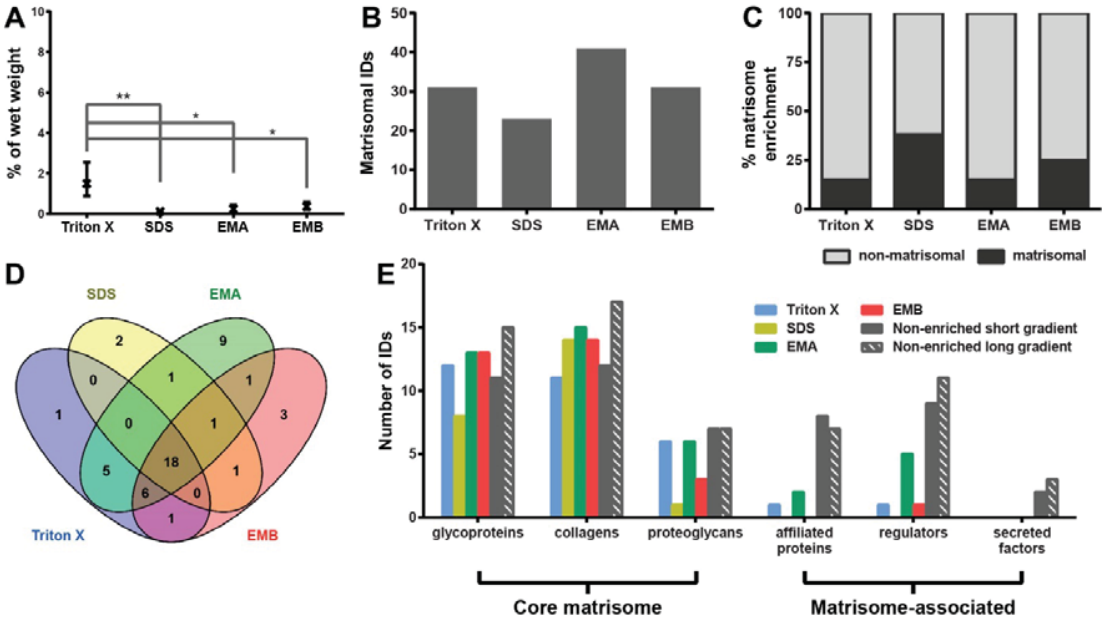

Figure 3: Comparison of methods for enrichment of the mammary gland matrisome. A) Percentage protein extraction yield after enrichment by individual methods. Results from 3 biological replicates are shown as mean and range (n=3). **p<0.01, *p<0.05. B) Number of identified matrisomal proteins after enrichment. C) Proportional representation of identified matrisomal and non-matrisomal proteins after enrichment. D) Venn diagram depicting overlap of identified matrisomal proteins across the individual enrichment methods. E) Distribution of identified proteins in matrisomal classes for enriched and non-enriched samples. Only proteins detected in at least 2 biological replicates were considered.

