Biology of Extracellular Matrix 7
Series Editor: Nikos K. Karamanos

Sylvie Ricard-Blum *Editor*

Extracellular Matrix Omics

Springer

Biology of Extracellular Matrix

Volume 7

Series Editor

Nikos K. Karamanos, Department of Chemistry/Laboratory of Biochemistry, University of Patras, Patras, Greece

Extracellular matrix (ECM) biology, which includes the functional complexities of ECM molecules, is an important area of cell biology. Individual ECM protein components are unique in terms of their structure, composition and function, and each class of ECM macromolecule is designed to interact with other macromolecules to produce the unique physical and signaling properties that support tissue structure and function. ECM ties everything together into a dynamic biomaterial that provides strength and elasticity, interacts with cell-surface receptors, and controls the availability of growth factors. Topics in this series include cellular differentiation, tissue development and tissue remodeling. Each volume provides an in-depth overview of a particular topic, and offers a reliable source of information for post-graduates and researchers alike. "Biology of Extracellular Matrix" is published in collaboration with the American Society for Matrix Biology.

More information about this series at http://www.springer.com/series/8422

Sylvie Ricard-Blum
Editor

Extracellular Matrix Omics

Springer

Editor
Sylvie Ricard-Blum
Institute of Molecular and Supramolecular
Chemistry and Biochemistry (ICBMS)
Université Claude Bernard Lyon 1
Villeurbanne Cedex, Lyon, France

ISSN 0887-3224 ISSN 2191-1959 (electronic)
Biology of Extracellular Matrix
ISBN 978-3-030-58329-3 ISBN 978-3-030-58330-9
https://doi.org/10.1007/978-3-030-58330-9

This Springer imprint is published by the registered company Springer Nature Switzerland AG.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

The extracellular matrix (ECM) comprises proteins (the matrisome) and complex, sulfated, polysaccharides, the glycosaminoglycans, which are covalently attached to core proteins to form hybrid molecules called proteoglycans. A non-sulfated glycosaminoglycan, hyaluronan, forms proteoglycan aggregates of high molecular weight mediated by its non-covalent interactions with link proteins. The matrisome is a small proteome encoded by ~1000 genes in humans, but it is still underexplored because of its specific features. A number of ECM proteins are multimeric, multidomain, and deposited in the extracellular matrix as insoluble and cross-linked supramolecular assemblies, which requires the adaptation of existing protocols and tools and/or the development of new protocols and tools to collect and analyze ECM -omic datasets. Furthermore, limited proteolysis of ECM proteins gives rise to bioactive fragments called matricryptins or matrikines, which have biological activities of their own, different from those of their parent proteins. This book aims at providing the readers with general and specific computational tools and resources to visualize and analyze ECM datasets, and with examples of -omic approaches unique to the ECM (e.g., glycosaminoglycomics and proteoglycanomics), and their use to assess ECM remodeling (degradomics) in health and diseases.

The matrisome, which comprises ECM and ECM-affiliated proteins, has been first defined *in silico* in human and model organisms (i.e., mice, *Caenorhabditis elegans*, and *Danio rerio*; see Chap. 2 by Gebauer and Naba), and then experimentally characterized by quantitative proteomics in a variety of healthy and diseased tissues, such as fibrotic liver (see Chap. 3 by Dolin et al.) and tumors (see Chap. 7 by Izzi et al.), and in biological processes (see Chap. 8 on ECM degradation by Kalogeropoulos et al.). Experimental protocols to collect -omic data are detailed in several chapters. The interpretation of these data requires the use of computational approaches, and the major general and ECM-specific bioinformatic tools and databases used to annotate ECM -omic data are listed by Naba and Ricard-Blum in the first chapter. Specific omics have been developed in addition to proteomics to take into account the other ECM components, namely proteoglycans (see Chap. 4 on proteoglycanomics by Koch and Apte) and glycosaminoglycans (see Chap. 5 on

glycosaminoglycomics by Sethi and Zaia). The ability of ECM proteins, proteoglycans, and glycosaminoglycans to form interaction networks within the ECM and with the cell surface is addressed in Chaps. 6 (Ricard-Blum) and 9 (Koeleman et al.). The integration of -omic data to build models of ECM-dependent signaling pathways is also illustrated (see Chap. 10 on integrative models for TGFβ signaling and ECM by Théret et al.).

Villeurbanne, France Sylvie Ricard-Blum

Contents

1 **The Extracellular Matrix Goes -Omics: Resources and Tools** 1
Alexandra Naba and Sylvie Ricard-Blum

2 **The Matrisome of Model Organisms: From In-Silico Prediction to Big-Data Annotation** 17
Jan M. Gebauer and Alexandra Naba

3 **Detecting Changes to the Extracellular Matrix in Liver Diseases** 43
Christine E. Dolin, Toshifumi Sato, Michael L. Merchant, and Gavin E. Arteel

4 **Characterization of Proteoglycanomes by Mass Spectrometry** 69
Christopher D. Koch and Suneel S. Apte

5 **Historical Overview of Integrated GAG-omics and Proteomics** 83
Manveen K. Sethi and Joseph Zaia

6 **Extracellular Matrix Networks: From Connections to Functions** 101
Sylvie Ricard-Blum

7 **Integration of Matrisome Omics: Towards System Biology of the Tumor Matrisome** 131
Valerio Izzi, Jarkko Koivunen, Pekka Rappu, Jyrki Heino, and Taina Pihlajaniemi

8 **Proteomic and Degradomic Analysis of Body Fluids: Applications, Challenges and Considerations** 157
Konstantinos Kalogeropoulos, Louise Bundgaard, and Ulrich auf dem Keller

9 Regulation of Cell-Matrix Adhesion Networks: Insights from Proteomics 183
Emma S. Koeleman, Alexander Loftus, Athanasia D. Yiapanas, and Adam Byron

10 Integrative Models for TGF-β Signaling and Extracellular Matrix 209
Nathalie Théret, Jérôme Feret, Arran Hodgkinson, Pierre Boutillier, Pierre Vignet, and Ovidiu Radulescu

Chapter 1
The Extracellular Matrix Goes -Omics: Resources and Tools

Alexandra Naba and Sylvie Ricard-Blum

Abstract The extracellular matrix (ECM) is the complex scaffold made of hundreds of proteins that governs the organization of cells and tissues in all multicellular organisms. It provides structural and mechanical properties to tissues. It also exerts signaling roles, either directly by interacting with cell surface receptors, or by interacting with growth factors and modulating their signaling activities, and by doing so regulates a multitude of cellular functions including cell-matrix interactions, cell proliferation, survival, and differentiation. The purpose of this introductory chapter is to present resources and tools developed to facilitate the identification and analysis of ECM genes and proteins across different conditions using high-throughput methodologies (i.e., genomics, transcriptomics, proteomics, and interactomics). Databases focused on specific ECM genes and ECM-related diseases including genetic diseases are highlighted in the second part of the chapter. The accessibility and standardization of -omic data are a prerequisite for the FAIR (Findability, Accessibility, Interoperability, and Reusability) guiding principles for scientific data management.

1.1 Introduction

The extracellular matrix (ECM) is the complex scaffold made of hundreds of proteins that governs the organization of cells and tissues in all multicellular organisms (Mecham 2011; Hynes and Yamada 2012). The ECM plays structural roles by conferring biomechanical properties to tissues and organs, controlling apico-basal cell polarity, and by serving as a substrate for cell migration. The ECM also exerts

A. Naba (✉)
Department of Physiology and Biophysics, University of Illinois at Chicago, Chicago, IL, USA
e-mail: anaba@uic.edu

S. Ricard-Blum (✉)
University of Lyon, Université Claude Bernard Lyon 1, CNRS, INSA Lyon, CPE, Institute of Molecular and Supramolecular Chemistry and Biochemistry, UMR 5246, Villeurbanne, France
e-mail: sylvie.ricard-blum@univ-lyon1.fr

S. Ricard-Blum (ed.), *Extracellular Matrix Omics*, Biology of Extracellular Matrix 7,
https://doi.org/10.1007/978-3-030-58330-9_1

signaling roles either directly by interacting with cell surface receptors including integrins, discoidin-domain receptors and heparan sulfate proteoglycans such as syndecans, or by interacting with growth factors and modulating their signaling activities (Hynes 2009). The ECM is also a source of bioactive fragments called matricryptins, which are released upon ECM remodeling and exert biological activities of their own (Parks and Mecham 2011; Ricard-Blum and Vallet 2019). Signals from the ECM have been shown to regulate gene expression and control cell proliferation, survival, cell mechanics, and differentiation (Rozario and DeSimone 2010). The pleiotropic roles of the ECM are at play in developmental processes such as gastrulation and cell specification (DeSimone and Mecham 2013; Dzamba and DeSimone 2018), physiological processes such as wound healing and aging (Karamanos et al. 2019), and pathological processes such as fibrosis and cancer (Brekken and Stupack 2017; Iozzo and Gubbiotti 2018; Zhou et al. 2018; Kai et al. 2019; Theocharis et al. 2019; Socovich and Naba 2019; Ricard-Blum and Miele 2019). Because of its supportive and instructive roles, the ECM is a central component of tissue engineering approaches for regenerative medicine (Berardi 2018). It also represents a source of biomarkers and potential novel therapeutic targets to respectively diagnose and predict, or alleviate and perhaps cure human diseases.

High-throughput profiling and screening methods that have emerged over the past two decades have radically transformed biomedical research. Such approaches rely on the unbiased identification of differences in gene expression or genetic states (genomics or transcriptomics) or protein abundance or protein states (proteomics) across different conditions. Once experimental data are acquired, their analysis can be divided in three major steps: (1) the definition of lists of genes or proteins of interest based on statistical cut-offs, (2) the identification of networks or pathways involving genes or proteins defined in (1), and (3) data visualization. Yet the true power of high-throughput approaches can only fully be harnessed if powerful computational tools and databases to comprehensively annotate the vast amount of data generated are available. To that end, freely available general databases such as UniProt (The UniProt Consortium 2019) or Gene Ontology (Attrill et al. 2019; The Gene Ontology Consortium 2019), and pathway analysis platforms such as the Database for Annotation, Visualization and Integrated Discovery (DAVID) (Jiao et al. 2012), Gene Set Enrichment Analysis (GSEA) and its Molecular Signature database (Subramanian et al. 2005, 2007; Liberzon et al. 2015), Cytoscape (Su et al. 2014), The Reactome Pathway Knowledgebase (Fabregat et al. 2018), or FunRich (Pathan et al. 2015), have been developed and include useful information on ECM genes and proteins (Table 1.1). However, these databases and tools rely on computational predictions and/or manual curation of experimental data from the literature. Consequently, systems or pathways more broadly studied will be more robustly and extensively annotated.

Having observed that ECM genes and proteins tended to be under-represented or mis-annotated in databases (Naba et al. 2012b), we previously proposed to define the "matrisome" as the compendium of genes encoding ECM and ECM-associated proteins (Hynes and Naba 2012; Naba et al. 2012a). Our work has revealed that 4% of the human genome encodes the matrisome. For more information, our readers

Table 1.1 Public databases for gene and protein annotations, enrichment analyses and pathway identifications

Resources/Tools	Description	References
Gene and proteins annotations		
Gene Ontology (GO) Resource http://geneontology.org/	Annotations of biological process, cellular component, and molecular function of genes and gene products.	Attrill et al. (2019), The Gene Ontology Consortium (2019)
UniProtKB https://www.uniprot.org/	The Universal Protein knowledgebase provides detailed annotations extracted from the literature by expert curators.	The UniProt Consortium (2019)
Bioinformatic tools to visualize and analyze omic data		
Cytoscape and Apps https://apps.cytoscape.org/ https://apps.cytoscape.org/apps/all	An open source platform for visualizing complex networks and integrating these with any type of attribute data. A lot of Apps are available.	Shannon et al. (2003), Saito et al. (2012), Su et al. (2014)
The Database for Annotation, Visualization and Integrated Discovery (DAVID) https://david.ncifcrf.gov/	DAVID provides a comprehensive set of functional annotation tools to understand biological meaning behind large list of genes	Huang et al. (2009), Jiao et al. (2012)
EnrichmentMap http://apps.cytoscape.org/apps/enrichmentmap	Visualizes enrichments of pathways as an enrichment map, a network representing overlaps among enriched pathways.	Merico et al. (2010)
Functional Enrichment analysis tool (FunRich) http://www.funrich.org/	Functional enrichment and interaction network analysis of genes and proteins	Pathan et al. (2015)
g:profiler https://biit.cs.ut.ee/gprofiler/gost	A web server for functional enrichment analysis and conversions of gene lists	Raudvere et al. (2019)
GSEA (gene set enrichment Analysis) https://www.gseamsigdb.org/	Suite of tools for the interpretation of gene expression data, enrichment analysis of gene lists	Subramanian et al. (2005, 2007)
The Omics Discovery Index (OmicsDI) http://www.omicsdi.org	An open source platform that enables access, discovery and dissemination of omics datasets	Perez-Riverol et al. (2017, 2019)
The Reactome Pathway Knowledgebase https://reactome.org/	A manually curated and peer-reviewed pathway database. Visualization, interpretation and analysis of pathway knowledge	Fabregat et al. (2018)

can find in Chap. 2 by Gebauer and Naba an overview of the computational approaches we and others have developed to predict the matrisome of different model organisms and illustrate their application for big data annotation. With this new framework, it has become easier to identify ECM genes and proteins in -omic datasets, and this has permitted ECM research to truly enter the -omics era (Naba et al. 2016).

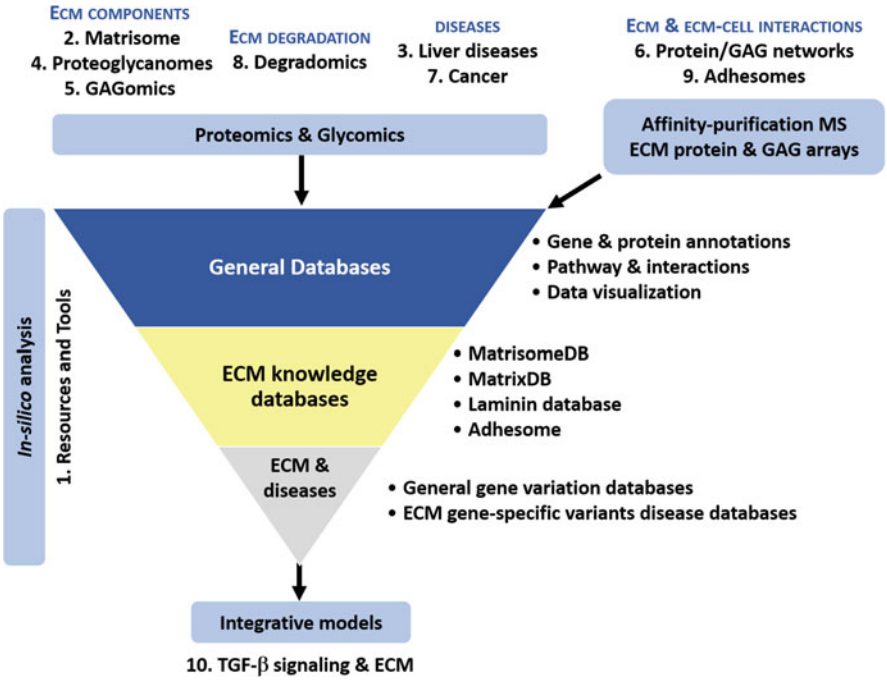

Fig. 1.1 Flow chart of the integrative multi-omic approach used to analyze the extracellular matrix (ECM), including the matrisome (proteomics), glycosaminoglycans (GAG, GAGomics), ECM and ECM-cell interactions (interactomics), ECM degradation (degradomics) and changes in the ECM associated with diseases. The flow chart also presents the tools used to analyze ECM -omic datasets (e.g. general and ECM-specific databases (DB)) in order to build integrative models of biological processes

The purpose of this introductory chapter is to present resources and tools developed to facilitate the identification and analysis of ECM genes and proteins using high-throughput methodologies (Fig. 1.1). In the second part of the chapter, we highlight databases focused on specific ECM genes and ECM-related diseases including genetic diseases and fibroses. Last, we briefly discuss the importance of the accessibility and standardization of -omic data, which is a prerequisite for the FAIR (Findability, Accessibility, Interoperability, and Reusability) guiding principles for scientific data management (Wilkinson et al. 2016).

1.2 ECM Knowledge Databases

Table 1.2 provides a list of ECM-focused databases and resources that can further assist the identification and annotation of ECM genes and proteins and their integration into interaction networks.

Table 1.2 Databases for ECM gene/protein annotations and ECM interaction data

Database name and web site	URL	References
Adhesome	http://www.adhesome.org	Winograd-Katz et al. (2014)
Gene Ontology (GO) Resource Extracellular matrix	http://www.informatics.jax.org/vocab/gene_ontology/GO:0031012	The Gene Ontology Consortium (2019)
The Laminin Database	http://www.lm.lncc.br	Golbert et al. (2014)
The Matrisome Project	http://matrisome.org	Naba et al. (2016)
MatrisomeDB	http://pepchem.org/matrisomedb	Shao et al. (2019)
MatrixDB	http://matrixdb.univ-lyon1.fr	Chautard et al. (2009, 2011), Launay et al. (2015), Clerc et al. (2019)

1.2.1 The Matrisome Project and MatrisomeDB

In order to distribute freely the bioinformatic tools and workflows we devised, which numerous groups have used to predict computationally the matrisomes of model organisms, we built The Matrisome Project website (http://matrisome.org) (Naba et al. 2016). The Matrisome Annotator application and the underlying open-access R-script available via The Matrisome Project have been used to annotate transcriptomic datasets which led to the identification ECM signatures of various diseased states as illustrated in recent studies (Hiebert et al. 2018; Izzi et al. 2018, 2019; Bin Lim et al. 2019; Etich et al. 2019) and as further discussed in Chap. 2 of this book by Gebauer and Naba.

ECM proteins have biochemical properties that distinguish them from intracellular proteins, such as their relative insolubility due to cross-linking and high levels of glycosylation. As a result, ECM proteins are largely under-represented in global proteomic datasets. New mass-spectrometry-based approaches have been specifically designed to study the ECM protein composition of healthy and pathological tissues. While it is beyond the scope of this chapter to detail these techniques, we invite our readers to refer to recent reviews that discuss technical aspects of ECM proteomic approaches (Randles et al. 2017; Lindsey et al. 2018; Raghunathan et al. 2019; Taha and Naba 2019). Over the past half-decade, proteomics has been applied to profile the protein composition of dozens of healthy and pathological tissues. During this same period, the scientific community has been increasingly supportive of the public release of -omic datasets to increase reproducibility. We thus developed MatrisomeDB (http://pepchem.org/matrisomedb), a searchable database compiling ECM proteomic datasets (Shao et al. 2019) that can be interrogated for specific proteins, protein signatures, tissues or diseases. It is our hope that this database will facilitate, if not accelerate, discoveries of ECM proteins playing so far unsuspected roles in pathophysiology.

1.2.2 The Laminin Database

Laminins are the main glycoproteins of basement membranes (Aumailley 2013; Hohenester and Yurchenco 2013), playing structural roles in the maintenance of tissue organization and integrity as well as signaling roles inducing changes in gene expression (Schéele et al. 2007). The laminin database (http://www.lm.lncc.br/) was devised by De Vasconcelos and colleagues (Golbert et al. 2014) and provides a comprehensive overview of this family of 16 isoforms in mammals, their receptors, and interacting proteins (Domogatskaya et al. 2012).

1.2.3 MatrixDB, the ECM Interaction Database

MatrixDB (http://matrixdb.univ-lyon1.fr/), the extracellular matrix database, is a member of the International Molecular Exchange (IMEx) consortium (Orchard et al. 2012), and its interaction data are also available via the IntAct database (Orchard et al. 2014). MatrixDB is focused on interactions between ECM proteins, proteoglycans, and glycosaminoglycans (Chautard et al. 2009, 2011; Launay et al. 2015; Clerc et al. 2019). MatrixDB manually extracts interaction data from publications using IMEX curation rules and a controlled vocabulary, which are regularly updated by the Molecular Interactions group of the Human Proteome Organization-Proteomics Standards Initiative (HUPO-PSI), and provides interaction data in MITAB format. This ensures consistency in data report and in data exchange between the databases of the consortium. MatrixDB curates interactions established by monomeric ECM proteins (e.g. isolated α chain from collagens, or elastin) identified by UniProt accession numbers (The UniProt Consortium 2019), ECM oligomeric proteins (e.g. native trimeric collagen or laminin molecules and the dimeric integrin receptors) identified by Complex Portal accession numbers (Meldal et al. 2019), and by bioactive fragments (i.e. matricryptins/matrikines) released upon ECM remodeling, which have biological activities and interaction repertoires of their own (Ricard-Blum and Vallet 2019), and are identified by the PRO feature of UniProtKB. In addition to protein-protein interactions, MatrixDB also curates protein-glycosaminoglycan interactions, which are of critical importance for ECM assembly and functions. Identifiers of Chemical Entities of Biological Interest, ChEBI, (Hastings et al. 2016) are used for glycosaminoglycans. The use of accession numbers of other publicly available databases allows cross-referencing and interoperability between databases. MatrixDB offers to the users the possibility to build and expand interaction networks using its graphical navigator, and to apply filters to build sub-networks based on a list of biomolecules, an interaction detection method and/or tissue expression level thanks to the integration of gene expression data (Launay et al. 2015; Clerc et al. 2019). The recent update of MatrixDB query interface allows users to build lists of proteins of interest by combining queries in order to build interaction networks specific of a disease, a biological process, or a

molecular function (Clerc et al. 2019). In its latest release, MatrixDB has integrated experimental proteomics data from MatrisomeDB to further provide an in-vivo context to interactions identified experimentally in vitro (Clerc et al. 2019) and Chap. 6 of this book by Ricard-Blum. The integration of proteomic data collected in various basement membranes into the comprehensive interaction network of basement membranes has allowed us to define the core interactome common to all the basement membranes studied, and the interactions, which are specific of each basement membrane (Clerc et al. 2019).

The AVEXIS Receptor Network with Integrated Expression database (ARNIE; https://www.sanger.ac.uk/resources/databases/arnie/) provides an extracellular protein interactome of zebrafish receptor and secreted proteins containing immunoglobulin and leucine-rich repeats with spatiotemporal expression patterns for all genes in order to identify signaling pathways playing a key role in zebrafish development (Martin et al. 2010).

1.2.4 The Consensus Adhesome

Proper cell-ECM interaction is critical to initiate ECM-dependent signal- and mechano-transduction leading to the modulation of gene expression and cellular phenotypes. Deciphering the molecular mechanisms involved in cell-ECM interactions is thus of utmost importance. The term "adhesome" was originally coined by Hynes and colleagues to define the "complement of adhesion-related genes and proteins" in echinoderms (Whittaker et al. 2006). Geiger and colleagues revisited the definition of this term to refer more specifically to the ensemble of proteins computationally predicted to localized at, or to regulate, integrin adhesion complexes (Zaidel-Bar et al. 2007). The adhesome website (http://www.adhesome.org/) provides a resource for exploring the components predicted to localize to focal adhesion complexes (Winograd-Katz et al. 2014). Recent advanced in proteomic methods have permitted the experimental characterization of the adhesomes of various integrins and be compared with the predicted adhesome (Horton et al. 2015, 2016) (see Chap. 9 of this book by Koeleman et al.).

1.3 Resources for the Study of ECM-Related Diseases

Mutations in ECM genes cause a broad spectrum of inherited diseases altering ECM structure, organization and functions, and targeting primarily skin, the neuromuscular, skeletal, and cardio-vascular systems (Bateman et al. 2009; Lamandé and Bateman 2019). ECM gene variations can be retrieved from general variant databases listed in Table 1.3. For example, ClinVar compiles public archives of reports on the relationships between human gene variations and phenotypes and includes 430,000 unique variants (Landrum et al. 2014, 2020). The Online

Table 1.3 General and ECM-specific gene variation databases and associated diseases

Databases and Resources	Description	References
General gene variation databases		
ClinVar https://www.ncbi.nlm.nih.gRelaov/clinvar/	Relationships among human variations and phenotypes, with supporting evidence.	Landrum et al. (2014, 2020)
LOVD[3] v.3.0 Leiden Open Variation Database https://www.lovd.nl/	Gene-centered collection and display of DNA variants, including genes coding for ECM proteins	Fokkema et al. (2011)
Online Mendelian Inheritance in Man (OMIM®) https://www.omim.org/	Information on all known Mendelian disorders and over 15,000 genes. OMIM focuses on the relationship between phenotype and genotype.	Amberger et al. (2019)
Effect of sequence changes and mutations of genes on molecular interactions		
The IMEx Consortium mutations data set https://www.ebi.ac.uk/intact/resources/datasets#mutationDs	Effect of sequences changes on experimental molecular interactions	IMEx Consortium Curators et al. (2019)
ECM gene-specific variants and disease databases		
Alport Database http://arup.utah.edu/database/ALPORT/	A database of variants in the *COL4A5* gene and their clinical significance.	Crockett et al. (2010)
COL7A1 gene mutations database http://www.col7a1-database.info/new/	Mutations of the *COL7A1* gene	Wertheim-Tysarowska et al. (2012)
The International registry of dystrophic epidermolysis bullosa patients and associated *COL7A1* mutations https://www.deb-central.org/	Mutations of the *COL7A1* gene in dystrophic epidermolysis bullosa patients	van den Akker et al. (2011)
Laminins and neuromuscular disorders http://www.lm.lncc.br		Golbert et al. (2014)
Osteogenesis imperfecta & Ehlers-Danlos syndrome variant database https://www.le.ac.uk/genetics/collagen/	Variants occurring in collagen I, III, and V genes, and in other genes coding for ECM proteins, and enzymes	Dalgleish (1997, 1998)
A structurally-integrated database for mutations of *PLOD* genes http://fornerislab.unipv.it/SiMPLOD/	Mutations in the PLOD/LH (Procollagen-Lysine 2-Oxoglutarate 5-Dioxygenase/Lysyl-Hydroxylase) enzyme family (LH1/PLOD1, LH2/PLOD2 and LH3/PLOD3)	Scietti et al. (2019)
The UMD-FBN1 mutations database http://www.umd.be/FBN1/	Mutations in fibrillin-1 (*FBN1*) gene (Marfan syndrome and associated disorders)	Collod-Béroud et al. (2003)

Mendelian Inheritance in Man (OMIM®), provides an online catalog of human genes and genetic disorders, containing over 24,600 entries, and 6259 molecularized phenotypes connected to 3961 genes (Amberger et al. 2019), and can be specifically interrogated for matrisome genes. Similarly, the Leiden Open-source Variation

Database (LOVD) is a freely available web-based interface compiling DNA variants compiled from various locus-specific databases (Fokkema et al. 2011). Gene-specific databases (e.g. focused on genes coding for ECM proteins) have been created with LOVD. The use of the databases maintained by the LOVD group is recommended by the Human Variome Project (https://www.humanvariomeproject.org/), an international non-governmental organisation working in collaboration with UNESCO to ensure that all information on genetic variations in the human genome and their effect on human health can be collected, curated, interpreted and shared (Burn and Watson 2016). In addition, the IMEx Consortium has built a mutation dataset generated from deep-curation, featuring 28,000 annotations describing the effect of small sequence changes on physical protein interactions (IMEx Consortium Curators et al. 2019). This data set of protein sequence changes or mutations and their effect over interaction outcome can be queried to evaluate the impact of sequence changes in ECM genes on ECM protein interactions.

Databases focused on ECMopathies, briefly described below, are also being developed (Table 1.3).

Collagenopathies refer to congenital disorders resulting from mutations in collagen genes that affect connective tissues (Jobling et al. 2014; Lamandé et al. 2017; Lamandé and Bateman 2018). A database reports ECM gene variants, including collagen genes, linked to osteogenesis imperfecta and Ehlers-Danlos syndromes (Dalgleish 1997, 1998). These syndromes can be caused by mutations in the human *PLOD1*, *PLOD2*, and *PLOD3* genes (procollagen-lysine, 2-oxoglutarate 5-dioxygenases 1, 2 and 3), that catalyzes the hydroxylation of lysine residues, a key post-translational modification of collagens. Their genetic variants are reported in the Structurally-integrated database for Mutations of *PLOD* genes (SiMPLOD), which allows the mapping of *PLOD* mutations on the experimentally determined X-ray structure of human PLOD3, and on the computational homology models of human PLOD1 and PLOD2 (Scietti et al. 2019),

The Alport database records 807 variants of *COL4A5* gene associated with X-linked Alport Syndrome, which primarily targets the kidney, and their clinical significance (Crockett et al. 2010). The COL7A1 gene variants database provides not only the possibility to search mutations but also a graphic view of mutations, general information on the COL7A1 gene sequence, the COL7A1 gene in other databases and other organisms, and bioinformatic tools to analyze mutations and design primers (Wertheim-Tysarowska et al. 2012).

Diseases caused by laminin mutations affect various systems in which the integrity of basement membranes is of paramount importance, including the muscular, nervous, and renal systems (Schéele et al. 2007; Chew and Lennon 2018). The laminin database presented above now integrates sections on the neuromuscular disorders resulting from laminin mutations, and miRNA-laminin relationships to account for the putative involvement of microRNAs in these disorders (Golbert et al. 2014). While many other diseases find their cause in ECM gene mutations, such as COMPopathies due to mutations in the *COMP* gene coding for the Cartilage Oligomeric Protein affecting the skeletal system (Posey et al. 2018) and Marfan syndrome (Meester et al. 2017), over-production or degradation of the ECM has also

been associated with the etiology of several diseases including fibroses, cardiovascular diseases, and cancers (Theocharis et al. 2019).

1.4 Conclusions and Perspectives

Expanding on this introductory chapter, our readers will find in this book examples of experimental -omic approaches devised to study proteoglycanomes (see Chap. 4 by Koch and Apte), and glycosaminoglycans (see Chap. 5 by Sethi and Zaia), biological processes such as ECM degradation (see Chap. 8 by Kalogeropoulos et al.) and specific pathological processes such as hepatic fibrosis (see Chap. 3 by Dolin et al.) and cancer (see Chap. 7 by Izzi et al.). Readers will also find examples of how to build interaction networks within the ECM (see Chap. 6 by Ricard-Blum) or at the interface between cell-ECM interactions (see Chap. 9 by Koeleman et al.) and methods to further integrate -omic level data to build models of ECM-dependent signaling pathways (see Chap. 10 by Théret et al.).

The generation of temporally and spatially-resolved data is now the next challenge of the -omic era, especially for ECM research (Bingham et al. 2020). The multi-omic approach, combining different omics (e.g. genomics, transcriptomics, proteomics, metabolomics and interactomics to name a few) is a pre-requisite to fully understand complex biological systems (Zhang and Kuster 2019). Tools allowing the integration of multi-omics data have started to emerge (Nagaraj et al. 2015; Ruggles et al. 2017; Subramanian et al. 2020) and the recently developed Knowledge Base Commons (https://kbcommons.org/) provides a universal framework for multi-omics data integration and biological discoveries (Zeng et al. 2019). The Omics Discovery Index (OmicsDI), that enables access, discovery and dissemination of 454,200 proteomics, genomics, metabolomics, transcriptomics, and multiomic datasets from 16 public resources is another valuable tool for multi-omic approaches (Perez-Riverol et al. 2017, 2019). An example of how powerful such integration can be is illustrated in a recent study from Schlotter and collaborators, where they correlated post-operative molecular imaging and pathology with proteomics, transcriptomics, and multi-dimensional network analysis to build the first integrated map of human calcific aortic valve disease (Schlotter et al. 2018). This spatiotemporal multi-omics mapping led to the identification of the first molecular regulatory networks in this disease, and showed that both structural ECM proteins (e.g. fibronectin, vitronectin and PCPE-2) and ECM proteins secreted by valvular interstitial cells (e.g. tenascin C and SPARC) contribute to the calcification propensity of the fibrosa layer (Schlotter et al. 2018).

The few examples reviewed above show how powerful the -omic approach coupled to data sharing is. With the decreased cost and increased availability of genomics, transcriptomics, and proteomics as well as computational tools to mine large datasets generated by these techniques, we can anticipate an increase in the number of datasets, including some focused on the ECM, that will become available to the scientific community. It is our hope that the community-built open-access and

open-source resources similar to those presented in this chapter will lead to faster biomedical breakthroughs.

Acknowledgments The authors would like to thank Martin Davis from the Naba lab for his critical reading of the manuscript.
Funding Sources

The work of the authors reported here was supported by a start-up fund from the Department of Physiology and Biophysics to and in part by a Catalyst Award from the Chicago Biomedical Consortium with support from the Searle Funds at the Chicago Community Trust (Grant C-088) to AN, and by the "Fondation pour la Recherche Médicale, France [grant number DBI20141231336 to SRB], the Institut Français de Bioinformatique [ANR-11-INBS-0013, Glycomatrix project, call 2015 to SRB], and the Groupement de Recherche (GDR) GagoSciences [CNRS, GDR 3739, Structure, Fonction et Régulation des Glycosaminoglycanes to SRB].

References[1]

Amberger JS, Bocchini CA, Scott AF, Hamosh A (2019) OMIM.org: leveraging knowledge across phenotype-gene relationships. Nucleic Acids Res 47:D1038–D1043. https://doi.org/10.1093/nar/gky1151

Attrill H, Gaudet P, Huntley RP, Lovering RC, Engel SR, Poux S, Van Auken KM, Georghiou G, Chibucos MC, Berardini TZ, Wood V, Drabkin H, Fey P, Garmiri P, Harris MA, Sawford T, Reiser L, Tauber R, Toro S (2019) Annotation of gene product function from high-throughput studies using the gene ontology. Database (Oxford) 2019. https://doi.org/10.1093/database/baz007

Aumailley M (2013) The laminin family. Cell Adh Migr 7:48–55. https://doi.org/10.4161/cam.22826

Bateman JF, Boot-Handford RP, Lamandé SR (2009) Genetic diseases of connective tissues: cellular and extracellular effects of ECM mutations. Nat Rev Genet 10:173–183. https://doi.org/10.1038/nrg2520

Berardi AC (ed) (2018) Extracellular matrix for tissue engineering and biomaterials. Humana Press, Totowa, NJ

Bin Lim S, Chua MLK, Yeong JPS, Tan SJ, Lim W-T, Lim CT (2019) Pan-cancer analysis connects tumor matrisome to immune response. NPJ Precis Oncol 3:15. https://doi.org/10.1038/s41698-019-0087-0

Bingham GC, Lee F, Naba A, Barker TH (2020) Spatial-omics: novel approaches to probe cell heterogeneity and ECM biology. Matrix Biol 91-92:152–166.

Brekken RA, Stupack D (eds) (2017) Extracellular matrix in tumor biology. Springer, Berlin

Burn J, Watson M (2016) The Human Variome Project. Hum Mutat 37:505–507. https://doi.org/10.1002/humu.22986

Chautard E, Ballut L, Thierry-Mieg N, Ricard-Blum S (2009) MatrixDB, a database focused on extracellular protein–protein and protein–carbohydrate interactions. Bioinformatics 25:690–691. https://doi.org/10.1093/bioinformatics/btp025

Chautard E, Fatoux-Ardore M, Ballut L, Thierry-Mieg N, Ricard-Blum S (2011) MatrixDB, the extracellular matrix interaction database. Nucleic Acids Res 39:D235–D240. https://doi.org/10.1093/nar/gkq830

Chew C, Lennon R (2018) Basement membrane defects in genetic kidney diseases. Front Pediatr 6. https://doi.org/10.3389/fped.2018.00011

[1]We apologize to the authors of studies we were not able to cite due to space limitations.

Clerc O, Deniaud M, Vallet SD, Naba A, Rivet A, Perez S, Thierry-Mieg N, Ricard-Blum S (2019) MatrixDB: integration of new data with a focus on glycosaminoglycan interactions. Nucleic Acids Res 47:D376–D381. https://doi.org/10.1093/nar/gky1035

Collod-Béroud G, Le Bourdelles S, Ades L, Ala-Kokko L, Booms P, Boxer M, Child A, Comeglio P, De Paepe A, Hyland JC, Holman K, Kaitila I, Loeys B, Matyas G, Nuytinck L, Peltonen L, Rantamaki T, Robinson P, Steinmann B, Junien C, Béroud C, Boileau C (2003) Update of the UMD-FBN1 mutation database and creation of an FBN1 polymorphism database. Hum Mutat 22:199–208. https://doi.org/10.1002/humu.10249

Crockett DK, Pont-Kingdon G, Gedge F, Sumner K, Seamons R, Lyon E (2010) The Alport syndrome COL4A5 variant database. Hum Mutat 31:E1652–E1657. https://doi.org/10.1002/humu.21312

Dalgleish R (1997) The human type I collagen mutation database. Nucleic Acids Res 25:181–187. https://doi.org/10.1093/nar/25.1.181

Dalgleish R (1998) The Human Collagen Mutation Database 1998. Nucleic Acids Res 26:253–255. https://doi.org/10.1093/nar/26.1.253

DeSimone DW, Mecham R (eds) (2013) Extracellular matrix in development. Springer, Berlin

Domogatskaya A, Rodin S, Tryggvason K (2012) Functional diversity of laminins. Annu Rev Cell Dev Biol 28:523–553. https://doi.org/10.1146/annurev-cellbio-101011-155750

Dzamba BJ, DeSimone DW (2018) Extracellular matrix (ECM) and the sculpting of embryonic tissues. Curr Top Dev Biol 130:245–274. https://doi.org/10.1016/bs.ctdb.2018.03.006

Etich J, Koch M, Wagener R, Zaucke F, Fabri M, Brachvogel B (2019) Gene expression profiling of the extracellular matrix signature in macrophages of different activation status: relevance for skin wound healing. Int J Mol Sci 20. https://doi.org/10.3390/ijms20205086

Fabregat A, Jupe S, Matthews L, Sidiropoulos K, Gillespie M, Garapati P, Haw R, Jassal B, Korninger F, May B, Milacic M, Roca CD, Rothfels K, Sevilla C, Shamovsky V, Shorser S, Varusai T, Viteri G, Weiser J, Wu G, Stein L, Hermjakob H, D'Eustachio P (2018) The Reactome Pathway Knowledgebase. Nucleic Acids Res 46:D649–D655. https://doi.org/10.1093/nar/gkx1132

Fokkema IFAC, Taschner PEM, Schaafsma GCP, Celli J, Laros JFJ, den Dunnen JT (2011) LOVD v.2.0: the next generation in gene variant databases. Hum Mutat 32:557–563. https://doi.org/10.1002/humu.21438

Golbert DCF, Santana-van-Vliet E, Mundstein AS, Calfo V, Savino W, de Vasconcelos ATR (2014) Laminin-database v.2.0: an update on laminins in health and neuromuscular disorders. Nucleic Acids Res 42:D426–D429. https://doi.org/10.1093/nar/gkt901

Hastings J, Owen G, Dekker A, Ennis M, Kale N, Muthukrishnan V, Turner S, Swainston N, Mendes P, Steinbeck C (2016) ChEBI in 2016: improved services and an expanding collection of metabolites. Nucleic Acids Res 44:D1214–D1219. https://doi.org/10.1093/nar/gkv1031

Hiebert P, Wietecha MS, Cangkrama M, Haertel E, Mavrogonatou E, Stumpe M, Steenbock H, Grossi S, Beer H-D, Angel P, Brinckmann J, Kletsas D, Dengjel J, Werner S (2018) Nrf2-mediated fibroblast reprogramming drives cellular senescence by targeting the matrisome. Dev Cell 46:145–161.e10. https://doi.org/10.1016/j.devcel.2018.06.012

Hohenester E, Yurchenco PD (2013) Laminins in basement membrane assembly. Cell Adh Migr 7:56–63. https://doi.org/10.4161/cam.21831

Horton ER, Byron A, Askari JA, Ng DHJ, Millon-Frémillon A, Robertson J, Koper EJ, Paul NR, Warwood S, Knight D, Humphries JD, Humphries MJ (2015) Definition of a consensus integrin adhesome and its dynamics during adhesion complex assembly and disassembly. Nat Cell Biol 17:1577–1587. https://doi.org/10.1038/ncb3257

Horton ER, Humphries JD, James J, Jones MC, Askari JA, Humphries MJ (2016) The integrin adhesome network at a glance. J Cell Sci 129:4159–4163. https://doi.org/10.1242/jcs.192054

Huang DW, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4:44–57. https://doi.org/10.1038/nprot.2008.211

Hynes RO (2009) The extracellular matrix: not just pretty fibrils. Science 326:1216–1219. https://doi.org/10.1126/science.1176009

Hynes RO, Naba A (2012) Overview of the matrisome - an inventory of extracellular matrix constituents and functions. Cold Spring Harb Perspect Biol 4:a004903–a004903. https://doi.org/10.1101/cshperspect.a004903

Hynes RO, Yamada KM (2012) Extracellular matrix biology, Cold Spring Harbor perspectives in biology. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

IMEx Consortium Curators, Del-Toro N, Duesbury M, Koch M, Perfetto L, Shrivastava A, Ochoa D, Wagih O, Piñero J, Kotlyar M, Pastrello C, Beltrao P, Furlong LI, Jurisica I, Hermjakob H, Orchard S, Porras P (2019) Capturing variation impact on molecular interactions in the IMEx Consortium mutations data set. Nat Commun 10:10. https://doi.org/10.1038/s41467-018-07709-6

Iozzo RV, Gubbiotti MA (2018) Extracellular matrix: the driving force of mammalian diseases. Matrix Biol 71–72:1–9. https://doi.org/10.1016/j.matbio.2018.03.023

Izzi V, Lakkala J, Devarajan R, Savolainen E-R, Koistinen P, Heljasvaara R, Pihlajaniemi T (2018) Expression of a specific extracellular matrix signature is a favorable prognostic factor in acute myeloid leukemia. Leuk Res Rep 9:9–13. https://doi.org/10.1016/j.lrr.2017.12.001

Izzi V, Lakkala J, Devarajan R, Kääriäinen A, Koivunen J, Heljasvaara R, Pihlajaniemi T (2019) Pan-Cancer analysis of the expression and regulation of matrisome genes across 32 tumor types. Matrix Biol Plus 1:100004. https://doi.org/10.1016/j.mbplus.2019.04.001

Jiao X, Sherman BT, Huang DW, Stephens R, Baseler MW, Lane HC, Lempicki RA (2012) DAVID-WS: a stateful web service to facilitate gene/protein list analysis. Bioinformatics 28:1805–1806. https://doi.org/10.1093/bioinformatics/bts251

Jobling R, D'Souza R, Baker N, Lara-Corrales I, Mendoza-Londono R, Dupuis L, Savarirayan R, Ala-Kokko L, Kannu P (2014) The collagenopathies: review of clinical phenotypes and molecular correlations. Curr Rheumatol Rep 16:394. https://doi.org/10.1007/s11926-013-0394-3

Kai F, Drain AP, Weaver VM (2019) The extracellular matrix modulates the metastatic journey. Dev Cell 49:332–346. https://doi.org/10.1016/j.devcel.2019.03.026

Karamanos NK, Theocharis AD, Neill T, Iozzo RV (2019) Matrix modeling and remodeling: a biological interplay regulating tissue homeostasis and diseases. Matrix Biol 75–76:1–11. https://doi.org/10.1016/j.matbio.2018.08.007

Lamandé SR, Bateman JF (2018) Collagen VI disorders: insights on form and function in the extracellular matrix and beyond. Matrix Biol 71–72:348–367. https://doi.org/10.1016/j.matbio.2017.12.008

Lamandé SR, Bateman JF (2019) Genetic disorders of the extracellular matrix. Anat Rec (Hoboken). https://doi.org/10.1002/ar.24086

Lamandé SR, Cameron TL, Savarirayan R, Bateman JF (2017) Molecular genetics of the cartilage collagenopathies. In: Grässel S, Aszódi A (eds) Cartilage, Pathophysiology, vol 2. Springer, Cham, pp 99–133

Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM, Maglott DR (2014) ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res 42:D980–D985. https://doi.org/10.1093/nar/gkt1113

Landrum MJ, Chitipiralla S, Brown GR, Chen C, Gu B, Hart J, Hoffman D, Jang W, Kaur K, Liu C, Lyoshin V, Maddipatla Z, Maiti R, Mitchell J, O'Leary N, Riley GR, Shi W, Zhou G, Schneider V, Maglott D, Holmes JB, Kattman BL (2020) ClinVar: improvements to accessing data. Nucleic Acids Res 48:D835–D844. https://doi.org/10.1093/nar/gkz972

Launay G, Salza R, Multedo D, Thierry-Mieg N, Ricard-Blum S (2015) MatrixDB, the extracellular matrix interaction database: updated content, a new navigator and expanded functionalities. Nucleic Acids Res 43:D321–D327. https://doi.org/10.1093/nar/gku1091

Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P (2015) The Molecular Signatures Database hallmark gene set collection. Cell Syst 1:417–425. https://doi.org/10.1016/j.cels.2015.12.004

Lindsey ML, Jung M, Hall ME, DeLeon-Pennell KY (2018) Proteomic analysis of the cardiac extracellular matrix: clinical research applications. Expert Rev Proteomics 15:105–112. https://doi.org/10.1080/14789450.2018.1421947

Martin S, Söllner C, Charoensawan V, Adryan B, Thisse B, Thisse C, Teichmann S, Wright GJ (2010) Construction of a large extracellular protein interaction network and its resolution by spatiotemporal expression profiling. Mol Cell Proteomics 9:2654–2665. https://doi.org/10.1074/mcp.M110.004119

Mecham R (ed) (2011) The extracellular matrix: an overview. Springer, Berlin

Meester JAN, Verstraeten A, Schepers D, Alaerts M, Van Laer L, Loeys BL (2017) Differences in manifestations of Marfan syndrome, Ehlers-Danlos syndrome, and Loeys-Dietz syndrome. Ann Cardiothorac Surg 6:582–594. https://doi.org/10.21037/acs.2017.11.03

Meldal BHM, Bye-A-Jee H, Gajdoš L, Hammerová Z, Horácková A, Melicher F, Perfetto L, Pokorný D, Lopez MR, Türková A, Wong ED, Xie Z, Casanova EB, Del-Toro N, Koch M, Porras P, Hermjakob H, Orchard S (2019) Complex Portal 2018: extended content and enhanced visualization tools for macromolecular complexes. Nucleic Acids Res 47:D550–D558. https://doi.org/10.1093/nar/gky1001

Merico D, Isserlin R, Stueker O, Emili A, Bader GD (2010) Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. PLoS ONE 5:e13984. https://doi.org/10.1371/journal.pone.0013984

Naba A, Clauser KR, Hoersch S, Liu H, Carr SA, Hynes RO (2012a) The matrisome: in silico definition and in vivo characterization by proteomics of normal and tumor extracellular matrices. Mol Cell Proteomics 11:M111.014647. https://doi.org/10.1074/mcp.M111.014647

Naba A, Hoersch S, Hynes RO (2012b) Towards definition of an ECM parts list: an advance on GO categories. Matrix Biol 31:371–372. https://doi.org/10.1016/j.matbio.2012.11.008

Naba A, Clauser KR, Ding H, Whittaker CA, Carr SA, Hynes RO (2016) The extracellular matrix: tools and insights for the "omics" era. Matrix Biol 49:10–24. https://doi.org/10.1016/j.matbio.2015.06.003

Nagaraj SH, Waddell N, Madugundu AK, Wood S, Jones A, Mandyam RA, Nones K, Pearson JV, Grimmond SM (2015) PGTools: a software suite for proteogenomic data analysis and visualization. J Proteome Res 14:2255–2266. https://doi.org/10.1021/acs.jproteome.5b00029

Orchard S, Kerrien S, Abbani S, Aranda B, Bhate J, Bidwell S, Bridge A, Briganti L, Brinkman FSL, Cesareni G, Chatr-aryamontri A, Chautard E, Chen C, Dumousseau M, Goll J, Hancock REW, Hannick LI, Jurisica I, Khadake J, Lynn DJ, Mahadevan U, Perfetto L, Raghunath A, Ricard-Blum S, Roechert B, Salwinski L, Stümpflen V, Tyers M, Uetz P, Xenarios I, Hermjakob H (2012) Protein interaction data curation: the International Molecular Exchange (IMEx) consortium. Nat Methods 9:345–350. https://doi.org/10.1038/nmeth.1931

Orchard S, Ammari M, Aranda B, Breuza L, Briganti L, Broackes-Carter F, Campbell NH, Chavali G, Chen C, del-Toro N, Duesbury M, Dumousseau M, Galeota E, Hinz U, Iannuccelli M, Jagannathan S, Jimenez R, Khadake J, Lagreid A, Licata L, Lovering RC, Meldal B, Melidoni AN, Milagros M, Peluso D, Perfetto L, Porras P, Raghunath A, Ricard-Blum S, Roechert B, Stutz A, Tognolli M, van Roey K, Cesareni G, Hermjakob H (2014) The MIntAct project--IntAct as a common curation platform for 11 molecular interaction databases. Nucleic Acids Res 42:D358–D363. https://doi.org/10.1093/nar/gkt1115

Parks WC, Mecham R (eds) (2011) Extracellular matrix degradation. Springer, Berlin

Pathan M, Keerthikumar S, Ang C-S, Gangoda L, Quek CYJ, Williamson NA, Mouradov D, Sieber OM, Simpson RJ, Salim A, Bacic A, Hill AF, Stroud DA, Ryan MT, Agbinya JI, Mariadason JM, Burgess AW, Mathivanan S (2015) FunRich: an open access standalone functional enrichment and interaction network analysis tool. Proteomics 15:2597–2601. https://doi.org/10.1002/pmic.201400515

Perez-Riverol Y, Bai M, da Veiga LF, Squizzato S, Park YM, Haug K, Carroll AJ, Spalding D, Paschall J, Wang M, Del-Toro N, Ternent T, Zhang P, Buso N, Bandeira N, Deutsch EW, Campbell DS, Beavis RC, Salek RM, Sarkans U, Petryszak R, Keays M, Fahy E, Sud M, Subramaniam S, Barbera A, Jiménez RC, Nesvizhskii AI, Sansone S-A, Steinbeck C, Lopez R,

Vizcaíno JA, Ping P, Hermjakob H (2017) Discovering and linking public omics data sets using the Omics Discovery Index. Nat Biotechnol 35:406–409. https://doi.org/10.1038/nbt.3790

Perez-Riverol Y, Zorin A, Dass G, Vu M-T, Xu P, Glont M, Vizcaíno JA, Jarnuczak AF, Petryszak R, Ping P, Hermjakob H (2019) Quantifying the impact of public omics data. Nat Commun 10:3512. https://doi.org/10.1038/s41467-019-11461-w

Posey KL, Coustry F, Hecht JT (2018) Cartilage oligomeric matrix protein: COMPopathies and beyond. Matrix Biol 71–72:161–173. https://doi.org/10.1016/j.matbio.2018.02.023

Raghunathan R, Sethi MK, Klein JA, Zaia J (2019) Proteomics, glycomics, and glycoproteomics of matrisome molecules. Mol Cell Proteomics. https://doi.org/10.1074/mcp.R119.001543

Randles MJ, Humphries MJ, Lennon R (2017) Proteomic definitions of basement membrane composition in health and disease. Matrix Biol 57–58:12–28. https://doi.org/10.1016/j.matbio.2016.08.006

Raudvere U, Kolberg L, Kuzmin I, Arak T, Adler P, Peterson H, Vilo J (2019) g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). Nucleic Acids Res 47:W191–W198. https://doi.org/10.1093/nar/gkz369

Ricard-Blum S, Miele AE (2019) Omic approaches to decipher the molecular mechanisms of fibrosis, and design new anti-fibrotic strategies. Semin Cell Dev Biol S1084952118302635. https://doi.org/10.1016/j.semcdb.2019.12.009

Ricard-Blum S, Vallet SD (2019) Fragments generated upon extracellular matrix remodeling: biological regulators and potential drugs. Matrix Biol 75–76:170–189. https://doi.org/10.1016/j.matbio.2017.11.005

Rozario T, DeSimone DW (2010) The extracellular matrix in development and morphogenesis: a dynamic view. Dev Biol 341:126–140. https://doi.org/10.1016/j.ydbio.2009.10.026

Ruggles KV, Krug K, Wang X, Clauser KR, Wang J, Payne SH, Fenyö D, Zhang B, Mani DR (2017) Methods, tools and current perspectives in proteogenomics. Mol Cell Proteomics 16:959–981. https://doi.org/10.1074/mcp.MR117.000024

Saito R, Smoot ME, Ono K, Ruscheinski J, Wang P-L, Lotia S, Pico AR, Bader GD, Ideker T (2012) A travel guide to Cytoscape plugins. Nat Methods 9:1069–1076. https://doi.org/10.1038/nmeth.2212

Schéele S, Nyström A, Durbeej M, Talts JF, Ekblom M, Ekblom P (2007) Laminin isoforms in development and disease. J Mol Med 85:825–836. https://doi.org/10.1007/s00109-007-0182-5

Schlotter F, Halu A, Goto S, Blaser MC, Body SC, Lee LH, Higashi H, DeLaughter DM, Hutcheson JD, Vyas P, Pham T, Rogers MA, Sharma A, Seidman CE, Loscalzo J, Seidman JG, Aikawa M, Singh SA, Aikawa E (2018) Spatiotemporal multi-omics mapping generates a molecular atlas of the aortic valve and reveals networks driving disease. Circulation 138:377–393. https://doi.org/10.1161/CIRCULATIONAHA.117.032291

Scietti L, Campioni M, Forneris F (2019) SiMPLOD, a structure-integrated database of collagen lysyl hydroxylase (LH/PLOD) enzyme variants. J Bone Miner Res 34:1376–1382. https://doi.org/10.1002/jbmr.3692

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13:2498–2504. https://doi.org/10.1101/gr.1239303

Shao X, Taha IN, Clauser KR, Gao Y (Tom), Naba A (2019) MatrisomeDB: the ECM-protein knowledge database. Nucleic Acids Res. https://doi.org/10.1093/nar/gkz849

Socovich AM, Naba A (2019) The cancer matrisome: from comprehensive characterization to biomarker discovery. Semin Cell Dev Biol 89:157–166. https://doi.org/10.1016/j.semcdb.2018.06.005

Su G, Morris JH, Demchak B, Bader GD (2014) Biological network exploration with Cytoscape 3. Curr Protoc Bioinformatics 47:8.13.1–8.1324. https://doi.org/10.1002/0471250953.bi0813s47

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP (2005) Gene set enrichment analysis: a

knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA 102:15545–15550. https://doi.org/10.1073/pnas.0506580102

Subramanian A, Kuehn H, Gould J, Tamayo P, Mesirov JP (2007) GSEA-P: a desktop application for gene set enrichment analysis. Bioinformatics 23:3251–3253. https://doi.org/10.1093/bioinformatics/btm369

Subramanian I, Verma S, Kumar S, Jere A, Anamika K (2020) Multi-omics data integration, interpretation, and its application. Bioinform Biol Insights 14:117793221989905. https://doi.org/10.1177/1177932219899051

Taha IN, Naba A (2019) Exploring the extracellular matrix in health and disease using proteomics. Essays Biochem:EBC20190001. https://doi.org/10.1042/EBC20190001

The Gene Ontology Consortium (2019) The Gene Ontology Resource: 20 years and still going strong. Nucleic Acids Res 47:D330–D338. https://doi.org/10.1093/nar/gky1055

The UniProt Consortium (2019) UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res 47:D506–D515. https://doi.org/10.1093/nar/gky1049

Theocharis AD, Manou D, Karamanos NK (2019) The extracellular matrix as a multitasking player in disease. FEBS J 286:2830–2869. https://doi.org/10.1111/febs.14818

van den Akker PC, Jonkman MF, Rengaw T, Bruckner-Tuderman L, Has C, Bauer JW, Klausegger A, Zambruno G, Castiglia D, Mellerio JE, McGrath JA, van Essen AJ, Hofstra RMW, Swertz MA (2011) The international dystrophic epidermolysis bullosa patient registry: an online database of dystrophic epidermolysis bullosa patients and their COL7A1 mutations. Hum Mutat 32:1100–1107. https://doi.org/10.1002/humu.21551

Wertheim-Tysarowska K, Sobczyńska-Tomaszewska A, Kowalewski C, Skroński M, Święćkowski G, Kutkowska-Kaźmierczak A, Woźniak K, Bal J (2012) The COL7A1 mutation database. Hum Mutat 33:327–331. https://doi.org/10.1002/humu.21651

Whittaker CA, Bergeron K-F, Whittle J, Brandhorst BP, Burke RD, Hynes RO (2006) The echinoderm adhesome. Dev Biol 300:252–266. https://doi.org/10.1016/j.ydbio.2006.07.044

Wilkinson MD, Dumontier M, Aalbersberg IjJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten J-W, Santos LB da S, Bourne PE, Bouwman J, Brookes AJ, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo CT, Finkers R, Gonzalez-Beltran A, Gray AJG, Groth P, Goble C, Grethe JS, Heringa J, Hoen PAC 't, Hooft R, Kuhn T, Kok R, Kok J, Lusher SJ, Martone ME, Mons A, Packer AL, Persson B, Rocca-Serra P, Roos M, Schaik R van, Sansone S-A, Schultes E, Sengstag T, Slater T, Strawn G, Swertz MA, Thompson M, Lei J van der, Mulligen E van, Velterop J, Waagmeester A, Wittenburg P, Wolstencroft K, Zhao J, Mons B (2016) The FAIR guiding principles for scientific data management and stewardship. Sci Data 3:1–9. https://doi.org/10.1038/sdata.2016.18

Winograd-Katz SE, Fässler R, Geiger B, Legate KR (2014) The integrin adhesome: from genes and proteins to human disease. Nat Rev Mol Cell Biol 15:273–288. https://doi.org/10.1038/nrm3769

Zaidel-Bar R, Itzkovitz S, Ma'ayan A, Iyengar R, Geiger B (2007) Functional atlas of the integrin adhesome. Nat Cell Biol 9:858–867. https://doi.org/10.1038/ncb0807-858

Zeng S, Lyu Z, Narisetti SRK, Xu D, Joshi T (2019) Knowledge Base Commons (KBCommons) v1.1: a universal framework for multi-omics data integration and biological discoveries. BMC Genomics 20:947. https://doi.org/10.1186/s12864-019-6287-8

Zhang B, Kuster B (2019) Proteomics is not an island: multi-omics integration is the key to understanding biological systems. Mol Cell Proteomics 18:S1–S4. https://doi.org/10.1074/mcp.E119.001693

Zhou Y, Horowitz JC, Naba A, Ambalavanan N, Atabai K, Balestrini J, Bitterman PB, Corley RA, Ding B-S, Engler AJ, Hansen KC, Hagood JS, Kheradmand F, Lin QS, Neptune E, Niklason L, Ortiz LA, Parks WC, Tschumperlin DJ, White ES, Chapman HA, Thannickal VJ (2018) Extracellular matrix in lung development, homeostasis and disease. Matrix Biol 73:77–104. https://doi.org/10.1016/j.matbio.2018.03.005

Chapter 2
The Matrisome of Model Organisms: From In-Silico Prediction to Big-Data Annotation

Jan M. Gebauer and Alexandra Naba

Abstract The extracellular matrix (ECM) is the architect of multicellular organisms. The use of various model organisms has helped decipher the mechanisms by which the ECM provides both chemical and mechanical cues that regulate fundamental cellular processes conserved during evolution such as cell migration, invasion, and differentiation.

