

Proteomic Technologies for Tissue Engineering Investigation and Post-
translational Modification Discovery

By

Qiyao Li

A dissertation submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy

(Chemistry)

at the

UNIVERSITY OF WISCONSIN-MADISON

2016

Date of final oral examination: 03/28/2016

The dissertation is approved by the following members of the Final Oral Committee:

Lloyd M. Smith, Professor, Chemistry

Nathan V. Welham, Associate Professor, Surgery

Lingjun Li, Associate Professor, Chemistry/Pharmacy

John C. Wright, Professor, Chemistry

Ying Ge, Associate Professor, Chemistry/Cell and Regenerative Biology

Acknowledgments

I would like to thank my Ph.D. advisor, Professor Lloyd Smith, for all his guidance and support while allowing me the freedom to pursue research endeavors. The knowledge, perspective, attitude, and scientific writing skills that I learnt from him over the years at the very start of my scientific career, have been and will continue to be a treasure to me.

I thank the former and current members of the Smith group. They fostered an open and enjoyable environment for me to grow. I wish to extend a special thank you to Dr. Brian Frey, who has been a great mentor in every aspect of my work and sometimes also in life. Apart from scientific support, his patience and kindness really brought sunshine to my daily work, and helped me through the difficult times. I thank Dr. Michael Shortreed for all the trust, guidance, and very helpful discussions, and Dr. Mark Scalf for maintaining and teaching me a lot about the mass spectrometers and sharing experiences with me. I thank Dr. Gloria Sheynkman, and Dr. Ranran Liu for teaching me their first-hand technical lessons and for being great friends. I also thank Dr. Julia Kennedy-Darling and Matt Holden, with whom I shared not only an office but also much of our colorful Ph.D. life.

I also gratefully acknowledge my collaborator, Dr. Nathan Welham, from Dept. of Surgery at UW. His broad knowledge, great attention to details, and ability to produce exceptionally beautiful scientific illustrations, have greatly impressed and benefited me. His talks are always inspirational, and it has been my great pleasure to collaborate with such an amazing and intelligent scientist. Much of my Ph.D. work would not have been done without him. I also thank Dr. Zhen Chang, Dr. Changying Ling, and Dr. Kohei Nishimoto in the Welham group, Dr.

Basak Uygun, Dr. Harold Ott, and Dr. Sarah Gilpin from Massachusetts General Hospital, for fruitful and enjoyable collaborations.

I also thank Professors John Wright, Lingjun Li, and Ying Ge, for taking the time to be on my committee. I have also taken classes from all of them during my first two years, and benefited a lot. Professor Lingjun Li was my very first contact with UW-Madison five years ago when I applied for graduate school. I would not have met all the amazing people or done the work at Madison without her. She has been a role model to me since then, with her great kindness and intelligence.

Last but not least, I would like to thank my parents for raising me and for their unconditional love. They support me in all of my life decisions, including my pursuit of scientific research, and are always there for me. Also very importantly, I thank my husband Luyao Lu. Meeting him was the best thing that ever happened to me. He is my best friend, my rock, and my most enthusiastic cheerleader. He has been driving to me every other weekend from Chicago to Madison, and then from Urbana-Champaign to Madison, for the past five years, which adds up to more than 50,000 miles! I would have been lost without his constant belief in me, his sunny optimism, and his inspiration and encouragement. To the family that I shared this entire wonderful journey with, I dedicate this thesis.

Table of Contents

Abstract.....	vi
Section I	1
Chapter 1 Introduction: Research Background and Summary	2
1.1 The Emergence and Basics of Tissue Engineering.....	3
1.2 Scaffolds for Tissue Engineering	4
1.3 ECM – the Primary Constituent of a Biological Scaffold	5
1.4 Cell Reseeding	7
1.4.1 Cell Sources	7
1.4.2 Cell Types	8
1.4.3 Cell Numbers	9
1.4.4 Reseeding Strategies	9
1.5 Using Proteomic Tools to Guide Tissue Engineering Efforts.....	10
1.6 Research Summary	11
1.7 Figures and Tables.....	12
1.8 References	17
Chapter 2 Understanding Decellularized Biological Scaffolds for Tissue Engineering	19
2.1 Introduction.....	20
2.2 Optimization of Proteomic Sample Preparation	22
2.2.1 Background	22
2.2.2 Methods and Results	22
2.3 Proteomic Analyses of Various Types of Biological Scaffolds	23
2.3.1 Background	23
2.3.2 Materials and Methods.....	24
2.3.3 Results.....	31
2.3.4 Discussion	38
2.4 Decellularization of Porcine Vocal Fold Mucosae.....	43
2.4.1 Methods and Results	43
2.4.2 Discussion	44
2.5 Figures and Tables.....	45

2.6 References	59
Chapter 3 Investigating Protein Turnover during in vitro Tissue Remodeling	63
3.1 Introduction.....	64
3.2 Materials and Methods	65
3.3 Results	72
3.4 Discussion.....	80
3.5 Figures and Tables	83
3.6 References	97
Chapter 4 Characterizing Bioengineered Vocal Fold Mucosa	100
4.1 Introduction.....	101
4.2 Materials and Methods	103
4.3 Results and Discussion.....	106
4.4 Figures and Tables	109
4.5 References	118
Section II	120
Chapter 5 Global Post-Translational Modification Discovery	121
5.1 Introduction.....	122
5.2 Experimental Procedures	123
5.3 Results and Discussion.....	127
5.4 Figures and Tables	136
5.5 References	146
Chapter 6 Conclusions, Ongoing Projects, and Future Directions.....	148
6.1 Conclusions	149
6.2 Ongoing Project #1 - Regeneration of Whole Canine Larynx	149
6.2.1 Objectives	150
6.2.2 Methods.....	150
6.2.3 Preliminary Results	151
6.3 Ongoing Project #2 - Optimizing Lung Scaffolds for Organ Regeneration	152
6.3.1 Objectives	152
6.3.2 Methods.....	152

6.3.3 Preliminary Results	152
6.4 Future Directions	153
6.5 Figures and Tables.....	155

Proteomic Technologies for Tissue Engineering Investigation and Post-translational Modification Discovery

Qiyao Li

Under the supervision of Professor Lloyd M. Smith

At the University of Wisconsin-Madison

Abstract

This dissertation is separated into two main sections. Section I includes Chapters 1 to 4, presenting the work in developing and applying a proteomics method to help understand the multiple stages of tissue engineering. The tissue engineering technique based on recellularization of decellularized biological scaffolds has shown great promise in regenerating and restoring functions of several tissues/organs, and has the potential to replace orthotopic transplantation as the treatment for severe tissue damage or end-stage organ failure. However, the system-wide changes that happen at the molecular level still remain largely unknown. Chapter 1 introduces the background of this tissue engineering strategy and the motivation of introducing proteomics into the toolbox for tissue engineering investigation. Chapters 2 to 4 first briefly discuss the optimization of a proteomics method and then focus on its applications in addressing some of the key questions regarding protein changes during decellularization and recellularization.

Section II includes Chapter 5, which is about developing a bioinformatics tool for global post-translational modification (PTM) discovery. Traditional PTM identification tools either

require efficient pre-enrichment strategies and a relatively large amount of starting material, study just one or a few PTM types at a time, or result in high false discovery rate (FDR). Chapter 5 presents a G-PTM-D strategy that identifies a rich variety of peptide modifications with high confidence in complex unenriched biological samples.

Finally, Chapter 6 presents a brief conclusion of Sections I and II, describes two ongoing projects with some preliminary data, and points out a few future directions.

Below is a list of papers that Qiyao Li contributed to from 09/2011 to 03/2016.

1. **Li, Q.**, Shortreed, M. R., Wenger, C. D., Scalf, M., Frey, B. L., Smith, L. M. Global post-translational modification discovery. (*Submitted for publication*).
2. **Li, Q.**, Chang, Z., Oliveira, G., Xiong, M., Smith, L. M., Frey, B. L., Welham, N. V. Protein turnover during *in vitro* tissue remodeling. *Biomaterials* 81, 104-113 (2016).
3. **Li, Q.**, Uygun, B., Geerts, S., Sinan, O., Scalf, M., Gilpin, S. E., Ott, H. C., Yarmush, M. L., Smith, L. M., Welham, N. V., Frey, B. L. Proteomic analysis of naturally-sourced biological scaffolds. *Biomaterials* 75, 37-46 (2016).
4. Ling, C., **Li, Q.**, Brown, M. E., Kishimoto, Y., Toya, Y., Devine, E. E., Choi, K., Nishimoto, K., Norman, I. G., Tsegay, T., Jiang, J. J., Burlingham, W. J., Gunasekaran, S., Smith, L. M., Frey, B. L., Welham, N. V. Bioengineered vocal fold mucosa for functional voice restoration. *Science Translational Medicine* 7: 314ra187 (2015).
5. Liu, R., **Li, Q.**, Smith, L. M. Detection of Large Ions in Time-of-Flight Mass Spectrometry: Effects of Ion Mass and Acceleration Voltage on Microchannel Plate Detector Response. *Journal of the American Society for Mass Spectrometry* 25, 1374-83 (2014).

Section I

Chapter 1

Introduction: Research Background and Summary

1.1 The Emergence and Basics of Tissue Engineering

When a tissue or organ has a congenital anomaly, or has experienced severe disease or trauma, such that conventional pharmaceutical treatments are no longer applicable, allogeneic transplantation becomes the first choice to reconstruct the devastated tissue or organ¹. Although transplantation has experienced remarkable advances and improved or saved countless lives in the past decades, it remains an imperfect solution². One major challenge it faces is the gap which continues to widen between organ demand and supply, despite advances in medicine and increased awareness of organ donation. According to the United Network for Organ Sharing, on average, one patient is added to the national transplant waiting list every 10 minutes, and 21 people die each day while waiting for a transplant; more than 120,000 people are in need of a lifesaving organ transplant, but less than 30,000 transplants are performed within a year. Those patients fortunate enough to receive an organ are burdened with the risk of chronic rejection or the morbidity associated with lifelong immunosuppressant therapy³.

Approximately three decades ago, a new concept – tissue engineering–, emerged as an alternative approach to allogeneic tissue or organ transplantation. A commonly applied definition of tissue engineering, as stated by Langer and Vacanti, is “an interdisciplinary field that applies the principles of engineering and life sciences toward the development of biological substitutes that restore, maintain, or improve tissue function”². A clear goal of tissue engineering is to regenerate the patient’s own tissues or organs with great biocompatibility and high bio-functionality¹. Cells, scaffolds, and growth-stimulating signals are usually referred to as the tissue engineering triad, the key components of engineered tissues (Fig. 1.1)⁴. General strategies for the creation of a new tissue or organ involve the use of 1) isolated cells or cell substitutes that

supply the needed function, or 2) tissue-inducing substances such as growth factors, or 3) cells placed on or within scaffolds². Powerful developments in the multidisciplinary field of tissue engineering have yielded a novel set of tissue replacement parts with strategy 1) or 2). However, their limitations include failure of the infused cells to maintain function in the recipient, purification, large-scale production, and delivery of the signaling molecules, and most importantly, inability to regenerate large-volume whole organs. The work presented in this thesis is focused on strategy 3) – cells placed on or within scaffolds.

1.2 Scaffolds for Tissue Engineering

Over the last three decades, four major scaffolding approaches for tissue engineering have evolved, namely implanting cell-seeded pre-made porous scaffolds, implanting cell-seeded decellularized allograft or xenograft extracellular matrix (ECM), implanting laminated cell sheets with secreted ECM, and injecting cell encapsulated self-assembled hydrogels. Their characteristics are highlighted in Table 1.1⁵.

Each approach has its own pros and cons and preferred applications. In this work, the ultimate biological goal is to regenerate whole tissues or organs such as vocal fold mucosa (or even the whole larynx), liver, or lung, for human use. Therefore, we chose strategy #2 in Table 1.1, which involves the use of an acellular naturally occurring three-dimensional biological scaffold, obtained by decellularization of allogeneic or xenogeneic whole tissues or organs. The scaffold is subsequently recellularized with autologous stem cells or differentiated cells. Self-assembly of the reseeded cells in the scaffold could result in the formation of functional tissues or organs. A major advantage of this approach is the most close-to-nature mechanical and

biological properties of the ECM scaffold. Apart from the excellent biocompatibility of the natural ECM, growth factors preserved in the decellularized scaffolds may further facilitate cell growth and remodeling activities. Other advantages include the ready availability of the off-the-shelf scaffolds and the obviation of immunosuppression due to the use of autologous cells. Furthermore, the retention of an intact vascular network with appropriately sized inflow and outflow conduits in the decellularized organs enable effective perfusion with nutrients and appropriate cues for cell behavior and convenient surgical connections to the recipient circulation.

1.3 ECM – the Primary Constituent of a Biological Scaffold

A decellularized biological scaffold is primarily composed of ECM, the non-cellular component present in all tissues and organs consisting of numerous structural and functional molecules. These molecules are arranged in a tissue-specific, three-dimensional manner, and are ideally suited to the tissue or organ from which the ECM is obtained. They affect the phenotype, proteome, and functionality of the resident cells. In turn, the structure and composition of the ECM are constantly changing in response to the resident cell activities and the mechanical demands of the tissue. This dynamic reciprocity between the ECM and the resident cell population is a major advantage for the use of the decellularized ECM scaffolds over synthetic materials and underscores the importance of maintaining as much of the native composition and ultrastructure as possible during preparation of these ECM scaffolds.

In general, the ECM molecules are highly conserved in eukaryotic organisms⁶⁻⁸, which minimizes adverse immune responses after xenotransplantation. Much attention has been given to certain antigens such as galactose- α -1,3-galactose (i.e., gal-epitope), which are present in

some xenogeneic ECM materials^{9, 10} and associated with hyperacute rejection. However, the matrix-associated gal-epitope antigens fail to activate complement or bind IgM antibody, presumably because of their relatively small amount and wide distribution³. In addition, a recent study suggests that xenogeneic ECM may elicit a necessary M2 phenotype profile of the responding macrophages, to support a constructive remodeling¹¹. Thus, use of decellularized scaffolds as a foundation for tissue/organ regeneration remains a viable option.

Researchers who are interested in the ECM are faced with a great challenge when it comes to its definition. The commonly used gene ontology (GO) categories prove to be not good at defining ECM proteins. Many ECM proteins are unannotated or sorted into diverse categories such “extracellular matrix,” “cell surface”, “basement membrane,” etc., with little logic or consistency¹². Therefore, there remains a pressing need for a better definition of the various ECM proteins. Molecular biology and protein chemistry have revealed that ECM proteins typically consist of repeated domains¹³. Recently, Naba *et al* proposed that it is the domain architecture, including composition, number, and order, that defines families of ECM proteins, and they developed a bioinformatics approach to predict the ensemble of genes encoding what they define as the “matrisome”, namely all components constituting the ECM (the “core matrisome”) and those associated with them (“matrisome-associated” proteins)¹².

As shown in Fig. 1.2, the core matrisome comprises three subcategories – collagens, ECM glycoproteins, and proteoglycans. These proteins are believed to assemble into ECM. The three subcategories of matrisome-associated proteins are more inclusively defined, including proteins known to associate with assembled ECM as well as proteins that may not. The secreted factors category includes diverse growth factors, cytokines and other secreted proteins. The ECM regulator category includes a list of proteases, protease inhibitors and ECM crosslinking

enzymes. The ECM-affiliated proteins category includes protein families that may be considered as ECM proteins (e.g., mucins, C-type lectins) or as secreted factors but which also associate with solid-phase complexes (e.g., collagen-related proteins such as C1q and homologs), and a few families that repeatedly appear in ECM-enriched preparations for currently unknown reasons (e.g., annexins, galectins).

1.4 Cell Reseeding

The source and type of cells are critical to the eventual functionality and clinical success of the engineered construct. An ideal cell population would be one that can proliferate as needed and give rise to the heterogeneous cells necessary to form a functional tissue/organ³.

1.4.1 Cell Sources

One decision to make is autologous versus allogeneic cells. Autologous cells are self-derived and thus limit the possibility for exposure to transmissible agents and are less likely to trigger an adverse immune response. In the case of a highly regenerative organ (e.g., liver), where organ biopsy is an option, autologous cells might be a feasible cell choice because they can be harvested and expanded (at least minimally) *in vitro*. However, in other organs (e.g., pancreas, lung, heart), autologous cells cannot easily be harvested or the numbers are often insufficient. In addition, not all autologous cells can be expanded to large numbers either because they do not readily divide (e.g., cardiocytes, neurons) or because they are difficult to maintain in an undifferentiated proliferative state (e.g., hematopoietic bone marrow cells)³.

Allogeneic cells, in contrast, can often be harvested in larger quantities and from healthier, younger individuals. They can be isolated, expanded, characterized, and generate an

off-the-shelf product that is available as needed. Therefore, whether to use autologous or allogeneic cells is often based on criteria such as (a) the ease of cell harvest and expansion, (b) the number of required cells and the timeframe in which they are needed, and (c) the ability to differentiate *in vitro*³.

Except in cases where specific organ-derived differentiated cell types can be isolated, expanded, and used, the most likely candidate to fulfill the demand of regenerating a functional organ is stem or progenitor cells.

Stem or progenitor cells utilized for most tissue-engineering approaches can be broadly categorized as embryonic stem cells, fetal cells, adult (including umbilical cord blood)-derived stem or progenitor cells, and adult-derived inducible pluripotent stem cells. Non-stem or progenitor cells used are usually parenchymal and supportive cells (e.g., fibroblasts). Alternatively, the cell sources can include vascular cells obtained from easily accessible sources such as peripheral blood or bone marrow (e.g., endothelial cells)³. The cells used in this work are primary cells isolated via biopsy or surgical harvest from the tissue of interest.

1.4.2 Cell Types

The type of cells for recellularization depends significantly on the tissue/organ. Parenchymal cells (e.g., hepatocytes in liver, cardiomyocytes in heart, epithelium in lung), the cell types responsible for the specific functions of the tissue/organ, are of obvious necessity. Non-parenchymal cells (e.g., fibroblasts, endothelial cells) enhance the functional phenotype of the parenchymal cells and contribute to the organization of the tissue architecture^{14, 15}. Fibroblasts secrete and remodel the ECM and improve parenchymal cell functions in co-cultures.

Although fibroblasts are recruited as necessary *in vivo*, they are also important during *in vitro* tissue/organ culture to facilitate ECM remodeling and cellular organization.

1.4.3 Cell Numbers

The number of cells initially required for seeding the scaffold depends on the organ type. For instance, the biomechanical duties of heart and lung require a functional whole organ at the time of implantation, and thus a high percentage of original cell numbers to be immediately present. By contrast, organs with primarily metabolic functions, such as pancreas and liver, can live with only a small portion of the native cell mass.

1.4.4 Reseeding Strategies

The recellularization process typically consists of cell seeding, the goal of which is redistribution of cells similar to their *in vivo* spatial configuration, followed by perfusion culture, aiming to prepare the cells for *in vivo* function. As with decellularization, the optimal reseeded method varies depending on tissue/organ types.

There are a variety of methods used for seeding cells into a three-dimensional scaffold with the aim of achieving uniform cell distribution¹⁶. In the passive methods of seeding, cells are laid on top of the scaffold and allowed to infiltrate into the porous scaffold over time. The main disadvantage, particularly for thick scaffolds, is that the infiltration rates can be very slow, resulting in poor loading and uneven distribution of cells within the scaffolds. In the active methods of seeding, an external force is applied to encourage the cells to infiltrate into the scaffold¹⁷. The nature of the force varies from centrifugal, compression force, to chemo-attraction. Nevertheless, it is still difficult to seed cells into high-density connective tissues which have few interstitial spaces.

1.5 Using Proteomic Tools to Guide Tissue Engineering Efforts

Although public and medical communities have great expectations on tissue engineering or regenerative medicine, reported human clinical trials are still scant. One of the reasons for the delayed progress in clinical applications of tissue engineering may be the lack of knowledge on the molecular-level changes that happen during the decellularization and recellularization processes. For example, regeneration and transplantation of a bio-artificial lung was successfully carried out in a rat model, but examination of the lung after transplantation indicated leakage of erythrocytes into the alveolar space, indicating a compromised capillary-endothelial barrier^{18, 19}. Such results highlight the potential of the “decellularization and recellularization” method for organ regeneration but also the need for improved molecular readouts to guide engineering efforts.

Efficient re-seeding of decellularized scaffolds depends on the structural integrity and elasticity of the scaffolds. Local variations in proteins in the scaffolds (collagens, laminins, fibronectins) have been correlated to variance in cell repopulation and subsequent proliferation²⁰. It has been postulated that scaffold components and ultrastructures may direct cells back to a tissue specific niche during reseeding, and that small changes in the abundance of these molecular cues can drastically affect the recellularization process. Current methods used to characterize the protein composition of native and decellularized tissues involve antibody or dye based staining, and hydroxyproline assays for collagen assessment. However, these methods fall short in the comprehensiveness in characterization and thus cannot provide a system-wide view or discover previously unknown components that are possibly crucial in the final clinical success.

In this work, we aimed to introduce proteomic techniques to the tissue engineering toolbox, in an attempt to shed new light on the decellularization and recellularization processes at the system-wide molecular level. Proteomics is the analysis of the entire protein complement expressed by a genome, or by a cell or tissue type. Proteomics is often performed with liquid chromatography tandem mass spectrometry (LC-MS/MS) and the typical workflow of shotgun proteomics is depicted in Fig. 1.3²¹. Given the vast amount of MS raw data that are acquired in a typical proteomics workflow, computational tools have been developed to extract qualitative and quantitative information from large-scale data sets in a streamlined manner (Fig. 1.4)²¹.

1.6 Research Summary

Section I of this thesis summarizes my research efforts in understanding tissue engineering processes with proteomics. In this first chapter, a brief introduction of the decellularization and recellularization-based tissue engineering technique was given. Chapter 2 first describes the method optimization of proteomic sample preparation, followed by application of the optimized method to a variety of native tissues and their decellularized scaffolds. Chapter 3 presents how we use a stable isotope labeling strategy to investigate protein turnover after repopulating the decellularized vocal fold mucosa with vocal fold fibroblasts. Chapter 4 studies a bioengineered vocal fold mucosa that is capable of voice restoration.

1.7 Figures and Tables

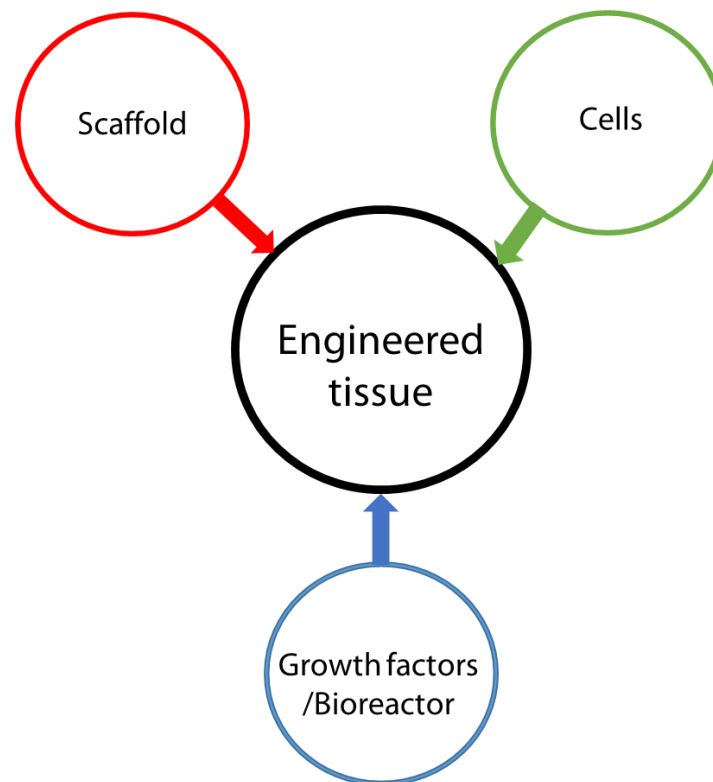


Figure 1.1. Tissue engineering triad of cells, signals (provided chemically by growth factors or physically by a bioreactor), and the scaffold acting as a template for tissue formation.

The Human Matrisome

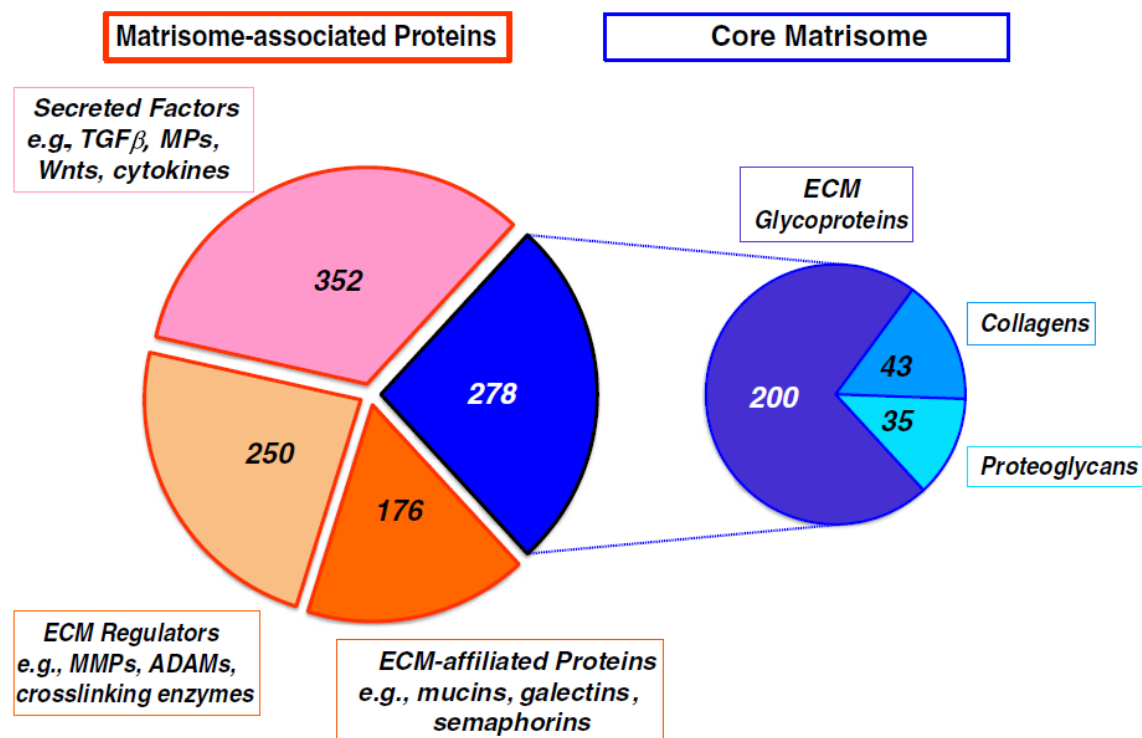


Figure 1.2. The human matrisome and its subcategories. Reprinted from ref 16, Copyright (2016), with permission from Elsevier.

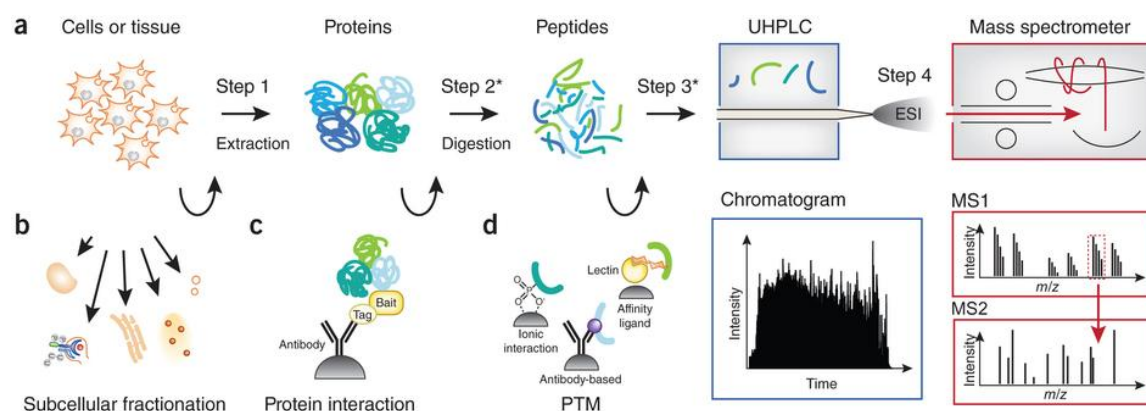


Figure 1.3. (a) The generic workflow of modern LC-MS/MS-based proteomics consists of four steps. In step 1, proteins are extracted from tissues, body fluids, cells or subcellular compartments. In step 2, proteins are proteolytically digested. In step 3, peptides are separated by UHPLC. In step 4, peptides are ionized by electrospray (electrospray ionization, ESI), and their masses and fragment masses are acquired in a mass spectrometer. (b–d) Variations of the workflow that include enrichment steps for proteins in subcellular compartments (b), for interacting proteins (c) and for peptides with PTMs (d). At steps 2 and 3 (*), additional fractionation of proteins or peptides is possible. Reprinted from ref 25, Copyright (2016), with permission from Nature Publishing Group.

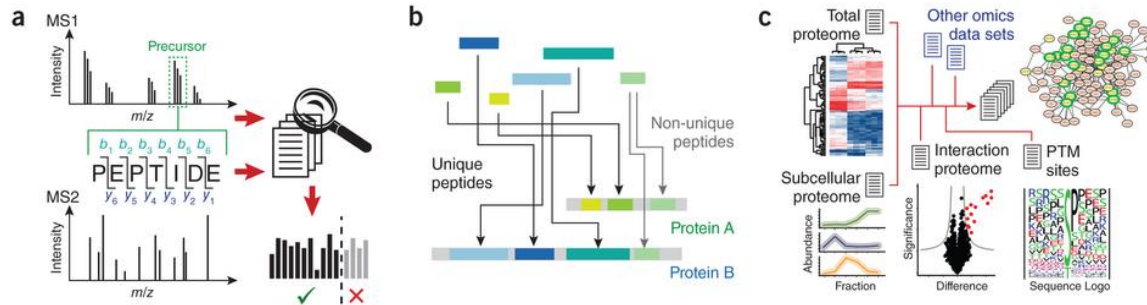


Figure 1.4. (a) Identification of peptides from the exact peptide mass determined in the MS1 scan (precursor mass) and its characteristic fragmentation patterns of the MS2 scan. Peptide and fragment masses are searched in an organism-specific sequence database and then scored. Identifications are evaluated with a target-decoy-based FDR threshold. (b) Peptides are mapped to proteins in an FDR-controlled approach. (c) Presentation and integration of data. Examples for the illustration of total proteome data with unsupervised hierarchical clustering, subcellular proteomes with correlation profiling, interaction proteomics with a volcano plot and peptide motif identification for PTMs. Data from all omics sources can be combined for systems-wide analysis of immune function. Reprinted from ref 25, Copyright (2016), with permission from Nature Publishing Group.

Table 1.1. Characteristics of different scaffolding approaches in tissue engineering.

<i>Scaffolding approach</i>	(1) Pre-made porous scaffolds for cell seeding	(2) Decellularized extracellular matrix for cell seeding	(3) Confluent cells with secreted extracellular matrix	(4) Cell encapsulated in self-assembled hydrogel
<i>Raw materials</i>	Synthetic or natural biomaterials	Allogenic or xenogenic tissues	Cells	Synthetic or natural biomaterials able to self-assemble into hydrogels
<i>Advantages</i>	Most diversified choices; precise design for microstructure and architecture	Most nature-simulating in terms of composition and mechanical properties	Biocompatible	Injectable, fast and simple one-step procedure; intimate cell and material interactions
<i>Disadvantages</i>	Time consuming cell seeding procedure; inhomogeneous distribution of cells	Inhomogeneous distribution of cells, difficulty in retaining all extracellular matrix, immunogenicity upon incomplete decellularization	Need multiple laminations	Soft structures
<i>Preferred applications</i>	Both soft and hard tissues; load-bearing tissues	Tissues with high ECM content; load-bearing tissues	Tissues with high cellularity, epithelial tissues, endothelial tissues, thin layer tissues	Soft tissues

1.8 References

1. Ikada, Y. Challenges in tissue engineering. *J R Soc Interface* **3**, 589-601 (2006).
2. Langer, R. & Vacanti, J.P. Tissue engineering. *Science* **260**, 920-926 (1993).
3. Badylak, S.F., Taylor, D. & Uygun, K. Whole-organ tissue engineering: decellularization and recellularization of three-dimensional matrix scaffolds. *Annu Rev Biomed Eng* **13**, 27-53 (2011).
4. O'Brien, F.J. Biomaterials & scaffolds for tissue engineering. *Materials Today* **14**, 88-95 (2011).
5. Chan, B.P. & Leong, K.W. Scaffolding in tissue engineering: general approaches and tissue-specific considerations. *Eur Spine J* **17**, 467-479 (2008).
6. Bernard, M.P. et al. Nucleotide sequences of complementary deoxyribonucleic acids for the pro alpha 1 chain of human type I procollagen. Statistical evaluation of structures that are conserved during evolution. *Biochemistry* **22**, 5213-5223 (1983).
7. Constantinou, C.D. & Jimenez, S.A. Structure of cDNAs encoding the triple-helical domain of murine alpha 2 (VI) collagen chain and comparison to human and chick homologues. Use of polymerase chain reaction and partially degenerate oligonucleotide for generation of novel cDNA clones. *Matrix* **11**, 1-9 (1991).
8. Exposito, J.Y., D'Alessio, M., Solursh, M. & Ramirez, F. Sea urchin collagen evolutionarily homologous to vertebrate pro-alpha 2(I) collagen. *J Biol Chem* **267**, 15559-15562 (1992).
9. McPherson, T.B., Liang, H., Record, R.D. & Badylak, S.F. Galalpha(1,3)Gal epitope in porcine small intestinal submucosa. *Tissue Eng* **6**, 233-239 (2000).
10. Raeder, R.H., Badylak, S.F., Sheehan, C., Kallakury, B. & Metzger, D.W. Natural anti-galactose alpha1,3 galactose antibodies delay, but do not prevent the acceptance of extracellular matrix xenografts. *Transpl Immunol* **10**, 15-24 (2002).
11. Badylak, S.F., Valentin, J.E., Ravindra, A.K., McCabe, G.P. & Stewart-Akers, A.M. Macrophage phenotype as a determinant of biologic scaffold remodeling. *Tissue Eng Part A* **14**, 1835-1842 (2008).
12. Naba, A., Hoersch, S. & Hynes, R.O. Towards definition of an ECM parts list: an advance on GO categories. *Matrix Biol* **31**, 371-372 (2012).
13. Hynes, R.O. & Naba, A. Overview of the matrisome--an inventory of extracellular matrix constituents and functions. *CSH Perspect Biol* **4**, a004903 (2012).
14. Morin, O. & Normand, C. Long-term maintenance of hepatocyte functional activity in co-culture: requirements for sinusoidal endothelial cells and dexamethasone. *J Cell Phys* **129**, 103-110 (1986).
15. Brutsaert, D.L. Cardiac endothelial-myocardial signaling: its role in cardiac growth, contractile performance, and rhythmicity. *Phys Rev* **83**, 59-115 (2003).
16. Thevenot, P., Nair, A., Dey, J., Yang, J. & Tang, L. Method to analyze three-dimensional cell distribution and infiltration in degradable scaffolds. *Tissue Eng Part C Methods* **14**, 319-331 (2008).
17. Cimetta, E. et al. Enhancement of viability of muscle precursor cells on 3D scaffold in a perfusion bioreactor. *Int J Artif Organs* **30**, 415-428 (2007).

18. Petersen, T.H. et al. Tissue-engineered lungs for in vivo implantation. *Science* **329**, 538-541 (2010).
19. Ott, H.C. et al. Regeneration and orthotopic transplantation of a bioartificial lung. *Nat Med* **16**, 927-933 (2010).
20. Daly, A.B. et al. Initial binding and recellularization of decellularized mouse lung scaffolds with bone marrow-derived mesenchymal stromal cells. *Tissue Eng Part A* **18**, 1-16 (2012).
21. Meissner, F. & Mann, M. Quantitative shotgun proteomics: considerations for a high-quality workflow in immunology. *Nat Immunol* **15**, 112-117 (2014).

Chapter 2

Understanding Decellularized Biological Scaffolds for Tissue Engineering

Part of this chapter is adapted from:

“Proteomic analysis of naturally-sourced biological scaffolds.” **Q. Li**, B. Uygun, S. Geerts, O. Sinan, M. Scalf, S. E. Gilpin, H. C. Ott, M. L. Yarmush, L. M. Smith, N. V. Welham, B. L. Frey. *Biomaterials*. 2016, 75, 37-46.

2.1 Introduction

A promising tissue engineering approach for functional organ replacement involves using synthetic or biological scaffolds, within which seeded cells, potentially from the transplant recipient, grow to form new tissues. In contrast to synthetic scaffolds, biological scaffolds are composed of naturally occurring extracellular matrix (ECM) and retain much of the correct anatomical structure, native ECM architecture, and biomechanical properties, as well as many of the cell adhesion ligands found in living tissues. These beneficial characteristics also modulate the immunological response, and thus biological scaffolds have been widely used in various cell culture and tissue reconstruction applications. These scaffolds range from simple purified proteins, to extracted protein mixtures, to very complex decellularized whole organs. A key challenge to achieving the ultimate therapeutic goal – creation of a functional organ – lies in understanding the process of removing cells and immunogenic material from a donor organ while maintaining the biochemical and biophysical properties of the scaffold that will promote growth of newly seeded cells.

The composition of biological scaffolds varies depending on their source, as well as the processes employed to decellularize them. The most effective agents for decellularization of each tissue and organ depend on many factors, such as the tissue's cellularity (e.g. liver vs. tendon), density (e.g. dermis vs. adipose tissue), lipid content (e.g. brain vs. urinary bladder), and thickness (e.g. dermis vs. pericardium)¹. Every cell removal method changes ECM composition and causes some degree of ultrastructure disruption. Minimization of these undesirable effects is the objective of decellularization. An overview of some commonly used agents (e.g. chemical, enzymatic, and physical) and their modes of action is provided in Table 2.1.

Current criteria for evaluating whole organ decellularization are based primarily on nucleic acids, as they are easy to quantify and have been directly correlated to adverse host responses¹. However, other cellular materials, such as proteins, are likely culprits in eliciting immunogenic responses²⁻⁴, and therefore additional analyses, such as proteomics, will greatly enhance understanding of the decellularization process and provide improved molecular readout to guide engineering efforts.

In the studies presented in this chapter, we first compared different sample preparation strategies for subsequent LC-MS/MS analysis in terms of proteome coverage. Filter-aided sample preparation (FASP)⁵ following protein extraction with SDS and DTT stood out as the one that provided in-depth coverage for both cellular and ECM proteins, making it the optimal choice for evaluating decellularized scaffolds. We then employed the FASP method of protein extraction and digestion to analyze four types of biological scaffolds: rat-tail type I collagen, growth factor reduced (GFR)-Matrigel, decellularized rat livers, and decellularized human lungs. We also characterized the early- versus late-stage of the decellularization process in the rat liver model by measuring protein abundances for naïve, partially-decellularized, and “fully” decellularized samples. In addition, we employed the optimized protocol to evaluate decellularization effectiveness of porcine vocal fold mucosae decellularized by five different strategies.