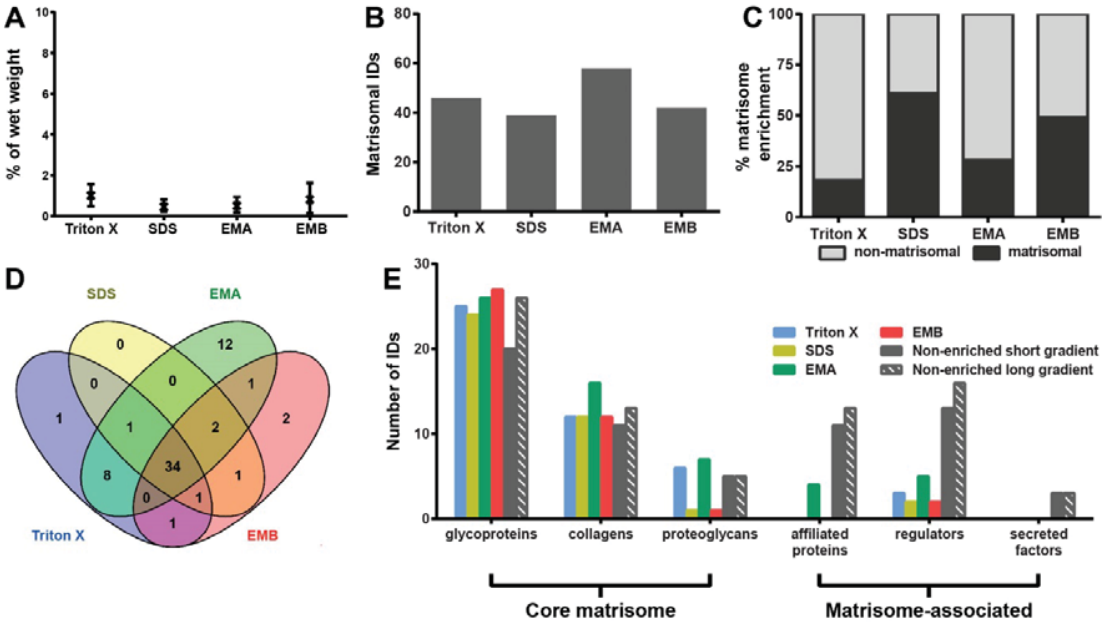

Figure 4: Comparison of methods for enrichment of the lung matrisome. A) Percentage protein extraction yield after enrichment by individual methods. Results from 3 biological replicates are shown as mean and range (n=3). B) Number of identified matrisomal proteins after enrichment. C) Proportional representation of identified matrisomal and non-matrisomal proteins after enrichment. D) Venn diagram depicting overlap of identified matrisomal proteins across the individual enrichment methods. E) Distribution of identified proteins in matrisomal classes for enriched and non-enriched samples. Only proteins detected in at least 2 biological replicates were considered.

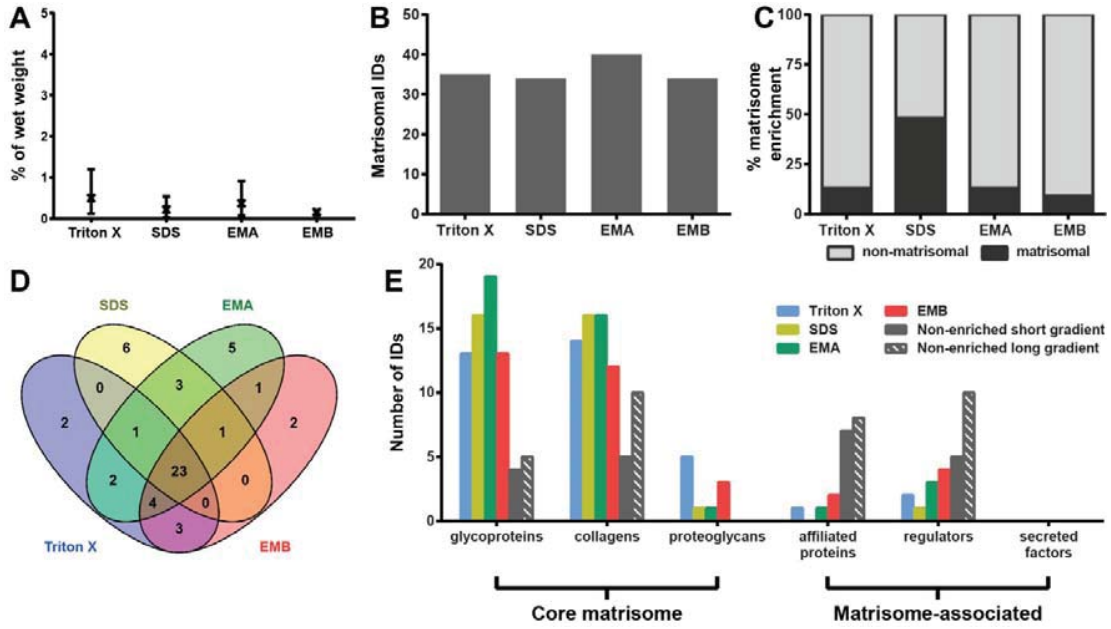

Figure 5: Comparison of methods for enrichment of the liver matrisome. A) Percentage protein extraction yield after enrichment by individual methods. Results from 3 biological replicates are shown as mean and range (n=3). B) Number of identified matrisomal proteins after enrichment. C) Proportional representation of identified matrisomal and non-matrisomal proteins after enrichment. D) Venn diagram depicting overlap of identified matrisomal proteins across the individual enrichment methods. E) Distribution of identified proteins in matrisomal classes for enriched and non-enriched samples. Only proteins detected in at least 2 biological replicates were considered.