High-throughput, or –omic, technologies has transfigured biomedical research. It is thus imperative to have a systematic way to identify ECM genes and proteins in large datasets. This requires having a comprehensive catalog of all the components that constitute the ECM. Here, we will describe the key structural features of ECM proteins (signal peptide, presence of protein domains, motifs, or repeats) that can be used to devise computational approaches to predict ECM proteins. We will then present fully automated machine-learning-based algorithms and approaches that have combined protein-sequence analysis and knowledge-based curation to define the matrisome of model organisms. Last, we provide examples of how the definition of the matrisome has facilitated the identification of ECM genes and proteins in –omic datasets and has advanced our understanding of the contribution of the ECM pathophysiological processes such as embryonic development, tissue regeneration, aging, and cancer.

2.1 Introduction

The extracellular matrix (ECM), a complex protein assembly, constitutes the architectural scaffold of all multicellular organisms. However, the collection of ECM proteins has greatly expanded throughout evolution to support the emergence of

J. M. Gebauer
Institute of Biochemistry, University of Cologne, Cologne, Germany
e-mail: jan.gebauer@uni-koeln.de

A. Naba (✉)
Department of Physiology and Biophysics, University of Illinois at Chicago, Chicago, IL, USA
e-mail: anaba@uic.edu

S. Ricard-Blum (ed.), *Extracellular Matrix Omics*, Biology of Extracellular Matrix 7,
https://doi.org/10.1007/978-3-030-58330-9_2

complex levels of cellular organization, and the development of novel tissues and specialized functions (Draper et al. 2019; Hynes 2012; Keeley and Mecham 2013; Özbek et al. 2010). Molecularly, this complexification has arisen from multiple mechanisms including exon shuffling, the apparition of novel protein domains, and expansion within families of genes (Hynes 2012; Keeley and Mecham 2013). Studying the ECM in various model organisms (mice, zebrafish, *Drosophila*, *C. elegans*, etc.) has been instrumental to understanding some of the fundamental mechanisms by which it controls evolutionarily conserved processes, such as cell proliferation, migration, invasion, lineage specification, and differentiation (Adams 2018; Brown 2011; Rozario and DeSimone 2010). Studying the ECM in model organisms has also shed light on its roles in cellular processes including wound healing and tissue repair, aging (Birch 2018), angiogenesis (Neve et al. 2014), fibrosis (Herrera et al. 2018), and cancer metastasis (Pickup et al. 2014).

The emergence of high-throughput screening technologies has transfigured biomedical research (*see Chap. 1 of this book*). It is thus imperative to have a systematic way to identify and annotate ECM genes and proteins in large datasets, if we want to be able to fully capture the extent of their involvement in pathophysiological contexts. This requires having a comprehensive catalog of all the components that constitute the ECM. In this chapter, we will discuss key structural features of ECM proteins that can be used to devise computational approaches to predict ECM proteins within proteomes. We will then present and discuss the limitations of fully automated machine-learning-based algorithms that have attempted to predict ECM proteins. Last, we will review approaches that have combined protein-sequence analysis and knowledge-based curation to define the human matrisome and the matrisomes of 6 model organisms: the mouse, the quail, zebrafish, *Drosophila*, *C. elegans*, and planarian. We will further briefly illustrate how the definition of these lists has greatly aided the identification of ECM genes and proteins in –omic datasets and has advanced our understanding of the contribution of the ECM to embryonic development and diseases.

2.2 Gene Ontology Annotations of ECM Proteins

The Gene Ontology (GO) database describes knowledge of proteins using "terms" with respect to three aspects: the cellular components they are found in, or localization; their molecular functions; and the biological processes they are involved in (The Gene Ontology Consortium 2019). Every annotation is based on either computational or phylogenetic evidences that may additionally be supported by experimental evidence (for more details, we invite our readers to refer to the GO website: http://geneontology.org). As of 2020, the database included more than 44,000 unique GO terms and provided annotations for over 1.3 million gene products or proteoforms. These annotations are cross-referenced to all major gene and protein databases. This colossal amount of data is instrumental for curating -omic datasets and identifying groups of co-regulated genes and proteins that localize to the same

compartment or play roles in the same cellular processes. Eventually, these annotations should support the building of system-wide views of given cellular states or physiological or pathological processes.

Query of the GO database for terms containing "extracellular matrix" retrieves 49 terms and an additional 137 terms are related to ECM biology. Some of these terms are as broad as "extracellular matrix" (GO:0031012) which is associated with over 2000 human gene products. Other terms remain vague, e.g. "positive regulation of extracellular matrix organization" (GO:1903055), and are associated to a smaller subset of human gene products (27, for GO:1903055, of which, some are encoding ECM proteins but some of them encoding intracellular components). However, if too broad, too granular, or too imprecise, such annotations can present limitations by either creating noise or failing to fully capture the entirety of a process and thus falls short in aiding with the extraction of relevant information in large datasets (Naba et al. 2012a). There is thus a need for a comprehensively-annotated compendium of ECM genes and proteins. To build it will require to take into account specific features of these proteins, including characteristic sequences found in ECM proteins, as well as their particular domain-based organization (see below).

2.3 Prototypical Organization of an ECM Protein

2.3.1 Protein Export and Signal-Peptide Prediction

As ECM proteins are a subset of secreted proteins, they share the same export mechanisms. Predicting protein export is thus an important first step in the identification of ECM proteins. Most secreted proteins are targeted to the extracellular space by an N-terminal signal peptide which is recognised by the signal recognition particle (SRP) during translation. Upon binding of the SRP, translation is halted and stalled ribosomes are transferred to the endoplasmatic reticulum (ER) where nascent protein chains are co-translationally secreted into the ER lumen. During or after translation, signal peptides are cleaved from mature proteins by a membrane standing protease. Many ECM proteins are post-translationally modified in the ER, which is essential for proper folding and function, before being secreted via the Golgi apparatus into the extracellular space (Viotti 2016). Signal peptide sequences are very diverse and have next to no detectable homology. However, they can all be described by a similar organisation consisting of three regions. The n-region is the most N-terminal part of the sequence. It consists of approximately 1 to 5 amino acids and is mostly positively charged. Next is the hydrophobic core (or h-region) which is 7 to 15 amino-acid long. This part of the peptide was shown to bind to the SRP in an α-helical form by (Keenan et al. 1998) and is of utmost importance for recognition by the SRP (Nilsson et al. 2015). Finally, the c-region often contains α-helix-breaking amino acids (either proline or glycine) and the cleavage site is usually surrounded by small uncharged amino acids (Martoglio and Dobberstein 1998).

Predicting the presence of N-terminal signal peptides has been attempted for nearly as long as the presence of such peptides was hypothesised (Blobel and Sabatini 1971). For readers interested in the development of this field, we recommend the excellent review by Nielsen and collaborators (Nielsen et al. 2019a), which illustrates the tremendous progress bioinformatics has made in recent years. One of the biggest challenges for all algorithms is the discrimination between proper signal peptides and N-terminal transmembrane helices (often called membrane anchors), since they both share a similar hydrophobic stretch of amino acids. Better algorithms are continuously being developed and, for that reason, we would encourage users to always use the newest algorithms available. At the time of writing of this chapter, three algorithms have similarly good level of performances: SignalP 5.0 (Armenteros et al. 2019), DeepSig (Savojardo et al. 2018), and SigUNET (Wu et al. 2019). All three use neural networks to predict signal peptides and take special precautions not to detect N-terminal transmembrane domains. SignalP (www.cbs.dtu.dk/services/SignalP/) and DeepSig (deepsig.biocomp.unibo.it) offer easy to use webservices, while SigUNET's source code is available via GitHub (github.com/mbilab/SigUNet), making it only accessible to more advanced users. Of note, while SignalP only reports the presence of signal peptides, DeepSig also detects N-terminal transmembrane helices.

In addition to these specialized algorithms, pipelines that detect both signal peptides and other protein features or cellular localizations are available. For our purpose, TOPCONS (topcons.cbr.su.se) (Tsirigos et al. 2015) is an interesting algorithm since it combines the results from many signal peptide predictors including [Poly]Phobius (Käll et al. 2004, 2005), Philius (Reynolds et al. 2008), and SP OCTOPUS (Viklund et al. 2008). Another tool is DeepLoc (www.cbs.dtu.dk/services/DeepLoc), a neural network which predicts the cellular location of a protein (Almagro Armenteros et al. 2017). Regardless of the algorithm used, we encourage our readers to apply more than one algorithm (e.g. SignalP and DeepSig) and compare the results obtained, as every prediction has a certain level of uncertainty.

In addition to the conventional protein secretion (CPS) characterised by the presence of an N-terminal signal peptide, a smaller number of proteins undergo secretion via unconventional pathways (Dimou and Nickel 2018). In comparison to the prediction of the presence of signal peptides, the prediction of proteins undergoing unconventional secretion is more difficult, partly because the number of well-characterized examples is limited. For an overview of predictors available before 2019, we invite our readers to refer to the recent re-evaluation by Nielsen and collaborators, which concludes that most predictors perform only moderately well (Nielsen et al. 2019b). Of note, a new predictor called OutCyte (www.outcyte.com) was recently published and reported promising results (Zhao et al. 2019). Interestingly, OutCyte also has a powerful module to predict signal peptides.

In addition to being secreted, most extracellular proteins are composed of multiple domains often present in repetition. As this modular domain-based organisation is quite a unique feature of ECM proteins, we briefly present below what are protein domains and how can they be computationally predicted.

2.3.2 *What Are Protein Domains and How Can We Predict Them?*

A domain is a consecutive stretch of amino acids that forms an individual folding unit. Additionally, domains are defined by their three-dimensional structure and are repeatedly used in different proteins. It is believed that all occurrences of a domain originate from a common ancestral "proto-domain". However, the sequence similarity between different occurrences of a particular domain can be very low.

2.3.2.1 The Example of the von Willebrand Factor A Domain

The von Willebrand Factor A (VWA) domain is a domain widespread in ECM proteins and structurally well defined (Gebauer et al. 2016; Springer 2006; Whittaker and Hynes 2002). A valuable tool to explore protein domains is the CATH database (cathdb.info) (Dawson et al. 2017). CATH compares all currently available crystal structures deposited in the protein data bank (PDB, http://www.wwpdb.org/) (Burley et al. 2019) and groups them based on structural similarity. VWA domains form the CATH Superfamily 3.40.50.410 and is composed of an alignment of 215 unique 3D structures. In the structural overlay the core domain, consisting of a central β-sheet surrounded by α-helices, is very easily recognizable (Fig. 2.1a). VWA domains are also present in intracellular proteins, but, with the VWA domain of collagen VI α3 (Col6a3-N5; PDB code 4IGI) and the VWA domain of the von Willebrand factor (VWF; PDB code 4DMU), we can compare two important members of the ECM

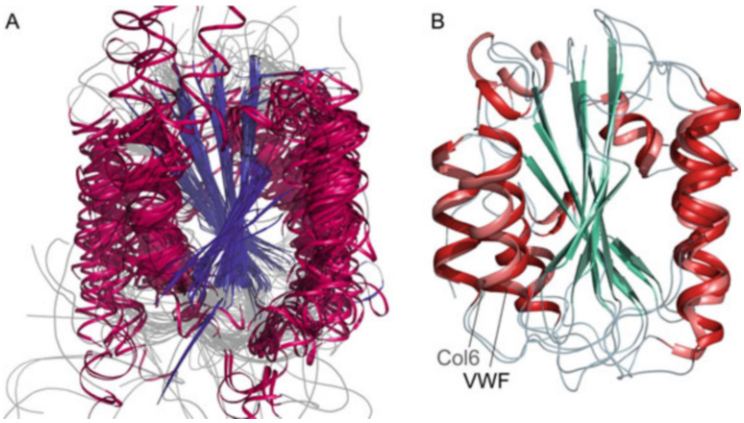

Fig. 2.1 Alignment of crystallised von Willebrand A domains. (**a**) Alignment of 30 representative structures for the CATH superfamily 3.40.50.410 as aligned by CATH (cath-superpose). Chains not belonging to the core domain were removed. α-helices are shown in red, β-sheets in blue and loops in transparent-grey. (**b**) Structural alignment of the Willebrand Factor type A domains of the von Willebrand factor (4DMU) and Collagen VI $\alpha 3$ (4IGI). Col6 is shown in muted colours. Alignment was generated using the "super" algorithm in PyMol

(Fig. 2.1b). Although the overall 3D structure is nearly identical (root-mean-square deviation of only ~1.3 Å), the total sequence identity is only 17%. Detecting such weak homologies is nearly impossible by direct sequence comparison but can still be done using a technique called profile hidden Markov models (HMM).

2.3.2.2 Profile Hidden Markov Models to Predict Domain Homology

Hidden Markov models (HMM) are sequence generators, which try to develop models explaining observed sequences (Eddy 1998; Franzese and Iuliano 2019). Computationally, the algorithm moves from left to right and emits a symbol—for the present purpose, one of the 20 natural amino acids—at every position of a sequence. Based on the multiple sequence alignment, the process knows the likelihood of occurrence of a particular amino acid at every position (emission probability). This process, by which every state stores the likelihood for emitting certain amino acids, is called a "match" state. With this in place, a new sequence can be compared to the built profile and the likelihood with which the new sequence could have been generated by the model can be calculated. However, this profile would only explain sequences, which have no insertion or deletions. To account for those, two new states called "insertion" and "deletion" can be added. During the process, the algorithm runs through a series of these states. For example, if the algorithm adopts a "match" state, it emits an amino acid, however, after emission there is a fair chance that the algorithm adopts a "delete" state for the next position and thereby skip the next match state, resulting in a "missing" amino acid in the alignment.

The probability of switching from one state to another is also derived from the multiple sequence alignment. As we can neither measure these states nor their transition frequency, but only infer those from the output (i.e. the observed sequence alignment) these states are called "hidden". In structural terms, we would expect to see higher probabilities for switching to insertion or deletion station in loop regions of domains connecting secondary structure elements (e.g. the black/grey loop structures in Fig. 2.1a, b).

These profiles can then be used to interrogate databases for other occurrences, which can further be used to refine the profile again. By this method the Pfam database (pfam.xfam.org) has identified over 17,929 different domains and grouped them into families and clans (El-Gebali et al. 2019). A similar approach led to the SMART database (smart.embl-heidelberg.de) (Letunic and Bork 2018). As these two databases use different manually curated sequence alignments, the profiles and predictions differ and may complement each other. A resource to search for both, as well as other predictions, is the InterPro web service (www.ebi.ac.uk/interpro/) (Mitchell et al. 2019). In addition to Pfam and SMART, it also includes PROSITE (prosite.expasy.org), Panther (www.pantherdb.org) and output from other predictors. For more details about protein domain prediction, our readers are directed to the specialised review by Chen and collaborators (Chen et al. 2018).

2.3.2.3 The Special Case of Repeats or Motifs

A special case of domain predictions are repeated motifs, which are very common in ECM proteins. The two most common motifs are Leucine-rich repeats (LRR) and collagen repeats. Both have relatively simple and strict necessities on their primary sequences and are best described by sequence patterns which allow repetitions. Collagens form a right-handed triple helix formed by three individual left-handed poly-proline type II helices. Due to the helicity, every third amino acid is directed to the centre of the triple-helix and due to space restrictions only glycine residues fit at these positions. Thus, the recognition sequence can be described as a repetitive pattern of $(GXY)_{n,}$ where n is usually greater than 10. Prolines are found to be overrepresented at the X and Y positions since they are necessary for the induction of the poly-proline type II helix formed by the individual chains. Detecting such simple patterns in a vast amount of data is a common problem in many fields of informatics and typically solved by the use of "regular expressions" (Aho 1990). For the collagen repeat, a regular expression would look like this: `/(G..){10,}/`, which would search for a GXY motif, with more than 10 occurrences. The advantage of this approach is that one can incorporate knowledge about the biochemistry of the protein directly into the pattern. For example, in our work on cuticular collagens in *C. elegans* (Teuscher et al. 2019), we used this possibility to group genes together based on the similarities of their collagenous domains. The rationale behind this decision was that trimers of collagen will more likely form between structurally similar collagenous domains and not necessarily between more homologous sequences. Cuticular collagens in *C. elegans* have collagenous domains of similar length but differ in the positions and lengths of their imperfections (parts of the sequence which do not adhere to the GXY patterns). For example, the C21a cluster has a collagen domain starting with 8 triplets of GXY, then 3 random amino acids followed by 12 triplets adjacent to a 2 amino acid spacer and finally a long stretch of 21 GXY triplets. This leads to a motif in pseudocode `(G..){8}...(G..){12}..(G..){21}` (the real regular expression is slightly more complex and can be found at cecoldb.permalink.cc/clades/C21a/) which identified 4 additional genes in the genome. Although the idea of using a simple pattern for prediction of domains is not new (see for example ProSite patterns (Hulo et al. 2008)), the ease of incorporating new biochemical knowledge in these patterns makes them useful to the day.

As presented, algorithms are available to predict signal peptides and distantly related protein domains, as well as to detect repetitive sequence. The question remains, can we also automatically predict ECM proteins?

2.3.3 *Predicting ECM Proteins by Machine-Learning-Based Approaches*

Trials to automatically predict ECM protein date back to 2010 (Jung et al. 2010). The idea behind these techniques is the notion that features in the primary protein sequence are responsible for the different sorting of proteins, for example to the ECM, and these should be detectable by machine learning algorithms. It is beyond the scope of this chapter to present the details of the different mechanisms and techniques used in the field of machine learning, instead we recommend specialised books and review articles (Husi 2019; Angermueller et al. 2016; Nielsen et al. 2019a). However, to evaluate proposed ECM predictors, our readers should be aware of the general procedures in machine learning, which can be divided into four steps: (1) data preparation (2) feature extraction (3) model fitting and (4) scoring and evaluation. A critical part of the machine learning process is the proper selection of training data. For ECM proteins, a dataset should include both *bona-fide* ECM proteins (positive set) and known non-ECM proteins (negative set). In brief, the algorithm reduces the data to a couple of features which are believed to be indicative of protein localisation (i.e. the ECM space) and then learns, based on the known output, how to adjust its internal parameter (i.e. the model) to explain the correlation between observed features and known outcomes (ECM protein vs. non-ECM protein). Finally, sequences that are not part of the training set are used to determine how well the algorithm predicts sequences it has never seen before. The quality of a machine learning algorithm is defined by (at least) two parameters: sensitivity and specificity. The first determines to what percentage the predictor correctly identified ECM proteins in the dataset in relation to the total amount of ECM proteins in the dataset. For example, a sensitivity of 10% means that only every 10th ECM protein fed to the algorithm is predicted correctly (true positive rate). Specificity is defined by the percentage the algorithm incorrectly predicts to be ECM proteins. For example, if there are 900 non-ECM proteins in a dataset and the algorithm predicted 90 of them to be ECM proteins (also in fact they are not, i.e. they are false positives), the algorithm would be 90% specific (it mis-predicted 10%; algorithm false negative rate).

Since 2013, most publications reporting the development of algorithms to predict ECM proteins are using a training dataset generate by Kandaswamy and collaborators (Kandaswamy et al. 2013). This dataset contains 445 proteins identified as ECM proteins and 4187 secreted non-ECM proteins from multiple species. Different combinations of feature extraction processes were tested including amino acid composition ([split] amino acid composition, pseudo amino acid composition, dipeptide composition, etc.), structural information such as the presence of secondary structure elements (Zhang et al. 2014), or domain predictions using above-mentioned databases (Jung et al. 2010). Using EcmPred, Kandaswamy and collaborators reported the identification of over 2000 ECM proteins in the ECM proteome (nearly twice the number of proteins defined as part of the matrisome, see below) (Kandaswamy et al. 2013). In a 2012 study, Rejimoan and collaborators

employed position-specific scoring matrices and suggested that they have identified 12 novel ECM protein candidates (CLDN4, DPCR1, Efemp1, IGHG1, IGHM, IGKV4-1, LAT2, LMBR1, MANSC1, NIPA1, NOTCH3, TNC) (Jose et al. 2012). Of these, 2, Efemp1 and TNC, are known ECM proteins while no structural or experimental evidence support the fact that the 10 others would be ECM proteins. The latest predictors published in 2017 (Guan et al. 2017) and 2018 (Kabir et al. 2018) claim to achieve sensitivity/specificity rates of 85.0%/86.5% and 82.3%/90.7%, respectively. However, the fact that the two algorithms, as well as the web interfaces of iECMP (Yang et al. 2015) and PECM (Zhang et al. 2014), are not publicly available makes it difficult to evaluate their performances. Additionally, inspection of the training dataset revealed that nearly 10% (50) sequences are Wnt proteins and approximately 10% are metalloproteases (26 MMPs and 22 ADAMs), mostly due to inclusion of the same protein from different species. Consequently, 5 out of 17 domains identified by Yang et al (Yang et al. 2015) to be indicative for ECM proteins were peptidase domains. As we will discuss below, this is in stark contrast to the ratio in the manually curated human matrisome (Naba et al. 2016), where ADAMs and MMPs constitute 6% of all sequences and Wnt proteins only 1.7%. This bias is likely to lead to over-training of the algorithms towards proteases and Wnt signalling proteins and to diminish the ability of these algorithms to detect other ECM proteins belonging to other families.

2.4 Combining Structural Features and Prior Knowledge to Define the Matrisome of Organisms

Despite significant advances in machine-learning-based processing to predict various protein features, purely computational approaches fail to satisfactorily predict the complete collection of proteins composing the ECM. This is partly due to the fact that proteins composing the ECM are very diverse structurally (see below). Yet, having such lists is instrumental to properly and comprehensively identify and classify ECM genes and proteins from datasets generated using high throughput approaches. To define, what is now called, the "matrisome", we devised a computational pipeline combining several sequence analysis predictors discussed above and manual curation based on prior knowledge (Fig. 2.2). The term matrisome was coined in reference to the term "basement membrane matrisome" originally proposed by George Martin and colleagues to describe protein complexes that includes type IV collagen, laminins, heparan sulphate proteoglycan and nidogens and that, upon assembly, form basement membranes (Martin et al. 1984). In 2011, we proposed to expand the list of proteins grouped under that term to collectively describe the ensemble of genes encoding proteins that are forming the structure of the ECM, as well as proteins present in the extracellular space and that have the ability to interact with and/or remodel ECM proteins (Naba et al. 2012b).

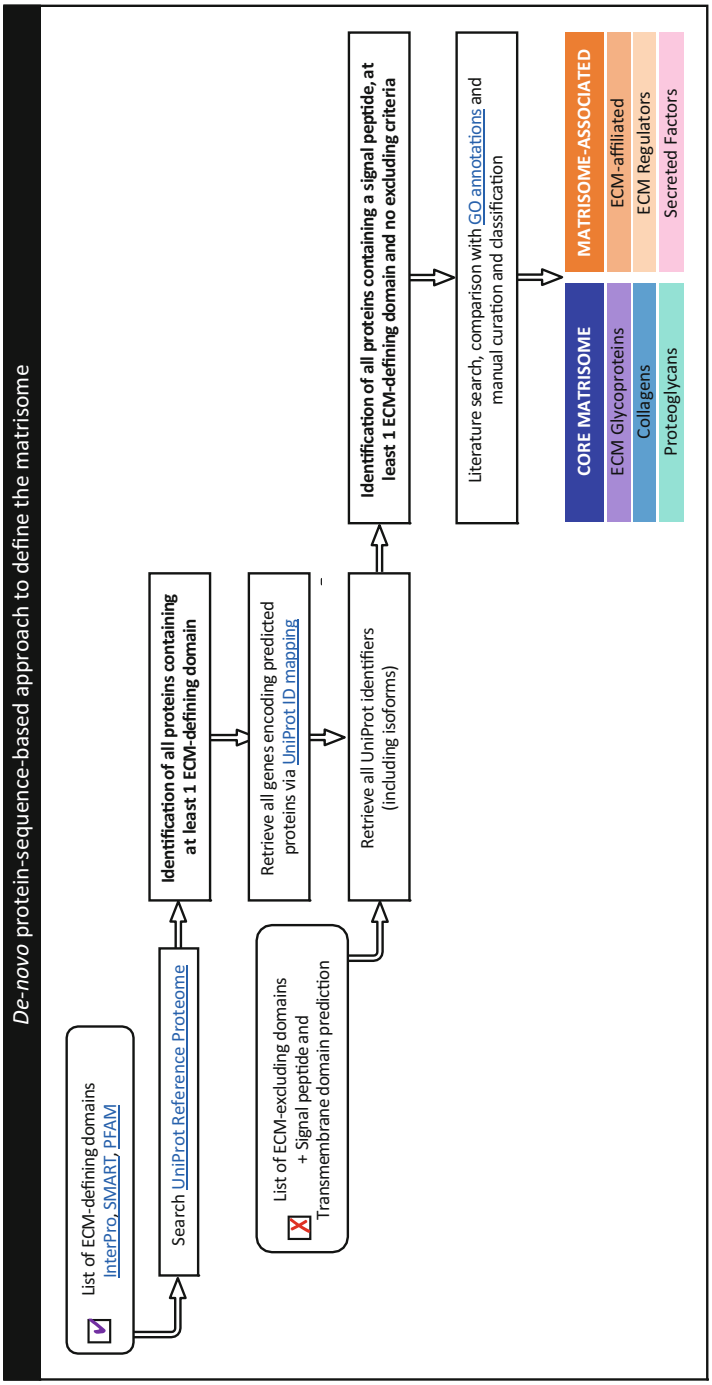

Fig. 2.2 Bioinformatic pipeline to predict the matrisome of model organisms

2.4.1 *Defining the Human and Murine Matrisomes*

To define the human and mouse matrisome, we first identify key structural features of ECM proteins, including the presence of a signal peptide (*see Sect. 2.3.1*) and the presence of at least one of 51 characteristic protein domains commonly found in ECM proteins (*see Sect. 2.3.2* and Table 2.1) (Adams and Engel 2007; Hohenester and Engel 2002; Whittaker and Hynes 2002; Whittaker et al. 2006). Screening of the human proteome for proteins displaying these features retrieves a very long list of proteins. Many of them were known ECM proteins, some were proteins for which no localization information were available, and many were clearly not ECM proteins. Indeed, the structural features used for the screen are also shared by other proteins, including transmembrane receptors and proteins involved in cell-cell adhesions. We thus then defined a set of features rarely found in ECM proteins, such as the presence of kinase or phosphatase catalytic domains, or 7-transmembrane domains found in G-protein-coupled receptors, and used these to eliminate non-ECM components (Naba et al. 2012b). Last, we conducted extensive literature search and database cross-referencing and curated the list to include all known ECM proteins in addition to putative novel ECM proteins (Fig. 2.2). Since this approach focused on identifying proteins considered to be "structural" components of the ECM, we termed this collection the "core matrisome" and further classified proteins of this collection into three categories: collagens, glycoproteins and proteoglycans (Hynes and Naba 2012; Naba et al. 2012b). While all 44 collagen genes of the human genome could be predicted by the presence of one single InterPro domain termed "Collagen triple helix repeat" (IPR008160; see Table 2.1), this domain also retrieved 32 other proteins in UniProt (release 2019_11). Manual examination of these led to their classification as ECM glycoproteins, if there was significant evidence for the protein to contribute to the structure of the ECM (Emid1, Emilin-1), as being associated to the matrisome (e.g. ficolins or collectins, *see below*), or excluded if they also presented a feature not expected to be found in ECM proteins (as for the transmembrane Macrophage receptor MARCO, or Scavenger receptors). Interestingly, although the vast majority of proteins identified with this approach were known ECM proteins, a few do not have a role yet, and can thus be considered "novel" ECM proteins.

ECM homeostasis is under the control of proteins, including enzymes, not traditionally considered to be part of the ECM. Similarly, the ECM is a reservoir for growth factors, also not traditionally considered as components of the ECM (Hynes 2009). However, these proteins and others are integral to ECM functions. To fully capture the complexity of the ECM, we reiterated the sequence-analysis-based pipeline using new sets of domains and knowledge-based curation to define a second collection of proteins, those that could associate with core ECM components (Naba et al. 2012b). We first defined ECM-affiliated proteins as proteins either structurally or functionally affiliated with core ECM components (e.g. galectins, mucins, surfactant proteins). We also defined a group termed ECM regulators which includes ECM-remodeling enzymes such as proteases and cross-linking enzymes as well as

Table 2.1 Protein domains commonly found in structural ECM proteins and used to predict the core matrisome

InterPro domain name	InterPro accession number	Example of core matrisome proteins containing said domain (Gene symbol)
Agrin NtA	IPR004850	Agrin (AGRN)
Amelin	IPR007798	Ameloblastin (AMBN)
Amelogenin	IPR004116	Amelogenin, X isoform (AMELX), Amelogenin, Y isoform (AMELY)
Anaphylatoxin/fibulin	IPR000020	Fibulins (FBLN1, FBLN2)
Bone sialoprotein II	IPR008412	Bone sialoprotein 2 (IBSP)
C-type lectin	IPR001304	C-type lectin domain family 18 member A (CLEC18A)
Collagen triple helix repeat	IPR008160	All 44 collagens
CUB	IPR000859	Procollagen C-endopeptidase enhancer 2 (PCOLCE2)
Cysteine-rich flanking region, C-terminal	IPR000483	Peroxidasin-like protein (PXDNL)
Dentin matrix 1	IPR009889	Dentin matrix acidic phosphoprotein 1 (DMP1)
EGF-like calcium-binding	IPR001881	Fibrillins (Fbn1-6), Perlecan (HSPG2)
EGF-like, laminin	IPR002049	Laminins, Agrin (AGRN), Multiple epidermal growth factor-like domains protein 6 (MEGF6)
EMI	IPR011489	Multimerin-1 (MMRN1), Periostin (POSTN), Transforming growth factor-beta-induced protein ig-h3 (TGFBI)
Endoglin/CD105 antigen	IPR001507	Alpha-tectorin (TECTA)
FAS1_domain	IPR000782	Periostin (POSTN), Transforming growth factor-beta-induced protein ig-h3 (TGFBI)
Fibrillar collagen, C-terminal	IPR000885	Collagen alpha-1(I) chain (COL1A1), Collagen alpha-1(II) chain (COL2A1), Collagen alpha-1(III) chain (COL3A1), Collagen alpha-2(V) chain (COL5A2)
Fibrinogen, alpha/beta/gamma chain, C-terminal globular	IPR002181	Fibrinogens (FGA, FGB, FGG), Tenascins (TNC, TNN, TNR, TNXB)
Fibronectin, type I	IPR000083	Fibronectin (FN1)
Fibronectin, type II	IPR000562	Fibronectin (FN1)
Fibronectin, type III	IPR003961	Fibronectin (FN1), Tenascins (TNC, TNN, TNR, TNXB)
Follistatin-like, N-terminal	IPR003645	Agrin (AGRN), Osteonectin (SPARC)
G2 nidogen and fibulin G2F	IPR006605	Nidogen-1 (NID1), Nidogen-2 (NID2), Hemicentin-1 (HMCN1), Hemicentin-2 (HMCN2)
Gamma-carboxyglutamic acid-rich (GLA)	IPR000294	Matrix Gla Protein (MGP), Osteocalcin (BGLAP)
Hemopexin/matrixin	IPR000585	Hemopexin (HPX), Vitronectin (VTN)

(continued)

Table 2.1 (continued)

InterPro domain name	InterPro accession number	Example of core matrisome proteins containing said domain (Gene symbol)
Insulin-like growth factor-binding protein, IGFBP	IPR000867	Insulin-like growth factor-binding proteins (IGFBP1-7), CCN proteins (CCN1-6)
Laminin G	IPR001791	Laminin alpha chains (LAMA1, LAMA2, LAMA3, LAMA4, LAMA5), Agrin (AGRN)
Laminin I	IPR009254	Laminin alpha chains (LAMA1, LAMA2, LAMA3, LAMA4, LAMA5)
Laminin, N-terminal	IPR008211	All laminins
LCCL	IPR004043	Cochlin (COCH), Vitrin (VIT)
Leucine-rich repeat, cysteine-rich flanking region, N-terminal	IPR000372	Prolargin (PRELP), Lumican (LUM)
Leucine-rich repeat, typical subtype	IPR003591	Peroxidasin-like protein (PXDNL)
Link	IPR000538	Neurocan core protein (NCAN), Hyaluronan and proteoglycan link protein 1 (HAPLN1), Versican core protein (VCAN), Aggrecan core protein (ACAN)
Micro-fibrillar-associated 1, C-terminal	IPR009730	Microfibrillar-associated protein 1 (MFAP1)
Microfibril-associated glycoprotein	IPR008673	Microfibrillar-associated proteins 2 and 5 (MFAP2, MFAP5)
Nidogen, extracellular region	IPR003886	Alpha-tectorin (TECTA), Nidogen-1 (NID1), Sushi, nidogen and EGF-like domain-containing protein 1 (SNED1)
Osteopontin	IPR002038	Osteopontin (SPP1)
Osteoregulin	IPR009837	Matrix extracellular phosphoglycoprotein (MEPE)
Reeler region	IPR002861	Reelin (RELN)
SEA	IPR000082	Agrin (AGRN), Perlecan (HSPG2)
Serglycin	IPR007455	Serglycin (SRGN)
Small leucine-rich proteoglycan, class I, decorin/asporin/ byglycan	IPR016352	Asporin (ASPN), Biglycan (BGN), Decorin (DCN)
Somatomedin B	IPR001212	Vitronectin (VTN)
Sushi/SCR/CCP	IPR000436	Neurocan core protein (NCAN), Sushi repeat-containing protein SRPX2 (SRPX2)
TB domain	IPR017878	Fibrillins (Fbn1-2), Latent-transforming growth factor beta-binding proteins (LTBP1-4)
Thrombospondin, C-terminal	IPR008859	Thrombospondins (THBS1, THBS2, THBS3, THBS4, COMP)
Thrombospondin, type 1 repeat	IPR000884	Thrombospoindins, SCO-spondin (SSPO), Papilin (PAPLN)
Thyroglobulin type-1	IPR000716	Insulin-like growth factor-binding proteins, Nidogen-1 (NID1)

(continued)

Table 2.1 (continued)

InterPro domain name	InterPro accession number	Example of core matrisome proteins containing said domain (Gene symbol)
Tropoelastin	IPR003979	Elastin (ELN)
von Willebrand factor, type A	IPR002035	Collagen VI (COL6A1, COL6A2, COL6A3, COL6A5, COL6A6), von Willebrand Factor (VWF)
von Willebrand factor, type C Domain	IPR001007	Peroxidasin-like protein (PXDNL), SCO-spondin (SSPO), von Willebrand factor C domain-containing protein 2-like (VWC2L), Alpha-tectorin (TECTA)
von Willebrand factor, type D	IPR001846	SCO-spondin (SSPO), Alpha-tectorin (TECTA), von Willebrand factor (VWF)

their regulators (e.g. matrix metalloproteinases and tissue inhibitors of metalloproteinases or cathepsins and cystatins); last, based on the increasing recognition that ECM proteins can modulate growth factor signaling, we also included secreted factors (e.g. growth factors, cytokines, morphogens). Overall, this computational pipeline identified 1027 human matrisome genes and 1110 murine matrisome genes (Naba et al. 2012b), representing approximately 5% of these genomes (Tables 2.2 and 2.3). The complete list of the human and murine matrisome genes can be accessed at http://matrisome.org (Naba et al. 2016).

The matrisome lists defined can be used to study the evolution of ECM genes (see below). They can also be used to annotate genomic, transcriptomic, and proteomic datasets and uncover novel or unsuspected roles for the ECM in pathophysiological processes such as cancer (Izzi et al. 2019; Pearce et al. 2018; Socovich and Naba 2019; Taha and Naba 2019; Tian et al. 2020; Yuzhalin et al. 2018) and fibrosis (Arteel and Naba 2020; Bingham et al. 2020; Dolin and Arteel 2020; Massey et al. 2017; Ricard-Blum and Miele 2020; Yu et al. 2018; Zhou et al. 2018), and eventually lead to the discovery of ECM biomarkers of predictive or prognostic values for patients (Izzi et al. 2019; Yuzhalin et al. 2018).

Thanks to the rapid development of sequencing technologies in the past decades, the genomes of a large number of model organisms used for biomedical research are now available through interfaces such as the Genome Research Consortium (https://www.ncbi.nlm.nih.gov/grc) or the Alliance of Genome Resources (https://www.alliancegenome.org/) (Agapite et al. 2020). Experimental proteomic data or *in-silico* predictions have also permitted to draft the proteomes of several of these model organisms, which are accessible via protein databases such as UniProt (https://www.uniprot.org/) (The UniProt Consortium 2019). It is thus now feasible to predict the matrisomes of other model organisms, using a pipeline similar to the one developed for the human and murine matrisomes.

Table 2.2 The matrisome of model organisms

	Human	Mouse	Zebrafish	Japanese quail	Drosophila	*C. elegans*	Planarian
Genome size (# of genes)	~20,000	~22,000	~26,000	~ 16,000	~14,000	~20,000	~30,000
# of matrisome genes	1027	1110	1002	706	641	719	256
Percentage of the genome encoding the matrisome	5%	5%	4.4%	4.4%	4.5%	4%	0.9%
Reference	Naba et al. (2012b)	Naba et al. (2012b)	Nauroy et al. (2018)	Huss et al. (2019)	Davis et al. (2019)	Teuscher et al. (2019)	Cote et al. (2019)

Table 2.3 Number of genes encoding proteins of different matrisome categories

	Human	Mouse	Zebrafish	Japanese quail	Drosophila	*C. elegans*	Planarian
Core matrisome	274	274	333	238	34	226	117
ECM glycoproteins	195	194	233	167	27	35	98
Collagens	44	44	58	40	4	181	19
Proteoglycans	35	36	42	31	3	10	–
Matrisome-associated	753	836	669	468	279	481	139
ECM-affiliated proteins	171	165	145	99	106	301	24
ECM regulators	238	304	219	155	98	128	71
Secreted factors	344	367	305	214	75	52	44
Organism-specific proteins	–	–	–	–	328	12	–
Complete matrisome	1027	1110	1002	706	641	719	256

2.4.2 The Zebrafish Matrisome

The zebrafish (*Danio rerio*) is a model organism broadly used in many areas of biomedical research, from developmental biology to the study of tissue regeneration and cancer metastasis (Meyers 2018; Parichy 2015). It is also an instrumental model system to help decipher the role of the ECM in these processes (Jessen 2015).The Zebrafish Information Network (ZFIN, https://zfin.org) (Ruzicka et al. 2019) provides an array of resources to the community, from genomic information to expression data and tools. In order to predict the zebrafish matrisome, we employed a sequence-orthology-based approach to retrieve all orthologs of human and murine matrisome genes present in the zebrafish genome (Nauroy et al. 2018). We identified 1002 matrisome genes in the zebrafish genome (about 4.4% of the whole genome, Table 2.2), 333 genes encoding core matrisome proteins and 669 genes encoding matrisome-associated proteins (Table 2.3). We showed that 68.8% of human matrisome genes (710 genes) have at least one ortholog in the zebrafish. We further evaluated the consequences of the teleost-specific whole-genome duplication on the matrisome and found that 44.4% of the matrisome genes have a "one-to-one" relationship between human and zebrafish, 22.3% had a "one-to-two" relationship between human and zebrafish, and 2.1% had a "one-to-many" relationship between human and zebrafish (Nauroy et al. 2018). This last number is to contrast the 15.2% of genes estimated to exist as multiple paralogs in zebrafish at the whole-genome level (Howe et al. 2013), suggesting that matrisome genes are differentially subjected to evolutionary pressure (Nauroy et al. 2018).

2.4.3 The Quail Matrisome

The quail (*Coturnix japonica*) is a model organism that has being extensively used to study developmental processes such as the behavior and differentiation of neural crest cells, through chick-quail graft experiments pioneered by Nicole Le Douarin (Ainsworth et al. 2010; Ribatti 2019), and has helped uncover the roles of several ECM proteins in embryonic development (Loganathan et al. 2016; Spence and Poole 1994; Zamir et al. 2008). Through sequence analysis, Huss and colleagues sought to identify quail orthologs of human matrisome genes (Huss et al. 2019) and predicted that 706 genes (or 4.4% of the quail genome) encoded matrisome proteins (Table 2.2) that can further be classified into 238 core matrisome genes and 468 matrisome-associated genes (Table 2.3). Overall, this study demonstrated that the orthology is greater for core matrisome genes than for matrisome-associated genes (Table 2.3). For example, 40 of the 44 human collagens have orthologs in the quail genome, with only COL5A3, COL6A5, COL11A2, COL26A1 missing (Table 2.3). With the quail matrisome defined and available to annotate big data, Huss and colleagues performed single-cell RNASeq (scRNASeq) to profile the expression of genes of primordial germ cells. These cells participate in

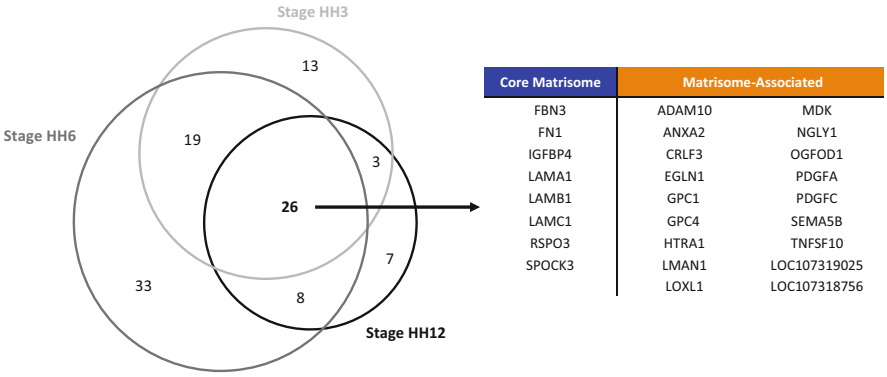

Fig. 2.3 Matrisome genes expressed by avian primordial germ cells at three stages of embryonic development (*Huss et al. Front. Cell Dev. Biol.,* 2019)

gonadogenesis and were profiled at 3 stages of quail embryo development, HH3, HH6 and HH12 (Huss et al. 2019). The study showed that primordial germ cells express a defined set of 26 matrisome and matrisome-associated genes throughout these 3 stages of development, but also identified stage-specific sets of matrisome genes (Fig. 2.3) (Huss et al. 2019).

2.4.4 The Drosophila *Matrisome*

Because of the conservation of protein domains during evolution, an orthology-based approach can also be applied to predict the matrisome of invertebrates. The fruit fly (*Drosophila melanogaster*) is a model organism broadly used to understand the fundamental mechanisms underlying ECM assembly and functions, cell-ECM interactions, and ECM-dependent cell polarization and morphogenesis (Brown 2011; Diaz-de-la-Loza et al. 2018; Ramos-Lewis and Page-McCaw 2019). It is also used to study human diseases (Cheng et al. 2018; Jennings 2011; Markow 2015). Similar to ZFIN, FlyBase (https://flybase.org) (Thurmond et al. 2019) provides resources to the scientific community on this model organism including genomic and orthology data for all *Drosophila* genes. To predict the *Drosophila* matrisome, we used a pipeline similar to the one we used to predict the zebrafish matrisome (see above), and first retrieved from FlyBase all orthologs of mammalian matrisome genes (Davis et al. 2019). In addition to ECM genes orthologous to mammalian genes, fruit flies similar to other arthropods, have additional specialized ECM proteins including the chitin-based cuticle (forming their exoskeleton) and the ECMs that line the lumens of the trachea, salivary glands and midgut; the eggshell; and the salivary glue (Davis et al. 2019). Since these do not have orthologs in mammals, we devised a *de-novo* discovery pipeline based on the presence of characteristic protein domains, as initially done to predict the human and murine

matrisomes (Davis et al. 2019). Altogether, we predicted that the *Drosophila* matrisome comprises 641 genes (Table 2.2), further classified into 34 core matrisome genes, 279 matrisome-associated genes, and 328 genes encoding proteins forming apical matrices (Table 2.3) (Davis et al. 2019). We also proposed a systematic classification based on sequence analysis and the presence of protein domains of the genes encoding proteins forming apical ECMs (Davis et al. 2019).

2.4.5 The C. elegans *Matrisome*

The nematode *Caenorhabditis elegans* is also broadly used to advance our understanding of biological and pathological processes (Frézal and Félix 2015; Kaletta and Hengartner 2006; Meneely et al. 2019).

WormBase (https://wormbase.org) (Harris et al. 2020) serves as the reference database for scientists working with this model organism. On the model of what we presented for the *Drosophila* matrisome, we undertook an approach combining ortholog identification using WormBase and *de-novo* characterization using protein-sequence features to predict the *C. elegans* matrisome. We identified 719 matrisome genes (Table 2.2), including 226 core matrisome genes, 481 matrisome-associated genes, and 12 genes encoding cuticlins, a family of proteins participating in the formation of the *C. elegans* cuticle (Table 2.3) (Teuscher et al. 2019). A unique characteristic of the matrisome of this model organism is the remarkable number, 185, of genes encoding collagen-domain-containing proteins. We further proposed to classify these genes into 4 groups based on sequence analysis. Group 1 comprises the vertebrate-like collagens *emb-9* and *let-2*, orthologs of the mammalian collagen IV, *cle-1*, orthologous to mammalian collagens XV and XVIII and *col-99*, orthologous to membrane-associated collagens with interrupted triple-helices (MACITs, collagen types XIII, XXIII, and XXV). Group 2 comprises 4 genes encoding collagen-domain-containing proteins orthologs to mammalian gliomedins and collectins. Group 3 comprises the 4 non-cuticular collagens with no clear orthology to mammalian collagens. Finally, group 4 includes the 173 cuticular collagens. We further proposed to sub-divide the cuticular collagens into 5 clusters based on their protein-domain organization, including the length of their collagenous domains (i.e. number of GXY repeats; *see 3.2.3*), the positions of interruptions, type of cysteine knot flanking the GXY repeats, and their prediction of being transmembrane or secreted (Teuscher et al. 2019).

Interestingly, the prediction of the matrisome of *Drosophil*a and *C. elegans* has revealed that a similar structure, the cuticle, exerting a similar protective function, is formed by different classes of proteins in different organisms, the *C. elegans* cuticle is mostly collagenous whereas in *Drosophila* it is composed of chitin-domain-containing proteins.

As briefly illustrated for the other matrisomes, the list of computationally-predicted *C. elegans* matrisome genes can further aid annotations of large datasets and identification of processes that are regulated in part by the ECM. In a recent

study, Ewald re-analyzed a previously published transcriptomic dataset aimed at characterizing changes in gene expression during *C. elegans* aging (Budovskaya et al. 2008) and showed that out of the 1200+ age-regulated genes, 150 were matrisome genes. Of them, 146 of which saw their level of expression decrease with aging (including 92 collagens) whereas only 4 saw their level of expression increase with aging (Ewald 2019).

2.4.6 The Planarian Matrisome

Planarians (*Schmidtea mediterranea*) are flatworms extensively studied for their high regenerative potential. Research using this model organisms has shed light on the molecular mechanisms underlying among other processes, stem cell biology, tissue regeneration and repair, and more recently pharmacology and toxicology (Elliott and Alvarado 2018; Gentile et al. 2011; Pagán 2017; Reddien 2018; Sánchez Alvarado 2015). Undertaking the same *de-novo* prediction approach based on sequence analysis and the presence of ECM-specific domains we used to predict the human matrisome, Cote and colleagues identified with high confidence the planarian matrisome as a collection of 256 genes, further divided into 117 core matrisome genes (including planarian orthologs to major structural ECM components, collagens, fibronectin, laminins, fibulins, etc.) and 139 matrisome-associated genes, including orthologs of ECM-affiliated proteins mucins and glypicans and of ECM regulators including ADAMTS and MMPs (see Tables 2.2 and 2.3) (Cote et al. 2019). A previous study from the Reddien lab had shown that muscle cells could provide positional instructions for the regeneration of any region of the planarian's body (Witchley et al. 2013). Since ECM proteins are capable of signaling this type of information to cells, Cote and colleagues sought to determine whether muscle cells contributed to the production of the ECM in planarian. Using the newly developed planarian matrisome to annotate previously published scRNASeq data from the planarian transcriptome atlas (Fincher et al. 2018), Cote and colleagues demonstrated that the main source of ECM proteins were indeed muscle cells. In particular, they showed that muscle cells expressed all 19 collagen genes indicating that it is muscle cells, and no other mesenchymal cell types, that act as a connective tissue in planarians (Cote et al. 2019).