2.2 Optimization of Proteomic Sample Preparation

2.2.1 Background

ECM proteomics is attracting increased attention due to the crucial role of the ECM in cell development and important physiological and pathological processes⁶⁻⁸. ECM proteins have been notoriously difficult to analyze because of their covalent crosslinking and numerous modifications^{6, 9}. Researchers have developed various protein extraction strategies and have performed extensive fractionation at the peptide and/or protein level, in order to obtain deeper coverage of ECM proteins. When evaluating scaffolds for tissue engineering applications, however, the identification of remnant cellular proteins is of equal importance to the analysis of ECM proteins.

2.2.2 Methods and Results

Protein extraction and digestion strategies were evaluated based on the amount of extracted protein (BCA Assay, 660 nm protein assay, data not shown) along with LC-MS/MS proteomic results (numbers of ECM proteins, all proteins, unique peptides, and peptide spectral matches – PSMs). Protein extraction methods based upon sonication in a chaotrope-containing buffer solution (8M urea) were adequate for many cellular proteins, but as anticipated, failed to solubilize many ECM proteins (Fig. 2.1). Thus, we tried sonication in a detergent-containing buffer “SDT” (4% SDS, 0.1M DTT, 0.1M Tris). To be specific, 150 μ L SDT extraction solution was added to 10-20 mg minced tissue pieces. Samples were subject to ultrasonic homogenation (300 V/T; Biologics, Manassas, VA) on ice for 6 min in total with 20s on and 20s off, and were placed in 95 $^{\circ}$ C water bath for 7 min. After being cooled to room temperature, samples were centrifuged at 16100 g for 5 min. This SDT-sonication method has yielded substantial

improvement over the chaotrope methods (Fig. 2.1). Protein extracts were prepared for LC-MS/MS analysis using several steps. The first step was digestion with trypsin, which turned out to outperform LysC (Fig. 2.2) using a protocol called filter-aided sample preparation (FASP)¹¹, which removes SDS and allows efficient enzymatic digestion and peptide recovery. In short, 30 μ L of supernatant from SDT-sonication was captured on a molecular weight cut-off filter (typically 30kDa). On the filter, detergent was washed away using 8M Urea. After an on-filter digestion, the generated peptides were small enough to pass through the filter and were collected for analysis. The SDT-FASP-trypsin method was also compared to collagenase I digestion followed by trypsin digestion, or collagenase I digestion followed by the SDT-FASP-trypsin method. The SDT-FASP-trypsin method still performed the best in terms of ECM and cellular proteome coverage.

2.3 Proteomic Analyses of Various Types of Biological Scaffolds

2.3.1 Background

Three types of biological scaffolds of increasing complexity – collagen, Matrigel, and decellularized organs – were studied in this work. Purified collagen is often employed in gel form as a very simple biological matrix, providing structure but limited molecular heterogeneity. Type I collagen gel has long been used in studies of cell proliferation and differentiation to sustain continuous cell growth^{10, 11}, or to promote expression of cell-specific morphology and function^{12, 13}. It has also been used for the reconstruction of various tissues such as thyroid follicles¹⁴, cornea¹⁵, and urinary bladder mucosa¹⁶. Another commonly used biological matrix is Matrigel, a mixture of basement membrane proteins extracted from the Engelbreth-Holm-Swarm

mouse tumor. It is often used to mimic the ECM in cancer research and in stem cell studies due to its ability to maintain stem cells in an undifferentiated state. Matrigel, either alone¹⁷⁻¹⁹ or in combination with collagen gel²⁰⁻²², has been shown to recapitulate certain tissue-like structures. The most sophisticated biological scaffolds are decellularized whole tissues or organs. Such decellularized scaffolds retain the organ's macroscopic three-dimensional architecture and organ-specific cues that can support the growth of cells to reconstruct the original organ type²³. After seeding these allogeneic or xenogeneic acellular scaffolds with organ-matched cell populations, researchers have created partially functional replacements for key organs including heart²⁴, liver²⁵, lung²⁶⁻²⁸, and kidney²⁹. These preclinical studies demonstrate the potential of this strategy to transform human medicine. A common characteristic of these naturally occurring scaffolds is their ability to support cell growth and even tissue regeneration. Although not yet well understood, it is very likely that both tissue structure and composition contribute to this capability. Thus, a deeper knowledge of the composition of these scaffolds is expected to have high value in elucidating the mechanisms that drive their functional capacity.

2.3.2 Materials and Methods

Tissue Acquisition and Decellularization—All animal care, handling and surgical procedures were performed in accordance with the guidelines set by the Institutional Animal Care and Use Committee (IACUC) at Massachusetts General Hospital. Livers were harvested from female Lewis rats (Charles River Laboratories, Wilmington, MA) weighing 175–200 g, gently flushed with and stored in 0.9% NaCl at –80 °C until utilized for decellularization. Organ decellularization was performed as previously described³⁰. Briefly, the liver was thawed at room temperature and washed by portal vein perfusion with PBS overnight at a flow rate of 1.0

mL/min to clear residual blood from the organ. Then, the liver was perfused with 0.01% (w/v) SDS (Sigma-Aldrich, St. Louis, MO) for increasing durations of 5, 10, 15 and 20 min with 60 min PBS washes in between. After the last PBS wash, the liver was perfused with 0.01% SDS for 24 h followed by perfusion with 0.1% (w/v) SDS for 24 h, and 0.2% (w/v) SDS for 2 h. Next, the liver was perfused with 0.1% (w/v) Triton X-100 (Sigma-Aldrich, St. Louis, MO) for 30 min to remove the residual components and rinsed with PBS for at least 2 h. Tissue biopsies (50–100 mg) were taken at three separate time points during this process: 1) immediately after harvest (naïve), 2) after the 24 h wash with 0.01% SDS (partially decellularized), and 3) at the end of the final PBS wash (decellularized). These liver tissues were flash frozen for DNA, total collagen (see Supplemental information) and proteomic analyses. Human lungs were obtained from the International Institute for the Advancement of Medicine (IIAM); they were recovered under sterile conditions within 60 min of cessation of cardiovascular circulation, but were otherwise unsuitable for transplantation. Perfusion decellularization was performed according to the previously published methodology^{31, 32}.

DNA Measurement—Genomic DNA from rat liver samples (naïve, partially decellularized and decellularized) were extracted using the PureLink® Genomic DNA Kit (Life Technologies, Grand Island, NY). The DNA concentration in the extracts was measured using the PicoGreen® Assay Kit (Life Technologies, Grand Island, NY).

Protein Extraction and Digestion—For rat-tail type I collagen and GFR-Matrigel (both from BD Bioscience, San Jose, CA), aliquots of approximately 200 µg of protein were added to 150 µL of SDT solution—4% SDS, 0.1 M Tris-HCl (pH 7.6) and 0.1 M dithiothreitol (DTT) (all reagents from Sigma-Aldrich, St. Louis, MO). For decellularized or naïve tissues, aliquots of approximately 15 mg of tissue were washed with ice-cold PBS solution (Invitrogen, Grand

Island, NY) and ground with disposable pellet pestles (Kimble Chase Kontes, Vineland, NJ) for 1 min in 1.5-mL tubes, followed by addition of 150 μ L SDT solution. Samples were then heated at 95 $^{\circ}$ C for 7 min. and sonicated on ice with a probe sonicator (Misonix XL2015, Misonix microtip PN/418, Farmingdale, NY)—alternating 20 seconds on and 20 seconds off for 6 min, followed by centrifugation at 22 $^{\circ}$ C for 5 min. at 16,100 g.

The FASP protocol was used for SDS removal and on-filter digestion. Briefly, a 30 μ L aliquot of the sample supernatant was mixed with 200 μ L of 8M urea/0.1 M Tris buffer pH 8.0 in a 30K MW Vivacon 500 filter (Sartorius, Bohemia, NY). The sample was washed, alkylated with iodoacetamide, washed further, then digested with trypsin (Promega, Madison, WI; protein:enzyme ratio of 50:1) overnight at 37 $^{\circ}$ C, and the digested peptides were collected by centrifugation. After the digestion was quenched with 10% trifluoroacetic acid (TFA) to a final concentration of 0.5% TFA, samples were desalted using Sep-Pak C18 1 cc Vac Cartridges (Waters, Milford, MA). The cartridge was wetted with 2 $\times$ 200 μ L acetonitrile (ACN) and equilibrated with 150 μ L H₂O. Tryptic peptides were loaded onto the sorbent, and then washed with 2 $\times$ 100 μ L 0.1% TFA and 50 μ L 0.1% formic acid (FA). Finally, peptides were eluted with 2 $\times$ 100 μ L of 75% ACN, 0.1% FA. The eluent was dried down using a Savant SpeedVac (Thermo Scientific, Waltham, MA) and reconstituted in 5% ACN, 2% FA.

High-pH Fractionation—As noted in the Results section, certain digested peptide samples were fractionated prior to LC-MS. In such cases, Sep-Pak desalting was skipped, and the quenched digests were subjected to high-pH fractionation on an HPLC system (Shimadzu, Columbia, MD) using a C18 Gemini 3 μ , 110Å, 3.0 $\times$ 150mm column (Phenomenex, Torrance, CA). Mobile phase A was aqueous 20 mM ammonium formate and mobile phase B was 20 mM ammonium formate in 70% ACN; the gradient of 0–100% mobile phase B occurred over 20 min.

The HPLC flow rate was 1 mL/min and the eluent was collected into 6 fractions, each of which was evaporated to dryness in the SpeedVac and reconstituted in 5% ACN, 2% FA.

Liquid Chromatography-Tandem Mass Spectrometry (LC-MS/MS)—Approximately 0.6 µg protein digest, estimated by BCA protein assay, was injected into a Waters nanoAcquity HPLC coupled to an ESI ion-trap/Orbitrap mass spectrometer (LTQ Orbitrap Velos, Thermo Scientific, Waltham, MA). Peptides were separated on a 100 µm inner diameter column packed with 20 cm of 3 µm MAGIC aqC18 beads (Bruker-Michrom, Auburn, CA), and eluted at 0.3 µL/min in 0.1% FA with a gradient of increasing ACN over 2.5 h. As noted in the Results section, certain samples were separated on a column packed with 20 cm of 1.7 µm BEH C18 particles (Waters, Milford, MA), and a heater cartridge was used to keep the capillary column at 60 °C. A full-mass scan (300-1500 m/z) was performed in the Orbitrap at a resolution of 60,000. The ten most intense peaks were selected for fragmentation by higher-energy collisional dissociation (HCD) at 42% collision energy, then analyzed with a resolution of 7500 and an isolation width of 2.5 m/z . Dynamic exclusion was enabled with a repeat count of 1 over 30 s and an exclusion duration of 120 s. All MS raw files may be downloaded from the PeptideAtlas data repository³³ by the following link: <http://www.peptideatlas.org/PASS/PASS00557>.

Proteomic Data Analysis—The acquired raw files were analyzed by MaxQuant^{34, 35} (version 1.4.1.2). The resulting peak lists were searched with Andromeda³⁶ against UniProt canonical protein databases for the appropriate organism (Rattus norvegicus: 7853 reviewed sequences downloaded on May 4, 2013 for rat-tail collagen search, and 33,607 sequences (reviewed plus unreviewed) downloaded on May 29, 2014 for rat liver searches; Mus musculus: 16,642 reviewed sequences downloaded on December 16, 2013; Homo sapiens: 20,278 reviewed sequences downloaded on December 5, 2013), each supplemented with 262 common

contaminants. Precursor and fragment ion mass tolerances were set to 4.5 ppm and 20 ppm, respectively. Static cysteine carbamidomethylation (+57.0215 Da) and up to 5 variable methionine and proline oxidations (+15.9949 Da) were specified. The modification parameters differed somewhat for the collagen gel search: oxidation of lysine (+15.9949 Da) and glycosylation of lysine (monosaccharide, +178.0473 Da, disaccharide, +340.0995 Da) were also included as variable modifications, and up to 7 variable modifications were allowed. A false discovery rate of 1% at both the peptide and the protein level was allowed. Up to two missed cleavages were allowed and a minimum of two unique peptides per protein was required. Protein groups containing matches to proteins from the reversed database or contaminants were discarded. Intensity-based absolute quantification (iBAQ) and label-free quantification (LFQ) algorithms embedded in the MaxQuant software package were employed. Only unique and razor peptides were used for quantification and a minimum count of two was required. We multiplied the LFQ intensity of the part-decell and the naïve sample by correction factors of 10.3 and 28.4, respectively, based on tissue wet and dry weight information (see paragraph below). Perseus software (version 1.5.0.15) was used for downstream statistical analyses. Proteins were filtered by requiring at least two valid values in at least one sample condition. The corrected intensities were log2 transformed and missing values were replaced using data imputation by employing a width of 0.3 and a downshift of 1.2.

Correction factors calculation for quantitative proteomics comparison of naïve, partially-decellularized (part-decell) and decellularized (decell) rat livers—Wet weight and dry weight were measured for the naïve, part-decell and decell rat livers with each having four biological replicates. The following data were derived from those measurements with linear fits:

$$\frac{\text{naïve wet weight}}{\text{naïve dry weight}} = \frac{1}{0.2392},$$

$$\frac{\text{partdecell wet weight}}{\text{partdecell dry weight}} = \frac{1}{0.093},$$

$$\frac{\text{decell wet weight}}{\text{decell dry weight}} = \frac{1}{0.0284},$$

$$\frac{\text{naïve dry weight}}{\text{decell dry weight}} = \frac{1}{0.0525},$$

$$\frac{\text{partdecell dry weight}}{\text{decell dry weight}} = \frac{1}{0.1329}$$

An estimate of 0.6 µg protein, which corresponded to 0.033 mg, 0.092 mg and 0.414 mg wet weight for the naïve, part-decell and decell tissue, respectively, was injected for each LC-MS/MS run. To normalize the intensity of each sample condition accounting for the large total protein and tissue weight differences among samples, we employed the following normalization calculation:

Naïve:

$$\frac{\text{naïve LFQ intensity}}{0.033} \times \frac{\text{naïve wet weight}}{\text{naïve dry weight}} \times \frac{\text{naïve dry weight}}{1}$$

Part-decell:

$$\frac{\text{partdecell LFQ intensity}}{0.092} \times \frac{\text{partdecell wet weight}}{\text{partdecell dry weight}} \times \frac{\text{partdecell dry weight}}{1}$$

Decell:

$$\frac{\text{decell LFQ intensity}}{0.414} \times \frac{\text{decell wet weight}}{\text{decell dry weight}} \times \frac{\text{decell dry weight}}{1}$$

So, Naïve / Decell = $\frac{\text{naïve LFQ intensity}}{\text{decell LFQ intensity}} \times \frac{28.4}{1}$;

$$\text{Part-decell / Decell} = \frac{\text{partdecell LFQ intensity}}{\text{decell LFQ intensity}} \times \frac{10.3}{1}$$

Therefore, correction factors of 28.4 and 10.3 were applied to the naïve and part-decell samples respectively, in order to make a fair quantitative comparison.

Total collagen measurement—Total collagen content of the rat liver scaffolds was measured using QuickZyme total collagen assay kit (Cedarlane Labs, Burlington, NC). The lyophilized samples were hydrolyzed in 6 M HCl for 20 h at 95 °C, and diluted 1:10 using 4 M HCl and assayed following manufacturer's instructions. The collagen amount per mg dry naïve tissue were 39±19, 43±9, and 32±1 µg, for the naïve, partially-decellularized, and decellularized tissues, respectively, indicating good preservation of collagens.

Western Blot Analysis— Protein from rat liver tissue was extracted with 150 µL SDT solution containing an aliquot of 1.5 µL Calbiochem Protease Inhibitor Cocktail Set I (Merck Millipore, Billerica, MA). The protein extract was incubated at 95 °C and sonicated as described above. A NanoDrop 1000 spectrophotometer (Thermo Scientific, Wilmington, DE) was used to estimate protein concentration. Reducing SDS-PAGE was performed using a pre-cast 4-20% acrylamide gel (Bio-Rad, Hercules, CA) with 142 µg, 52 µg, and 5 µg total protein load for the naïve, part-decell, and decell samples, respectively; these relative amounts were based on the correction factors reported above. Following transfer, polyvinylidene fluoride membranes were blotted using the following primary antibodies: rabbit anti-fibronectin (1:1500); mouse anti-GAPDH (1:20,000) (both from Sigma Aldrich); goat anti-serpina3k (1:200; Santa Cruz Biotechnology, Dallas, TX); and rabbit anti-filamin A (1:1000; Cell Signaling Technology, Beverly, MA). Blots were detected using the Clarity western ECL substrate kit (Bio-Rad) with relevant horseradish peroxidase (HRP)-conjugated anti-mouse (1:20,000), anti-rabbit (1:20,000)

(both from Bio-Rad), and anti-goat (1:5,000) (Santa Cruz Biotechnology) IgG secondary antibodies, according to the manufacturer's instructions. An ImageQuant LAS 4000 mini system (GE Healthcare, Pittsburgh, PA) was used to produce digital images of the chemiluminescent membranes.

Statistical Analysis—Gene Ontology (GO) enrichment analyses were performed using BiNGO 2.44 (hypergeometric model and Benjamini-Hochberg (BH) correction)³⁷ and REVIGO (SimRel cutoff = 0.4)³⁸. Two-sample t-tests with BH correction were performed to quantitatively compare the LFQ values of the naïve and decellularized rat liver samples.

2.3.3 Results

We performed proteomic analysis on four types of biological scaffolds. The first two types, collagen I gel and GFR-Matrigel, are relatively less complex than decellularized organs and thus were employed to test the proteomics workflow. This workflow, in brief, was as follows: samples were extracted with an SDS/DTT/Tris buffer at 95 °C, sonicated, subjected to FASP to remove detergent and digest the proteins into peptides, and finally analyzed by LC-MS/MS. After success with these simpler scaffolds, we quantitatively compared the proteins found in decellularized, partially-decellularized, and naïve rat livers. The last type of biological scaffold studied, decellularized human lungs, demonstrated the applicability of the method to different tissue types.

Rat-Tail Type I Collagen—Type I collagen (collagen I) is one of the primary constituents of the ECM. An important analytical consideration is that collagen I is comprised of a large fraction of proline residues (~18%), which are frequently posttranslationally modified by prolyl 4-hydroxylase to generate hydroxyproline residues that increase the stability of the collagen

triple helix. It is thus crucial to include proline hydroxylation (oxidation +15.9949 Da) as a variable modification during proteomic database searching. In addition, hydroxylation of lysine and glycosylation (mono- and disaccharide) of hydroxylysine were included as variable modifications, as they are known to occur on collagen I.

Proteomic analysis of rat-tail type I collagen yielded numerous unique peptides and high sequence coverage (~80%) for collagen I alpha-1 and alpha-2 chains, as shown in Table 2.2. The ~20% of undetected protein sequences for these collagen chains were mostly in the signal and propeptide regions. These results are superior to those reported in previous collagen proteomic studies^{39, 40}, indicating that the FASP method is suitable for proteomics of collagenous samples.

Numerous other proteins were identified in this rat-tail type I collagen sample, and their relative abundances were measured by intensity-based absolute quantification (iBAQ) using the MaxQuant software package⁴¹. iBAQ values are normalized by the number of theoretical peptides for each protein, which allows for comparison of different proteins within a single sample⁴². The most abundant proteins apart from collagen I include serum albumin, fibromodulin, and decorin (shown in Table 2.2). Note that the iBAQ% values for the most abundant proteins (e.g. collagen I) are an underestimate, as our mass spectrometry settings facilitate the observation of lower-abundance proteins. Although tendons have limited blood supply, some plasma proteins were relatively abundant among the 101 co-extracted proteins, including serum albumin, hemoglobin, and serotransferrin, which is reportedly able to stimulate cell proliferation⁴³. Several small leucine-rich proteoglycans were present, including fibromodulin, decorin, prolargin, and biglycan, which associates with fibrillar collagens to help regulate collagen fibrillogenesis⁴⁴. Procollagen C-endopeptidase enhancer, known to be highly expressed in rat-tail tendon⁴⁵, was also co-extracted. Additional identified proteins included vimentin, which associates with and

stabilizes collagen I mRNAs, thus contributing to high levels of collagen expression⁴⁶; fibronectin and thrombospondin-4, two glycoproteins that contribute to cell proliferation, migration and adhesion⁴⁷; and complement component C9, a component of the membrane attack complex that plays a key role in the innate and adaptive immune response by forming pores in the plasma membrane of target cells⁴⁸. Various muscle proteins and proteases were also identified. Besides collagen I, other fibril-forming collagens were present, including collagens type II, III, V, and XI, as well as collagen chaperone protein SERPINH1 at low abundance.

Growth factor reduced (GFR) Matrigel—Given the increased proteomic complexity of the GFR-Matrigel sample, as compared to the type I collagen sample, we chose to examine the utility of enhanced separation conditions. We optionally employed high-pH reverse-phase HPLC for offline fractionation of the peptides and also compared 1.7 versus 3.0 μm particle sizes for the low-pH online reverse-phase HPLC-MS analysis. The proteomic results from this comparison are given in Fig. 2.3A. Fractionation led to a greater than 3-fold increase in the number of proteins and unique peptides. This deep proteomic coverage may provide understanding of less abundant but important components of biological scaffolds, but it comes at a cost of 6X increased instrument time. The smaller 1.7 μm packing particles also increased the number of protein identifications (by about 27%). With both the fractionation step and the smaller particle size column, a total of 953 proteins were identified among the three replicates.

The importance of matrix structure and function to biological scaffolds led us to employ a recently developed categorization approach offered by Naba et al. called the “matrisome”, which encompasses ECM and ECM-associated proteins⁴⁹. The current mouse matrisome inventory consists of 1122 protein entries (note that Matrigel is derived from mouse). The percentage of iBAQ intensity attributed to matrisome proteins is shown in Fig. 2.3A for each separation

condition. The decrease in the percentage of matrisome protein intensity after fractionation may reflect improved detection of low-abundance cellular (i.e. non-matrisome) proteins.

Among the 953 proteins identified in GFR-Matrigel, the most abundant ones (based on iBAQ values) were laminin, nidogen, vimentin, basement membrane-specific heparan sulfate proteoglycan core protein, and fibrinogen. Numerous cellular proteins were observed, which is in agreement with previous report⁵⁰. A total of 83 proteins were assigned to the matrisome category and their qualitative and quantitative distribution among subcategories is depicted in Fig. 2.3B. The GFR-Matrigel matrisome was composed primarily of abundant ECM glycoproteins, which possess domains and motifs that promote cell adhesion and form interactions involving ECM assembly⁵¹. There also existed a large number of low-abundance ECM regulators, which are important for matrix modification and remodeling.

Naïve, Partially-decellularized and Decellularized Rat Livers—With encouraging proteomic results from collagen and Matrigel, we next applied the proteomic approach to analysis of decellularized rat livers. Three biological replicates for each of three conditions (naïve, part-decell, and decell) were analyzed. For this study, peptides were not pre-fractionated off-line, but they were on-line separated using the more efficient chromatographic condition (1.7 μ m particle size at 60 °C) prior to MS/MS analysis.

Proteomic results obtained for the three sample conditions are shown in Fig. 2.4. As expected, fewer distinct proteins were detected as decellularization progressed, dropping from 1019 to 854 and then 517 protein identifications in the naïve, part-decell and decell conditions, respectively. For a closer look at ECM composition, we again employed the matrisome/non-matrisome categorization, but in this case we first needed to generate a list of rat matrisome proteins, since curated matrisome inventories currently only exist for mouse and human. It is not

immediately possible to create a completely accurate rat matrisome database by the methodology previously used for mouse and human because the rat genome sequence is less complete than that of the other two species. Therefore, as a reasonable approximation, we generated a matrisome protein list for rat by combining those previously curated for human (1062 entries) and mouse (1122 entries)⁴⁹. The resulting 1284 gene symbols (900 were shared) were used to categorize matrisome proteins in our rat liver proteomic results. The number of identified matrisome proteins was comparable for the naïve livers and the part-decell livers (32 and 33 proteins, respectively), but increased substantially to 58 proteins for the decell livers. Nonetheless, the “fully” decellularized rat liver proteome was still dominated by non-matrisome or cellular proteins, as shown in Fig 2.4.

DNA quantification, a classic assay for evaluating decellularization effectiveness, was also performed for each sample condition, and the results are given in Fig 2.4. The DNA content dropped dramatically in the first stage of decellularization, from 962 to 74 ng per mg of tissue (dry weight), and then decreased further to 20 ng/mg for the decell samples. This value for the decell liver satisfies one of the traditional decellularization criteria—less than 50 ng DNA per mg of dry tissue weight¹. These DNA results contrast drastically with the proteomic results where the non-matrisome iBAQ% remained at 99% for the part-decell and 75% for the decell samples, indicating that the majority of the protein content remaining after decellularization is not related to the ECM.

A total of 58 matrisome proteins were identified in at least two biological replicates of the decellularized livers: the most abundant was collagen VI, followed by transglutaminase 2, biglycan, fibrinogen, and fibronectin. Collagen VI is a network-forming collagen that anchors interstitial structures such as blood vessels and collagen fibers into surrounding connective

tissues⁵². It has been demonstrated to promote mesenchymal cell proliferation and is also a potent inhibitor of apoptosis⁵³, making it an important component of biological scaffolds used for tissue reconstruction. The other 14 collagen forms identified included: fibrillar collagens I, II, III, and V, which contribute tensile strength; network-forming collagen IV and multiplexin collagen XVIII, which anchor and stabilize other basement membrane components; and fibril-associated collagens XII, XIV, and XVI, which regulate fibrillar collagen growth. Several ECM glycoproteins were observed, including fibrinogen, fibronectin, dermatopontin, laminin, vitronectin, and elastin. A number of proteoglycans, ECM-affiliated proteins, ECM regulators, and secreted factors were also identified. These matrisome proteins together drive the enrichment of Gene Ontology (GO) terms that are favorable for subsequent cell seeding and tissue regeneration, such as the terms: protein complex assembly, liver development, extracellular matrix organization, and cell-substrate adhesion.

Conversely, the 459 identified cellular proteins included ones that are unfavorable, such as the previously reported xenoantigens serum albumin⁵⁴, hemoglobin alpha-1 chain⁵⁵, and peroxiredoxin 1^{56, 57}, as well as tyrosine-protein kinase Lyn, a protein with an inflammatory response function⁵⁸. The detection of a large number of remnant cellular proteins reveals that the decellularized rat livers are not as free of cellular material as indicated by the DNA assay. The presence of these proteins, along with the possibility of other intracellular molecules, may provide clues to explain why some decellularized scaffolds contribute to tissue/organ rejection upon in vivo transplantation.

To quantitatively characterize differential protein removal and preservation during the decellularization process, we performed label-free quantification based on the LFQ intensity generated by the MaxQuant software. The heatmap in Fig. 2.5A shows the abundance change of

each protein in each biological sample. The majority of proteins decreased in abundance as decellularization progressed, but a number did not (rows at the bottom of the heatmap). Fig. 2.5B includes a volcano plot showing fold change in protein abundance (decell versus naïve) plotted against statistical p-value. The purple rectangle denotes cutoff criteria for protein underrepresentation (fold change $< 1/8$; p-value < 0.01) in the decellularized sample, i.e., proteins that were significantly removed after decellularization. The orange rectangle denotes proteins that were retained after decellularization (fold change $> 1/8$ compared to naïve). A number of ECM regulators and ECM-affiliated proteins were significantly depleted, such as cathepsins B, D, and Z, serpins A1, A3K, and H1, and annexins 2, 4, 6, and 7. However, the “core-matrisome,” which is critical to overall scaffold structure and facilitates cell attachment and growth, was well preserved as indicated by retention of: collagens I, IV, VI, XII, and XIV; ECM glycoproteins including fibrinogen, dermatopontin, and fibronectin; as well as proteoglycans including biglycan and decorin. Seventeen non-matrisome, or cellular, proteins were retained in the decellularized samples, including: filamin A, a large (280 kDa) protein that binds to numerous proteins such as transmembrane receptors and signaling molecules⁵⁹; LASP1, an actin-binding protein that may play a role in cytoskeletal organization⁶⁰; vimentin, an intermediate filament protein that anchors the organelles in the cytosol and regulates focal contact⁶¹; as well as a few other membrane and ATP-binding-related proteins.

We validated these proteomic results using western blots for four different proteins, one from each major category (“matrisome” versus “non-matrisome” and “retained” versus “removed”). The western blots are shown in Fig. 2.6, and in each case they confirm our quantitative proteomic data: (i) fibronectin (FN) – a matrisome protein – was retained; serine protease inhibitor A3K (SERPINA3K) – a matrisome protein – was removed; filamin alpha

(FLNA) – a non-matrisome protein – was retained; and glyceraldehyde 3-phosphate dehydrogenase (GAPDH) – a non-matrisome protein – was removed.

Naïve and Decellularized Human Lungs—Another type of decellularized organ scaffold, human lung, was analyzed and compared to its naïve counterpart. Peptides from these samples were off-line separated into 6 fractions using high-pH HPLC, and then each fraction was analyzed by LC-MS/MS (3 μ m particle-size, room temperature). The total and matrisome protein numbers as well as the iBAQ% arising from the non-matrisome proteins are shown in Fig. 2.7. The number of protein identifications dramatically decreased from 2147 to 384 after decellularization. Matrisome proteins were well retained during decellularization, comprising 152 of the 384 proteins and 59% of the total iBAQ intensity. These results demonstrated the universal applicability of our proteomic workflow to different tissue types, as well as its capacity for achieving in-depth proteome coverage for both cellular and matrisome proteins.

2.3.4 Discussion

Besides identifying the myriad of ECM components that play important roles in facilitating the engraftment, survival, and long-term function of reseeded cells, a comprehensive profile of the remnant cellular components in biological scaffolds is also imperative in revealing specific molecules involved in the host immune response after transplantation. The few published reports of proteome composition of decellularized biological scaffolds focus primarily on ECM proteins⁶²⁻⁶⁴. In this work, we have shown that protein extraction with SDS/DTT/Tris buffer followed by FASP cleanup and digestion enables identification and quantification of numerous low-abundance cellular proteins that represent the residual component of all the cell-related complexity, along with a variety of ECM proteins that are crucial in tissue reconstruction.

The quantitative proteomic results are reassuring in that most of the remnant cellular proteins have significantly decreased in abundance, whereas the structural matrix proteins are well preserved after decellularization. A further benefit of the FASP method is that it efficiently removes the sodium dodecyl sulfate (SDS) routinely employed for tissue/organ decellularization, as SDS is quite detrimental to MS-based proteomics. This concomitant benefit of FASP is especially important in evaluating tissues that are sampled during the perfusion process (such as the partially-decellularized rat livers presented in this work) or ones that have not gone through extensive washing after SDS perfusion.

Collagen I and Matrigel are used extensively in biological research and medical applications. We identified 103 different proteins in “purified” collagen I, and many of the observed minor components may play a role in supporting cell growth or differentiation when collagen I is employed as a scaffold. GFR-Matrigel is a variant of Matrigel that has reduced levels of bioreactive growth factors such as fibroblast growth factor, epidermal growth factor, insulin-like growth factor 1, and transforming growth factor-beta⁵⁰. The absence of these growth factors in our proteomic results confirms the quality of this “growth factor reduced” matrix, but hepatoma-derived growth factor and insulin-like growth factor 2 mRNA-binding protein 1 were still identified, although at very low abundance. The identification of nearly 1000 additional cellular and matrisome proteins, beyond the classically reported basement membrane-associated proteins laminin, nidogen, fibronectin and collagen IV, suggests caution in interpreting the functional consequences of cell culture or tissue regeneration experiments using Matrigel.

Collagen gel, Matrigel and decellularized liver have all been used as biological scaffolds for hepatocyte culture. Decellularized liver matrix is superior to the other two matrices in terms of supporting hepatocyte attachment and long-term survival^{65, 66}. To potentially explain these

differences, we compared and contrasted their matrisome proteome composition; the Venn diagram is shown in Fig. 2.8. A total of 34 matrisome proteins in the decellularized liver were also identified in either collagen or Matrigel, or both. This group represents a key set of proteins that may be essential for fundamental cell attachment, growth, and differentiation of hepatocytes. There were 24 matrisome proteins exclusively identified in the decellularized liver but absent in rat-tail collagen and Matrigel; these represent the uniqueness of the decellularized rat liver compared to collagen gel and Matrigel. Of note, COL4A1 and COL4A2 were shared by the three sample types, whereas COL4A3, A5, and A6 were unique to the decellularized rat livers. This observation is consistent with the fact that the “major” A1 and A2 chains of collagen IV are present ubiquitously in basement membranes, whereas the “minor” A3 through A6 chains are more spatially and temporally restricted⁵². The 24 unique components in decellularized livers could be critical determinants for directing constructive tissue remodeling, especially when combined with the original organ architecture and preserved microvascular network, thereby making them superior scaffolds for hepatocyte culture and liver regeneration.

The decellularized ECM scaffold composition varies depending upon the tissue/organ source and the decellularization protocol. This observation is illustrated in Fig. 2.9, where we show the iBAQ% attributed to each matrisome subcategory for the decellularized rat liver and human lung samples. Compared with decellularized human lung, decellularized rat liver has higher collagen and proteoglycan content, but proportionally lower abundance of ECM glycoproteins and secreted factors. These distinct matrisome compositions, together with differences in ECM architecture, help explain the fact that decellularized liver tissue may be the preferred ECM substrate for hepatocytes⁶⁶, while ECM from the lung may be preferred for respiratory epithelial cells⁶⁷.

Proteomic analysis revealed that the rat liver prior to decellularization has a higher fraction of cellular protein than the human lung – iBAQ values of 99% for the naïve liver and 86% for the naïve lung (Figures 2.4 & 2.7). This difference in baseline cellularity may partially explain why the decellularized lung sample appears more decellularized than the liver, when evaluated at the proteomic level – iBAQ values of 75% non-matrisome for the liver and only 41% for the lung. As mentioned above, these iBAQ values probably overestimate the cellular protein abundance, as our proteomics method is somewhat biased toward them, especially in tissues that possess compact ECM structures that are difficult to solubilize completely with sonication in SDS/DTT/Tris solution. Nonetheless, the relative comparison of iBAQ values (e.g. higher cellular percentage for liver than lung) should be valid. Notably, Hill et al. recently reported a method that involves protein fractionation and stable isotope labeled peptides to allow unbiased protein characterization and quantification⁶⁴. Our present work employs a simple sample preparation method to reveal numerous ECM proteins that can be used to evaluate ECM preservation and loss, along with numerous low-abundance cellular proteins, which could have significant clinical consequences.

During creation of a decellularized scaffold, it is crucial to preserve the organ-specific ECM “zip codes” that direct or support site-appropriate cell attachment and differentiation. A total of 22 matrisome proteins were significantly removed during either the first stage (naïve to part-decell), the second stage (part-decell to decell), or both stages of rat liver decellularization. Only three of these are “core matrisome” proteins: fibulin-5, laminin subunit beta-1, and vitronectin. The Space of Disse, lying between sinusoidal endothelial cells (SEC) and hepatocytes, has a unique matrix that is essential for maintaining the differentiation of these two types of cells⁶⁸. The Space of Disse primarily consists of abundant fibronectin and collagen VI;

common basement membrane proteins such as collagen IV and laminin; as well as fibrillar collagens I, III, and V⁶⁹. These proteins were preserved during the decellularization process in the overall liver matrix, and thus it is reasonable to conclude that they were also preserved in the Space of Disse. Although it is unknown to what extent the loss of three core matrisome proteins and 19 other ECM regulators and ECM-affiliated proteins will affect subsequent cell seeding and proliferation, it is encouraging that the majority of matrisome proteins are preserved in this complex matrix, especially those associated with the biologically important Space of Disse.

In this work, we showed that cellular proteins are more difficult to remove than DNA and still constitute a large fraction of the decellularized scaffold proteome. The rat liver data presented in Fig. 2.4 and 2.10 (two volcano plots: naïve vs. part-decell; part-decell vs. decell) reveal that more cellular proteins were removed during the latter stage of the decellularization process than during the earlier stage. This is likely due to the fact that the SDS concentration was higher (0.1-0.2%) during the latter stage (part-decell to decell), as compared to only 0.01% initially (naïve to part-decell). Although the decellularized rat livers are still dominated by remnant cellular proteins in terms of protein numbers (confirming that elimination of DNA does not assure adequate removal of all cellular material), quantitative comparison shows that most of these cellular proteins have significantly decreased abundance after decellularization. The majority of the 17 well-retained cellular proteins were large plasma membrane or intermediate filament proteins. These results suggest that although vascular perfusion decellularization is a powerful technique for preserving key structural components, disrupting cell membranes, and removing nucleic acids, it remains a challenge to remove all of the large anchoring and adhesion complexes⁷⁰. The association between the relative abundance of remnant cellular proteins and subsequent in vivo host response requires further systematic investigation that combines

proteomic analysis with transplantation studies. It is also possible that reseeded cells will facilitate degradation of these cellular remnants, and research is underway to study the turnover of these proteins.

Biological scaffold-based tissue engineering is an important approach within the sphere of regenerative medicine. A crucial step is to acquire scaffolds that provide microenvironments hospitable to reseeded cells and conducive to the growth of three-dimensional structured tissue. Our study provides the first insight into the complexity of the proteome of biological scaffolds and sheds new light on the quantitative protein changes that occur as decellularization progresses. We believe proteomics will be an extremely useful tool in facilitating understanding of the connection between the scaffold composition and future clinical transplantation outcomes.

2.4 Decellularization of Porcine Vocal Fold Mucosae

2.4.1 Methods and Results

Porcine vocal fold mucosae (VFM) samples were harvested and subjected to five different decellularization methods with simultaneous mechanical agitation: 1% SDS for 24 h (“SDS”); 1% CHAPS for 24 h (“CHAPS”); osmotic stress/enzymatic approach (“Long”); osmotic/enzymatic approach plus 1% SDS for 24 h (“LongSDS”); osmotic/enzymatic approach plus 1% CHAPS for 24 h (“LongCHAPS”). We utilized the optimized proteomics protocol to analyze the five samples from the five decellularization methods as well as a native porcine VFM as the control.

Our goal was to determine which decellularization method removes cellular proteins while retaining ECM proteins most effectively. However, we observed little difference between

the various decellularization conditions and between decellularized and control sample (intact tissue) (Fig. 2.11).

2.4.2 Discussion

Several possible reasons including saturation effect, or the control sample being not really intact, etc., were explored. However, none of them was able to explain the little differences in the proteomic results among the different sample conditions. Based on the fact that marked differences between decellularized and native rat livers and human lungs were observed, we hypothesized that the porcine VFM decellularization with mechanical agitation was not effective enough to decellularize VFM tissues that are compact and sticky, and other decellularization approaches such as perfusion-based decellularization may be a better choice. The ongoing perfusion decellularization study is presented in Chapter 6 of this thesis.