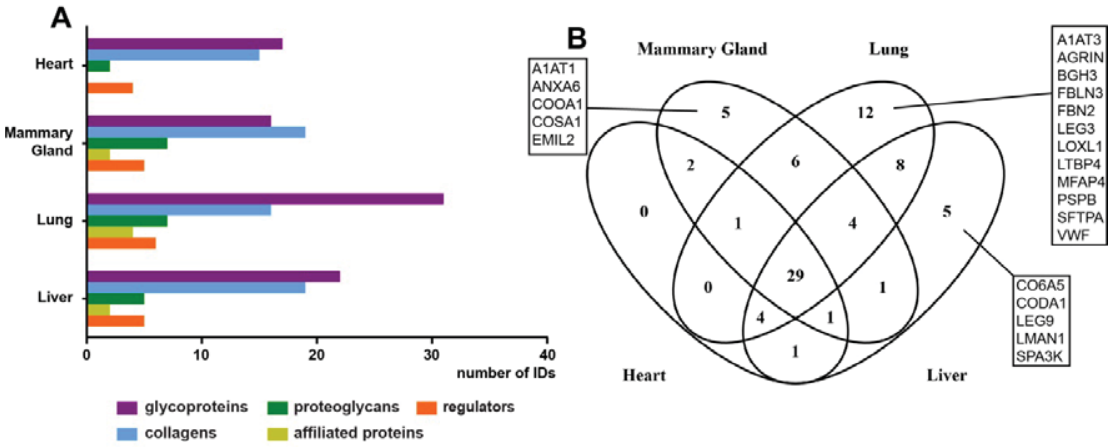

Figure 6: Comparative assessment of matrisomal proteins across organs. A) Distribution of identified proteins in matrisomal classes enriched by all methods across the four organs. Proteins identified in 2 or more biological replicates were considered. B) Venn diagram depicting overlap of identified matrisomal proteins across the individual organs.

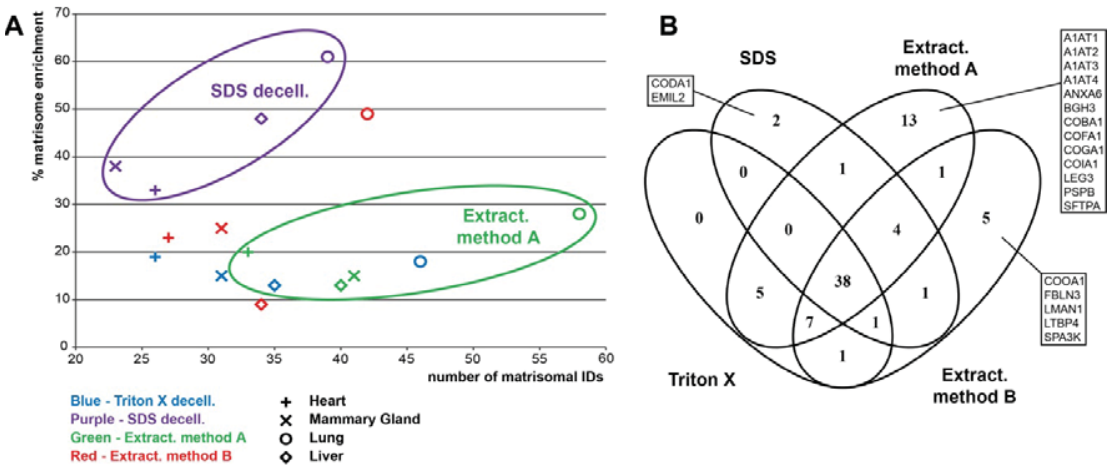

Figure 7: Comparative assessment of matrisomal proteins across enrichment methods. A) A scatter plot of the number of identified matrisomal proteins and enrichment capability. Clustering of samples extracted by EMA method and after SDS decellularisation are highlighted. The percentage matrisome enrichment is defined as the percentage of identified proteins which are catalogued in the Matrisome Project Database. Only proteins detected in at least 2 biological replicates were considered. B) Venn diagram depicting overlap of identified matrisomal proteins across the individual enrichment methods.

.