2.5 Conclusion and Future Directions

Despite advances in machine learning, predicting ECM proteins remains challenging. As discussed in this chapter, the matrisome *per se* is not a homogenous group of proteins and is likely to have several different evolutionary origins. It is thus conceivable that its functional and structural diversity, will make it difficult or potentially impossible to be predicted, at once, with a single machine learning

algorithm. Yet, for the past decade, our knowledge of the matrisomes of various model organisms has vastly expanded. We thus propose that novel machine-learning-based approaches could be trained using, as positive dataset, one or several of the manually curated matrisomes already defined (e.g. human and mouse). This proposal has recently been implemented independently by Liu et al. to develop the ECMPride algorithm (Liu et al. 2020). The algorithm was trained on the human matrisome list as the positive dataset and additionally used extracted sequence features as well as the ECM domain list we compiled (Naba et al. 2012b) as predictive features. This resulted in the prediction of 779 so-called "new" ECM proteins. However, examination of the list of proteins revealed the presence of proteins that are clearly not components of the ECM, for example transmembrane receptors (e.g. LDLR) or proteins belonging the blood coagulation cascade (e.g. THBD). The overestimation of the number of ECM proteins might be caused by the exclusive usage of intracellular proteins as the negative training dataset. Inclusion of transmembrane receptors sharing common sequence features with ECM proteins as well as extracellular proteins not belonging to the ECM might help train the algorithm to better differentiate between the secretome and the matrisome. Furthermore, we suggest that annotated matrisomes from other species not used in the training set (e.g. *C. elegans*), might serve as good free test datasets and should be used to evaluate the power of the algorithm. In particular, it would be very interesting to determine how well future predictors will identify those proteins in the test dataset that do not have clear homologues in the training set (e.g. cuticular collagens). If found to be highly sensitive and specific, the algorithm could then be further used to predict the matrisome of evolutionarily distant organisms that have seen emerged ECM innovations (Draper et al. 2019; Hynes 2012; Özbek et al. 2010; Shoemark et al. 2019).

Accurate and comprehensive big data annotation is only the first step toward discovering ECM genes and proteins involved in pathophysiological processes. Collectively, we should work toward understanding the underlying cellular and molecular mechanisms these genes and proteins control with the ultimate goal of advancing fundamental scientific knowledge and improving patient diagnosis and treatment. We propose that this will be greatly facilitated by the systematic dissemination and sharing of –omic datasets and by efforts aimed at enhancing accessibility of such datasets to non-specialists. To this end, we launched MatrisomeDB, a database collating mass-spectrometry-based proteomics data characterizing the ECM of normal and diseased tissues as well as on the ECM produced by cells in culture (Shao et al. 2020). Along the same line, Dr. Ricard-Blum and collaborators have developed MatrixDB, an inventory of curated ECM protein-protein and protein-glycsoaminoglycan interaction networks (Clerc et al. 2019). With the democratization of –omic technologies, it is our hope that such endeavors will expand and integrate with each other, to eventually provide a system-wide view of the ECM, fully representative of its complexity.

Acknowledgments The authors would like to thank Martin Davis from the Naba lab for his critical reading of the manuscript.

Funding Sources This work was supported by a start-up fund from the Department of Physiology and Biophysics to AN and through project B11 of the SFB829 by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—Project-ID 73111208.

References

Adams JC (2018) Matricellular proteins: functional insights from non-mammalian animal models. Curr Top Dev Biol 130:39–105

Adams JC, Engel J (2007) Bioinformatic analysis of adhesion proteins. Methods Mol Biol 370:147–172

Agapite J, Albou L-P, Aleksander S, Argasinska J, Arnaboldi V, Attrill H, Bello SM, Blake JA, Blodgett O, Bradford YM et al (2020) Alliance of Genome Resources Portal: unified model organism research platform. Nucleic Acids Res 48:D650–D658

Aho AV (1990) CHAPTER 5 - Algorithms for finding patterns in strings. In: Van leeuwen J (ed) Algorithms and complexity. Elsevier, Amsterdam, pp 255–300

Ainsworth SJ, Stanley RL, Evans DJR (2010) Developmental stages of the Japanese quail. J Anat 216:3–15

Almagro Armenteros JJ, Sønderby CK, Sønderby SK, Nielsen H, Winther O (2017) DeepLoc: prediction of protein subcellular localization using deep learning. Bioinformatics 33:3387–3395

Angermueller C, Pärnamaa T, Parts L, Stegle O (2016) Deep learning for computational biology. Mol Syst Biol 12:878

Armenteros JJA, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, von Heijne G, Nielsen H (2019) SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat Biotechnol 37:420–423

Arteel GE, Naba A (2020) The liver matrisome, looking beyond collagens. JHEP Rep 2(4):100115 S2589-5559(20):30049–30045

Bingham, G.C., Lee, F., Naba, A., and Barker, T.H. (2020). Spatial-omics: novel approaches to probe cell heterogeneity and ECM biology. Matrix Biol. pii: S0945-053X(20)30049-4

Birch HL (2018) Extracellular matrix and ageing. Subcell Biochem 90:169–190

Blobel G, Sabatini D (1971) Dissociation of mammalian polyribosomes into subunits by puromycin. Proc Natl Acad Sci U S A 68:390–394

Brown NH (2011) Extracellular matrix in development: insights from mechanisms conserved between invertebrates and vertebrates. Cold Spring Harb Perspect Biol:3

Budovskaya YV, Wu K, Southworth LK, Jiang M, Tedesco P, Johnson TE, Kim SK (2008) An elt-3/elt-5/elt-6 GATA transcription circuit guides aging in *C. elegans*. Cell 134:291–303

Burley SK, Berman HM, Bhikadiya C, Bi C, Chen L, Costanzo LD, Christie C, Duarte JM, Dutta S, Feng Z et al (2019) Protein Data Bank: the single global archive for 3D macromolecular structure data. Nucleic Acids Res 47:D520–D528

Chen J, Guo M, Wang X, Liu B (2018) A comprehensive review and comparison of different computational methods for protein remote homology detection. Brief Bioinform 19:231–244

Cheng L, Baonza A, Grifoni D (2018) Drosophila models of human disease. Biomed Res Int:2018

Clerc O, Deniaud M, Vallet SD, Naba A, Rivet A, Perez S, Thierry-Mieg N, Ricard-Blum S (2019) MatrixDB: integration of new data with a focus on glycosaminoglycan interactions. Nucleic Acids Res 47:D376–D381

Cote LE, Simental E, Reddien PW (2019) Muscle functions as a connective tissue and source of extracellular matrix in planarians. Nat Commun 10:1592

Davis MN, Horne-Badovinac S, Naba A (2019) In-silico definition of the *Drosophila melanogaster* matrisome. Matrix Biol Plus 100015

Dawson NL, Lewis TE, Das S, Lees JG, Lee D, Ashford P, Orengo CA, Sillitoe I (2017) CATH: an expanded resource to predict protein function through structure and sequence. Nucleic Acids Res 45:D289–D295

Diaz-de-la-Loza M-C, Ray RP, Ganguly PS, Alt S, Davis JR, Hoppe A, Tapon N, Salbreux G, Thompson BJ (2018) Apical and basal matrix remodeling control epithelial morphogenesis. Dev Cell 46:23–39.e5

Dimou E, Nickel W (2018) Unconventional mechanisms of eukaryotic protein secretion. Curr Biol 28:R406–R410

Dolin CE, Arteel GE (2020) The matrisome, inflammation, and liver disease. Semin Liver Dis. 40 (2):180–188

Draper GW, Shoemark DK, Adams JC (2019) Modelling the early evolution of extracellular matrix from modern Ctenophores and Sponges. Essays Biochem. 63:389–405

Eddy SR (1998) Profile hidden Markov models. Bioinformatics 14:755–763

El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A et al (2019) The Pfam protein families database in 2019. Nucleic Acids Res 47:D427–D432

Elliott SA, Alvarado AS (2018) Planarians and the history of animal regeneration: paradigm shifts and key concepts in biology. Methods Mol Biol 1774:207–239

Ewald CY (2019) The matrisome during aging and longevity: a systems-level approach toward defining matreotypes promoting healthy aging. GER:1–9

Fincher CT, Wurtzel O, de Hoog T, Kravarik KM, Reddien PW (2018) Cell type transcriptome atlas for the planarian *Schmidtea mediterranea*. Science 360

Franzese M, Iuliano A (2019) Hidden Markov models. In: Ranganathan S, Gribskov M, Nakai K, Schönbach C (eds) Encyclopedia of bioinformatics and computational biology. Academic Press, Oxford, pp 753–762

Frézal L, Félix M-A (2015) *C. elegans* outside the Petri dish. ELife 4:e05849

Gebauer JM, Kobbe B, Paulsson M, Wagener R (2016) Structure, evolution and expression of collagen XXVIII: lessons from the zebrafish. Matrix Biol. 49:106–119

Gentile L, Cebrià F, Bartscherer K (2011) The planarian flatworm: an in vivo model for stem cell biology and nervous system regeneration. Dis Model Mech 4:12–19

Guan L, Zhang S, Xu H (2017) BAMORF: a novel computational method for predicting the extracellular matrix proteins. IEEE Access 5:18498–18505

Harris TW, Arnaboldi V, Cain S, Chan J, Chen WJ, Cho J, Davis P, Gao S, Grove CA, Kishore R et al (2020) WormBase: a modern model organism information resource. Nucleic Acids Res 48: D762–D767

Herrera J, Henke CA, Bitterman PB (2018) Extracellular matrix as a driver of progressive fibrosis. J Clin Invest 128:45–53

Hohenester E, Engel J (2002) Domain structure and organisation in extracellular matrix proteins. Matrix Biol 21:115–128

Howe K, Clark MD, Torroja CF, Torrance J, Berthelot C, Muffato M, Collins JE, Humphray S, McLaren K, Matthews L et al (2013) The zebrafish reference genome sequence and its relationship to the human genome. Nature 496:498–503

Hulo N, Bairoch A, Bulliard V, Cerutti L, Cuche BA, de Castro E, Lachaize C, Langendijk-Genevaux PS, Sigrist CJA (2008) The 20 years of PROSITE. Nucleic Acids Res 36:D245–D249

Husi H (2019) Computational biology. Codon Publications

Huss DJ, Saias S, Hamamah S, Singh JM, Wang J, Dave M, Kim J, Eberwine J, Lansford R (2019) Avian primordial germ cells contribute to and interact with the extracellular matrix during early migration. Front Cell Dev Biol 7:35

Hynes RO (2009) The extracellular matrix: not just pretty fibrils. Science 326:1216–1219

Hynes RO (2012) The evolution of metazoan extracellular matrix. J Cell Biol 196:671–679

Hynes RO, Naba A (2012) Overview of the matrisome - an inventory of extracellular matrix constituents and functions. Cold Spring Harb Perspect Biol 4:a004903

Izzi V, Lakkala J, Devarajan R, Kääriäinen A, Koivunen J, Heljasvaara R, Pihlajaniemi T (2019) Pan-cancer analysis of the expression and regulation of matrisome genes across 32 tumor types. Matrix Biol Plus 1:100004

Jennings BH (2011) Drosophila – a versatile model in biology & medicine. Mater Today 14:190–195
Jessen JR (2015) Recent advances in the study of zebrafish extracellular matrix proteins. Dev Biol 401:110–121
Jose A, Rejimoan R, Sivakumar Kc, Mundayoor S (2012) Prediction of extracellular matrix proteins using SVMhmm classifier. IJCA ACCTHPCA (Spl Iss) (1):7–11. https://www.ijcaonline.org/specialissues/accthpca/number1/7548-1002
Jung J, Ryu T, Hwang Y, Lee E, Lee D (2010) Prediction of extracellular matrix proteins based on distinctive sequence and domain characteristics. J Comput Biol 17:97–105
Kabir M, Ahmad S, Iqbal M, Khan Swati ZN, Liu Z, Yu D-J (2018) Improving prediction of extracellular matrix proteins using evolutionary information via a grey system model and asymmetric under-sampling technique. Chemometr Intell Lab Syst 174:22–32
Kaletta T, Hengartner MO (2006) Finding function in novel targets: *C. elegans* as a model organism. Nat Rev Drug Discov 5:387–399
Käll L, Krogh A, Sonnhammer ELL (2004) A combined transmembrane topology and signal peptide prediction method. J Mol Biol 338:1027–1036
Käll L, Krogh A, Sonnhammer ELL (2005) An HMM posterior decoder for sequence feature prediction that includes homology information. Bioinformatics 21(Suppl 1):i251–i257
Kandaswamy KK, Pugalenthi G, Kalies K-U, Hartmann E, Martinetz T (2013) EcmPred: prediction of extracellular matrix proteins based on random forest with maximum relevance minimum redundancy feature selection. J Theor Biol 317:377–383
Keeley FW, Mecham R (2013) Evolution of extracellular matrix. Springer, Berlin
Keenan RJ, Freymann DM, Walter P, Stroud RM (1998) Crystal structure of the signal sequence binding subunit of the signal recognition particle. Cell 94:181–191
Letunic I, Bork P (2018) 20 years of the SMART protein domain annotation resource. Nucleic Acids Res 46:D493–D496
Liu B, Leng L, Sun X, Wang Y, Ma J, Zhu Y (2020) ECMPride: prediction of human extracellular matrix proteins based on the ideal dataset using hybrid features with domain evidence. PeerJ 8: e9066
Loganathan R, Rongish BJ, Smith CM, Filla MB, Czirok A, Bénazéraf B, Little CD (2016) Extracellular matrix motion and early morphogenesis. Development 143:2056–2065
Markow TA (2015) The secret lives of Drosophila flies. ELife 4:e06793
Martin GR, Kleinman HK, Terranova VP, Ledbetter S, Hassell JR (1984) The regulation of basement membrane formation and cell-matrix interactions by defined supramolecular complexes. Ciba Found Symp 108:197–212
Martoglio B, Dobberstein B (1998) Signal sequences: more than just greasy peptides. Trends Cell Biol 8:410–415
Massey VL, Dolin CE, Poole LG, Hudson SV, Siow DL, Brock GN, Merchant ML, Wilkey DW, Arteel GE (2017) The hepatic "matrisome" responds dynamically to injury: characterization of transitional changes to the extracellular matrix in mice. Hepatology 65:969–982
Meneely PM, Dahlberg CL, Rose JK (2019) Working with worms: *Caenorhabditis elegans* as a model organism. Curr Protoc Essent Lab Tech 19:e35
Meyers JR (2018) Zebrafish: development of a vertebrate model organism. Curr Protoc Essent Lab Tech 16:e19
Mitchell AL, Attwood TK, Babbitt PC, Blum M, Bork P, Bridge A, Brown SD, Chang H-Y, El-Gebali S, Fraser MI et al (2019) InterPro in 2019: improving coverage, classification and access to protein sequence annotations. Nucleic Acids Res 47:D351–D360
Naba A, Hoersch S, Hynes RO (2012a) Towards definition of an ECM parts list: an advance on GO categories. Matrix Biol 31:371–372
Naba A, Clauser KR, Hoersch S, Liu H, Carr SA, Hynes RO (2012b) The matrisome: in silico definition and in vivo characterization by proteomics of normal and tumor extracellular matrices. Mol Cell Proteomics 11:M111.014647

Naba A, Clauser KR, Ding H, Whittaker CA, Carr SA, Hynes RO (2016) The extracellular matrix: tools and insights for the "omics" era. Matrix Biol 49:10–24

Nauroy P, Hughes S, Naba A, Ruggiero F (2018) The in-silico zebrafish matrisome: a new tool to study extracellular matrix gene and protein functions. Matrix Biol 65:5–13

Neve A, Cantatore FP, Maruotti N, Corrado A, Ribatti D (2014) Extracellular matrix modulates angiogenesis in physiological and pathological conditions. Biomed Res Int 2014:756078

Nielsen H, Tsirigos KD, Brunak S, von Heijne G (2019a) A brief history of protein sorting prediction. Protein J. 38:200–216

Nielsen H, Petsalaki EI, Zhao L, Stühler K (2019b) Predicting eukaryotic protein secretion without signals. Biochim Biophys Acta Proteins Proteomics 1867:140174

Nilsson I, Lara P, Hessa T, Johnson AE, von Heijne G, Karamyshev AL (2015) The code for directing proteins for translocation across ER membrane: SRP cotranslationally recognizes specific features of a signal sequence. J Mol Biol 427:1191–1201

Özbek S, Balasubramanian PG, Chiquet-Ehrismann R, Tucker RP, Adams JC (2010) The evolution of extracellular matrix. Mol Biol Cell 21:4300–4305

Pagán OR (2017) Planaria: an animal model that integrates development, regeneration and pharmacology. Int J Dev Biol 61:519–529

Parichy DM (2015) Advancing biology through a deeper understanding of zebrafish ecology and evolution. ELife 4:e05635

Pearce OMT, Delaine-Smith RM, Maniati E, Nichols S, Wang J, Böhm S, Rajeeve V, Ullah D, Chakravarty P, Jones RR et al (2018) Deconstruction of a metastatic tumor microenvironment reveals a common matrix response in human cancers. Cancer Discov 8:304–319

Pickup MW, Mouw JK, Weaver VM (2014) The extracellular matrix modulates the hallmarks of cancer. EMBO Rep 15:1243–1253

Ramos-Lewis W, Page-McCaw A (2019) Basement membrane mechanics shape development: lessons from the fly. Matrix Biol 75–76:72–81

Reddien PW (2018) The cellular and molecular basis for planarian regeneration. Cell 175:327–345

Reynolds SM, Käll L, Riffle ME, Bilmes JA, Noble WS (2008) Transmembrane topology and signal peptide prediction using dynamic Bayesian networks. PLOS Comput Biol 4:e1000213

Ribatti D (2019) Nicole Le Douarin and the use of quail-chick chimeras to study the developmental fate of neural crest and hematopoietic cells. Mech Dev 158:103557

Ricard-Blum S, Miele AE (2020) Omic approaches to decipher the molecular mechanisms of fibrosis, and design new anti-fibrotic strategies. Semin Cell Dev Biol 101:161–169

Rozario T, DeSimone DW (2010) The extracellular matrix in development and morphogenesis: a dynamic view. Dev Biol 341:126–140

Ruzicka L, Howe DG, Ramachandran S, Toro S, Van Slyke CE, Bradford YM, Eagle A, Fashena D, Frazer K, Kalita P et al (2019) The Zebrafish Information Network: new support for non-coding genes, richer gene ontology annotations and the alliance of genome resources. Nucleic Acids Res 47:D867–D873

Sánchez Alvarado A (2015) Unravelling a can of worms. ELife 4:e07431

Savojardo C, Martelli PL, Fariselli P, Casadio R (2018) DeepSig: deep learning improves signal peptide detection in proteins. Bioinformatics 34:1690–1696

Shao X, Taha IN, Clauser KR, Gao, Y. (Tom), and Naba, A. (2020) MatrisomeDB: the ECM-protein knowledge database. Nucleic Acids Res. 48:D1136–D1144

Shoemark DK, Ziegler B, Watanabe H, Strompen J, Tucker RP, Özbek S, Adams JC (2019) Emergence of a thrombospondin superfamily at the origin of metazoans. Mol Biol Evol 36:1220–1238

Socovich AM, Naba A (2019) The cancer matrisome: from comprehensive characterization to biomarker discovery. Semin Cell Dev Biol 89:157–166

Spence SG, Poole TJ (1994) Developing blood vessels and associated extracellular matrix as substrates for neural crest migration in Japanese quail, *Coturnix coturnix japonica*. Int J Dev Biol 38:85–98

Springer TA (2006) Complement and the multifaceted functions of VWA and integrin I domains. Structure 14:1611–1616

Taha IN, Naba A (2019) Exploring the extracellular matrix in health and disease using proteomics. Essays Biochem:EBC20190001

Teuscher AC, Jongsma E, Davis MN, Statzer C, Gebauer JM, Naba A, Ewald CY (2019) The in-silico characterization of the *Caenorhabditis elegans* matrisome and proposal of a novel collagen classification. Matrix Biol Plus 1:100001

The Gene Ontology Consortium (2019) The Gene Ontology Resource: 20 years and still GOing strong. Nucleic Acids Res 47:D330–D338

The UniProt Consortium (2019) UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res 47:D506–D515

Thurmond J, Goodman JL, Strelets VB, Attrill H, Gramates LS, Marygold SJ, Matthews BB, Millburn G, Antonazzo G, Trovisco V et al (2019) FlyBase 2.0: the next generation. Nucleic Acids Res 47:D759–D765

Tian C, Öhlund D, Rickelt S, Lidström T, Huang Y, Hao L, Zhao RT, Franklin O, Bhatia SN, Tuveson DA et al (2020) Cancer-cell-derived matrisome proteins promote metastasis in pancreatic ductal adenocarcinoma. Cancer Res. 80:1461–1474

Tsirigos KD, Peters C, Shu N, Käll L, Elofsson A (2015) The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides. Nucleic Acids Res. 43:W401–W407

Viklund H, Bernsel A, Skwark M, Elofsson A (2008) SPOCTOPUS: a combined predictor of signal peptides and membrane protein topology. Bioinformatics 24:2928–2929

Viotti C (2016) ER to Golgi-dependent protein secretion: the conventional pathway. In: Pompa A, De Marchis F (eds) Unconventional protein secretion: methods and protocols. Springer, New York, NY, pp 3–29

Whittaker CA, Hynes RO (2002) Distribution and evolution of von Willebrand/integrin A domains: widely dispersed domains with roles in cell adhesion and elsewhere. Mol Biol Cell 13:3369–3387

Whittaker CA, Bergeron K-F, Whittle J, Brandhorst BP, Burke RD, Hynes RO (2006) The echinoderm adhesome. Dev Biol 300:252–266

Witchley JN, Mayer M, Wagner DE, Owen JH, Reddien PW (2013) Muscle cells provide instructions for planarian regeneration. Cell Rep 4:633–641

Wu J-M, Liu Y-C, Chang DT-H (2019) SigUNet: signal peptide recognition based on semantic segmentation. BMC Bioinformatics 20:677

Yang R, Zhang C, Gao R, Zhang L (2015) An ensemble method with hybrid features to identify extracellular matrix proteins. PLoS ONE 10:e0117804

Yu G, Ibarra GH, Kaminski N (2018) Fibrosis: lessons from OMICS analyses of the human lung. Matrix Biol. 68–69:422–434

Yuzhalin AE, Urbonas T, Silva MA, Muschel RJ, Gordon-Weeks AN (2018) A core matrisome gene signature predicts cancer outcome. Br J Cancer 118:435–440

Zamir EA, Rongish BJ, Little CD (2008) The ECM moves during primitive streak formation—computation of ECM versus cellular motion. PLOS Biol 6:e247

Zhang J, Sun P, Zhao X, Ma Z (2014) PECM: Prediction of extracellular matrix proteins using the concept of Chou's pseudo amino acid composition. J Theor Biol 363:412–418

Zhao L, Poschmann G, Waldera-Lupa D, Rafiee N, Kollmann M, Stühler K (2019) OutCyte: a novel tool for predicting unconventional protein secretion. Sci Rep 9:1–9

Zhou Y, Horowitz JC, Naba A, Ambalavanan N, Atabai K, Balestrini J, Bitterman PB, Corley RA, Ding B-S, Engler AJ et al (2018) Extracellular matrix in lung development, homeostasis and disease. Matrix Biol 73:77–104

Chapter 3
Detecting Changes to the Extracellular Matrix in Liver Diseases

Christine E. Dolin, Toshifumi Sato, Michael L. Merchant, and Gavin E. Arteel

Abstract Liver disease, regardless of etiology, shares a similar natural history of disease progress and is common worldwide. This spectrum of diseases progresses from simple steatosis (fat accumulation), to inflammation, and eventually to fibrosis and cirrhosis. Hepatic fibrosis is primarily characterized by robust accumulation of collagen ECM that leads to organ dysfunction and decompensation. The role of the ECM in early stages of chronic liver disease is less well-understood, but recent studies have demonstrated that a number of changes in the hepatic ECM in early-stage liver disease may also contribute to disease progression. The purpose of this review is to summarize the established and proposed changes to the hepatic extracellular matrix (ECM) that may contribute to inflammation during earlier stages of disease development, and to discuss potential mechanisms by which these changes may mediate the progression of the disease.

C. E. Dolin
Department of Pharmacology and Toxicology, University of Louisville Health Sciences Center, Louisville, KY, USA

T. Sato
Division of Gastroenterology, Hepatology and Nutrition, Department of Medicine, University of Pittsburgh, Pittsburgh, PA, USA

Pittsburgh Liver Research Center, University of Pittsburgh, Pittsburgh, PA, USA

M. L. Merchant
Division of Nephrology & Hypertension, Department of Medicine, University of Louisville Health Sciences Center, Louisville, KY, USA

G. E. Arteel (✉)
Division of Gastroenterology, Hepatology and Nutrition, Department of Medicine, University of Pittsburgh, Pittsburgh, PA, USA

Pittsburgh Liver Research Center, University of Pittsburgh, Pittsburgh, PA, USA

University of Pittsburgh School of Medicine, Pittsburgh, PA, USA
e-mail: gearteel@pitt.edu

S. Ricard-Blum (ed.), *Extracellular Matrix Omics*, Biology of Extracellular Matrix 7,
https://doi.org/10.1007/978-3-030-58330-9_3

3.1 The Extracellular Matrix (ECM) of the Liver

The ECM is best viewed as a dynamic compartment that encompasses a diverse range of components that work bi-directionally with surrounding tissue to regulate cell and tissue homeostasis. Although the classic meaning of the ECM referred to only proteins directly involved in generating the ECM structure, such as collagens, proteoglycans and glycoproteins, the definition of the ECM is now broader, and includes all components associated with this compartment, including ECM affiliated proteins (e.g., collagen-related proteins), ECM regulator/modifier proteins (e.g., lysyl oxidases and proteases) and secreted factors that bind to the ECM (e.g., TGFβ and other cytokines) (Naba et al. 2016; Arteel and Naba 2020). This updated definition has been coined the 'matrisome' (Naba et al. 2012). Although the canonical function of the ECM is structural, it is also a key storage unit for signaling molecules (e.g., growth factors and cytokines), as well as serving as a sensing mechanism for outside-in signaling and vice-versa (Hynes 2009).

In most tissues, there are two distinct structural ECM components: the interstitial matrix and the basement membrane (Martinez-Hernandez and Amenta 1993). Interstitial matrix proteins (e.g., fibronectins, elastin, and fibrillar collagens) form networks that provide support to the overall superstructure that shapes and encapsulates the organ (Friedman 2010; Arteel and Naba 2020). In most tissues, the basement membrane is a thin, electron-dense sheet of ECM that is the foundation for epithelial and endothelial cells (Arteel and Naba 2020). Similar to the interstitial matrix, the basement membrane comprises many structural ECM proteins that facilitate structure and growth of the cells. The basement membrane in most tissues is a true barrier between the epithelial/endothelial cells and the adjacent parenchymal cells. In contrast, the basement layer in the liver is fenestrated and much loosely organized (Friedman 2010) (Fig. 3.1). Although it possesses similar ECM as more clearly-defined basement membranes [e.g., collagen type IV and laminin (Griffiths et al. 1991)], this region acts more as a structural filter that facilitates bidirectional exchange of proteins and xenobiotics between the sinusoidal blood and hepatocytes (Arteel and Naba 2020). Although it is clear that liver does not have a *basal lamina*, whether or not the ECM found in the space of Disse should be considered a basement membrane is a subject of a histological, rather than functional, debate (Martinez-Hernandez and Amenta 1993; Arteel and Naba 2020).

3.2 Balance and Imbalance of ECM Turnover in the Liver

As mentioned above, the ECM is a dynamic compartment that responds to stress and changes. Under normal conditions, these responses assist in maintaining organ homeostasis and help mediate responses to injury/stress. Subcutaneous wound healing is a canonical illustration of functional changes ECM in response to damage; the tightly regulated deposition and remodeling of the ECM not only mediates

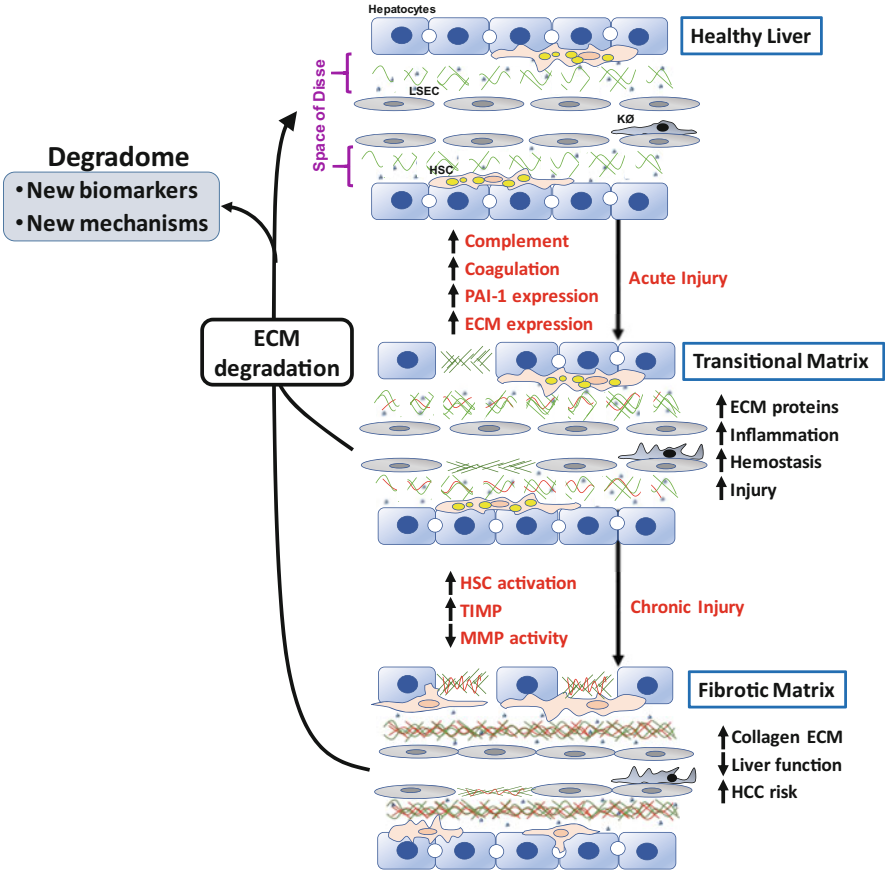

Fig. 3.1 Transitional remodeling of the hepatic ECM. Acute injury quantitatively and qualitatively changes the local ECM. This transitional/provisional ECM plays a key role in mediating the early inflammatory response at the site of injury. These changes often resolve if the insult is removed, and is also hypothesized to facilitate resolution and recovery from that injury. In contrast, when the insult is chronic (e.g., chronic liver diseases), these transitional changes to the ECM may progress to a collagenous/fibrotic ECM. Although this injury can also resolve, it does so less readily. The increase in turnover of the ECM both during injury and resolution release degraded proteins into the blood that may serve as extrahepatic signals, as a well as be a potential source of biomarkers ("degradome"). Abbreviations: *LSEC* liver sinusoidal endothelial cells; *HSC* hepatic stellate cell; *Kø* Kupffer cell; *PAI-1* plasminogen activator inhibitor-1

wound closure, but also recovery and healing (Sun et al. 2014). However, failure when these responses are dysregulated, the changes to the ECM can be maladaptive (Bonnans et al. 2014). For example, 'aging' of the ECM (i.e., increased crosslinking) is hypothesized to contribute to dysfunction in several organ systems, including the liver (Harvey et al. 2016; Sessions and Engler 2016; Sacca et al. 2016; Phillip et al. 2015; Arteel and Naba 2020). The key levels of ECM regulation, both adaptive or

maladaptive, include de novo synthesis, post-translational modifications and degradation.

3.2.1 De Novo Synthesis

Almost all hepatocellular cells play a role in generating the basal ECM found in the normal liver; for example, parenchymal and a nonparenchymal (e.g., cholangiocytes and hepatic sinusoidal endothelial cells) all produce components of the fibrillar ECM (Martinez-Hernandez and Amenta 1995). Although Kupffer cells do not generate fibrillar matrix proteins under basal conditions, they do produce several factors that are associated with the ECM, such as cytokines (see below). It is not clear if hepatic stellate cells (HSC) produce any ECM proteins under basal conditions. However, once HSCs activate and transdifferentiate into a myofibroblast-like phenotype, they are responsible for a large portion of the collagenous ECM generated during fibrosis (Friedman 2010) (Fig. 3.1). Other myofibroblast-like cells also contribute to collagen production (e.g., fibrocytes and periportal fibroblasts) during fibrogenesis (Cassiman et al. 2002; Zeisberg et al. 2007; Robertson et al. 2007; Omenetti et al. 2008). The spectrum and amount of ECM proteins generated by these various cell types change in response to damage and dyshomeostasis. The contribution of extrahepatic sources to the hepatic ECM via *de novo* synthesis is unclear, but these compartments clearly contribute to ECM via other mechanisms of homeostasis (Arteel and Naba 2020).

3.2.2 Maturation of ECM Through Post-Translational Modifications

Post-translational modifications of ECM proteins regulate the formation of oligomeric fibers that comprise mature ECM. For example, prolyl 4-hydroxylase modifies proline amino acids on collagen monomers to facilitate the formation of collagen fibers and helices (Kagan 2000); lysyl oxidases and transglutaminases also play key roles in ECM cross-linking (Liu et al. 2016; Tatsukawa et al. 2016). These post-translational modifications are critical to stabilize and the ECM and protect them from degradation; however, these enzymatic modifications may play key roles in excessive ECM accumulation during fibrogenesis and ECM 'aging' (Liu et al. 2016). Furthermore, although it is now understood that fibrosis is potentially reversible (Poynard et al. 2002), highly crosslinked ECM appears to persist in recovered livers (Issa et al. 2004; Schuppan et al. 2018). The ECM is also post-translationally modified by nonezymatic mechanisms; for example, diabetic 'aging' of ECM is thought to be mediated via adduction of ECM proteins with advanced glycation endproducts (AGEs) (Huijberts et al. 2008).

3.2.3 ECM Degradation

The regulated degradation of ECM proteins by specific proteases is a key factor in maintaining appropriate levels of turnover. Protein families that degrade ECM include serine proteases (e.g., thrombin, cathepsins and plasmin) and metalloproteinases [e.g., MMPs, ADAMs and ADAMTS (Duarte et al. 2015; Edwards et al. 2008; Dubail and Apte 2015; Brix et al. 2015; Beier and Arteel 2012)]. The activity of these proteases are also kept in check by protease inhibitors. For example, MMPs are regulated by tissue inhibitors of metalloproteinases (TIMPs); the elevation of TIMP activity has been shown to partially contribute to excess collagen accumulation during fibrogenesis (Kisseleva and Brenner 2006). MMP-12 activity is similarly regulated by its inhibitor Timp-1, which in toto controls elastin turnover in response to liver injury and/or during fibrogenesis (Pellicoro et al. 2012). Likewise, plasminogen activator inhibitors (e.g., PAI-1) inhibit the activity of the plasminogen activators (uPA/tPA); elevation of PAI-1 levels are sufficient to cause accumulation of fibrin ECM during hepatic injury, even in the absence of increase fibrin ECM deposition [e.g., by thrombin activation (Beier and Arteel 2012)]. Several ECM proteins exist in soluble precursors; their cleavage and activation by proteases can also thereby regulate the deposition of hepatic ECM. The regulation of the deposition insoluble fibrin by the coagulation cascade serves as a canonical example of this point (Beier and Arteel 2012).

3.3 Critical Role of Inflammation in Chronic Liver Disease

The liver is strategically located between the intestinal tract and the rest of the body, which makes it an important physical and biochemical filter between the portal and systemic blood supply. However, in the process of serving as this barrier, the liver is often damaged (Preziosi and Monga 2017; Luedde et al. 2014). To counter this constant exposure to potential hepatotoxicants, the liver has tremendous capacity to regenerate (Preziosi and Monga 2017; Michalopoulos and DeFrances 2005). This ability is unique compared to other solid organs that are far less capable to regenerate when damaged. Liver regeneration involves a highly orchestrated response to facilitate the regenerative process. Perturbation of this complex regenerative response can impair normal tissue recovery after injury or damage (Forbes and Newsome 2016). In the context of repeated injury, if recovery from each event is incomplete, damage can accumulate, which leads to the chronicity of liver disease (Michalopoulos and DeFrances 2005) (Fig. 3.1).

Chronic liver diseases are driven by numerous primary (e.g., diet, alcohol abuse and viral infection) and risk modifying (e.g., genetic variation and environmental exposures, etc.) factors (Kirpich et al. 2015; Lieber et al. 1965; Ganesan et al. 2018; Hajarizadeh et al. 2013; Morrison and Kowdley 2000). Despite divergent underlying factors, chronic liver diseases all share a well-known pathological spectrum, ranging

initially from simple steatosis (fat accumulation), to inflammation and necrosis (steatohepatitis), to fibrosis and cirrhosis (Altamirano and Bataller 2011; Schwartz and Reinus 2012; Poole et al. 2017; Seth et al. 2011). Although the progression of liver disease is well-understood, there is no FDA-approved therapy to halt or reverse this process in humans (Singh et al. 2017). Fibrosis can reverse with successful removal of the primary cause (Fig. 3.1), with the caveat that it is more difficult to reverse severe fibrosis/cirrhosis (Iredale et al. 1998). Indeed, cirrhosis is often considered an end-stage liver disease that will eventually kill the patient, absent a liver transplant. Even in the case of HCV, where removal of the primary causative factor (viral infection) is now nearly 100%, reportedly 30–60% of cirrhotic livers do not recover histologically (Vispo et al. 2009; Grgurevic et al. 2017). Moreover, the clinically-relevant sequelae of severe/decompensated cirrhosis (e.g., portal hypertension) do not appear to reverse after successful removal of the HCV infection (Libanio and Marinho 2017). Moreover, maintenance of compensated cirrhosis (i.e., "stable cirrhotics") greatly increases the risk of development of hepatocellular carcinoma (HCC). These limitations of therapy for chronic liver diseases translate to over 2 million people dying from complications of cirrhosis and related diseases (e.g., HCC) each year (Byass 2014).

Given the above concerns with treating fibrosis/cirrhosis, there is a great need to improve identification of at-risk individuals and prevention of liver diseases progression during earlier phases, especially inflammation. Upregulation of inflammation is a key step delineating between benign liver (i.e., steatosis) and progressive liver diseases. The inflammatory response during chronic liver diseases involves both the innate and adaptive immune responses (Hensley and Deng 2018; Li et al. 2018). In contrast to acute liver injury, inflammation during chronic liver disease is relatively low-grade, in which innate immune cells are activated, and the adaptive immune response is dysregulated (Wree and Marra 2016; Gao et al. 2019; Dong et al. 2019; Pellicoro et al. 2014; Robinson et al. 2016). When this injury overwhelms the ability of the liver to repair/recover from said damage, the chronicity of liver diseases ensues.

3.4 The Role of the Extracellular Matrix in Liver Diseases-More than Fibrosis and Collagen

The study of the role of the extracellular matrix in liver disease has focused predominantly on fibrosis. This is not necessarily a surprise, given that chronic liver disease often does not manifest symptoms until very late in disease progression (Sweet et al. 2017), and that clinical presentation is often only during sequelae associated with end-stage liver disease (see above). Moreover, fibrosis is an overt pathological change that can be observed even macroscopically in the liver. However, the ECM is a dynamic compartment that changes in response to stress well before fibrosis. This concept, as well as the nature and impact of these changes, is

well-understood in some fields (e.g., subcutaneous wound healing), but has been somewhat lagging in the liver field (Sun et al. 2014).

Work by this group and others have shown that acute liver injury also causes dramatic changes to the hepatic ECM (Poole et al. 2017) (Fig. 3.1). Indeed, the acute phase response in the liver involves several ECM proteins, such as fibrin, osteopontin, and fibronectin (Beier et al. 2009; Gillis and Nagy 1997; Thiele et al. 2005). These acute, subhistologic changes to the ECM/matrisome appear to be transitional and resolve after resolution of acute injury (Massey et al. 2017; Poole and Arteel 2016) (Fig. 3.1). In contrast, the ECM associated with chronic injury is collagenous scarring, which does not resolve as readily. This pattern of ECM changes during acute (transitional and temporary) and chronic (scarring and more permanent) injury is in-line qualitatively with subcutaneous wound healing (Sun et al. 2014) (Fig. 3.1). These differences also parallel the changes observed during early-stage and progressive liver disease. In recent years, the quality and frequency of early referrals, as well as improved detection methods, have increased the rate of detection of early-stage asymptomatic liver diseases (Srivastava et al. 2019). This improvement facilitates the opportunity for mechanism-based therapies to halt disease progression during earlier (i.e., prefibrotic) phases of the disease progression (Arteel and Naba 2020).

Research on the hepatic ECM changes during liver disease has primarily focused on the regulation and deposition of collagen. Given that the accumulation of collagen is robust during fibrosis and cirrhosis, and that it is easy to detect histochemically, this focus is not necessarily surprising. However, there is a myriad of ECM proteins that qualitatively and quantitative change during fibrogenesis (Gressner et al. 2007; Gressner and Bachem 1990), and their role(s) in disease progression is not understood well. Moreover, the expanded definition of the ECM to encompass non-fibrillar proteins found in that microenvironmental niche (i.e., matrisome) has not been explored in detail in the context of liver disease (Naba et al. 2016; Arteel and Naba 2020). Taken together, hepatic fibrogenesis is far more complicated than simply collagen accumulation.

3.5 The Hepatic Matrisome and the Control of Inflammation

As mentioned above, inflammation is a gateway pathology to progressive liver disease. Given that inflammatory injury is much more reversible than fibrotic changes, this pathologic stage is a key target that being explored for new therapies and diagnoses. In this context, the hepatic matrisome represents a potential therapeutic target or detect liver disease.

3.5.1 *Maintenance of Structure*

The ECM plays a key direct role in maintaining the overall architecture of the liver, partially through definition of organ boundaries and zones (Federman et al. 2002) (Fig. 3.2). Furthermore, the ECM also indirectly characterizes liver morphology (Julich et al. 2015); during branching morphogenesis of the liver, the matrisome and the glycocalyx on the cell surface coordinate to regulate growth factor-hepatic cell interactions, which drives phenotype of the eventual mature organ (Patel et al. 2017; Rozario and DeSimone 2010). The variation of the ECM within the hepatic lobule is proposed to help define intralobular zones (McClelland et al. 2008; Lee-Montiel et al. 2017). The structure of the ECM defines properties that regulate inflammation. For example, although the basement membrane usually physically impedes inflammatory cell transmigration, the changes in this ECM in response to injury orchestrate homing of inflammatory to the site of injury (Wang et al. 2006).

Invasive cells can also secrete matrix metalloproteinases that degrade ECM, and thereby mediate their extravasation during liver injury [see below (Hamada et al. 2008)]. This degradation also exposes self-antigens (e.g., basement membrane collagens) that is a feed-forward signal for inflammatory cell recruitment (Mak et al. 2016). Moreover, even in situations in which ECM is accumulating, there is an overall increase ECM turnover (Roderfeld 2018). The enhanced rate of ECM

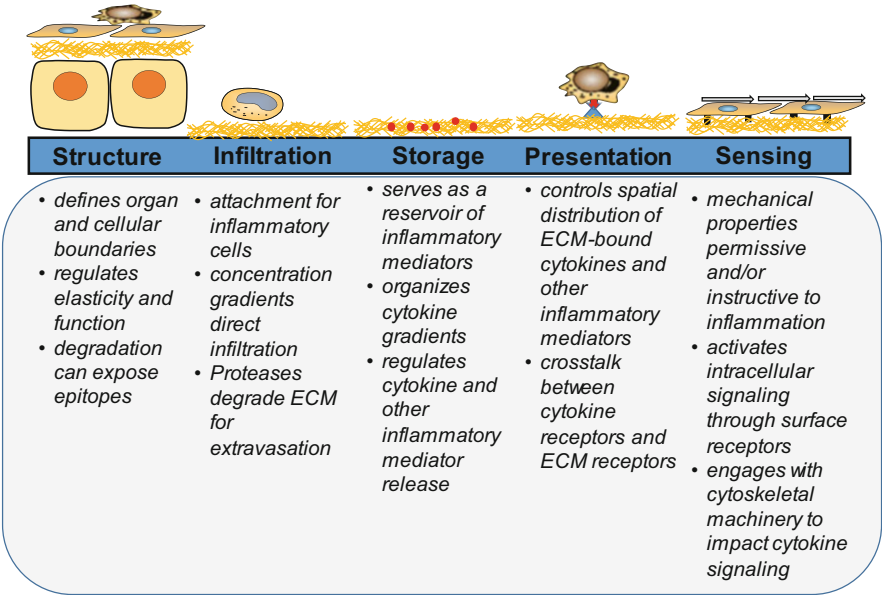

Fig. 3.2 Contributions of the hepatic matrisome/ECM to inflammation. The ECM plays a myriad of roles that directly and indirectly regulate inflammation and injury in the liver. These roles can be generally categorized as functions related to structure, infiltration, storage, presentation and signaling

turnover/degradation can release ECM peptide fragments act as chemotractants to inflammatory (Song et al. 1993; Santambrogio and Rammensee 2019) (Fig. 3.1).

The structure of the ECM also contributes to the regulation of inflammation indirectly via altering the integrity and/or elasticity to the liver. Remodeling of the hepatic ECM in response to injury can alter the overall structure of the ECM, which can translate to changes in elasticity (Klaas et al. 2016), and inflammation directly increases ECM stiffness in organs (Wu and Birukov 2019; Karki and Birukova 2018; Mammoto et al. 2013; Hsu et al. 2016). Even acute liver injury alters ECM structural components and impacts organ elasticity (Klaas et al. 2016). Indeed, although liver stiffness is most often associated with fibrogenic changes to the liver, inflammation also impacts assessments of liver stiffness (e.g., transient elastography); these changes are often viewed as 'false positive' signals during these measurements (Coco et al. 2007; Grgurevic et al. 2017). However, it is likely that the increase in liver stiffness measurements caused by inflammation is at least in part, a true signal (Dolin and Arteel 2020). Emerging technologies, such as three-dimensional magnetic resonance elastography (3D-MRE) and magnetic resonance imaging proton density fat fraction (MRI-PDFF), are being developed to facilitate noninvasive differentiation between inflammation and fibrosis (Allen et al. 2018).

3.5.2 Facilitation of Infiltration

Hepatic inflammation after injury (i.e., "sterile" inflammation) involves innate immune cells (e.g., natural killer cells, natural killer T cells, dendritic cells, neutrophils, eosinophils and monocytes) that are recruited to the liver (Oliveira et al. 2018; Karlmark et al. 2009). These immune cells bind to several ECM proteins through several types of receptors, including integrins and surface glycoproteins (e.g., CD54, CD44 and CD26) that are arrayed on the recruited cells (Shimizu and Shaw 1991) (Fig. 3.2). These interactions have important implications in liver disease and their modulation can be employed as a therapeutic strategy; for example, inhibition of T-cell binding to fibronectin is considered to be, at least in part, responsible for the anti-inflammatory effect of the drug, pentoxifylline (Shirin et al. 1998).

Interactions between leukocytes and the ECM play key roles in the adhesion, transmigration and phenotype of infiltrating leukocytes (Ley et al. 2007). ECM receptors on the surface of leukocytes direct their migration through interaction with the ECM (Shimizu and Shaw 1991). The regulation of expression and location of these receptors is critical for the rapid phenotypic change between adhesive and nonadhesive states of immune cells required during inflammation (Shimizu and Shaw 1991). Selectins (CD62) and other receptors mediate initial leukocyte capture and rolling in the microvasculature (Lee and Kubes 2008; Ley et al. 2007; Wong et al. 1997; Fox-Robichaud and Kubes 2000). Leukocyte adhesion is predominantly dependent on the interaction between β_1- and β_2-integrins and the ECM (Lee and Kubes 2008), as well as CD44 and vascular adhesion protein-1 (Lee and Kubes 2008). The interaction between the ECM and cell infiltration is not unidirectional,

but rather dynamic, in that as leukocytes respond to cues from the ECM, they in turn release matrix-degrading proteases (Woodfin et al. 2010) that alter the extracellular composition and allow for easier extravasation.

Chemokines also play an important role in the activation step of leukocyte adhesion by interacting with the ECM (Ley et al. 2007; Proudfoot et al. 2003). The release of these chemokines via direct or indirect (e.g., ECM proteolysis) creates a haptotactic gradient that is critical for immune cell chemotraction (Monneau et al. 2016). Almost all hepatic cells secrete chemokines under basal conditions and in response to injury (Oliveira et al. 2018). Chemokines are retained in the matrisome via binding to the glycosaminoglycan (GAG)/heparin sulfate components found in the space of Disse (Monneau et al. 2016; Heydtmann et al. 2005). ECM proteins themselves may also possess chemotactic functions. For example, osteopontin is chemotactic to natural killer T cells, neutrophils, and macrophages (Ramaiah and Rittling 2008), while fibronectin activates macrophage and directs monocyte and neutrophil translocation (Godfrey 1990).

3.5.3 Management of Storage, Presentation and Sensing

The ECM also is a reservoir for signaling molecules, such as growth factors, cytokines and chemokines, that maintain homeostasis and respond to dyshomeostasis (Fig. 3.2). The ECM stores these proteins, which predominantly bind to glycosaminoglycans (GAG), which shield them from targeted degradation (Lipowsky 2018). Injury activates proteases (e.g., MMPs and ADAM) that cleave these linkages, thereby rapidly releasing these mediators (Karsdal et al. 2015; Sorokin 2010; Vempati et al. 2014; Lipowsky 2018). The localized release of these mediators also contributes to the above-mentioned haptotactic gradient that directs inflammatory cells to the origin of the injury (Monneau et al. 2016; Vempati et al. 2014; Wasmuth et al. 2010). Indeed, the initial release acute phase proteins in response to injury/dyshomeostasis is via proteolysis of stored precursors rather than by de novo synthesis. The interactions between these factors and the ECM also serve to present or restrict access of ligands to receptors, to modulate the spatial distribution of growth factors or to create chemotactic gradients (Rozario and DeSimone 2010; Dolin and Arteel 2020).

The ECM behaves dynamically as a signaling moiety that mediates both outside-in and inside-out signaling between the cell and the environment. As a family, the integrins play key roles in mediating these interactions. Integrins transfer information from the ECM to the cell, allowing rapid and dynamic responses to changes in the extracellular environment (Humphries et al. 2006). Integrins play a myriad of roles within the body, including proliferation/angiogenesis, maintenance of differentiation, as well as inflammation and apoptosis (Hodivala-Dilke et al. 2003; Zhou et al. 2009). Altered/dysregulated integrin signaling is hypothesized to be involved in all stages of the progression of chronic liver diseases (Patsenker and Stickel 2011). There are also several non-integrin receptors involved in signaling between the ECM

and the cell. For example, CD44, a type I transmembrane glycoprotein has been demonstrated to be involved in liver disease and inflammation via interacting with its canonical ligand, hyaluronic acid (HA) (Seth et al. 2014; Patouraux et al. 2017). Interactions between this ECM glycosaminoglycan and CD44 are known to facilitate migration of leukocytes to inflamed tissue, as well as the progression of inflammatory injury (McDonald and Kubes 2015). Interestingly, CD44 has also been implicated in the resolution of injury by facilitating the migration of hematopoietic stem cells to the injured liver (Crosby et al. 2009).

The interaction between cells and the surrounding ECM can also impact downstream signaling cascades that mediate both injurious and restorative signals (Dolin and Arteel 2020). This control can be at mediated via altering receptor affinity, or changes to downstream signaling cascades. Under basal conditions, receptors for these mediators are generally dispersed on the plasma membrane in lipid/lipoprotein-rich regions (i.e., lipid rafts); the relatively close proximity of receptor monomers facilitates ligand binding, receptor dimerization and subsequent downstream signaling (Simons and Toomre 2000). It has been recently suggested that ECM proteins contribute to this 2-dimensional organization on the plasma membrane (Sadeghi and Vink 2015). Signal integration between integrins and extracellular signaling factors also varies with interactions with the ECM stratum. This influence of ECM on signaling has best been described for cellular responses to growth factors, and is categorized as concomitant signaling, collaborative activation, direct activation, amplification and negative regulation (Ivaska and Heino 2011; Schnittert et al. 2018). Chronic inflammation impairs growth factor signaling, in part by altering the make-up of the ECM surrounding the cell (Ozaki et al. 2011). Moreover, ECM interactions qualitatively and quantitatively influence the response of TLR and TNFα signaling (Gay and Gangloff 2007).