2.5 Figures and Tables

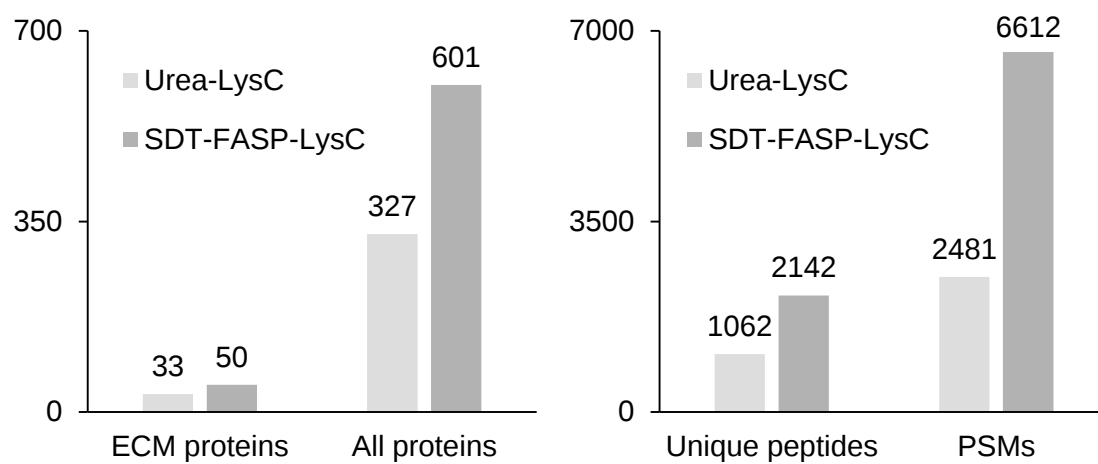


Figure 2.1. Comparison of the proteomic identifications obtained from protein extraction with 8M urea or the SDT solution followed by LysC digestion.

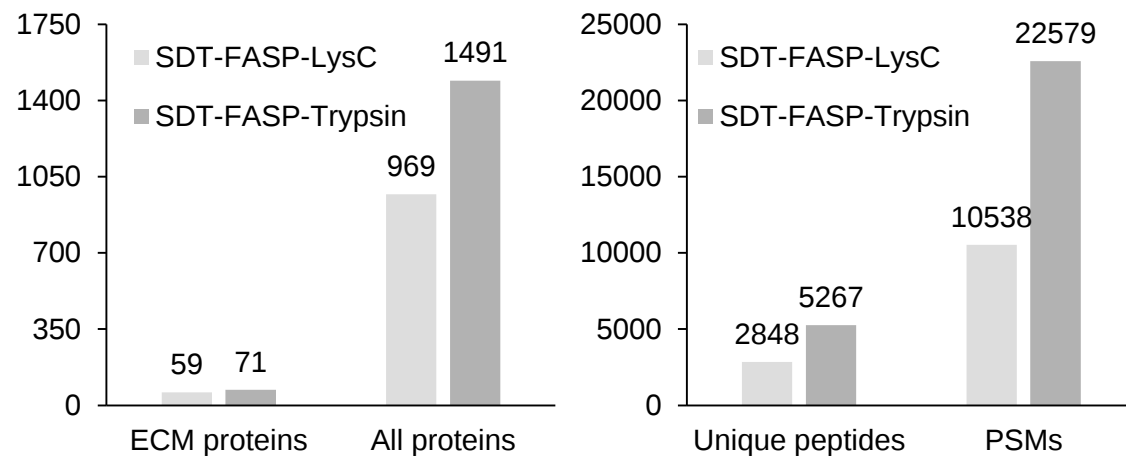


Figure 2.2. Comparison of the proteomic identifications obtained from LysC or trypsin digestion.

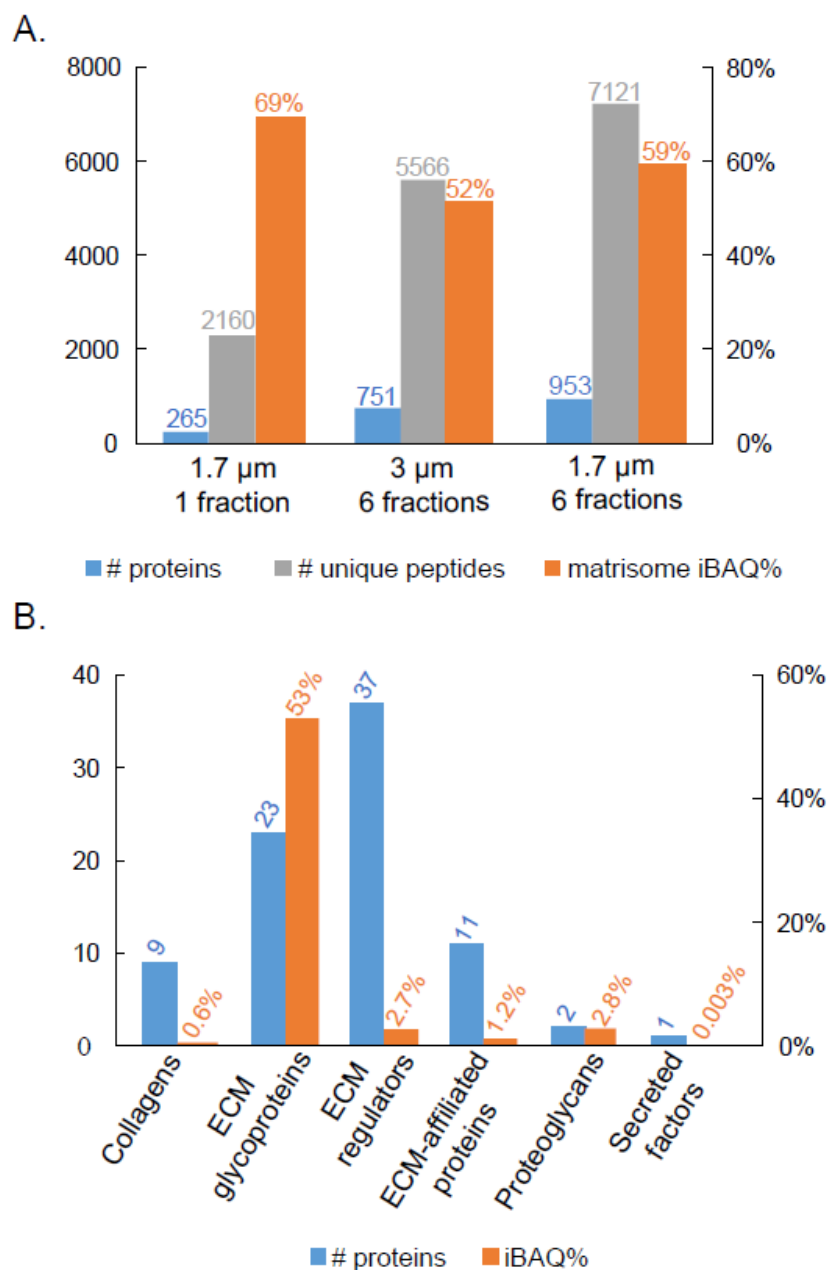


Figure 2.3. Proteomic results for GFR-Matrigel samples. A) Numbers of proteins and unique peptides identified, along with the percentage of iBAQ intensity (iBAQ%) arising from matrisome proteins. B) Under the condition of 1.7 μ m column and 6 peptide fractions, the number of proteins and the percentage of iBAQ intensity for each matrisome subcategory. N = 3 bio reps.

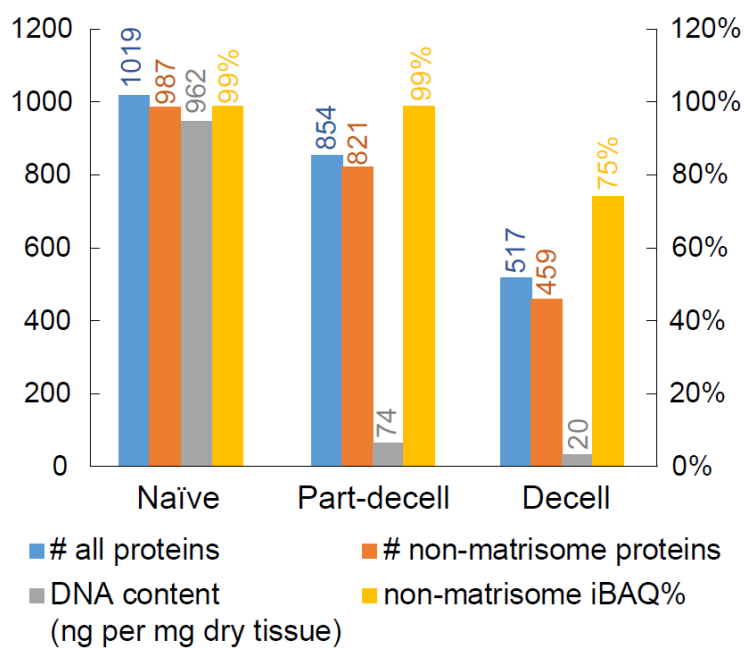


Figure 2.4. Proteomic results and DNA content for the naïve, partially-decellularized (part-decell), and decellularized (decell) rat livers. N = 3 bio reps.

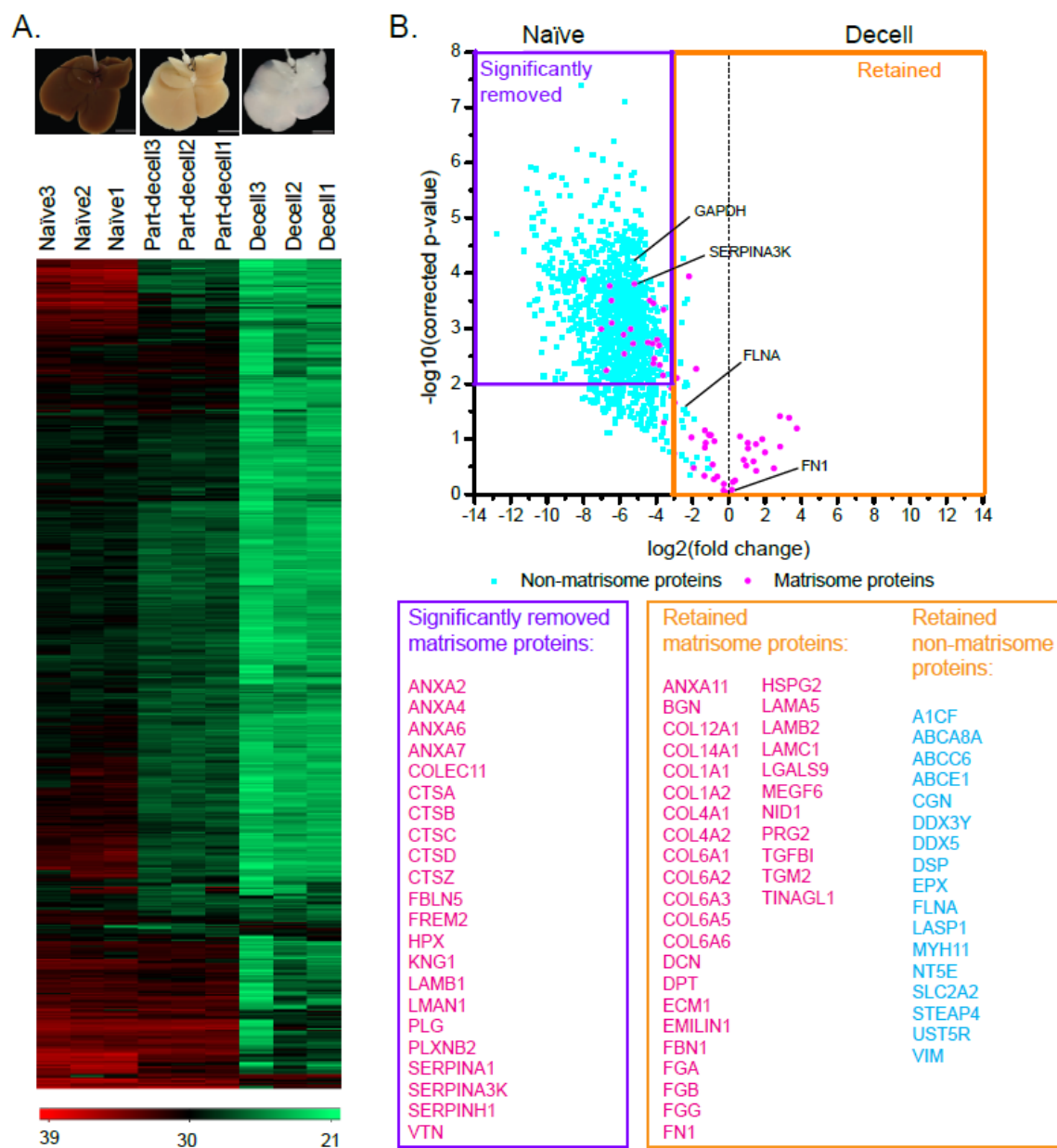


Figure 2.5. Quantitative proteomic results for rat livers. A) Images of naïve, part-decell, and decell livers (scale bar: 10 mm), along with a heatmap showing changes in protein abundances for these three conditions. Numbers in the color code are log-transformed-corrected LFQ intensity. B) Volcano plots showing quantitative comparison of the naïve versus the decell rat livers. The purple rectangle denotes cutoff criteria for significant removal of a protein in the decell sample compared to the naïve sample (fold change $< 1/8$; Benjamini Hochberg-adjusted p-value < 0.01). The orange rectangle encompasses the proteins that are considered retained in the decell sample (fold change $> 1/8$). Four proteins that were chosen for further validation are highlighted in the volcano plot. N = 3 bio reps.

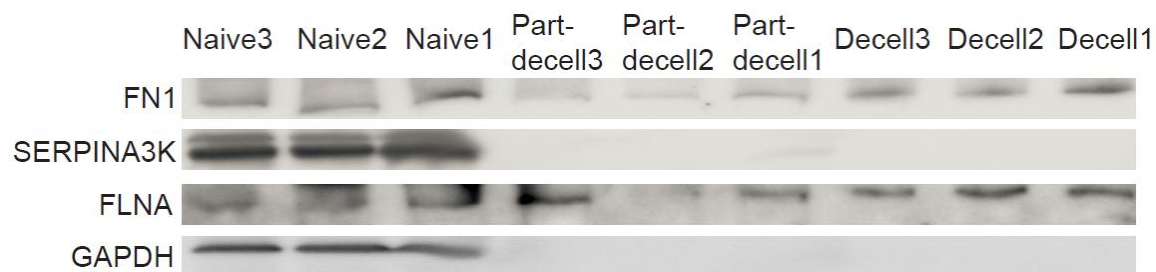


Figure 2.6. Western blots of four proteins – FN1, SERPINA3K, FLNA, GAPDH – in the naïve, partially-decellularized (part-decell), and decellularized (decell) rat livers. N = 3 bio reps.

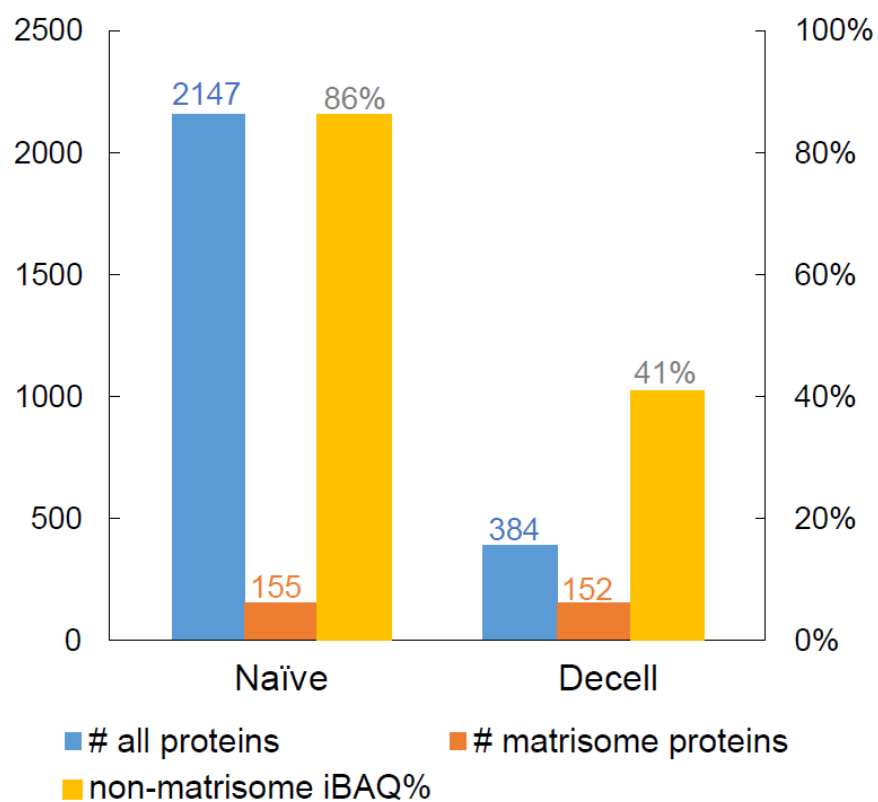


Figure 2.7. Proteomic results for naïve and decellularized (decell) human lung. Both samples were subjected to peptide pre-fractionation (6 fractions.)

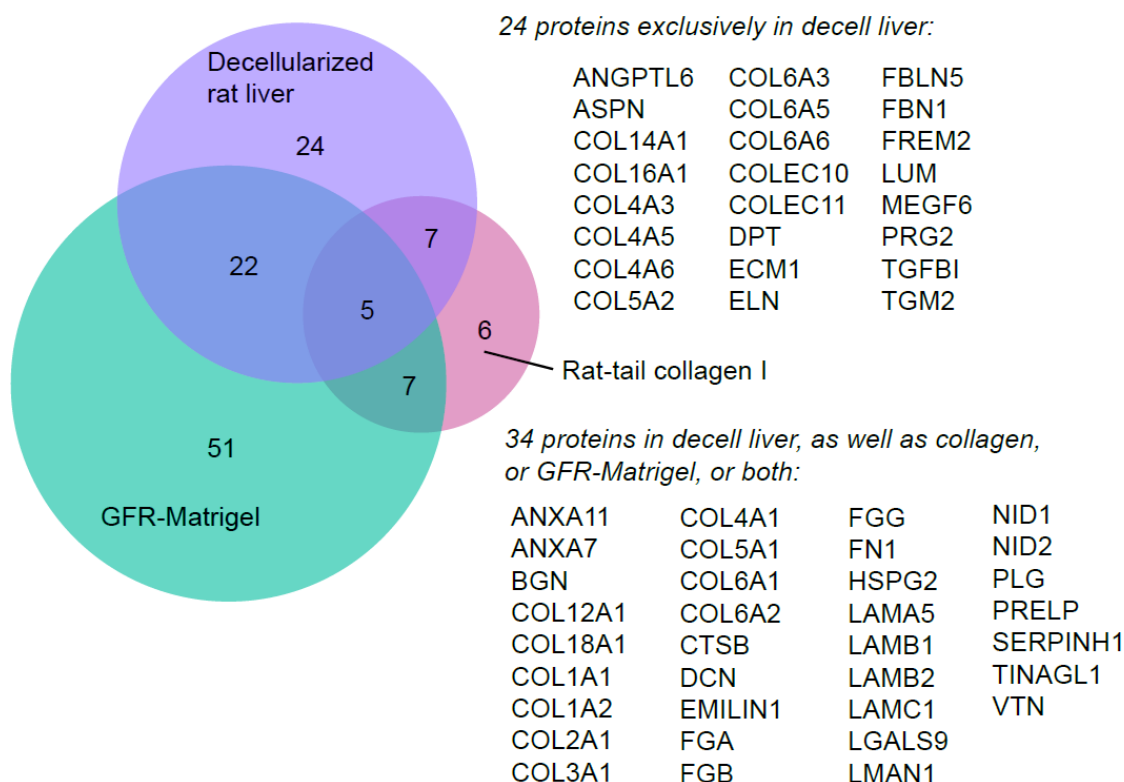


Figure 2.8. Venn diagram with the number of matrisome protein identifications for rat-tail type I collagen, GFR-Matrigel, and decellularized rat liver. Note that the GFR-Matrigel samples were subjected to peptide fractionation (6 fractions), while the collagen gel and rat liver samples were not, which could account for the relatively large number of proteins only detected in the GFR-Matrigel sample.

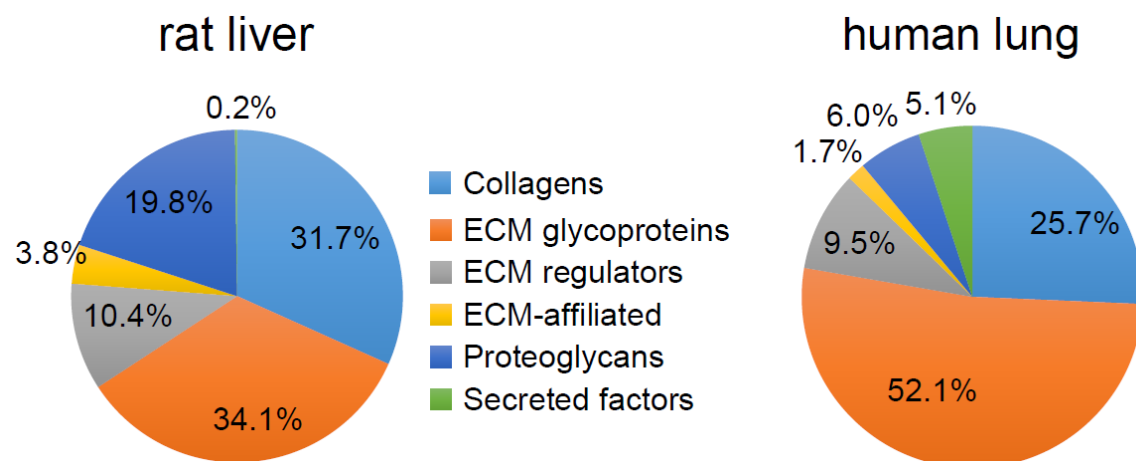


Figure 2.9. Matrisome compositions by sub-category for decellularized rat liver and decellularized human lung.

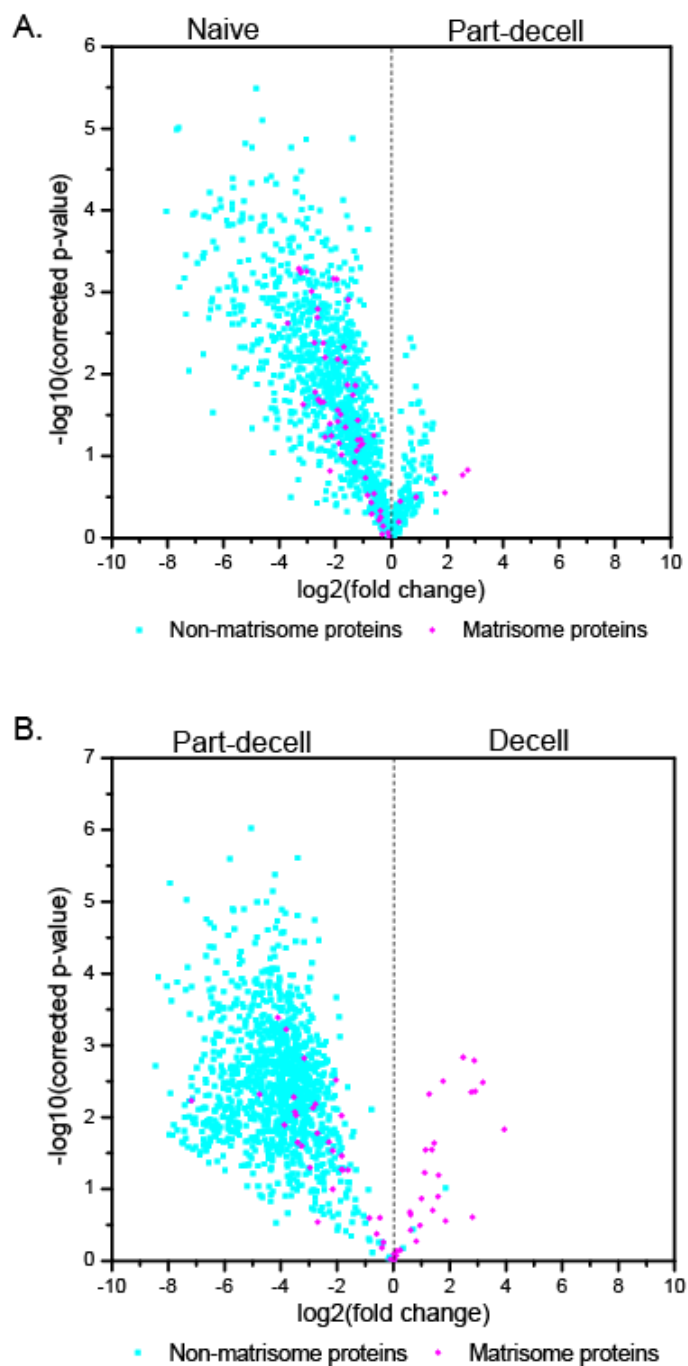


Figure 2.10. Volcano plots showing quantitative comparison of A) the naïve versus the partially-decellularized (part-decell) and B) partially-decellularized (part-decell) versus the decellularized (decell) rat livers.

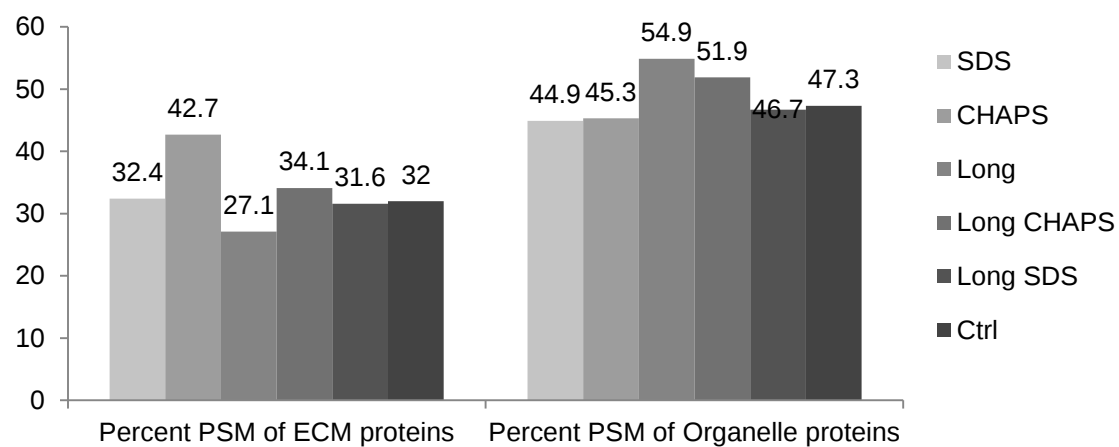


Figure 2.11. Comparison of the percent PSMs of ECM and organelle proteins resulted from the five decellularization strategies.

Table 2.1. Selected agents and techniques for tissue decellularization.

Agent/Technique	Mode of action
<i>Chemical Agents</i>	
Acids and bases	Solubilizes cytoplasmic components of cells, disrupts nucleic acids, tend to denature proteins
Hypotonic and hypertonic solutions	Cell lysis by osmotic shock, disrupt DNA-protein interactions
Non-ionic detergents	Disrupt DNA-protein interactions, disrupt lipid-lipid and lipid-protein interactions and to a lesser degree protein-protein interactions
Ionic detergents	Solubilize cell and nucleic membranes, tend to denature proteins
Zwitterionic detergents	Exhibit properties of non-ionic and ionic detergents
Solvents	Cell lysis by dehydration, solubilize and remove lipids
<i>Biologic Agents</i>	
Enzymes	
- Nucleases	Catalyze the hydrolysis of ribonucleotide and deoxyribonucleotide chains
- Trypsin	Cleaves peptide bonds on the C-side of Arg and Lys
- Dispace	Cleaves specific peptides, mainly fibronectin and collagen IV
Chelating Agents (EDTA, EGTA)	Chelating agents bind metallic ions, thereby disrupting cell adhesion to ECM
<i>Physical and Miscellaneous Agents</i>	
Temperature (freezing and thawing)	Intracellular ice crystals disrupt cell membrane
Direct application of force	Removal of tissue eliminates cells and force can burst remaining cells
Pressure	Pressure can burst cells and aid in removal of cellular material
Electroporation	Pulsed electrical fields disrupt cell membranes
<i>Techniques to Apply Agents</i>	
Perfusion	Facilitates chemical exposure and removal of cellular material
Pressure gradient across tissue	Facilitates chemical exposure and removal of cellular material
Supercritical fluid	Pressure can burst cells, supercritical fluid facilitates chemical exposure and removal of cellular material
Agitation	Can lyse cells, but more commonly used to facilitate chemical exposure and removal of cellular material

Table 2.2. Proteomic results for rat tail type I collagen. (n = 3 bio reps)

Protein description	Unique peptides	Sequence coverage (%)	iBAQ (%)
Collagen I alpha-2 chain	108	81.3	26.4
Collagen I alpha-1 chain	101	78.8	16.8
Serum albumin	33	53.5	15.8
Fibromodulin	8	25.8	7.6
Decorin	10	29.1	4.1
98 other protein groups	/	/	29.3

2.6 References

1. Crapo, P.M., Gilbert, T.W. & Badylak, S.F. An overview of tissue and whole organ decellularization processes. *Biomaterials* **32**, 3233-3243 (2011).
2. Griffiths, L.G., Choe, L.H., Reardon, K.F., Dow, S.W. & Orton, E.C. Immunoproteomic identification of bovine pericardium xenoantigens. *Biomaterials* **29**, 3514-3520 (2008).
3. Beretta, B. et al. Antigenic determinants of bovine serum albumin. *Int Arch Allergy Immunol* **126**, 188-195 (2001).
4. Benjamin, D.C. et al. The Antigenic Structure of Proteins - a Reappraisal. *Annual Review of Immunology* **2**, 67-101 (1984).
5. Wisniewski, J.R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. *Nature methods* **6**, 359-362 (2009).
6. Hansen, K.C. et al. An in-solution ultrasonication-assisted digestion method for improved extracellular matrix proteome coverage. *Mol Cell Proteomics* **8**, 1648-1657 (2009).
7. Angel, P.M. et al. Networked-based Characterization of Extracellular Matrix Proteins from Adult Mouse Pulmonary and Aortic Valves. *J Proteome Res* **10**, 812-823 (2011).
8. Wilson, R. et al. Changes in the Chondrocyte and Extracellular Matrix Proteome during Post-natal Mouse Cartilage Development. *Mol Cell Proteomics* **11** (2012).
9. Wilson, R. The extracellular matrix: an underexplored but important proteome. *Expert Review of Proteomics* **7**, 803-806 (2010).
10. Yang, J., Elias, J.J., Petrakis, N.L., Wellings, S.R. & Nandi, S. Effects of Hormones and Growth-Factors on Human Mammary Epithelial-Cells in Collagen Gel Culture. *Cancer Research* **41**, 1021-1027 (1981).
11. Rocha, V., Ringo, D.L. & Read, D.B. Casein Production during Differentiation of Mammary Epithelial-Cells in Collagen Gel Culture. *Experimental Cell Research* **159**, 201-210 (1985).
12. Sirica, A.E., Richards, W., Tsukada, Y., Sattler, C.A. & Pitot, H.C. Fetal Phenotypic Expression by Adult Rat Hepatocytes on Collagen Gel-Nylon Meshes. *Proc Natl Acad Sci U S A* **76**, 283-287 (1979).
13. Balasubramanian, P., Prabhakaran, M.P., Sireesha, M. & Ramakrishna, S. in *Polymer Composites - Polyolefin Fractionation - Polymeric Peptidomimetics - Collagens*, Vol. 251. (eds. A. Abe, H.H. Kausch, M. Moller & H. Pasch) 173-206 (2013).
14. Toda, S. & Sugihara, H. Reconstruction of Thyroid-Follicles from Isolated Porcine Follicle Cells in 3-Dimensional Collagen Gel Culture. *Endocrinology* **126**, 2027-2034 (1990).
15. Minami, Y., Sugihara, H. & Oono, S. Reconstruction of Cornea in 3-Dimensional Collagen Gel Matrix Culture. *Investigative Ophthalmology & Visual Science* **34**, 2316-2324 (1993).
16. Fujiyama, C., Masaki, Z. & Sugihara, H. Reconstruction of the Urinary-Bladder Mucosa in 3-Dimensional Collagen Gel Culture - Fibroblast-Extracellular Matrix Interactions on the Differentiation of Transitional Epithelial-Cells. *Journal of Urology* **153**, 2060-2067 (1995).

17. Ito, K.I. et al. Expression of Tissue-Type Plasminogen-Activator and Its Inhibitor Couples with Development of Capillary Network by Human Microvascular Endothelial-Cells on Matrigel. *Journal of Cellular Physiology* **162**, 213-224 (1995).
18. Braun, H., Buhnemann, C., Neumann, J. & Reymann, K.G. Preparation of a tissue-like cortical primary culture from embryonic rats using Matrigel and serum free Start V Medium. *Journal of Neuroscience Methods* **157**, 32-38 (2006).
19. Eldred, J.R., Rogers, K. & Boughner, D. Tissue Engineering the Aortic Valve Spongiosa Using Matrigel-Cell-Scaffold-Composites (MCSCs). *Faseb Journal* **22** (2008).
20. Lu, S.H. et al. Reconstruction of Engineered Uterine Tissues Containing Smooth Muscle Layer in Collagen/Matrigel Scaffold In Vitro. *Tissue Engineering Part A* **15**, 1611-1618 (2009).
21. Lu, S.H. et al. Self-assembly of renal cells into engineered renal tissues in collagen/Matrigel scaffold in vitro. *Journal of Tissue Engineering and Regenerative Medicine* **6**, 786-792 (2012).
22. Zhou, J. et al. The Spatiotemporal Development of Intercalated Disk in Three-Dimensional Engineered Heart Tissues Based on Collagen/Matrigel Matrix. *Plos One* **8** (2013).
23. Badylak, S.F., Taylor, D. & Uygun, K. Whole-organ tissue engineering: decellularization and recellularization of three-dimensional matrix scaffolds. *Annu Rev Biomed Eng* **13**, 27-53 (2011).
24. Ott, H.C. et al. Perfusion-decellularized matrix: using nature's platform to engineer a bioartificial heart. *Nature medicine* **14**, 213-221 (2008).
25. Uygun, B.E. et al. Organ reengineering through development of a transplantable recellularized liver graft using decellularized liver matrix. *Nature medicine* **16**, 814-820 (2010).
26. Petersen, T.H. et al. Tissue-engineered lungs for in vivo implantation. *Science* **329**, 538-541 (2010).
27. Price, A.P., England, K.A., Matson, A.M., Blazar, B.R. & Panoskaltsis-Mortari, A. Development of a Decellularized Lung Bioreactor System for Bioengineering the Lung: The Matrix Reloaded. *Tissue Engineering Part A* **16**, 2581-2591 (2010).
28. Ott, H.C. et al. Regeneration and orthotopic transplantation of a bioartificial lung. *Nature Medicine* **16**, 927-U131 (2010).
29. Song, J.J. et al. Regeneration and experimental orthotopic transplantation of a bioengineered kidney. *Nat Med* **19**, 646-651 (2013).
30. Uygun, B.E. et al. Decellularization and Recellularization of Whole Livers. *Journal of Visualized Experiments : JoVE*, 2394 (2011).
31. Gilpin, S.E. et al. Perfusion decellularization of human and porcine lungs: Bringing the matrix to clinical scale. *Journal of Heart and Lung Transplantation* **33**, 298-308 (2014).
32. Guyette, J.P. et al. Perfusion decellularization of whole organs. *Nature Protocols* **9**, 1451-1468 (2014).
33. Desiere, F. et al. The PeptideAtlas project. *Nucleic Acids Res.* **34**, D655-D658 (2006).
34. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature biotechnology* **26**, 1367-1372 (2008).

35. Cox, J. et al. A practical guide to the MaxQuant computational platform for SILAC-based quantitative proteomics. *Nat Protoc* **4**, 698-705 (2009).
36. Cox, J. et al. Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment. *J Proteome Res* **10**, 1794-1805 (2011).
37. Maere, S., Heymans, K. & Kuiper, M. BiNGO: a Cytoscape plugin to assess overrepresentation of Gene Ontology categories in Biological Networks. *Bioinformatics* **21**, 3448-3449 (2005).
38. Supek, F., Bošnjak, M., Škunca, N. & Šmuc, T. REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. *PLoS ONE* **6**, e21800 (2011).
39. Hansen, K.C. et al. An in-solution ultrasonication-assisted digestion method for improved extracellular matrix proteome coverage. *Mol Cell Proteomics* **8**, 1648-1657 (2009).
40. Pflieger, D. et al. Comparative proteomic analysis of extracellular matrix proteins secreted by two types of skin fibroblasts. *Proteomics* **6**, 5868-5879 (2006).
41. Schwanhauser, B. et al. Corrigendum: Global quantification of mammalian gene expression control. *Nature* **495**, 126-127 (2013).
42. Shin, J.B. et al. Molecular architecture of the chick vestibular hair bundle. *Nature neuroscience* **16**, 365-374 (2013).
43. Ekblom, P., Thesleff, I., Saxen, L., Miettinen, A. & Timpl, R. Transferrin as a fetal growth factor: acquisition of responsiveness related to embryonic induction. *Proceedings of the National Academy of Sciences of the United States of America* **80**, 2651-2655 (1983).
44. Matheson, S., Larjava, H. & Hakkinen, L. Distinctive localization and function for lumican, fibromodulin and decorin to regulate collagen fibril organization in periodontal tissues. *Journal of periodontal research* **40**, 312-324 (2005).
45. Kessler, E., Mould, A.P. & Hulmes, D.J. Procollagen type I C-proteinase enhancer is a naturally occurring connective tissue glycoprotein. *Biochemical and biophysical research communications* **173**, 81-86 (1990).
46. Challa, A.A. & Stefanovic, B. A novel role of vimentin filaments: binding and stabilization of collagen mRNAs. *Molecular and cellular biology* **31**, 3773-3789 (2011).
47. Pankov, R. & Yamada, K.M. Fibronectin at a glance. *J Cell Sci* **115**, 3861-3863 (2002).
48. Andrews, N.W., Abrams, C.K., Slatin, S.L. & Griffiths, G. A T. cruzi-secreted protein immunologically related to the complement component C9: evidence for membrane pore-forming activity at low pH. *Cell* **61**, 1277-1287 (1990).
49. Naba, A. et al. The matrisome: in silico definition and in vivo characterization by proteomics of normal and tumor extracellular matrices. *Mol Cell Proteomics* **11**, M111 014647 (2012).
50. Hughes, C.S., Postovit, L.M. & Lajoie, G.A. Matrigel: a complex protein mixture required for optimal growth of cell culture. *Proteomics* **10**, 1886-1890 (2010).
51. Hynes, R.O. & Naba, A. Overview of the matrisome--an inventory of extracellular matrix constituents and functions. *Cold Spring Harbor perspectives in biology* **4**, a004903 (2012).
52. Keene, D.R., Engvall, E. & Glanville, R.W. Ultrastructure of type VI collagen in human skin and cartilage suggests an anchoring function for this filamentous network. *The Journal of cell biology* **107**, 1995-2006 (1988).