Table I: List of identified matrisomal proteins

Protein Class*	Protein name	Uniprot Entry Name	Gene Name	Heart				Mammary Gland				Lung				Liver			
				Triton	SDS	EMA	EMB	Triton	SDS	EMA	EMB	Triton	SDS	EMA	EMB	Triton	SDS	EMA	EMB
GP	Agrin	AGRIN_MOUSE	Agrn									x	x	x	x				
GP	TGF-beta-induced protein ig-h3	BGH3_MOUSE	Tgfbi											x					
GP	Dermatopontin	DERM_MOUSE	Dpt							x	x			x	x				
GP	Elastin	ELN_MOUSE	Eln		x					x			x	x	x		x	x	
GP	Emilin-1	EMIL1_MOUSE	Emilin1		x	x			x			x	x	x	x	x	x	x	x
GP	Fibulin-3	FBLN3_MOUSE	Fbln3												x				
GP	Fibulin-5	FBLN5_MOUSE	Fbln5	x								x	x	x	x	x		x	
GP	Fibrillin-1	FBN1_MOUSE	Fbn1	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
GP	Fibrillin-2	FBN2_MOUSE	Fbn2										x		x				
GP	Fibrinogen alpha chain	FIBA_MOUSE	Fga	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
GP	Fibrinogen beta chain	FIBB_MOUSE	Fgb	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
GP	Fibrinogen gamma chain	FIBG_MOUSE	Fgg	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
GP	Fibronectin	FINC_MOUSE	Fn1	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
GP	Laminin subunit alpha-2	LAMA2_MOUSE	Lama2	x	x	x	x	x		x	x	x	x	x					
GP	Laminin subunit alpha-3	LAMA3_MOUSE	Lama3									x	x	x	x		x	x	
GP	Laminin subunit alpha-4	LAMA4_MOUSE	Lama4			x	x					x	x	x	x			x	
GP	Laminin subunit alpha-5	LAMA5_MOUSE	Lama5	x	x	x	x				x	x	x	x	x	x	x	x	x
GP	Laminin subunit beta-1	LAMB1_MOUSE	Lamb1	x	x	x	x	x		x	x	x	x	x	x			x	
GP	Laminin subunit beta-2	LAMB2_MOUSE	Lamb2	x	x	x	x	x		x	x	x	x	x	x	x	x	x	x
GP	Laminin subunit beta-3	LAMB3_MOUSE	Lamb3									x	x	x	x		x		
GP	Laminin subunit gamma-1	LAMC1_MOUSE	Lamc1	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
GP	Laminin subunit gamma-2	LAMC2_MOUSE	Lamc2									x	x	x	x		x		
GP	Latent-transforming growth factor beta-binding protein 1	LTBP4_MOUSE	Ltbp4												x		x		
GP	Microfibril-associated glycoprotein 4	MFAP4_MOUSE	Mfap4											x					
GP	Nidogen-1	NID1_MOUSE	Nid1	x		x	x	x		x	x	x	x	x	x	x		x	x

Protein Class*	Protein name	Uniprot Entry Name	Gene Name	Heart				Mammary Gland				Lung				Liver			
				Triton	SDS	EMA	EMB	Triton	SDS	EMA	EMB	Triton	SDS	EMA	EMB	Triton	SDS	EMA	EMB
GP	Nidogen-2	NID2_MOUSE	Nid2			x		x				x		x		x		x	x
GP	Nephronectin	NPNT_MOUSE	Npnt									x	x	x	x		x	x	x
GP	Periostin	POSTN_MOUSE	Postn		x	x		x	x	x	x	x	x	x	x		x	x	
GP	Tubulointerstitial nephritis antigen-related protein	TINAL_MOUSE	Tinagl1									x			x	x		x	x
GP	von Willebrand factor	VWF_MOUSE	Vwf									x	x	x	x				
COL	Collagen alpha-1 (I) chain	CO1A1_MOUSE	Col1a1	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
COL	Collagen alpha-2 (I) chain	CO1A2_MOUSE	Col1a2	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
COL	Collagen alpha-1 (II) chain	CO2A1_MOUSE	Col2a1	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
COL	Collagen alpha-1 (III) chain	CO3A1_MOUSE	Col3a1	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
COL	Collagen alpha-1 (IV) chain	CO4A1_MOUSE	Col4a1	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
COL	Collagen alpha-2 (IV) chain	CO4A2_MOUSE	Col4a2	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
COL	Collagen alpha-3 (IV) chain	CO4A3_MOUSE	Col4a3						x			x	x	x	x	x	x	x	
COL	Collagen alpha-4 (IV) chain	CO4A4_MOUSE	Col4a4								x	x	x	x	x		x		
COL	Collagen alpha-1 (V) chain	CO5A1_MOUSE	Col5a1	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
COL	Collagen alpha-2 (V) chain	CO5A2_MOUSE	Col5a2	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
COL	Collagen alpha-1 (VI) chain	CO6A1_MOUSE	Col6a1	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
COL	Collagen alpha-2 (VI) chain	CO6A2_MOUSE	Col6a1	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
COL	Collagen alpha-5 (VI) chain	CO6A5_MOUSE	Col6a5													x	x	x	x
COL	Collagen alpha-6 (VI) chain	CO6A6_MOUSE	Col6a6		x	x	x									x		x	
COL	Collagen alpha-1 (VII) chain	CO7A1_MOUSE	Col7a1				x		x	x	x						x		
COL	Collagen alpha-1 (VIII) chain	CO8A1_MOUSE	Col8a1						x		x			x					
COL	Collagen alpha-1 (XI) chain	COBA1_MOUSE	Col11a1							x				x					
COL	Collagen alpha-1 (XIV) chain	CODA1_MOUSE	Col13a1														x		
COL	Collagen alpha-1 (XIII) chain	COEA1_MOUSE	Col14a1					x		x						x	x	x	x
COL	Collagen alpha-1 (XV) chain	COFA1_MOUSE	Col15a1			x				x									
COL	Collagen alpha-1 (XVI) chain	COGA1_MOUSE	Col16a1			x								x				x	