In addition to directly and indirectly influencing signaling transduction cascades, integrin/ECM complexes form linkages to B-actin, and other components of the cytoskeleton (Harburger and Calderwood 2009; Iwamoto and Calderwood 2015). These complexes facilitate the clustering of integrins into focal adhesions, which further influences normal responses to development, growth and maintenance signals (Lorenz et al. 2018; Harburger and Calderwood 2009; Iwamoto and Calderwood 2015). Loss of control of this process is a key step to permit unregulated clonal expansion of mutated cells during carcinogenesis [i.e., anchorage independent growth (Hamidi and Ivaska 2018; Reddig and Juliano 2005)]. This vertical integration of ECM with the cytoskeleton via integrins is also a key component of mechanosensing, and likely is responsible, at least in part, for the impact of ECM rigidity on the inflammatory response [see above; (Lorenz et al. 2018)].

Inflammation and ECM remodeling likely perpetuate a positive "vicious cycle" (Sorokin 2010). The ECM modulates and regulates immune cell chemotaxis, transmigration and differentiation. These activated immune cells alter the ECM composition by triggering both de novo ECM deposition, as well as proteolytic degradation. Cleaved ECM produced by this increase in ECM turnover can be proinflammatory in their own right, and serve as alarmins to distal targets. Although these processes can be adaptive and beneficial in wound healing and recover,

aberrant ECM accumulation/alteration and overproduction of ECM degradation products can perpetuate a maladaptive inflammatory response, such as is observed during chronic hepatic inflammation (Schuster et al. 2018). Thus, the ECM not only plays a key role in mediating inflammation, but also in the resolution of inflammation [i.e., catabasis (Widgerow 2012; Franitza et al. 2000; Canedo-Dorantes and Canedo-Ayala 2019)].

3.6 ECM Remodeling and the "Degradome"

The ECM is a dynamic compartment subject to constant protein turnover. This turnover can undergo more dramatic, rapid changes (i.e. remodeling) during inflammation and disease (Fig. 3.1). One important means of regulation of ECM turnover is via activation of ECM proteases (e.g. MMPs; see above). All hepatic cells release different types of proteases and protease inhibitors in response to normal/abnormal conditions (Benyon and Arthur 2001; Calabro et al. 2014; Zang et al. 2015; Ramachandran et al. 2012). This altered turnover not only impacts the makeup of the ECM/matrisome, but also the changes the pattern of degraded ECM peptides found in biological fluids [e.g., blood (Sand et al. 2016)]. The activities of these changes have the potential to be experimental imputed through use of algorithms such as Proteasix (http://www.proteasix.org/) that relies of compiled protease and peptidase databases (Merops database, https://www.ebi.ac.uk/merops/) for substrate specificity to impute endo- and exopeptide activity contributing to observed peptide amino- and carboxy-termini. The potential of these protease degradation products (i.e. the 'degradome') to serve as indices of various diseases is becoming increasingly understood (Fig. 3.1). For example, the rate of ECM turnover of anchored ECM into soluble ECM have been demonstrated to strongly influence tumor growth and morphology (Nargis et al. 2018). Understanding the role of the ECM degradome in disease is facilitated by modern mass spectrometry methods that allow widespread characterization of the degradome (i.e. peptidomics or 'degradomics') (Randles and Lennon 2015). Even if degradation products do not play a critical role in disease mechanisms, they may be useful surrogate biomarkers.

3.7 Proteomic Analysis of the Hepatic Matrisome

3.7.1 Overview

Formation and stabilization of the extracellular matrix (ECM) proteome is guided by protein-protein specific interactions and stabilized by the presence of protein post-translational modifications including glycosylation or intra- and inter-molecular protein cross-linking. Collectively these interactions yield a structurally stabilized three-dimensional (3D) matrix that significantly contributes to both the mechanical

and biologic properties of tissues. These stabilizing features while biological beneficial contribute to the limitations to the proteomic experiment: low solubility, high prevalence of structural proteins such as collagens, the apparent spatial stochasticism of low abundance proteins, post-translational modifications and truncations, and protein-protein cross-linking. The current advances in the field of extracellular matrix proteomics has developed from approaches to improve segmenting the core versus associated matrix proteins as well approaches to enhance protein identification and quantification (Naba et al. 2012; Hynes and Naba 2012).

The effective isolation of the ECM from cellular and non-matrisomal extracellular proteomes is a critical component for comparative proteomic studies. Whole tissue studies traditionally have relied on biochemical methods such as differential extraction (Massey et al. 2017) while physical methods such as laser capture dissection methods allow for spatially resolved isolation of histologically specific tissue compartments (Hobeika et al. 2017). Differential ECM extraction from whole tissue has been based on methods adopted from regenerative medicine that required purified ECM as molecular scaffolds to support artificial organ growth (Gilbert et al. 2006; Sullivan et al. 2012; Willemse et al. 2020; Verstegen et al. 2017). Classically these methods have heavily utilized neutral ionic (phosphate buffered sodium dodecyl sulfate) or ammonical non-ionic (ammonium hydroxide buffered triton X-100) detergent-based decellularization buffers to solubilize and extract all cellular proteins. Several groups including ours have adapted and improved these chemical methods based on differential solubilization approaches to sequentially isolate the structural and associated ECM proteomes (Massey et al. 2017; Didangelos et al. 2011; de Castro Bras et al. 2013). Following sample decellularization, the insoluble fraction is extracted with salts, acids or chaotropes (guanidine hydrochloride) and then enzymatically deglycosylated prior to proteomic analyses. These steps yields fractions used to understand the disease associated pathobiology resolved into soluble, ECM-associated, and an insoluble ECM proteomic components.

Proteomics has been used to address two fundamental questions in ECM biology: (a) what proteins are present and (b) what are the relative abundances within the ECM. A thorough discussion of proteomic methods is beyond the scope of this review but are available elsewhere (Ankney et al. 2018; Cox and Mann 2011; Rauniyar and Yates 2014). The majority of all published ECM proteomic studies defining the matrisome (Naba et al. 2012) have utilized sequential extraction, proteolytic digestion, and one-dimensional low pH, reversed phase liquid chromatography-mass spectrometry analysis using a LTQ-Orbitrap hybrid mass analyzer (Shao et al. 2020). These studies have approached the proteomics with a label-free approach with a data dependent acquisition method based on rank ordered lists of peptide signal intensity to select ions for fragmentation. The peptide relative quantification based on high resolution mass spectrometry data is derived from peak signal intensity or a peptide extracted area under the curve. The deduced amino acid sequence of the peptide was based on spectrum matching to theoretical or empirically established spectra.

3.7.2 3-Step ECM Extraction (Fig. 3.3)

Sequential extraction of the hepatic ECM was modified from the protocol described by de Castro Bras et al. for heart tissue (de Castro Bras et al. 2013). The original description of this technique was first published in detail in a Dissertation (Massey 2014).

Sample Preparation and Wash Snap frozen liver tissue (75–100 mg) was immediately added to ice-cold phosphate-buffered saline (pH 7.4) wash buffer containing commercially available protease and phosphatase inhibitors (Sigma Aldrich) and 25 mM EDTA to inhibit proteinase and metalloproteinase activity, respectively. While immersed in wash buffer, liver tissue was diced into small fragments using a scalpel. The diced sample was washed 5 times to remove contaminants. Between washes, samples were pelleted by centrifugation (12,000×*g*, 5 min), and wash buffer was decanted.

NaCl Extraction Diced samples were incubated in 10 volumes of 0.5 M NaCl buffer, containing 10 mM Tris HCl (pH 7.5), proteinase/phosphatase inhibitors, and 25 mM EDTA. The samples were mildly mixed on a plate shaker (800 rpm) overnight at room temperature. The following day, the remaining tissue pieces

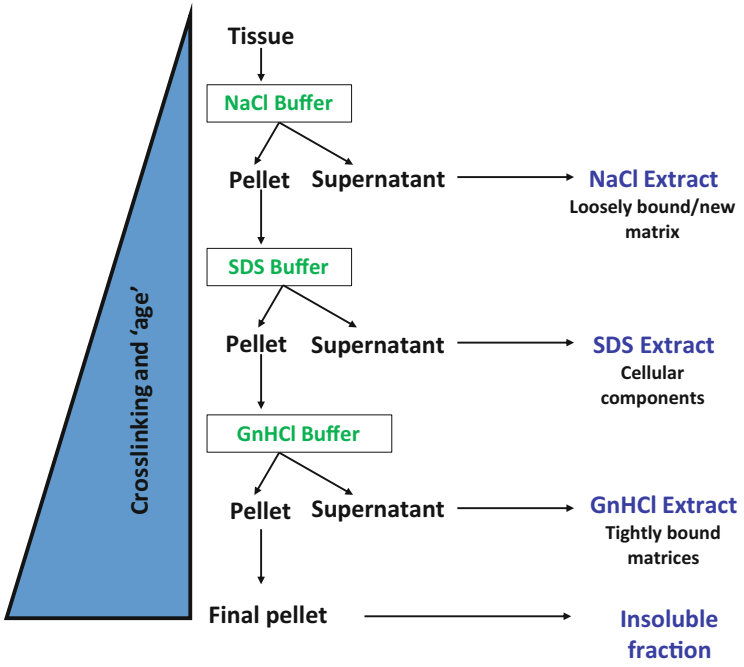

Fig. 3.3 Sequential extraction of the hepatic ECM. By using increasingly rigorous extraction solutions, the matrisome can be separated into components based on their solubility, which is inversely proportional to the level of crosslinking and 'age'

were pelleted by centrifugation (10,000×*g* for 10 min). The pellet was used for the SDS extraction (see below). The supernatant was collected and desalted using ZebaSpin columns (Pierce) according to manufacturer's instructions. To precipitate proteins, desalted supernatant was incubated with 5× supernatant volume of 100% acetone overnight at −20 °C, centrifuged (16,000×*g*, 45 min), and dried in a rotary evaporator. Proteins were resuspended in deglycosylation buffer.

SDS Extraction The pellet from the NaCl extraction was subsequently incubated in 10 volumes (based on original weight) of 1% SDS solution, containing proteinase/phosphatase inhibitors and 25 mM EDTA. The samples were mildly mixed on a plate shaker (800 rpm) overnight at room temperature. The following day, the remaining tissue pieces were pelleted by centrifugation at 10,000×*g* for 10 min. The pellet was used for the GnHCl extraction (see below). The supernatant was collected and desalted using ZebaSpin columns (Pierce) according to manufacturer's instructions. To precipitate proteins, desalted supernatant was incubated with 5× supernatant volume of 100% acetone overnight at −20 °C, centrifuged (16,000×*g*, 45 min), and dried in a rotary evaporator. Proteins were resuspended in deglycosylation buffer.

Guanidine HCl Extraction The pellet from the SDS extraction was incubated with 5 volumes (based on original weight) of a denaturing guanidine buffer containing 4 M guanidine HCl (pH 5.8), 50 mM sodium acetate, 25 mM EDTA, and proteinase/phosphatase inhibitors. The samples were vigorously mixed on a plate shaker at 1200 rpm for 48 h at room temperature; vigorous shaking is necessary at this step to aid in the mechanical disruption of ECM components. The remaining insoluble components were pelleted by centrifugation at 10,000×*g* for 10 min. This insoluble pellet was retained and solubilized as described below. To precipitate proteins, the supernatant was mixed with 6× supernatant volume of 100% ice cold ethanol overnight at 20 °C, centrifuged (16,000×*g*, 45 min), and washed with 90% ethanol. Pellets were dried in a rotary evaporator and resuspended in deglycosylation buffer.

Deglycosylation and Solubilization The supernatants from each extraction were dried in a rotary evaporator and resuspended in deglycosylation buffer containing 150 mM NaCl, 50 mM sodium acetate, 10 mM EDTA, and proteinase/phosphatase inhibitors. Resuspended samples were desalted using ZebaSpin columns (Pierce) according to manufacturer's instructions. The desalted extracts were then mixed with 5 volumes of 100% acetone and stored at −20 °C overnight to precipitate proteins. The precipitated proteins were pelleted by centrifugation at 16,000×*g* for 45 min. Acetone was evaporated by vacuum drying in a rotary evaporator for 1 h. Dried protein pellets were resuspended in 500 μL deglycosylation buffer containing 150 mM NaCl, 50 mM sodium acetate, pH 6.8, 10 mM EDTA, and proteinase/phosphatase inhibitors that contained chondroitinase ABC (*P. vulgaris*; 0.025 U/sample), endo-beta-galactosidase (*B. fragilis*; 0.01 U/sample) and heparitinase II (*F. heparinum*; 0.025 U/sample). Samples were incubated overnight at 37 °C. 20 μL DMSO was added to the insoluble fraction (pellet from guanidine HCl extraction) to aide in solubilization. Protein concentrations were estimated by absorbance at

280 nm using bovine serum albumin (BSA) in deglycosylation buffer for reference standards.

3.7.3 Sample Cleanup and Preparation for Liquid Chromatography and Mass Spectrometry

Liver ECM extracts in deglycosylation buffer were pooled by experimental group and subsequently analyzed by the University of Louisville Proteomics Biomarkers Discovery Core (PBDC). At the PBDC, samples in deglycosylation buffer were thawed to room temperature and clarified by centrifugation at 5000×*g* for 5 min at 4 °C. 50 μL (25 μg) of each sample were reduced by adding 5.55 μL of 1 M DTT and incubating at 60 °C for 30 min. 144.45 μL of 8 M urea in 0.1 M Tris-HCl, pH 8.5, was added to each sample. Each reduced and diluted sample was digested with a modified Filter-Aided Sample Preparation (FASP) method developed by Jacek R. Wisniewski et al. (2009). Recovered material was dried in a rotary evaporator and redissolved in 200 μL of 2% (v/v) acetonitrile (ACN)/0.4% formic acid (FA). The samples were then trap-cleaned with a C18 PROTO™ 300 Å Ultra MicroSpin Column (The Nest Group, Southborough, MA). The sample eluates were stored at −80 °C for 30 min, dried in a rotary evaporator, and stored at −80 °C. Before liquid chromatography, dried samples were warmed to room temperature and dissolved in 2% (v/v) ACN/0.1% FA to a final concentration of 0.25 μg/μL. 16 μL (4 μg) of sample was injected into the Orbitrap Elite.

3.7.4 Liquid Chromatography and Tandem Mass Spectrometry

At the PBDC, liver digest samples were separated on a Dionex Acclaim PepMap 100 75 μm × 2 cm nanoViper (C18, 3 μm, 100 Å) trap and Dionex Acclaim PepMap RSLC 50 μM × 15 cm nanoViper (C18, 2 μm, 100 Å) separating columns. An EASY n-LC (Thermo, Waltham, MA) UHPLC system was used with buffer A = 2% (v/v) acetonitrile/0.1% (v/v) formic acid and buffer B = 80% (v/v) acetonitrile/0.1% (v/v) formic acid as mobile phases. Following injection of the sample onto the trap, separation was accomplished with a 140 min linear gradient from 0% B to 50% B, followed by a 30 min linear gradient from 50% B to 95% B, and lastly a 10 min wash with 95% B. A 40 mm stainless steel emitter (Thermo, Waltham, MA; P/N ES542) was coupled to the outlet of the separating column. A Nanospray Flex source (Thermo, Waltham, MA) was used to position the end of the emitter near the ion transfer capillary of the mass spectrometer. The ion transfer capillary temperature of the mass spectrometer was set at 225 °C, and the spray voltage was set at 1.6 kV.

An Orbitrap Elite—ETD mass spectrometer (Thermo) was used to collect data from the LC eluate. An Nth Order Double Play with ETD Decision Tree method was created in Xcalibur v2.2. Scan event one of the method obtained an FTMS MS1 scan for the range 300–2000 m/z. Scan event two obtained ITMS MS2 scans on up to ten peaks that had a minimum signal threshold of 10,000 counts from scan event one. A decision tree was used to determine whether collision induced dissociation (CID) or electron transfer dissociation (ETD) activation was used. An ETD scan was triggered if any of the following held: an ion had charge state 3 and m/z less than 650, an ion had charge state 4 and m/z less than 900, an ion had charge state 5 and m/z less than 950, or an ion had charge state greater than 5; a CID scan was triggered in all other cases. The lock mass option was enabled (0% lock mass abundance) using the 371.101236 m/z polysiloxane peak as an internal calibrant.

3.7.5 Informatics

The hepatic ECM mass spectrometry data were analyzed at the University of Louisville PBDC using Proteome Discoverer v1.4.0.288. The database used in Mascot v2.4 and SequestHT searches was the 6/2/2015 version of the UniprotKB *Mus musculus* reference proteome canonical and isoform sequences. +57 on C (Carbamidomethylation) was selected as a fixed modification, and +1 on N (Asn- > Asp) and +16 on MP (Oxidation) were selected as variable modifications. A maximum of two missed cleavages were allowed. A Target Decoy PSM Validator node was included in the Proteome Discoverer workflow in order to estimate the false discovery rate (FDR).

The Proteome Discoverer analysis workflow allows for extraction of MS2 scan data from the Xcalibur RAW file, separate searches of CID and ETD MS2 scans in Mascot and Sequest, and collection of the results into a single file (.msf extension). The resulting.msf files from Proteome Discoverer were loaded into Scaffold Q + S v4.3.2. Scaffold was used to calculate the FDR using the Peptide and Protein Prophet algorithms. Protein identification probability of the sequences was set to >95% on the software. The results were annotated with mouse gene ontology (GO) information from the Gene Ontology Annotations Database.

3.8 Summary and Conclusions

In conclusion, the ECM should not be viewed as simply a structural element of the liver, but rather as a microenvironmental niche that dynamically responds to changes in homeostasis. Although it is well understood in some areas that the ECM changes during hepatic injury, most work to date in the liver has focused on changes to the ECM during collagenic fibrosis. Although it is clear that fibrosis is highly relevant to clinical liver disease, it is best viewed as an end-stage of disease pathogenesis, and is

arguably the least sensitive to external interventions (Mehal and Schuppan 2015). Recent improvements in detecting earlier stages of liver diseases enhance the viability of blunting/reversing disease progression well before fibrosis. Inflammation is a key stage of progression that could be targeted therapeutically in this context. Changes in the ECM/matrisome during inflammation are key to regulate and mediate the inflammatory response. However, there are critical gaps in our knowledge on the role of changes to the hepatic ECM in chronic inflammation. There is an opportunity to cross-fertilize our understanding from other fields in which the ECM and inflammation are more well described (Hyldig et al. 2017; Lumelsky et al. 2018; Rousselle et al. 2018; Dolin and Arteel 2020). These other fields may provide new therapies that can be repurposed for chronic liver diseases (Pritchard and McCracken 2015).

References

Allen AM, Shah VH, Therneau TM, Venkatesh SK, Mounajjed T, Larson JJ, Mara KC, Schulte PJ, Kellogg TA, Kendrick ML, McKenzie TJ, Greiner SM, Li J, Glaser KJ, Wells ML, Chen J, Ehman RL, Yin M (2018) The role of three-dimensional magnetic resonance elastography in the diagnosis of nonalcoholic steatohepatitis in obese patients undergoing bariatric surgery. Hepatology. https://doi.org/10.1002/hep.30483

Altamirano J, Bataller R (2011) Alcoholic liver disease: pathogenesis and new targets for therapy. Nat Rev Gastroenterol Hepatol 8(9):491–501. https://doi.org/10.1038/nrgastro.2011.134

Ankney JA, Muneer A, Chen X (2018) Relative and absolute quantitation in mass spectrometry-based proteomics. Annu Rev Anal Chem (Palo Alto, Calif) 11(1):49–77. https://doi.org/10.1146/annurev-anchem-061516-045357

Arteel GE, Naba A (2020) The liver matrisome, looking beyond collagens. JHEP Rep. https://doi.org/10.1016/j.jhepr.2020.100115

Beier JI, Arteel GE (2012) Alcoholic liver disease and the potential role of plasminogen activator inhibitor-1 and fibrin metabolism. Exp Biol Med (Maywood) 237(1):1–9. https://doi.org/10.1258/ebm.2011.011255

Beier JI, Luyendyk JP, Guo L, von Montfort C, Staunton DE, Arteel GE (2009) Fibrin accumulation plays a critical role in the sensitization to lipopolysaccharide-induced liver injury caused by ethanol in mice. Hepatology 49(5):1545–1553. https://doi.org/10.1002/hep.22847

Benyon RC, Arthur MJ (2001) Extracellular matrix degradation and the role of hepatic stellate cells. Semin Liver Dis 21(3):373–384. https://doi.org/10.1055/s-2001-17552

Bonnans C, Chou J, Werb Z (2014) Remodelling the extracellular matrix in development and disease. Nat Rev Mol Cell Biol 15(12):786–801. https://doi.org/10.1038/nrm3904

Brix K, McInnes J, Al-Hashimi A, Rehders M, Tamhane T, Haugen MH (2015) Proteolysis mediated by cysteine cathepsins and legumain-recent advances and cell biological challenges. Protoplasma 252(3):755–774. https://doi.org/10.1007/s00709-014-0730-0

Byass P (2014) The global burden of liver disease: a challenge for methods and for public health. BMC Med 12:159. https://doi.org/10.1186/s12916-014-0159-5

Calabro SR, Maczurek AE, Morgan AJ, Tu T, Wen VW, Yee C, Mridha A, Lee M, d'Avigdor W, Locarnini SA, McCaughan GW, Warner FJ, McLennan SV, Shackel NA (2014) Hepatocyte produced matrix metalloproteinases are regulated by CD147 in liver fibrogenesis. PLoS One 9(7):e90571. https://doi.org/10.1371/journal.pone.0090571

Canedo-Dorantes L, Canedo-Ayala M (2019) Skin acute wound healing: a comprehensive review. Int J Inflamm 2019:3706315. https://doi.org/10.1155/2019/3706315

Cassiman D, Libbrecht L, Desmet V, Denef C, Roskams T (2002) Hepatic stellate cell/myofibroblast subpopulations in fibrotic human and rat livers. J Hepatol 36(2):200–209. https://doi.org/10.1016/s0168-8278(01)00260-4

Coco B, Oliveri F, Maina AM, Ciccorossi P, Sacco R, Colombatto P, Bonino F, Brunetto MR (2007) Transient elastography: a new surrogate marker of liver fibrosis influenced by major changes of transaminases. J Viral Hepat 14(5):360–369. https://doi.org/10.1111/j.1365-2893.2006.00811.x

Cox J, Mann M (2011) Quantitative, high-resolution proteomics for data-driven systems biology. Annu Rev Biochem 80:273–299. https://doi.org/10.1146/annurev-biochem-061308-093216

Crosby HA, Lalor PF, Ross E, Newsome PN, Adams DH (2009) Adhesion of human haematopoietic (CD34(+)) stem cells to. Human liver compartments is integrin and CD44 dependent and modulated by CXCR3 and CXCR4. J Hepatol 51(4):734–749. https://doi.org/10.1016/j.jhep.2009.06.021

de Castro Bras LE, Ramirez TA, DeLeon-Pennell KY, Chiao YA, Ma Y, Dai Q, Halade GV, Hakala K, Weintraub ST, Lindsey ML (2013) Texas 3-step decellularization protocol: looking at the cardiac extracellular matrix. J Proteome 86:43–52. https://doi.org/10.1016/j.jprot.2013.05.004

Didangelos A, Yin X, Mandal K, Saje A, Smith A, Xu Q, Jahangiri M, Mayr M (2011) Extracellular matrix composition and remodeling in human abdominal aortic aneurysms: a proteomics approach. MCP 10(8):M111 008128. https://doi.org/10.1074/mcp.M111.008128

Dolin CE, Arteel GE (2020) The matrisome, inflammation and liver disease. Semin Liver Dis 40:180–188. https://doi.org/10.1055/s-0039-3402516

Dong X, Liu J, Xu Y, Cao H (2019) Role of macrophages in experimental liver injury and repair in mice. Exp Ther Med 17(5):3835–3847. https://doi.org/10.3892/etm.2019.7450

Duarte S, Baber J, Fujii T, Coito AJ (2015) Matrix metalloproteinases in liver injury, repair and fibrosis. Matrix Biol 44-46:147–156. https://doi.org/10.1016/j.matbio.2015.01.004

Dubail J, Apte SS (2015) Insights on ADAMTS proteases and ADAMTS-like proteins from mammalian genetics. Matrix Biol 44-46:24–37. https://doi.org/10.1016/j.matbio.2015.03.001

Edwards DR, Handsley MM, Pennington CJ (2008) The ADAM metalloproteinases. Mol Asp Med 29(5):258–289. https://doi.org/10.1016/j.mam.2008.08.001

Federman S, Miller LM, Sagi I (2002) Following matrix metalloproteinases activity near the cell boundary by infrared micro-spectroscopy. Matrix Biol 21(7):567–577

Forbes SJ, Newsome PN (2016) Liver regeneration - mechanisms and models to clinical application. Nat Rev Gastroenterol Hepatol 13(8):473–485. https://doi.org/10.1038/nrgastro.2016.97

Fox-Robichaud A, Kubes P (2000) Molecular mechanisms of tumor necrosis factor alpha-stimulated leukocyte recruitment into the murine hepatic circulation. Hepatology 31(5):1123–1127. https://doi.org/10.1053/he.2000.6961

Franitza S, Hershkoviz R, Kam N, Lichtenstein N, Vaday GG, Alon R, Lider O (2000) TNF-alpha associated with extracellular matrix fibronectin provides a stop signal for chemotactically migrating T cells. J Immunol 165(5):2738–2747. https://doi.org/10.4049/jimmunol.165.5.2738

Friedman SL (2010) Extracellular Matrix. In: Dufour JF, Clavien PA (eds) Signaling pathways in liver diseases. Springer, Berlin, pp 93–104. https://doi.org/10.1007/978-3-642-00150-5_6

Ganesan M, Poluektova LY, Kharbanda KK, Osna NA (2018) Liver as a target of human immunodeficiency virus infection. World J Gastroenterol 24(42):4728–4737. https://doi.org/10.3748/wjg.v24.i42.4728

Gao B, Ahmad MF, Nagy LE, Tsukamoto H (2019) Inflammatory pathways in alcoholic steatohepatitis. J Hepatol 70(2):249–259. https://doi.org/10.1016/j.jhep.2018.10.023

Gay NJ, Gangloff M (2007) Structure and function of toll receptors and their ligands. Annu Rev Biochem 76:141–165. https://doi.org/10.1146/annurev.biochem.76.060305.151318

Gilbert TW, Sellaro TL, Badylak SF (2006) Decellularization of tissues and organs. Biomaterials 27(19):3675–3683. https://doi.org/10.1016/j.biomaterials.2006.02.014

Gillis SE, Nagy LE (1997) Deposition of cellular fibronectin increases before stellate cell activation in rat liver during ethanol feeding. Alcohol Clin Exp Res 21(5):857–861

Godfrey HP (1990) T cell fibronectin: an unexpected inflammatory lymphokine. Lymphokine Res 9 (3):435–447
Gressner AM, Bachem MG (1990) Cellular sources of noncollagenous matrix proteins: role of fat-storing cells in fibrogenesis. Semin Liver Dis 10(1):30–46. https://doi.org/10.1055/s-2008-1040455
Gressner OA, Weiskirchen R, Gressner AM (2007) Evolving concepts of liver fibrogenesis provide new diagnostic and therapeutic options. Comp Hepatol 6:7. https://doi.org/10.1186/1476-5926-6-7
Grgurevic I, Bozin T, Madir A (2017) Hepatitis C is now curable, but what happens with cirrhosis and portal hypertension afterwards? Clin Exp Hepatol 3(4):181–186. https://doi.org/10.5114/ceh.2017.71491
Griffiths MR, Keir S, Burt AD (1991) Basement membrane proteins in the space of Disse: a reappraisal. J Clin Pathol 44(8):646–648. https://doi.org/10.1136/jcp.44.8.646
Hajarizadeh B, Grebely J, Dore GJ (2013) Epidemiology and natural history of HCV infection. Nat Rev Gastroenterol Hepatol 10(9):553–562. https://doi.org/10.1038/nrgastro.2013.107
Hamada T, Fondevila C, Busuttil RW, Coito AJ (2008) Metalloproteinase-9 deficiency protects against hepatic ischemia/reperfusion injury. Hepatology 47(1):186–198. https://doi.org/10.1002/hep.21922
Hamidi H, Ivaska J (2018) Every step of the way: integrins in cancer progression and metastasis. Nat Rev Cancer 18(9):533–548. https://doi.org/10.1038/s41568-018-0038-z
Harburger DS, Calderwood DA (2009) Integrin signalling at a glance. J Cell Sci 122(Pt 2):159–163. https://doi.org/10.1242/jcs.018093
Harvey A, Montezano AC, Lopes RA, Rios F, Touyz RM (2016) Vascular fibrosis in aging and hypertension: molecular mechanisms and clinical implications. Can J Cardiol 32(5):659–668. https://doi.org/10.1016/j.cjca.2016.02.070
Hensley MK, Deng JC (2018) Acute on chronic liver failure and immune dysfunction: a mimic of sepsis. Semin Respir Crit Care Med 39(5):588–597. https://doi.org/10.1055/s-0038-1672201
Heydtmann M, Lalor PF, Eksteen JA, Hubscher SG, Briskin M, Adams DH (2005) CXC chemokine ligand 16 promotes integrin-mediated adhesion of liver-infiltrating lymphocytes to cholangiocytes and hepatocytes within the inflamed human liver. J Immunol 174 (2):1055–1062. https://doi.org/10.4049/jimmunol.174.2.1055
Hobeika L, Barati MT, Caster DJ, McLeish KR, Merchant ML (2017) Characterization of glomerular extracellular matrix by proteomic analysis of laser-captured microdissected glomeruli. Kidney Int 91(2):501–511. https://doi.org/10.1016/j.kint.2016.09.044
Hodivala-Dilke KM, Reynolds AR, Reynolds LE (2003) Integrins in angiogenesis: multitalented molecules in a balancing act. Cell Tissue Res 314(1):131–144. https://doi.org/10.1007/s00441-003-0774-5
Hsu JJ, Lim J, Tintut Y, Demer LL (2016) Cell-matrix mechanics and pattern formation in inflammatory cardiovascular calcification. Heart 102(21):1710–1715. https://doi.org/10.1136/heartjnl-2016-309667
Huijberts MS, Schaper NC, Schalkwijk CG (2008) Advanced glycation end products and diabetic foot disease. Diabetes Metab Res Rev 24(Suppl 1):S19–S24. https://doi.org/10.1002/dmrr.861
Humphries JD, Byron A, Humphries MJ (2006) Integrin ligands at a glance. J Cell Sci 119 (19):3901–3903. https://doi.org/10.1242/jcs.03098
Hyldig K, Riis S, Pennisi CP, Zachar V, Fink T (2017) Implications of extracellular matrix production by adipose tissue-derived stem cells for development of wound healing therapies. Int J Mol Sci 18(6). https://doi.org/10.3390/ijms18061167
Hynes RO (2009) The extracellular matrix: not just pretty fibrils. Science 326(5957):1216–1219
Hynes RO, Naba A (2012) Overview of the matrisome—an inventory of extracellular matrix constituents and functions. Cold Spring Harb Perspect Biol 4(1):a004903. https://doi.org/10.1101/cshperspect.a004903
Iredale JP, Benyon RC, Pickering J, McCullen M, Northrop M, Pawley S, Hovell C, Arthur MJ (1998) Mechanisms of spontaneous resolution of rat liver fibrosis. Hepatic stellate cell apoptosis

and reduced hepatic expression of metalloproteinase inhibitors. J Clin Invest 102(3):538–549. https://doi.org/10.1172/jci1018

Issa R, Zhou X, Constandinou CM, Fallowfield J, Millward-Sadler H, Gaca MD, Sands E, Suliman I, Trim N, Knorr A, Arthur MJ, Benyon RC, Iredale JP (2004) Spontaneous recovery from micronodular cirrhosis: evidence for incomplete resolution associated with matrix cross-linking. Gastroenterology 126(7):1795–1808. https://doi.org/10.1053/j.gastro.2004.03.009

Ivaska J, Heino J (2011) Cooperation between integrins and growth factor receptors in signaling and endocytosis. Annu Rev Cell Dev Biol 27:291–320. https://doi.org/10.1146/annurev-cellbio-092910-154017

Iwamoto DV, Calderwood DA (2015) Regulation of integrin-mediated adhesions. Curr Opin Cell Biol 36:41–47. https://doi.org/10.1016/j.ceb.2015.06.009

Julich D, Cobb G, Melo AM, McMillen P, Lawton AK, Mochrie SG, Rhoades E, Holley SA (2015) Cross-scale integrin regulation organizes ECM and tissue topology. Dev Cell 34(1):33–44. https://doi.org/10.1016/j.devcel.2015.05.005

Kagan HM (2000) Intra- and extracellular enzymes of collagen biosynthesis as biological and chemical targets in the control of fibrosis. Acta Trop 77(1):147–152

Karki P, Birukova AA (2018) Substrate stiffness-dependent exacerbation of endothelial permeability and inflammation: mechanisms and potential implications in ALI and PH (2017 Grover conference series). Pulm Circ 8(2). https://doi.org/10.1177/2045894018773044

Karlmark KR, Weiskirchen R, Zimmermann HW, Gassler N, Ginhoux F, Weber C, Merad M, Luedde T, Trautwein C, Tacke F (2009) Hepatic recruitment of the inflammatory Gr1+ monocyte subset upon liver injury promotes hepatic fibrosis. Hepatology 50(1):261–274. https://doi.org/10.1002/hep.22950

Karsdal MA, Manon-Jensen T, Genovese F, Kristensen JH, Nielsen MJ, Sand JM, Hansen NU, Bay-Jensen AC, Bager CL, Krag A, Blanchard A, Krarup H, Leeming DJ, Schuppan D (2015) Novel insights into the function and dynamics of extracellular matrix in liver fibrosis. Am J Physiol Gastrointest Liver Physiol 308(10):G807–G830. https://doi.org/10.1152/ajpgi.00447.2014

Kirpich IA, Marsano LS, McClain CJ (2015) Gut-liver axis, nutrition, and non-alcoholic fatty liver disease. Clin Biochem 48(13–14):923–930. https://doi.org/10.1016/j.clinbiochem.2015.06.023

Kisseleva T, Brenner DA (2006) Hepatic stellate cells and the reversal of fibrosis. J Gastroenterol Hepatol 21(Suppl 3):S84–S87. https://doi.org/10.1111/j.1440-1746.2006.04584.x

Klaas M, Kangur T, Viil J, Maemets-Allas K, Minajeva A, Vadi K, Antsov M, Lapidus N, Jarvekulg M, Jaks V (2016) The alterations in the extracellular matrix composition guide the repair of damaged liver tissue. Sci Rep 6:27398. https://doi.org/10.1038/srep27398

Lee WY, Kubes P (2008) Leukocyte adhesion in the liver: distinct adhesion paradigm from other organs. J Hepatol 48(3):504–512. https://doi.org/10.1016/j.jhep.2007.12.005

Lee-Montiel FT, George SM, Gough AH, Sharma AD, Wu J, DeBiasio R, Vernetti LA, Taylor DL (2017) Control of oxygen tension recapitulates zone-specific functions in human liver microphysiology systems. Exp Biol Med (Maywood) 242(16):1617–1632. https://doi.org/10.1177/1535370217703978

Ley K, Laudanna C, Cybulsky MI, Nourshargh S (2007) Getting to the site of inflammation: the leukocyte adhesion cascade updated. Nat Rev Immunol 7(9):678–689. https://doi.org/10.1038/nri2156

Li B, Selmi C, Tang R, Gershwin ME, Ma X (2018) The microbiome and autoimmunity: a paradigm from the gut-liver axis. Cell Mol Immunol 15(6):595–609. https://doi.org/10.1038/cmi.2018.7

Libanio D, Marinho RT (2017) Impact of hepatitis C oral therapy in portal hypertension. World J Gastroenterol 23(26):4669–4674. https://doi.org/10.3748/wjg.v23.i26.4669

Lieber CS, Jones DP, Decarli LM (1965) Effects of prolonged ethanol intake: production of fatty liver despite adequate diets. J Clin Invest 44:1009–1021. https://doi.org/10.1172/JCI105200

Lipowsky HH (2018) Role of the glycocalyx as a barrier to leukocyte-endothelium adhesion. Adv Exp Med Biol 1097:51–68. https://doi.org/10.1007/978-3-319-96445-4_3

Liu SB, Ikenaga N, Peng ZW, Sverdlov DY, Greenstein A, Smith V, Schuppan D, Popov Y (2016) Lysyl oxidase activity contributes to collagen stabilization during liver fibrosis progression and limits spontaneous fibrosis reversal in mice. FASEB J 30(4):1599–1609. https://doi.org/10.1096/fj.14-268425

Lorenz L, Axnick J, Buschmann T, Henning C, Urner S, Fang S, Nurmi H, Eichhorst N, Holtmeier R, Bodis K, Hwang JH, Mussig K, Eberhard D, Stypmann J, Kuss O, Roden M, Alitalo K, Haussinger D, Lammert E (2018) Mechanosensing by beta1 integrin induces angiocrine signals for liver growth and survival. Nature 562(7725):128–132. https://doi.org/10.1038/s41586-018-0522-3

Luedde T, Kaplowitz N, Schwabe RF (2014) Cell death and cell death responses in liver disease: mechanisms and clinical relevance. Gastroenterology 147(4):765–783. https://doi.org/10.1053/j.gastro.2014.07.018

Lumelsky N, O'Hayre M, Chander P, Shum L, Somerman MJ (2018) Autotherapies: enhancing endogenous healing and regeneration. Trends Mol Med 24(11):919–930. https://doi.org/10.1016/j.molmed.2018.08.004

Mak KM, Png CY, Lee DJ (2016) Type V collagen in health, disease, and fibrosis. Anat Rec (Hoboken) 299(5):613–629. https://doi.org/10.1002/ar.23330

Mammoto A, Mammoto T, Kanapathipillai M, Wing Yung C, Jiang E, Jiang A, Lofgren K, Gee EP, Ingber DE (2013) Control of lung vascular permeability and endotoxin-induced pulmonary oedema by changes in extracellular matrix mechanics. Nat Commun 4:1759. https://doi.org/10.1038/ncomms2774

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

Martinez-Hernandez A, Amenta PS (1995) The extracellular matrix in hepatic regeneration. FASEB J 9(14):1401–1410. https://doi.org/10.1096/fasebj.9.14.7589981

Massey VL (2014) Extracellular matrix proteins and the liver-lung axis in disease. Electronic theses and dissertations. Paper 1752. https://doi.org/10.18297/etd/1752

Massey VL, Dolin CE, Poole LG, Hudson SV, Siow DL, Brock GN, Merchant ML, Wilkey DW, Arteel GE (2017) The hepatic "matrisome" responds dynamically to injury: characterization of transitional changes to the extracellular matrix in mice. Hepatology 65(3):969–982. https://doi.org/10.1002/hep.28918

McClelland R, Wauthier E, Uronis J, Reid L (2008) Gradients in the liver's extracellular matrix chemistry from periportal to pericentral zones: influence on human hepatic progenitors. Tissue Eng A 14(1):59–70. https://doi.org/10.1089/ten.a.2007.0058

McDonald B, Kubes P (2015) Interactions between CD44 and Hyaluronan in leukocyte trafficking. Front Immunol 6:68

Mehal WZ, Schuppan D (2015) Antifibrotic therapies in the liver. Semin Liver Dis 35(2):184–198. https://doi.org/10.1055/s-0035-1550055

Michalopoulos GK, DeFrances M (2005) Liver regeneration. Adv Biochem Eng Biotechnol 93:101–134

Monneau Y, Arenzana-Seisdedos F, Lortat-Jacob H (2016) The sweet spot: how GAGs help chemokines guide migrating cells. J Leukoc Biol 99(6):935–953. https://doi.org/10.1189/jlb.3MR0915-440R

Morrison ED, Kowdley KV (2000) Genetic liver disease in adults. Early recognition of the three most common causes. Postgrad Med 107(2):147–152. https://doi.org/10.3810/pgm.2000.02.872

Naba A, Clauser KR, Hoersch S, Liu H, Carr SA, Hynes RO (2012) The matrisome: in silico definition and in vivo characterization by proteomics of normal and tumor extracellular matrices. Mol Cell Proteomics 11(4). https://doi.org/10.1074/mcp.M111.014647

Naba A, Clauser KR, Ding H, Whittaker CA, Carr SA, Hynes RO (2016) The extracellular matrix: tools and insights for the "omics" era. Matrix Biol 49:10–24. https://doi.org/10.1016/j.matbio.2015.06.003

Nargis NN, Aldredge RC, Guy RD (2018) The influence of soluble fragments of extracellular matrix (ECM) on tumor growth and morphology. Math Biosci 296:1–16. https://doi.org/10.1016/j.mbs.2017.11.014

Oliveira THC, Marques PE, Proost P, Teixeira MMM (2018) Neutrophils: a cornerstone of liver ischemia and reperfusion injury. Lab Investig 98(1):51–62. https://doi.org/10.1038/labinvest.2017.90

Omenetti A, Porrello A, Jung Y, Yang L, Popov Y, Choi SS, Witek RP, Alpini G, Venter J, Vandongen HM, Syn WK, Baroni GS, Benedetti A, Schuppan D, Diehl AM (2008) Hedgehog signaling regulates epithelial-mesenchymal transition during biliary fibrosis in rodents and humans. J Clin Invest 118(10):3331–3342. https://doi.org/10.1172/JCI35875

Ozaki I, Hamajima H, Matsuhashi S, Mizuta T (2011) Regulation of TGF-beta1-induced pro-apoptotic signaling by growth factor receptors and extracellular matrix receptor integrins in the liver. Front Physiol 2:78. https://doi.org/10.3389/fphys.2011.00078

Patel VN, Pineda DL, Hoffman MP (2017) The function of heparan sulfate during branching morphogenesis. Matrix Biol 57–58:311–323. https://doi.org/10.1016/j.matbio.2016.09.004

Patouraux S, Rousseau D, Bonnafous S, Lebeaupin C, Luci C, Canivet CM, Schneck AS, Bertola A, Saint-Paul MC, Iannelli A, Gugenheim J, Anty R, Tran A, Bailly-Maitre B, Gual P (2017) CD44 is a key player in non-alcoholic steatohepatitis. J Hepatol 67(2):328–338. https://doi.org/10.1016/j.jhep.2017.03.003

Patsenker E, Stickel F (2011) Role of integrins in fibrosing liver diseases. Am J Physiol Gastrointest Liver Physiol 301(3):G425–G434. https://doi.org/10.1152/ajpgi.00050.2011

Pellicoro A, Aucott RL, Ramachandran P, Robson AJ, Fallowfield JA, Snowdon VK, Hartland SN, Vernon M, Duffield JS, Benyon RC, Forbes SJ, Iredale JP (2012) Elastin accumulation is regulated at the level of degradation by macrophage metalloelastase (MMP-12) during experimental liver fibrosis. Hepatology 55(6):1965–1975. https://doi.org/10.1002/hep.25567

Pellicoro A, Ramachandran P, Iredale JP, Fallowfield JA (2014) Liver fibrosis and repair: immune regulation of wound healing in a solid organ. Nat Rev Immunol 14(3):181–194. https://doi.org/10.1038/nri3623

Phillip JM, Aifuwa I, Walston J, Wirtz D (2015) The mechanobiology of aging. Annu Rev Biomed Eng 17:113–141. https://doi.org/10.1146/annurev-bioeng-071114-040829

Poole LG, Arteel GE (2016) Transitional remodeling of the hepatic extracellular matrix in alcohol-induced liver injury. Biomed Res Int 2016:3162670. https://doi.org/10.1155/2016/3162670

Poole LG, Dolin CE, Arteel GE (2017) Organ-organ crosstalk and alcoholic liver disease. Biomol Ther 7(3). https://doi.org/10.3390/biom7030062

Poynard T, McHutchison J, Manns M, Trepo C, Lindsay K, Goodman Z, Ling MH, Albrecht J (2002) Impact of pegylated interferon alfa-2b and ribavirin on liver fibrosis in patients with chronic hepatitis C. Gastroenterology 122(5):1303–1313. https://doi.org/10.1053/gast.2002.33023

Preziosi ME, Monga SP (2017) Update on the mechanisms of liver regeneration. Semin Liver Dis 37(2):141–151. https://doi.org/10.1055/s-0037-1601351

Pritchard MT, McCracken JM (2015) Identifying novel targets for treatment of liver fibrosis: what can we learn from injured tissues which heal without a scar? Curr Drug Targets 16 (12):1332–1346

Proudfoot AE, Handel TM, Johnson Z, Lau EK, LiWang P, Clark-Lewis I, Borlat F, Wells TN, Kosco-Vilbois MH (2003) Glycosaminoglycan binding and oligomerization are essential for the in vivo activity of certain chemokines. Proc Natl Acad Sci USA 100(4):1885–1890. https://doi.org/10.1073/pnas.0334864100

Ramachandran P, Pellicoro A, Vernon MA, Boulter L, Aucott RL, Ali A, Hartland SN, Snowdon VK, Cappon A, Gordon-Walker TT, Williams MJ, Dunbar DR, Manning JR, van Rooijen N, Fallowfield JA, Forbes SJ, Iredale JP (2012) Differential Ly-6C expression identifies the recruited macrophage phenotype, which orchestrates the regression of murine liver fibrosis. Proc Natl Acad Sci USA 109(46):E3186–E3195. https://doi.org/10.1073/pnas.1119964109

Ramaiah SK, Rittling S (2008) Pathophysiological role of osteopontin in hepatic inflammation, toxicity, and cancer. Toxicol Sci 103(1):4–13. https://doi.org/10.1093/toxsci/kfm246
Randles M, Lennon R (2015) Applying proteomics to investigate extracellular matrix in health and disease. Curr Top Membr 76:171–196. https://doi.org/10.1016/bs.ctm.2015.06.001
Rauniyar N, Yates JR (2014) Isobaric labeling-based relative quantification in shotgun proteomics. J Proteome Res 13(12):5293–5309. https://doi.org/10.1021/pr500880b
Reddig PJ, Juliano RL (2005) Clinging to life: cell to matrix adhesion and cell survival. Cancer Metastasis Rev 24(3):425–439. https://doi.org/10.1007/s10555-005-5134-3
Robertson H, Kirby JA, Yip WW, Jones DE, Burt AD (2007) Biliary epithelial-mesenchymal transition in posttransplantation recurrence of primary biliary cirrhosis. Hepatology 45 (4):977–981. https://doi.org/10.1002/hep.21624
Robinson MW, Harmon C, O'Farrelly C (2016) Liver immunology and its role in inflammation and homeostasis. Cell Mol Immunol 13(3):267–276. https://doi.org/10.1038/cmi.2016.3
Roderfeld M (2018) Matrix metalloproteinase functions in hepatic injury and fibrosis. Matrix Biol 68-69:452–462. https://doi.org/10.1016/j.matbio.2017.11.011
Rousselle P, Braye F, Dayan G (2018) Re-epithelialization of adult skin wounds: cellular mechanisms and therapeutic strategies. Adv Drug Deliv Rev. https://doi.org/10.1016/j.addr.2018.06.019
Rozario T, DeSimone DW (2010) The extracellular matrix in development and morphogenesis: a dynamic view. Dev Biol 341(1):126–140. https://doi.org/10.1016/j.ydbio.2009.10.026
Sacca SC, Gandolfi S, Bagnis A, Manni G, Damonte G, Traverso CE, Izzotti A (2016) From DNA damage to functional changes of the trabecular meshwork in aging and glaucoma. Ageing Res Rev 29:26–41. https://doi.org/10.1016/j.arr.2016.05.012
Sadeghi S, Vink RL (2015) Membrane sorting via the extracellular matrix. Biochim Biophys Acta 1848(2):527–531. https://doi.org/10.1016/j.bbamem.2014.10.035
Sand JM, Leeming DJ, Byrjalsen I, Bihlet AR, Lange P, Tal-Singer R, Miller BE, Karsdal MA, Vestbo J (2016) High levels of biomarkers of collagen remodeling are associated with increased mortality in COPD - results from the ECLIPSE study. Respir Res 17(1):125. https://doi.org/10.1186/s12931-016-0440-6
Santambrogio L, Rammensee HG (2019) Contribution of the plasma and lymph degradome and peptidome to the MHC ligandome. Immunogenetics 71(3):203–216. https://doi.org/10.1007/s00251-018-1093-z
Schnittert J, Bansal R, Storm G, Prakash J (2018) Integrins in wound healing, fibrosis and tumor stroma: high potential targets for therapeutics and drug delivery. Adv Drug Deliv Rev 129:37–53. https://doi.org/10.1016/j.addr.2018.01.020
Schuppan D, Ashfaq-Khan M, Yang AT, Kim YO (2018) Liver fibrosis: direct antifibrotic agents and targeted therapies. Matrix Biol 68-69:435–451. https://doi.org/10.1016/j.matbio.2018.04.006
Schuster S, Cabrera D, Arrese M, Feldstein AE (2018) Triggering and resolution of inflammation in NASH. Nat Rev Gastroenterol Hepatol 15(6):349–364. https://doi.org/10.1038/s41575-018-0009-6
Schwartz JM, Reinus JF (2012) Prevalence and natural history of alcoholic liver disease. Clin Liver Dis 16(4):659–666. https://doi.org/10.1016/j.cld.2012.08.001
Sessions AO, Engler AJ (2016) Mechanical regulation of cardiac aging in model systems. Circ Res 118(10):1553–1562. https://doi.org/10.1161/CIRCRESAHA.116.307472
Seth D, Haber PS, Syn WK, Diehl AM, Day CP (2011) Pathogenesis of alcohol-induced liver disease: classical concepts and recent advances. J Gastroenterol Hepatol 26(7):1089–1105. https://doi.org/10.1111/j.1440-1746.2011.06756.x
Seth D, Duly A, Kuo PC, McCaughan GW, Haber PS (2014) Osteopontin is an important mediator of alcoholic liver disease via hepatic stellate cell activation. World J Gastroenterol 20 (36):13088–13104. https://doi.org/10.3748/wjg.v20.i36.13088
Shao X, Taha IN, Clauser KR, Gao YT, Naba A (2020) MatrisomeDB: the ECM-protein knowledge database. Nucleic Acids Res 48(D1):D1136–D1144. https://doi.org/10.1093/nar/gkz849

Shimizu Y, Shaw S (1991) Lymphocyte interactions with extracellular matrix. FASEB J 5 (9):2292–2299. https://doi.org/10.1096/fasebj.5.9.1860621

Shirin H, Bruck R, Aeed H, Frenkel D, Kenet G, Zaidel L, Avni Y, Halpern Z, Hershkoviz R (1998) Pentoxifylline prevents concanavalin A-induced hepatitis by reducing tumor necrosis factor alpha levels and inhibiting adhesion of T lymphocytes to extracellular matrix. J Hepatol 29 (1):60–67. https://doi.org/10.1016/s0168-8278(98)80179-7

Simons K, Toomre D (2000) Lipid rafts and signal transduction. Nat Rev Mol Cell Biol 1(1):31–39. https://doi.org/10.1038/35036052

Singh S, Osna NA, Kharbanda KK (2017) Treatment options for alcoholic and non-alcoholic fatty liver disease: a review. World J Gastroenterol 23(36):6549–6570. https://doi.org/10.3748/wjg.v23.i36.6549

Song KS, Kim HS, Park KE, Kwon OH (1993) The fibrinogen degradation products (FgDP) levels in liver disease. Yonsei Med J 34(3):234–238. https://doi.org/10.3349/ymj.1993.34.3.234

Sorokin L (2010) The impact of the extracellular matrix on inflammation. Nat Rev Immunol 10 (10):712–723. https://doi.org/10.1038/nri2852

Srivastava A, Jong S, Gola A, Gailer R, Morgan S, Sennett K, Tanwar S, Pizzo E, O'Beirne J, Tsochatzis E, Parkes J, Rosenberg W (2019) Cost-comparison analysis of FIB-4, ELF and fibroscan in community pathways for non-alcoholic fatty liver disease. BMC Gastroenterol 19 (1):122. https://doi.org/10.1186/s12876-019-1039-4

Sullivan DC, Mirmalek-Sani SH, Deegan DB, Baptista PM, Aboushwareb T, Atala A, Yoo JJ (2012) Decellularization methods of porcine kidneys for whole organ engineering using a high-throughput system. Biomaterials 33(31):7756–7764. https://doi.org/10.1016/j.biomaterials.2012.07.023

Sun BK, Siprashvili Z, Khavari PA (2014) Advances in skin grafting and treatment of cutaneous wounds. Science 346(6212):941–945. https://doi.org/10.1126/science.1253836

Sweet PH, Khoo T, Nguyen S (2017) Nonalcoholic fatty liver disease. Prim Care 44(4):599–607. https://doi.org/10.1016/j.pop.2017.07.003

Tatsukawa H, Furutani Y, Hitomi K, Kojima S (2016) Transglutaminase 2 has opposing roles in the regulation of cellular functions as well as cell growth and death. Cell Death Dis 7(6):e2244. https://doi.org/10.1038/cddis.2016.150

Thiele GM, Duryee MJ, Freeman TL, Sorrell MF, Willis MS, Tuma DJ, Klassen LW (2005) Rat sinusoidal liver endothelial cells (SECs) produce pro-fibrotic factors in response to adducts formed from the metabolites of ethanol. Biochem Pharmacol 70(11):1593–1600. https://doi.org/10.1016/j.bcp.2005.08.014

Vempati P, Popel AS, Mac Gabhann F (2014) Extracellular regulation of VEGF: isoforms, proteolysis, and vascular patterning. Cytokine Growth Factor Rev 25(1):1–19. https://doi.org/10.1016/j.cytogfr.2013.11.002

Verstegen MMA, Willemse J, van den Hoek S, Kremers GJ, Luider TM, van Huizen NA, Willemssen F, Metselaar HJ, JNM IJ, van der Laan LJW, de Jonge J (2017) Decellularization of whole human liver grafts using controlled perfusion for transplantable organ bioscaffolds. Stem Cells Dev 26(18):1304–1315. https://doi.org/10.1089/scd.2017.0095

Vispo E, Barreiro P, Del Valle J, Maida I, de Ledinghen V, Quereda C, Moreno A, Macias J, Castera L, Pineda JA, Soriano V (2009) Overestimation of liver fibrosis staging using transient elastography in patients with chronic hepatitis C and significant liver inflammation. Antivir Ther 14(2):187–193

Wang S, Voisin MB, Larbi KY, Dangerfield J, Scheiermann C, Tran M, Maxwell PH, Sorokin L, Nourshargh S (2006) Venular basement membranes contain specific matrix protein low expression regions that act as exit points for emigrating neutrophils. J Exp Med 203(6):1519–1532. https://doi.org/10.1084/jem.20051210

Wasmuth HE, Tacke F, Trautwein C (2010) Chemokines in liver inflammation and fibrosis. Semin Liver Dis 30(3):215–225. https://doi.org/10.1055/s-0030-1255351

Widgerow AD (2012) Cellular resolution of inflammation—catabasis. Wound Repair Regen 20 (1):2–7. https://doi.org/10.1111/j.1524-475X.2011.00754.x

Willemse J, Verstegen MMA, Vermeulen A, Schurink IJ, Roest HP, van der Laan LJW, de Jonge J (2020) Fast, robust and effective decellularization of whole human livers using mild detergents and pressure controlled perfusion. Mater Sci Eng C Mater Biol Appl 108:110200. https://doi.org/10.1016/j.msec.2019.110200

Wisniewski JR, Zougman A, Mann M (2009) Combination of FASP and StageTip-based fractionation allows in-depth analysis of the hippocampal membrane proteome. J Proteome Res 8 (12):5674–5678. https://doi.org/10.1021/pr900748n

Wong J, Johnston B, Lee SS, Bullard DC, Smith CW, Beaudet AL, Kubes P (1997) A minimal role for selectins in the recruitment of leukocytes into the inflamed liver microvasculature. J Clin Invest 99(11):2782–2790. https://doi.org/10.1172/JCI119468

Woodfin A, Voisin MB, Nourshargh S (2010) Recent developments and complexities in neutrophil transmigration. Curr Opin Hematol 17(1):9–17. https://doi.org/10.1097/MOH.0b013e3283333930

Wree A, Marra F (2016) The inflammasome in liver disease. J Hepatol 65(5):1055–1056. https://doi.org/10.1016/j.jhep.2016.07.002

Wu D, Birukov K (2019) Endothelial cell mechano-metabolomic coupling to disease states in the lung microvasculature. Front Bioeng Biotechnol 7:172. https://doi.org/10.3389/fbioe.2019.00172

Zang S, Wang L, Ma X, Zhu G, Zhuang Z, Xun Y, Zhao F, Yang W, Liu J, Luo Y, Liu Y, Ye D, Shi J (2015) Neutrophils play a crucial role in the early stage of nonalcoholic steatohepatitis via neutrophil Elastase in mice. Cell Biochem Biophys 73(2):479–487. https://doi.org/10.1007/s12013-015-0682-9

Zeisberg M, Yang C, Martino M, Duncan MB, Rieder F, Tanjore H, Kalluri R (2007) Fibroblasts derive from hepatocytes in liver fibrosis via epithelial to mesenchymal transition. J Biol Chem 282(32):23337–23347. https://doi.org/10.1074/jbc.M700194200

Zhou HF, Chan HW, Wickline SA, Lanza GM, Pham CT (2009) Alphavbeta3-targeted nanotherapy suppresses inflammatory arthritis in mice. FASEB J 23(9):2978–2985. https://doi.org/10.1096/fj.09-129874

Chapter 4
Characterization of Proteoglycanomes by Mass Spectrometry

Christopher D. Koch and Suneel S. Apte

Abstract As composites of a core protein and several chemically distinct types of glycosaminoglycan (GAG) chains, proteoglycans are diverse molecules that occupy a unique niche in biology. They have varied and essential roles as structural and regulatory molecules in numerous physiological processes and disease pathology. In regard to cellular context, some link the interior of the cell to the extracellular matrix (ECM) as transmembrane or membrane-anchored molecules with a major role in cell adhesion and signal transduction. Others reside in pericellular matrix, where they influence crucial aspects of cell behavior, and several reside in interstitial ECM as components of structural macromolecular networks. Because of their unique composition, they can be challenging to identify and characterize using conventional biochemical or antibody-based methods. In contrast, the GAG component, despite its immense chemical diversity, typically carries a strong net negative charge which can be exploited to advantage for affinity-isolation and enrichment of proteoglycans from any biological system in a core protein-, GAG-, tissue-, and species-agnostic manner by anion exchange chromatography. This method, when coupled with high resolution liquid-chromatography tandem mass spectrometry (LC-MS/MS) can be used to define the proteoglycanome of any cell type, tissue or organism. A proteoglycanomics strategy can be further refined by inclusion of additional orthogonal affinity steps or fractionation for greater specificity and to deliver proteoglycans with distinct specified characteristics. Moreover, elimination of the GAG chain chemically and/or obliteration of the core protein enables glycomics characterization of GAG structure. Enzymatic digestion of GAGs on tryptic peptides allows mapping of glycopeptides, which has been used for identification of novel proteoglycans and

C. D. Koch
Department of Laboratory Medicine, Yale-New Haven Health, New Haven, CT, USA

Department of Biomedical Engineering, Cleveland Clinic Lerner Research Institute, Cleveland, OH, USA
e-mail: c.koch@yale.edu

S. S. Apte (✉)
Department of Biomedical Engineering, Cleveland Clinic Lerner Research Institute, Cleveland, OH, USA
e-mail: aptes@ccf.org

S. Ricard-Blum (ed.), *Extracellular Matrix Omics*, Biology of Extracellular Matrix 7,
https://doi.org/10.1007/978-3-030-58330-9_4

to precisely define sites of GAG attachment. Recent application of proteoglycanomics to human aorta and human aortic aneurysms demonstrated its potential to identify tissue and disease proteoglycanomes and the detailed method that was used is provided here for application to other tissues or biological systems.