53. Atkinson, J.C., Ruhl, M., Becker, J., Ackermann, R. & Schuppan, D. Collagen VI regulates normal and transformed mesenchymal cell proliferation in vitro. *Experimental cell research* **228**, 283-291 (1996).
54. Beretta, B. et al. Antigenic determinants of bovine serum albumin. *International archives of allergy and immunology* **126**, 188-195 (2001).
55. Griffiths, L.G., Choe, L.H., Reardon, K.F., Dow, S.W. & Christopher Orton, E. Immunoproteomic identification of bovine pericardium xenoantigens. *Biomaterials* **29**, 3514-3520 (2008).
56. Chang, J.W. et al. Peroxiredoxin-I is an autoimmunogenic tumor antigen in non-small cell lung cancer. *FEBS letters* **579**, 2873-2877 (2005).
57. Karasawa, R., Ozaki, S., Nishioka, K. & Kato, T. Autoantibodies to peroxiredoxin I and IV in patients with systemic autoimmune diseases. *Microbiology and immunology* **49**, 57-65 (2005).
58. Xu, Y., Harder, K.W., Huntington, N.D., Hibbs, M.L. & Tarlinton, D.M. Lyn tyrosine kinase: accentuating the positive and the negative. *Immunity* **22**, 9-18 (2005).
59. Kajita, M. et al. Filamin acts as a key regulator in epithelial defence against transformed cells. *Nature communications* **5**, 4428 (2014).
60. Schreiber, V. et al. Lasp-1, a novel type of actin-binding protein accumulating in cell membrane extensions. *Molecular medicine (Cambridge, Mass.)* **4**, 675-687 (1998).
61. Tsuruta, D. & Jones, J.C. The vimentin cytoskeleton regulates focal contact size and adhesion of endothelial cells subjected to shear stress. *J Cell Sci* **116**, 4977-4984 (2003).
62. Marcal, H., Ahmed, T., Badylak, S.F., Tottey, S. & Foster, L.J. A comprehensive protein expression profile of extracellular matrix biomaterial derived from porcine urinary bladder. *Regenerative medicine* **7**, 159-166 (2012).
63. Gilpin, S.E. et al. Perfusion decellularization of human and porcine lungs: bringing the matrix to clinical scale. *The Journal of heart and lung transplantation : the official publication of the International Society for Heart Transplantation* **33**, 298-308 (2014).
64. Hill, R.C., Calle, E.A., Dzieciatkowska, M., Niklason, L.E. & Hansen, K.C. Quantification of extracellular matrix proteins from a rat lung scaffold to provide a molecular readout for tissue engineering. *Mol Cell Proteomics* **14**, 961-973 (2015).
65. Moghe, P.V. et al. Culture matrix configuration and composition in the maintenance of hepatocyte polarity and function. *Biomaterials* **17**, 373-385 (1996).
66. Sellaro, T.L. et al. Maintenance of human hepatocyte function in vitro by liver-derived extracellular matrix gels. *Tissue engineering. Part A* **16**, 1075-1082 (2010).
67. Lin, Y.M., Zhang, A., Rippon, H.J., Bismarck, A. & Bishop, A.E. Tissue engineering of lung: the effect of extracellular matrix on the differentiation of embryonic stem cells to pneumocytes. *Tissue engineering. Part A* **16**, 1515-1526 (2010).
68. Schuppan, D., Ruehl, M., Somasundaram, R. & Hahn, E.G. Matrix as a modulator of hepatic fibrogenesis. *Seminars in liver disease* **21**, 351-372 (2001).
69. Zhu, C., Coombe, D.R., Zheng, M.H., Yeoh, G.C. & Li, L. Liver progenitor cell interactions with the extracellular matrix. *Journal of tissue engineering and regenerative medicine* **7**, 757-766 (2013).
70. Badylak, S.F., Weiss, D.J., Caplan, A. & Macchiarini, P. Engineered whole organs and complex tissues. *Lancet* **379**, 943-952 (2012).

Chapter 3

Investigating Protein Turnover during *in vitro* Tissue Remodeling

Adapted from:

“Protein turnover during *in vitro* tissue engineering.” **Q. Li**, Z. Chang, G. Oliveira, M. Xiong, L. M. Smith, B. L. Frey, N. V. Welham. *Biomaterials*. 2016, 81, 104-113.

3.1 Introduction

One of the results presented in Chapter 2 is that decellularized biological scaffolds still contain a large number of cellular proteins, and a number of matrisome proteins have been significantly removed during decellularization. A natural question to ask is, what happens to the removed matrisome proteins and retained cellular proteins after cell reseeded, or how do the reseeded cells turn over the scaffold proteome?

The acellular scaffold contains tissue-specific extracellular matrix (ECM) that has been shown to direct stem and progenitor cells towards a target fate^{1, 2}, as well as maintain the functional phenotype of somatic cells in extended culture^{3, 4}. Beyond regulating cell behavior, the scaffold itself is also continuously remodeled by its resident cells. This dynamic reciprocity constitutes an advantage of acellular biological scaffolds over synthetic materials for tissue reconstruction⁵. Prior work on matrix remodeling has focused primarily on accumulation of individual structural matrix proteins (e.g., collagens)^{6, 7} and/or cellular secretion of known matrix-remodeling enzymes (e.g., matrix metalloproteinases)^{8, 9}. However, the current human matrisome (i.e., all ECM and ECM-associated proteins) consists of >1000 proteins¹⁰. This large number of proteins, especially when combined with the various complex interactions and signaling networks formed between the ECM and its resident cells, creates a significant analytical challenge. A proteome-wide analysis addresses this challenge by characterizing the complex and synergistic biological events that comprise the remodeling process. Moreover, since tissue remodeling is a dynamic process, it is desirable to reveal protein turnover by differentiating between original and the newly-synthesized proteins.

Stable isotope labeling with amino acids in cell culture (SILAC) is a quantitative proteomics method^{11, 12}, wherein two cell populations are cultured in media that are identical except that one contains a “heavy” and the other a “light” form of a particular amino acid (e.g., $^{13}\text{C}_6$ - versus $^{12}\text{C}_6$ -lysine, respectively). These isotopically labeled amino acids are metabolically incorporated into each cell’s proteome, and the two populations are mixed prior to MS sample preparation and analysis. The resulting MS peak ratios between “heavy” and “light” forms indicate relative protein abundances. SILAC has been used to study protein turnover in cells¹³, animals¹⁴ and plants¹⁵, and protein half-lives can be calculated¹⁶.

In the work presented in this chapter, using vocal fold mucosa (VFM) as the model system, we compared different decellularization and recellularization approaches. We then developed a novel strategy using SILAC to differentiate between proteins originally present in the acellular scaffolds and newly synthesized ones, thereby assessing active protein synthesis and *in vitro* remodeling of the ECM. The entire workflow is summarized in Fig. 3.1. This study is the first to analyze the dynamic relationship between the matrix and its resident cells, providing biological system-wide insight into the protein turnover that is central to tissue remodeling.

3.2 Materials and Methods

Porcine and human VFM preparation—Porcine larynges were harvested from female market pigs (age 6-8 mo) and snap frozen within 2 h of death. Human larynges were harvested from female cadavers (age 27-73 y) under IRB exemption and snap frozen within 3-48 h of death. Prior to experimentation, larynges were thawed overnight at 4 °C and each VFM specimen (epithelium and lamina propria) was microdissected from its underlying thyroarytenoid muscle.

VFM decellularization—Porcine VFM were assigned to five decellularization protocols, as detailed in Fig. 3.2a. Strategies 1 and 2 consisted of immersion in 1% CHAPS or 1% SDS, respectively, for 24 h at room temperature (R.T., 22 °C), followed by PBS wash for 24 h at R.T. Strategy 3 was previously reported by Xu *et al*¹⁷. Briefly, osmotic stress was first applied by immersing the samples in a highly hypertonic 3M NaCl solution for 24 h at R.T. Samples were then treated with 25 µg/mL DNase I and 10 µg/mL RNase A in an isotonic PBS-containing EDTA-free mini-protease inhibitor cocktail for 24 h at 37 °C, followed by 70% ethanol for 24 h at R.T. Another round of DNase and RNase digestion (at the concentrations noted above) was performed for 48 h at 37 °C, followed by PBS wash for 24 h at R.T. Strategies 4 and 5 involved the addition of either 1% CHAPS or 1% SDS treatment to strategy 3 following the second round of enzyme treatment. For all conditions, a shaker applied continuous mechanical agitation, and 1000 U/mL penicillin and 1 mg/mL streptomycin in PBS were added at each step to mitigate potential bacterial contamination of the decellularized tissue.

Following the initial experiment, strategy 4 was selected for decellularization of porcine and human VFM specimens in subsequent recellularization experiments. In these later experiments, DNase I and RNase A concentrations were adjusted to 500 U/mL and 20 µg/mL respectively.

Recellularization of decellularized VFM with immortalized Vocal Fold Fibroblasts (VFFs)—A previously characterized VFF cell line¹⁸ was used for all recellularization experiments. Five cell seeding protocols were evaluated, as illustrated in Fig. 3.3a. Each decellularized scaffold was placed in the apical chamber of a culture insert with either its luminal (strategy 1) or deep LP (strategies 2-5) surface facing upwards. DMEM (1.7 mL, containing 10% fetal bovine serum, 100 U/mL penicillin, 100 µg/mL streptomycin, 0.25 µg/mL amphotericin B)

was added to the basolateral chamber and 5×10^5 VFFs in 0.5 mL DMEM was pipetted onto the seeding surface of each scaffold. Strategies 1 and 3 included a post-seeding centrifugation step, performed at 170 *g* for 8 min. Strategy 4 involved placement of a platelet-derived growth factor (PDGF)-infused gel into the basolateral chamber at the time of cell seeding, followed by replacement of the PDGF gel at 24 h and 1 w post-seeding. The PDGF-infused gel was prepared by first making 50 ng/mL PDGF (R&D systems, Minneapolis, MN) in type I collagen solution (pH 7.2), and incubated at 37 °C for 2 hours to allow gel formation. Strategy 5 involved soaking the scaffold in 1 mL type I collagen solution (pH 7.5) with agitation at 4 °C overnight, and incubating in 37 °C 5% CO₂ incubator for 2 hours for gel formation. All seeded scaffolds were first incubated at 37 °C in 5% CO₂ overnight, transferred to a new well after 24 h, and cultured for 6 w. Half of each sample, corresponding to the anterior VFM, was harvested under sterile conditions at 3 w. Unseeded scaffolds (retained as negative control samples) were subject to the same culture conditions.

Hydroxyproline and sulfated glycosaminoglycan (sGAG) assays—Hydroxyproline content was measured using a commercial detection kit (BioVision, Mountain View, CA) following sample hydrolysis in 12 N HCl for 3 h at 120 °C. Absorbance was measured at 560 nm. sGAG content was measured using the Blyscan assay (Biocolor, Carrickfergus, UK) following processing with a detergent removal spin column (Thermo Scientific) and papain extraction, according to the manufacturer's instructions. Absorbance was measured at 656 nm.

VFF metabolic labeling, reseeded and culturing—For the metabolic labeling experiment, VFFs were cultured and expanded for 7 d in Dulbecco's Modified Eagle Medium (DMEM) containing 100 mg/L ¹³C₆-lysine and 100 mg/L ¹³C₆-arginine (SILAC Protein Quantitation kit; Thermo Scientific). One concern associated with SILAC is the metabolic conversion of heavy

arginine into heavy proline^{19, 20}, resulting in satellite peaks for proline-containing peptides and consequently inaccurate quantification. We minimized this effect by supplementing the SILAC media with 100 mg/L ¹²C₆-proline. We measured the extent of the problem by including heavy proline and oxidized heavy proline as variable modifications during the database search, and we found less than 0.1% of all peptides contain either of the two modifications. Therefore, the effect of arginine to proline conversion was negligible.

The heavy-labeled human VFFs were analyzed by LC-MS/MS proteomics as described in the Materials and Methods section. A total of 6894 peptides were identified at 1% false discovery rate (FDR), which matched to 907 proteins when requiring at least two unique peptides per protein. Among these identified peptides, 6733 of them were quantified, meaning they had valid intensity values. Of those, 1794 of them had a heavy-to-light ratio (H/L), and thus the percentage of the heavy form (H%) could be calculated. The other 4939 had zero light intensity and thus invalid H% values, which were replaced with 100%. The overall average H% for the 6733 peptides is 98.9%, indicating near-complete metabolic labeling.

Heavy-labeled cells were seeded on decellularized human VFM scaffolds using recellularization strategy 3 and then cultured. Samples were harvested for proteomic assays at 1, 2, 3, 4, 5 and 6 week (w) post-seeding; biopsies of the 3 and 6 w samples were additionally processed for histology.

Although the reseeded scaffolds were cultured in heavy media, it is worth considering whether light amino acids from degradation of the scaffold were incorporated during *de novo* protein synthesis. If that were to happen, there would be proteins containing both light and heavy arginine/lysine²¹. We searched for this possibility by considering all peptide spectral matches (PSMs) having one or two missed cleavages, as these are the only peptides containing more than

one arginine/lysine. Of these 7387 PSMs in the 6 w sample (4 biological replicates $\times$ 2 technical replicates), only 65 PSMs ($< 1\%$) from 16 peptides have mixed heavy and light arginine/lysine. Furthermore, the median posterior error probability (PEP) for these 16 peptides is 1.17×10^{-1} , whereas the PEP of all peptides is 2.24×10^{-7} , which indicates these 16 peptides are likely to be false positive hits. Thus, we find very little evidence of protein synthesis using light amino acids derived from the breakdown of scaffold proteins.

Histology, cell migration analysis, and cell number estimation—Samples intended for histology were rinsed in PBS and fixed in 1 mL of 4% paraformaldehyde (PFA) at 4 °C for 1 h, incubated in 1 mL of 25% sucrose at 4 °C overnight, and embedded in Tissue-Tek Optimum Cutting Temperature compound (Sakura Finetek, Tokyo, Japan). Serial frozen sections (8 μ m thickness) were prepared in the coronal plane using a cryostat. For the decellularization and recellularization strategy comparison experiments, 200 serial sections (representing 1.6 mm total tissue thickness) were prepared from each sample, beginning at the midmembranous VFM transection plane and moving towards either the anterior or posterior pole. Three sections of every 10 were stained with hematoxylin and eosin (H&E), Alcian blue (pH 2.5) and Movat's pentachrome. For the metabolic labeling experiment, 20 serial sections were prepared from each biopsy sample and stained with H&E. All sections were imaged using standard light microscopy.

For the recellularization strategy comparison experiment, cell migration was quantified using H&E-stained sections and Metamorph 7.5 (Molecular Devices, Downingtown, PA) by measuring the shortest distance between each individual cell and the epithelial seeding surface. To minimize possible artifacts due to uneven cell distribution throughout the scaffold, image analysis was performed on at least 10 sections per sample, selected from each 1.6 mm span of serial sections.

Additional cell counting was performed on 3 and 6 w sections obtained from the validation of recellularization strategy 3 using decellularized human VFM scaffolds. The mean number of cells per 8 μm section at 3 and 6 w post-seeding was used to estimate the total number of cells in an entire human female VFM scaffold with average length of 10 mm²². The cell numbers were 8.2×10^4 and 2.9×10^5 , respectively. By logarithmic growth curve extrapolation, the number of cells that actually engrafted in the scaffold at the time of seeding was 2.4×10^4 (~5% of the cells we used for reseeding). Since the decellularized scaffold we analyzed was half of the entire scaffold, we mixed a certain portion of the heavy cell peptide solution with a certain portion of the scaffold peptide solution to achieve the “scaffold plus 1.2×10^4 cells”, “scaffold plus 4.1×10^4 cells” and “scaffold plus 1.5×10^4 cells” conditions, and used them as the 0 w sample, the 3 w control, and the 6 w control, respectively.

Proteomics sample preparation—For heavy-labeled VFF, protein was extracted with 150 μL of SDT solution containing 4% SDS, 0.1 M Tris-HCl (pH 7.6) and 0.1 M dithreothreitol. For intact, decellularized, and reseeded human VFM, approximately 15 mg tissue pieces were washed with ice cold PBS, then ground with disposable pellet pestles (Kimble Chase Kontes, Vineland, NJ), before 150 μL of SDT solution was added. Samples were then heated at 95 $^{\circ}\text{C}$ for 7 min. and sonicated on ice with a probe sonicator—alternating 20 seconds on and 20 seconds off for 6 min, followed by centrifugation at 20 $^{\circ}\text{C}$ for 5 min at 16,100 g. The FASP (Filter-Aided Sample Preparation) protocol was used for SDS removal and on-filter digestion²³. Briefly, a 30 μL aliquot of the supernatant was added to a 30K MW Vivacon 500 filter (Sartorius, Bohemia, NY), washed, alkylated, and digested with trypsin (protein:enzyme ratio of 50:1) overnight at 37 $^{\circ}\text{C}$. Finally, the digest was collected by centrifugation. After the digestion was quenched with 10% trifluoroacetic acid (TFA) to a final concentration of 0.5% TFA, samples

were desalted using Sep-Pak C18 1 cc Vac Cartridges (Waters, Milford, MA), according to the manufacturer's instructions. Eluate was dried down and reconstituted in 5% ACN and 2% FA.

Mass spectrometry and data analysis—Approximately 1 µg protein digest (estimated by BCA assay) of the heavy-labeled fibroblasts, or 0.05 mg tissue equivalent was injected into a Waters nanoAcquity HPLC coupled to an ESI ion-trap/orbitrap mass spectrometer (LTQ Orbitrap Velos, Thermo Scientific, Waltham, MA). Peptides were separated on a 100 µm inner diameter column packed with BEH C18 particles (Waters, Milford, MA), and eluted at 0.3 µL/min in 0.1% FA with a gradient of increasing ACN over 2.5 h. A heater cartridge was used to keep the capillary column at 60 °C. A full-mass scan (300-1500 m/z) was performed in the orbitrap at a resolution of 60,000 and acquired in profile mode. The ten most intense peaks were selected for fragmentation by high-energy collisional dissociation (HCD) at 42% collision energy, with a resolution of 7500, and isolation width of 2.5 m/z . Dynamic exclusion was enabled with a repeat count of 2 over 30 s and an exclusion duration of 120 s. Each sample had four biological replicates and the 1w to 6w samples were subjected to LC-MS/MS analysis twice for each biological replicate.

The acquired raw files were analyzed by MaxQuant (version 1.4.1.2)²⁴. The derived peak lists were searched with Andromeda against the UniProt canonical protein database (*Homo sapiens*: 20278 sequences downloaded on December 5, 2013) supplemented with common contaminants. All biological and technical replicates for a particular sample were searched together. Precursor and fragment ion mass tolerances were set to 4.5 ppm and 20 ppm, respectively. Static cysteine carbamidomethylation (+57.0215 Da) and up to 7 variable methionine and proline oxidation (+15.9949 Da) were specified. A false discovery rate (FDR) of 1% at both the peptide and the protein level was allowed. Up to two missed cleavages were

allowed and a minimum of two unique peptides per protein was required. The “match between runs” function was enabled. A minimum of two unique and razor peptide ratio counts was required and only unmodified peptides were used to quantify a protein. Protein groups containing matches to proteins from the reversed database or contaminants were discarded. The mass spectrometry proteomics data have been deposited in the ProteomeXchange Consortium²⁵ via the PRIDE partner repository with the dataset identifier PXD002734.

Statistical analyses—Technical replicates were averaged and all statistical comparisons were performed using independent biological replicates. Gene ontology term enrichment analysis was performed using the BiNGO²⁶ (hypergeometric model with Benjamini-Hochberg correction) and REVIGO²⁷ (SimRel cutoff = 0.4) algorithms. Ontology term enrichment schematics were generated using Cytoscape 2.8.2²⁸. The differences among multiple groups were analyzed with one-way ANOVA, and independent samples t -test was performed between two groups. P value of <0.05 was considered statistically significant.

3.3 Results

Comparison of five decellularization strategies—The first step towards creating a functional tissue is to engineer a scaffold that retains much of the ECM composition and architecture of the original tissue while removing any cellular remnants that hold the potential to trigger a maladaptive immune response. Previous work has shown that optimal decellularization is tissue- or organ-specific⁵. The vocal fold mucosa (VFM) is an attractive model for matrix remodeling studies because its unique ECM is biomechanically tuned for voice production;

disordered VFM is recalcitrant to current clinical therapies and so represents a substantial, albeit challenging, tissue engineering need²⁹.

We therefore decellularized porcine VFM using five different strategies, each consisting of isolated or sequential treatments with a zwitterionic (CHAPS) or anionic (SDS) detergent, osmotic stress, and nuclease digestion (Fig. 3.2a), in order to identify an optimal procedure. A hydroxyproline assay showed preservation of native collagen abundance with each of the five strategies ($P > 0.05$; Fig. 3.2b), suggesting maintenance of scaffold structural integrity and strength. sGAG abundance was significantly depleted with strategies 3 and 5 ($P < 0.05$; Fig. 3.2b), indicating probable impairment of tissue viscosity and biomechanical performance³⁰ but consistent with decellularization outcomes in other tissues³¹. The lower concentrations of both hydroxyproline and sGAG with strategy 2 compared to 1, as well as with strategy 5 compared to 4, suggest that that SDS is more disruptive to the VFM ECM than CHAPS. Hematoxylin and eosin (H&E), Alcian blue and Movat's pentachrome histological stains were used to evaluate the effectiveness of cell removal and preservation of native ECM architecture. The consistent absence of cell nuclei in histologic sections from strategies 3-5 (Fig. 3.2c; Fig. 3.3) support the benefit of osmotic stress and nuclease digestion in lysing cells and removing nuclear material *in situ*. Strategy 4 was selected for decellularization of porcine and human VFM in subsequent recellularization and isotopic labeling experiments, based on its effective depletion of cells and superior quantitative preservation of collagen and sGAGs.

Comparison of five reseeding strategies—To find a seeding condition that promotes maximum cell infiltration into the scaffold, we compared five reseeding strategies in decellularized porcine VFM (Fig. 3.4a). Vocal fold fibroblasts (VFFs) were seeded on either the luminal or deep lamina propria (LP) surface of the scaffold and treated with either gentle

centrifugation, chemoattraction using PDGF, or encapsulation in a collagen I gel³²⁻³⁴. H&E staining and cell migration analysis were performed 3 w after reseeding (Fig. 3.4b and 3.4c). Seeding on the deep LP surface combined with either centrifugation (strategy 3) or chemoattraction (strategy 4) resulted in superior cell migration, compared to all other strategies ($P < 0.05$). With strategies 3 and 4, VFFs infiltrated the scaffold to a mean depth of $\sim 200 \mu\text{m}$ and a maximum depth of $\sim 800 \mu\text{m}$, which approximates the mean thickness of porcine VFM ($\sim 900 \mu\text{m}$)²². Given the expense and possibility of unanticipated off-target effects from extended PDGF treatment, we selected strategy 3 for VFF reseeding in subsequent experiments. We next repeated the reseeding experiment in decellularized *human* VFM and extended the culture time to 6 w. Fig. 3d shows a substantial increase in cell density from 3 w to 6 w, suggesting ongoing VFF proliferation throughout the culture period. The final recellularized construct at 6 w contained VFFs of comparable morphology and density to those of native VFM (Fig. 3.5).

Proteomic characterization of native, decellularized and recellularized human VFM—To further characterize the selected decellularization and recellularization strategies, we performed LC-MS/MS-based proteomic analysis of native, decellularized, and recellularized (6 w post-seeding) human VFM. A total of 1028, 509 and 704 proteins, respectively, were identified using a 1% false discovery rate (FDR) (Fig. 3.6a). Gene Ontology (GO) enrichment analyses were performed on the set of 429 proteins that were exclusively identified in the native condition, as well as the set of 160 proteins that were exclusively identified in the recellularized condition. Native VFM was characterized by enrichment of an array of biological process (BP) terms including those associated with defense response and muscle contraction, consistent with the presence of epithelial and immune cells³⁵ that were not used in our recellularization experiments, as well as residual muscle cells and fibers that remain in the native VFM despite careful

microdissection during sample preparation³⁶. Recellularized VFM was characterized by enrichment of BP terms associated with various metabolic, signaling, transport and regulatory functions, as well as a set of interconnected BP terms associated with biogenesis, morphogenesis and developmental processes (Fig. 3.6b). Interestingly, a number of these BP terms relate to specific tissue substructures, such as the epithelium, vascular and nervous systems, implying that VFFs in the recellularized VFM respond to regional cues and engage in remodeling these subspecialized ECMs, in addition to the primary ECM of the lamina propria.

We also specifically examined the effects of decellularization and recellularization on matrisome proteins. Note that the non-matrisome proteins, remnant cellular proteins from incomplete removal during decellularization, still constitute a large portion of the proteome of decellularized VFM (Figs. 3.6a, 3.6c); this has been observed in other types of acellular scaffolds³⁷. A total of 46 matrisome proteins were removed by decellularization (Fig. 3.6c). The majority of these proteins were ECM regulators, ECM-affiliated proteins, and secreted factors: their removal resulted in a corresponding decrease in the relative MS intensity attributed to these three matrisome categories (Fig. 3.6d). In contrast, core matrisome collagens, glycoproteins and proteoglycans showed generally well-preserved MS intensity following decellularization, consistent with our previous hydroxyproline and histological data (Fig. 3.2b and 3.2c). Sixteen of the 46 decellularization-removed proteins were replenished by recellularization (Fig. 3.6c). Most of these newly synthesized matrisome proteins were ECM regulators that contribute to matrix remodeling (e.g., cathepsins B and C, serine protease HTRA1, matrix metalloproteinase [MMP]-10, serpin E2) and glycoproteins that support cell adhesion (e.g., laminins, EMILIN-2). Of the 34 matrisome proteins that were exclusively identified in recellularized VFM (Fig. 3.6c), the most abundant was thrombospondin-1, an adhesion glycoprotein that supports fibroblast migration and

interaction with the ECM. The most abundant ECM regulators were the matrix metalloproteinases MMP1 and MMP2, as well as the inhibitor TIMP1, all of which are associated with matrix turnover and cell migration. Overall, the proteomics data show that decellularized VFM contains a well-preserved core matrisome, and that the recellularized VFM is characterized by tissue-appropriate biogenesis and matrix remodeling.

Matrisome and cellular protein turnover—Vocal fold fibroblasts are responsible for ECM maintenance and turnover in the functionally important lamina propria region. To investigate these aspects of the tissue engineering process, we developed and implemented a SILAC-based strategy to differentiate between original and newly synthesized proteins. The goal of this experiment, therefore, was to quantitatively measure the rate and extent to which seeded VFFs remodel the decellularized VFM *in vitro*. As illustrated in Fig. 3.1, we isotopically labeled the proteome of cultured VFFs using medium containing $^{13}\text{C}_6$ -lysine (Lys) and $^{13}\text{C}_6$ -arginine (Arg) (99% incorporation of $^{13}\text{C}_6$ -Lys/Arg to the VFF proteome achieved by 6 d). We then seeded these heavy isotope-labeled cells onto decellularized VFM containing naturally occurring $^{12}\text{C}_6$ -Lys/Arg. We continued VFF culture with $^{13}\text{C}_6$ -Lys/Arg supplementation for 6 w; samples were harvested weekly for LC-MS/MS proteomic analysis.

To evaluate overall protein turnover, we first looked at the MS intensity of four broad classes of proteins: heavy (newly synthesized) and light (remnant) for both matrisome and non-matrisome (cellular) protein categories (Fig. 3.7a and 3.7b). The percentages of these 4 classes add up to 100% for each time point, but redistribution occurs over time. The percentage of newly synthesized matrisome proteins (black squares in Fig. 3.7a) increased steadily across the 6 w culture period ($P < 0.05$); the percentage of newly synthesized cellular proteins (black squares in Fig. 3.7b) increased and then plateaued ($P < 0.05$). This observation suggests that VFFs engage

in more active and sustained synthesis of matrisome, versus cellular, proteins during extended *in vitro* culture within decellularized VFM. Analysis of remnant protein forms showed no change ($P > 0.05$) in the normalized MS intensity of matrisome proteins (red circles in Fig. 3.7a) but ~50% degradation of remnant cellular proteins (red circles in Fig. 3.7b) over the 6 w period ($P = 0.05$). Follow-up analysis of these remnant cellular proteins revealed the fastest degradation rates for nuclear and cytoskeletal proteins, moderate degradation of cytoplasmic proteins, and little change in already-low-abundance cell membrane proteins (Fig. 3.8).

Next, we evaluated the extent to which interaction between VFFs and the decellularized VFM contributes to protein synthesis and degradation kinetics. To accomplish this, we collected tissue biopsies from 3 and 6 w samples and employed histology-based cell counts and logarithmic growth curve extrapolation to estimate the total number of cells within each scaffold throughout the experiment (including an estimate of initial VFF engraftment at 0 w). Using these cell counts, we mixed peptide digests isolated from $^{13}\text{C}_6$ -Lys/Arg VFF with peptide digests from decellularized VFM at appropriate ratios to match those of the recellularized tissue at each time point of interest. As these control samples were prepared at the peptide level, there was no opportunity for cell-scaffold interaction. Analysis at 3 and 6 w showed significantly greater MS intensity for heavy-labeled matrisome proteins in the *in vitro* culture samples compared to controls ($P < 0.01$; Fig. 3.7a), suggesting that exposure to the native ECM enhances matrisome synthesis by VFFs. We observed a comparable effect for heavy-labeled cellular proteins at 3 w ($P < 0.01$; Fig. 3.7b), as well as significantly lower MS intensity for remnant cellular proteins in the *in vitro* culture samples compared to controls at 3 w ($P < 0.05$; Fig. 3.7b). Overall, these data indicate that interaction between VFFs and the decellularized VFM promotes sustained matrisome synthesis, early-phase cellular activity that involves an uptick in cellular protein

synthesis (consistent with VFF activation and proliferation), and early-phase enhancement of remnant cellular protein degradation.

Studies on degradation of collagens and other ECM proteins^{38, 39} have suggested two key pathways: a principal intracellular pathway involving phagocytosis with subsequent lysosomal cathepsin digestion, and an extracellular pathway involving secretion of matrix metalloproteinases by fibroblasts^{40, 41}. The light-form matrisome proteins in our *in vitro* system were apparently resistant to degradation (Fig. 3.7a), which may be due to their extensive crosslinking, inadequate culture duration, or the absence of more effective phagocytes that would be present *in vivo*. In contrast, we observed gradual degradation of remnant cellular proteins over time. Little has been reported on the degradation mechanism of remnant cellular proteins with direct exposure to the matrix. Therefore, we fit both zero- and first-order kinetic functions to our MS intensity data (Fig. 3.7c), as both have been experimentally demonstrated for protein degradation. Note that the discrepancy between these functions can be explained by the assumption that the degradation rate follows a Michaelis-Menten function, and the reaction order depends upon the relative magnitude of dissociation constant and substrate concentration⁴². These functions predict complete remnant cellular protein degradation by VFFs at 16 w or beyond in our experimental system.

Matrisome protein synthesis—Based on our observation of sustained matrisome synthesis by VFFs across the 6 w culture period, we performed a follow-up analysis of each matrisome subcategory and its constituent proteins (Fig. 3.9 and 3.10). As we found no evidence of significant remnant matrisome protein degradation over time (Fig. 3.7a), most of the change in the percentage of MS intensity could be attributed to the net output from synthesis and degradation of heavy-labeled matrisome protein by resident VFFs. ECM regulators and

glycoproteins, important players in fibroblast migration and interaction with the matrix, exhibited the fastest increase in MS intensity during the first 4 w. Collagens and proteoglycans, extensively crosslinked core matrisome proteins that were well preserved in the original decellularized scaffold (Fig. 3.2b and 3.2c; Fig. 3.6d), increased more gradually during the same time period, then plateaued. Secreted factors, which include key growth factors and signaling molecules that bind to ECM proteins/glycans and were severely depleted during decellularization (Fig. 3.2b and 3.2c; Fig. 3.6d), exhibited an exponential increase in MS intensity from 2-5 w.

We examined the specific proteins driving MS intensity changes within the core matrisome (Fig. 3.9b - d). Most net collagen synthesis was attributable to the primary fibrillar isoform collagen I (COL1A1) as well as the collagen I-associated isoforms VI (COL6A3) and XII (COL12A1) (Fig. 3.9b). At the final 6 w time point, ~25% of total COL1A1 intensity was due to new protein synthesis. Several ECM glycoproteins exhibited sharp increases in net synthesis during the first week of culture (Fig. 3.9c). The greatest initial increase was seen for heavy-labeled fibronectin (FN1), which accounted for 75-80% of total FN1 intensity during the remainder of the experiment. Early and rapid FN1 synthesis is consistent with its critical roles facilitating the deposition of other ECM proteins and maintaining cell-ECM adhesion sites⁴³. The small leucine-rich proteoglycan (SLRP) family members decorin (DCN), lumican (LUM) and biglycan (BGN), which are regulators of collagen fibril assembly and associated tissue strength⁴⁴,⁴⁵, followed a pattern of gradually increasing net synthesis from 1-5 w, then a decrease at 6 w (Fig. 3.9d).

The primary analytical advantage of our SILAC-based method in a tissue engineering context is the ability to differentiate original and newly synthesized proteins. This is particularly important when a high rate of synthesis but little overall turnover results in a dramatic difference

in a given protein's heavy versus total MS intensity change. To further illustrate this, we plotted heavy, light and total MS intensity fold changes for the proteoglycan DCN at 0, 3 and 6 w (Fig. 3.9e), and compared these data with immunoblots of total decorin (DCN) abundance at 1, 3 and 6 w (Fig. 3.9f). We observed a ~100-fold increase ($P = 0.00029$) in heavy DCN intensity over the 6 w culture period but statistically insignificant change in light ($P = 0.48$) and total ($P = 0.74$) DCN intensity, or DCN abundance. This observation demonstrates the value of isotopic labeling for capturing aspects of protein synthesis kinetics that would otherwise be masked by traditional protein detection methods.

3.4 Discussion

Tissue engineering using acellular biological scaffolds is a popular and promising technique that is dependent upon a synergistic relationship between the scaffold ECM and its seeded cells. Despite the importance of this relationship, traditional assays have been unable to capture the dynamic remodeling events that are presumably responsible for the engineered tissue function reported in prior studies^{31, 46}. This deficiency, in part, is due to the difficulty of interrogating the cell-matrix system comprehensively and quantitatively. In response to this challenge, we adapted a SILAC-based proteomics method to differentiate original and newly synthesized proteins in a tissue engineering context, applied the method to decellularized VFM, and herein present the first broad analysis of protein turnover after reseeding cells on an acellular scaffold.

Across the proteome, and over time, proteins in certain categories were actively degraded whereas others were actively synthesized. Our initial analysis of the decellularized VFM

proteome attributed substantial MS intensity to remnant cellular proteins, despite the absence of cellular structures in histology. We observed ongoing degradation of these remnant proteins during the 6 w culture period, which theoretically corresponds to a decrease in the immunogenic potential of the final engineered tissue^{47, 48}. Further, isotopic labeling enabled measurement of the degradation rates of these cellular proteins *in vitro*; we suspect the rates would be faster *in vivo*.

Biological scaffolds are not intended as permanent implants; rather, they should be biodegradable and subject to ECM turnover by resident cells⁴⁹. In this work with VFM, we observed little degradation of the scaffold's core matrisome; it would be helpful to examine whether longer *in vitro* culture times and/or *in vivo* conditions result in the breakdown and replacement of these large structural proteins. In contrast, isotopic labeling revealed active and sustained synthesis of a myriad of matrisome proteins across the experiment. Different synthesis rates were observed among categories, with ECM glycoproteins, ECM regulators, and secreted factors exhibiting the fastest rates, confirming their importance to tissue remodeling. Notably, remnant cellular protein degradation and matrisome protein synthesis were both significantly increased compared to control samples that lacked interplay between the scaffold and its seeded cells.

In summary, the ability to differentiate residual from newly synthesized proteins at the biological system level provides a more complete understanding of ECM turnover during tissue engineering. The analytical strategy developed here is directly applicable to other tissue/organ types as well as different engineering techniques. Isotopic labeling experiments can also be conducted *in vivo*, whereby an entire organism's proteome is labeled via sustained dietary intake of heavy amino acids⁵⁰. Such an experimental setup could be used to evaluate remodeling of a

tissue engineered graft by host cells. Beyond the realm of tissue engineering, this method could also be applied to a variety of three-dimensional and organotypic culture systems, such as are used in developmental biology⁵¹ and cancer biology⁵².