Protein Class*	Protein name	Uniprot Entry Name	Gene Name	Heart				Mammary Gland				Lung				Liver			
				Triton	SDS	EMA	EMB	Triton	SDS	EMA	EMB	Triton	SDS	EMA	EMB	Triton	SDS	EMA	EMB
COL	Collagen alpha-1 (XVIII) chain	COIA1_MOUSE	Col18a1											x				x	
COL	Collagen alpha-1 (XXIV) chain	COOA1_MOUSE	Col24a1								x								
COL	Collagen alpha-1 (XXVIII) chain	COSA1_MOUSE	Col28a1						x	x									
PG	Asporin	ASPN_MOUSE	Aspn					x		x				x					
PG	Lumican	LUM_MOUSE	Lum					x		x		x		x		x			
PG	Mimecan	MIME_MOUSE	Ogn					x		x		x		x					
PG	Perlecan	PGBM_MOUSE	Hspg2	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
PG	Byglican	PGS1_MOUSE	Bgn					x			x	x		x		x			x
PG	Decorin	PGS2_MOUSE	Dcn	x				x		x	x	x		x		x			
AP	Prolargin	PRELP_MOUSE	Prelp							x		x		x		x			x
AP	Annexin A2	ANXA2_MOUSE	Anxa2					x		x				x					
AP	Annexin A6	ANXA6_MOUSE	Anxa6							x									
AP	Galectin-3	LEG3_MOUSE	Lgals3											x					
AP	Galectin-9	LEG9_MOUSE	Lgals9													x		x	x
AP	Protein ERGIC-53	LMAN1_MOUSE	Lman1																x
AP	Pulmonary surfactant-associated protein B	PSPB_MOUSE	Sftpb											x					
AP	Pulmonary surfactant-associated protein A	SFTPA_MOUSE	Sftpa1											x					
REG	Alpha-1-antitrypsin 1-1	A1AT1_MOUSE	Serpina1a							x									
REG	Alpha-1-antitrypsin 1-2	A1AT2_MOUSE	Serpina1b			x				x				x				x	
REG	Alpha-1-antitrypsin 1-3	A1AT3_MOUSE	Serpina1c											x					
REG	Alpha-1-antitrypsin 1-4	A1AT4_MOUSE	Serpina1d			x				x									
REG	Lysyl oxidase homolog 1	LOXL1_MOUSE	Loxl1										x	x	x				
REG	Plasminogen	PLMN_MOUSE	Plg									x				x			x
REG	Serpin H1	SERPH_MOUSE	Serpinh1	x		x				x		x		x				x	x
REG	Serine protease inhibitor A3K	SPA3K_MOUSE	Serpina3k																x
REG	Protein-glutamine-glutamyltransferase 2	TGM2_MOUSE	Tgm2	x	x	x	x	x		x	x	x	x	x	x	x	x	x	x

* GP: Glycoprotein; COL: Collagen; PG: Proteoglycan; AP: Affiliated protein; REG: Regulator