4.1 Introduction

Proteoglycans (PGs) are composite molecules in whom glycosaminoglycan (GAG) chains are covalently attached to a polypeptide backbone referred to as a proteoglycan core protein (Iozzo and Schaefer 2015). The attachment typically occurs to Ser residues adjacent to a Gly residue and usually within an acidic sequence context, although keratan sulfate attachment can occur not only to Ser, but also to Thr and Asn through distinct linkages (Brinkmann et al. 1997; Funderburgh 2002). Proteoglycans are integral components of the extracellular matrix (ECM) of most tissues, and some, such as syndecans and glypicans, are specialized and important transmembrane and membrane-anchored cell-surface components, respectively (Filmus and Capurro 2014; Mitsou et al. 2017; Gondelaud and Ricard-Blum 2019). Hyaluronan-binding aggregating chondroitin sulfate proteoglycans (CSPGs) such as aggrecan are quantitatively abundant in structural tissues such as cartilage and intervertebral disc, where they provide unique biophysical properties (Hascall and Heinegard 1974; Heinegard and Saxne 2011). Specifically, the ability of the aggregates to hold large amounts of water and thus exert an internal tissue swelling pressure, as well as electrostatic charge repulsion between the aggregates makes them indispensable for compression load-bearing in these tissues (Buschmann and Grodzinsky 1995). Aggregating proteoglycans in pericellular ECM of mesenchymal cells have the potential to control focal adhesion formation, cell shape and genetic programs (Mead et al. 2018). In the brain, a diverse group of aggregating PGs (aggrecan, versican, brevican, neurocan) (Zimmermann and Dours-Zimmermann 2008) are prominent components of the limited amount of ECM that is present, and their swelling pressure may ensure mechanical buffering within the cranium; furthermore, perineuronal nets of a subset of neurons have PGs as a major component, where they are thought to insulate the soma (cell body) of the neuron against rewiring of established neuronal circuits (Fawcett et al. 2019). Aggrecan, for example, is critical in this regard (Rowlands et al. 2018). In the view of the Nobel laureate Roger Tsien, holes in perineuronal nets, which are formed on conclusion of the juvenile critical period and mark the closure of neuronal plasticity, could be the seat of long-term memory (Tsien 2013). Major roles in developmental and cancer cell signaling by heparan sulfate proteoglycans (HSPGs) are attributed to the binding of growth factors and morphogens to the highly sulfated GAG chains (Bandari et al. 2015; Ortmann et al. 2015; Sarrazin et al. 2011; van Wijk and van Kuppevelt 2014; Yu and Woessner Jr. 2000). Similar roles are reprised during inflammation with respect to cytokines (Gondelaud and Ricard-Blum 2019; Bartlett et al. 2007). In addition, some PGs or their fragments can act as damage-associated molecular

patterns to stoke inflammation (Nastase et al. 2018). In many tissues such as tendons, skeletal muscle, lungs, cornea and blood vessels, proteoglycans are quantitatively minor components but serve crucial roles such as compression load-bearing, regulation of collagen fibril assembly and sequestration of growth factors (Ezura et al. 2000; Robinson et al. 2017; Zhang et al. 2006). One reason why proteoglycan steady state levels may be low in many cells and matrices is that they are among the most dynamic components of ECM and thus turned over quite rapidly, particularly in cell-proximate ECM.

Proteoglycan core proteins can be very large, e.g., perlecan, aggrecan and versican, or quite small, e.g., the leucine-rich proteoglycans. PG sub-groups are usually defined by a combination of size and other unique properties such as the chemical nature of their GAG chains, i.e., whether they are chemically defined chondroitin sulfate (CS), heparan sulfate (HS) or keratan sulfate (KS). Dermatan sulfate (DS) sometimes termed CS-B, is chemically similar to CS but contains iduronate (Thelin et al. 2013). This classification of PGs is imperfect, since different types of GAG can be present on the same core protein, e.g., aggrecan typically contains KS chains, yet the CS chains dominate and it is usually considered a CSPG. Moreover, some proteoglycan core proteins are not constitutively modified, and these could be regarded as "part-time" proteoglycans. However, for the purpose of this chapter, any molecule that carries a covalently attached GAG chain, if only in a small proportion of the core protein, is operationally defined as a proteoglycan.

Most GAGs are sulfated and thus share the useful property of having a net negative charge, enabling their isolation using anion exchange chromatography (AEC). This characteristic of PGs is also the basis of their detection in tissue sections using basic dyes such as alcian blue and safranin O or toluidine blue metachromasia of highly negatively charged GAGs such as heparin (Scott 1985). Their staining intensity is also a useful guide, albeit neither absolute nor specific, to GAG abundance. Tissue identification of individual GAGs and PG gene products is typically done using specific antibodies to the GAG chain or core protein. The former has the advantage over anti-core protein antibodies of being species-agnostic. Immunostaining and western blot can be used to detect the core proteins but relies on antibodies with high specificity. Additionally, the GAG-dense environment in which a core protein epitope may reside may mask antibody reactivity in tissue sections, requiring an intimate understanding of the molecules in order to properly design epitope retrieval strategies. However, the use of core protein and GAG antibodies as a targeted approach to examine the entire proteoglycan landscape of a tissue is neither practical nor efficient.

An untargeted approach, such as shotgun mass spectrometry would allow simultaneous detection of many proteoglycans within a complex cell or tissue extract while circumventing the need for quality antibodies and epitope retrieval protocols. Analysis of proteins by mass spectrometry is, however, limited by the caveat that high abundance proteins are preferentially detected during data-dependent analysis of mass spectra, and low abundance or highly modified components like proteoglycans may go undetected. A solution is to divide the sample into multiple fractions, using one or preferably, two orthogonal fractionation methods (Ly and Wasinger

2011). However, analysis of multiple fractions increases the instrument run time and requires analysis of data from multiple MS runs. An alternative to routine fractionation for reduction of sample complexity is enriching for the desired components, or excluding undesired ones (Ly and Wasinger 2011). We capitalized on the net negative charge of the GAGs to isolate and identify PG core proteins by mass spectrometry from the aorta (Cikach et al. 2018), the largest artery in the body, which contains a large repertoire of ECM molecules sandwiched in the space between elastic lamellae and arrays of smooth muscle cells.

Our approach utilized a well-characterized technique for proteoglycan enrichment from aortic tissue, i.e., isolation by AEC prior to analysis by LC-MS/MS (Cikach et al. 2018). AEC elution conditions can be adjusted to maximize proteoglycan yields and successful elution of proteoglycans can be evaluated using a variety of biochemical techniques such as fluorophore-assisted carbohydrate electrophoresis (providing precise delineation of the GAG type), safranin O staining (non-specific, but provides quantifiable staining intensity and indicates abundance of the GAGs in the eluted fractions) and western blot with a specific core protein or GAG antibody (Cikach et al. 2018). Combinations of these orthogonal methods can be used to determine exactly which PGs and GAGs were enriched, and LC-MS/MS can be subsequently used for unbiased identification of PGs extracted from essentially any tissue. Here, we describe the approach that was used to determine the proteoglycanome of the human aorta (Cikach et al. 2018). The proteoglycan isolation and quantitation methods are similar to those previously described by Carrino and colleagues (Carrino et al. 1991, 1994).

4.2 Extraction of Proteoglycans from Tissue

1. Snap freeze tissue in liquid nitrogen immediately after collection and store at −80 °C until use.
2. Finely dice tissue with a scalpel or surgical scissors in a clean petridish on ice. Weigh diced tissue and add 1 mL ice-cold proteoglycan extraction buffer per 100 mg tissue. The extraction buffer is: 4 M guanidine hydrochloride, 2% 3-[(3-cholamidopropyl) dimethylammonio]-1-propanesulfonate [CHAPS], 50 mM sodium acetate, adjusted to pH 6.0 with HCl or NaOH. 1 tablet of complete Mini EDTA-free Protease Inhibitor (Roche) is added per 6–10 mL.
3. Homogenize tissue in the extraction buffer with a mechanical homogenizer such as an Ultra-Turrax T2 or T10 homogenizer (IKA Works Inc.), taking care to keep the sample on ice as homogenization will quickly warm the sample.
4. Rotate homogenized tissue end-over-end at 4 °C for a minimum of 16 h.
5. Clarify the homogenate by centrifugation for 15 min at 20,000×*g*.
6. Retain the supernatant as it contains both cellular and many extracellular matrix proteins including proteoglycans. Highly crosslinked proteins, including many collagens and elastic fibers, will remain in the sample pellet unless extracted by additional steps.

The extraction process would be similar for a cell culture monolayer, with the difference that after the medium is aspirated, the cells should be washed several times with serum-free medium prior to addition of the extraction buffer. This minimizes ion suppression from abundant serum proteins such albumin during LC-MS/MS.

4.3 Proteoglycan Isolation by Anion Exchange Chromatography

1. Buffer exchange: While guanidine hydrochloride allows for efficient extraction of proteoglycans and proteins from tissue, it is not compatible with AEC due to the presence of chloride as a counterion. Therefore, buffer exchange to another chaotropic agent such as urea must precede the AEC step. This can be accomplished by a number of methods including dialysis and centrifugal filtration; we use gel filtration, which is an efficient, rapid and economical method using columns that are inexpensive and easy to prepare.

   a. Prepare columns by cutting the tops off 5 mL, 10 mL or 25 mL plastic serologic pipettes and lightly pack the tips with glass wool. Connect a stopcock to the bottom of the pipette for volume control, such as a length of surgical tubing with removable clamp. Swell Sephadex G-50 fine resin in G50 buffer at a ratio of 20–25 mL/g of resin for at least 24 h. G50 buffer is 8 M urea, 0.5% CHAPS, 50 mM sodium acetate, 150 mM NaCl adjusted to pH 7.0 with HCl or NaOH. Heating on a hot plate-stirrer aids urea dissolution.
   b. Load swollen Sephadex G-50 fine resin into the column (Table 4.1) and pack by gravity flow, adding G-50 buffer as needed. Once the column is packed, allow the G50 buffer to almost completely enter the column, i.e., leaving a meniscus to prevent the resin from drying.
   c. Taking care not to disturb the resin bed, add the appropriate sample volume (see Table 4.1) to the column and allow it to enter the resin by gravity. Close the stopcock when the last of the sample has just entered the resin.

Table 4.1 Volumes for buffer exchange using Sephadex G50 fine. The sample volume dictates the size of column(s) required. Buffer exchange can be accomplished with larger sample volumes through the use of sample aliquots and multiple columns. The pre-V_0 volume will not contain any proteins and is obtained immediately upon the sample fully entering the resin. Proteins will elute immediately following the pre-V_0 volume and are expected to be fully contained in the corresponding V_0 volume, although volumes for specific applications may vary and should be determined experimentally

Pipette/column size (mL)	Sephadex G50 fine (mL)	Sample (mL)	Pre-V_0 (mL)	V_0 (mL)
5	4	1	0.25	1.5
10	8	2	0.5	3
25	24	6	1.5	9

d. Carefully add G50 buffer to the column and resume gravity flow. Immediately start collecting the eluate as pre-V_0. This eluate will be G50 buffer without proteins and can be discarded. The volume is specifically adjusted for the column sizes and sample volumes referenced in Table 4.1.
e. Starting immediately after the pre-V_0 volume, collect the appropriate V_0 volume. This fraction will deliver the proteoglycans and proteins in the AEC-compatible G50 buffer.
f. Columns can be reused for subsequent samples after washing with 2–4 column volumes of G50 buffer; however single use may be more economical given the volume of G50 buffer that may be required for this.

2. Anion Exchange Chromatography

a. Swell diethylaminoethyl (DEAE)-Sephacel with G50 buffer. Pack 4 mL swollen DEAE resin by gravity flow of G50 buffer into a glass column fitted with a stopcock as described above. Once the column is packed, allow the G50 buffer to almost completely enter the column taking care not to let the resin dry. Add the sample to the column without disturbing the resin bed and allow it to enter the resin by gravity flow. Collect the flow-through and retain for future analysis or discard. Wash the column with five column volumes (20 mL) of wash buffer (8 M urea, 0.5% CHAPS, 50 mM sodium acetate, 250 mM NaCl, pH 7.0). This fraction will contain weakly anionic proteins. It can be collected and retained for future analysis but is usually discarded.
b. Add elution buffer (same as wash buffer above, but with 0.5–1 M NaCl) to the column and collect 2 mL fractions. Most proteoglycans will elute with 0.5 M NaCl (Fig. 4.1), however proteoglycans containing high anion charge densities such as versican and aggrecan may elute best at higher concentrations (up to 1 M) of NaCl. In our experience, 1 M NaCl is sufficient to elute all proteoglycans with the highest concentrations eluting in fractions 1–3 (6 mL total eluate volume).

3. Quantitation of isolated proteoglycans. The isolated fractions first undergo buffer exchange by dialysis to 20 mM HEPES, 150 mM NaCl, pH 7.2, following which a safranin O staining and spectrophotometry method (Carrino et al. 1991) is applied to quantify proteoglycan content of the fractions. For total protein, sample absorbance at 280 nM is used. A representative safranin O-based quantitation profile of fractions is shown in Fig. 4.2.

a. Prepare 0.45 μm nitrocellulose by soaking in water for at least 1 min and mount it into a dot blot apparatus.
b. Pipette 25 μL (1 part) of each DEAE eluate fraction (isolated proteoglycans) into 250 μL (10 parts) 0.02% safranin O into each well. Mix briefly by pipetting and allow the mixture to stand for 1 min. Apply a vacuum to the dot blot apparatus to collect the precipitate onto the nitrocellulose.
c. Wash each well three times with 200 μL water, drawing water through the nitrocellulose under vacuum. Wells with significant precipitation (high proteoglycan/glycosaminoglycan content) may require higher vacuum pressure

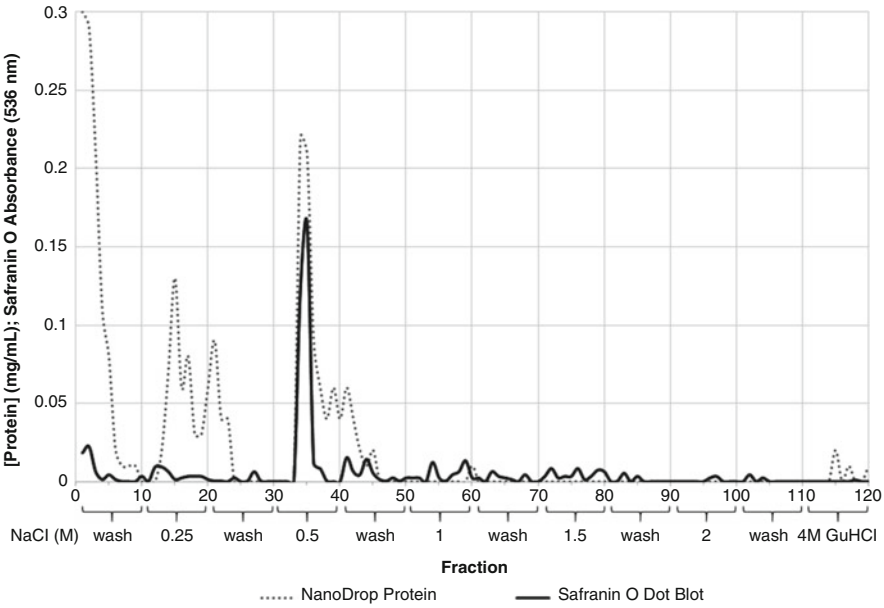

Fig. 4.1 AEC elution profile of aortic GAGs. Aortic proteoglycans were eluted from a 4 mL DEAE-Sephacel column with G50 buffer containing increasing concentrations of NaCl. Ten 1 mL fractions were collected for each elution condition. Each elution was preceded by a 10 mL wash with G50 buffer containing 150 mM NaCl. A final elution was performed with 4 M GuHCl to ensure no proteoglycans remained on the column after the 2 M NaCl elution. Total protein and relative GAG concentration were determined for each fraction using a NanoDrop (absorbance at 280 nm) and safranin O dot blot, respectively. Most proteoglycans eluted with 500 mM NaCl, however the large proteoglycans versican and aggrecan may require 1 M NaCl for complete elution

and a longer time to empty fully. To prevent bleeding of precipitate into adjacent wells, dry the nitrocellulose in the dot blot apparatus for a minimum of 1 h.

d. Remove the nitrocellulose from the dot blot apparatus and allow it to dry completely at room temperature. Cut or punch out the precipitate spots from the membrane and place each in a 1.5 mL microcentrifuge tube. Add 1 mL 10% cetylpyridinium chloride (CPC) to each tube and vortex vigorously. Incubate the tubes at 37 °C for 10 min, vortex vigorously, and incubate at 37 °C for an additional 10 min. Vortex vigorously after incubation and transfer a portion of the CPC solution from each tube into a 96 well plate or cuvette for absorbance measurement at 536 nm.

   i. Fluorophore-assisted carbohydrate analysis, if available, is a specialized technique that can be used to identify and quantify hyaluronan and GAGs (Calabro et al. 2001; Midura et al. 2018).
   ii. Western blot/ELISA using antibodies against specific GAGs and core proteins, or the GAG stubs left behind after enzymatic release can be

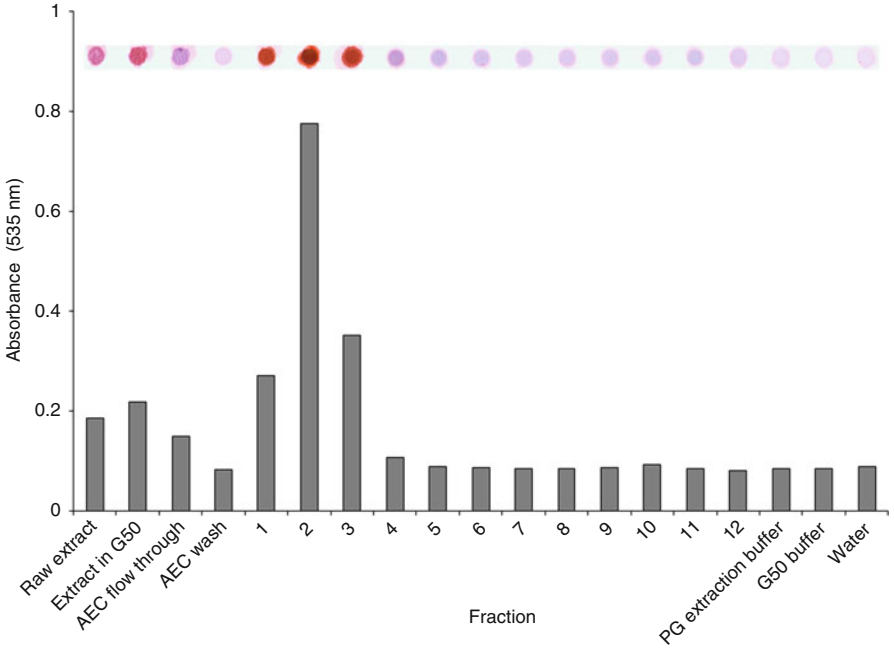

Fig. 4.2 Quantitation of aortic glycosaminoglycans isolated using AEC. GAG concentration, as a surrogate marker for proteoglycans, was determined for each sample using the safranin O dot blot assay. 2 mL fractions were collected from a column containing 4 mL DEAE-Sephacel. Representative data from a single sample is shown. "Raw extract" refers to the initial sample in proteoglycan extraction buffer, and "Extract in G50" refers to the same sample after buffer exchange

used for identifying specific components. Enzymatic removal of GAGs from proteoglycans with the appropriate lyase is typically required to ensure full mobility in polyacrylamide gels and may be essential for proper epitope recognition by some antibodies.

4.4 Analysis of Isolated Proteoglycans by Mass Spectrometry

1. Sample preparation for mass spectrometry

   a. Lyophilize 20 μg total protein in a SpeedVac evaporator.
   b. Reconstitute the dried protein in 50 μL 6 M urea, 100 mM Tris, pH 7.0.
   c. Reduce cysteine bonds by adding 2.5 μL 200 mM dithiothreitol (prepared fresh) and incubate at room temperature for 15 min.
   d. Alkylate the proteins by adding 10 μL 200 mM iodoacetamide (prepared fresh) and incubate at room temperature for 20 min, protecting from light.

e. Quench the excess iodoacetamide by adding 10 μL 200 mM dithriothreitol and incubate at room temperature for 15 min.
f. Reduce the urea concentration to approximately 1.2 M by diluting the sample with 160 μL water.
g. Adjust the pH to >8.0 with 100 mM ammonium bicarbonate. This will likely take 20–30 μL of ammonium bicarbonate; verify the correct pH by blotting a 1–2 μL sample on litmus paper.
h. Add trypsin at an enzyme:protein ratio of 1:20. Incubate at room temperature for 24 h or at 37 °C for 8–16 h.
i. Desalt the trypsinized peptides using a C18 column such as a Pierce C18 spin column (ThermoFisher Scientific) following the manufacturer's instructions.
j. Fully lyophilize the C18 eluate in a SpeedVac evaporator.
k. Reconstitute the peptides in 30 μL 1% acetic acid.
Note: This protocol leads to injection of approximately 1 μg of total peptides in 5 μL on the liquid chromatography column. The starting amount of total protein (step a) and the final reconstitution volume (step k) can be adjusted to increase or decrease the final injection concentration.

2. Mass spectrometry

Tryptic peptides can be identified using a number of modern mass spectrometry instruments. We used an LTQ-Orbitrap Elite hybrid mass spectrometer (Thermo Fisher Scientific) for its high resolution and sensitivity.

a. Separate peptides with an in-line liquid chromatography system, e.g., Dionex Ultimate 3000 nanoflow ultrahigh pressure liquid chromatography (UHPLC) system using a 75 μm × 15 cm, 3 μm particle size, Acclaim PepMap 100 C18 column (Thermo Fisher Scientific) at a flow rate of 0.3 μL/min.
b. Elute peptides over 2 h using buffers A (0.1% formic acid in water) and B (0.1% formic acid in acetonitrile) with the following LC conditions:

Time (min)	Buffer B (%)
0–5	2
5–110	Linear gradient: 2–40
110–115	Linear gradient: 40–80
115–120	80

c. The UHPLC column is coupled to a nanospray source through a PicoTip emitter (FS360-20-15-N-20-C15, New Objective).
d. Collect spectra using a full-ion scan at a resolution of 60,000 over the mass/charge range 300–2000. MS^2 scans using collision-induced dissociation (CID) can be performed on the 20 most abundant precursor ions from MS^1 scans using the data-dependent mode with dynamic exclusion.

3. Bioinformatics—The precise bioinformatics approach used will vary according to the facility or user preference. The following approach was used with data from the LTQ-Orbitrap Elite hybrid mass spectrometer

   a. MS^2 spectra were matched to the UniProtKB/Swiss-Prot human database using ProteomeDiscoverer (Thermo Scientific).
   b. The percolator function was utilized to select only matches with a Q value < 0.01 (<1% false discovery rate [FDR]).
   c. The mass tolerance was 10 ppm for precursor ions and 0.8 Da for MS^2.
   d. Only fully tryptic peptides were considered with a maximum of three missed tryptic cleavage sites.
   e. Carbamidomethylation of cysteine was set as a fixed modification and oxidation of methionine was set as a variable modification.

      Notes:

   i. Additional modifications (i.e. phosphorylation) can be included, which may increase peptide identification and improve core protein coverage.
   ii. Tryptic peptides containing GAG attachment sites will not be recognized unless the GAGs are enzymatically removed prior to analysis and the appropriate modifications are sought during spectral analysis (Noborn et al. 2015, 2016).

This detailed method provides a workhorse approach for isolating proteoglycans and can serve as the foundation for a proteomics study of core proteins, glycopeptide analysis or glycomics analysis. In all these applications, it ensures reduction of sample complexity and brings the PGs to the forefront. Using this approach to define the proteoglycanome of the ascending thoracic aorta, we identified 20 distinct proteoglycan core proteins (Cikach et al. 2018).

Pursuing a similar rationale, Talusan and colleagues used ion exchange chromatography isolation of proteoglycans coupled to proteomics analysis by gel-LC-MS/MS, in which the fractions obtained by urea extraction and High Q support strong AEC were electrophoresed on gels prior to tryptic digestion of the gel slices and mass spectrometry (Talusan et al. 2005). They found a strong correlation between proteoglycan abundance and species in an atherosclerosis-prone artery (internal carotid) and an atherosclerosis-resistant artery (internal thoracic) (Talusan et al. 2005). Vijayagopal et al. (1996) isolated LDL binding proteoglycans from atherosclerotic human arteries using a combination of orthogonal chromatographies. First, dissociative extraction and ion-exchange chromatography, similar to that described above, were used to isolate proteoglycans. Next proteoglycans were sub-fractionated on an LDL affinity column and proteoglycan fractions with high-affinity binding to LDL were analyzed for GAG species. The extracts from atheromatous plaques contained a high proportion of chondroitin and heparan sulfate proteoglycans, whereas normal aorta contained more diverse GAG species. Although core proteins were not evaluated, the data suggested enrichment of a small subset of proteoglycan species in atheromas, among which the CSPGs versican and biglycan have been shown to both bind LDL and be present in plaques (Didangelos et al. 2012; Wight

and Merrilees 2004). The CSPG aggrecan, typically a component of cartilage, was also reported in atheromatous plaques and stented coronary arteries (Talusan et al. 2005; Suna et al. 2018).

Mass spectrometry (MS) is used extensively for analysis of glycans, including GAGs. These applications digest the core protein to completion to preserve the GAGs and rely heavily on sophisticated bioinformatics approaches. This is because MS identifies molecules by mass and cannot readily distinguish between isomeric structures, e.g., glucose, mannose or galactose, although fragmentation of the glycan by different methods, coupled with other analytic techniques can help resolve its structure (Rojas-Macias et al. 2019). Profiling of glycans by MS requires dedicated isolation, derivatization and characterization techniques as well as high-resolution MS instruments (Rojas-Macias et al. 2019).

Whereas glycomics methods obliterate the core protein structure, glycoproteomics methods preserve the glycan-peptide linkage and can identify peptides to which glycans are attached, but these methods have limited ability to provide structural detail of the GAG. As an example of a glycoproteomics strategy, Noborn and colleagues have developed approaches to characterize HS and CS linkage regions, attachment sites, and identify novel proteoglycans (Noborn et al. 2015, 2016, 2018; Gomez Toledo et al. 2015). As in the method we describe here, their approach first requires isolation and enrichment of glycopeptides after trypsin digestion, using strong AEC. For identification of CS-attachment sites, for example, CS chains were depolymerized with chondroitinase ABC, leaving a residual hexameric structure composed of the linkage region and a GlcA-GalNAc disaccharide dehydrated on the terminal GlcA residue attached to the tryptic peptide. LC-MS/MS was then used to define the mass of the peptide with the residual hexasaccharide (Noborn et al. 2015).

4.5 Conclusions and Future Perspectives

In summary, we have briefly reviewed the immense biological significance of proteoglycans that justifies the need for specialized -omics approaches to obtain complete characterization, but it is obvious that this is a complex endeavor requiring multidisciplinary combinations of specialized techniques, instruments and skills. Regardless of whether the final goal is a proteoglycanome, glycoproteome or glycome, the strategy will employ some measure of enrichment of proteoglycans as an initial step. One can thus imagine several potential applications of the basic approaches we have outlined here. An important application is to identify the complete repertoire of "part-time" proteoglycans, since GAG-attachment contributes substantially to the functional diversity of the genome. Akin to our delineation of the aortic proteoglycanome, it is possible to use the approach to identify the proteoglycanome of any tissue or organ in development, health and disease. For example, taken together with isolation of specific organ structures such as renal glomeruli, or neural ganglia, proteoglycanomes of specific structures in organs can

also be achieved. The emerging explosion of single cell RNA sequence data for many tissues and organs, both adult and at many stages of embryonic development, will further allow a very refined determination of the proteoglycan repertoire. Another application that has been explored to only a limited extent is to define how proteoglycanomes differ in various diseases. This would be most informative if a quantitative analysis could be applied. One such approach would be to digest extracts of normal and disease tissue with trypsin, label the tryptic peptides with isobaric tags, and then combine the samples for glycopeptide enrichment steps prior to LC-MS/MS. In this way, variables introduced by preparation of each sample separately are eliminated, and moreover, the combined enriched glycopeptide pool can be analyzed in the same mass spectrometry run, avoiding run-to-run variation as well. The improved sensitivity of modern instruments and informatics tools that can combine glycopeptide analysis in multi-omics strategies, e.g., with single cell RNA sequencing, will likely make this approach important in understanding many diseases in the future.

Acknowledgments The authors express their gratitude to David A. Carrino, PhD for providing instruction in the proteoglycan isolation and quantitation methods used here and Belinda Willard PhD and Ling Li PhD of the Lerner Research Institute Proteomics and Metabolomics core for guidance in mass spectrometry. This work was supported by the Allen Distinguished Investigator Program, through support made by The Paul G. Allen Frontiers Group and the American Heart Association (to S.S.A).

References

Bandari S et al (2015) Sweet on Hedgehogs: regulatory roles of heparan sulfate proteoglycans in Hedgehog-dependent cell proliferation and differentiation. Curr Protein Pept Sci 16(1):66–76

Bartlett AH, Hayashida K, Park PW (2007) Molecular and cellular mechanisms of syndecans in tissue injury and inflammation. Mol Cells 24(2):153–166

Brinkmann T, Weilke C, Kleesiek K (1997) Recognition of acceptor proteins by UDP-D-xylose proteoglycan core protein beta-D-xylosyltransferase. J Biol Chem 272(17):11171–11175

Buschmann MD, Grodzinsky AJ (1995) A molecular model of proteoglycan-associated electrostatic forces in cartilage mechanics. J Biomech Eng 117(2):179–192

Calabro A et al (2001) Fluorophore-assisted carbohydrate electrophoresis (FACE) of glycosaminoglycans. Osteoarthritis Cartilage 9(Suppl A):S16–S22

Carrino DA, Arias JL, Caplan AI (1991) A spectrophotometric modification of a sensitive densitometric Safranin O assay for glycosaminoglycans. Biochem Int 24(3):485–495

Carrino DA et al (1994) Identity of the core proteins of the large chondroitin sulphate proteoglycans synthesized by skeletal muscle and prechondrogenic mesenchyme. Biochem J 298(Pt 1):51–60

Cikach FS et al (2018) Massive aggrecan and versican accumulation in thoracic aortic aneurysm and dissection. JCI Insight 3(5)

Didangelos A et al (2012) Novel role of ADAMTS-5 protein in proteoglycan turnover and lipoprotein retention in atherosclerosis. J Biol Chem 287(23):19341–19345

Ezura Y et al (2000) Differential expression of lumican and fibromodulin regulate collagen fibrillogenesis in developing mouse tendons. J Cell Biol 151(4):779–788

Fawcett JW, Oohashi T, Pizzorusso T (2019) The roles of perineuronal nets and the perinodal extracellular matrix in neuronal function. Nat Rev Neurosci 20(8):451–465

Filmus J, Capurro M (2014) The role of glypicans in Hedgehog signaling. Matrix Biol 35:248–252
Funderburgh JL (2002) Keratan sulfate biosynthesis. IUBMB Life 54(4):187–194
Gomez Toledo A et al (2015) Positive mode LC-MS/MS analysis of chondroitin sulfate modified glycopeptides derived from light and heavy chains of the human inter-alpha-trypsin inhibitor complex. Mol Cell Proteomics 14(12):3118–3131
Gondelaud F, Ricard-Blum S (2019) Structures and interactions of syndecans. FEBS J 286 (15):2994–3007
Hascall VC, Heinegard D (1974) Aggregation of cartilage proteoglycans. I. The role of hyaluronic acid. J Biol Chem 249(13):4232–4241
Heinegard D, Saxne T (2011) The role of the cartilage matrix in osteoarthritis. Nat Rev Rheumatol 7 (1):50–56
Iozzo RV, Schaefer L (2015) Proteoglycan form and function: a comprehensive nomenclature of proteoglycans. Matrix Biol 42:11–55
Ly L, Wasinger VC (2011) Protein and peptide fractionation, enrichment and depletion: tools for the complex proteome. Proteomics 11(4):513–534
Mead TJ et al (2018) ADAMTS9-regulated pericellular matrix dynamics governs focal adhesion-dependent smooth muscle differentiation. Cell Rep 23(2):485–498
Midura RJ et al (2018) Quantification of hyaluronan (HA) using a simplified fluorophore-assisted carbohydrate electrophoresis (FACE) procedure. Methods Cell Biol 143:297–316
Mitsou I, Multhaupt HAB, Couchman JR (2017) Proteoglycans, ion channels and cell-matrix adhesion. Biochem J 474(12):1965–1979
Nastase MV et al (2018) Small leucine-rich proteoglycans in renal inflammation: two sides of the coin. J Histochem Cytochem 66(4):261–272
Noborn F et al (2015) Identification of chondroitin sulfate linkage region glycopeptides reveals prohormones as a novel class of proteoglycans. Mol Cell Proteomics 14(1):41–49
Noborn F et al (2016) Site-specific identification of heparan and chondroitin sulfate glycosaminoglycans in hybrid proteoglycans. Sci Rep 6:34537
Noborn F et al (2018) Expanding the chondroitin glycoproteome of Caenorhabditis elegans. J Biol Chem 293(1):379–389
Ortmann C et al (2015) Sonic hedgehog processing and release are regulated by glypican heparan sulfate proteoglycans. J Cell Sci 128(12):2374–2385
Robinson KA et al (2017) Decorin and biglycan are necessary for maintaining collagen fibril structure, fiber realignment, and mechanical properties of mature tendons. Matrix Biol 64:81–93
Rojas-Macias MA et al (2019) Towards a standardized bioinformatics infrastructure for N- and O-glycomics. Nat Commun 10(1):3275
Rowlands D et al (2018) Aggrecan directs extracellular matrix-mediated neuronal plasticity. J Neurosci 38(47):10102–10113
Sarrazin S, Lamanna WC, Esko JD (2011) Heparan sulfate proteoglycans. Cold Spring Harb Perspect Biol 3(7)
Scott JE (1985) Proteoglycan histochemistry–a valuable tool for connective tissue biochemists. Coll Relat Res 5(6):541–575
Suna G et al (2018) Extracellular matrix proteomics reveals interplay of aggrecan and aggrecanases in vascular remodeling of stented coronary arteries. Circulation 137(2):166–183
Talusan P et al (2005) Analysis of intimal proteoglycans in atherosclerosis-prone and atherosclerosis-resistant human arteries by mass spectrometry. Mol Cell Proteomics 4(9):1350–1357
Thelin MA et al (2013) Biological functions of iduronic acid in chondroitin/dermatan sulfate. FEBS J 280(10):2431–2446
Tsien RY (2013) Very long-term memories may be stored in the pattern of holes in the perineuronal net. Proc Natl Acad Sci U S A 110(30):12456–12461
van Wijk XM, van Kuppevelt TH (2014) Heparan sulfate in angiogenesis: a target for therapy. Angiogenesis 17(3):443–462

Vijayagopal P et al (1996) Isolation and characterization of a proteoglycan variant from human aorta exhibiting a marked affinity for low density lipoprotein and demonstration of its enhanced expression in atherosclerotic plaques. Atherosclerosis 127(2):195–203

Wight TN, Merrilees MJ (2004) Proteoglycans in atherosclerosis and restenosis: key roles for versican. Circ Res 94(9):1158–1167

Yu WH, Woessner JF Jr (2000) Heparan sulfate proteoglycans as extracellular docking molecules for matrilysin (matrix metalloproteinase 7). J Biol Chem 275(6):4183–4191

Zhang G et al (2006) Decorin regulates assembly of collagen fibrils and acquisition of biomechanical properties during tendon development. J Cell Biochem 98(6):1436–1449

Zimmermann DR, Dours-Zimmermann MT (2008) Extracellular matrix of the central nervous system: from neglect to challenge. Histochem Cell Biol 130(4):635–653

Chapter 5
Historical Overview of Integrated GAG-omics and Proteomics

Manveen K. Sethi and Joseph Zaia

Abstract The ECM is a complex molecular network that surrounds all cells and consists of proteins, glycoproteins, hyaluronan, glycosaminoglycans (GAGs), and proteoglycans (PGs). GAGs and PGs play vital roles in ECM-related processes such as cell migration, proliferation, adhesion, and differentiation. Among the *omics* technologies, including genomics, transcriptomics, and proteomics, glycomics is the least mature. Over the past two decades, with efforts from glycoscientists around world, and the advent of new glyco-techonologies, databases, tools, and methods, much progress has been made. Now, the focus is to integrate proteomics and glycomics domains in a new platform to analyze and characterize biomolecule classes and define their structural and functional roles. Towards this end, we have developed approaches that integrate analysis of GAGs and proteins towards an end goal of elucidating pathophysiological mechanisms to inform development of disease therapies and regenerative medicine. In this chapter, we provide a historical overview of our groups' methods for glycomics, glycoproteomics, and proteomics of key ECM constituents, i.e., GAGs and PGs, as reported over the past decade.

Abbreviations

CE	Capillary electrophoresis
CS	Chondroitin sulfate
DS	Dermatan Sulfate
ECM	Extracellular matrix
GAGs	Glycosaminoglycans
GalNAc	*N*-acetylgalactosamine
GlcA	Glucuronic acid
GlcNAc	*N*-acetylglucosamine

M. K. Sethi · J. Zaia (✉)
Center for Biomedical Mass Spectrometry, Department of Biochemistry, Boston University School of Medicine, Boston, MA, USA
e-mail: jzaia@bu.edu

S. Ricard-Blum (ed.), *Extracellular Matrix Omics*, Biology of Extracellular Matrix 7,
https://doi.org/10.1007/978-3-030-58330-9_5

HA	Hyaluronic acid
HEP	Heparin
HILIC	Hydrophilic interaction chromatography
HS	Heparan sulfate
IdoA	Iduronic acid
IMS	Imaging mass spectrometry
IPRP	Ion-pair reversed-phase chromatography
KS	Keratan sulfate
LC	Liquid chromatography
LIF	Laser-induced fluorescence
MALDI	Matrix-assisted laser desorption/ionization
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
MWCO	Molecular weight cut-off filters
PGs	Proteoglycans
PNNs	Perineuronal nets
PTMs	Post-translational modifications
SEC	Size exclusion chromatography
TOF	Time-of-flight

5.1 Introduction

Glycosylation is required for all life forms and abundant, with over 50% of mammalian proteins being glycosylated (Apweiler et al. 1999). In comparison to other major biomolecules such as proteins and DNA, the biological roles of carbohydrates remain poorly understood. Carbohydrates, being complex and remarkably diverse in nature, are arduous to synthesize, characterize, and analyze. However, over the past decade, with the advent of new technologies, experimental techniques, and instrumentation, analysis of glycans and glycoproteins, also formally known as 'glycomics' and 'glycoproteomics', respectively, has gained momentum.

Proteoglycans (PGs) and glycosaminoglycans (GAGs) are ubiquitous components of the ECM and play essential roles in all areas of physiology, including cell signaling, cell adhesion, and cell functions (Afratis et al. 2012). PGs are composed of core proteins to which GAG chains are attached. GAGs are linear polysaccharides consisting of repeating disaccharide units of hexosamine (*N*-acetylglucosamine or *N*-acetylgalactosamine), and hexuronic acid (glucuronic acid or iduronic acid) that are covalently attached. GAGs can be divided into categories based on the repeated disaccharide unit, i.e., heparan sulfate (HS), chondroitin sulfate (CS), dermatan sulfate (DS), keratan sulfate (KS) and hyaluronic acid (or hyaluronan) (HA) (Table 5.1) (Sethi and Zaia 2017). They are heterogeneous concerning chain length and subsequent modifications, including sulfation, acetylation, and uronic acid epimerization of disaccharide units. GAG structure is spatially and temporally

Table 5.1 Type of glycosaminoglycan (GAG), and its repeating disaccharide unit

Type of GAG	Disaccharide unit
Hyaluronan (HA)	GlcNAc and GlcA
Chondroitin sulfate (CS)	GalNAc and GlcA
Dermatan sulfate (DS)	GalNAc and GlcA/IdoA
Heparin sulfate (HS)	GlcNAc and GlcA/IdoA
Keratan sulfate (KS)	GlcNAc and Gal

GlcNAc *N*-acetylglucosamine, *GalNAc* *N*-acetylgalactosamine, *GlcA* glucuronic acid, *IdoA* iduronic acid, *Gal* galactose

regulated and plays specific and distinct functional roles during development and disease onset (Iozzo and Schaefer 2015). Unlike other GAGs, HA does not contain sulfate and is not bound to a core protein; rather it exists as a molecular backbone for extracellular matrix complexes consisting of glycoproteins, proteoglycans, collagens and other interacting molecules.

The ECM is a complex molecular network that surrounds all cells, occupying approximately a 20% volume fraction of the adult brain (Sykova and Nicholson 2008). Its main components include hyaluronan, proteoglycans, glycoproteins, and a variety of posttranslational remodeling proteases, such as matrix metalloproteinases (MMPs), which cleave ECM molecules, allowing for highly dynamic functional adaptations (Muir et al. 2002; Rivera et al. 2010). Organized forms of ECM, namely perineuronal nets (PNNs), composed of hyaluronan, proteoglycans, glycoproteins, and collagen, surround the synapse and interact with cell surface receptors. In pathologies, including cancers, cardiovascular diseases, fibrosis, neurodevelopmental and neuropsychiatric diseases, ECM structure and function becomes dysregulated. Thus, characterizing the ECM structure is central to the understanding of physiology and pathophysiology in many diseases (Raghunathan et al. 2019a). The matrisome is defined as the supramolecular complexes, consisting of proteoglycans, glycoproteins, collagens, and hyaluronan, that form the functional units of the ECM (Martin et al. 1984), including the associated molecules. According to the matrisome project, the core matrisome consists of 195 glycoproteins, 44 collagens, and 35 proteoglycans (Shao et al. 2019).

Large scale proteomics studies have quantified, and cataloged expression patterns of various ECM and associated proteins (Byron et al. 2013; Chang et al. 2016; Goddard et al. 2016; Hill et al. 2015; Lindsey et al. 2016; Naba et al. 2012, 2015, 2017), but have not defined the glycosylation patterns of these proteins. It is essential to profile the glycosylation of the matrisome molecules to understand its structural, functional, and biological role in critical molecular mechanisms necessary to understand biomolecular deregulation related to a disease or condition.

Our group has developed methods for performing GAG glycomics, proteomics, glycoproteomics in both separate and integrated forms employing multiple experimental techniques such as *in solution* and *on slide* tissue digestion followed by liquid chromatography based-tandem mass spectrometry (LC-MS/MS). In this chapter, we provide a historical overview of these methods for glycomics, glycoproteomics, and proteomics of key ECM constituents, i.e., GAGs and PGs, as reported over the past

decade (Bielik and Zaia 2010, 2011; Bowman and Zaia 2010; Gill et al. 2013; Hitchcock et al. 2008a, b; Huang et al. 2011; Khatri et al. 2014, 2016; Klein et al. 2018; Leymarie et al. 2012; Raghunathan et al. 2019b; Shao et al. 2013a, b; Shi et al. 2012; Staples et al. 2009, 2010; Turiak et al. 2014). We describe the experimental approaches and their optimization to achieve higher coverage and better quality data.

5.2 Glycosaminoglycan Analysis/GAG-omics

5.2.1 Overview of GAGs

The GAG classes include unsulfated hyaluronan (HA), and sulfated heparin/heparin sulfate (HS), chondroitin/dermatan sulfate (CS/DS), and keratan sulfate (KS) (Fig. 5.1a–d). HS and CS are unbranched polymers composed of ~20–200 repeating disaccharide units; *N*-acetylgalactosamine (GalNAc) or *N*-acetylglucosamine (GlcNAc) and uronic acid, e.g., glucuronate (GlcA) or iduronate (IdoA) attached to serine or threonine residue of core protein through a characteristic tetrasaccharide linker (Kjellen and Lindahl 1991). In contrast, HA is not covalently attached to a core protein and is not sulfated, but consists of repeating disaccharide units of GlcA and GlcNAc attached via alternating β1,3- and β1,4-glycosidic linkage (Sethi and Zaia 2017). KS is composed of repeating disaccharide units of Gal, and GlcNAc via alternating β1,4 and β1,3-glycosidic linkage. The KS GAG chain may be attached to the core protein in three ways: KSI, where the GAG is a sulfated lactosamine chain attached to an *N*-glycan, KSII where the sulfated lactosamine chains are attached to *O*-linked glycans on serine/threonine residues, and KSIII which the GAG chains are attached to the core protein through mannose-Ser linkage (Funderburgh 2000). Identified over 100 years ago, GAGs are found in mast cell granules, cell surfaces, basement membrane, and extracellular matrix (ECM) (Zaia 2008). The sulfated GAGs consist of repeating disaccharide units that become modified biosynthetically via a series of enzymatic events, including deacetylation, sulfation, and epimerization. Figure 5.1e shows CS and HS biosynthesis. These spatial and temporal variations in GAG structure give rise to context-specific interactions with protein partners, growth factors, receptors, ligands responsible for critical biological processes such as cell signaling, adhesion, and interaction in normal and pathological conditions. These modifications are also responsible for the heterogeneous, anionic, and complex nature of GAGs that make GAG analytically challenging to study (Zaia 2005).