3.5 Figures and Tables

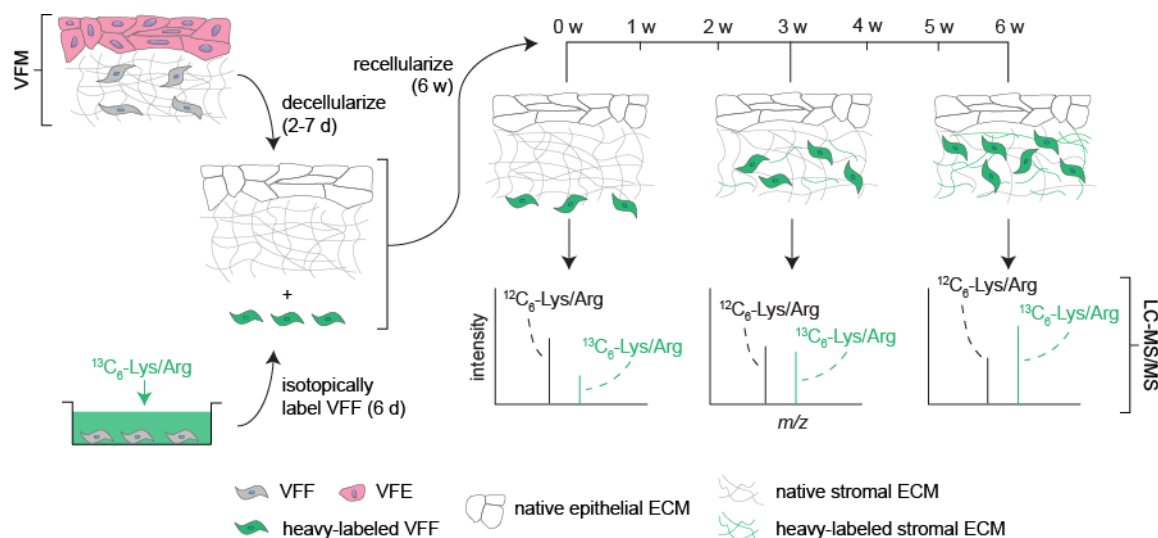


Figure 3.1. A brief depiction of the entire experimental workflow. Vocal fold mucosae (VFM) are decellularized using one of five strategies for 2 to 7 d. Vocal fold fibroblasts (VFFs) are isotopically labeled for sufficient time to ensure full-proteome incorporation of $^{13}\text{C}_6$ -Lys and $^{13}\text{C}_6$ -Arg. Next, the labeled VFFs are seeded and cultured for up to 6 w on the decellularized VFM matrix, with liquid chromatography-tandem mass spectrometry (LC-MS/MS)-based analysis at each of 6 w. Representative cell proliferation and new ECM synthesis are shown only for 0, 3 and 6 w timepoints. (VFE: vocal fold epithelial cell; ECM: extracellular matrix)

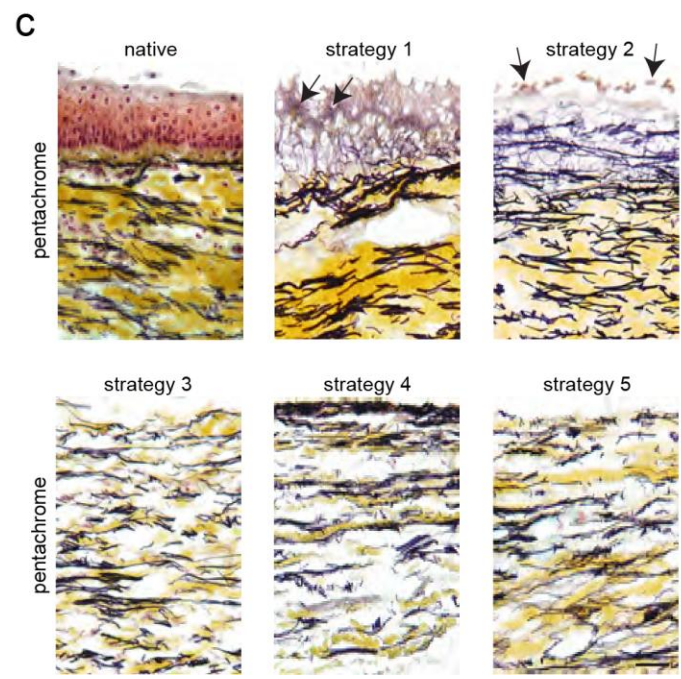
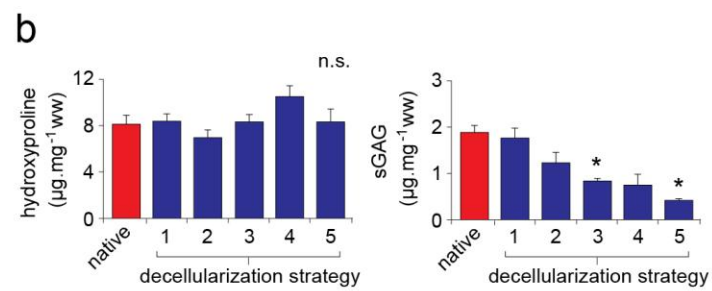
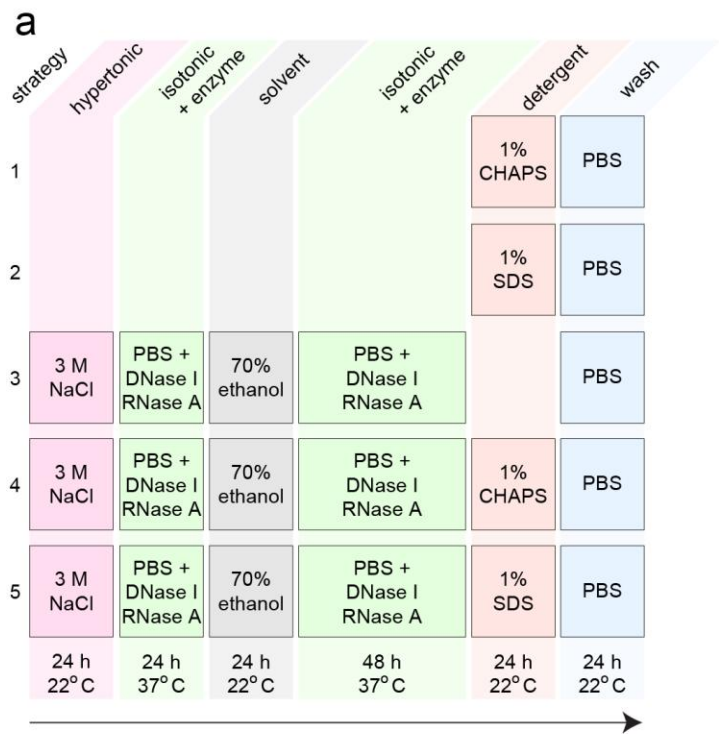


Figure 3.2. Comparison of five decellularization strategies. a) Graphic illustration of the five strategies. b) Quantitative hydroxyproline and sulfated glycosaminoglycan (sGAG) assay results for the native condition and each decellularization condition. $n = 3$ biological replicates, with 3 technical replicates for each. *, $P < 0.05$ versus native condition; n.s., non-significant difference. Error bars, s.e.m. c) Pentachrome-stained sections of porcine VFM, either native or decellularized with one of the five strategies. Arrows indicate residual cells. Yellow color represents collagen and black color represents elastin. Scale bar, 50 μm .

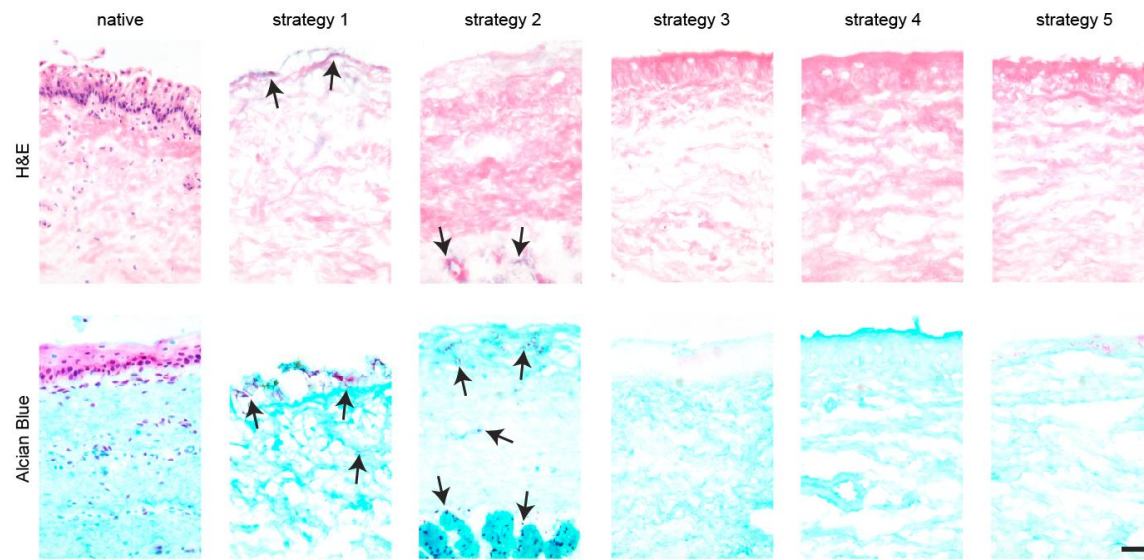


Figure 3.3. H&E- and Alcian Blue-stained sections of porcine VFM, either native or decellularized by each of the five different strategies. Black arrows indicate residual cells. Scale bar, 50 μm .

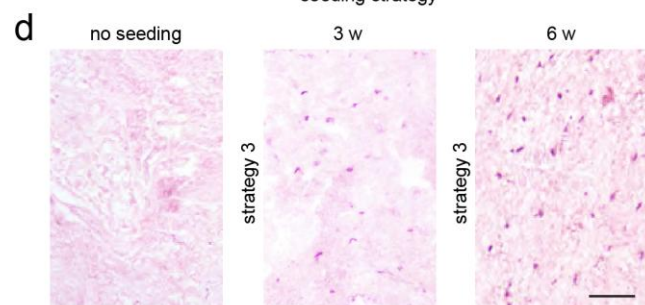
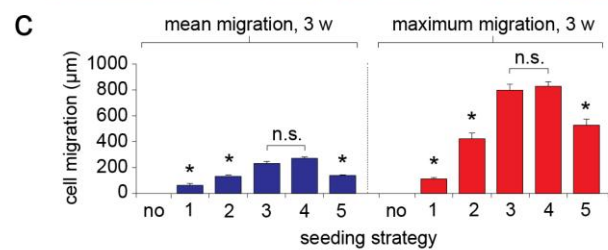
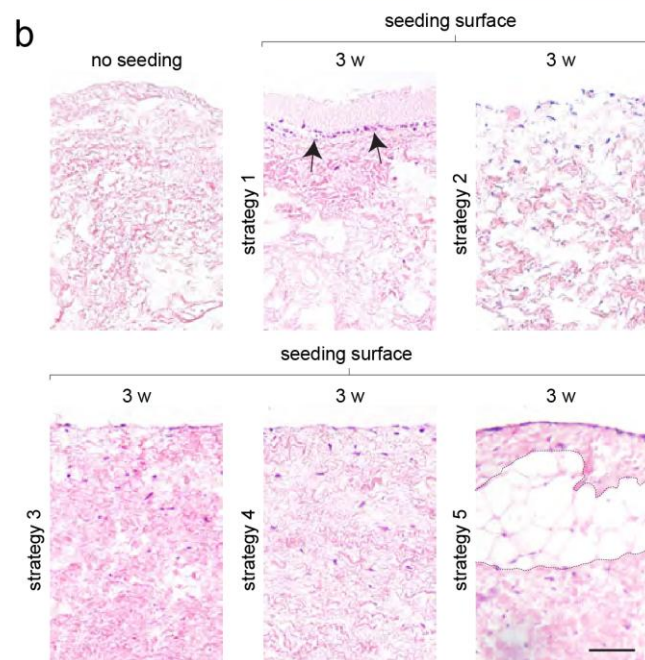
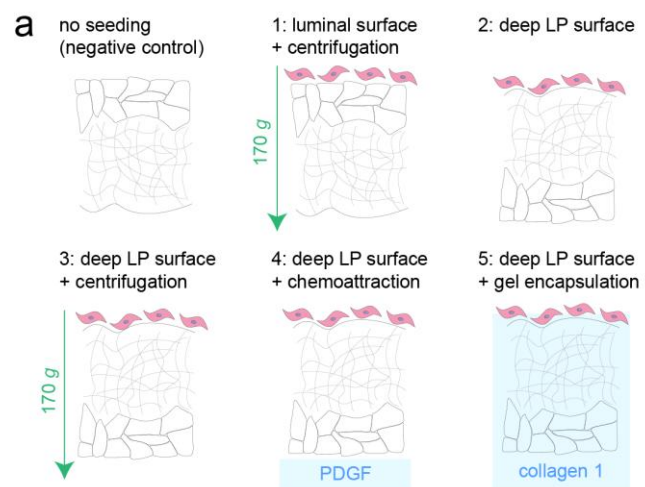


Figure 3.4. Comparison of five reseeding strategies. a) Graphic illustration of the five strategies. b) H&E-stained sections of porcine VFM for the no-seeding (decellularized) control and the samples recellularized (3w post-seeding) with one of the five strategies. The direction of the tissues are the same as panel a), with cells seeded from the top of the images. Arrows show representative cells stuck at basement membrane; dashed lines show polymerized collagen blob. Scale bar, 100 μm . c) Cell migration analysis results for the porcine VFM, either decellularized or recellularized (3w post-seeding) with one of the five strategies. $n = 3$ biological replicates, with 3 technical replicates for each. *, $P < 0.05$ compared to both strategy 3 and 4; n.s., non-significant difference. Error bars, s.e.m. d) H&E-stained sections of human VFM, either decellularized or recellularized with strategy #3 for 3 or 6 w post-seeding. Scale bar, 100 μm .

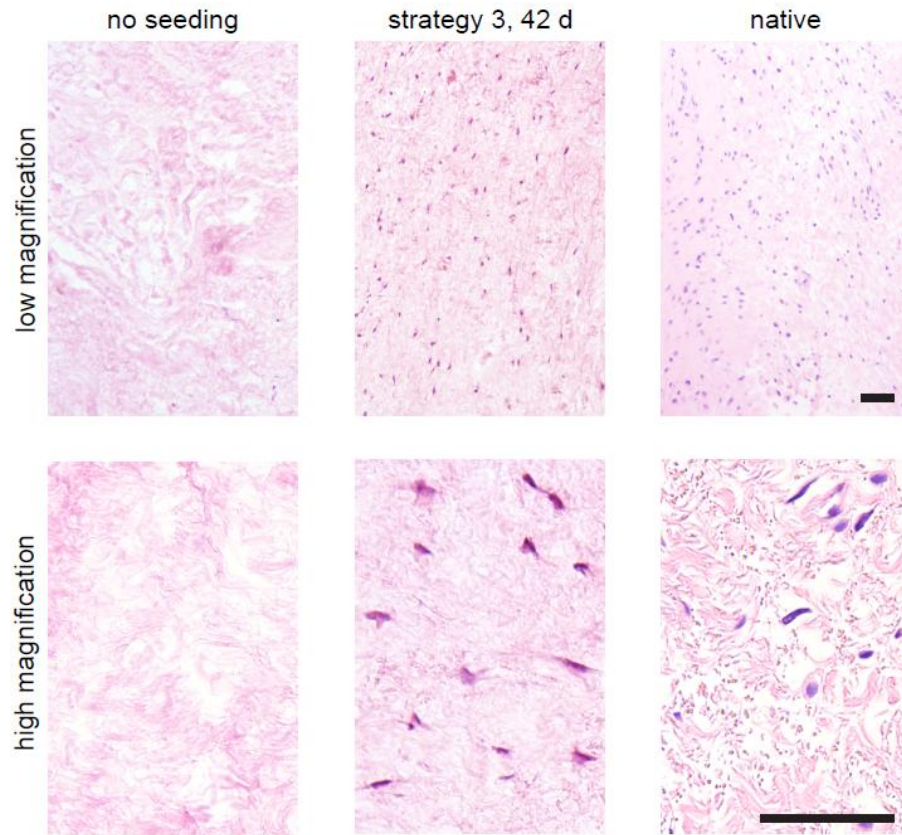


Figure 3.5. H&E-stained sections of human VFM showing comparable morphological features of the recellularized (with strategy 3 for 42 d) tissue with native tissue. Scale bar, 100 μm .

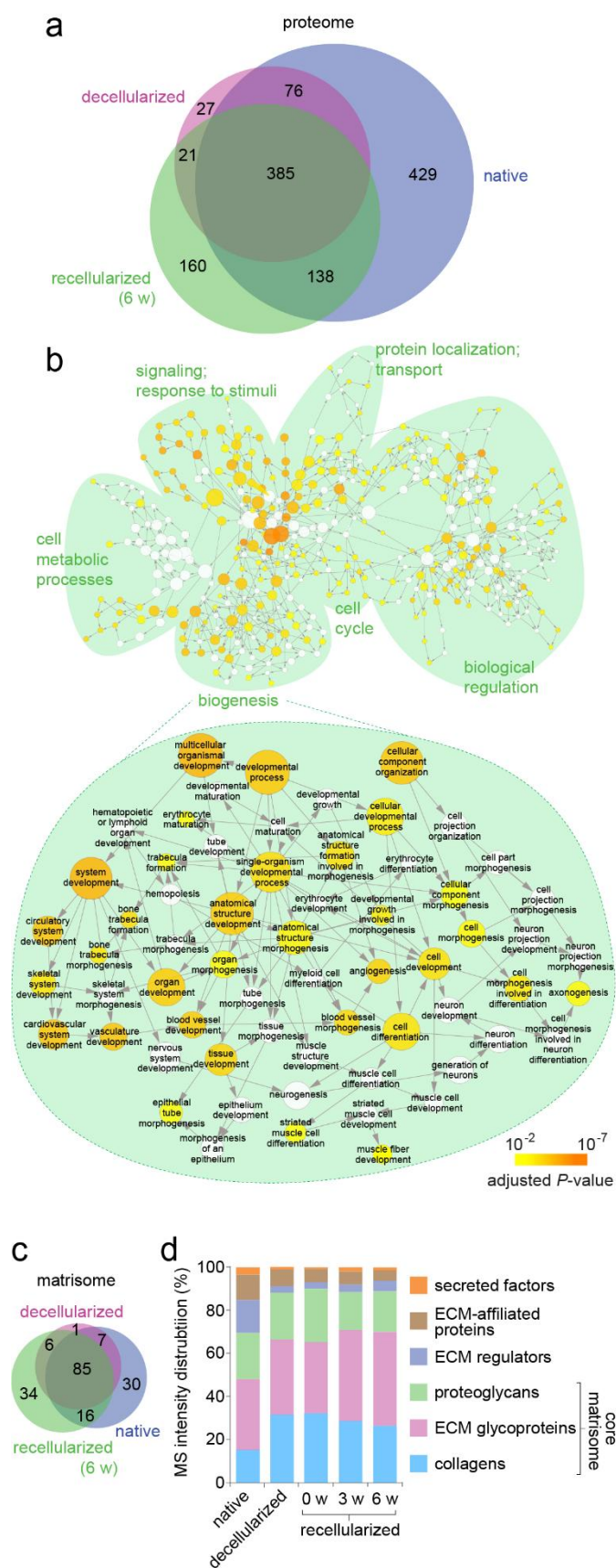


Figure 3.6. Proteomic results comparing the native, decellularized, and recellularized (6 w post-seeding) human VFM. a) Venn diagram with numbers of protein identifications for the whole proteome. b) Functional enrichment analysis of the 160 proteins that were exclusively identified in the recellularized condition. Enriched gene ontology terms are depicted as nodes connected by arrows that represent hierarchies and relationships between terms. Node size is proportional to the number of proteins assigned to a given term; node color represents the Benjamini Hochberg-corrected P-value corresponding to enrichment of the term. Functionally related ontology terms are labeled and grouped using green ovals. Biogenesis functional groups are enlarged for better visualization of individual terms. c) Venn diagram with numbers of protein identifications for the matrisome. d) Mass spectrometry intensity distribution of the six matrisome subcategories.

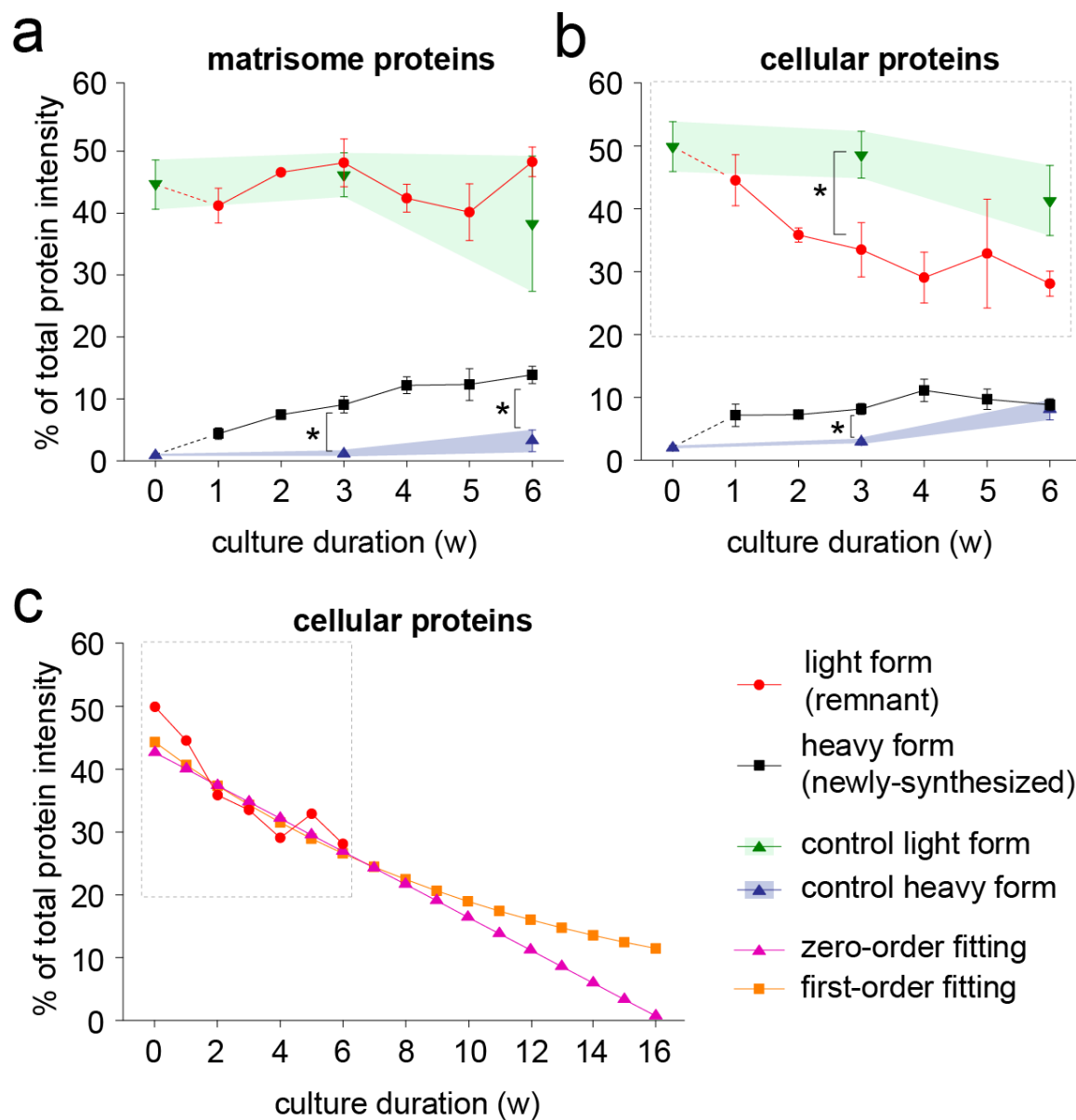


Figure 3.7. Percentage of light or heavy protein intensity, out of the total protein intensity for the entire sample, at each time point for a) matrisome proteins, and b) & c) cellular proteins. The 0 w, and 3 w and 6 w control samples were all from mixing cells and scaffolds at the peptide level. $n = 4$ biological replicates, with 2 technical replicates for each. *, $P < 0.05$. Error bars, s.e.m.

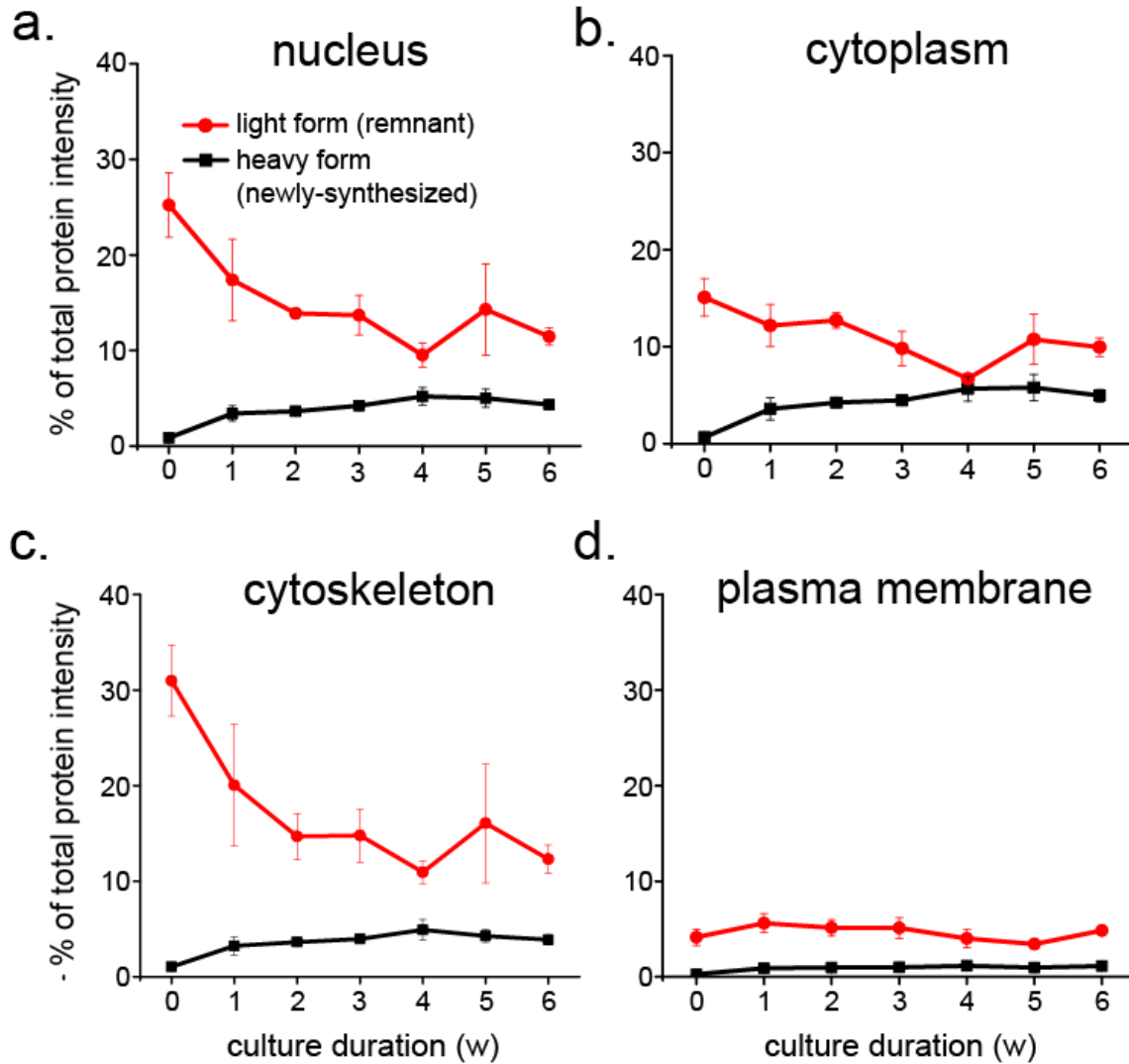


Figure 3.8. Percentage of light or heavy protein intensity, out of total protein intensity in each sample at each time point, for a) nucleus proteins, b) cytoplasm proteins, c) cytoskeleton proteins, and d) plasma membrane proteins. $n = 4$ biological replicates, with 2 technical replicates for each. Error bars, s.e.m.

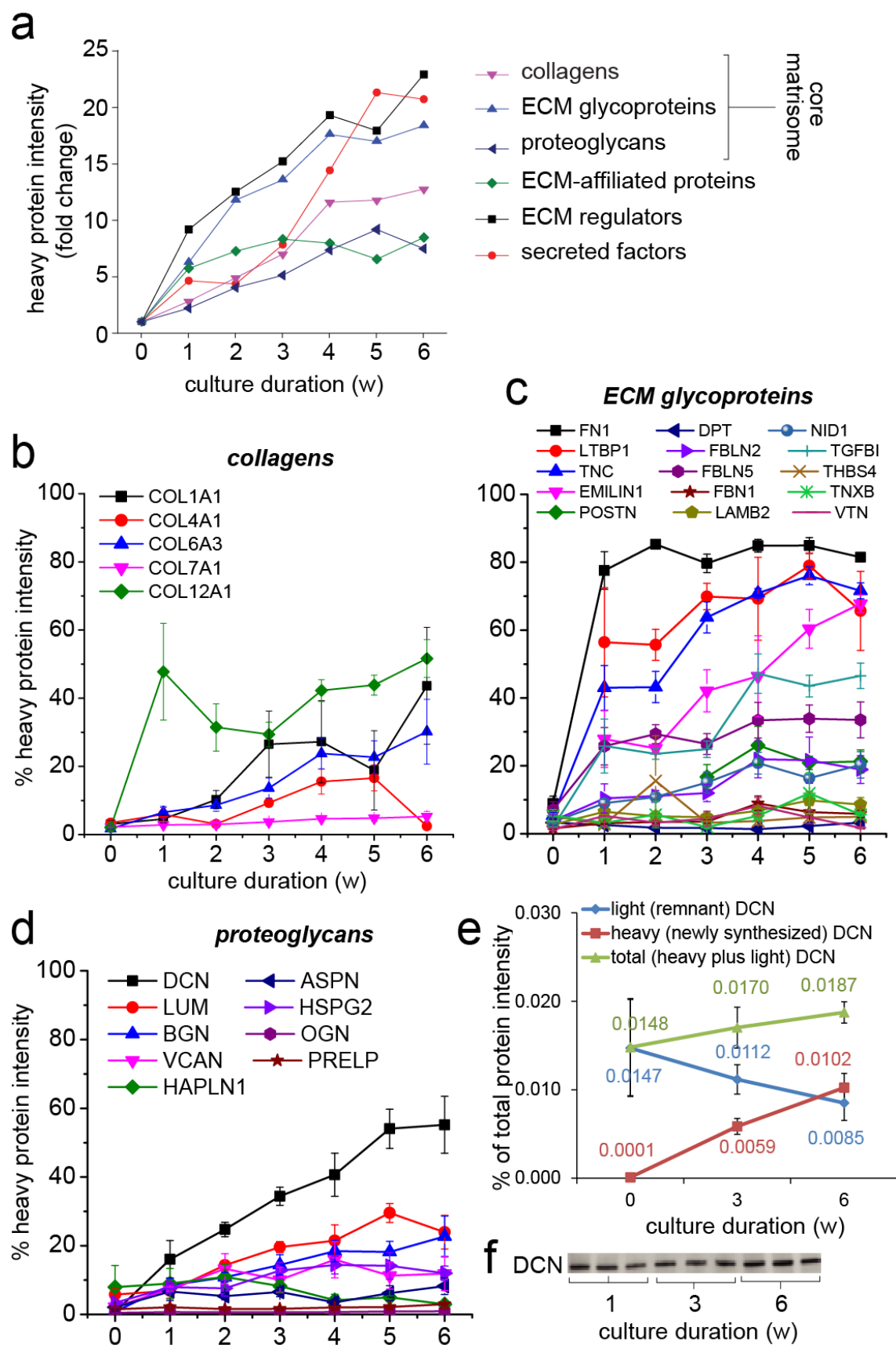


Figure 3.9. New synthesis of matrisome proteins. a) Fold change in heavy (newly synthesized) protein intensity for each of the six matrisome subcategories compared to the 0 w control. Protein turnover is shown for specific proteins from the core matrisome subcategories: b) collagens, c) ECM glycoproteins, and d) proteoglycans. (Note: this turnover is plotted as the percentage of heavy intensity out of the total intensity for each individual protein). e) Percentage of heavy, light, or total (heavy plus light) MS intensity for the proteoglycan decorin, out of the total protein intensity of the particular sample. n = 4 biological replicates, with 2 technical replicates for each. Error bars, s.e.m. f) Immunoblots showing total abundance of decorin. n = 3 biological replicates. (Note: the discrepancy in initial time point for e) and f) was because the 0 w MS data were generated following peptide-level sample preparation, meaning that no directly comparable protein samples were available for immunoblotting.)

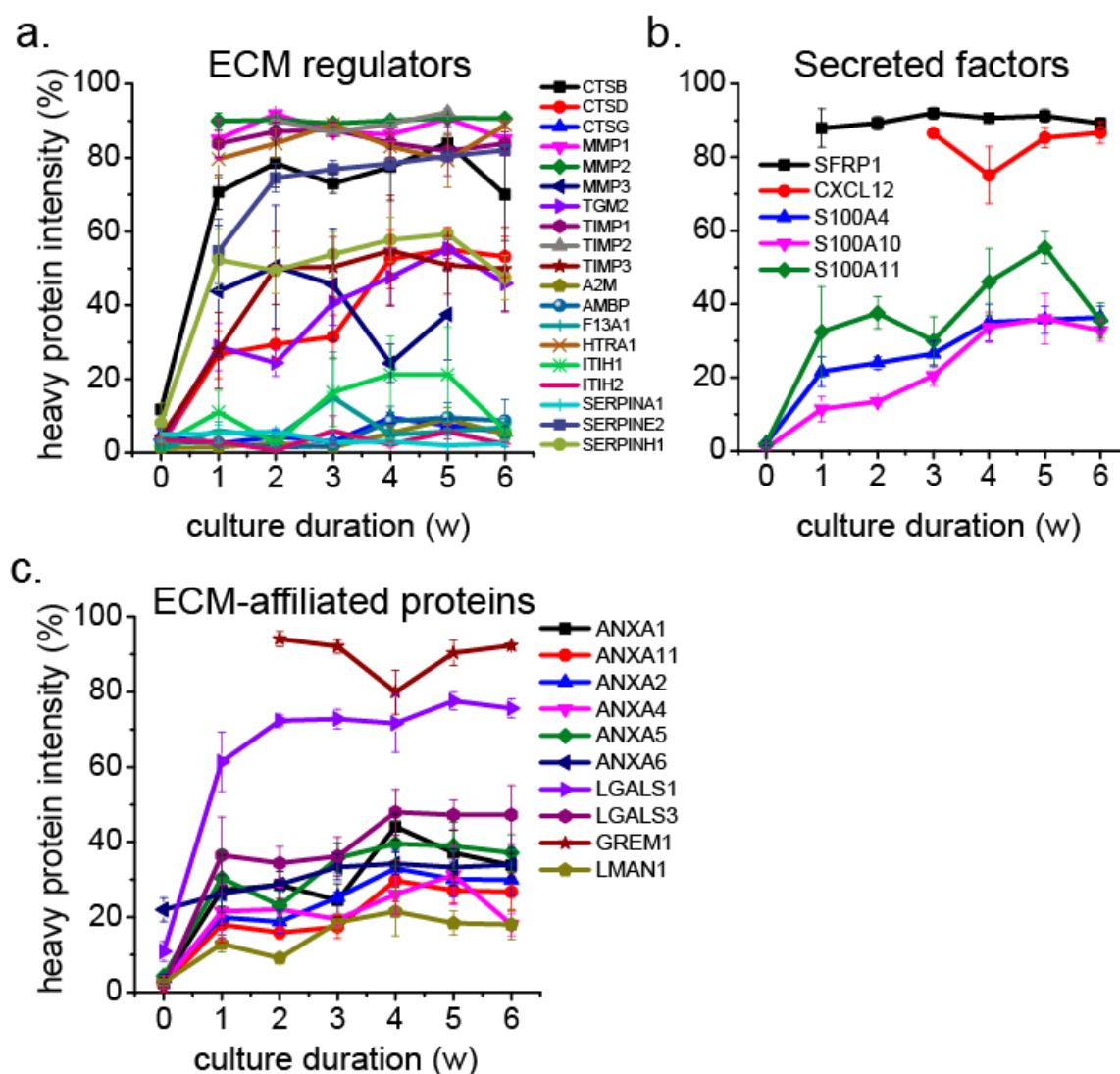


Figure 3.10. Protein turnover for specific proteins from the matrisome-associated subcategories: a) ECM regulators, b) secreted factors, and c) ECM-affiliated proteins. (Note: this turnover is plotted as the percentage of heavy intensity out of the total intensity for each individual protein). n = 4 biological replicates, with 2 technical replicates for each. Error bars, s.e.m.

3.6 References

1. Cortiella, J. et al. Influence of acellular natural lung matrix on murine embryonic stem cell differentiation and tissue formation. *Tissue engineering. Part A* 16, 2565-2580 (2010).
2. Ross, E.A. et al. Embryonic stem cells proliferate and differentiate when seeded into kidney scaffolds. *Journal of the American Society of Nephrology : JASN* 20, 2338-2347 (2009).
3. Sellaro, T.L. et al. Maintenance of human hepatocyte function in vitro by liver-derived extracellular matrix gels. *Tissue engineering. Part A* 16, 1075-1082 (2010).
4. DeQuach, J.A., Yuan, S.H., Goldstein, L.S. & Christman, K.L. Decellularized porcine brain matrix for cell culture and tissue engineering scaffolds. *Tissue engineering. Part A* 17, 2583-2592 (2011).
5. Badylak, S.F., Taylor, D. & Uygun, K. Whole-organ tissue engineering: decellularization and recellularization of three-dimensional matrix scaffolds. *Annual review of biomedical engineering* 13, 27-53 (2011).
6. Weber, K.T. et al. Collagen remodeling of the pressure-overloaded, hypertrophied nonhuman primate myocardium. *Circulation research* 62, 757-765 (1988).
7. Cleutjens, J.P., Verluyten, M.J., Smiths, J.F. & Daemen, M.J. Collagen remodeling after myocardial infarction in the rat heart. *The American journal of pathology* 147, 325-338 (1995).
8. Legrand, C. et al. Airway epithelial cell migration dynamics. MMP-9 role in cell-extracellular matrix remodeling. *The Journal of cell biology* 146, 517-529 (1999).
9. Godin, D., Ivan, E., Johnson, C., Magid, R. & Galis, Z.S. Remodeling of carotid artery is associated with increased expression of matrix metalloproteinases in mouse blood flow cessation model. *Circulation* 102, 2861-2866 (2000).
10. Naba, A. et al. The matrisome: in silico definition and in vivo characterization by proteomics of normal and tumor extracellular matrices. *Mol Cell Proteomics* 11, M111 014647 (2012).
11. Ong, S.E. et al. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* 1, 376-386 (2002).
12. Ong, S.E. & Mann, M. A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC). *Nature protocols* 1, 2650-2660 (2006).
13. Doherty, M.K., Hammond, D.E., Clague, M.J., Gaskell, S.J. & Beynon, R.J. Turnover of the human proteome: determination of protein intracellular stability by dynamic SILAC. *J Proteome Res* 8, 104-112 (2009).
14. Doherty, M.K. & Beynon, R.J. Protein turnover on the scale of the proteome. *Expert review of proteomics* 3, 97-110 (2006).
15. Piques, M. et al. Ribosome and transcript copy numbers, polysome occupancy and enzyme dynamics in Arabidopsis. *Molecular systems biology* 5, 314 (2009).
16. Cambridge, S.B. et al. Systems-wide proteomic analysis in mammalian cells reveals conserved, functional protein turnover. *J Proteome Res* 10, 5275-5284 (2011).

17. Xu, C.C., Chan, R.W. & Tirunagari, N. A biodegradable, acellular xenogeneic scaffold for regeneration of the vocal fold lamina propria. *Tissue engineering* 13, 551-566 (2007).
18. Chen, X. & Thibeault, S.L. Novel isolation and biochemical characterization of immortalized fibroblasts for tissue engineering vocal fold lamina propria. *Tissue Eng Part C Methods* 15, 201-212 (2009).
19. Van Hoof, D. et al. An experimental correction for arginine-to-proline conversion artifacts in SILAC-based quantitative proteomics. *Nature methods* 4, 677-678 (2007).
20. Bendall, S.C. et al. Prevention of amino acid conversion in SILAC experiments with embryonic stem cells. *Mol Cell Proteomics* 7, 1587-1597 (2008).
21. Yee, J.C. et al. Global assessment of protein turnover in recombinant antibody producing myeloma cells. *Journal of biotechnology* 148, 182-193 (2010).
22. Mackenzie, I.J. Vocal Cord Physiology: Contemporary Research and Clinical Issues. *Journal of Anatomy* 139, 176-176 (1984).
23. Wisniewski, J.R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. *Nature methods* 6, 359-362 (2009).
24. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature biotechnology* 26, 1367-1372 (2008).
25. Vizcaino, J.A. et al. ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nature biotechnology* 32, 223-226 (2014).
26. Maere, S., Heymans, K. & Kuiper, M. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics (Oxford, England)* 21, 3448-3449 (2005).
27. Supek, F., Bosnjak, M., Skunca, N. & Smuc, T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* 6, e21800 (2011).
28. Cline, M.S. et al. Integration of biological networks and gene expression data using Cytoscape. *Nature protocols* 2, 2366-2382 (2007).
29. Titze, I.R. The physics of small-amplitude oscillation of the vocal folds. *The Journal of the Acoustical Society of America* 83, 1536-1552 (1988).
30. Hahn, M.S., Jao, C.Y., Faquin, W. & Grande-Allen, K.J. Glycosaminoglycan composition of the vocal fold lamina propria in relation to function. *The Annals of otology, rhinology, and laryngology* 117, 371-381 (2008).
31. Uygun, B.E. et al. Organ reengineering through development of a transplantable recellularized liver graft using decellularized liver matrix. *Nature medicine* 16, 814-820 (2010).
32. Thevenot, P., Nair, A., Dey, J., Yang, J. & Tang, L. Method to analyze three-dimensional cell distribution and infiltration in degradable scaffolds. *Tissue Eng Part C Methods* 14, 319-331 (2008).
33. Minehara, H. et al. A new technique for seeding chondrocytes onto solvent-preserved human meniscus using the chemokinetic effect of recombinant human bone morphogenetic protein-2. *Cell and tissue banking* 12, 199-207 (2011).
34. Martinello, T. et al. Successful recellularization of human tendon scaffolds using adipose-derived mesenchymal stem cells and collagen gel. *Journal of tissue engineering and regenerative medicine* 8, 612-619 (2014).