5.2.2 Analytical Challenges of GAG Analysis

Routinely, a GAG oligosaccharide is subjected to a series of chemical and enzymatic degradation steps and analyzed using chromatographic, electrophoretic, or mass

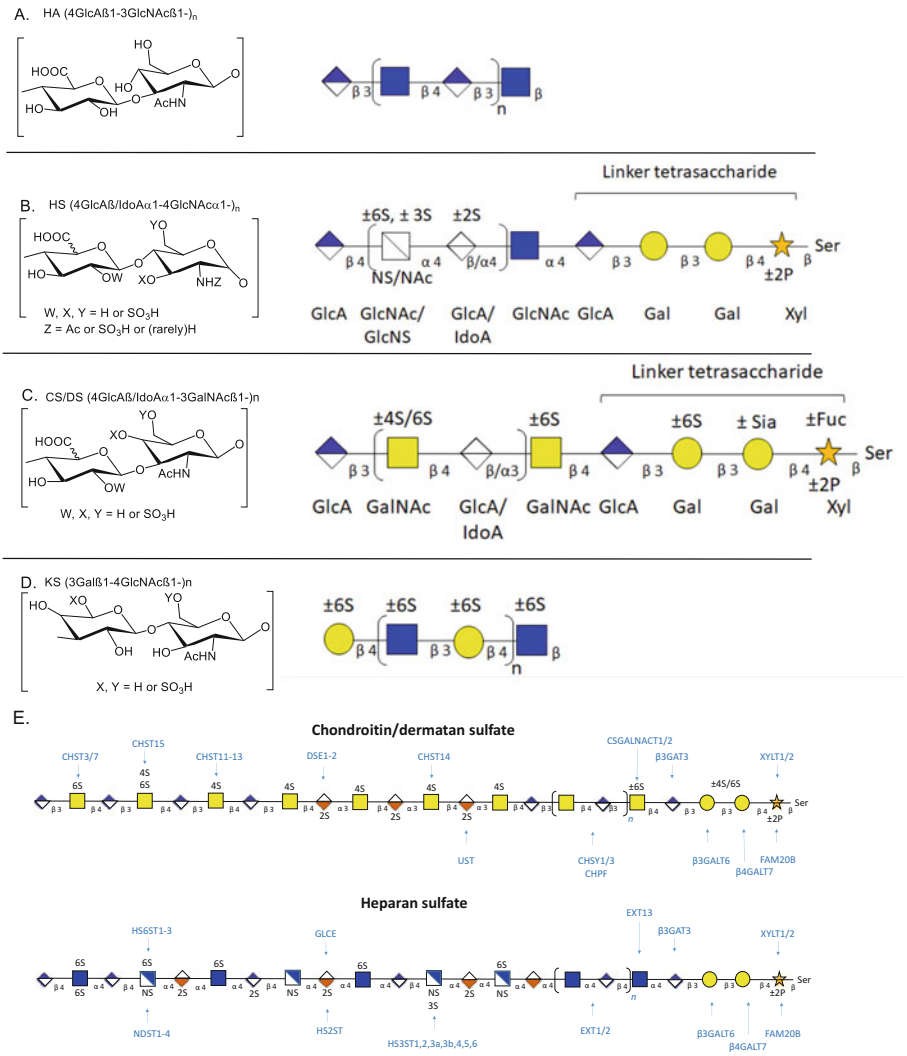

Fig. 5.1 Structure of key Glycosaminoglycans (GAGs). (**a**) Heparan sulfate (HS) disaccharide unit, HS chains and linker tetrasaccharide structure. (**b**) Chondroitin sulfate (CS) sulfate disaccharide unit, CS chains and linker tetrasaccharide structure. (**c**) Hyaluronic acid (HA) disaccharide unit. (**d**) Keratan sulfate (KS) disaccharide unit and chains (**e**), CS and HS biosynthesis (adapted from Lindahl et al. 2015). *GlcA* glucuronic acid, *GalNAc* *N*-acetylgalactosamine, *GlcNAc* *N*-acetylglucosamine, *Ac* acetyl, *S0$_3$H* sulfate

spectrometric methods (Conrad 1997; Turnbull et al. 1999; Venkataraman et al. 1999; Zaia 2009). Mass spectrometry serves as an essential tool for structural analysis of GAGs with high sensitivity and versatility. Over the years, several MS-based methods for GAG-glycomics or GAG-omics have been reported, including matrix-assisted laser desorption ionization (MALDI)-MS, size exclusion chromatography (SEC)-MS, and ion-pair reversed-phase chromatography (IPRP)-MS, and HILIC-MS (Henriksen et al. 2004; Hitchcock et al. 2008b; Laremore and Linhardt 2007; Liu et al. 2019; Shao et al. 2013b; Venkataraman et al. 1999; Wang et al. 2012). Success in MS analysis of GAGs depends largely on the extraction and workup methods used. In particular, it is important to remove salts, contaminants, nucleic acids, or lipids that could interfere with further analysis (Zaia 2009). Other analytical challenges include problems with recovery of GAGs from liquid chromatography (LC) system as charged glycans may stick to the titanium containing metallic loops, filters or transfer lines. In addition, it is necessary to use mass spectrometer fragile ion tuning parameters to minimize the extent to which sulfated ions dissociate during desolvation and ion transfer prior to mass analysis (Staples and Zaia 2011). We and others have developed effective analytical techniques to overcome these challenges (Bodet et al. 2017; Henriksen et al. 2004; Hitchcock et al. 2008a; Laremore and Linhardt 2007; Liu et al. 2019; Shao et al. 2013a; Solakyildirim 2019; Staples et al. 2009, 2010; Wang et al. 2012).

5.2.3 GAG LC-MS/MS Analysis Using SEC and Amide-HILIC

In 2006, we demonstrated successful LC-MS/MS platform with a compatible extraction method for quantifying CS GAGs using a size exclusion column (SEC) with on-line MS detection (Hitchcock et al. 2006). Despite its robustness and reliability, SEC is a low-resolution technique. Thus, in 2008 we implemented LC-MS/MS platforms utilizing amide-hydrophilic interaction chromatography (HILIC) instead of SEC for analyzing CS GAGs from connective tissues. We were able to profile the GAG chain non-reducing end, the linker region, and Δ-unsaturated interior oligosaccharide domains of the CS chains. The GAGs were extracted from the core protein using sequential β-elimination, C-18 cleanup to remove hydrophobic molecules, and finally, anion exchange spin columns to remove cationic molecules. The eluted anionic GAG mixture was then partially depolymerized with chondroitinase enzymes, and further differentially stable isotope-labeled by reductive amination using 2-anthranilic acid—d_0 and d_4, and subjected to amide-HILIC on-line LC-MS/MS analysis (Hitchcock et al. 2008b). One limiting factor of using amide-HILIC LC-MS/MS was the stability of the spray interface as conventional silica sprayers clogged in negative mode, and thus, required time consuming and extensive optimization. This problem was solved by

using a non-silica sprayer such as provided by the Agilent Chip Cube and the Advion NanoMate robot.

5.2.4 GAG CE-LIF Analysis

We then ventured into capillary electrophoresis (CE) coupled with laser-induced fluorescence (LIF) for GAG disaccharide compositional analysis, an essential step towards understanding the GAG structure-function relationship. We first reported a method that utilized capillary electrophoresis (CE) with laser-induced fluorescence (LIF) to analyze GAG disaccharides in the biological samples. This method made several improvements to existing methods including, optimization of reductive amination conditions, an increase in sensitivity by using cellulose cleanup for derivatization, and optimization of separation for reproducibility and robustness (Hitchcock et al. 2008a). CE has various benefits over other analytical methods, including high resolving power and separation efficiency for disaccharide structural isomers differing in sulfation position, economical (use of less buffer and sample), faster, automated and reproducible analysis, but has not been widely used mainly because of disaccharides recovery issues after derivatization workup. Our method eliminated noise background and improved quantification of biological samples by 100-fold, and thus, enabled disaccharide quantification of HS and CS GAGs from biologically relevant PGs and intact tissue samples. This method was not, however, compatible with on-line MS detection.

5.2.5 GAG HILIC-CHIP-MS Based Analysis

In order to improve the chromatographic resolution of the LC-MS method for GAGs, we utilized a novel chip-based amide-HILILC system for negative ion LC-MS/MS of partially depolymerized heparin/HS, and CS/DS GAGs. The chip-based trapping cartridges assisted in the removal of contaminating proteins, lipids, nucleic acids, and acidic non-GAG carbohydrates by focusing the analyte in the MS while allowing contaminants to flow through with minimum interaction with the stationary phase. We were able to achieve robust positioning of the spray needle and the analysis of GAGs isolated from complex biological and chemical samples (Staples et al. 2009). In this work, we noted that there was a physical limitation for analysis of highly sulfated (polar) GAG oligosaccharides such as HS dp10s that start to elute when the source voltage is not able to maintain the electrospray.

To overcome this problem, we optimized the novel amide-HILIC HPLC CHIP platform with an introduction of makeup flow (MUF) (Staples et al. 2010). The MUF chips allowed electrospray in high aqueous conditions during negative-ion mode LC-MS, thus, eliminating the need to raise spray voltages as aqueous content increased. We used this chip for analysis of highly modified GAG domains involved

in various biological processes. We were able to analyze dp10-dp14 HS and dp14-18 heparin oligosaccharides, which was not possible with a standard amide-HILIC HPLC chips (Staples et al. 2010). The HILIC-Chip based platform was a unique platform for analysis of GAGs but the ions observed were low in charge, resulting in undesirable sulfate loss from precursor ion during collision-induced dissociation (CID). To overcome sulfate losses, we used metal cation adducts to stabilize sulfate groups or nonvolatile polar compounds such as sulfolane to supercharge proteins could be added. Thus, we utilized microfluidic novel pulsed makeup flow (MUF) HPLC-chips that enabled controlled application of additives during a given chromatographic window and thus, reduced the nonvolatile additive build up in the ion source. Using these chips, the tandem-MS of these supercharged precursor ions showed significant decrease in sulfate loss (Huang et al. 2011). We further worked to improve tandem mass spectrometry of GAGs by reducing sulfate loss and generating better product ion profiles (Bielik and Zaia 2011; Leymarie et al. 2012; Shi et al. 2012).

5.2.6 *Tetraplex Stable Isotope-Coded Based Quantitative GAG Glycomics*

We demonstrated an effective method for tetraplex stable isotope-labeled reductive amination tags for quantitative glycomics of chondroitin sulfate proteoglycans (CSPGs), pharmaceutical heparins, and *N*-glycans from glycoproteins subjected to an online LC-MS platform as well as tandem mass spectrometry which was used or comparison of isomeric glycan fine structures from various samples. This method provided not only a precise compositional profiling of GAGs but also fine structural compositions together with multiplexing benefits for high-throughput (Bowman and Zaia 2010).

5.2.7 *GAG Disaccharide Analysis Using HILIC LC-MS*

Using a single LC-MS platform to generate complete disaccharide profiles for GAG, we utilized HILIC-MS for quantification of both enzyme-derived and nitrous acid depolymerization products for structural analysis of HS and CS/DS GAGs (Gill et al. 2013). HILIC is one of the most widely used separation tools for glycans. It offers several advantages such as shorter sample preparation time, ultrafast analysis due to low column backpressure and improved MS sensitivity. HILIC with online ESI-MS has been used widely for the analysis of released glycans (Luo et al. 2009; Mauko et al. 2011; Ruhaak et al. 2008; Zauner et al. 2011), glycopeptides (Calvano et al. 2008; Wohlgemuth et al. 2009; Zauner et al. 2010), GAG oligosaccharides (Huang et al. 2011; Kailemia et al. 2014; Staples et al. 2010). For GAG disaccharide

analysis the challenge with HILIC is to find mobile phase conditions that achieve efficient retention of disaccharides containing a range of 0–4 sulfate groups. Generally speaking, chromatographic resolution is best for 2.1 and 1.0 mm internal diameter columns. It is typically necessary to use a tandem MS step to differentiate isomeric disaccharides that co-elute using HILIC (Gill et al. 2013).

5.3 On-Slide Tissue Digestion Coupled with LC-MS/MS for Integrated Glycomics and Proteomics

We innovated a novel *on slide* digestion platform in our lab that utilized serial enzyme digestions from surfaces of fresh frozen or fixed tissue sections (Raghunathan et al. 2019b; Shao et al. 2013a; Turiak et al. 2014). To understand the biological roles played by GAGs and PGs expression during pathogenesis, it is crucial to detect and profile GAGs and proteins at the histological scale to minimize cell heterogeneity and potentially inform diagnosis and prognosis. This method provided a readout of HA, CS, HS GAG quantities, domain structures, and non-reducing end structures as well as *N*-glycans, and proteins using a simple workflow of application of enzyme and extraction of biomolecules with minimal need for workup (Fig. 5.2). The method was able to quantify different biomolecules and perform integrated *omics* for tissue volumes of 10 nL or greater, corresponding to a 1 μL droplet of enzyme solution applied to a 1 mm diameter target on a 10 μm thick tissue slide. Using this method allowed the staining of parallel sections or immunohistochemistry to guide the selection of the target area on an unstained tissue section. This method provides a targeted approach to analyze a specific tissue area, for example, tumor vs. non-tumor, myelin vs. non-myelin, etc., and uncover detailed structural profiles and establish a functional relationship to understand the disease or normal pathology.

Compared to *in solution* digestion (Ji et al. 2015; Wisniewski 2016), on slide digestion is more economical in terms of time required per sample. On slide digestion also requires less post-digestion cleanup prior to the LC-MS step. The LC step results in higher dynamic range of detection for GAGs and proteins than can be achieved using MALDI imaging mass spectrometry (IMS) (Raghunathan et al. 2019b; Shao et al. 2013a; Turiak et al. 2014). By contrast, MALDI-IMS has the advantage of higher tissue spatial resolution than the *on slide* digestion method (Drake et al. 2017, 2018a).

In 2013, we reported this method for comparative glycomics profiling of HS disaccharides from human astrocytoma, and glioblastoma tissues (Shao et al. 2013a). Later, in 2014, we modified the technique to include various compound classes GAGs, *N*-glycans, and proteins/peptides using the bovine cortex and mouse brain tissue sections (Turiak et al. 2014). The data from a small 1.5 mm diameter tissue spot was consistent with previously published bulk mouse, liver, and brain tissue demonstrating the power of our method. More recently, we reduced the number of

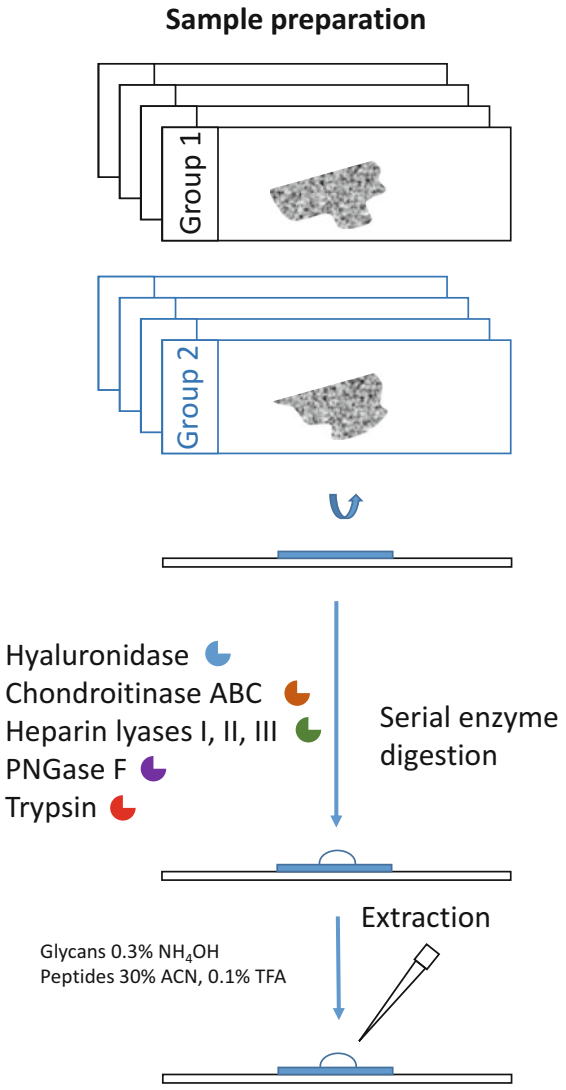

Fig. 5.2 Schematic representation of *On slide* tissue digestion workflow. Taken and modified from Raghunathan et al. (2019b)

processing steps by digesting HS disaccharides, and *N*-glycans together (Raghunathan et al. 2019b).

We have applied this *state-of-the-art* platform to understand various brain pathologies, including glioblastoma (Shao et al. 2013a), aging (Raghunathan et al. 2018), schizophrenia (unpublished), and Parkinson's disease (unpublished), and have uncovered several dysregulated GAGs, ECM related proteins and pathways.

5.4 In Solution Tissue Digestion for Integrated Proteomics and Glycomics

In the past, we have performed *in solution* tissue digestions to characterize GAGs and proteins but not in a sequential and/or an integrated *omics* manner (Jacobsen et al. 2019; Shao et al. 2013b). Recently, we developed a streamlined serial *in solution* protocol to analyze GAGs and proteins from the brain or other tissues (Fig. 5.3) (*manuscript submitted*). Compared to our *on-slide* digestion protocol that provides a selection of target area on a tissue slides (Raghunathan et al. 2018, 2019b; Shao et al. 2013a; Turiak et al. 2014) and MALDI-imaging method for glycans that offers higher spatial resolution (Drake et al. 2017, 2018a, b), this method can be applied to free-floating or frozen tissues and provides a high depth of coverage. This platform is more rapid (time-effective) and efficient (single-pot) than the currently used parallel approach, i.e., a multi-pot simultaneous enzyme application method (Chen et al. 2017; Shao et al. 2013b; Turnbull et al. 2010). The removal of GAGs also facilitates protein identification of the remaining deglycosylated PGs with higher peptide-coverage using conventional-proteomics (Klein et al. 2018), compared to current studies achieving only low PG-coverage (Donovan et al. 2012; Hondius et al. 2016). The protocol follows a filter-aided sample preparation (FASP) type (Wiśniewski et al. 2009) serial *in-solution* digestion using molecular weight cut-off (MWCO) membrane filters as a reactor to digest glycosaminoglycan (GAG) classes, including HA, CS, and HS, and collect it as a flow-through, and finally collect proteins to perform trypsin digestion to generate peptides from tissue or cell lysates. We have applied this workflow to mouse brain tissue, and human healthy and Alzheimer's brain tissue.

5.5 Deep Sequencing of Proteoglycans

The peptide sequence coverage for large and highly complex PGs containing a high degree of glycosylation arising from GAGs, *N*-glycans, and mucin *O*-glycans are poorly-annotated by conventional MS analysis. Thus, little is known about the role of site-specific glycosylation of PGs in normal and disease pathologies. We developed a workflow (Fig. 5.4) to improve sequence coverage and identification of

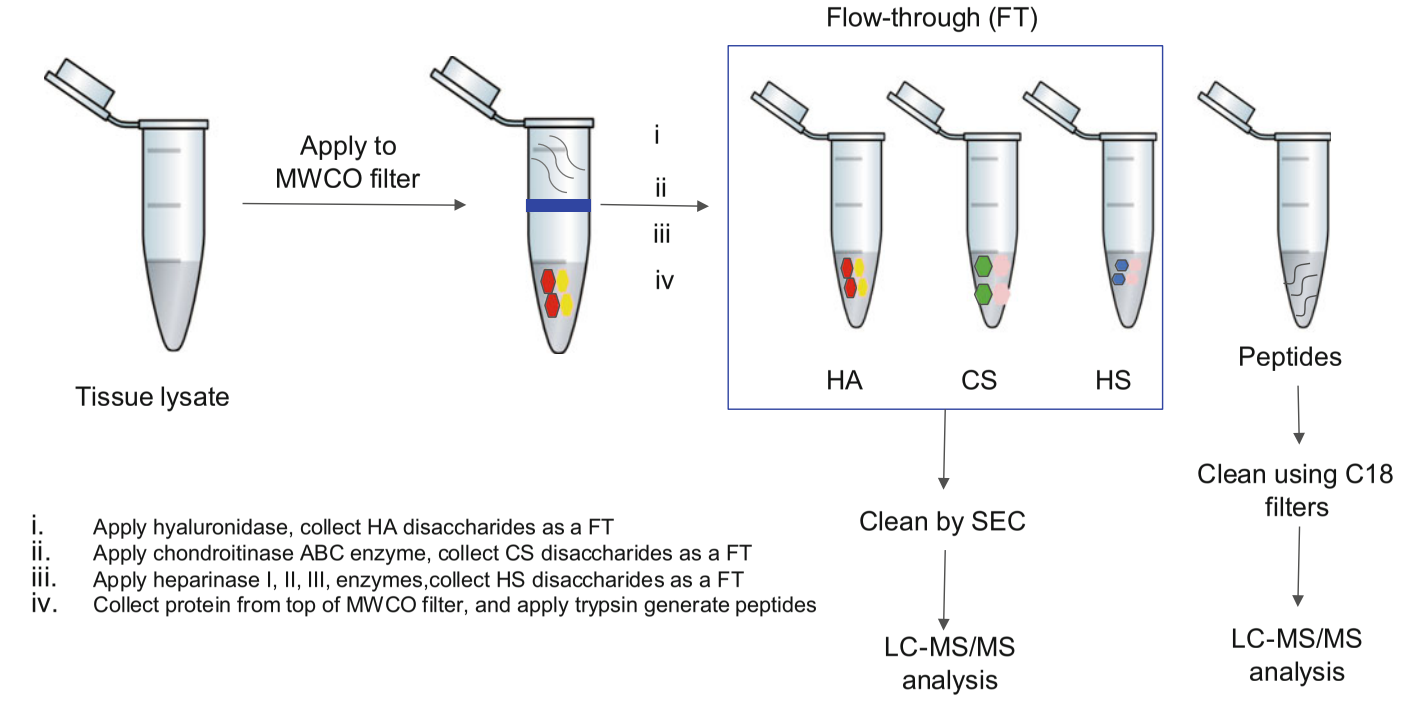

Fig. 5.3 Schematic representation of *In solution* tissue digestion workflow

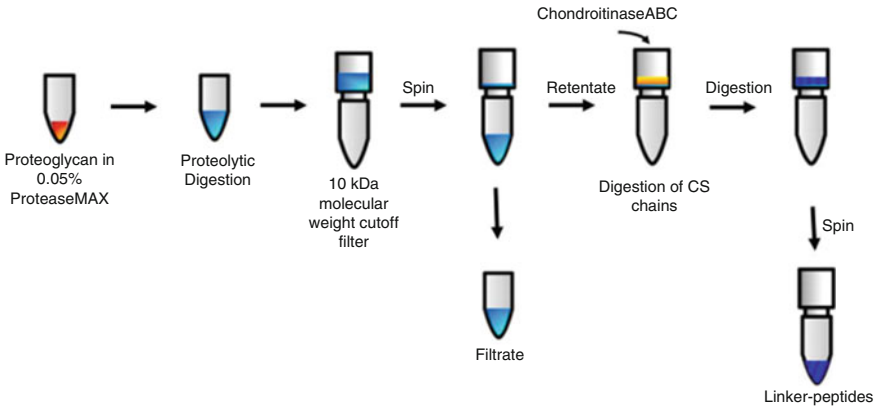

Fig. 5.4 Schematic representation of the workflow for enrichment of proteoglycan linker-peptides taken from Klein et al. (2018)

glycosylated peptides in biologically relevant proteoglycans (PGs), including small leucine-rich proteoglycan (SLRP) decorin and three hyalectan proteoglycans: neurocan, brevican, and aggrecan necessary to understand their role in pathophysiology (Klein et al. 2018). Using the workflow, we were able to identify linker-glycosite (created by removal of GAGs that leaves a linker tetrasaccharide plus one disaccharide to the protein/peptide), and 3 *N*-glycosylation sites for decorin, densely glycosylated mucin like region in the extended domain for neurocan and brevican, and 50 linker-glycosites and mucin-type *O*-glycosites in the extended region and *N*-glycosites in the globular domains for Aggrecan, many of which were not previously identified or reported.

5.6 Conclusions

Over the past decade, glycoscientists around the world have created a vast pool of knowledge, glyco-databases, and glyco-technologies to characterize and analyse glycans, to define their structural composition, and relate their biological functions. Mass spectrometry has played a major role in determining the structural compositions of various biomolecules, and multiple disciplines *viz.* genomics, transcriptomics, proteomics, glycomics, and glycoproteomics have been integrated *omics* to address biologically relevant questions for the understanding of the biological system. Towards this end, we have developed various platforms for integrated *omics* approach and gain insights into development, disease, therapy, and regenerative medicine.

At this time, there are effective analytical methods for the combined analysis of GAGs and proteins from tissue samples. These methods employ digestion steps prior to electrospray mass spectral analysis. The first system developed employed

SEC-MS, a system that is extremely robust but of limited sensitivity. The use of HILIC-MS allows effective profiling of GAG oligosaccharide mixtures and is the preferred method for disaccharide analysis. That HILIC-MS can be reduced in scale allows it to be used for detection of GAGs released using *on slide* digestion. We have analyzed tissue cohorts of several dozen samples using this approach. In order to improve the depth and sensitivity of PG coverage, we optimized *in solution* enrichment and digestion protocols. Looking ahead, there is no barrier to quantitative profiling of GAGs and proteins from tissue. Analytical throughput would be improved by application of robotic automation. The use of robotics may also reduce the volume of tissue required for analysis.

Acknowledgements This work was supported by NIH grants P41GM104602, R01MH105608, and R01GM133963.

References

Afratis N, Gialeli C, Nikitovic D, Tsegenidis T, Karousou E, Theocharis AD, Pavao MS, Tzanakakis GN, Karamanos NK (2012) Glycosaminoglycans: key players in cancer cell biology and treatment. FEBS J 279:1177–1197

Apweiler R, Hermjakob H, Sharon N (1999) On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. Biochim Biophys Acta 1473:4–8

Bielik AM, Zaia J (2010) Extraction of chondroitin/dermatan sulfate glycosaminoglycans from connective tissue for mass spectrometric analysis. Methods Mol Biol 600:215–225

Bielik AM, Zaia J (2011) Multistage tandem mass spectrometry of chondroitin sulfate and dermatan sulfate. Int J Mass Spectrom 305:131–137

Bodet PE, Salard I, Przybylski C, Gonnet F, Gomila C, Ausseil J, Daniel R (2017) Efficient recovery of glycosaminoglycan oligosaccharides from polyacrylamide gel electrophoresis combined with mass spectrometry analysis. Anal Bioanal Chem 409:1257–1269

Bowman MJ, Zaia J (2010) Comparative glycomics using a tetraplex stable-isotope coded tag. Anal Chem 82:3023–3031

Byron A, Humphries JD, Humphries MJ (2013) Defining the extracellular matrix using proteomics. Int J Exp Pathol 94:75–92

Calvano CD, Zambonin CG, Jensen ON (2008) Assessment of lectin and HILIC based enrichment protocols for characterization of serum glycoproteins by mass spectrometry. J Proteome 71:304–317

Chang CW, Dalgliesh AJ, Lopez JE, Griffiths LG (2016) Cardiac extracellular matrix proteomics: challenges, techniques, and clinical implications. Proteomics Clin Appl 10:39–50

Chen J, Kawamura T, Sethi MK, Zaia J, Repunte-Canonigo V, Sanna PP (2017) Heparan sulfate: resilience factor and therapeutic target for cocaine abuse. Sci Rep 7:13931

Conrad HE (1997) Heparin-binding proteins. Elsevier

Donovan LE, Higginbotham L, Dammer EB, Gearing M, Rees HD, Xia Q, Duong DM, Seyfried NT, Lah JJ, Levey AI (2012) Analysis of a membrane-enriched proteome from postmortem human brain tissue in Alzheimer's disease. Proteomics Clin Appl 6:201–211

Drake RR, Powers TW, Jones EE, Bruner E, Mehta AS, Angel PM (2017) MALDI mass spectrometry imaging of N-linked glycans in cancer tissues. Adv Cancer Res 134:85–116

Drake RR, Powers TW, Norris-Caneda K, Mehta AS, Angel PM (2018a) In situ imaging of N-glycans by MALDI imaging mass spectrometry of fresh or formalin-fixed paraffin-embedded tissue. Curr Protocols Protein Sci 94:e68

Drake RR, West CA, Mehta AS, Angel PM (2018b) MALDI mass spectrometry imaging of N-linked glycans in tissues. Adv Exp Med Biol 1104:59–76

Funderburgh JL (2000) MINI REVIEW Keratan sulfate: structure, biosynthesis, and function. Glycobiology 10:951–958

Gill VL, Aich U, Rao S, Pohl C, Zaia J (2013) Disaccharide analysis of glycosaminoglycans using hydrophilic interaction chromatography and mass spectrometry. Anal Chem 85:1138–1145

Goddard ET, Hill RC, Barrett A, Betts C, Guo Q, Maller O, Borges VF, Hansen KC, Schedin P (2016) Quantitative extracellular matrix proteomics to study mammary and liver tissue microenvironments. Int J Biochem Cell Biol 81:223–232

Henriksen J, Ringborg LH, Roepstorrf P (2004) On-line size-exclusion chromatography/mass spectrometry of low molecular mass heparin. J Mass Spectrom 39:1305–1312

Hill RC, Calle EA, Dzieciatkowska M, Niklason LE, Hansen KC (2015) Quantification of extracellular matrix proteins from a rat lung scaffold to provide a molecular readout for tissue engineering. Mol Cell Proteomics 14:961–973

Hitchcock AM, Costello CE, Zaia J (2006) Glycoform quantification of chondroitin/dermatan sulfate using a liquid chromatography-tandem mass spectrometry platform. Biochemistry 45:2350–2361

Hitchcock AM, Bowman MJ, Staples GO, Zaia J (2008a) Improved workup for glycosaminoglycan disaccharide analysis using CE with LIF detection. Electrophoresis 29:4538–4548

Hitchcock AM, Yates KE, Costello CE, Zaia J (2008b) Comparative glycomics of connective tissue glycosaminoglycans. Proteomics 8:1384–1397

Hondius DC, van Nierop P, Li KW, Hoozemans JJ, van der Schors RC, van Haastert ES, van der Vies SM, Rozemuller AJ, Smit AB (2016) Profiling the human hippocampal proteome at all pathologic stages of Alzheimer's disease. Alzheimer's Dementia 12:654–668

Huang Y, Shi X, Yu X, Leymarie N, Staples GO, Yin H, Killeen K, Zaia J (2011) Improved liquid chromatography-MS/MS of heparan sulfate oligosaccharides via chip-based pulsed makeup flow. Anal Chem 83:8222–8229

Iozzo RV, Schaefer L (2015) Proteoglycan form and function: a comprehensive nomenclature of proteoglycans. Mat Biol 42:11–55

Jacobsen A, Shi X, Shao C, Eysturskardelta J, Mikalsen SO, Zaia J (2019) Characterization of glycosaminoglycans in gaping and intact connective tissues of farmed Atlantic Salmon (Salmo salar) fillets by mass spectrometry. ACS Omega 4:15337–15347

Ji IJ, Hua S, Shin DH, Seo N, Hwang JY, Jang IS, Kang MG, Choi JS, An HJ (2015) Spatially-resolved exploration of the mouse brain glycome by tissue glyco-capture (TGC) and nano-LC/MS. Anal Chem 87:2869–2877

Kailemia MJ, Ruhaak LR, Lebrilla CB, Amster IJ (2014) Oligosaccharide analysis by mass spectrometry: a review of recent developments. Anal Chem 86:196–212

Khatri K, Staples GO, Leymarie N, Leon DR, Turiak L, Huang Y, Yip S, Hu H, Heckendorf CF, Zaia J (2014) Confident assignment of site-specific glycosylation in complex glycoproteins in a single step. J Proteome Res 13:4347–4355

Khatri K, Klein JA, White MR, Grant OC, Leymarie N, Woods RJ, Hartshorn KL, Zaia J (2016) Integrated omics and computational glycobiology reveal structural basis for influenza A virus glycan microheterogeneity and host interactions. Mol Cell Proteomics 15:1895–1912

Kjellen L, Lindahl U (1991) Proteoglycans: structures and interactions. Annu Rev Biochem 60:443–475

Klein JA, Meng L, Zaia J (2018) Deep sequencing of complex proteoglycans: a novel strategy for high coverage and site-specific identification of glycosaminoglycan-linked peptides. Mol Cell Proteomics 17:1578–1590

Laremore TN, Linhardt RJ (2007) Improved matrix-assisted laser desorption/ionization mass spectrometric detection of glycosaminoglycan disaccharides as cesium salts. Rapid Commun Mass Spectrom 21:1315–1320

Leymarie N, McComb ME, Naimy H, Staples GO, Zaia J (2012) Differential characterization and classification of tissue specific glycosaminoglycans by tandem mass spectrometry and statistical methods. Int J Mass Spectrom 312:144–154

Lindahl U, Couchman J, Kimata K, Esko JD (2015) Proteoglycans and sulfated glycosaminoglycans. In: Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, Darvill AG, Kinoshita T, Packer NH et al (eds) Essentials of glycobiology. Cold Spring Harbor, New York, pp 207–221

Lindsey ML, Hall ME, Harmancey R, Ma Y (2016) Adapting extracellular matrix proteomics for clinical studies on cardiac remodeling post-myocardial infarction. Clin Proteomics 13:19

Liu H, Joshi A, Chopra P, Liu L, Boons G-J, Sharp JS (2019) Salt-free fractionation of complex isomeric mixtures of glycosaminoglycan oligosaccharides compatible with ESI-MS and microarray analysis. Sci Rep 9:16566

Luo Q, Rejtar T, Wu S-L, Karger BL (2009) Hydrophilic interaction 10μm I.D. porous layer open tubular columns for ultratrace glycan analysis by liquid chromatography–mass spectrometry. J Chromatogr A 1216:1223–1231

Martin GR, Kleinman HK, Terranova VP, Ledbetter S, Hassell JR (1984) The regulation of basement membrane formation and cell-matrix interactions by defined supramolecular complexes. CIBA Found Symp 108:197–212

Mauko L, Pelzing M, Dolman S, Nordborg A, Lacher NA, Haddad PR, Hilder EF (2011) Zwitterionic-type hydrophilic interaction nano-liquid chromatography of complex and high mannose glycans coupled with electrospray ionisation high resolution time of flight mass spectrometry. J Chromatogr A 1218:6419–6425

Muir EM, Adcock KH, Morgenstern DA, Clayton R, von Stillfried N, Rhodes K, Ellis C, Fawcett JW, Rogers JH (2002) Matrix metalloproteases and their inhibitors are produced by overlapping populations of activated astrocytes. Brain Res Mol Brain Res 100:103–117

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

Naba A, Clauser KR, Hynes RO (2015) Enrichment of extracellular matrix proteins from tissues and digestion into peptides for mass spectrometry analysis. J Visual Exp: JoVE e53057

Naba A, Pearce OMT, Del Rosario A, Ma D, Ding H, Rajeeve V, Cutillas PR, Balkwill FR, Hynes RO (2017) Characterization of the extracellular matrix of normal and diseased tissues using proteomics. J Proteome Res 16:3083–3091

Raghunathan R, Polinski NK, Klein JA, Hogan JD, Shao C, Khatri K, Leon D, McComb ME, Manfredsson FP, Sortwell CE et al (2018) Glycomic and proteomic changes in aging brain nigrostriatal pathway. Mol Cell Proteomics 17:1778–1787

Raghunathan R, Sethi MK, Klein JA, Zaia J (2019a) Proteomics, glycomics, and glycoproteomics of matrisome molecules. Mol Cell Proteomics

Raghunathan R, Sethi MK, Zaia J (2019b) On-slide tissue digestion for mass spectrometry based glycomic and proteomic profiling. MethodsX 6:2329–2347

Rivera S, Khrestchatisky M, Kaczmarek L, Rosenberg GA, Jaworski DM (2010) Metzincin proteases and their inhibitors: foes or friends in nervous system physiology? J Neurosci Off J Soc Neurosci 30:15337–15357

Ruhaak LR, Huhn C, Waterreus W-J, de Boer AR, Neusüss C, Hokke CH, Deelder AM, Wuhrer M (2008) Hydrophilic interaction chromatography-based high-throughput sample preparation method for N-glycan analysis from total human plasma glycoproteins. Anal Chem 80:6119–6126

Sethi MK, Zaia J (2017) Extracellular matrix proteomics in schizophrenia and Alzheimer's disease. Anal Bioanal Chem 409:379–394

Shao C, Shi X, Phillips JJ, Zaia J (2013a) Mass spectral profiling of glycosaminoglycans from histological tissue surfaces. Anal Chem 85:10984–10991

Shao C, Shi X, White M, Huang Y, Hartshorn K, Zaia J (2013b) Comparative glycomics of leukocyte glycosaminoglycans. FEBS J 280:2447–2461

Shao X, Taha IN, Clauser KR, Gao Y, Naba A (2019) MatrisomeDB: the ECM-protein knowledge database. Nucleic Acids Res 48:D1136–D1144

Shi X, Huang Y, Mao Y, Naimy H, Zaia J (2012) Tandem mass spectrometry of heparan sulfate negative ions: sulfate loss patterns and chemical modification methods for improvement of product ion profiles. J Am Soc Mass Spectrom 23:1498–1511

Solakyildirim K (2019) Recent advances in glycosaminoglycan analysis by various mass spectrometry techniques. Anal Bioanal Chem 411:3731–3741

Staples GO, Zaia J (2011) Analysis of glycosaminoglycans using mass spectrometry. Curr Proteomics 8:325–336

Staples GO, Bowman MJ, Costello CE, Hitchcock AM, Lau JM, Leymarie N, Miller C, Naimy H, Shi X, Zaia J (2009) A chip-based amide-HILIC LC/MS platform for glycosaminoglycan glycomics profiling. Proteomics 9:686–695

Staples GO, Naimy H, Yin H, Kileen K, Kraiczek K, Costello CE, Zaia J (2010) Improved hydrophilic interaction chromatography LC/MS of heparinoids using a chip with postcolumn makeup flow. Anal Chem 82:516–522

Sykova E, Nicholson C (2008) Diffusion in brain extracellular space. Physiol Rev 88:1277–1340

Turiak L, Shao C, Meng L, Khatri K, Leymarie N, Wang Q, Pantazopoulos H, Leon DR, Zaia J (2014) Workflow for combined proteomics and glycomics profiling from histological tissues. Anal Chem 86:9670–9678

Turnbull JE, Hopwood JJ, Gallagher JT (1999) A strategy for rapid sequencing of heparan sulfate and heparin saccharides. Proc Natl Acad Sci U S A 96:2698–2703

Turnbull JE, Miller RL, Ahmed Y, Puvirajesinghe TM, Guimond SE (2010) Glycomics profiling of heparan sulfate structure and activity. Methods Enzymol 480:65–85

Venkataraman G, Shriver Z, Raman R, Sasisekharan R (1999) Sequencing complex polysaccharides. Science 286:537–542

Wang Z, Zhang F, Dordick JS, Linhardt RJ (2012) Molecular mass characterization of glycosaminoglycans with different degrees of sulfation in bioengineered heparin process by size exclusion chromatography. Curr Anal Chem 8:506–511

Wisniewski JR (2016) Quantitative evaluation of filter aided sample preparation (FASP) and multienzyme digestion FASP protocols. Anal Chem 88:5438–5443

Wiśniewski JR, Zougman A, Nagaraj N, Mann M (2009) Universal sample preparation method for proteome analysis. Nat Methods 6:359

Wohlgemuth J, Karas M, Eichhorn T, Hendriks R, Andrecht S (2009) Quantitative site-specific analysis of protein glycosylation by LC-MS using different glycopeptide-enrichment strategies. Anal Biochem 395:178–188

Zaia J (2005) Principles of mass spectrometry of glycosaminoglycans. J Biomacromol Mass Spectrom 1:3–36

Zaia J (2008) Mass spectrometry and the emerging field of glycomics. Chem Biol 15:881–892

Zaia J (2009) On-line separations combined with MS for analysis of glycosaminoglycans. Mass Spectrom Rev 28:254–272

Zauner G, Koeleman CAM, Deelder AM, Wuhrer M (2010) Protein glycosylation analysis by HILIC-LC-MS of Proteinase K-generated N- and O-glycopeptides. J Sep Sci 33:903–910

Zauner G, Deelder AM, Wuhrer M (2011) Recent advances in hydrophilic interaction liquid chromatography (HILIC) for structural glycomics. Electrophoresis 32:3456–3466

Chapter 6
Extracellular Matrix Networks: From Connections to Functions

Sylvie Ricard-Blum

Abstract The extracellular matrix (ECM) forms of a three dimensional interaction network mostly comprised of proteins, collagens being the most abundant ones (Ricard-Blum, Cold Spring Harb Perspect Biol 3:a004978, 2011), glycosaminoglycans (GAGs) and proteoglycans (PGs) (Iozzo and Schaefer, Matrix Biol 42:11–55, 2015; Karamanos et al., Chem Rev 118:9152–9232, 2018). We review here the major methods used to identify and characterize ECM protein, glycosaminoglycan, and proteoglycan interactions with a focus on high-throughput methods able to identify a number of interactions simultaneously such as yeast two hybrid assays, ECM protein and GAG arrays, and affinity purification coupled to mass spectrometry (MS). The use of large experimental interaction datasets publicly available, and of interaction databases to retrieve interaction data required to build interaction networks is discussed. The interest of the data generated from the functional and structural analyses of interactomes to decipher molecular mechanisms of biological processes, to design further functional experiments, and to select ECM proteins or GAGs and/or their biomolecular interactions as therapeutic targets is illustrated by several examples. The ultimate goal of these studies is to build three-dimensional ECM networks, integrating the 3D structure of individual ECM molecules and their complexes.

6.1 Introduction

The extracellular matrix (ECM) forms of a three dimensional interaction network mostly comprised of proteins, glycosaminoglycans (GAGs) and proteoglycans (PGs). ECM proteins (e.g. collagens, laminins, fibronectin, elastin) and associated proteins, which comprise the matrisome (Chap. 2), self-organize in association with proteoglycans (Iozzo and Schaefer 2015; Theocharis et al. 2019) to form

S. Ricard-Blum (✉)
University of Lyon, Université Claude Bernard Lyon 1, CNRS, INSA Lyon, CPE, Institute of Molecular and Supramolecular Chemistry and Biochemistry, UMR 5246, Villeurbanne, France
e-mail: sylvie.ricard-blum@univ-lyon1.fr

S. Ricard-Blum (ed.), *Extracellular Matrix Omics*, Biology of Extracellular Matrix 7,
https://doi.org/10.1007/978-3-030-58330-9_6

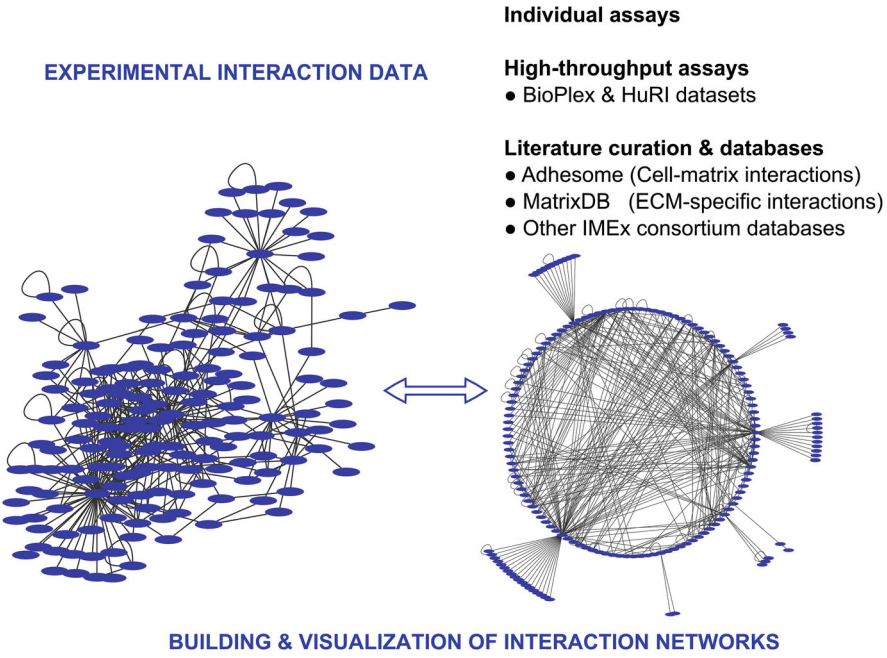

Fig. 6.1 Sources of experimental interaction data used to build an interaction network visualized with two Cytoscape layouts (Shannon et al. 2003) (https://cytoscape.org/). The large datasets are BioPlex (Huttlin et al. 2015, 2017, 2020) (https://bioplex.hms.harvard.edu/), and HuRI (a reference map of the human binary protein interactome, (Luck et al. 2020). Interactions can also be retrieved from databases such as MatrixDB, an ECM database (http://matrixdb.univ-lyon1.fr/), the adhesome, a focal adhesion network (www.adhesome.org), and those from the International Molecular Exchange (IMEx) consortium (https://www.imexconsortium.org/) described in this chapter and in the first chapter

supramolecular assemblies, which are molecular alloys (Bruckner 2010). In addition to the interactions they establish to promote ECM assembly, ECM proteins and proteoglycans exert their biological activities in a concerted and regulated manner by binding to a variety of partners. These dynamic interactions influence each other in vivo and form networks, which are rewired depending on the physiopathological context, and orchestrate the functions of the ECM. The first step to decipher the molecular mechanisms of ECM assembly, homeostasis and changes induced by diseases (Bonnans et al. 2014; Theocharis et al. 2019), to identify new therapeutic targets and to design new therapies is to make an inventory of the biomolecular interactions connecting the ECM constituents (Fig. 6.1), and then to add on this scaffold the information required to build subnetworks specific of a tissue, a molecular function, a biological process or a disease (Fig. 6.2). This process, often referred to as contextualization, can be carried out by integrating experimental data related to ECM proteins or proteoglycans such as genomic, transcriptomic and quantitative proteomic data collected in healthy or diseased tissues. Keywords from the

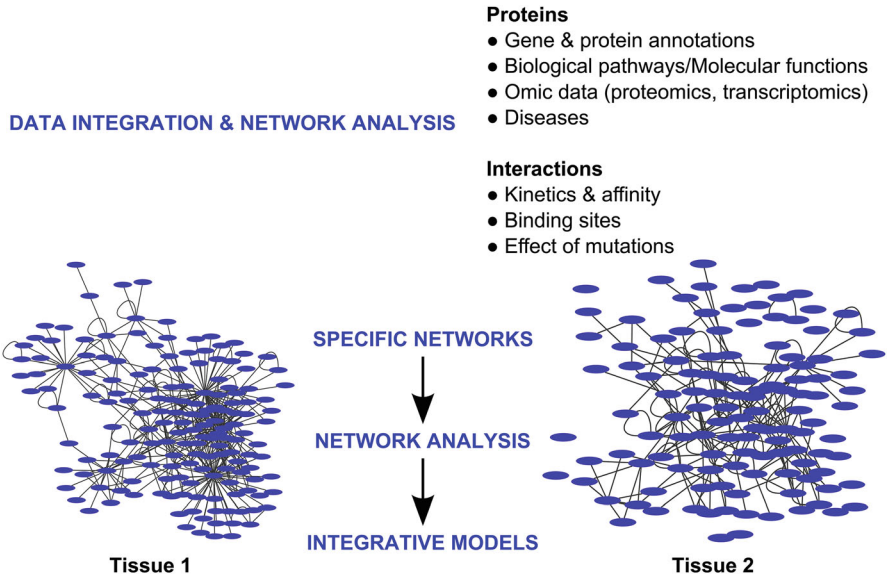

Fig. 6.2 Contextualization of interaction networks: integration of gene/protein annotations and -omic data to build specific interaction networks (e.g. tissue-specific or disease-specific interaction networks), of kinetics and affinity to rank interaction in networks, and of binding sites to discriminate competitive interactions from those occurring simultaneously

UniProtKB database (UniProt Consortium 2019), and annotations from Gene Ontology (The Gene Ontology Consortium 2017) (Chap. 1) can also be integrated in the networks to provide information on the cell/tissue locations, molecular functions and the biological processes ECM proteins and proteoglycans are involved in. The parameters governing interactions (kinetics, affinity and binding sites) can also be added to discriminate transient from stable interactions, rank the interactions within the networks and identify those competing for the same binding site(s) on an extracellular protein or a proteoglycan (Fig. 6.2).

We review here the major methods used to identify and characterize ECM protein, glycosaminoglycan, and proteoglycan interactions with a focus on high-throughput methods able to identify a number of interactions simultaneously such as yeast two hybrid assays, ECM protein and GAG arrays, and affinity purification coupled to mass spectrometry (MS). We also discuss the use of large experimental interaction datasets publicly available, and of interaction databases to retrieve interaction data required to build interaction networks. Text-mining based computational approaches to extract protein-protein interactions from biomedical literature (Papanikolaou et al. 2015; Yu et al. 2018) and the prediction of protein-protein interactions (Kotlyar et al. 2017) are both beyond the scope of this chapter. This review also provides selected examples of interaction networks of individual ECM protein, glycosaminoglycan or proteoglycans, ECM families (e.g. syndecans), and of

ECM networks contextualized by integrating expression or proteomic data, which are specific of a particular biological process or disease. We illustrate the usefulness of these networks and how the data generated from their functional and structural analyses can be used to decipher molecular mechanisms of biological processes such as ECM proteostasis (Doan et al. 2019) and diseases (Bonnans et al. 2014; Theocharis et al. 2019), to design further functional experiments, and to select ECM proteins or GAGs and/or their biomolecular interactions as therapeutic targets. The perspectives offered by the collection of interaction data at large scale, and by the building of interactomes are discussed in the conclusion, together with the ultimate goal to build three-dimensional ECM networks, integrating 3D structure of individual ECM molecules and their complexes, which will be the first step towards 3D models of the extracellular matrix.

6.2 Experimental Identification of ECM Protein and Proteoglycan Interactions

Experimental methods for interactome mapping together with computational methods to predict and map interaction networks have been recently reviewed (Snider et al. 2015). To draw relevant conclusions from the analysis of an interaction network it is important to get an extensive coverage of the interactions, and thus to use high-throughput methods to identify biomolecular interactions. The human ECM proteome, defined as the matrisome (Naba et al. 2016) is small (about 1100 proteins) but its specific features make the identification of ECM interactions with certain high throughput methods challenging. Indeed, the major proteins of the ECM are multimers. The 28 collagen types and the laminin isoforms are trimeric, fibronectin is a dimer, the matricellular proteins thrombospondins and tenascins are multimeric, and several ECM receptors (e.g. discoidin-domain receptors) or coreceptors (syndecans) are able to dimerize. Furthermore, ECM proteins and proteoglycans undergo proteolytic cleavages leading to the release in vivo of bioactive protein fragments called matricryptins or matrikines (Monboisse et al. 2014; Ricard-Blum and Vallet 2016a, 2019), which are biological regulators and potential drugs and exert their activities by interacting with several receptors (Ricard-Blum and Vallet 2016b). Gene-centric methods such as two-hybrid approaches are thus not appropriate to identify interactions established by multimeric proteins, ECM bioactive fragments which have interaction repertoires of their own, and glycosaminoglycans (GAG) chains, which are either covalently attached to the protein cores of proteoglycans (heparin, heparan sulfate, chondroitin sulfate, dermatan sulfate and keratan sulfate) or associate in a non-covalent manner to proteoglycans and link proteins to form proteoglycan aggregates (hyaluronan).

Most ECM interactions have been identified in small- or medium-scale studies, and large-scale approaches to screen binary ECM protein or GAG pairs are still underused. A high-throughput screening procedure called AVEXIS (avidity-based

extracellular interaction screen) has been developed to identify weak extracellular receptor-ligand interactions (t(1/2) $\leq$ 0.1 s), and used to investigate signaling pathways in early vertebrate (zebrafish) development mediated by the immunoglobulin superfamily and leucine-rich repeat families (Bushell et al. 2008). This method has been extended to other domain families and more secreted proteins, leading to a large screen including 16,544 potential interactions in zebrafish, resulting in the identification of 96 novel interactions and of the first extracellular ligands for 15 proteins (Martin et al. 2010). AVEXIS has also been used to identify extracellular host-pathogen interactions (Martinez-Martin 2017). It should be noted that this approach is not appropriate to compare the strength of the interactions because it is based on the artificial pentamerization of the prey by a peptide from the cartilage oligomeric matrix protein (Kerr and Wright 2012), which also prevents the use of this assay to explore the interactions established by multimeric ECM proteins.

6.2.1 Yeast Two-Hybrid Assays

This method has allowed the identification of the binding partners of ECM domains such as the NC1 domain of collagen VII, a globular domain of the laminin β3 chain (Aho and Uitto 1998), the N- and C-terminal domains of collagen VI chains (Kuo et al. 1997), the N-propeptide of the procollagen α1chain of collagen V (12 partners) (Symoens et al. 2011), and of ECM proteins such as thrombospondin-1 (Aho and Uitto 1998), microfibril-associated glycoprotein-2 (Nehring et al. 2005), the extracellular matrix protein-1 (ECM-1) (Fujimoto et al. 2006), and tuftelin, an enamel matrix protein (Paine et al. 1998). Interacting sequences can also be identified by this approach as shown for thrombopondin-1 (Aho and Uitto 1998) or ECM-1 (Fujimoto et al. 2006). Among the interaction detected via two-hybrid assays, some have been selected for functional characterization. The interactions between lysyl oxidase (LOX) and a disintegrin and metalloproteinase with thrombospondin motifs (ADAMTS) families have been first reported in a yeast two-hybrid screen performed with prolysyl oxidase as a bait and a human placenta library (Aviram et al. 2019). Laminins and fibronectin have been also identified as putative binding partners of human lysyl oxidase by a yeast two-hybrid assay (Kraft-Sheleg et al. 2016).

6.2.2 ECM Protein, Peptide and Glycosaminoglycan Arrays

We have designed ECM protein and glycosaminoglycan arrays to identify their binding partners by spotting them onto a gold affinity chip. They are physically adsorbed on the gold surface, and their potential protein partners are recirculated on the chip surface for 20–30 min for proteins and proteoglycans (Faye et al. 2009; Salza et al. 2014; Vallet et al. 2018) up to 120 min for intact parasites (Fatoux-Ardore et al. 2014), and probed by surface plasmon resonance imaging. This

approach allowed us to detect new partners of endostatin, a matricryptin of collagen XVIII (Faye et al. 2009), recombinant collagen V homotrimer pro[α1(V)]$_3$ (Bonod-Bidaud et al. 2012), procollagen C-proteinase enhancer-1 (PCPE-1) (Salza et al. 2014), the propeptide of prolysyl oxidase (Vallet et al. 2018), and to investigate the interactions of the ECM with the amyloid Aβ42 peptide, Aβ42 oligomers and fibrils (Salza et al. 2017), and parasite using intact, living, *Leishmania* parasites (Fatoux-Ardore et al. 2014). ECM proteins, GAGs and controls are spotted in triplicates and about 130 biomolecule pairs can be screened in parallel (Faye et al. 2009). The spotting of individual domains of ECM components on the arrays has been used to identify those encompassing binding sites (Bonod-Bidaud et al. 2012; Salza et al. 2014).

Libraries of 56 and 57 triple-helical peptides spanning the length of the triple helix of collagens II and III helices have been synthesized to set up solid-phase binding assays (Farndale 2019). These Collagen Toolkits allow to locate sites where collagen receptors and ECM components bind to collagens II and III. The use of Collagen II toolkit has shown that the binding of 30 proteins to 170 sites on collagen II is not random, but displays a periodicity of approximately 28 nm, and that a single site close to the collagenase-cleavage site interacts with 20 proteins (Farndale 2019).

Arrays of natural or synthetic GAGs, reviewed in (Karamanos et al. 2018) have been developed to identify GAG-binding proteins. They include neoglycolipid technology-based microarrays (Li and Feizi 2018), allyl amine coated microplates, which have been used to characterize the binding of tumor necrosis factor-inducible gene 6 protein, complement factor H, fibrillin-1, and versican to GAGs (Marson et al. 2009), and plastic microarray with GAG trapped by their reducing end, used to characterize GAG interactions with growth factors and anti-glycosaminoglycan antibodies (Takada et al. 2013). Poly-L-lysine-coated slides have been used to capture GAGs and make GAG arrays for the rapid detection of GAG-protein interactions (Rogers and Hsieh-Wilson 2012).

Commercially available protein arrays comprising 8268 human proteins (Protoarray version 4, Invitrogen) have been used to identify binding partners of keratan sulfate (KS), chondroitin sulfate A (CSA), and hyaluronan (HA) (Conrad et al. 2010). The binding of biotinylated GAGs to spotted proteins has been detected by streptavidin conjugated to a fluorophore. 217 KS-binding proteins, including 75 kinases, 24 CSA-binding proteins and 6 HA-binding proteins have been identified by this method (Conrad et al. 2010). Novel GAG-binding proteins have been identified using *E. coli* proteome chips made of 4300 *E. coli* proteins leading to the identification of 185 heparin-, 62 HS-, 98 chondroitin sulfate B (CSB)—and 101 chondroitin sulfate C (CSC)-interacting proteins, and of YcbS as a novel bacterial virulence factor (Hsiao et al. 2016).

6.2.3 *Affinity Purification: Mass Spectrometry (AP-MS)*

147 heparin/heparan sulfate-binding proteins from rat liver plasma membranes have been captured by affinity chromatography on a heparin column, and identified by shot-gun mass spectrometry (Ori et al. 2011). Interactions promoting elastic fiber supramolecular assembly and those mediating its integration in the ECM were investigated by affinity capture and mass spectrometry (Cain et al. 2009). The bait-prey complexes were purified by affinity chromatography from conditioned serum-free media and solubilized ECM via the tag added to the bait proteins and 112 prey proteins interacting with seven baits including MAGP-1, fibulin-5, lysyl oxidase, fibrillin-1 and collagen VIII were identified by MS/MS (Cain et al. 2009).

Mass spectrometry-based proteomics has been also used to define the intracellular interactome of collagen I (Doan et al. 2019). This approach is based on the inducible expression of collagen α1 and α2 chains bearing different tags in HT-1080 human cells, followed by covalent cross-linking by dithiobis(succinimidyl propionate) to stabilize transient interactions, cell lysis, co-immunoprecipitation with anti-tag antibodies and quantitative MS proteomics using iTRAQ (Isobaric Tagging Reagents for Quantitative Proteomic Analysis) or tandem-mass tags (Doan et al. 2019). This approach, which could be used to build the intracellular interactomes of other collagen types, allowed the identification of about 30 new partners of collagen I (Doan et al. 2019).

AP-MS is not an appropriate method to identify binding partners of native ECM proteins when they are oligomeric. Furthermore, the interactions detected using this approach may be direct or indirect (i.e. involving a bridging molecule), which should be indicated in the interaction maps as it has been done for the interactomes of elastic fibers (Shin and Yanagisawa 2019), syndecans 1 and 4 (Roper et al. 2012) and the four human syndecans (Gondelaud and Ricard-Blum 2019).

6.2.4 *Interaction Databases and Large Interaction Datasets*

Interaction data issued from manual curation of the literature, publicly available interaction databases and/or large interaction datasets can be added to interaction networks to increase their coverage. The Biological General Repository for Interaction Datasets (BioGRID) contains protein, genetic, and chemical interactions, and records for >700,000 post-translational modification sites (Oughtred et al. 2019). A number of ECM networks includes data from BioGRID (Rivera et al. 2011; Zhan et al. 2017; Izzi et al. 2019). The Search Tool for the Retrieval of Interacting Genes (STRING) database includes physical, functional interactions, experimental and predicted interactions (Szklarczyk et al. 2017). STRING data have been integrated into protein-protein interaction networks of the proteoglycan serglycin (Korpetinou et al. 2014), heparin/heparan sulfate-binding proteins (Meneghetti et al. 2015), ECM proteins involved in models of systemic inflammation (Bhan et al. 2019), and

chondroitin sulfate effects on cartilage ECM proteins (Calamia et al. 2012). They have also been used to build and analyze the network of adhesome and degradome proteins in cancer (Rizwan et al. 2015), the ECM interaction network in colorectal cancer (Qi and Ding 2018), lysyl oxidase-like 2 (LOXL2), actin, cytoplasmic 1 (ACTB) and actin, cytoplasmic 2 (ACTG1) interactome in esophageal squamous cell carcinoma (Zhan et al. 2017), and the protein-protein interaction network of genes differentially expressed in tumor cells during the angiogenic response (Yang et al. 2019). BioGRID and STRING databases have also been used to build the mammalian interactome of lysyl oxidase including 70 partners and 210 protein-protein interactions (Okkelman et al. 2014).

MatrixDB, the interaction database we have developed (Chautard et al. 2009a, 2011; Launay et al. 2015; Clerc et al. 2019) (Chap. 1) is focused on the extracellular matrix, and takes into account its specific features for the manual curation process, namely the oligomeric nature of a number of ECM proteins, the existence of ECM bioactive fragments (matricryptins/matrikines) released upon ECM remodeling, and the presence of glycosaminoglycans. MatrixDB is one of the very few databases, if not the only one, to report glycosaminoglycan-protein interactions (Clerc et al. 2019). MatrixDB is a member of the International Molecular Exchange consortium (IMEX, https://www.imexconsortium.org/), which provides a non-redundant set of physical molecular interaction data from a broad taxonomic range of organisms through manual curation of the literature (Orchard et al. 2012). MatrixDB data have been used to build interactomes of fibrillar collagens I, II and III (An et al. 2016; An and Brodsky 2016), procollagen C-proteinase enhancer-1 (PCPE-1) (Salza et al. 2014), the propeptide of lysyl oxidase (Vallet et al. 2018), GAGs (Vallet et al. 2017; Vallet et al. 2020), heparin/heparan sulfate (Ori et al. 2011; Karamanos et al. 2018), decorin (Gubbiotti et al. 2016), 30 proteoglycans (Peysselon et al. 2012), glomerular ECM (Lennon et al. 2014), and other basement membranes (Clerc et al. 2019). In addition, large, publicly available, interaction datasets can be queried to retrieve interactions involving ECM proteins. We used the BioPlex 2.0 dataset generated by affinity purification-mass spectrometry (AP-MS) and comprising more than 56,000 candidate interactions (Huttlin et al. 2017) to identify further binding partners of the four human syndecans (Gondelaud and Ricard-Blum 2019). A new BioPlex dataset collected from human embryonic kidney 293 T cells, BioPlex 3.0, comprises 118,162 interactions among 14,586 proteins (Huttlin et al. 2020). Biomolecular interactions being context-dependent, the AP-MS approach has been used to profile interactions in a human colon cancer cell line, HCT116, resulting in a network comprising 71,000 interactions among 10,531 proteins (Huttlin et al. 2020) (https://bioplex.hms.harvard.edu/). Other large interaction datasets include the reference map of the human binary protein interactome (HuRI) comprised of ~53,000 protein–protein interactions (Luck et al. 2020).

Ingenuity Pathway Analysis (IPA) includes a knowledgebase and provides tools to analyze omic data (https://www.qiagenbio informatics.com/products/ingenuity-pathway-analysis). IPA has been used to study the molecular network of elastic fiber formation (Shin and Yanagisawa 2019), the network of the 15 ECM proteins, which are differentially abundant in collagen IX knock-out and wild-type cartilage with a

focus on the network of the collagens involved (Brachvogel et al. 2013), the remodelling of cellular microenvironment due to loss of collagen VII (Küttner et al. 2013), and the ECM network in breast cancer (Angel et al. 2019), and cancer progression (Reuter et al. 2009). IPA has also been used to investigate ECM remodeling and matrikine signals downstream of Cx43/matrix metalloproteinase-3 (MMP-3)/osteopontin in glioblastoma multiforme (Aftab et al. 2019), and gene interaction networks during the embryo-to-hatchling transition in chicken (Cogburn et al. 2018).

6.3 Extracellular Matrix Interaction Networks

Interaction networks can be built from a list of proteins of interest generated by the users or from experimental data, namely proteins expressed by over-expressed and/or down-regulated genes in development (Martin et al. 2010), or in diseases such as recessive dystrophic epidermolyis bullosa (Küttner et al. 2013) and fibrosis (Ricard-Blum and Miele 2019). The networks provide insights into the structural and functional cross-talk between proteins and glycosaminoglycans/proteoglycans within the ECM, between the ECM and the cell surface in the pericellular matrix. Signaling pathways involved in zebrafish development have been identified by constructing an extracellular interactome containing 92 proteins and 188 extracellular interactions between receptors belonging to the immunoglobulin and leucine-rich repeat protein families (Martin et al. 2010). The connectivity of this network follows a power law, and the secreted proteins are twice more connected than membrane receptors

The interaction networks can be visualized as 1D or 2D maps, where molecules are visualized by symbols of different color and shapes corresponding to specific features (e.g. their location, their molecular function, the biological process they are involved in, the diseases they are associated with, their up-or down regulation in a pathological situation such as cancer (Reuter et al. 2009), their relative abundance measured by quantitative MS (Küttner et al. 2013) or phosphorylation status as shown for the fibronectin-induced phospho-adhesome (Robertson et al. 2015). The interactions can be visualized with various types of lines as shown in few examples below. Lines of different thickness and/or color have been used to reflect the confidence of interaction data (Bhan et al. 2019), the interaction strength (i.e. the affinity of the interactions) (Peysselon and Ricard-Blum 2014; Thomson et al. 2019), which can be used to rank interactions within the network. Affinity values have been integrated in the interaction network of a bioactive fragment of collagen XVIII (endostatin) with its partners located at the surface of endothelial cells where it regulates angiogenesis, and in the secondary network comprised of endostatin partners and heparin/heparan sulfate (Peysselon and Ricard-Blum 2014). The values of equilibrium dissociation constants have also been added in the interaction network of fibrillin-1 and tropoelastin with their partners (Thomson et al. 2019). Colored or dotted lines may discriminate direct and indirect interactions (Angel et al. 2019;

Gondelaud and Ricard-Blum 2019), experimentally determined interactions and those retrieved from curated databases (Manninen and Varjosalo 2017), experimentally supported and predicted interactions (Paladin et al. 2015) or the techniques used to identify the interactions (Thomson et al. 2019). The links between proteins may also represent a functional link as in the human protease web where they represent cleavages (Fortelny et al. 2014). For a comprehensive overview of the computational methods to map and visualize interactomes, the basics of protein interaction network mapping and the major interactome resources the readers are referred to a recent review (Snider et al. 2015).

Further information, referred to as data integration, can be added to interaction networks to perform global analyses by interconnecting molecular information layers of a cell or a tissue i.e. molecular and functional networks based on genomics, transcriptomics, proteomics, metabolomics, and phenotypes. The methods for integration of biological data are detailed in a recent review (Gligorijević and Pržulj 2015). Expression data collected by whole zebrafish embryo in situ hybridization at five stages of embryonic development for 164 genes have been integrated in the extracellular protein interaction network to switch from a static interaction network to dynamic tissue- and stage-specific subnetworks (Martin et al. 2010). They are publicly and freely available via the on-line database called ARNIE (AVEXIS Receptor Network with Integrated Expression, www.sanger.ac.uk/arnie). Expression data can also be collected by manual curation of the literature to build protein-protein interaction networks specific of a tissue or a cell type as done for the interaction network of matricryptins with their receptors located at the surface of endothelial and tumor cells (Ricard-Blum and Vallet 2016b). This network highlights the complex cross-talks occurring between matricryptins, some of them being able to interact, and between matricryptins and their receptors. Indeed, a matricryptin may bind to several receptors belonging to different families (e.g. tyrosine kinase receptors and integrins), and a receptor may be activated by several matricryptins (Ricard-Blum and Vallet 2016b).

The combination of omic data and their integrated analysis are a powerful approach to decipher the molecular mechanisms of physiopathological processes. The study of Cx43 secretome and interactome has shown the existence of synergistic networks and mechanisms for glioma migration and MMP-3 activation (Aftab et al. 2019), and in fibrosis (Ricard-Blum and Miele 2019). The interplay between the degradome and the adhesome has been investigated in breast cancer metastasis by building the interaction network of major cell adhesion and degradome proteins differentially expressed in metastatic *versus* non-metastatic breast cancer cell lines, with a focus on matrix metalloproteinases (MMPs), integrins and E-cadherin (Rizwan et al. 2015).

6.3.1 Interaction Networks of Individual ECM Components

The interactions of several ECM proteins and proteoglycans such as thrombospondin-1 (Resovi et al. 2014) and the small leucine-rich proteoglycan decorin (Gubbiotti et al. 2016) have been reviewed to provide the readers with an updated list of their partners. The distribution of the ~30 partners of human collagen III on its triple helix has led to the identification of a cell interaction domain, a fibrillogenesis and enzyme cleavage domain, three major ligand-binding regions, and hemostasis domains (Parkin et al. 2017). Known and predicted ligand-binding sites have been mapped onto the collagen IV heterotrimers (α1α1α2, α3α4α5, and α5α5α6), leading to the definition of functional (angiogenesis and haemostasis), and disease (autoimmunity, infection, glycation, tumor growth and inhibition) domains (Parkin et al. 2011).

Building and analyzing the interaction network of a single ECM protein, proteoglycan, matricryptin or glycosaminoglycan also give new insights on its molecular functions and biological roles. The study of the interactome of the tissue inhibitor of metalloproteinase-1 (TIMP-1) has highlighted its interrelated functions as a protease inhibitor and signaling molecule associated with its two-domain structure (Grünwald et al. 2019). The domain composition of the partners of a protein may be used to design binding assays in order to identify additional binding partners, functions or biological process the protein is involved in. For example, a first series of experiments has identified partners of the ECM protein procollagen C-proteinase enhancer-1 containing an epidermal growth factor (EGF) domain (Salza et al. 2014). This prompted us to screen other EGF-containing proteins for their ability to interact with PCPE-1, which allowed us to identify further partners of this protein (Salza et al. 2014). Computational analysis of Gene Ontology (GO) terms annotating the partners of a protein may provide clues on its function. The analysis of GO terms associated with PCPE-1 partners showed their involvement in tumor growth, neurodegenerative diseases and angiogenesis. Then we analyzed the effect of two biological fragments of PCPE-1 (the netrin and the CUB1-CUB2 domains) on angiogenesis in vitro, and demonstrated that the CUB1-CUB2 domain inhibits the formation of tubes/angiogenesis in vitro (Salza et al. 2014). Domains mediating interactions may be visualized in interaction maps to highlight their roles as done for the PDZ domain in the interactomes of syndecans 1 and 4 where it mediates interactions with intracellular partners of syndecans (Roper et al. 2012). The presence of EGF in the interactome of the propeptide of lysyl oxidase suggests that the propeptide might regulate EGF-mediated signaling, and its binding to two cross-linking enzymes, transglutaminase-2 and LOXL2, might affect ECM cross-linking and/or the activities of these enzymes (Vallet et al. 2018).

The first draft of the interactome of endostatin, an anti-angiogenic, anti-tumoral and anti-fibrotic matricryptin of collagen XVIII comprised glycosaminoglycans, matricellular proteins (thrombospondin-1 and SPARC), collagens (I, IV, and VI), transglutaminase-2, and the Aβ42 amyloid peptide, which forms amyloid fibrils in patients with Alzheimer's disease (AD) (Faye et al. 2009). This last interaction led us

to explore the role of endostatin in Alzheimer's disease. We demonstrated the presence of endostatin in the cerebrospinal fluid of patients with Alzheimer's disease and other dementia, and reported that the calculation of its ratio relative to well-established AD markers improves the diagnosis of behavioral frontotemporal dementia and the discrimination of patients with Alzheimer's disease from those with other dementia (Salza et al. 2015).

The search for binding partners of a protein might be restricted to a specific compartment. The intracellular partners of collagen I have been identified in HT 1080 human cells to understand the molecular mechanisms of its intracellular folding, processing, assembly, and quality control leading to the building the collagen I proteostasis network comprising 48 proteins, and the discovery of apparent aspartyl hydroxylation as a post-translational modification of the N-propeptide of collagen I (DiChiara et al. 2016).

The interactomes of several glycosaminoglycans have been drafted. The interaction network of heparin/heparan sulfate comprises 435 proteins identified experimentally and by literature curation (Ori et al. 2011), and the interactome of keratan sulphate 217 proteins including 75 kinases, numerous cytoskeletal and nerve function proteins, and several membrane or secreted proteins (Conrad et al. 2010). Fifteen partners of the keratan sulfate proteoglycan mimecan/osteoglycin have been identified too (Conrad et al. 2010). In addition, several protein-protein interaction networks of heparin/heparan sulfate-binding proteins have built, including those of anti-thrombin and fibroblast growth factor-1 (Meneghetti et al. 2015), and of the 435 heparin-binding proteins identified by (Ori et al. 2011).

6.3.2 Interactomes of ECM Protein, Glycosaminoglycan or Proteoglycan Families

The interaction network of the major fibrillar collagens types (I, II and III) has been built and their partners divided into several categories, namely matrix proteins, receptors, enzymes/chaperones, and GAG/proteoglycans (An et al. 2016).

A tentative interaction network of five out of six mammalian GAGs, chondroitin sulfate, dermatan sulfate, heparan sulfate, heparin and hyaluronan, has been created using MatrixDB interaction data (Vallet et al. 2017). A number of proteins interact with several GAGs, connecting their interactomes. Heparin has the highest number of specific partners, whereas all DS-binding proteins are able to also interact with other GAGs (Vallet et al. 2017, Vallet et al. 2020). We have recently built the first comprehensive draft of the GAG interactome comprised of 832 biomolecules (827 proteins and 5 GAGs) and 932 protein-GAG interactions (Vallet et al. 2020) and involved in ECM organization, signal transduction, hemostasis and development (Vallet et al. 2020). However, there might be a bias in the number of proteins specifically assayed for their GAG-binding properties, and in the most studied GAGs, which are by far heparin and heparan sulfate. This emphasizes the

need to design glycosaminoglycome- and proteome-wide assays to increase interactome coverage allowing to draw firm conclusions from interactome analyses.

The interaction network of 30 proteoglycans, comprised of 209 molecules and 557 interactions, has been built to investigate the role of intrinsic disorder (Peysselon et al. 2012). Several proteoglycans connected to more than 10 partners, mostly hyalectans (e.g. aggrecan, versican, and brevican), are enriched (≥30%) in intrinsic disorder (Peysselon et al. 2012). The literature-curated interactomes of syndecans 1 and 4 partially overlap with the ones of integrins with which they cooperate to promote cell adhesion (Roper et al. 2012). The first global interaction network of the four syndecans comprising 351 partners (131, 103, 84 and 33 for syndecans 1, 2, 4 and 3) has been completed (Gondelaud and Ricard-Blum 2019). Beside the expected role of the syndecan network in signal transduction and cell communication, the functional analysis of the syndecan interactome has shown that about 40% of the partners of syndecans 1 and 4 are associated with exosomes, which play a role in cancer development (Gondelaud and Ricard-Blum 2019). The consensus syndecan interactome (i.e. the partners shared by the four syndecans), corresponding to their canonical functions, includes 14 proteins, 12 of them being intracellular or membrane proteins, namely protein kinases, proteins involved in signaling, actin cytoskeleton organization and microtubule formation (Gondelaud and Ricard-Blum 2019).

6.3.3 Networks of ECM Supramolecular Assemblies and Basement Membranes

Some networks take into account the oligomeric structure of native ECM proteins but the next step will be to look for interactions established by the ECM supramolecular assemblies such as collagen fibrils, beaded filaments, anchoring fibrils or hexagonal networks (Ricard-Blum 2011) and elastic fibers (Kozel and Mecham 2019).

Collagen I has more than 100 partners. A 2-dimensional map of the partner-binding sites and disease-associated mutations of human collagen I fibril D-period has been drafted in 2002. It has shown the existence of hot spots for partner interactions and the location of most of the binding sites in its C-terminal half. Mutations associated with genetic diseases showed non-random distribution patterns within the fibril (Di Lullo et al. 2002). The distribution of functional sites and mutations suggests that the collagen fibril is organized into two domains, the cell interaction domain and the matrix interaction domain mediating collagen cross-linking, proteoglycan interactions, and tissue mineralization. Functional sites have been superimposed from the 2D fibril map onto a 3D X-ray structure of the collagen microfibril by molecular modeling revealing the existence of domains in the native fibril (Sweeney et al. 2008; Orgel et al. 2011). Binding site accessibility have been analyzed to determine how the partners of fibrils access their buried recognition

motifs. The internal dynamics of collagen fibrils might regulate their interactions with their partners (Hoop et al. 2017).

Interactions between major constituents of elastic fibers (fibrillin-1, MAGP-1, fibulin-5, and lysyl oxidase) have been identified together with secondary interacting networks with fibronectin and heparan sulfate-associated molecules, which contribute to the formation of microfibrils (Cain et al. 2009). The interactomes of fibrillin-1 and tropoelastin are connected (Thomson et al. 2019). The interaction network among core elastogenic genes/gene products, namely elastin, fibrillin-1, fibulin-4, fibulin-5 and LTBP4, and an extended interaction network among elastic fiber-associated genes/gene products comprised of 40 direct interactions with additional indirect interactions have been built (Shin and Yanagisawa 2019).

Basement membranes are thin ECM sheets mostly comprised of collagen IV and laminins supramolecular assemblies. Jürgen Engel wrote in 2007 an article entitled "Vision for novel biophysical elucidations of extracellular matrix networks" (Engel 2007). He concluded his review with a dream about two papers published in 2017. One of his dreams was the building of a schematic representation "for a specific region of a basement membrane in tissue at atomic resolution together with the dynamics of the various interactions, showing a concerted action of several components". Part of his dream becomes true thanks to the proteomic analysis of the glomerular extracellular matrix, which identified 144 structural and regulatory ECM proteins, and the building of the glomerular ECM interactome using several data sources (Lennon et al. 2014). The glomerular ECM interactome contains a core of highly connected structural components formed by basement membrane proteins and structural ECM proteins. They are involved in more interactions than ECM-associated proteins as shown by the distribution of the number of interactions per protein in the interactome (Lennon et al. 2014). We have built a global human basement membrane interactome using interaction data stored in MatrixDB database, and its advanced search tool (Clerc et al. 2019). Quantitative proteomic data generated from various basement membranes by Naba et al. (Naba et al. 2016) has been integrated to get interactomes specific of a basement membrane (e.g. human glomerular, retinal vascular, and lens capsule basement membranes). This allowed us to define a consensus interactome of basement membranes including proteins and interactions common to all the basement membranes studied, and to identify those, which are restricted to a given basement membrane (Clerc et al. 2019). The above examples illustrate the interest of combining omic data, namely proteomics and interactomics in this study.

6.3.4 Networks Associated with ECM Degradation and ECM-Cell Interactions

ECM undergoes remodeling in health and diseases (Bonnans et al. 2014) thanks to an interconnected network of proteases and their inhibitors, the protease degradome

(Kappelhoff et al. 2017). Several proteases act on the same substrate(s), and proteases interact with each other and may regulate each other forming the protease web, which comprises 340 proteins and 1264 interactions in human (Overall and Kleifeld 2006; Fortelny et al. 2014) and Chap. 7.

ECM-cell interaction networks or adhesomes have been extensively characterized as summarized below, and detailed in Chap. 9. The binding of integrins to ECM proteins triggers intracellular signaling pathways that regulate cell functions via the formation of integrin adhesion complexes. These large, dynamic, multiprotein complexes referred to as the integrin adhesome link the ECM to the cytoskeleton (Zaidel-Bar et al. 2007; Winograd-Katz et al. 2014; Horton et al. 2016b; Manninen and Varjosalo 2017). The current version of the network includes 150 components and 64 interactions but the integration of 82 associated components increases the number of interactions 6542 interactions (the adhesome: a focal adhesion network, www.adhesome.org).

Proteomic analysis of integrin adhesomes from different cell types enabled the building of a meta-adhesome of 2412 proteins (Horton et al. 2016b). A consensus integrin adhesome of 60 proteins, comprising the core cell adhesion machinery, and four axes have been defined, namely integrin-linked protein kinase (ILK)—Particularly interesting new Cys-His protein 1 (PINCH)—kindlin, focal adhesion kinase (FAK)—paxillin, talin—vinculin and α-actinin—zyxin—vasodilator-stimulated phosphoprotein (VASP) (Horton et al. 2015, 2016b). Knockdown of consensus components core components FAK, ILK, TLN1 and ZYX showed enlarged focal adhesions in MCF7 breast cancer cells (Fokkelman et al. 2016). ECM, integrins and their associated adhesion complexes regulate the mechanosignalling pathways. The role of the consensus adhesome in mechanotransduction, which affects cell differentiation and proliferation has been explored (Horton et al. 2016a)). The phosphoadhesome induced by fibronectin has been characterized by MS-based proteomic and phosphoproteomic analysis of adhesion complexes (Robertson et al. 2015, 2017). The desmo-adhesome (desmosome network) of keratinocytes, comprised of 59 proteins (30 intrinsic and 29 accessory components) and 128 direct interactions, is robust against random perturbations but susceptible to targeted attacks (Cirillo and Prime 2009). The desmosome components targeted in human disease, and the integration in this network of non-desmosomal regulatory proteins provide new insights in desmosomal diseases and associated pathways (Celentano et al. 2017). Integrin adhesion complexes have been extensively characterized but other receptors such as discoidin-domain receptors 1 and 2 (Leitinger and Saltel 2018), and syndecans (Couchman et al. 2015) contribute to cell adhesion and cell-ECM interactions. Proteomic analyses of the adhesion complexes associated with these receptors are needed to get a comprehensive overview of global cell adhesomes.

6.3.5 *ECM Interaction Networks Associated with Physiological and Pathological Processes*

Building and analysis of interaction networks are useful to decipher the molecular mechanisms of diseases as recently reviewed for fibrosis (Ricard-Blum and Miele 2019), to identify therapeutic targets and/or o design therapeutic molecules. The integrin interactome has been successfully targeted to control immune diseases whereas this approach failed in cancer (Vicente-Manzanares and Sánchez-Madrid 2018).

6.3.5.1 Development and Aging

Protein-protein interaction networks involving ECM components have been created and analyzed for physiological processes such as development in chicken (Cogburn et al. 2018) and aging (Chautard et al. 2010). The investigation of chicken development by transcriptional profiling of liver during the embryo-to-hatchling transition period has led to the identification interactions of blood clotting factors and von Willebrand factor with several collagen genes coding for the α1 chains of collagens I, III and XVIII, fibulin 1–2, SPARC, laminin subunits β1 and β3 and lysyl oxidase (Cogburn et al. 2018).

We have shown that "response to wounding" and "tissue regeneration" are overrepresented biological processes in the ECM network associated with aging, and that calcium and protein binding are overrepresented molecular functions in the aging ECM network. Furthermore, heparin is one the two most connected ECM component in this network (Chautard et al. 2010). Given that intrinsically disordered proteins are associated with age-related diseases, and that the extracellular proteome is significantly enriched in disorder compared to the entire human proteome (Peysselon et al. 2011), the role of disorder in ECM aging warrants further investigation (Peysselon and Ricard-Blum 2011).

6.3.5.2 Cancer and Angiogenesis

The cancer matrisome has been characterized in different tumor types and microenvironmental niches, leading to the identification of protein signatures discriminating tumors from healthy tissues, different tumor stages, and primary from secondary tumors (Socovich and Naba 2019). The pan-cancer analysis of the expression and regulation of 820 matrisome genes across 32 tumor types in more than 10,000 patients has identified 919 transcriptional modules differentially governing the matrisome components and 40 master regulators governing the matrisome of each cancer studied (Izzi et al. 2019).

An ECM interaction network of the physical, enzymatic, and transcriptional interactions connecting 282 biomolecules involved in cancer progression has been

identified. Its topology predicts that tumor development depends on specific ECM-interacting network hubs such as β1 integrin, and it is disrupted by the blockade of β1integrin, which reduces tumorigenesis (Reuter et al. 2009). An esophageal squamous cell carcinoma (ESCC)-specific protein-protein interactome involving LOXL2 and the actin-related proteins, actin β and actin γ1, has been generated. It comprises 362 proteins (the 3 above proteins, 14 core proteins, and 345 common proteins) and 2580 protein-protein interactions (Zhan et al. 2017). A three-gene signature, including LOXL2, cadherin-1 and fibronectin, has been shown to predict a poor clinical outcome in ESCC patients (Zhan et al. 2017).

The interaction network of 38 proteins identified by LC–MS/MS after collagenase 3 digestion of breast cancer tissue includes 34 ECM proteins, 13 of which are collagens (Angel et al. 2019). These proteins are strongly associated with breast cancer progression. FAS (tumor necrosis factor receptor superfamily member 6), AHR (aryl hydrocarbon receptor), Brd4 (bromodomain-containing protein 4), SPDEF (SAM pointed domain-containing Ets transcription factor), IGFBP2 (insulin-like growth factor-binding protein 2), TP53 (cellular tumor antigen p53), COLQ (collagen Q or acetylcholinesterase collagenic tail peptide), and TGFB1 (transforming growth factor β-1 proprotein) as among the main potential regulators of this network (Angel et al. 2019). Protein-protein interaction network of differentially expressed genes in the ECM has been constructed to investigate changes occurring in signaling molecules during the metastasis of colorectal cancer. Fibronectin and MMP-2 are the most connected proteins of this network (18 partners), the key regulatory pathway for extracellular signal transmission was Fibronectin → Secreted Protein Acidic Rich in Cysteine → COL1A1 → MMP-2, and the key ECM complex is comprised of vascular endothelial growth factor A and C-C motif chemokine 7/C-C chemokine receptor type 3 (Qi and Ding 2018).

Increased angiogenesis is a hallmark of cancer and of a number of other diseases. Heparin and heparan sulfate interactions with proteins regulating angiogenesis have been compiled to create the "angiogenesis glycomic interactome" comprised of canonical (9), non-canonical growth factors and other regulators (23), proangiogenic receptors (6), angiogenesis inhibitors (14) and effectors (4) (Chiodelli et al. 2015). This is a useful resource to design new therapeutic strategies aiming at angiogenesis regulation and vascular homeostasis. Syndecans 1, 2 and 4, MMP-9, CD44, and versican are at the center of the interaction network of three angiogenesis-associated protein families, namely collagen IV, CXC chemokines and thrombospondin-1-containing proteins (Rivera et al. 2011).

6.3.5.3 Alzheimer's Disease

We have built experimentally the interaction network of ECM proteins and GAGs with the amyloid-β42 peptide (Aβ42) under various molecular and supramolecular forms (i.e. monomeric peptide, oligomers, fibrils and fibrillary aggregates) to identify the molecular functions, biological processes, and pathways targeted by the Aβ42 peptide in the ECM. We have shown that the multimerization state of this

peptide governs its interactions with the ECM. Aβ42 oligomers, but not the monomeric peptide, bind to cell surface receptors. Aβ42 fibrils, but not oligomers interact with GAGs, proteoglycans, enzymes, and growth factors, whereas Aβ42 fibrillar aggregates bind to further membrane proteins (syndecan-4, TEM-8, and discoidin-domain receptor-2) (Salza et al. 2017).

6.3.5.4 Host ECM-Pathogen Interactions

Several high-throughput methods have been used to study so-called extracellular protein interactions between viruses or bacteria and their hosts, but the extracellular proteins included in these studies are i membrane-tethered proteins and secreted molecules but not ECM proteins (Martinez-Martin 2017). However, in addition to cell surface-associated ECM receptors, host ECM is also targeted by a number of pathogens. A single ECM protein, fibronectin, interacts with numerous pathogens including bacteria such as *Staphylococcus aureus*, streptococci, enterococci, mycobacteria, *E. coli*, and *Salmonella enterica* (Vaca et al. 2019). The characterization of host ECM-pathogen interaction repertoires is a prerequisite to understand the molecular mechanisms of infection, and to design new therapeutic strategies to prevent or block it.

The interactions of parasites of the host ECM have been investigated at a large scale for 24 strains and six species of *Leishmania* using protein and glycosaminoglycan arrays probed by surface plasmon resonance imaging (Fatoux-Ardore et al. 2014). About 70 ECM proteins, GAGs, growth factors, and cell surface receptors have been spotted onto Gold affinity chips, and reacted with living parasites as described in paragraph 2.2. We have identified 23 new protein and 4 new GAG partners of procyclic promastigotes and 18 partners (15 proteins, 3 GAGs) of three species of stationary-phase promastigotes. The diversity of *Leishmania*-ECM interactions reflects the dynamic and complex host-parasite interplay, which depends mostly on the species and strains of the parasite. Numerous *Leishmania* strains interacts with tropoelastin and angiogenesis regulators, including antiangiogenic factors (endostatin, anastellin) and proangiogenic factors (ECM-1, VEGF, and tumor-endothelial marker-8). Heparin and heparan sulfate bind to most *Leishmania* strains tested, and 6-O-sulfate groups play a crucial role in these interactions (Fatoux-Ardore et al. 2014).

A number of pathogens bind to GAGs (Aquino and Park 2016). Systematic protein interactome analysis of glycosaminoglycans revealed YcbS as a novel bacterial virulence factor A systematic study on microbial proteome-mammalian GAG interactions has been carried out with *E. coli* proteome chips to probe it interactions with 4 GAGs (HP, HS, CSB, and CSC). 185 heparin-, 62 HS-, 98 CSB-, and 101 CSC-interacting proteins have been identified using this approach, and their analyses has demonstrated the functions of HP- and HS-specific binding proteins in glycine, serine, and threonine metabolism (Hsiao et al. 2016).

Host cell heparan sulfate contribute to the attachment and invasion of *Trypanosoma cruzi* amastigote, which causes Chagas disease (Bambino-Medeiros

et al. 2011). The interaction of *Trypanosoma cruzi* with the host ECM is the first step of infection before the parasite invades cells in different tissues. An ECM-focused interactome of early *Trypanosoma cruzi* infection has been drafted using thrombospondin-1, and galectin-3, which are up-regulated by infective trypomastigotes, laminin γ1, which is upregulated by the parasite trans-sialidase gp83, and ERK1/2 as seed nodes. The interaction network of the laminin γ1 chain is connected to thrombospondin-1 through plasminogen, the α1 chain of collagen VII and MMP-2, suggesting that this network could ease parasite motility (Cardenas et al. 2010). This ECM-centered interactome is regulated by gp83, which also regulates trypanosome usage of the ECM during early infection to enhance cellular infection (Nde et al. 2012). This network, containing 104 protein coding genes connected by 218 biological interactions, has been useful to determine the molecular mechanisms of *T. cruzi* infection, and to identify new therapeutic approaches to treat Chagas disease (Cardenas et al. 2010).

6.3.5.5 Heritable Diseases

Numerous mutations in the genes coding for ECM proteins are associated with heritable diseases such as osteogenesis imperfecta, Ehlers Danlos syndromes (EDS) and epidermolysis bullosa. Regions rich in lethal mutations leading to osteogenesis imperfecta align with binding sites for integrins and proteoglycans in the helical domain of collagen I (Marini et al. 2007), highlighting the interest to include mutations and their effect on interactions in interactomes to predict how they will be rewired in genetic diseases. Collagen V mutations are associated with EDS, and the interactome of its α1 chain has been used to study genotype-phenotype correlations in classic Ehlers-Danlos syndrome (Paladin et al. 2015).

Collagen VII loss occurs in patients with recessive dystrophic epidermolysis bullosa (RDEB). The effect of its loss has been investigated in proteins secreted by fibroblasts from healthy subjects and RDEB patients leading to the identification of 214 proteins significantly altered in the ECM of all RDEB samples. The analysis of the corresponding interaction network made of 64 proteins shows a decrease in basement membrane proteins (laminin α4, β1, β2 and γ1 chains, collagen IV and several integrins) and an increase in collagens III, V and VI, and MMPs although MMP-14 protease activity is reduced in RDEB cells (Küttner et al. 2013).

6.4 Concluding Remarks and Perspectives

A number of ECM networks published so far contain only interactions between protein/proteoglycan(s) or GAG(s) of interest but do not include the connections between the partners, which is a prerequisite to study the architecture and topology of networks and to draw meaningful conclusions from their analysis. Furthermore, networks specific of a biomolecule, a biological process or a disease should be

analyzed in the context of the complete ECM interactome to determine how this molecule, biological process or disease is connected to, and affects, the ECM architecture and function. This emphasize the need to build a complete ECM interactome covering the whole matrisome. Two drafts of the ECM interaction network have been built (Chautard et al. 2009b; Cromar et al. 2012) but their coverage of both the matrisome and its interactions should be increased. To further explore the connections between the ECM, the cell surface and the intracellular compartment it would be also interesting to map the ECM interactome to the most comprehensive experimental map of the human interactome available today, namely the Bioplex 3.0 interactome comprised of 120,000 interactions among nearly 15,000 proteins (Huttlin et al. 2020) (https://bioplex.hms.harvard.edu/), and to the reference map of the human binary protein interactome (Luck et al. 2020).

Another challenge will be to switch from 1D or 2D ECM interaction maps to 3D networks by integrating structural data of individual ECM biomolecules and/or their complexes in the ECM interactome, and disordered regions of ECM proteins, which should provide the network with structural flexibility and interaction rewiring. The building of 3D models of tissue-specific ECM using mesoscopic rigid body modelling (Wong et al. 2018) will be very useful to reach this goal by giving the possibility to map the ECM interactome onto a structural scaffold, and to take into account the strength of ECM interactions, the physical constraints imposed to the ECM by its covalent cross-linking, and the dynamics of conformational changes of ECM components upon binding to their partners, which are increaseimortant in intrinsically disordered regions. The integration in these models of the ECM cross-linkome mediated by the lysyl oxidase family, transgutaminase-2 and peroxidasin (Vallet and Ricard-Blum 2019) will be challenging inasmuch as elastin cross-linking is a stochastic process (Hedtke et al. 2019). A 3D model will be useful to provide the ECM interactome with spatial organization. Another perspective is to integrate in the ECM network splicing isoforms, which might have different interaction repertoires, and mutations, which could decrease or increase the interactions, and rewire the ECM interactome. Mutations of interest will be retrieved from the IMEx consortium mutations data set (IMEx Consortium Curators et al. 2019), and from ECM-specific variant databases (Chap. 1).

A last challenge will be to fully integrate GAGs in the ECM interactomes by adding protein-GAG interactions to protein-protein interactions, the identification of which is favored by the use of gene-centric methods, and to collect quantitative data on GAGs at large scale (glycosaminoglycanomics) (Ricard-Blum and Lisacek 2017), which should be possible thanks to recent developments in GAG analysis in tissues or biological fluids (Chap. 5), and to integrate them in *in-silico* ECM networks to better reflect the in vivo ECM networks.

Acknowledgements The work performed by the authors was supported by the "Fondation de la Recherche Médicale", France [grant number DBI20141231336 to SRB], the Institut Français de Bioinformatique [ANR-11-INBS-0013, Glycomatrix project, call 2015 to SRB], and the Groupement de Recherche (GDR) GagoSciences [CNRS, GDR 3739, Structure, Fonction et Régulation des Glycosaminoglycanes to SRB].

References

Aftab Q, Mesnil M, Ojefua E, Poole A, Noordenbos J, Strale P-O, Sitko C, Le C, Stoynov N, Foster LJ, Sin W-C, Naus CC, Chen VC (2019) Cx43-associated Secretome and Interactome reveal synergistic mechanisms for glioma migration and MMP3 activation. Front Neurosci 13:143. https://doi.org/10.3389/fnins.2019.00143

Aho S, Uitto J (1998) Two-hybrid analysis reveals multiple direct interactions for thrombospondin 1. Matrix Biol 17:401–412. https://doi.org/10.1016/s0945-053x(98)90100-7

An B, Brodsky B (2016) Collagen binding to OSCAR: the odd couple. Blood 127:521–522. https://doi.org/10.1182/blood-2015-12-682476

An B, Lin Y-S, Brodsky B (2016) Collagen interactions: drug design and delivery. Adv Drug Deliv Rev 97:69–84. https://doi.org/10.1016/j.addr.2015.11.013

Angel PM, Schwamborn K, Comte-Walters S, Clift CL, Ball LE, Mehta AS, Drake RR (2019) Extracellular matrix imaging of breast tissue pathologies by MALDI-imaging mass spectrometry. Proteomics Clin Appl 13:e1700152. https://doi.org/10.1002/prca.201700152

Aquino RS, Park PW (2016) Glycosaminoglycans and infection. Front Biosci (Landmark Ed) 21:1260–1277

Aviram R, Zaffryar-Eilot S, Hubmacher D, Grunwald H, Mäki JM, Myllyharju J, Apte SS, Hasson P (2019) Interactions between lysyl oxidases and ADAMTS proteins suggest a novel crosstalk between two extracellular matrix families. Matrix Biol 75–76:114–125. https://doi.org/10.1016/j.matbio.2018.05.003

Bambino-Medeiros R, Oliveira FOR, Calvet CM, Vicente D, Toma L, Krieger MA, Meirelles MN, Pereira MCS (2011) Involvement of host cell heparan sulfate proteoglycan in Trypanosoma cruzi amastigote attachment and invasion. Parasitology 138:593–601. https://doi.org/10.1017/S0031182010001678

Bhan C, Dash SP, Dipankar P, Kumar P, Chakraborty P, Sarangi PP (2019) Investigation of extracellular matrix protein expression dynamics using murine models of systemic inflammation. Inflammation 42:2020–2031. https://doi.org/10.1007/s10753-019-01063-5

Bonnans C, Chou J, Werb Z (2014) Remodelling the extracellular matrix in development and disease. Nat Rev Mol Cell Biol 15:786–801. https://doi.org/10.1038/nrm3904

Bonod-Bidaud C, Roulet M, Hansen U, Elsheikh A, Malbouyres M, Ricard-Blum S, Faye C, Vaganay E, Rousselle P, Ruggiero F (2012) In vivo evidence for a bridging role of a collagen V subtype at the epidermis-dermis interface. J Invest Dermatol 132:1841–1849. https://doi.org/10.1038/jid.2012.56

Brachvogel B, Zaucke F, Dave K, Norris EL, Stermann J, Dayakli M, Koch M, Gorman JJ, Bateman JF, Wilson R (2013) Comparative proteomic analysis of normal and collagen IX null mouse cartilage reveals altered extracellular matrix composition and novel components of the collagen IX interactome. J Biol Chem 288:13481–13492. https://doi.org/10.1074/jbc.M112.444810

Bruckner P (2010) Suprastructures of extracellular matrices: paradigms of functions controlled by aggregates rather than molecules. Cell Tissue Res 339:7–18. https://doi.org/10.1007/s00441-009-0864-0

Bushell KM, Söllner C, Schuster-Boeckler B, Bateman A, Wright GJ (2008) Large-scale screening for novel low-affinity extracellular protein interactions. Genome Res 18:622–630. https://doi.org/10.1101/gr.7187808

Cain SA, McGovern A, Small E, Ward LJ, Baldock C, Shuttleworth A, Kielty CM (2009) Defining elastic fiber interactions by molecular fishing: an affinity purification and mass spectrometry approach. Mol Cell Proteomics 8:2715–2732. https://doi.org/10.1074/mcp.M900008-MCP200

Calamia V, Lourido L, Fernández-Puente P, Mateos J, Rocha B, Montell E, Vergés J, Ruiz-Romero C, Blanco FJ (2012) Secretome analysis of chondroitin sulfate-treated chondrocytes reveals anti-angiogenic, anti-inflammatory and anti-catabolic properties. Arthritis Res Ther 14:R202. https://doi.org/10.1186/ar4040

Cardenas TC, Johnson CA, Pratap S, Nde PN, Furtak V, Kleshchenko YY, Lima MF, Villalta F (2010) REGULATION of the EXTRACELLULAR MATRIX INTERACTOME by Trypanosoma cruzi. Open Parasitol J 4:72–76. https://doi.org/10.2174/1874421401004010072

Celentano A, Mignogna MD, McCullough M, Cirillo N (2017) Pathophysiology of the desmoadhesome. J Cell Physiol 232:496–505. https://doi.org/10.1002/jcp.25515

Chautard E, Ballut L, Thierry-Mieg N, Ricard-Blum S (2009a) MatrixDB, a database focused on extracellular protein-protein and protein-carbohydrate interactions. Bioinformatics 25:690–691. https://doi.org/10.1093/bioinformatics/btp025

Chautard E, Thierry-Mieg N, Ricard-Blum S (2009b) Interaction networks: from protein functions to drug discovery. A review. Pathol Biol (Paris) 57:324–333. https://doi.org/10.1016/j.patbio.2008.10.004

Chautard E, Thierry-Mieg N, Ricard-Blum S (2010) Interaction networks as a tool to investigate the mechanisms of aging. Biogerontology 11:463–473. https://doi.org/10.1007/s10522-010-9268-5

Chautard E, Fatoux-Ardore M, Ballut L, Thierry-Mieg N, Ricard-Blum S (2011) MatrixDB, the extracellular matrix interaction database. Nucleic Acids Res 39:D235–D240. https://doi.org/10.1093/nar/gkq830

Chiodelli P, Bugatti A, Urbinati C, Rusnati M (2015) Heparin/heparan sulfate proteoglycans glycomic interactome in angiogenesis: biological implications and therapeutical use. Molecules 20:6342–6388. https://doi.org/10.3390/molecules20046342

Cirillo N, Prime SS (2009) Desmosomal interactome in keratinocytes: a systems biology approach leading to an understanding of the pathogenesis of skin disease. Cell Mol Life Sci 66:3517–3533. https://doi.org/10.1007/s00018-009-0139-7

Clerc O, Deniaud M, Vallet SD, Naba A, Rivet A, Perez S, Thierry-Mieg N, Ricard-Blum S (2019) MatrixDB: integration of new data with a focus on glycosaminoglycan interactions. Nucleic Acids Res 47:D376–D381. https://doi.org/10.1093/nar/gky1035

Cogburn LA, Trakooljul N, Chen C, Huang H, Wu CH, Carré W, Wang X, White HB (2018) Transcriptional profiling of liver during the critical embryo-to-hatchling transition period in the chicken (*Gallus gallus*). BMC Genomics 19:695. https://doi.org/10.1186/s12864-018-5080-4

Conrad AH, Zhang Y, Tasheva ES, Conrad GW (2010) Proteomic analysis of potential keratan sulfate, chondroitin sulfate a, and hyaluronic acid molecular interactions. Invest Ophthalmol Vis Sci 51:4500–4515. https://doi.org/10.1167/iovs.09-4914

Couchman JR, Gopal S, Lim HC, Nørgaard S, Multhaupt HAB (2015) Fell-Muir lecture: Syndecans: from peripheral coreceptors to mainstream regulators of cell behaviour. Int J Exp Pathol 96:1–10. https://doi.org/10.1111/iep.12112

Cromar GL, Xiong X, Chautard E, Ricard-Blum S, Parkinson J (2012) Toward a systems level view of the ECM and related proteins: a framework for the systematic definition and analysis of biological systems. Proteins 80:1522–1544. https://doi.org/10.1002/prot.24036

Di Lullo GA, Sweeney SM, Korkko J, Ala-Kokko L, San Antonio JD (2002) Mapping the ligand-binding sites and disease-associated mutations on the most abundant protein in the human, type I collagen. J Biol Chem 277:4223–4231. https://doi.org/10.1074/jbc.M110709200

DiChiara AS, Taylor RJ, Wong MY, Doan N-D, Rosario AMD, Shoulders MD (2016) Mapping and exploring the collagen-I Proteostasis network. ACS Chem Biol 11:1408–1421. https://doi.org/10.1021/acschembio.5b01083

Doan N-D, DiChiara AS, Del Rosario AM, Schiavoni RP, Shoulders MD (2019) Mass spectrometry-based proteomics to define intracellular collagen Interactomes. Methods Mol Biol 1944:95–114. https://doi.org/10.1007/978-1-4939-9095-5_7

Engel J (2007) Visions for novel biophysical elucidations of extracellular matrix networks. Int J Biochem Cell Biol 39:311–318. https://doi.org/10.1016/j.biocel.2006.08.003

Farndale RW (2019) Collagen-binding proteins: insights from the collagen toolkits. Essays Biochem 63:337–348. https://doi.org/10.1042/EBC20180070

Fatoux-Ardore M, Peysselon F, Weiss A, Bastien P, Pratlong F, Ricard-Blum S (2014) Large-scale investigation of Leishmania interaction networks with host extracellular matrix by surface plasmon resonance imaging. Infect Immun 82:594–606. https://doi.org/10.1128/IAI.01146-13

Faye C, Chautard E, Olsen BR, Ricard-Blum S (2009) The first draft of the endostatin interaction network. J Biol Chem 284:22041–22047. https://doi.org/10.1074/jbc.M109.002964

Fokkelman M, Balcıoğlu HE, Klip JE, Yan K, Verbeek FJ, EHJ D, van de Water B (2016) Cellular adhesome screen identifies critical modulators of focal adhesion dynamics, cellular traction forces and cell migration behaviour. Sci Rep 6:31707. https://doi.org/10.1038/srep31707

Fortelny N, Cox JH, Kappelhoff R, Starr AE, Lange PF, Pavlidis P, Overall CM (2014) Network analyses reveal pervasive functional regulation between proteases in the human protease web. PLoS Biol 12:e1001869. https://doi.org/10.1371/journal.pbio.1001869

Fujimoto N, Terlizzi J, Aho S, Brittingham R, Fertala A, Oyama N, McGrath JA, Uitto J (2006) Extracellular matrix protein 1 inhibits the activity of matrix metalloproteinase 9 through high-affinity protein/protein interactions. Exp Dermatol 15:300–307. https://doi.org/10.1111/j.0906-6705.2006.00409.x

Gligorijević V, Pržulj N (2015) Methods for biological data integration: perspectives and challenges. J R Soc Interface 12:20150571. https://doi.org/10.1098/rsif.2015.0571

Gondelaud F, Ricard-Blum S (2019) Structures and interactions of syndecans. FEBS J 286:2994–3007. https://doi.org/10.1111/febs.14828

Grünwald B, Schoeps B, Krüger A (2019) Recognizing the molecular multifunctionality and Interactome of TIMP-1. Trends Cell Biol 29:6–19. https://doi.org/10.1016/j.tcb.2018.08.006

Gubbiotti MA, Vallet SD, Ricard-Blum S, Iozzo RV (2016) Decorin interacting network: a comprehensive analysis of decorin-binding partners and their versatile functions. Matrix Biol 55:7–21. https://doi.org/10.1016/j.matbio.2016.09.009

Hedtke T, Schräder CU, Heinz A, Hoehenwarter W, Brinckmann J, Groth T, Schmelzer CEH (2019) A comprehensive map of human elastin cross-linking during elastogenesis. FEBS J 286:3594–3610. https://doi.org/10.1111/febs.14929

Hoop CL, Zhu J, Nunes AM, Case DA, Baum J (2017) Revealing accessibility of cryptic protein binding sites within the functional collagen fibril. Biomol Ther 7:76. https://doi.org/10.3390/biom7040076

Horton ER, Byron A, Askari JA, Ng DHJ, Millon-Frémillon A, Robertson J, Koper EJ, Paul NR, Warwood S, Knight D, Humphries JD, Humphries MJ (2015) Definition of a consensus integrin adhesome and its dynamics during adhesion complex assembly and disassembly. Nat Cell Biol 17:1577–1587. https://doi.org/10.1038/ncb3257

Horton ER, Astudillo P, Humphries MJ, Humphries JD (2016a) Mechanosensitivity of integrin adhesion complexes: role of the consensus adhesome. Exp Cell Res 343:7–13. https://doi.org/10.1016/j.yexcr.2015.10.025

Horton ER, Humphries JD, James J, Jones MC, Askari JA, Humphries MJ (2016b) The integrin adhesome network at a glance. J Cell Sci 129:4159–4163. https://doi.org/10.1242/jcs.192054

Hsiao FS-H, Sutandy FR, Syu G-D, Chen Y-W, Lin J-M, Chen C-S (2016) Systematic protein interactome analysis of glycosaminoglycans revealed YcbS as a novel bacterial virulence factor. Sci Rep 6:28425. https://doi.org/10.1038/srep28425

Huttlin EL, Ting L, Bruckner RJ, Gebreab F, Gygi MP, Szpyt J, Tam S, Zarraga G, Colby G, Baltier K, Dong R, Guarani V, Vaites LP, Ordureau A, Rad R, Erickson BK, Wühr M, Chick J, Zhai B, Kolippakkam D, Mintseris J, Obar RA, Harris T, Artavanis-Tsakonas S, Sowa ME, De Camilli P, Paulo JA, Harper JW, Gygi SP (2015) The BioPlex network: a systematic exploration of the human Interactome. Cell 162:425–440. https://doi.org/10.1016/j.cell.2015.06.043

Huttlin EL, Bruckner RJ, Paulo JA, Cannon JR, Ting L, Baltier K, Colby G, Gebreab F, Gygi MP, Parzen H, Szpyt J, Tam S, Zarraga G, Pontano-Vaites L, Swarup S, White AE, Schweppe DK, Rad R, Erickson BK, Obar RA, Guruharsha KG, Li K, Artavanis-Tsakonas S, Gygi SP, Harper JW (2017) Architecture of the human interactome defines protein communities and disease networks. Nature 545:505–509. https://doi.org/10.1038/nature22366