35. Catten, M., Gray, S.D., Hammond, T.H., Zhou, R. & Hammond, E. Analysis of cellular location and concentration in vocal fold lamina propria. *Otolaryngology--head and neck surgery : official journal of American Academy of Otolaryngology-Head and Neck Surgery* 118, 663-667 (1998).
36. Welham, N.V., Chang, Z., Smith, L.M. & Frey, B.L. Proteomic analysis of a decellularized human vocal fold mucosa scaffold using 2D electrophoresis and high-resolution mass spectrometry. *Biomaterials* 34, 669-676 (2013).
37. Li, Q. et al. Proteomic analysis of naturally-sourced biological scaffolds. *Biomaterials* 75, 37-46 (2016).
38. Hausser, H., Ober, B., Quentin-Hoffmann, E., Schmidt, B. & Kresse, H. Endocytosis of different members of the small chondroitin/dermatan sulfate proteoglycan family. *The Journal of biological chemistry* 267, 11559-11564 (1992).
39. Shapiro, S.D. Matrix metalloproteinase degradation of extracellular matrix: biological consequences. *Current opinion in cell biology* 10, 602-608 (1998).
40. Knowles, G.C., McKeown, M., Sodek, J. & McCulloch, C.A. Mechanism of collagen phagocytosis by human gingival fibroblasts: importance of collagen structure in cell recognition and internalization. *J Cell Sci* 98 (Pt 4), 551-558 (1991).
41. Dupuy, A.G. & Caron, E. Integrin-dependent phagocytosis: spreading from microadhesion to new concepts. *J Cell Sci* 121, 1773-1783 (2008).
42. Xu, L. & Qu, Z. Roles of protein ubiquitination and degradation kinetics in biological oscillations. *PLoS One* 7, e34616 (2012).
43. Sottile, J. & Hocking, D.C. Fibronectin polymerization regulates the composition and stability of extracellular matrix fibrils and cell-matrix adhesions. *Molecular biology of the cell* 13, 3546-3559 (2002).
44. Liu, X.J. et al. Lumican Accelerates Wound Healing by Enhancing alpha2beta1 Integrin-Mediated Fibroblast Contractility. *PLoS One* 8, e67124 (2013).
45. Tufvesson, E. & Westergren-Thorsson, G. Biglycan and decorin induce morphological and cytoskeletal changes involving signalling by the small GTPases RhoA and Rac1 resulting in lung fibroblast migration. *J Cell Sci* 116, 4857-4864 (2003).
46. Ott, H.C. et al. Perfusion-decellularized matrix: using nature's platform to engineer a bioartificial heart. *Nature medicine* 14, 213-221 (2008).
47. Brown, B.N., Valentin, J.E., Stewart-Akers, A.M., McCabe, G.P. & Badylak, S.F. Macrophage phenotype and remodeling outcomes in response to biologic scaffolds with and without a cellular component. *Biomaterials* 30, 1482-1491 (2009).
48. Badylak, S.F. & Gilbert, T.W. Immune response to biologic scaffold materials. *Seminars in immunology* 20, 109-116 (2008).
49. O'Brien, F.J. Biomaterials & scaffolds for tissue engineering. *Materials Today* 14, 88-95 (2011).
50. Zanivan, S., Krueger, M. & Mann, M. In vivo quantitative proteomics: the SILAC mouse. *Methods in molecular biology (Clifton, N.J.)* 757, 435-450 (2012).
51. Spence, J.R. et al. Directed differentiation of human pluripotent stem cells into intestinal tissue in vitro. *Nature* 470, 105-109 (2011).
52. Ridky, T.W., Chow, J.M., Wong, D.J. & Khavari, P.A. Invasive three-dimensional organotypic neoplasia from multiple normal human epithelia. *Nature medicine* 16, 1450-1455 (2010).

Chapter 4

Characterizing Bioengineered Vocal Fold Mucosa

Adapted from:

“Bioengineered vocal fold mucosa for functional voice restoration.” C. Ling, **Q. Li**, M. E. Brown, Y. Kishimoto, Y. Toya, E. E. Devine, K. Choi, K. Nishimoto, I. G. Norman, T. Tsegay, J. J. Jiang, W. J. Burlingham, S. Gunasekaran, L. M. Smith, B. L. Frey, N. V. Welham. *Sci. Trans. Med.* 2015; 7(314):314ra187.

4.1 Introduction

Chapter 2 and 3 presented studies on the decellularization and recellularization processes of tissue engineering, respectively. In this chapter, a vocal fold mucosa was engineered and studied.

Vocal folds (VF) are folds of tissue located within the larynx at the top of the trachea and have three important functions: to vibrate the air coming up from the lungs to form sound, to prevent choking on material in the throat, and to control the amount of air sent to and released from the lungs. Therefore, impaired vocal folds can have significant effects on individual health and social function.

The human VF is a laminated structure composed of epithelium, basement membrane, lamina propria and thyroarythnoid muscle (Fig. 4.1)¹. VF epithelial (VFE) cells, together with basement membrane, act as a barrier to the entry of inhaled and systemic challenges². They are also of great significance in water/ion transport for maintaining VF hydration, which is critical to VF oscillation³. Furthermore, because the VFE cells represent the most superficial layer of the VF mucosa (VFM), they play an important role in local immune response by signaling professional immune cells via expression of major histocompatibility complex (MHC) class I and II proteins⁴. The lamina propria is occupied by ECM and is primarily consisted of reticular, collagenous and elastic fibers. These fibers serve as scaffolds for the maintenance of structure, strength and resilience. The most abundant cell type in the lamina propria is the vocal fold fibroblast (VFF). These cells are largely evenly spread throughout the depth of the tissue. During a wound healing response, vocal fold fibroblasts migrate to the site of injury and remodel the tissue by generating ECM. The epithelium and superficial lamina propria form the VFM, which

vibrates during phonation and is the most common site for injury. Therefore, the target of most bioengineered constructs for the larynx is the vocal fold mucosa.

Voice impairment (dysphonia) affects an estimated 20 million people in the United States, resulting in reduced general and disease-specific quality of life⁵, reduced occupational performance and attendance⁶, and direct health care costs exceeding \$11 billion per year⁷. Between 60 and 80% of voice complaints in the treatment-seeking population involve changes to VFM⁸; severe mucosal impairment or loss due to trauma, disease, or disease resection often culminates in fibrosis and deterioration of VF vibratory capacity for voice⁹. Patients with significant VFM damage have limited treatment options. Medialization of the impaired VF, achieved by delivering an implant or injectate to the paraglottic space¹⁰, can improve VF closure and therefore voice, but does not address fibrotic changes within the ECM. Superficial injection of regenerative biomaterials offers an alternative means to improve VF viscoelasticity and vibratory function¹¹; however, most biomaterials are not specifically engineered for the VF biomechanical environment, have limited residence time, and are not suited for large deficits involving extensive tissue loss. Creation of an organotypic bioengineered VF mucosa could theoretically bypass these challenges by providing on-demand tissue for transplantation that is both biomechanically appropriate for use as a dynamic sound source for voice production and capable of maintaining barrier function at the boundary of the upper and lower airways.

Tissue engineering of partial and complete VF mucosae has been attempted using decellularized ECM-based¹² and collagen and fibrin¹³ gel-based scaffolds, seeded with embryonic stem cell derivatives, adult stem cells, and terminally differentiated cells. These organotypic culture approaches have generated engineered mucosae with desirable histologic features; however, to date, there is no benchmark culture system based solely on human-sourced

VF primary cells against which stem cell-based approaches can be evaluated, no direct comparisons showing equivalency with native human VF mucosa, and most importantly, limited progress towards the restoration of physiologic function. Significant advances have been hampered by the near-unavailability of disease-free primary human VF mucosal cells¹⁴ and limited attention to the intricate protein- and anatomic substructure-level complexity that characterizes mucosal morphogenesis.

The primary objective of this study was to engineer human-sourced VF mucosae with functional performance comparable to that of native tissue. Initial cell isolation and characterization experiments were performed using primary cells isolated from cadavers and surgical patients. The engineering strategy involved organotypic coculture of primary VFF and VFE with type I collagen gel as scaffold. Proteomic analysis was performed comparing scaffold only, VFF in scaffold, and VFE on scaffold. The engineered VFM was also compared with the native tissue.

4.2 Materials and Methods

3D organotypic culture—Purified rat tail collagen, type I (BD Biosciences) was prepared to a final concentration of 2.4 mg.mL^{-1} according to the manufacturer's instructions and seeded with $2 \times 10^5 \text{ VFF.mL}^{-1}$. The cell-scaffold mixture was added to the apical chamber of a culture insert (0.14 mL per well of a 24-well insert with $0.4 \text{ }\mu\text{m}$ pore size, 0.64 cm diameter, 0.3 cm^2 surface area; or 2.0 mL per well of a 6-well insert with $0.4 \text{ }\mu\text{m}$ pore size, 2.31 cm diameter, 4.2 cm^2 surface area; BD Biosciences). The collagen-based scaffold was then polymerized at $37 \text{ }^\circ\text{C}$ for 40 min. Fibroblast-oriented medium was added to both apical and basolateral chambers and

the cells were cultured for 24 h. Next, VFE were seeded on the polymerized scaffold surface within the apical chamber (2×10^5 VFE per well of a 24-well insert; 2×10^6 VFE per well of a 6-well insert) and cultured in epithelial cell-oriented medium; a 1:1 ratio of fibroblast- and epithelial cell-oriented media was added to the basolateral chamber. After an additional 48 h, the epithelial cell-oriented medium was aspirated from the apical chamber, leaving VFE at the air-liquid interface. We continued organotypic culture for a total of 8-28 d with basolateral chamber media change every 48 h. Schematic illustrating the experimental approach is shown in Fig. 4.2. As initial histological assessment showed no difference in engineered VF mucosa morphology at 14 and 28 d, we performed all subsequent proteomic assays on samples harvested at 14 d. Experimental comparisons involving the scaffold only, VFF in scaffold and VFE on scaffold involved identical culture conditions for the entire 14 d period.

Protein extraction and digestion—Proteins were extracted from each sample ($n = 3$ biological replicates per condition) by first adding 150 μ L of 4% SDS, 0.1 M Tris-HCl (pH 7.6) and 0.1 M dithiothreitol. Next, samples were sonicated (alternating 20 s on/off cycles for 6 min) using a probe sonicator (XL2015 with PN/418 microtip; Misonix), heated to 95 °C for 7 min, and then centrifuged at 16,100 g at 20 °C for 5 min. A 30 μ L aliquot of the supernatant was processed according to the filter-aided sample preparation (FASP) protocol for SDS removal and on-filter digestion. Briefly, the supernatant was added to a 30,000 MWCO Vivacon 500 filter (Sartorius), washed, alkylated, and digested with trypsin (50:1 w/w protein-to-trypsin ratio) at 37 °C overnight. The digested sample was desalted using a Sep-Pak C18 1 cc Vac cartridge (Waters), evaporated to dryness in a vacuum centrifuge, and reconstituted in 5% acetonitrile and 2% formic acid in water.

MS parameters—The following MS experiment was performed using two technical replicates per biological replicate. Approximately 1.2 µg of the protein digest, as estimated by a BCA assay (Pierce), was injected into a Waters nanoAcquity HPLC coupled to an ESI iontrap/orbitrap mass spectrometer (LTQ Orbitrap Velos; Thermo Scientific). Peptides were separated on a 100 µm-inner-diameter column packed with 20 cm of 3 µm MAGIC aqC18 beads (Bruker-Michrom), which were packed against an in-house laser-pulled tip, and eluted at 0.3 µL.min⁻¹ in 0.1% formic acid with a gradient of increasing acetonitrile, over 2.5 h. A full-mass scan (300-1500 m/z) was performed in the orbitrap at a resolution of 60,000. The ten most intense peaks were then selected for fragmentation by higher energy collisional dissociation (HCD) at 42% collision energy, with a resolution of 7500 and isolation width of 2.5 m/z. Dynamic exclusion was enabled with a repeat count of 2 over 30 s and an exclusion duration of 120 s.

MS data processing—We searched the mass spectra against appropriate organism protein databases (*Homo sapiens* and *Rattus norvegicus*; UniProt) using the SEQUEST algorithm within Proteome Discoverer (Thermo Scientific). We allowed two missed cleavages, required at least two unique peptides per protein identification, and filtered the results using a 1% peptide false discovery rate. Precursor mass tolerance was set to 25 ppm and 0.05 Da for fragment ion tolerance. Variable methionine and proline oxidation (+15.995 Da) and static carbamidomethylation of cysteines (+57.021 Da) were also used. Spectral counting-based protein quantification was performed using the normalized spectral abundance factor (NSAF) approach¹⁵. Further normalization, based on an assumption of comparable degradation of rat collagen across experimental conditions, was performed using a correction factor calculated from the NSAF values of the 10 most abundant proteins in the collagen, type I-based scaffold.

Statistical analyses—Technical replicates were averaged and all statistical comparisons were performed using independent biological replicates. NSAF-based quantitative proteomic data were analyzed using a two-tailed Student's *t* test with implementation of Benjamini-Hochberg correction¹⁶ to account for multiple testing. Gene ontology term enrichment analysis was performed using the BiNGO¹⁷ (hypergeometric model with Benjamini-Hochberg correction) and REVIGO¹⁸ (SimRel cutoff = 0.4) algorithms. Ontology term enrichment schematics were generated using Cytoscape 2.8.2¹⁹. An initial (pre-correction) type I error rate of 0.01 was used for all statistical testing; quantitative proteomic data were subject to an additional fold change cutoff of 4. All *P*-values were two-sided.

4.3 Results and Discussion

Comparing the engineered with the native VFM—We conducted discovery proteomic analysis of the engineered VFM using LC-MS/MS to characterize its biological complexity. Using a 1% false discovery rate, we identified 762 unique proteins in the engineered mucosa, compared with 908 in the native mucosa and 32 in the collagen, type I-based scaffold (Fig. 4.3). Gene ontology-based enrichment analysis of the engineered VF mucosa proteome, as compared to the full human protein database (Uniprot), revealed a wide complement of functional protein sets (ontology terms) associated with various metabolic, catalytic, transport, binding and signaling processes; spanning an array of subcellular and extracellular locations (Fig. 4.4). We identified organogenesis/morphogenesis-specific ontology term enrichment, consistent with successful organotypic culture, as well as ECM terms indicative of protein complexes and anatomic substructures that are characteristic of native VF mucosa, such as the basal lamina,

anchoring collagen and fibrillar collagen. Normalized spectral abundance factor (NSAF)-based quantitative analysis showed that ECM protein abundances in the engineered mucosa were more similar to the native mucosa than the scaffold [z score shift, 0.46 ± 1.35 versus 3.65 ± 1.68 (means $\pm$ SD, $n = 76$ proteins); $P < 0.0001$, Student's t test] (Fig. 4.5), confirming that the engineered mucosa ECM is mostly attributable to new protein synthesis by VFF and VFE.

We identified ECM proteins and glycoconjugates that are considered critical to the biomechanical function of native VF mucosa^{1, 20}, including: multiple collagen isoforms; the elastin conduit fibrillin 1 and elastin microfibril interface-located protein (EMILIN) 1; the small leucine-rich repeat proteoglycans decorin, lumican and biglycan; and the glycoproteins fibronectin, fibulin 1 and tenascin X. These observations suggest that while the engineered lamina propria appeared immature on histologic analysis (data not shown), its developing ECM is populated by a core set of protein constituents with the potential to support vibratory function. This conclusion was further supported by rheologic experiments showing that the viscoelastic profile of the engineered mucosa was more similar to that of the native mucosa than that of the scaffold (data not shown).

Comparing the engineered VFM with its isolated subcomponents—We pursued further quantitative proteomic analyses to identify protein complexes and functionality that were unique to the engineered VF mucosa, compared to its isolated subcomponents: VFF in collagen, type I scaffold and VFE on collagen, type I scaffold. Proteome coverage and overlap were generally comparable, with 528 proteins common to all three conditions and 59 proteins unique to the engineered mucosa (Fig. 4.6A). NSAF-based analysis showed that the majority of differentially abundant proteins were overrepresented in engineered VF mucosa compared to the other experimental conditions (fold change > 4 ; adjusted $P < 0.01$; Fig. 4.6B), suggesting upregulation

of a specific protein set due to VFF-VFE synergy in organotypic culture. Subsequent pairwise comparisons indicated that the engineered mucosa proteome was most similar to VFE on scaffold and least similar to VFF in scaffold (69 versus 139 differentially abundant proteins, respectively) (Fig. 4.7), consistent with VFE being the predominant cell type in the engineered mucosa.

Next, we performed functional enrichment analysis of the protein set that was either exclusively identified or quantitatively overrepresented in engineered VF mucosa compared to both VFF in scaffold and VFE on scaffold. The most highly represented biological process ontology terms (adjusted $P < 4 \times 10^{-5}$) indicated that the engineered mucosa was uniquely engaged in macromolecule catabolism, protein localization, and cellular component organization or biogenesis (Fig. 4.8), suggesting dynamic structural assembly. Further evaluation of the deep ontology output revealed significant enrichment of additional terms relevant to mucosal assembly: cell-substrate junction assembly (adjusted $P = 3 \times 10^{-3}$), epidermis development (adjusted $P = 5.3 \times 10^{-3}$), adherens junction organization (adjusted $P = 7.4 \times 10^{-3}$) and cell junction assembly (adjusted $P = 8.7 \times 10^{-3}$). Enrichment of these ontology terms was driven by a common protein set that was significantly overrepresented in engineered VF mucosa compared to the other experimental conditions, consisting of the basal lamina constituent laminin 5, the basal epithelial cell marker keratin 5, the desmosome constituents junction plakoglobin (also known as γ -catenin) and periplakin, as well as collagen, type V.

4.4 Figures and Tables

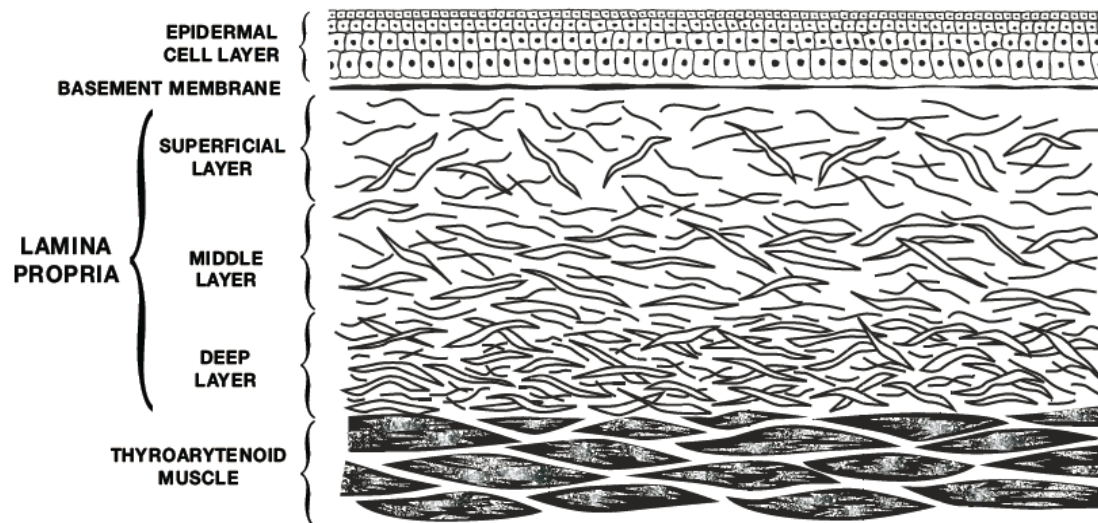


Figure 4.1. Diagram of the layered structure of the vocal fold. Reprinted from ref 1, with permission from SAGE Publications.

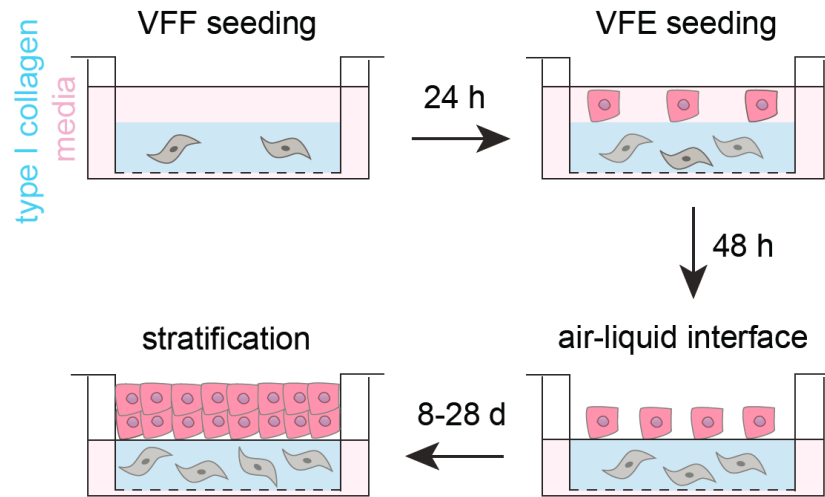


Figure 4.2. Assembly of engineered VF mucosa.

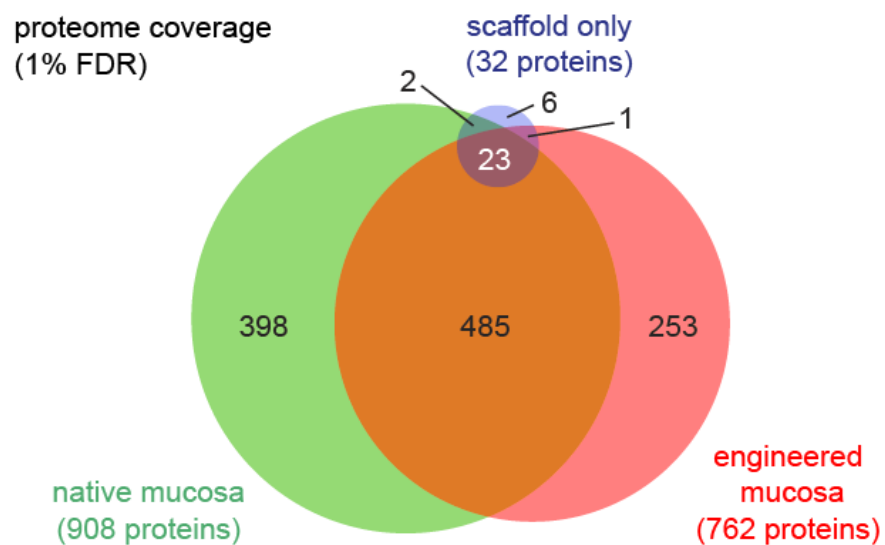


Figure 4.3. Venn diagram summarizing proteome coverage in engineered mucosa compared to native mucosa and scaffold only. FDR, false discovery rate.

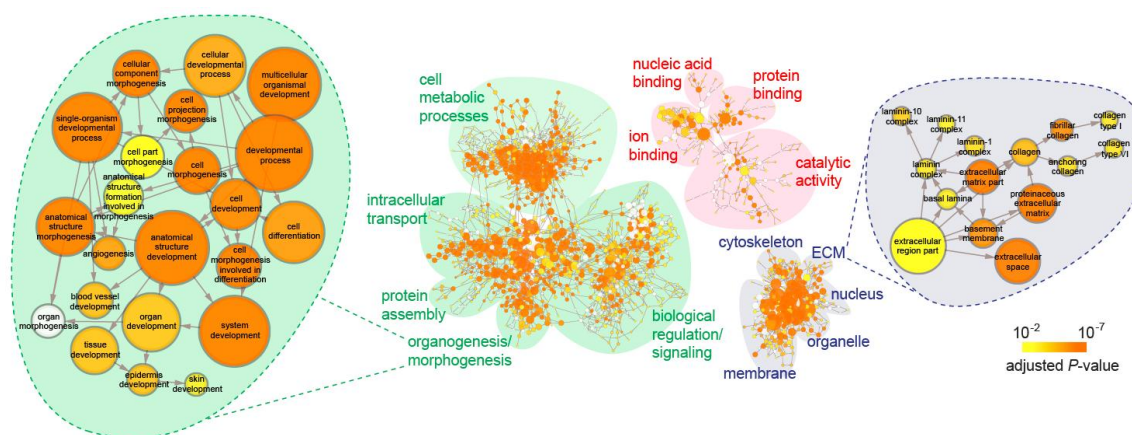


Figure 4.4. Functional enrichment analysis of the engineered mucosa proteome. Enriched gene ontology terms are depicted as nodes connected by arrows that represent hierarchies and relationships between terms. Node size is proportional to the number of proteins assigned to a given term; node color represents the Benjamini Hochberg-corrected P-value corresponding to enrichment of the term. Functionally related ontology terms are labeled and grouped using colored ovals (biological process terms in green; molecular function terms in red; cellular component terms in blue). Organogenesis/morphogenesis and ECM functional groups are enlarged for better visualization of individual terms.

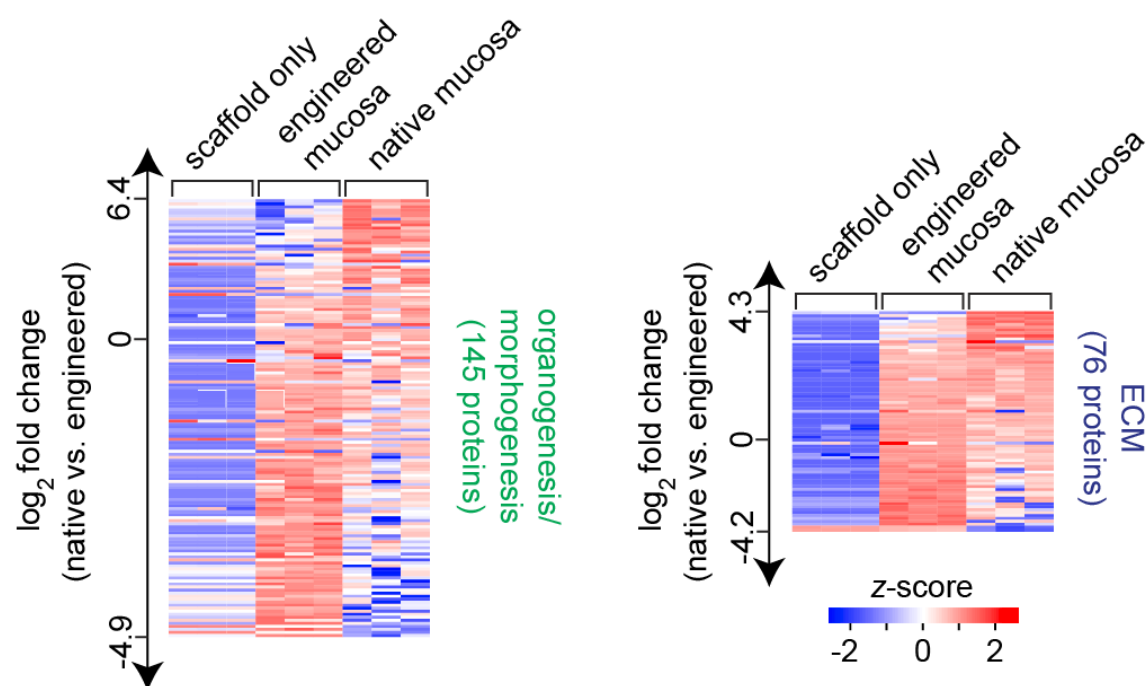


Figure 4.5. Heatmaps summarizing normalized spectral abundance factor-based quantification of proteins associated with the ontology terms highlighted in Fig 3.2.

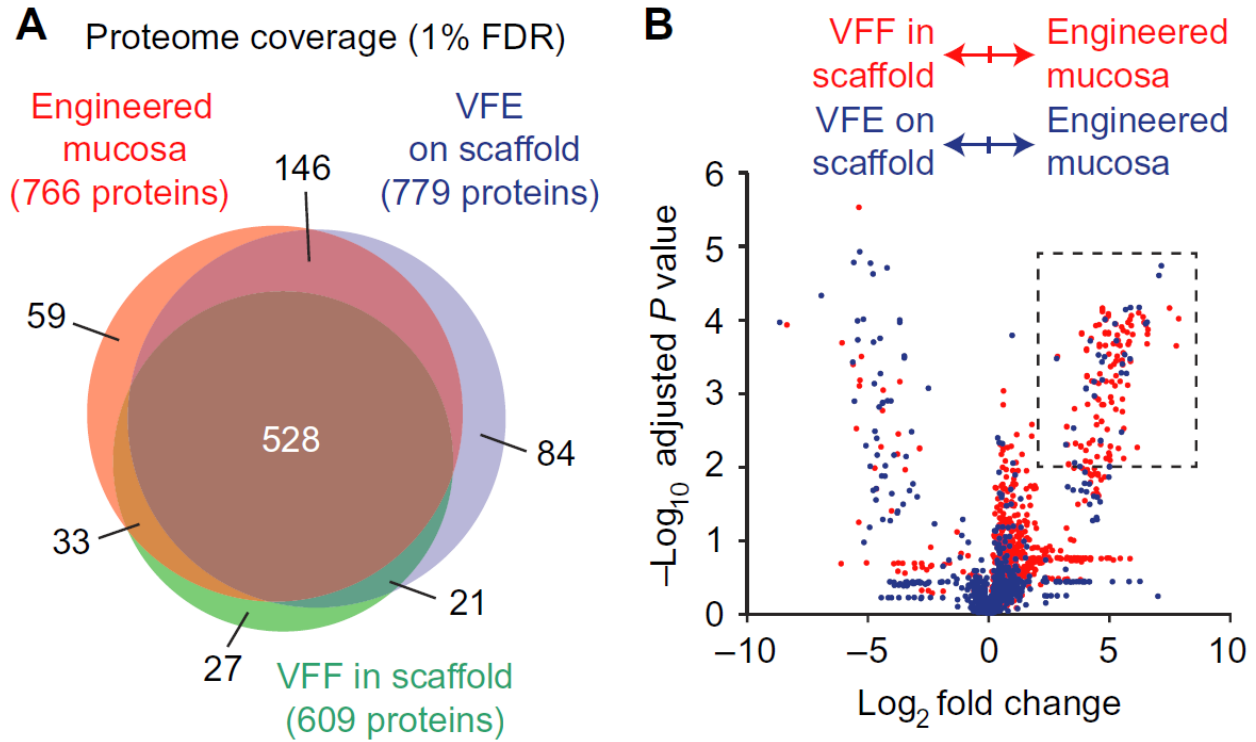


Figure 4.6. Proteomic-based analysis of engineered VF mucosa compared to its isolated subcomponents. (A) Venn diagram summarizing proteome coverage and overlap in protein identifications across experimental conditions. FDR, false discovery rate. (B) Volcano plot summarizing NSAF-based protein quantification in engineered mucosa versus VFF in scaffold (red) and VFE on scaffold (blue). The dashed rectangle denotes cutoff criteria for protein overrepresentation in engineered mucosa compared to the other conditions. Adjusted P values were calculated using a Student's t test ($n = 3$).

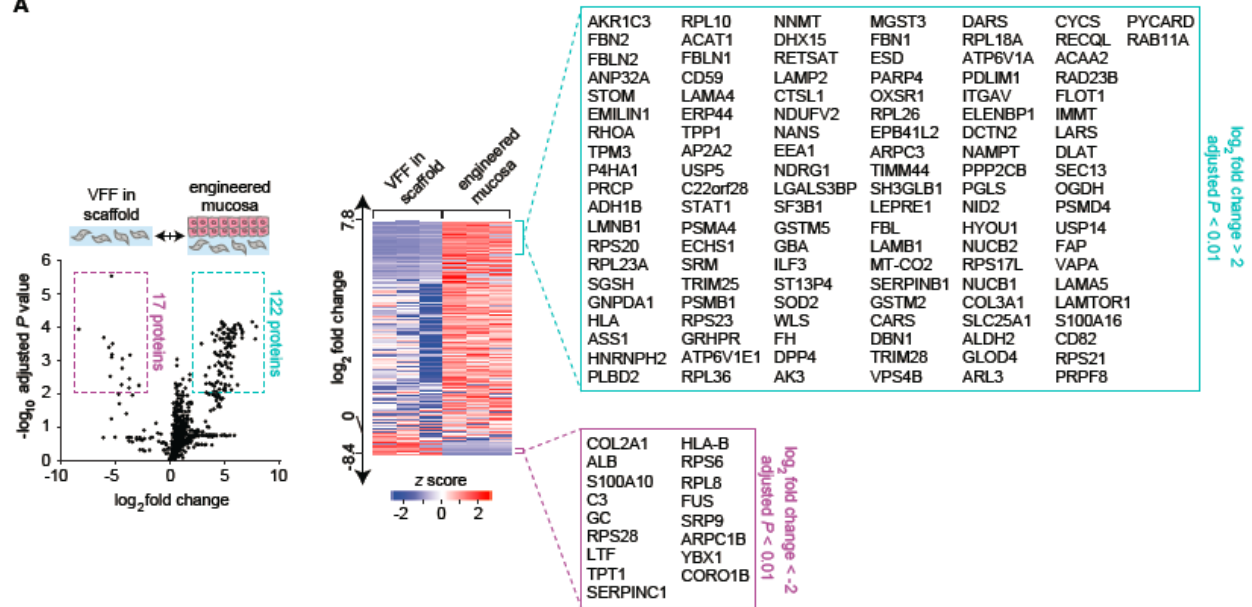
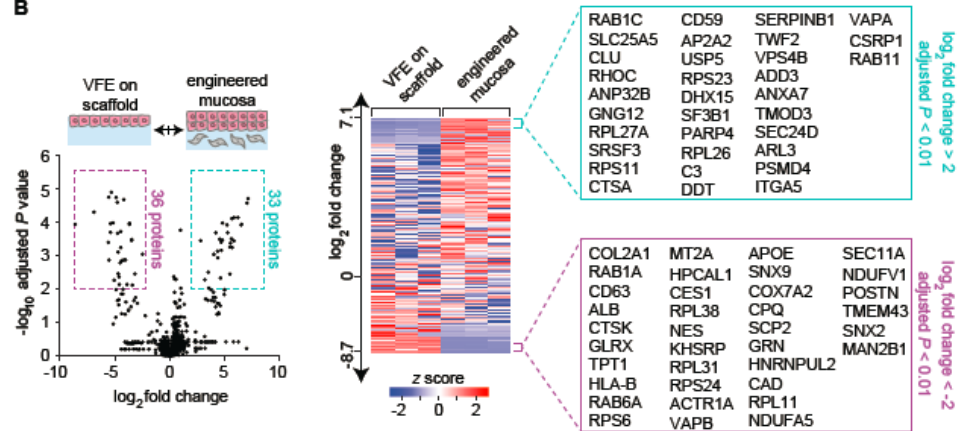
A**B**

Figure 4.7. Additional quantitative proteomic analysis of engineered VF mucosa compared to its isolated subcomponents. (A) Volcano plot and heatmap showing NSAF-based protein quantification in engineered mucosa versus VFF in scaffold. The cyan dashed rectangle (Volcano plot) and square bracket (heatmap) denote cutoff criteria for protein overrepresentation in engineered mucosa compared to VFF in scaffold; the magenta dashed rectangle (Volcano plot) and square bracket (heatmap) denote cutoff criteria for protein overrepresentation in VFF in scaffold compared to engineered mucosa. (B) Parallel analysis of engineered mucosa versus VFE on scaffold. The cyan dashed rectangle (Volcano plot) and square bracket (heatmap) denote cutoff criteria for protein overrepresentation in engineered mucosa compared to VFE on scaffold; the magenta dashed rectangle (Volcano plot) and square bracket (heatmap) denote cutoff criteria for protein overrepresentation in VFE on scaffold compared to engineered mucosa. Adjusted P values were calculated using a Student's t test ($n = 3$).

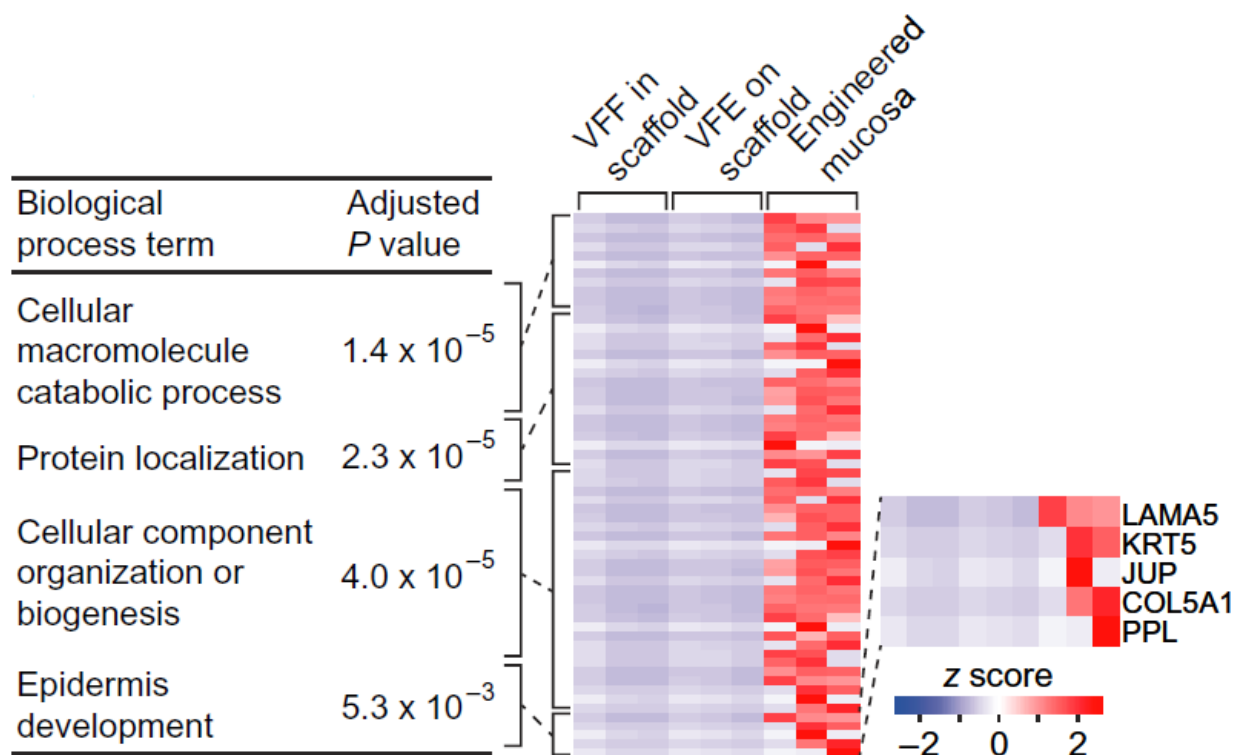


Figure 4.8. Summary of enriched biological process terms associated with the protein set exclusive to engineered mucosa or overrepresented in engineered mucosa compared to both VFF in scaffold and VFE on scaffold. The table lists the three most highly represented terms (adjusted *P* values were calculated using BiNGO, $n = 3$; postprocessing was performed using REVIGO), as well as the mechanistically relevant epidermis (in the context of mucosa, epithelium) development term. The heat map shows the relative abundance of overrepresented proteins that map to these terms of interest.

4.5 References

1. Gray, S.D., Titze, I.R., Alipour, F. & Hammond, T.H. Biomechanical and histologic observations of vocal fold fibrous proteins. *The Annals of otology, rhinology, and laryngology* **109**, 77-85 (2000).
2. Erickson, E. & Sivasankar, M. Simulated reflux decreases vocal fold epithelial barrier resistance. *The Laryngoscope* **120**, 1569-1575 (2010).
3. Lodewyck, D., Menco, B. & Fisher, K. Immunolocalization of aquaporins in vocal fold epithelia. *Archives of Otolaryngology-Head & Neck Surgery* **133**, 557-563 (2007).
4. Barker, E., Haverson, K., Stokes, C.R., Birchall, M. & Bailey, M. The larynx as an immunological organ: immunological architecture in the pig as a large animal model. *Clinical and experimental immunology* **143**, 6-14 (2006).
5. Cohen, S.M., Dupont, W.D. & Courey, M.S. Quality-of-life impact of non-neoplastic voice disorders: a meta-analysis. *The Annals of otology, rhinology, and laryngology* **115**, 128-134 (2006).
6. Cohen, S.M., Kim, J., Roy, N., Asche, C. & Courey, M. The impact of laryngeal disorders on work-related dysfunction. *The Laryngoscope* **122**, 1589-1594 (2012).
7. Cohen, S.M., Kim, J., Roy, N., Asche, C. & Courey, M. Direct health care costs of laryngeal diseases and disorders. *The Laryngoscope* **122**, 1582-1588 (2012).
8. Coyle, S.M., Weinrich, B.D. & Stemple, J.C. Shifts in relative prevalence of laryngeal pathology in a treatment-seeking population. *Journal of voice : official journal of the Voice Foundation* **15**, 424-440 (2001).
9. Benninger, M.S. et al. Vocal fold scarring: current concepts and management. *Otolaryngology--head and neck surgery : official journal of American Academy of Otolaryngology-Head and Neck Surgery* **115**, 474-482 (1996).
10. Ford, C.N., Bless, D.M. & Loftus, J.M. Role of injectable collagen in the treatment of glottic insufficiency: a study of 119 patients. *The Annals of otology, rhinology, and laryngology* **101**, 237-247 (1992).
11. Caton, T., Thibeault, S.L., Klemuk, S. & Smith, M.E. Viscoelasticity of hyaluronan and nonhyaluronan based vocal fold injectables: implications for mucosal versus muscle use. *The Laryngoscope* **117**, 516-521 (2007).
12. Xu, C.C., Chan, R.W. & Tirunagari, N. A biodegradable, acellular xenogeneic scaffold for regeneration of the vocal fold lamina propria. *Tissue engineering* **13**, 551-566 (2007).
13. Long, J.L. et al. Functional testing of a tissue-engineered vocal fold cover replacement. *Otolaryngology--head and neck surgery : official journal of American Academy of Otolaryngology-Head and Neck Surgery* **142**, 438-440 (2010).
14. Chen, X. & Thibeault, S.L. Novel isolation and biochemical characterization of immortalized fibroblasts for tissue engineering vocal fold lamina propria. *Tissue Eng Part C Methods* **15**, 201-212 (2009).
15. Zybaylov, B. et al. Statistical analysis of membrane proteome expression changes in *Saccharomyces cerevisiae*. *J Proteome Res* **5**, 2339-2347 (2006).
16. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* **57**, 289-300 (1995).

17. Maere, S., Heymans, K. & Kuiper, M. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics (Oxford, England)* **21**, 3448-3449 (2005).
18. Supek, F., Bosnjak, M., Skunca, N. & Smuc, T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* **6**, e21800 (2011).
19. Cline, M.S. et al. Integration of biological networks and gene expression data using Cytoscape. *Nature protocols* **2**, 2366-2382 (2007).
20. Gray, S.D., Titze, I.R., Chan, R. & Hammond, T.H. Vocal fold proteoglycans and their influence on biomechanics. *The Laryngoscope* **109**, 845-854 (1999).

Section II

Chapter 5

Global Post-Translational Modification Discovery

Adapted from:

“Global Post-Translational Modification Discovery.” **Q. Li**, M. R. Shortreed; C. D. Wenger; B. L. Frey; M. Scalf; L. M. Smith. (*Submitted for publication.*)

5.1 Introduction

Protein post-translational modifications (PTMs) modulate critical biological processes such as protein signaling, localization, and degradation and have been implicated in a wide variety of pathologies. In spite of their importance, the comprehensive identification and discovery of PTMs in complex biological samples has continued to pose a difficult challenge for proteomics technologies¹.

We have recently described a global PTM (G-PTM) identification strategy which enables the rapid and confident identification of numerous PTM types in a single-pass database search^{2,3}. Identification is accomplished by searching for the presence or absence of PTMs exclusively at curated sites designated in the UniProt repository. However, as such lists of curated PTMs are at present quite incomplete, the G-PTM strategy necessarily misses many important PTMs. For example, hydroxyproline is known to be a prevalent PTM in type I collagen^{4,5}, but only four of many hydroxyproline sites in type I collagen are included in the human UniProt database. Furthermore, despite containing 470 PTM types as of May 2015 (<http://www.uniprot.org/docs/ptmlist>), the UniProt database is still missing many additional novel PTMs, chemical derivatives, and sample-specific amino acid variants, which thereby precludes the G-PTM strategy from identifying these modifications.

In order to search for both known and unknown PTM types at once, several unrestrictive PTM identification approaches, including wide mass tolerance searches, have been devised⁶⁻¹². However, these strategies suffer from limitations such as not being readily applicable to large proteomic datasets, requiring detection of both modified and unmodified forms of a peptide in the sample, or sacrificing either sensitivity or confidence of modified peptide detection.

We describe here a global PTM discovery (G-PTM-D) strategy that combines the ability to discover uncurated/unexpected modifications offered by an unrestrictive approach, with the high confidence afforded by the G-PTM approach for identifying curated PTMs. G-PTM-D searches for modifications only at amino acid residue positions corresponding to either curated PTMs from the UniProt repository or potential modifications discovered on specific peptides from an initial search using a wide mass tolerance. Limiting modifications to defined positions in this manner enables discovery of a large variety of PTMs while maintaining high confidence for all peptide and PTM identifications.

5.2 Experimental Procedures

Datasets

Three separate datasets with deep proteome coverage were used to evaluate the performance of the G-PTM-D strategy.

Human Jurkat Cell Lysate (with 28 peptide fractions). Sample preparation and MS analysis of human Jurkat cells were previously reported¹³. The 28 MS raw files consisting of 490,057 MS/MS scans are available via FTP from the PeptideAtlas data repository¹⁴ by accessing the following link: <http://www.peptideatlas.org/PASS/PASS00215>. The curated database was the Homo sapiens reference proteome from UniProt (downloaded on December 23, 2013), limited to those proteins with mRNA transcript abundances exceeding 0.1 Transcripts Per Million¹³. This dataset was used to illustrate the G-PTM-D workflow in detail, and to compare G-PTM-D with a variable phosphorylation search and a G-PTM search.

Four Common Human Cell Lysates (with 6 peptide fractions × 3 biological replicates for each cell lysate). Sample preparation and MS analyses of human HEK293, A549, HeLa, and K562 cell lysates were previously reported¹⁵. The four cell lines were randomly chosen from the eleven cell lines reported. The curated database was the Homo sapiens reference proteome from UniProt (downloaded on May 15, 2015). These datasets were analyzed with the streamlined G-PTM-D workflow (*i.e.* automatically adding 11 regular modification types in the supplemented databases, without construction of the ΔM histogram), and were used to demonstrate the wide applicability of G-PTM-D and its improved performance compared with the G-PTM strategy.

Matrigel (with 6 peptide fractions). Sample preparation and MS analysis of Matrigel were previously reported¹⁶. The MS raw files consisting of 107,162 MS/MS scans from 6 peptide fractions are available via FTP from the PeptideAtlas data repository by accessing the following link: <http://www.peptideatlas.org/PASS/PASS00557>. (The 6 raw files used are the ones with “Matrigel01” in the file name.) The curated database was the Mus musculus reference proteome from UniProt (downloaded on April 20, 2015). This dataset was used to compare the performance of G-PTM-D with pMatch and MODa.

G-PTM-D Search

Stage 1 – G-PTM search with ultra-wide precursor mass tolerance. A PTM-curated database in Extensible Markup Language (XML) format was first downloaded from <http://www.uniprot.org/proteomes/>. A Perl script *xml_trimming.pl* was first run to delete irrelevant information in the database and retain only the sequence and PTM information, in order to speed up the downstream search. The software program Morpheus¹⁷ (revision 149) (freely-available at <http://cwenger.github.io/Morpheus/>) utilized the trimmed UniProt XML database along with data files in either .raw or .mzML format. When an XML database is

specified in Morpheus, all curated modifications are automatically extracted, added to the variable modifications box, and selected. During the search process, all protein sequences are read, along with the locations of selected UniProt variable modifications. The Precursor Mass Tolerance (monoisotopic) was set to either ± 1000 Da or ± 200 Da for Jurkat, ± 200 Da for four common human cell lines, and ± 1000 Da for Matrigel. Other settings were as follows: Protease = trypsin (no proline rule); Maximum Missed Cleavages = 2; Initiator Methionine Behavior = variable; Fixed Modifications = carbamidomethylation of C; Variable Modifications = oxidation of M, and others automatically selected after adding the XML database; Maximum Variable Modification Isoforms Per Peptide = 1024; Precursor Monoisotopic Peak Correction = disabled; Product Mass Tolerance = ± 0.01 Da (monoisotopic); Maximum False Discovery Rate = 1%.

Stage 2 – Translating ΔM values into potential modifications and incorporating them into a new XML database. The Morpheus output file called *PSMs.tsv* generated during stage 1 contains each peptide spectral match (PSM) with its associated attributes including the actual precursor mass error (ΔM). A ΔM histogram was constructed with a bin size of 0.002 Da. Peaks in the histogram were examined as possible modifications. A *sub_ptmlist.txt* file was created and contained a list of the chosen modifications, including the name of the modification type (e.g., Phosphoserine), the modified amino acid full name (e.g., Serine), as well as the monoisotopic mass difference (e.g., 79.966330). An *AA_Name_to_letter.txt* file contained a list of the 20 common amino acid full names (e.g., Serine) and corresponding abbreviations (e.g., S). The *PSMs.tsv* file, the two *.txt* files, and the original XML database were all used as inputs to a Perl script *xml_AddOpenSearchResult.pl*. The Perl script reads in each PSM along with its ΔM . If the ΔM of a peptide is within ± 0.02 Da from a PTM type (e.g., Phosphoserine, Phosphothreonine, and Phosphotyrosine) specified in the *sub_ptmlist.txt* file, it assigns this PTM type (e.g.,

phosphorylation) to every amino acid (e.g., serine, threonine, and tyrosine) in that particular peptide by writing these potential modification identities and positions in the new XML database.

Stage 3 – Second-round search with the new XML database and narrow precursor mass tolerance. The output from running the Perl script described in stage 2 is a new XML file that contains not only the UniProt curated PTMs, but also potential modifications identified by the wide precursor tolerance search. A second-round Morpheus search was performed with this new database, ± 10 ppm precursor mass tolerance, and ± 0.01 Da product mass tolerance. All other search parameters were kept the same as in stage 1.

pMatch search

The Matrigel spectra were first searched against a protein database (downloaded from UniProt in FASTA format) with the pFind search engine¹⁸. Carbamidomethylation of cysteine was specified as a fixed modification, and oxidation of methionine as variable. Next, pMatch (version 1.5) was used for library construction from the identified spectra, followed by search of all spectra against the library, with precursor mass tolerance of ± 500 Da. Default values were used for other parameters.

MODa search

MODa version 1.23 was used to search the Matrigel spectra with the following parameters: auto parent mass correction enabled; fragment ion mass tolerance = 0.01 Da; minimum/maximum modification size = -200/+200 Da; enzyme = Trypsin, KR/C; fixed modification = C, 57.0215; HighResolution = ON. Default values were used for all other parameters. Significant peptide identifications were obtained using anal_moda.jar.

FDR and PEP calculation

A 1% global FDR at the PSM level was applied when reporting the results. This means that the ratio of the number of decoy PSMs to the number of target PSMs is 0.01.

The FDR for the modified peptides is the ratio of the number of modified decoy PSMs to the number of modified target PSMs, from the list of all PSMs meeting the 1% global FDR cutoff.

The PEP for the modified peptides at a certain score is the ratio of the number of modified decoy PSMs to the number of modified target PSMs among the PSMs that have a score within half of the score bin size from that particular score. Peptides that contain only carbamidomethylation of cysteine or oxidation of methionine were not considered as “modified”.

5.3 Results and Discussion

Global PTM Discovery (G-PTM-D) Search Strategy

In the G-PTM-D strategy, a protein database that contains both the protein sequences and the curated PTMs is employed for a first-round G-PTM search using a wide (e.g. ± 1000 Da) precursor mass tolerance (Figure 1). As described previously², the G-PTM search approach differs from the traditional variable modification search approach in that it considers only previously curated PTMs at specific amino acid residue positions, evaluating the data for either the presence or absence of the PTMs at those specific residues. The output file from this first-round wide tolerance search contains each peptide spectral match (PSM) with its associated attributes including the precursor mass error ΔM (i.e. the difference between the measured experimental mass of the peptide and the theoretical mass of the highest scoring peptide from the database). A histogram of all ΔM s from the entire search reveals numerous peaks, which

correspond to various modification types. For example, peptides with ΔM corresponding to the +79.966 Da peak in the ΔM histogram are identified as having a probable phosphorylation. For each of those peptides, a phosphorylation site is added to the original database for each serine, threonine, and tyrosine, in that peptide. This process is repeated for all of the peaks in the histogram having a ΔM readily attributable to a modification. Finally, the modified database, containing both the UniProt-curated PTMs of the original search and the newly added potential modifications (with both identity and possible locations), is used to conduct a second-round G-PTM search with the usual narrow precursor mass tolerance, resulting in identification of myriad site-specific modifications.

The performance of the G-PTM-D search strategy was evaluated by searching a deep proteomic dataset obtained from human Jurkat cells (Experimental Procedures). The ΔM histogram shown in Fig. 5.1 is from the G-PTM-D search of this dataset with ± 1000 Da precursor tolerance. Dozens of peaks in the histogram rise up well above the noise and are readily matched to the masses of known modifications. This first-round wide precursor mass tolerance search yielded 45,198 newly identified positions of potential modification, which were added to the 22,550 curated PTM positions already present in the original UniProt repository. This "supplemented" database was then used for the second-round search, yielding 16,677 site-specific modifications, comprising 27 different types (Fig. 5.1). In addition to modifications that sometimes result from sample handling (e.g., deamidation) and electrospray ionization mass spectrometry (e.g., ammonia loss, water loss, metal adducts); numerous biologically significant modifications were observed, including PTMs (e.g., phosphorylation, formylation, methylation, and acetylation), co-translational modifications (e.g., N-terminal acetylation), and amino acid variants.

G-PTM-D can provide amino acid specificity information for unknown modifications. In the Jurkat ΔM histogram (Fig. 5.1), several ΔM s (e.g. -91.007, -73.001, +249.978, +301.978, +323.960) did not correspond to any known modifications to our knowledge. To find out whether they are specific to certain amino acid (AA) residues, we assigned them to any AA in the peptides that had the respective ΔM , then incorporated them into the new XML database, and performed the second-round search with regular mass tolerances. About half of the peptides with the -91.007 Da or -73.001 Da mass differences had the modification on cysteine, which had carbamidomethylation (+57.021 Da) as a fixed modification due to the addition of iodoacetamide during sample preparation. Therefore, the modification -91.007 Da could be $-91.007 + 57.021 = -33.986$ Da, corresponding to the conversion of cysteine to dehydroalanine (DHA). The -73.001 Da modification likely results from the addition of H₂O to DHA. The mass differences of +249.978 Da, +301.978 Da, and +323.960 Da, occur more frequently on hydrophobic (Ile, Val, Phe) and acidic (Asp, Glu) AAs. The molecular formula for 301.978 could be C₈H₆N₄O₅S₂, but its chemical structure is unknown. The mass shift of +323.960 Da is likely +301.978 Da plus a sodium adduct. These examples illustrate the use of G-PTM-D to provide amino acid site localization information for previously unknown modification types.

Comparison of G-PTM-D with Variable Phosphorylation and G-PTM Searches

The Jurkat dataset was further used to compare performance of G-PTM-D with a variable phosphorylation search (“vPhospho”), and a “G-PTM” search that only uses the curated PTM information from UniProt. Compared with the other two searches, G-PTM-D identified nearly triple the number of modified proteins and unique peptides, and increased the number of modified peptide PSMs by more than six fold (Fig. 5.3a). Notably, the false discovery rate (FDR) for modified peptides identified by G-PTM-D was only 0.43% (Fig. 5.3b), which is even below

the global FDR of 1% (for all peptides, both modified and unmodified). In contrast, the FDR for phosphorylated peptides identified using the variable phosphorylation strategy was much worse (11%). Posterior error probability (PEP) values were calculated from the numbers of target and decoy spectral matches having nearly the same Morpheus score (Experimental Procedures). These PEP values are plotted in Fig. 5.3c as a function of the Morpheus score¹⁷—the peptide spectrum matching score provided by the Morpheus search algorithm. Note that PEP is a local false discovery rate, representing the probability that individual peptides with a given score are false¹⁹. At the lowest Morpheus score (9) that meets 1% global FDR, the probability is only 0.036 (3.6%) that the modified peptides from G-PTM-D are incorrect, whereas the probability is substantially higher (60.1%) for phosphorylated peptides from the variable modification search.

All three of these search strategies employ the target-decoy approach for calculations of FDR. Briefly, decoy protein sequences are generated on-the-fly by reversing the order of the amino acid residues (unmodified or modified) for each protein sequence, and PTMs move with their companion amino acid. This results in an equal number of target and decoy sequences. We searched datasets from four additional human cell lines, HEK293, A549, HeLa, and K562 (Experimental Procedures), in order to demonstrate the target-decoy approach with G-PTM-D. The numbers of unique peptide hits (all or modified, target or decoy) are plotted against Morpheus score in Fig. 5.4. The distribution of scores for target and modified-target peptides is bimodal, with the lower group of scores overlapping with the decoy peptide distributions, while the higher group of scores corresponds primarily to target peptides. A 1% FDR criterion corresponds to a Morpheus score of approximately 9, which falls between the decoy and target groups of scores.

The results from the four cell lines were also examined to illustrate the general improvement afforded by G-PTM-D for identification of modified peptides. For these data, we implemented the streamlined G-PTM-D workflow (*i.e.* automatically adding 11 regular modification types in the supplemented databases, without manual construction of the ΔM histogram). On average, ~7400 proteins were identified for each cell line. G-PTM-D revealed additional modifications, compared to G-PTM searches (Fig. 5.5). For methylation, dimethylation, and hydroxylation, which are not as well studied or curated as phosphorylation and acetylation, G-PTM-D provided a remarkable increase in the number of modified peptide identifications. These results suggest that modified peptides are more common than previously recognized and that many peptides are routinely missed or misassigned in proteomics experiments on unenriched cell lysate samples where the database search algorithm does not consider PTMs.

Apart from a remarkable improvement in the number of modified peptide identifications, G-PTM-D also delivered increased/better Morpheus scores than the G-PTM search. Among the PSMs identified by G-PTM-D with 1% FDR in the Jurkat dataset, 15% of them had higher scores for the G-PTM-D search compared with the G-PTM search, while the other 85% had the same score (Fig. 5.6). All of the PSMs with increased Morpheus scores were modified, and 98.7% of them were reassigned to different base peptide sequences by the G-PTM-D strategy compared with the G-PTM search. These spectra would have been incorrectly identified without the more complete list of modification types provided by G-PTM-D. Among the relatively few spectra that were assigned the same base peptide but with increased score, some were found to have the same type and number of modifications, but at different locations (see example in Fig. 5.7). These results indicate that the curated site information used in a G-PTM search does not always yield the correct site, and G-PTM-D may find a better match.

Comparison to Other PTM Identification Strategies

Performing the second-round search with narrow precursor mass tolerance and evaluating the data for either the presence or absence of the modifications only at specific locations, is crucial to the confident spectral identification afforded by G-PTM-D compared with other wide precursor mass tolerance search strategies^{11, 12}. In the Jurkat cell dataset, 25% (62,289) of the 249,223 PSMs identified at 1% FDR in the second-round G-PTM-D search, had not been identified in the first-round wide-tolerance search. The majority of these "rescued" PSMs (44,530) corresponded to unmodified peptides, consistent with the ~20% loss of unmodified PSMs reported by Gygi and co-workers in their wide-tolerance searches¹². The balance of the "rescued" PSMs was comprised of 17,759 PSMs for modified peptides. For example, one spectrum was matched to a *decoy* (false) peptide during the first-round search with a ΔM of +8.9263 Da, but in the second-round search, it was identified as a phosphorylated peptide from the chromosome alignment-maintaining phosphoprotein 1 with a ΔM of -0.0029 Da (annotated spectrum in Fig. 5.8). Thus, the narrow-tolerance second-round search of G-PTM-D rescues both unmodified and modified peptides, and it even corrects some assignment errors that are introduced by the wide-tolerance search.

Finally, we compared the performance of G-PTM-D with two other unrestrictive modification search tools – pMatch⁹ and MODa¹¹. This comparison was performed on an alternative and simpler dataset obtained from a Matrigel sample (Experimental Procedures) in order to demonstrate the applicability of G-PTM-D to different datasets, and also to limit the CPU hours needed for running pMatch and MODa. pMatch is based on an open MS/MS spectral library search, and MODa uses multiple sequence tags and a dynamic programming spectral alignment algorithm. At 1% FDR, G-PTM-D identifies almost double and triple the number of

PSMs compared with pMatch and MODa, respectively, for both modified and unmodified peptides (Fig. 5.9). The FDR and PEP for modified PSMs identified by G-PTM-D were also smaller than those for pMatch and MODa (Fig. 5.9b and 5.9c), indicating higher confidence by G-PTM-D. In addition, G-PTM-D detected more modification types compared with pMatch and MODa. The modification masses in MODa are in 1 Da intervals, limiting its ability to distinguish different modifications with close mass shifts. Because pMatch uses a spectral library constructed from identified spectra, instead of the entire database, to perform the wide-tolerance search, its search speed is much faster than that of G-PTM-D. However, the use of the spectral library prevents the detection of modified peptides when the unmodified forms are not identified in advance.

Search time

The first-round search of the Jurkat 28 peptide fractions (490,057 MS/MS scans) with ± 1000 Da and ± 200 Da precursor mass tolerances took 13 days and 3 days, respectively, on a Dell Precision workstation with Intel Xeon CPU, 2.70 GHz, and a maximum of 24 threads. The second-round search only took 1.4 hours. A precursor tolerance of ± 200 Da is good enough to capture the vast majority of the important and known modifications, and a total analysis time of 3 days on nearly half a million MS/MS scans is reasonable, considering the wealth of modification information acquired. We searched 3 out of the 28 fractions with ± 1000 Da precursor tolerance with either Proteome Discoverer (the search program used by Gygi *et al.*) or Morpheus, with the same protein database in either FASTA or XML format. The search times for Morpheus (73 h) and for Proteome Discoverer (62 h) were comparable. In cases where computation time/resources are limited, one could choose to perform a G-PTM search with normal precursor mass tolerance, and then use only the proteins identified in the G-PTM search

or reduce the “Maximum Variable Modification Isoforms per Peptide” in the Morpheus graphical user interface, to perform the wide precursor tolerance search. Another option for saving time on repeated analyses of similar types of samples is to only perform the wide-tolerance search on a representative subset of the samples. Then, one could use the XML database of modified peptides generated from this set for narrow-tolerance G-PTM type searches of all other similar samples.

G-PTM-D offers a number of additional advantages compared to other PTM identification software tools. The Morpheus algorithm employed with G-PTM-D accepts up to ten modifications per peptide, while the maximum number we detected in all the datasets was seven. Thus, the search algorithm does not effectively limit the identification of multiply-modified peptides. The fact that G-PTM-D can detect peptides that contain more than one modification extends its analytical capability, especially in cases where different modification types coexist within a single peptide. In addition, most other software tools (such as MS-Alignment²⁰, ModifiComb⁶, MODa¹¹, PeaksPTM²¹, DeltAMT¹⁰, and pMatch⁹) require the coexistence of the modified and unmodified forms of a peptide in the sample, while G-PTM-D does not. Furthermore, ModifiComb and DeltAMT cannot detect modifications when the modified and unmodified forms are offline separated into different fractions, while G-PTM-D is able to combine any number of fractions in the same search.

In summary, the G-PTM-D search strategy is able to reveal a wide variety of site-specific modifications with high confidence in deep proteomic datasets from unenriched samples. It achieves this by searching for the presence or absence of modifications only at either already curated or at potential new sites discovered in a wide tolerance search. This search strategy greatly reduces the search space compared to conventional PTM variable searches, and provides

increased confidence in the identification of modified peptides. Importantly, no pre-enrichment of samples for particular PTM types is required, thus providing a broad and unbiased view of a wide range of modifications. G-PTM-D provides a powerful new tool for the identification and discovery of protein variation.

5.4 Figures and Tables

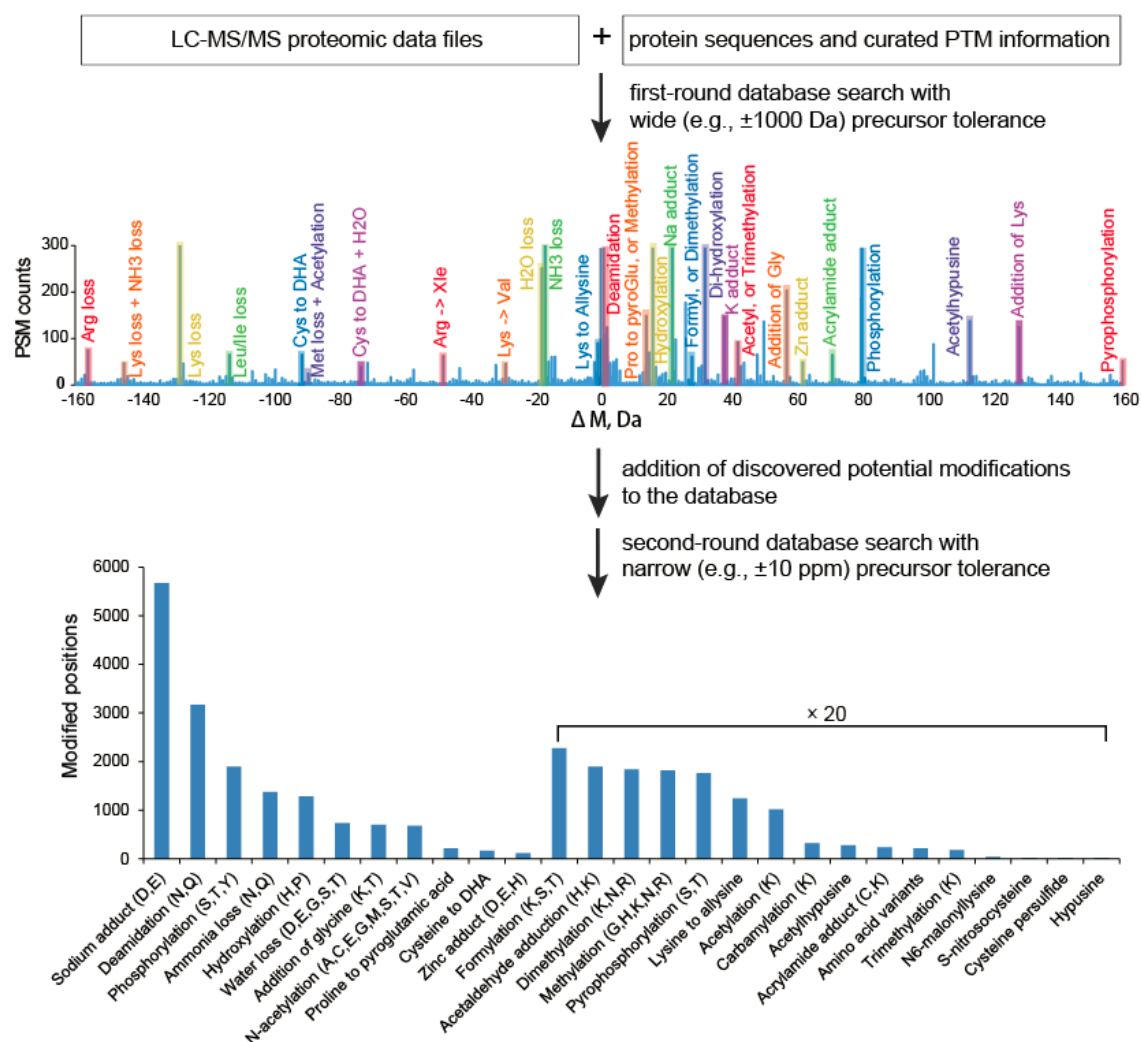


Figure 5.1. G-PTM-D workflow, illustrated with results from the Jurkat cell dataset. An expanded view (± 160 Da) of the ΔM histogram is displayed here; the full histogram (± 1000 Da) and a comparison to a ± 200 Da search are shown in Figure 5.2.

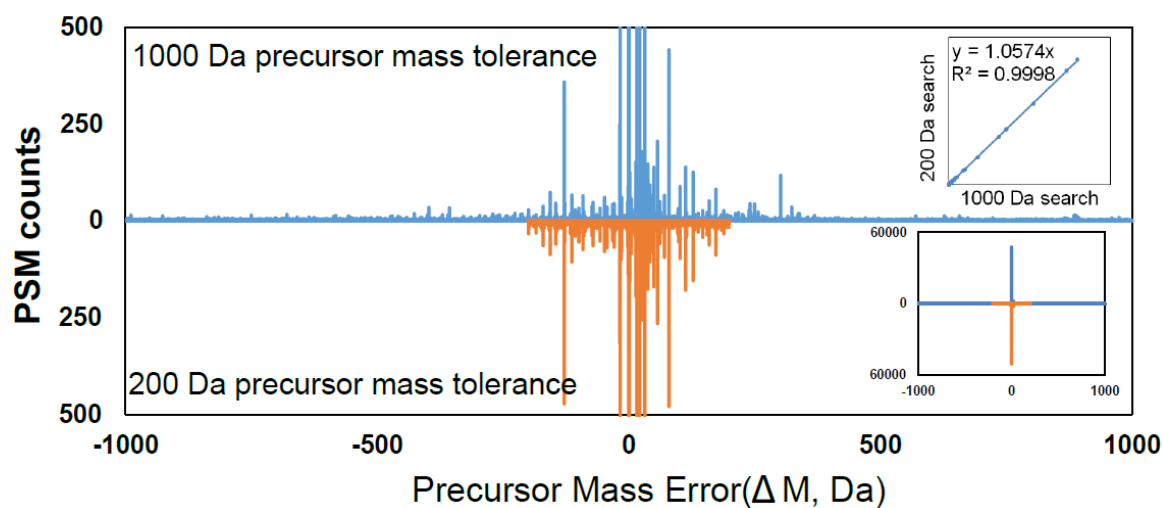


Figure 5.2. Histograms of precursor mass error (ΔM) searched with ± 1000 Da or ± 200 Da precursor mass tolerances. The lower-right inset shows the unexpanded plot. The upper right inset shows excellent correlation between the peak heights obtained from the two searches, indicating that the sensitivity in identifying potential modifications is not compromised by the size of the mass tolerance, although modifications exceeding 200 Da cannot be observed in the ± 200 Da search.

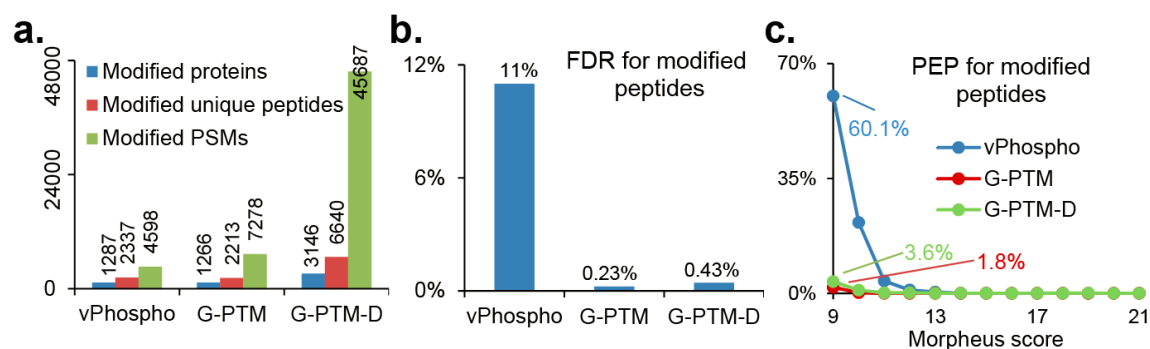


Figure 5.3. Results from three types of searches of the Jurkat cell dataset: a vPhospho search (using the UniProt FASTA database with phosphorylation as a variable modification), a G-PTM search (using the PTM-curated UniProt database), and a G-PTM-D search. a) Numbers of modified proteins, unique peptides, and PSMs for each search. b) False discovery rate (FDR) for modified peptides. c) Posterior error probability (PEP) for modified peptides as a function of the Morpheus score. All results are based on 1% global FDR.

—●— Target hits —●— Decoy hits —●— Modified target hits —●— Modified decoy hits

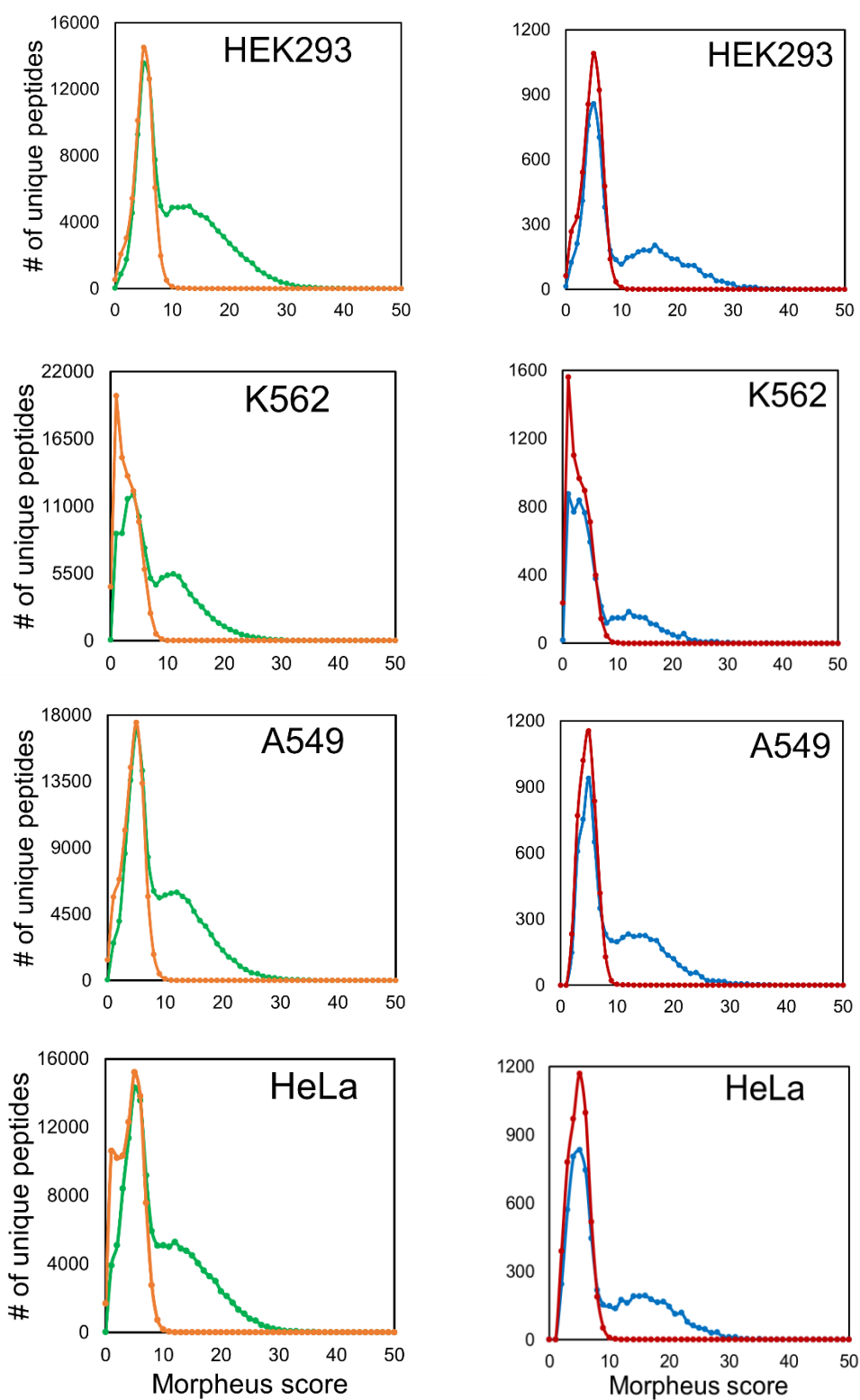


Figure 5.4. Histograms of the number of unique peptides (all or modified, target or decoy) against Morpheus score, for four human cell lines (score bin size = 1). Note that in situations where the score of a spectrum match is identical for both decoy and target peptide, Morpheus automatically assigns the match to decoy, which is why the number of decoy hits exceeds the number of target hits at low scores.

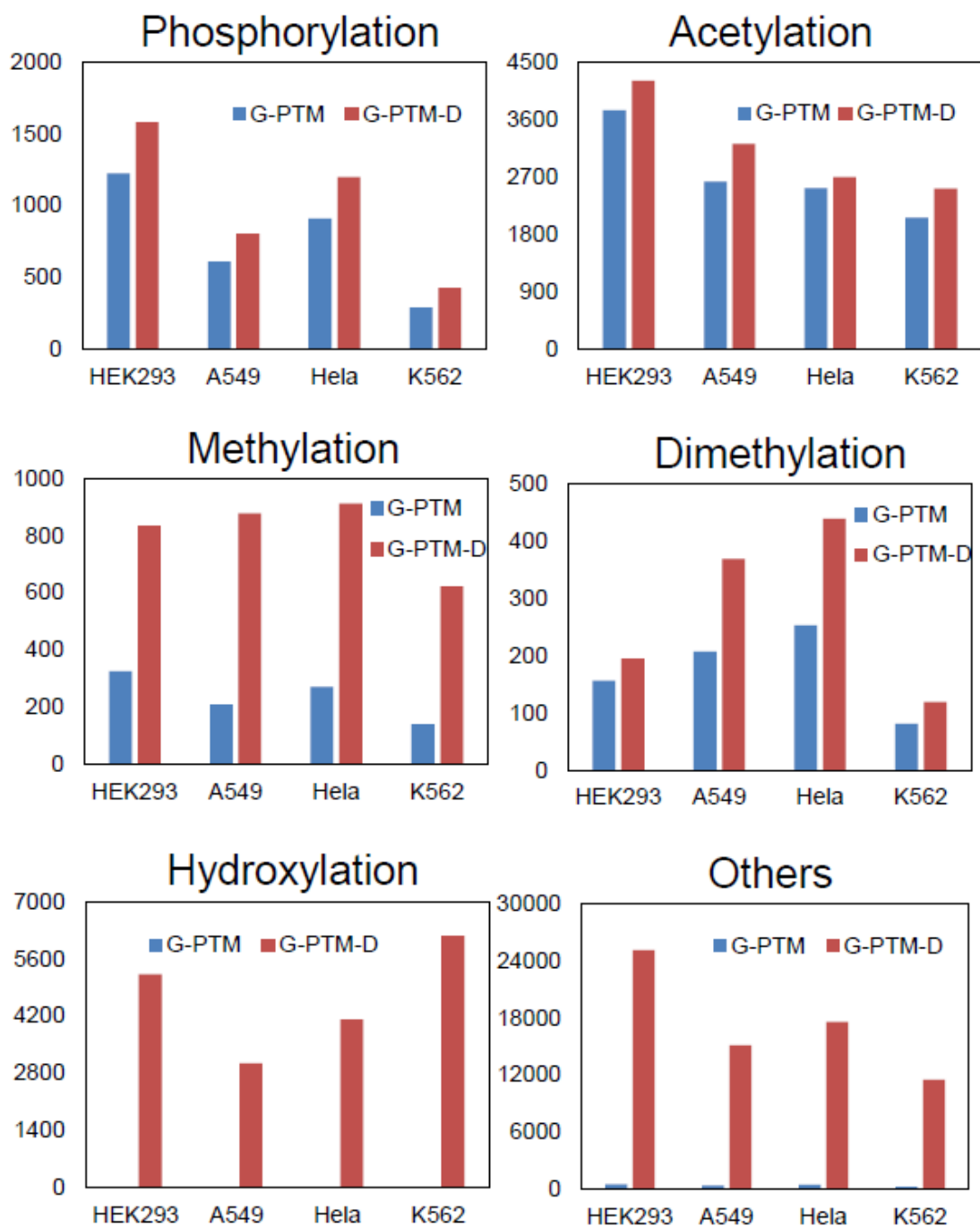


Figure 5.5. Numbers of peptides with modifications identified by G-PTM or G-PTMD for the four human cell lines. “Others” include trimethylation, carboxylation, sulfation, water loss, ammonia loss, and deamidation.

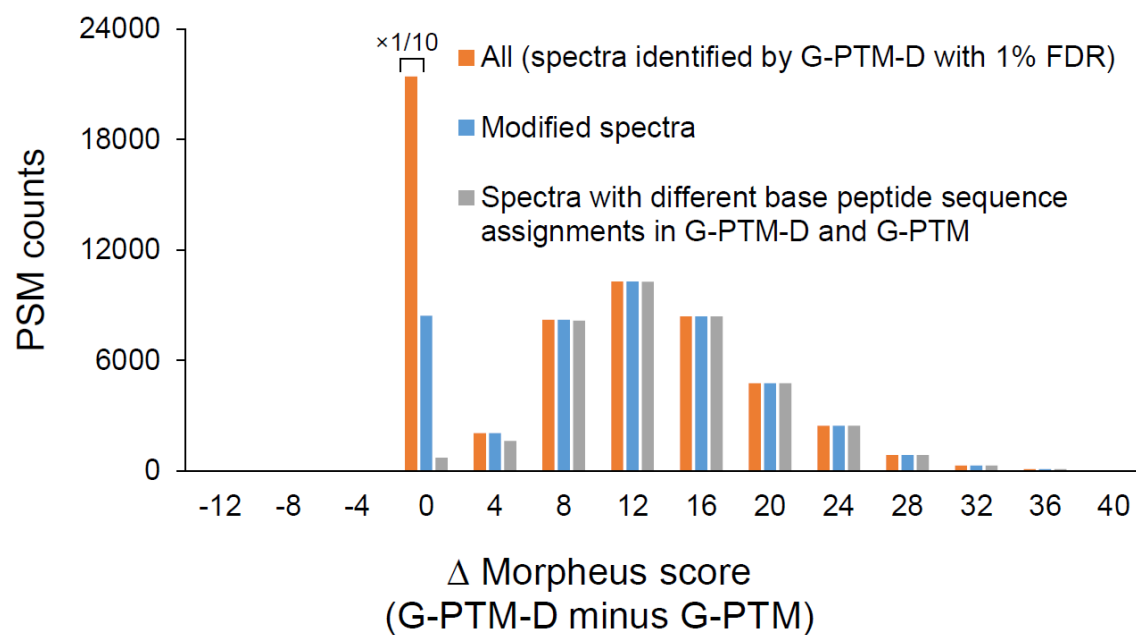
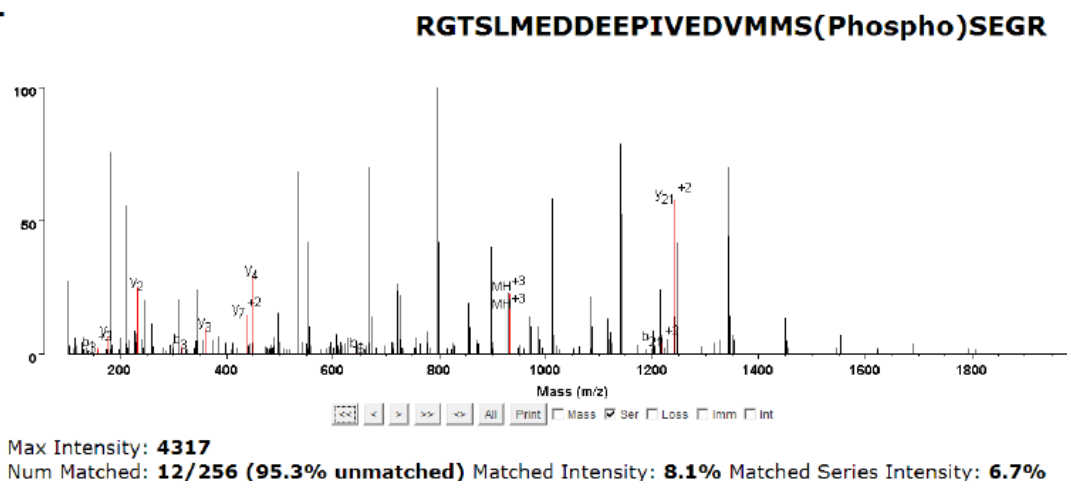


Figure 5.6. Histogram of Δ Morpheus score - the difference between the Morpheus score by G-PTM-D and the Morpheus score by G-PTM, for each of the Jurkat spectra that were identified by G-PTM-D with 1% FDR. A Δ Morpheus score of zero indicates no difference between the two types of searches. The positive Δ Morpheus scores (15% of all assignments) indicate G-PTM-D found a better match, and all of these improved cases were for modified spectra.

a.



b.

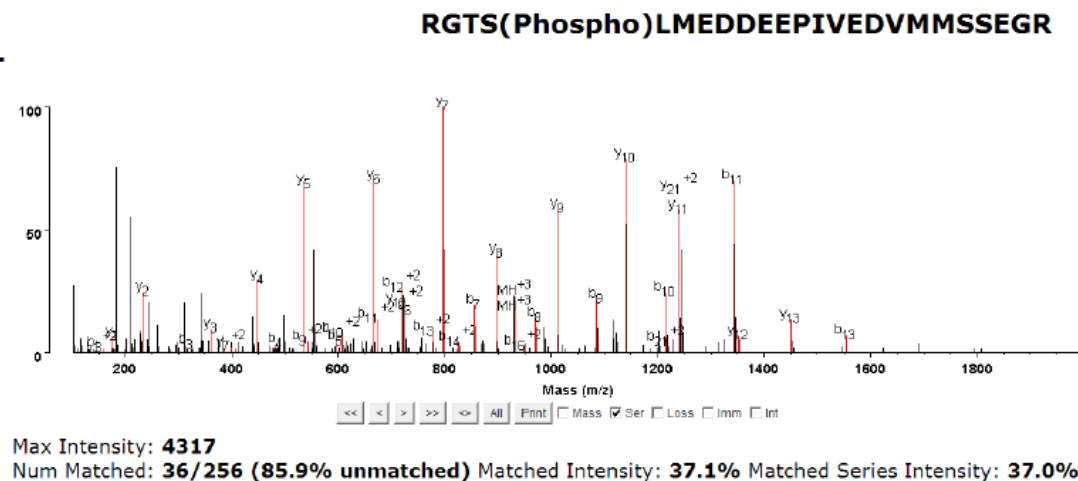


Figure 5.7. Annotation of the same spectrum from a) G-PTM identification, which includes phosphorylation of serine 20, and b) G-PTM-D identification, which yields a much better match to fragment ions for this phosphorylation of serine 4. Note that the G-PTM search employed the curated phosphorylation sites from UniProt, which only included serine 20 for this peptide. G-PTM-D, however, was able to reassign this spectrum to phosphorylation of serine 4, which is likely the correct modification site, given the substantial improvement in fragment ion matches.

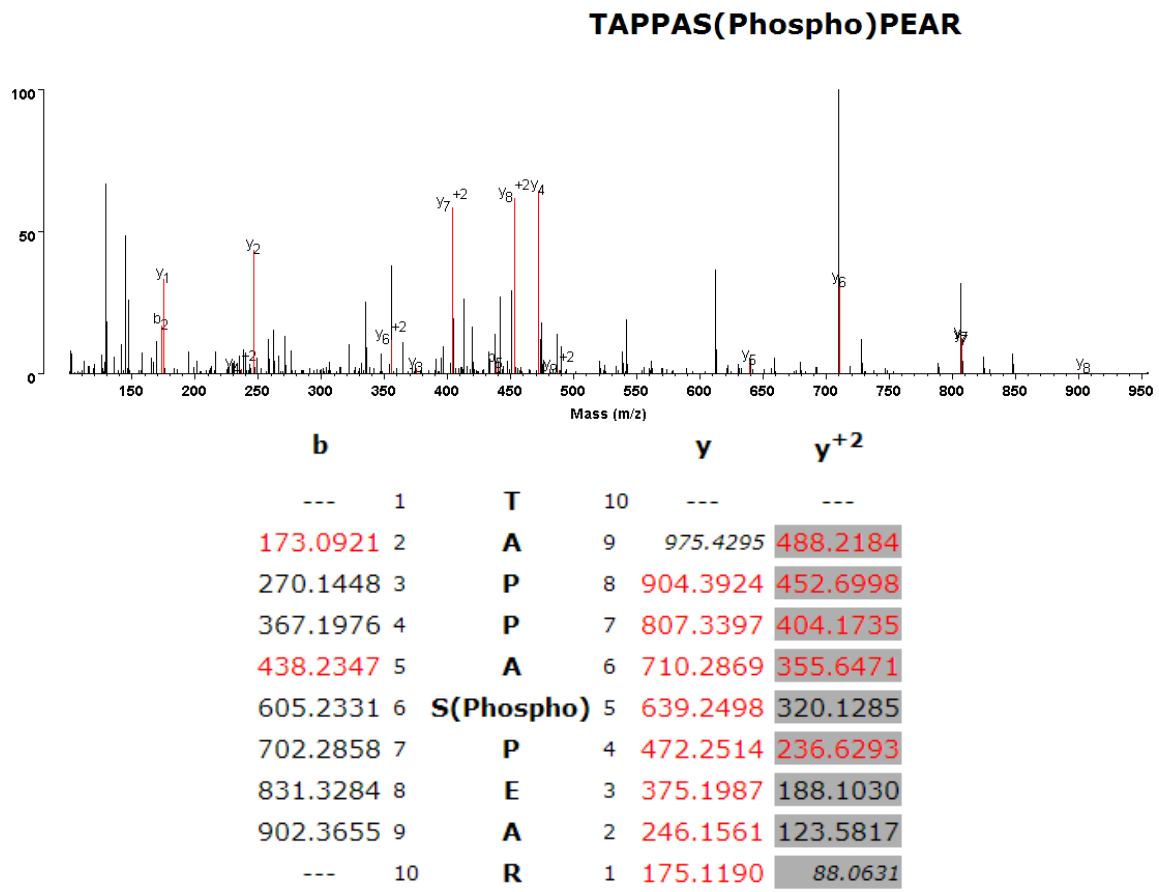


Figure 5.8. An annotated spectrum that is assigned as TAPPAS(Phospho)PEAR in the second-round search, but was assigned (incorrectly) as a decoy peptide in the wide tolerance first-round search of G-PTM-D of the Jurkat dataset. Red font is used to represent ion matches.

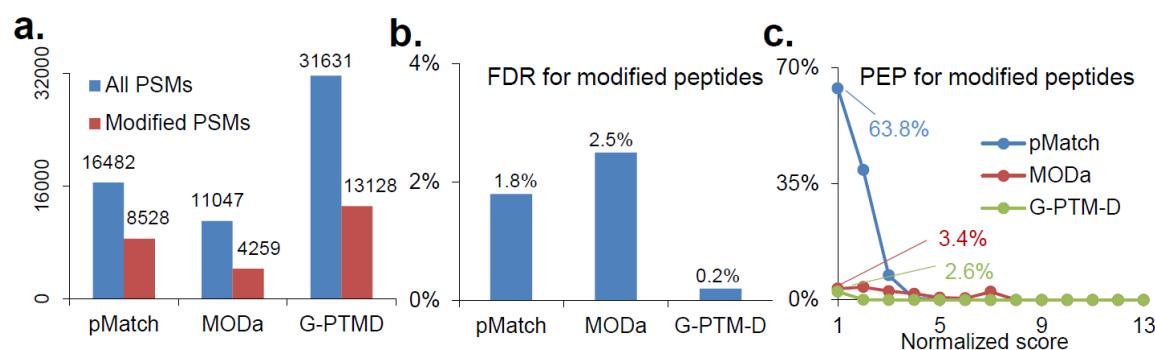


Figure 5.9. Results from searches of the Matrigel dataset using pMatch, MODa, and G-PTM-D.

a) Numbers of identified PSMs (all and modified). b) False discovery rate (FDR) for modified peptides. c) Posterior error probability (PEP) for modified peptides as a function of score. Note that the PEP values were plotted versus normalized scores to account for the different score scales of Morpheus, pMatch, and MODa (scores were normalized to a scale from 1 to 13). All results in this figure met a 1% global FDR.

5.5 References

1. Olsen, J.V. & Mann, M. Status of large-scale analysis of post-translational modifications by mass spectrometry. *Mol Cell Proteomics* **12**, 3444-3452 (2013).
2. Shortreed, M.R. et al. Global Identification of Protein PTMs in a Single-pass Database Search. *J Proteome Res* (2015).
3. Cesnik, A.J., Shortreed, M.R., Sheynkman, G.M., Frey, B.L. & Smith, L.M. Human Proteomic Variation Revealed by Combining RNA-Seq Proteogenomics and Global Post-Translational Modification (G-PTM) Search Strategy. *J Proteome Res* (2016).
4. Gelse, K., Poschl, E. & Aigner, T. Collagens--structure, function, and biosynthesis. *Advanced drug delivery reviews* **55**, 1531-1546 (2003).
5. Shoulders, M.D. & Raines, R.T. Collagen structure and stability. *Annual review of biochemistry* **78**, 929-958 (2009).
6. Savitski, M.M., Nielsen, M.L. & Zubarev, R.A. ModifiComb, a new proteomic tool for mapping substoichiometric post-translational modifications, finding novel types of modifications, and fingerprinting complex protein mixtures. *Mol Cell Proteomics* **5**, 935-948 (2006).
7. Baumgartner, C., Rejtar, T., Kullolli, M., Akella, L.M. & Karger, B.L. SeMoP: a new computational strategy for the unrestricted search for modified peptides using LC-MS/MS data. *J Proteome Res* **7**, 4199-4208 (2008).
8. Chen, Y., Chen, W., Cobb, M.H. & Zhao, Y. PTMap--a sequence alignment software for unrestricted, accurate, and full-spectrum identification of post-translational modification sites. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 761-766 (2009).
9. Ye, D. et al. Open MS/MS spectral library search to identify unanticipated post-translational modifications and increase spectral identification rate. *Bioinformatics (Oxford, England)* **26**, i399-i406 (2010).
10. Fu, Y. et al. DeltAMT: a statistical algorithm for fast detection of protein modifications from LC-MS/MS data. *Mol Cell Proteomics* **10**, M110 000455 (2011).
11. Na, S., Bandeira, N. & Paek, E. Fast multi-blind modification search through tandem mass spectrometry. *Mol Cell Proteomics* **11**, M111 010199 (2012).
12. Chick, J.M. et al. A mass-tolerant database search identifies a large proportion of unassigned spectra in shotgun proteomics as modified peptides. *Nat Biotech* **advance online publication** (2015).
13. Sheynkman, G.M., Shortreed, M.R., Frey, B.L. & Smith, L.M. Discovery and Mass Spectrometric Analysis of Novel Splice-junction Peptides Using RNA-Seq. *Mol Cell Proteomics* **12**, 2341-2353 (2013).
14. Desiere, F. et al. The PeptideAtlas project. *Nucleic acids research* **34**, D655-658 (2006).
15. Geiger, T., Wehner, A., Schaab, C., Cox, J. & Mann, M. Comparative proteomic analysis of eleven common cell lines reveals ubiquitous but varying expression of most proteins. *Mol Cell Proteomics* **11**, M111 014050 (2012).
16. Li, Q. et al. Proteomic Analysis of Naturally-Sourced Biological Scaffolds. *Biomaterials*.
17. Wenger, C.D. & Coon, J.J. A Proteomics Search Algorithm Specifically Designed for High-Resolution Tandem Mass Spectra. *J Proteome Res* **12**, 1377-1386 (2013).

18. Wang, L.H. et al. pFind 2.0: a software package for peptide and protein identification via tandem mass spectrometry. *Rapid communications in mass spectrometry : RCM* **21**, 2985-2991 (2007).
19. Käll, L., Storey, J.D., MacCoss, M.J. & Noble, W.S. Posterior Error Probabilities and False Discovery Rates: Two Sides of the Same Coin. *J Proteome Res* **7**, 40-44 (2008).
20. Tsur, D., Tanner, S., Zandi, E., Bafna, V. & Pevzner, P.A. Identification of post-translational modifications by blind search of mass spectra. *Nature biotechnology* **23**, 1562-1567 (2005).
21. Han, X., He, L., Xin, L., Shan, B. & Ma, B. PeaksPTM: Mass spectrometry-based identification of peptides with unspecified modifications. *J Proteome Res* **10**, 2930-2936 (2011).

Chapter 6

Conclusions, Ongoing Projects, and Future Directions

6.1 Conclusions

We developed new mass spectrometry-based proteomic assays that allowed us to fully characterize, at the biological system level, both the decellularization and the repopulation processes of a tissue engineering technique that has the potential to transform human medicine. These assays achieved optimal proteomic coverage of both ECM and cellular proteins in native, decellularized, and engineered tissues. We quantitatively measured the turnover of hundreds of matrix and cellular proteins, the first study evaluating dynamic proteome changes during the remodeling of decellularized scaffolds by reseeded cells *in vitro*. These proteomic techniques and resulting data have advanced our understanding of tissue biology, enabled us to optimize each step of the tissue engineering approach, and created a fundamental knowledge base for guiding future tissue engineering efforts.

A bioinformatics tool for global protein modification discovery has also been developed. This strategy, named G-PTM-D, eliminates the pre-enrichment step required in traditional PTM identification methods, results in low false discovery rate, and discovers a myriad of modifications that are present in biological samples. G-PTM-D extracts an additional rich set of information, which has been neglected, from regular proteomic datasets. We believe that this new strategy, with its open-source implementation, will provide crucial new information to biologists who seek to understand the nature and diversity of proteomic variation in their systems of interest.

6.2 Ongoing Project #1 - Regeneration of Whole Canine Larynx

6.2.1 Objectives

The main purpose of this study is to evaluate the progressive decellularization of whole dog larynx via arterial perfusion, and then regenerate the whole larynx by cell reseeded.

6.2.2 Methods

Fresh dog larynges were heparinized, decellularized, and washed, via sequential arterial perfusion of sodium dodecyl sulfate (for 12 or 72 h), Triton-X 100 (for 12 h), and saline (for 12 h). Three larynges were used as controls. Vocal fold mucosae (VF), thyroarytenoid muscles (TA), cricoid cartilages (CC), and thyroid gland (TG) were evaluated using histology and immunohistochemistry (IHC), quantitative DNA, collagen, elastin and sulfated glycosaminoglycan (sGAG) assays, and mass spectrometry-based proteomics described as follows.

Protein extraction and digestion were similar to what has been described in previous chapters. Briefly, each tissue sample weighing between 4 mg to 50 mg was minced, added with 150 μ L to 500 μ L of SDT solution, sonicated, and centrifuged, followed by the FASP protocol for SDS removal and on-filter digestion. After the digestion was quenched with 10% trifluoroacetic acid (TFA) to a final concentration of 0.5% TFA, samples were desalted using Sep-Pak C18 1 cc Vac Cartridges. The eluent was dried down and reconstituted in 5% ACN, 2% FA. Approximately 0.17 μ g protein digest, estimated by BCA protein assay, was injected into a Waters nanoAcquity HPLC coupled to an ESI ion-trap/Orbitrap mass spectrometer. Peptides were separated on a 100 μ m inner diameter column packed with 20 cm of 1.7 μ m BEH C18 particles, with a heater cartridge used to keep the capillary column at 60 $^{\circ}$ C. Peptides were eluted at 0.3 μ L/min in 0.1% FA with a gradient of increasing ACN over 2.5 h.

For each tissue type, the acquired MS raw files from the three decellularization conditions, each with six replicates, were analyzed together by MaxQuant (version 1.5.2.8). The UniProt database used contained 28,436 protein sequences (*canis lupus familiaris*) downloaded on Feb 21, 2014, supplemented with 262 common contaminants. A minimum of two unique peptides per protein was required. Protein groups containing matches to proteins from the reversed database or contaminants were discarded. The intensity-based absolute quantification (iBAQ) algorithm embedded in the MaxQuant software package was employed.

6.2.3 Preliminary Results

The larynges became progressively translucent from native to 12h to 72 h. Histology and IHC indicated loss of nuclei and cytoskeletal protein β -actin, alongside preservation of collagen types I and IV and hyaluronic acid, particularly in VF and TA (data not shown). DNA and elastin were depleted in both VF and TA ($p < 0.05$); collagen and sGAG were preserved in VF and CC ($p > 0.05$) (data not shown).

Proteomic analyses showed that, for all four tissue types, although there were still a large number of cellular proteins (Fig. 6.1), the intensity of them significantly went down (Fig. 6.2). Thyroid gland was the easiest to decellularize, based on the observations that the number of cellular proteins went from 1207 to 588 to 327, and most of the cellular protein removal happened in the first 12 hours (Fig. 6.1). TG was directly perfused by the carotid artery and probably with the highest arterial flow rate, which may have made it easier for protein removal. Collagens and ECM glycoproteins were generally well preserved in all tissue types, but other categories (proteoglycans, ECM-affiliated proteins, secreted factors, and ECM regulators) were more easily removed (Fig. 6.3). CC was relatively stable over time, predominantly due to

superior preservation of proteoglycans. More conclusions will be drawn with further analysis and interpretation of Figs. 6.1-6.4. Recellularization of the whole larynx will be performed and studied.

6.3 Ongoing Project #2 - Optimizing Lung Scaffolds for Organ Regeneration

6.3.1 Objectives

One objective of this study was to study the differences in scaffolds of lungs at different ages, and to determine how these differences affect cell reseeded and therefore how to make a better scaffold for lung regeneration.

6.3.2 Methods

Six lungs from pediatric donors at different ages (2 days, 6 days, 7 days, 3 months, 9 months, 7 years) and three lungs from adult donors were perfusion decellularized. Proteomic analyses were performed, with HPLC offline pre-separation of each sample into six fractions.

6.3.3 Preliminary Results

The 3 months (m) sample showed abnormally high abundance of collagens, consistent with the fact that it had a mutation in the ABCA3 gene, a consequence of which is increased lung stiffness and thickened alveolar septa. Because of the mutation, the 3 m sample was excluded from further comparative analyses.

The 2 d, 6 d, 7 d were considered as three biological replicates for the neonatal group; the 9 m and 7 y were considered as two biological replicates of the pediatric group; and the three adult samples were the adult group. Fig. 6.5 shows that the percentage of cellular proteins in the

scaffolds was notably higher in the adult group than in the neonatal and pediatric group, indicating that the adult lungs were more difficult to decellularize. Fig. 6.6 shows the percentage of matrisome subcategories out of all matrisome proteins. The younger groups have higher collagen intensity. As the lung grows, glycoproteins and proteoglycans accumulate. The ECM regulators and secreted factors intensities go up in the later stage of the development (i.e., no significant difference between neonatal and pediatric, but a lot higher in the adult group). The next step in the experimental stage is to find protein targets that are differentially present in different groups, manipulate the culturing media for reseeded cells, and determine how these proteins affect cell reseeded.

6.4 Future Directions

We will continue to promote the adoption of proteomic methodologies in tissue engineering investigations, and wide application of the G-PTM-D strategy in the biological and proteomics community.

A large portion of the work presented in this thesis is focused on analysis of proteins in tissues. Since glycan is another important component of many tissues, we will develop a robust method for glycomic analysis, and examine tissue remodeling at both the proteome and glycome levels, in order to capture the full biological system-level output and be able to appropriately and completely relate these changes to tissue functionality. We will perform isotopic labeling glycan turnover studies by replacing the naturally occurring “light” form of glutamine (amide- ^{14}N -Gln; the sole nitrogen source for synthesis of both N- and O-linked glycans) with the “heavy” form

(amide- ^{15}N -Gln), which will allow us to identify and quantify new glycans synthesized by the seeded cells (will be “heavy” labeled) from residual scaffold glycans (will be unlabeled).

We are also interested in understanding how remodeling of the engineered tissue continues *in vivo*. We will isotopically label mice by feeding a “heavy” diet with $^{13}\text{C}_6$ -Lys/Arg and amide- ^{15}N -Gln. We will monitor systemic $^{13}\text{C}_6$ -Lys/Arg and amide- ^{15}N -Gln abundance via mass spectrometry analysis of weekly blood draws and proceed to tissue implantation once mice exhibit >95% incorporation of each isotope. After implantation of an engineered vocal fold mucosa, we will use mass spectrometry to distinguish and quantify proteins and glycans introduced by the host (“heavy” labeled) from those of the original graft (unlabeled).

For the G-PTM-D strategy, one current limitation lies in the long CPU hours needed for performing wide-tolerance searches. To mitigate this issue, we could modify the source code of the search engine so that the wide tolerances are restricted to multiple discrete windows, instead of one continuous window. Another limitation of the current G-PTM-D implementation is that it is not designed for optimal site localization of modifications. This may be remedied by combining the G-PTM-D strategy with other spectrum identification and score reporting algorithms that take into consideration the presence and intensity of site-determining ions in MS/MS spectra.

6.5 Figures and Tables

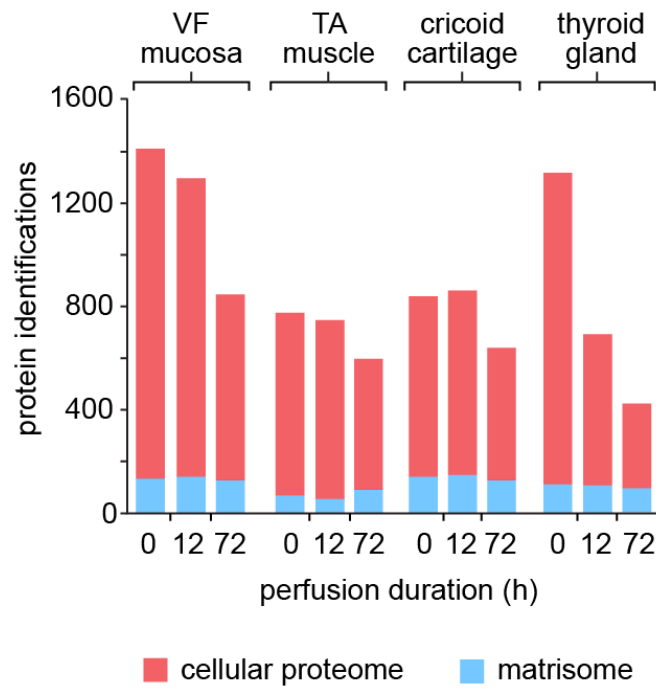


Figure 6.1. The numbers of matrisome and cellular proteins in each experimental condition, for each of the four tissue types within the larynges.

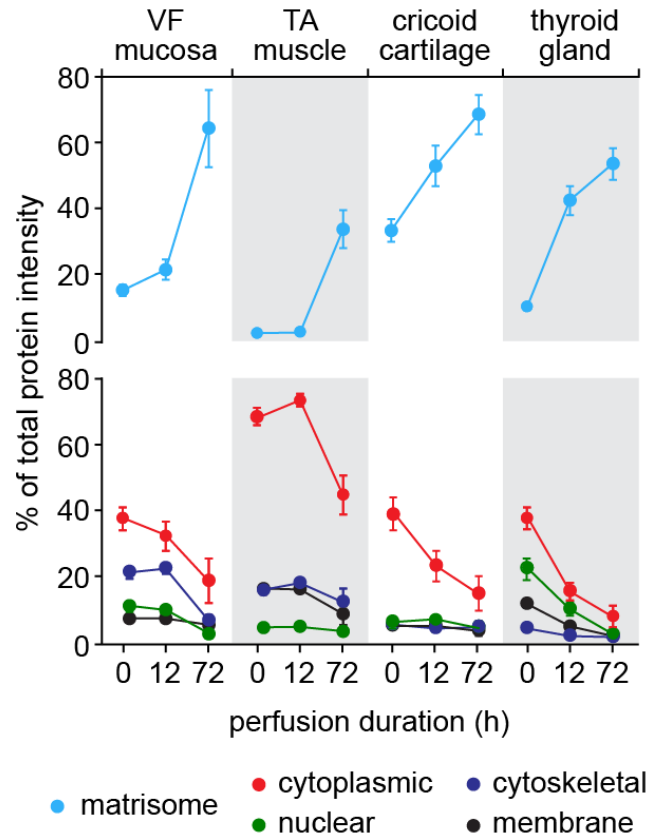


Figure 6.2. The percentage of MS intensity attributed to the matrisome, cytoplasmic, cytoskeletal, nuclear, and membrane proteins, out of the total MS intensity, for each experimental condition. N = 6. Error bars, s.e.m.

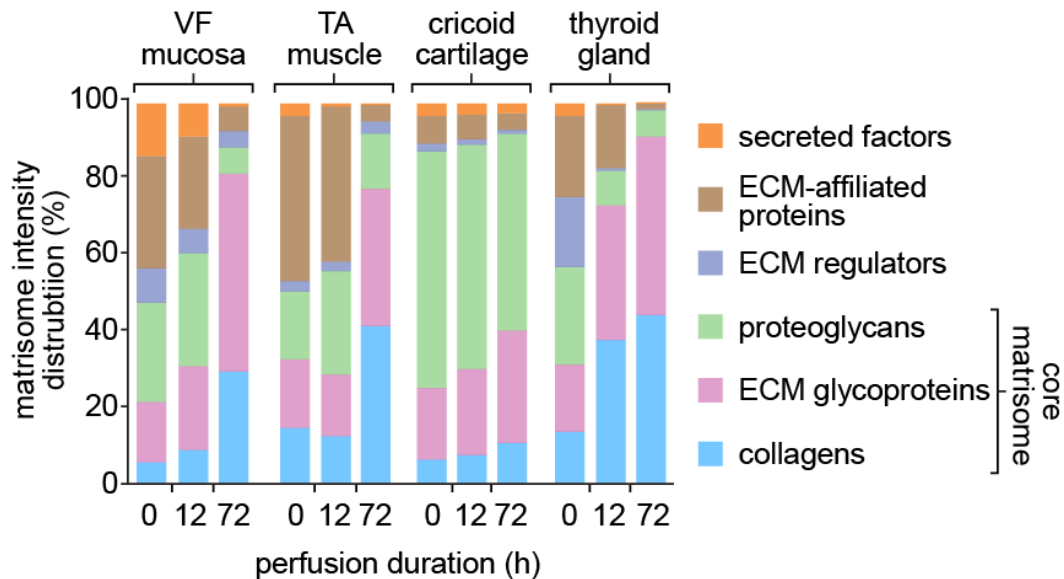


Figure 6.3. The percentage of MS intensity attributed to each of the six matrisome sub-categories, out of the total matrisome protein intensity, for each experimental condition.

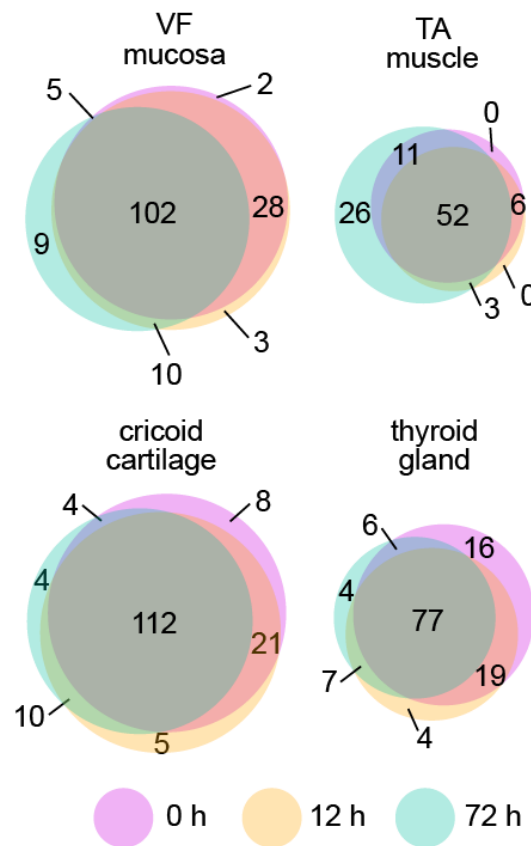


Figure 6.4. Venn diagrams of matrisome protein identifications.

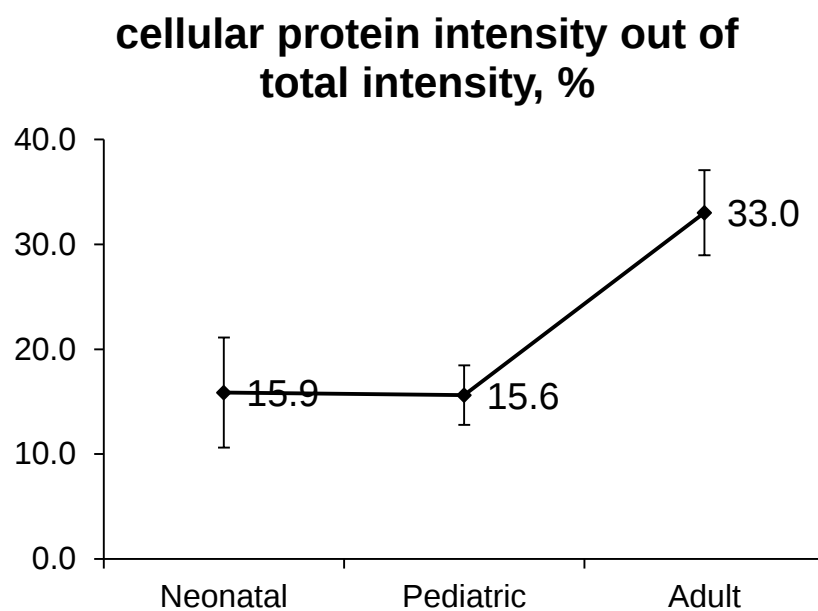


Figure 6.4. The percentage of MS intensity attributed to cellular proteins, out of the total sample intensity, for each experimental group – neonatal, pediatric, and adult. Error bars, s.e.m.

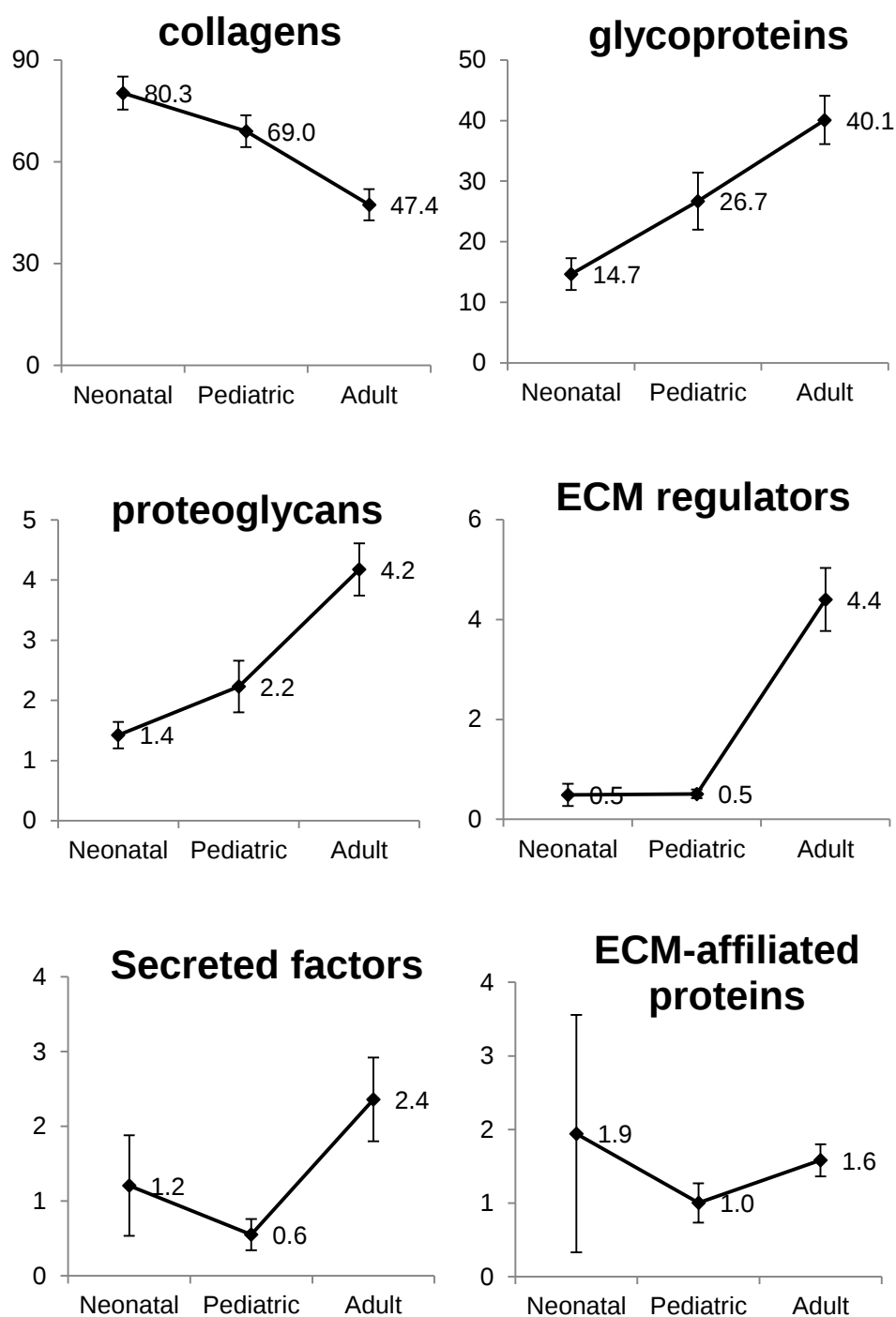


Figure 6.5. The percentage of MS intensity attributed to the six matrisome categories, out of the total matrisome intensity, for each experimental group. Error bars, s.e.m.