Huttlin EL, Bruckner RJ, Navarrete-Perea J, Cannon JR, Baltier K, Gebreab F, Gygi MP, Thornock A, Zarraga G, Tam S, Szpyt J, Panov A, Parzen H, Fu S, Golbazi A, Maenpaa E, Stricker K, Thakurta SG, Rad R, Pan J, Nusinow DP, Paulo JA, Schweppe DK, Vaites LP,

Harper JW, Gygi SP (2020) Dual proteome-scale networks reveal cell-specific remodeling of the human interactome. bioRxiv. https://doi.org/10.1101/2020.01.19.905109

IMEx Consortium Curators, Del-Toro N, Duesbury M, Koch M, Perfetto L, Shrivastava A, Ochoa D, Wagih O, Piñero J, Kotlyar M, Pastrello C, Beltrao P, Furlong LI, Jurisica I, Hermjakob H, Orchard S, Porras P (2019) Capturing variation impact on molecular interactions in the IMEx consortium mutations data set. Nat Commun 10:10. https://doi.org/10.1038/s41467-018-07709-6

Iozzo RV, Schaefer L (2015) Proteoglycan form and function: a comprehensive nomenclature of proteoglycans. Matrix Biol 42:11–55. https://doi.org/10.1016/j.matbio.2015.02.003

Izzi V, Juho L, Raman D, Anna K, Jarkko K, Ritva H, Taina P (2019) Pan-Cancer analysis of the expression and regulation of matrisome genes across 32 tumor types. Matrix Biol Plus 1:100004

Kappelhoff R, Puente XS, Wilson CH, Seth A, López-Otín C, Overall CM (2017) Overview of transcriptomic analysis of all human proteases, non-proteolytic homologs and inhibitors: organ, tissue and ovarian cancer cell line expression profiling of the human protease degradome by the CLIP-CHIP™ DNA microarray. Biochim Biophys Acta, Mol Cell Res 1864:2210–2219. https://doi.org/10.1016/j.bbamcr.2017.08.004

Karamanos NK, Piperigkou Z, Theocharis AD, Watanabe H, Franchi M, Baud S, Brézillon S, Götte M, Passi A, Vigetti D, Ricard-Blum S, Sanderson RD, Neill T, Iozzo RV (2018) Proteoglycan chemical diversity drives multifunctional cell regulation and therapeutics. Chem Rev 118:9152–9232. https://doi.org/10.1021/acs.chemrev.8b00354

Kerr JS, Wright GJ (2012) Avidity-based extracellular interaction screening (AVEXIS) for the scalable detection of low-affinity extracellular receptor-ligand interactions. J Vis Exp 61:e3881. https://doi.org/10.3791/3881

Korpetinou A, Skandalis SS, Labropoulou VT, Smirlaki G, Noulas A, Karamanos NK, Theocharis AD (2014) Serglycin: at the crossroad of inflammation and malignancy. Front Oncol 3:327. https://doi.org/10.3389/fonc.2013.00327

Kotlyar M, Rossos AEM, Jurisica I (2017) Prediction of protein-protein interactions. Curr Protoc Bioinformatics 60:8.2.1–8.2.14. https://doi.org/10.1002/cpbi.38

Kozel BA, Mecham RP (2019) Elastic fiber ultrastructure and assembly. Matrix Biol 84:31–40. https://doi.org/10.1016/j.matbio.2019.10.002

Kraft-Sheleg O, Zaffryar-Eilot S, Genin O, Yaseen W, Soueid-Baumgarten S, Kessler O, Smolkin T, Akiri G, Neufeld G, Cinnamon Y, Hasson P (2016) Localized LoxL3-dependent fibronectin oxidation regulates myofiber stretch and integrin-mediated adhesion. Dev Cell 36:550–561. https://doi.org/10.1016/j.devcel.2016.02.009

Kuo HJ, Maslen CL, Keene DR, Glanville RW (1997) Type VI collagen anchors endothelial basement membranes by interacting with type IV collagen. J Biol Chem 272:26522–26529. https://doi.org/10.1074/jbc.272.42.26522

Küttner V, Mack C, Rigbolt KT, Kern JS, Schilling O, Busch H, Bruckner-Tuderman L, Dengjel J (2013) Global remodelling of cellular microenvironment due to loss of collagen VII. Mol Syst Biol 9:657. https://doi.org/10.1038/msb.2013.17

Launay G, Salza R, Multedo D, Thierry-Mieg N, Ricard-Blum S (2015) MatrixDB, the extracellular matrix interaction database: updated content, a new navigator and expanded functionalities. Nucleic Acids Res 43:D321–D327. https://doi.org/10.1093/nar/gku1091

Leitinger B, Saltel F (2018) Discoidin domain receptors: multitaskers for physiological and pathological processes. Cell Adhes Migr 12:398–399. https://doi.org/10.1080/19336918.2018.1491495

Lennon R, Byron A, Humphries JD, Randles MJ, Carisey A, Murphy S, Knight D, Brenchley PE, Zent R, Humphries MJ (2014) Global analysis reveals the complexity of the human glomerular extracellular matrix. J Am Soc Nephrol 25:939–951. https://doi.org/10.1681/ASN.2013030233

Li Z, Feizi T (2018) The neoglycolipid (NGL) technology-based microarrays and future prospects. FEBS Lett 592:3976–3991. https://doi.org/10.1002/1873-3468.13217

Luck K, Kim D-K, Lambourne L, Spirohn K, Begg BE, Bian W, Brignall R, Cafarelli T, Campos-Laborie FJ, Charloteaux B, Choi D, Coté AG, Daley M, Deimling S, Desbuleux A, Dricot A,

Gebbia M, Hardy MF, Kishore N, Knapp JJ, Kovács IA, Lemmens I, Mee MW, Mellor JC, Pollis C, Pons C, Richardson AD, Schlabach S, Teeking B, Yadav A, Babor M, Balcha D, Basha O, Bowman-Colin C, Chin S-F, Choi SG, Colabella C, Coppin G, D'Amata C, De Ridder D, De Rouck S, Duran-Frigola M, Ennajdaoui H, Goebels F, Goehring L, Gopal A, Haddad G, Hatchi E, Helmy M, Jacob Y, Kassa Y, Landini S, Li R, van Lieshout N, MacWilliams A, Markey D, Paulson JN, Rangarajan S, Rasla J, Rayhan A, Rolland T, San-Miguel A, Shen Y, Sheykhkarimli D, Sheynkman GM, Simonovsky E, Taşan M, Tejeda A, Tropepe V, Twizere J-C, Wang Y, Weatheritt RJ, Weile J, Xia Y, Yang X, Yeger-Lotem E, Zhong Q, Aloy P, Bader GD, De Las RJ, Gaudet S, Hao T, Rak J, Tavernier J, Hill DE, Vidal M, Roth FP, Calderwood MA (2020) A reference map of the human binary protein interactome. Nature 580:402–408. https://doi.org/10.1038/s41586-020-2188-x

Manninen A, Varjosalo M (2017) A proteomics view on integrin-mediated adhesions. Proteomics 17:1600022. https://doi.org/10.1002/pmic.201600022

Marini JC, Forlino A, Cabral WA, Barnes AM, San Antonio JD, Milgrom S, Hyland JC, Körkkö J, Prockop DJ, De Paepe A, Coucke P, Symoens S, Glorieux FH, Roughley PJ, Lund AM, Kuurila-Svahn K, Hartikka H, Cohn DH, Krakow D, Mottes M, Schwarze U, Chen D, Yang K, Kuslich C, Troendle J, Dalgleish R, Byers PH (2007) Consortium for osteogenesis imperfecta mutations in the helical domain of type I collagen: regions rich in lethal mutations align with collagen binding sites for integrins and proteoglycans. Hum Mutat 28:209–221. https://doi.org/10.1002/humu.20429

Marson A, Robinson DE, Brookes PN, Mulloy B, Wiles M, Clark SJ, Fielder HL, Collinson LJ, Cain SA, Kielty CM, McArthur S, Buttle DJ, Short RD, Whittle JD, Day AJ (2009) Development of a microtiter plate-based glycosaminoglycan array for the investigation of glycosaminoglycan-protein interactions. Glycobiology 19:1537–1546. https://doi.org/10.1093/glycob/cwp132

Martin S, Söllner C, Charoensawan V, Adryan B, Thisse B, Thisse C, Teichmann S, Wright GJ (2010) Construction of a large extracellular protein interaction network and its resolution by spatiotemporal expression profiling. Mol Cell Proteomics 9:2654–2665. https://doi.org/10.1074/mcp.M110.004119

Martinez-Martin N (2017) Technologies for Proteome-Wide Discovery of extracellular host-pathogen interactions. J Immunol Res 2017:2197615. https://doi.org/10.1155/2017/2197615

Meneghetti MCZ, Hughes AJ, Rudd TR, Nader HB, Powell AK, Yates EA, Lima MA (2015) Heparan sulfate and heparin interactions with proteins. J R Soc Interface 12:0589. https://doi.org/10.1098/rsif.2015.0589

Monboisse JC, Oudart JB, Ramont L, Brassart-Pasco S, Maquart FX (2014) Matrikines from basement membrane collagens: a new anti-cancer strategy. Biochim Biophys Acta 1840:2589–2598. https://doi.org/10.1016/j.bbagen.2013.12.029

Naba A, Clauser KR, Ding H, Whittaker CA, Carr SA, Hynes RO (2016) The extracellular matrix: tools and insights for the "omics" era. Matrix Biol 49:10–24. https://doi.org/10.1016/j.matbio.2015.06.003

Nde PN, Lima MF, Johnson CA, Pratap S, Villalta F (2012) Regulation and use of the extracellular matrix by Trypanosoma cruzi during early infection. Front Immunol 3:337. https://doi.org/10.3389/fimmu.2012.00337

Nehring LC, Miyamoto A, Hein PW, Weinmaster G, Shipley JM (2005) The extracellular matrix protein MAGP-2 interacts with Jagged1 and induces its shedding from the cell surface. J Biol Chem 280:20349–20355. https://doi.org/10.1074/jbc.M500273200

Okkelman IA, Sukaeva AZ, Kirukhina EV, Korneenko TV, Pestov NB (2014) Nuclear translocation of lysyl oxidase is promoted by interaction with transcription repressor p66β. Cell Tissue Res 358:481–489. https://doi.org/10.1007/s00441-014-1972-z

Orchard S, Kerrien S, Abbani S, Aranda B, Bhate J, Bidwell S, Bridge A, Briganti L, Brinkman FSL, Brinkman F, Cesareni G, Chatr-aryamontri A, Chautard E, Chen C, Dumousseau M, Goll J, Hancock REW, Hancock R, Hannick LI, Jurisica I, Khadake J, Lynn DJ, Mahadevan U, Perfetto L, Raghunath A, Ricard-Blum S, Roechert B, Salwinski L, Stümpflen V, Tyers M,

Uetz P, Xenarios I, Hermjakob H (2012) Protein interaction data curation: the international molecular exchange (IMEx) consortium. Nat Methods 9:345–350. https://doi.org/10.1038/nmeth.1931

Orgel JPRO, San Antonio JD, Antipova O (2011) Molecular and structural mapping of collagen fibril interactions. Connect Tissue Res 52:2–17. https://doi.org/10.3109/03008207.2010.511353

Ori A, Wilkinson MC, Fernig DG (2011) A systems biology approach for the investigation of the heparin/heparan sulfate interactome. J Biol Chem 286:19892–19904. https://doi.org/10.1074/jbc.M111.228114

Oughtred R, Stark C, Breitkreutz B-J, Rust J, Boucher L, Chang C, Kolas N, O'Donnell L, Leung G, McAdam R, Zhang F, Dolma S, Willems A, Coulombe-Huntington J, Chatr-Aryamontri A, Dolinski K, Tyers M (2019) The BioGRID interaction database: 2019 update. Nucleic Acids Res 47:D529–D541. https://doi.org/10.1093/nar/gky1079

Overall CM, Kleifeld O (2006) Tumour microenvironment – opinion: validating matrix metalloproteinases as drug targets and anti-targets for cancer therapy. Nat Rev Cancer 6:227–239. https://doi.org/10.1038/nrc1821

Paine CT, Paine ML, Snead ML (1998) Identification of tuftelin- and amelogenin-interacting proteins using the yeast two-hybrid system. Connect Tissue Res 38:257–267. https://doi.org/10.3109/03008209809017046. discussion 295–303

Paladin L, Tosatto SCE, Minervini G (2015) Structural in silico dissection of the collagen V interactome to identify genotype-phenotype correlations in classic Ehlers-Danlos syndrome (EDS). FEBS Lett 589:3871–3878. https://doi.org/10.1016/j.febslet.2015.11.022

Papanikolaou N, Pavlopoulos GA, Theodosiou T, Iliopoulos I (2015) Protein-protein interaction predictions using text mining methods. Methods 74:47–53. https://doi.org/10.1016/j.ymeth.2014.10.026

Parkin JD, San Antonio JD, Pedchenko V, Hudson B, Jensen ST, Savige J (2011) Mapping structural landmarks, ligand binding sites, and missense mutations to the collagen IV heterotrimers predicts major functional domains, novel interactions, and variation in phenotypes in inherited diseases affecting basement membranes. Hum Mutat 32:127–143. https://doi.org/10.1002/humu.21401

Parkin JD, San Antonio JD, Persikov AV, Dagher H, Dalgleish R, Jensen ST, Jeunemaitre X, Savige J (2017) The collαgen III fibril has a "flexi-rod" structure of flexible sequences interspersed with rigid bioactive domains including two with hemostatic roles. PLoS One 12:e0175582. https://doi.org/10.1371/journal.pone.0175582

Peysselon F, Ricard-Blum S (2011) Understanding the biology of aging with interaction networks. Maturitas 69:126–130. https://doi.org/10.1016/j.maturitas.2011.03.013

Peysselon F, Ricard-Blum S (2014) Heparin-protein interactions: from affinity and kinetics to biological roles. Application to an interaction network regulating angiogenesis. Matrix Biol 35:73–81. https://doi.org/10.1016/j.matbio.2013.11.001

Peysselon F, Xue B, Uversky VN, Ricard-Blum S (2011) Intrinsic disorder of the extracellular matrix. Mol BioSyst 7:3353–3365. https://doi.org/10.1039/c1mb05316g

Peysselon F, Marie F-A, Sylvie R-B (2012) From binary interactions of glycosaminoglycans and proteoglycans to interaction networks. In: Balazs EA (ed) Structure and function of biomatrix. Control of cell behavior and gene expression. Hyaluronan from basic science to clinical applications, vol 5. Matrix Biology Institute, Edgewater, New Jersey, pp 277–298

Qi L, Ding Y (2018) Construction of key signal regulatory network in metastatic colorectal cancer. Oncotarget 9:6086–6094. https://doi.org/10.18632/oncotarget.23710

Resovi A, Pinessi D, Chiorino G, Taraboletti G (2014) Current understanding of the thrombospondin-1 interactome. Matrix Biol 37:83–91. https://doi.org/10.1016/j.matbio.2014.01.012

Reuter JA, Ortiz-Urda S, Kretz M, Garcia J, Scholl FA, Pasmooij AMG, Cassarino D, Chang HY, Khavari PA (2009) Modeling inducible human tissue neoplasia identifies an extracellular matrix

interaction network involved in cancer progression. Cancer Cell 15:477–488. https://doi.org/10.1016/j.ccr.2009.04.002
Ricard-Blum S (2011) The collagen family. Cold Spring Harb Perspect Biol 3:a004978. https://doi.org/10.1101/cshperspect.a004978
Ricard-Blum S, Lisacek F (2017) Glycosaminoglycanomics: where we are. Glycoconj J 34:339–349. https://doi.org/10.1007/s10719-016-9747-2
Ricard-Blum S, Miele AE (2019) Omic approaches to decipher the molecular mechanisms of fibrosis, and design new anti-fibrotic strategies. Semin Cell Dev Biol 101:161–169. https://doi.org/10.1016/j.semcdb.2019.12.009
Ricard-Blum S, Vallet SD (2016a) Proteases decode the extracellular matrix cryptome. Biochimie 122:300–313. https://doi.org/10.1016/j.biochi.2015.09.016
Ricard-Blum S, Vallet SD (2016b) Matricryptins network with Matricellular receptors at the surface of endothelial and tumor cells. Front Pharmacol 7:11. https://doi.org/10.3389/fphar.2016.00011
Ricard-Blum S, Vallet SD (2019) Fragments generated upon extracellular matrix remodeling: biological regulators and potential drugs. Matrix Biol 75–76:170–189. https://doi.org/10.1016/j.matbio.2017.11.005
Rivera CG, Bader JS, Popel AS (2011) Angiogenesis-associated crosstalk between collagens, CXC chemokines, and thrombospondin domain-containing proteins. Ann Biomed Eng 39:2213–2222. https://doi.org/10.1007/s10439-011-0325-2
Rizwan A, Cheng M, Bhujwalla ZM, Krishnamachary B, Jiang L, Glunde K (2015) Breast cancer cell adhesome and degradome interact to drive metastasis. NPJ Breast Cancer 1:15017. https://doi.org/10.1038/npjbcancer.2015.17
Robertson J, Jacquemet G, Byron A, Jones MC, Warwood S, Selley JN, Knight D, Humphries JD, Humphries MJ (2015) Defining the phospho-adhesome through the phosphoproteomic analysis of integrin signalling. Nat Commun 6:6265. https://doi.org/10.1038/ncomms7265
Robertson J, Humphries JD, Paul NR, Warwood S, Knight D, Byron A, Humphries MJ (2017) Characterization of the Phospho-adhesome by mass spectrometry-based proteomics. Methods Mol Biol 1636:235–251. https://doi.org/10.1007/978-1-4939-7154-1_15
Rogers CJ, Hsieh-Wilson LC (2012) Microarray method for the rapid detection of glycosaminoglycan-protein interactions. Methods Mol Biol 808:321–336. https://doi.org/10.1007/978-1-61779-373-8_22
Roper JA, Williamson RC, Bass MD (2012) Syndecan and integrin interactomes: large complexes in small spaces. Curr Opin Struct Biol 22:583–590. https://doi.org/10.1016/j.sbi.2012.07.003
Salza R, Peysselon F, Chautard E, Faye C, Moschcovich L, Weiss T, Perrin-Cocon L, Lotteau V, Kessler E, Ricard-Blum S (2014) Extended interaction network of procollagen C-proteinase enhancer-1 in the extracellular matrix. Biochem J 457:137–149. https://doi.org/10.1042/BJ20130295
Salza R, Oudart J-B, Ramont L, Maquart F-X, Bakchine S, Thoannès H, Ricard-Blum S (2015) Endostatin level in cerebrospinal fluid of patients with Alzheimer's disease. J Alzheimers Dis 44:1253–1261. https://doi.org/10.3233/JAD-142544
Salza R, Lethias C, Ricard-Blum S (2017) The multimerization state of the amyloid-β42 amyloid peptide governs its interaction network with the extracellular matrix. J Alzheimers Dis 56:991–1005. https://doi.org/10.3233/JAD-160751
Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13:2498–2504. https://doi.org/10.1101/gr.1239303
Shin SJ, Yanagisawa H (2019) Recent updates on the molecular network of elastic fiber formation. Essays Biochem 63:365–376. https://doi.org/10.1042/EBC20180052
Snider J, Kotlyar M, Saraon P, Yao Z, Jurisica I, Stagljar I (2015) Fundamentals of protein interaction network mapping. Mol Syst Biol 11:848. https://doi.org/10.15252/msb.20156351
Socovich AM, Naba A (2019) The cancer matrisome: from comprehensive characterization to biomarker discovery. Semin Cell Dev Biol 89:157–166. https://doi.org/10.1016/j.semcdb.2018.06.005

Sweeney SM, Orgel JP, Fertala A, McAuliffe JD, Turner KR, Di Lullo GA, Chen S, Antipova O, Perumal S, Ala-Kokko L, Forlino A, Cabral WA, Barnes AM, Marini JC, San Antonio JD (2008) Candidate cell and matrix interaction domains on the collagen fibril, the predominant protein of vertebrates. J Biol Chem 283:21187–21197. https://doi.org/10.1074/jbc.M709319200

Symoens S, Renard M, Bonod-Bidaud C, Syx D, Vaganay E, Malfait F, Ricard-Blum S, Kessler E, Van Laer L, Coucke P, Ruggiero F, De Paepe A (2011) Identification of binding partners interacting with the α1-N-propeptide of type V collagen. Biochem J 433:371–381. https://doi.org/10.1042/BJ20101061

Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, Doncheva NT, Roth A, Bork P, Jensen LJ, von Mering C (2017) The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. Nucleic Acids Res 45:D362–D368. https://doi.org/10.1093/nar/gkw937

Takada W, Fukushima M, Pothacharoen P, Kongtawelert P, Sugahara K (2013) A sulfated glycosaminoglycan array for molecular interactions between glycosaminoglycans and growth factors or anti-glycosaminoglycan antibodies. Anal Biochem 435:123–130. https://doi.org/10.1016/j.ab.2013.01.004

The Gene Ontology Consortium (2017) Expansion of the gene ontology knowledgebase and resources. Nucleic Acids Res 45:D331–D338. https://doi.org/10.1093/nar/gkw1108

Theocharis AD, Manou D, Karamanos NK (2019) The extracellular matrix as a multitasking player in disease. FEBS J 286:2830–2869. https://doi.org/10.1111/febs.14818

Thomson J, Singh M, Eckersley A, Cain SA, Sherratt MJ, Baldock C (2019) Fibrillin microfibrils and elastic fibre proteins: functional interactions and extracellular regulation of growth factors. Semin Cell Dev Biol 89:109–117. https://doi.org/10.1016/j.semcdb.2018.07.016

UniProt Consortium (2019) UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res 47:D506–D515. https://doi.org/10.1093/nar/gky1049

Vaca DJ, Thibau A, Schütz M, Kraiczy P, Happonen L, Malmström J, Kempf VAJ (2019) Interaction with the host: the role of fibronectin and extracellular matrix proteins in the adhesion of gram-negative bacteria. Med Microbiol Immunol 209:277–299. https://doi.org/10.1007/s00430-019-00644-3

Vallet SD, Ricard-Blum S (2019) Lysyl oxidases: from enzyme activity to extracellular matrix cross-links. Essays Biochem 63:349–364. https://doi.org/10.1042/EBC20180050

Vallet SD, Lisette D, Arnaud V, Romain S, Faye C, Attila A, Nicolas T-M, Sylvie R-B (2017) Strategies for building protein–glycosaminoglycan interaction networks combining SPRi, SPR, and BLI. In: Schasfoort RBM (ed) Handbook of surface plasmon resonance, 2nd edn. The Royal Society of Chemistry, London, pp 398–414

Vallet SD, Miele AE, Uciechowska-Kaczmarzyk U, Liwo A, Duclos B, Samsonov SA, Ricard-Blum S (2018) Insights into the structure and dynamics of lysyl oxidase propeptide, a flexible protein with numerous partners. Sci Rep 8:11768. https://doi.org/10.1038/s41598-018-30190-6

Vallet SD, Clerc O, Ricard-Blum S (2020) Glycosaminoglycan-protein interactions: the first draft of the glycosaminoglycan interactome. J Histochem Cytochem Aug 6:22155420946403. https://doi.org/10.1369/0022155420946403

Vicente-Manzanares M, Sánchez-Madrid F (2018) Targeting the integrin interactome in human disease. Curr Opin Cell Biol 55:17–23. https://doi.org/10.1016/j.ceb.2018.05.010

Winograd-Katz SE, Fässler R, Geiger B, Legate KR (2014) The integrin adhesome: from genes and proteins to human disease. Nat Rev Mol Cell Biol 15:273–288. https://doi.org/10.1038/nrm3769

Wong H, Prévoteau-Jonquet J, Baud S, Dauchez M, Belloy N (2018) Mesoscopic rigid body modelling of the extracellular matrix self-assembly. J Integr Bioinform 15:20180009. https://doi.org/10.1515/jib-2018-0009

Yang M, Liu J, Wang F, Tian Z, Ma B, Li Z, Wang B, Zhao W (2019) Lysyl oxidase assists tumor-initiating cells to enhance angiogenesis in hepatocellular carcinoma. Int J Oncol 54:1398–1408. https://doi.org/10.3892/ijo.2019.4705

Yu K, Lung P-Y, Zhao T, Zhao P, Tseng Y-Y, Zhang J (2018) Automatic extraction of protein-protein interactions using grammatical relationship graph. BMC Med Inform Decis Mak 18:42. https://doi.org/10.1186/s12911-018-0628-4

Zaidel-Bar R, Itzkovitz S, Ma'ayan A, Iyengar R, Geiger B (2007) Functional atlas of the integrin adhesome. Nat Cell Biol 9:858–867. https://doi.org/10.1038/ncb0807-858

Zhan X-H, Jiao J-W, Zhang H-F, Li C-Q, Zhao J-M, Liao L, Wu J-Y, Wu B-L, Wu Z-Y, Wang S-H, Du Z-P, Shen J-H, Zou H-Y, Neufeld G, Xu L-Y, Li E-M (2017) A three-gene signature from protein-protein interaction network of LOXL2- and actin-related proteins for esophageal squamous cell carcinoma prognosis. Cancer Med 6:1707–1719. https://doi.org/10.1002/cam4.1096

Chapter 7
Integration of Matrisome Omics: Towards System Biology of the Tumor Matrisome

Valerio Izzi, Jarkko Koivunen, Pekka Rappu, Jyrki Heino, and Taina Pihlajaniemi

Abstract The tumor matrisome, the collection of structural and functional extracellular matrix (ECM) proteins found in tumors together with ECM-associated proteins, growth factors, cytokines and chemokines is an extremely dynamic entity whose composition and regulation depends on both local, cell-specific processes and overarching pan-cancer microenvironmental landscapes. A deeper knowledge and understanding of the tumor matrisome would grant us novel diagnostic, prognostic and therapeutic possibilities but to reach that point there is a pressing need to combine fine-grained molecular studies with system biology approaches to build a complete map of what actually happens in the tumor microenvironment. In this chapter, we will review the most salient findings on the tumor matrisome at the transcriptomics, regulomics and proteomics level and show how the use of big data enables the identification of biological features and characteristics that cross different tumors and the discovery of novel molecular mechanisms connecting the different omics level. Also, we will provide a simple walkthrough to jumpstart matrisome data analysis for complete beginners in a bid to ignite more research in this fascinating field.

V. Izzi (✉)
Faculty of Biochemistry and Molecular Medicine, University of Oulu, Oulu, Finland

Finnish Cancer Institute, Helsinki, Finland
e-mail: valerio.izzi@oulu.fi

J. Koivunen · T. Pihlajaniemi
Faculty of Biochemistry and Molecular Medicine, University of Oulu, Oulu, Finland
e-mail: jarkko.koivunen@oulu.fi; taina.pihlajaniemi@oulu.fi

P. Rappu · J. Heino
Department of Biochemistry, University of Turku, Turku, Finland
e-mail: pekrappu@utu.fi; jyheino@utu.fi

S. Ricard-Blum (ed.), *Extracellular Matrix Omics*, Biology of Extracellular Matrix 7,
https://doi.org/10.1007/978-3-030-58330-9_7

7.1 Introduction

Tumorigenesis is a complex and dynamic process and, at all stages (initiation, progression and metastasis), neoplastic cells are physically and chemically embedded in a rich milieu of immune and stromal cells, growth factors, enzymes, signaling molecules and the structural and functional proteins constituting the extracellular matrix (ECM) (Henke et al. 2020; Hinshaw and Shevde 2019; Walker et al. 2018; Wang et al. 2017). Supported by a large amount of biochemical and functional data about the interaction between the non-cellular elements of the tumor microenvironment (TME) and the fundamental delineative work of Naba et al. in 2012, researchers have now learned to rethink the ECM and the other extracellular proteins and growth factors as an ensemble, the "matrisome", that is intimately linked to cancer progression (Jinka et al. 2012; Malandrino et al. 2018; Naba et al. 2012; Poltavets et al. 2018; Socovich and Naba 2019).

The advent of sequencing technologies and their rapid spread through laboratories worldwide has irrevocably changed the world of cancer research ([Editorial] 2020). The unprecedented possibilities offered by large, harmonized patient cohorts analyzed by several "omics" approaches (genomics, epigenomics, transcriptomics, proteomics, etc.), such as those offered by The Cancer Genome Atlas (TCGA) (Gao et al. 2019), have opened up to fundamental discoveries in all the most crucial aspects of cancer, including driver and passenger mutations, cellular signaling pathways, stemness and differentiation instances, and the mechanisms that regulate entire "layers" of the TME (Bailey et al. 2018; Huang et al. 2018; Malta et al. 2018; Sanchez-Vega et al. 2018; Thorsson et al. 2018).

Similar efforts have also been surfacing in respect to the matrisome, whose in-depth characterization has, however, verged more on the biochemical properties of the proteins identified, their functions and their eventual interactions rather than on the systematic classification of which matrisome elements characterize a given TME and what regulatory systems decree the presence or the absence of that specific element.

In this chapter, we will review the most salient findings on the tumor matrisome at the transcriptomics, regulomics and proteomics level and show how:

1. The use of big data enables the identification of biological features and characteristics that cross different tumors and are foundation stones of neoplastic diseases, and
2. These seemingly disconnected "levels" of information can be integrated together to discover novel ties among them.

Also, as we surmise that one of the barriers to the evaluation of the matrisome at a more systematic level is represented by the need for specific bioinformatics and coding abilities, we will provide a simple walkthrough that will enable even the most unexperienced users to jumpstart their investigations without the need for professional help (*though you should definitely contact a bioinformatician at some point!*).

7.2 Transcriptomics of the Tumor Matrisome

7.2.1 Integrative Expression Profiling and "Classical" Gene Signature Studies

In respect to the sizeable number of publications investigating the matrisome at the proteomics level, the transcriptional landscape of the tumor matrisome has not yet been investigated to the same depth and this difference becomes even more evident when one considers studies focusing on Pan-Cancer analyses.

The "Pan-Cancer" mode of analysis is particularly relevant to this chapter and needs a bit of elaboration before proceeding. Fortunately, the seminal works by, e.g., (Hoadley et al. 2014, 2018; Nawy 2018) are a perfect example. Thanks to them, we now know that tumors apparently different at the histopathological level are nevertheless quite similar, as they are shaped by common cellular and genetic processes which originate from similar, if not the same, cell of origin. Such a result can only be obtained by comparing multiple tumor types at once, eventually incorporating tens of clinical and biological subtypes, and enables the observation of recurrent "themes" that transverse entire cohorts of tumors and that might be hidden under other, more prominent transcriptional features if the observer's perspective is not wide enough to capture them. In a typical Pan-Cancer analysis, furthermore, the investigational approach is bottom-up (or, to be more precise, unsupervised) and the observer begins with a database as large as possible, subjecting none or minimal filters on the types and quantities of "features" (may they be genes, pathways, processes, etc.) to be studied. This lets the most salient elements connecting different tumors "emerge" from the context and guarantees them some sort of "priority" over the rest of the transcriptomic background.

Integrative Analysis
The purpose of integrative Pan-Cancer analyses is to identify similarities and differences between different tumor types across some fundamental genetic or cellular process. The discovery is generally driven by the intersection of different "omics" or the deep "excavation" of such data (see also Fig. 7.1).

In tumor matrisome field, most of the available data target one or few, related neoplasms. These studies are, of course, invaluable for the determination of which matrisome genes are characteristic of a given tumor type/subtype, yet they fall short in bringing Pan-Cancer similarities to light, and so questions like "does the transcriptional architecture of matrisome in different tumor types differ?" or "are there similarities in the matrisome of different tumors?" remain largely unanswered. Furthermore, there are surprisingly few studies devoted to understanding and classifying the tumor matrisome transcriptome and its dynamics and many interesting findings are buried under other results in articles with different aims, making it all the more difficult to provide a comprehensive and coherent view of this field of research.

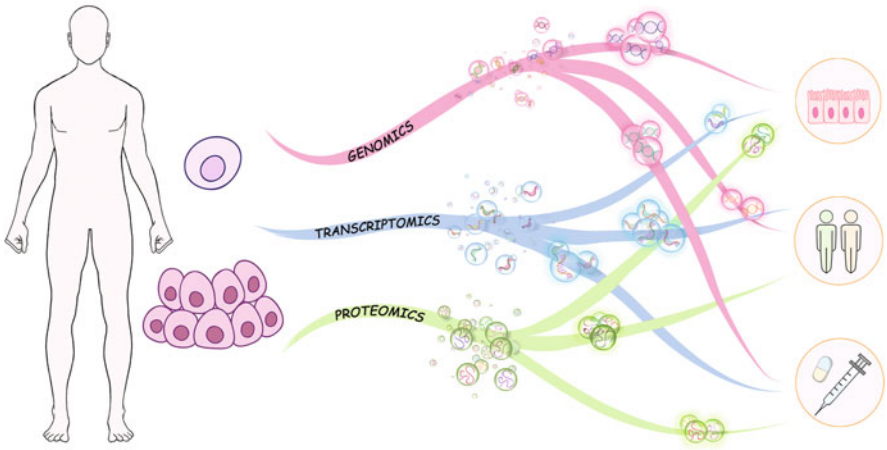

Fig. 7.1 Integrative approaches to study the tumor matrisome make use of samples and data from cells, tissues and entire cohorts of patients. Analyses can focus on a single "level" (genomics, transcriptomics and proteomics) to identify groups of features (*signatures*, represented as bubbles of different size and importance) shared across the level. Different levels can be also integrated "vertically" and used to identify multidimensional signatures of, e.g., tissue and cell identity, disease state and response/resistance to therapy. Figure courtesy of Monica Bassignana (www.monicabassignana.com)

Finally, it should be added that the incredible diversity in functions, abundance and cellular origin of matrisome components is a divisive factor for reviewing and condensing knowledge since many times, for example, cytokines and growth factors will be cited in manuscripts according to their functions and independently of ECM genes eventually found among the same results. All this said, there are still excellent examples of integrated tumor matrisome transcriptomics to be found.

One such is the recent work of Lim et al. (2017), in which a large database was built by the integration of single pancreatic, ovarian, renal, gastric, prostate, liver, bladder, melanoma, colorectal lung and breast cancer datasets and used to unravel connections between matrisome gene expression and clinical, cytopathological and immune features of cancer. The authors deployed a strongly supervised approach, as they decided *a priori* which matrisome genes to analyze and what type of analyses and comparisons to perform to substantiate their hypothesis. More specifically, not the whole span of the matrisome (that enlists 1028 entries corresponding to 1026 genes according to the Molecular Signatures Database, www.gsea-msigdb.org/gsea/msigdb/) was analyzed, but rather a restricted set of 29 genes that constitute a signature called Tumor Matrisome Index (TMI) previously validated in early-stage non-small cell lung cancer (NSCLC) (Lim et al. 2017, 2018) and mostly comprised of non-core matrisome genes. TMI was checked recursively for its ability to tell apart tumor samples from healthy donors and, upon confirmation of such discriminative power, investigated in uni- and multi-variable survival models and for association

with tumor stages, cytological findings and immunotherapeutic targets such as programmed death protein *PD-1*. What remains unanswered in this study, such as the existence of other similar matrisome signatures or the eventual similarities and differences between the neoplasms, is still keenly investigable by researchers thanks to the egregious work made by the authors in sharing the code needed to regenerate the database (and available at doi.org/10.6084/m9.figshare.7878086).

Another example of integrated cancer matrisome transcriptomics profiling is the work by Yuzhalin et al. who, by comparing tumor samples from esophagus, intestine, stomach, lung, breast and ovary with respective normal samples, identified a set of 9 core matrisome genes largely overexpressed in neoplasms (Yuzhalin et al. 2018a, b). Further on, the authors confirm the expression of these genes at the protein level in matching cases and demonstrate the association of this signature with patient survival in all tumor types and its association with cellular processes such as epithelial-mesenchymal transition (EMT), hypoxia, etc. in colorectal cancers. In respect to the work of Lim et al. mentioned above, this latter manuscript deploys a more unsupervised approach to data analysis, looking at all core matrisome genes expressed in a given cancer, selecting for those highly overexpressed in respect to healthy tissues and then comparing results to identify genes with a similar behavior. Still, how these tumor types have been selected remains unclear and one is left to wonder whether by adding other neoplasms, this signature would have shown some form of tissue specificity or whether other signatures would have emerged. Importantly, the work of Yuzhalin et al. shows the ease of use of modern data browsers (such as Oncomine in this case, www.oncomine.org) to access big data from omics study and build a knowledge base that can be confidently tested in downstream applications.

A different take on the problem of identifying matrisome signatures is brought forward by Chakravarthy et al. (2018), who recently curated an analysis in which tumor vs. healthy tissues for 15 different neoplasms were compared for gene expression and the results intersected with the parent term for ECM in gene ontology ("extracellular matrix"—GO: 0031032), thus identifying a set of 58 matrisome genes as cancer-associated and showing that this signature associates with transforming growth factor b (TGF-b), the presence of fibroblasts within the TME and the outcome of immunotherapy.

Matrisome gene sets can also be derived by "fishing" in the transcriptome using a "bait" with some biological meaning. This intriguing approach has been lately presented by Peeney et al. (2019), who explored the association of tissue inhibitor of metalloproteinase-2 (TIMP-2) with breast and lung neoplastic cancer and found that the expression of TIMP-2 could be used to identify genes that strongly correlate with it and potentially modulate its expression and function. Unsurprisingly, most of these genes are matrisome genes, both core ECM and ECM-associated ones, with the remaining part covering TGF-b and EMT pathways. A very similar strategy has been successfully applied by Jia et al. (2016) who used collagen XI gene (*COL11A1*) as a bait for identifying transcriptionally co-regulated genes in cancer-associated fibroblasts (CAFs), yielding a large signature of 195 genes supposedly marking these cells and largely composed of matrisome genes. Interestingly, as noted in a comment

to this work by Anastassiou D. (2017), this signature might actually be a mixture of specific and nonspecific signals, as it crosses with the 64-gene "invasiveness signature" that the authors have reported (Kim et al. 2010) and supplements it with other gene sets, such as a 10-gene signature of ovarian cancers (Cheon et al. 2014), whose specificity to CAFs is also unsure. Despite the controversy, it is important to notice that also both the 64-gene signature by Kim et al. and the 10-gene signature by Cheon et al. are almost entirely made of matrisome genes (*core structural ECM proteins—such as collagens—for a good part, too!*), evidencing how crucial the correct classification of the tumor matrisome is for understating the processes and mechanisms of TME.

Recently, we have performed the largest cancer matrisome analysis of this kind so far, by comparing gene expression patterns across more than 10,000 patients and 32 tumor types and inferring the transcriptional architecture (transcription factors and master regulators) governing the tumor matrisome as well as the therapeutic possibilities offered by this fine-mapping strategy (Izzi et al. 2019). In this work we used a completely unsupervised approach, not performing any pre-selection of matrisome genes on biological or computational grounds nor selecting for tumor types *a priori*, and we identified a clear organ-of-origin mechanism that is responsible for similarities between tumors from the same anatomical location. Such mechanism is loosely pyramidal, with master regulators (cancer drivers and the cellular and supracellular processes they supervise) shared across tumors from similar origins connecting to an overarching sparse network of transcription factors (highly tissue- and cell-specific) which control the expression of similar matrisome genes. To our knowledge this is the first and, till now, the only study of this type in the field, and provides the only "regulatory map" of the tumor matrisome so far. Furthermore, it is interesting to notice how it crosses with other studies reported above. E.g., comparing signatures from different tumors, we find agreement with the same conclusions as Lim and Yuzhalin, we confirm the expected role of TGF-b signaling and hypoxia in instructing matrisome expression patterns, and identify tumor-specific and Pan-Cancer regulatory networks which can be pharmacologically targeted and impinge on patient survival.

The analyses presented above, albeit with their different aims and approaches, are tied together by the vision of tumor matrisome as an ensemble of elements whose activity and regulation goes beyond a single type of cancer and is explainable based on biological factors that the different tumors share. This vision is, of course, innovative and is getting momentum now that big data from large omics experiments are becoming increasingly available. Still, integrative Pan-Cancer studies of the tumor matrisome are in their infancy as compared with classical gene expression profiling studies (one tumor type, eventually stratified by clinical and/or biological subtypes), whose count outnumbers integrative studies by far. An extensive (*or even barely appropriate!*) discussion on these more "classical" studies identifying matrisome expression and regulation in cancer would be largely outside the page constraints of this chapter, so we will only briefly touch on the subject, leaving it to the interested readers to search in-depth the relevant literature (maybe by trying, at first, data mining approaches to PubMed such as inputting the string "*extracellular*

matrix AND cancer gene expression" in PubVenn—pubvenn.appspot.com—or in PubTator Central (PTC)—www.ncbi.nlm.nih.gov/research/pubtator).

A most typical example of this type of studies is provided by the pioneering works of Bergamaschi et al. (2008) and Triulzi et al. (2013), which demonstrate the possibility of stratifying breast cancer patients according to matrisome gene expression and placing one of the four groups identified, the "ECM3" group, at the highest risk of disease progression and, hence, marking it with poor survival likeliness. Interestingly, the "ECM3" group is entirely made of Grade-III breast carcinomas (high-grade dedifferentiated cancer), suggesting specialization of the tumor ECM concurrent with neoplastic progression (a hypothesis also supported by our recent study presented above). On the same note, Bao et al. (2019) as well as Helleman et al. (2008) and Jansen et al. (2004) demonstrated a clear association between the expression of matrisome or matrisome/receptors signatures and breast cancer patient stratification for survival and other clinical endpoints, strongly supporting the idea that precise matrisome regulation is a crucial element along the oncogenic development processes of the breast. Adding to this, Brechbuhl et al. (2020) recently showed that a subset of CD146 (Mucin 18)-negative CAFs are crucial to the pro-metastatic behavior of human breast cancer cells xenoimplanted in mice as their secretome includes various collagens and laminins as well as fibronectin 1 (*FN1*) and tenascin C (*TNC*), all necessary to cell motility as well as the enzymatic machinery of the lysyl oxidases needed to crosslink collagens and stiffen the ECM, a hallmark of increased invasiveness (Najafi et al. 2019).

Similar findings have been reported for ovary tumors (e.g., Cheon et al. 2014; Finkernagel et al. 2016; Pearce et al. 2018a, b; Zhang et al. 2013), all demonstrating the roles of different matrisome and ECM signatures in driving tumor metastatization, resistance to therapy and progression, and pointed to their clinical usefulness in predicting patient trajectories and, hence, devise more accurate treatment schemes.

Comparably, several reports exist about the importance of matrisome signatures or even single matrisome components, both at the gene expression and the protein level (also including a few circulating ECM proteins), for the growth, metastatization and chemoresistance of the neoplasms of the lung (Andriani et al. 2018; Frezzetti et al. 2019; Lim et al. 2018), colon and liver (Naba et al. 2014b, 2017a, b), pancreas (Naba et al. 2017a, b; Tian et al. 2019), prostate (Pang et al. 2019; Peixoto et al. 2019; Penet et al. 2017), brain (Ciasca et al. 2016; Simão et al. 2018; Sood et al. 2019) and even leukemias, lymphomas and myelomas (Brandt et al. 2013; Cerchietti et al. 2019; Foroushani et al. 2017; Glavey et al. 2017; Izzi et al. 2017a, b, c; Kobayashi et al. 2020).

As mentioned before, this list is far from exhaustive and only serves the purpose of priming readers to the vast field of matrisome signatures in cancer, as opposite to integrative Pan-Cancer studies that still are minority. Rather than dragging on with the list of available tumor matrisome studies, we prefer to provide an example of simple integrative analysis workflow that readers can execute completely online, without any programming skill nor sophisticated instrumentation and that will hopefully stimulate more investigations in this direction.

7.2.2 *Walkthrough: A Guided Example of Pan-Cancer Integrative Analysis of the Tumor Matrisome Using CBioPortal*

The diffusion of omics techniques and the consequent exponential accumulation of big data has spurred an unprecedented need for data analysts and bioinformaticians in biomedical sciences. Typically, integrative Pan-Cancer analyses require coding skills, data management abilities and a solid foundation in statistics, but investigators with no programming knowledge can also approach these analyses leveraging on trusted, free portals optimized for a "click-and-forget" experience. A colossus in this field is CBioPortal, a freely available one-stop portal for cancer omics analyses (Cerami et al. 2012) hosted by the Center for Molecular Oncology at Memorial Sloan Kettering Cancer Center and maintained by an international consortium. We will here make use of the wonderful set of features and analytical power offered by CBioPortal to build a simple integrative analysis and show that, with minimal effort and the help of Excel spreadsheets, any researcher can explore big data. It is important to notice, however, that any responsive database or web-based resource comes with a compromise between ease of access and freedom in terms of analytical possibilities, and long sessions of code writing and data wrangling in various programming languages remain an unavoidable step in any in-depth analysis. A short list of useful resources is provided in Table 7.1.

- Log in at https://www.cbioportal.org/. At this stage, if you haven't already, is important to create an account that will allow you to perform some of the analyses in this walkthrough. To this aim, click the Login button at top-right corner of the page and follow the instructions.
- The CBioPortal homepage (version 3.3.2, database version 2.12.4 at the moment of writing) has three tabs on top (Query, Quick Search (Beta!) and Download) to choose the mode of analysis. Below these tabs you will find the study selector, with a by-organ box on the left and a detailed by-study view in the center. On the right you will find the news box with social media integration and some general details on the amount of studies available.
- In Query mode, CBioPortal has a convenient TCGA PanCancer Atlas Studies button in the upper frame of the central box. Clicking it will select 32 studies and 10,967 samples. Also, different studies that compose the Pan-Cancer atlas can be manually selected by ticking their respective boxes.
- After selecting at least one cancer type (or all the Pan-Cancer studies in our example!), the two buttons on the bottom (Query By Gene and Explore Selected Studies) become active, enabling to either scan selected studies by a list of genes of interest or to access an agglomerated view of all data in selected studies, which can be used to launch group-level analyses.
- Click Explore Selected Studies and a new page with demographics and general clinical and genetic findings will open. One of the (many) tables within is the Genomic Profile Sample counts box, where you can further subset data based on

Table 7.1 Programming languages and resources for integrative analyses of tumor matrisome transcriptomics

	Type of resource	Use	Web-based	Restricted access	Website
R	Programming language	Analysis[#]	No	No	http://www.r-project.org/
Python	Programming language	Analysis[#]	No	No	http://www.python.org/
CBioPortal	Portal	Analysis and data download[§]	Yes	No	http://www.cbioportal.org/
Xena	Portal	Analysis and data download[§]	Yes	No	http://xenabrowser.net/
Oncomine	Portal	Analysis[a]	Yes	Yes	http://www.oncomine.org/
Gepia 2	Portal	Analysis[a]	Yes	No	http://gepia2.cancer-pku.cn/
Genetic data common (GDC)	Portal	Data download and minimal analysis	Yes	No/Yes (depending on data type)	http://portal.gdc.cancer.gov/
Cosmic	Portal	Analysis and data download[§]	Yes	No	http://cancer.sanger.ac.uk/cosmic
CanEvolve	Portal	Analysis[a]	Yes	No	http://www.canevolve.org/
KM plotter	Portal	Analysis[a]	Yes	No	http://kmplot.com/
PROGgene V2	Portal	Analysis[a]	Yes	No	http://genomics.jefferson.edu/proggene/

[a]We refer here to the possibility of performing data analysis and downloading results, different from the complete analytical solutions (and freedom) offered by programming languages ([#]) and also different from the possibility to download entire slices of data together with analysis results etc. offered by other tools ([§])

the available type of information. In this case, focusing on transcriptomics, we want to tick the mRNA Expression, RSEM (Batch normalized from Illumina HiSeq_RNASeqV2) box to exclude samples without gene expression data.

- Click the button Select Samples that appears at the bottom of the box and the data will be filtered accordingly
- Now move to the box Cancer Studies and select, e.g., brca_tcga_pan_can_atlas_2018. This will restrict our analysis to breast cancer samples within the PanCancer Atlas dataset.
- Now, moving to the box Cancer Type Detailed, you will notice that only the subtypes of breast cancer samples are available. We will use these to make groups to compare. Let's for example select Breast Invasive Ductal Carcinoma by ticking its box. We can now click the Groups button in the upper right corner and Create a

new group from selected samples. The name automatically suggested is the same as the subtype, but you can change it. Conclude by clicking Create.

- Head back to the box Cancer Type Detailed, deselect the previous breast cancer subtype and choose Breast Invasive Lobular Carcinoma instead. Repeat the group creation step above to have a second group. Now that you have at least two groups within the same study you can run comparisons!
- From the Groups button select the two groups we have just created and press Compare. You will land into a comparison page with various result tabs. You can glance at, or even study, them all but for this walkthrough we will focus on the last tab, mRNA, that returns the association between a gene and one the cancer subtypes selected, if any. To simplify, let's tick the selector Significant only that is on the upper frame of the result table.
- In the right corner of the upper frame there is a Download TSV button, marked by a cloud icon with a downward arrow. Click it and store your results in your preferred folder.
- Now we need to read these data in Excel. Open a blank spreadsheet and, from the top ribbon, choose Data and, on the left, the From Text/CSV button (alternatively, from the File menu, choose the Open function). A window to locate your file will open, but your file (being in. TSV format) won't appear unless you choose All Files (its default is Text Files) from the file format selector in the bottom right corner. Choose the cell where your imported file will begin, and you are set.
- Now we can ask "which of these differentially-regulated genes are actually matrisome genes?". To this aim, we will open a new browser window and point to The Matrisome Project (http://matrisome.org/). The list we need is included in the Excel file "Human Matrisome (Updated August 2014)" which you can access by clicking its icon (the filename is matrisome_hs_masterlist).
- Clicking the Excel icon will download the list of matrisome genes with its categories. Open the file when downloaded.
- To filter the CBioPortal results by the matrisome genes we need small additional modifications of both the files. Excel, in fact, needs perfect matching between names of the table range you want to filter, of the selection criteria you want to apply and of the destination where you want to put the results. To enable this, you will need to:
- Modify the header of the first column of the results from Gene to Gene Symbol (as in the matrisome list file),
- Get back at the results, select the whole table (Ctrl + A on Windows), click Data from the upper ribbon and from the Filter section choose Advanced.
- Set as follows:
  - Action: Filter the list, in-place
  - List range: leave as it is (the whole table should be selected)
  - Criteria range: click the upper arrow button, move to the matrisome file you downloaded earlier and select all the entries in the Gene Symbol column plus the header. There are some deprecated entries at the bottom of this list. You might want to avoid selecting them at this step!

- Press Enter and click OK. From more than 9000 entries your list has now reduced to 411 matrisome genes. You can sort them by their association with a given tumor subtype by the last column of the results.
- Select only the genes that have highest expression in Breast Invasive Ductal Carcinoma and you will have a 85-gene signature. You could copy the selected gene symbols back into the CBioPortal and launch a real Pan-Cancer analysis as follows:
  - Browse back to the first CBioPortal page (not the same where differential expression results are)
  - Remove the cancer subtype filter from the upper left side of the page
  - At the upper right corner, just below your sign-in box, click the search box "Click gene symbols below or enter here" and paste your 85 genes
  - Click Query and observe the results. You might, eventually, filter them again by tumor types and subtypes and study them down to individual genes in individual tumors. Also, you could study these genes in single tumors by going back to the Query By Gene mode, or study them in other fantastic resources such as the Xena Browser (http://xenabrowser.net/) or GEPIA2 (http://gepia2.cancer-pku.cn/#index).

7.2.3 From Transcriptomics to Regulomics

The enormous analytical power offered by the integrative analysis of tumor transcriptomes finds a natural application in the evaluation of what lies upstream and downstream of genes and how these elements connect to the transcriptome. Oversimplifying, we could say that cis- and trans-acting mutations, single-nucleotide variations and copy-number alterations, but also epigenetic regulators, micro- and long non-coding RNAs (miRNAs and lncRNAs, respectively), transcription factors and their own regulators are all examples of upstream regulators of the transcriptomics layer, while the whole apparatus for protein translation, post-translational modification and inter and intracellular trafficking being downstream of it (Bansal et al. 2019; Bradner et al. 2017; Casamassimi and Ciccodicola 2019; Ghedira 2018; Lavanya et al. 2019; Lelli et al. 2012; Nath et al. 2015; Qin et al. 2015; Tak et al. 2019).

If a paucity of integrative studies can be observed for the tumor matrisome transcriptome, it should not be surprising to find virtually no studies on matrisome regulomics at the transcriptional layer. In fact, while much is known at the proteomics/interactomics level downstream of tumor matrisome genes (see Sect. 7.3), what lies upstream is still largely a black hole, with an astonishing lack of targeted studies and only scattered information to be drawn by studies dedicated to the investigation of other genomics/transcriptomics elements.

To date, our attempt at classifying the transcriptional architecture of the tumor matrisome (Izzi et al. 2019) remains the most in-depth characterization of upstream regulomics. Moreover, we have recently tried to expand this survey to a more general concept linking the regulatory architecture of tumor matrisome to that of stem cells and suggesting several potential regulators that the two cell states might share, at least for specific collagens (Izzi et al. 2020). Notably, we and others seem to have come to the same conclusion on different matrisome elements too (Izzi et al. 2017a, b, c; Praktiknjo et al. 2020; Zhu et al. 2017), strengthening the bond between stemness and tumorigenesis in matrisome regulomics. Recent findings from positional and single-cell transcriptomics analyses, however, suggest an even more complex scenario where different areas of the same tumor and their surroundings—as well as different circulating tumor cells shed from primary neoplastic sites—express different matrisome genes (Bartoschek et al. 2018; Lim et al. 2019; Sathe et al. 2020; Ting et al. 2014), strongly advocating for a greater local heterogeneity than currently imagined and, hence, for a plethora of regulatory mechanisms!

Though extremely sparse, a literature on upstream regulomics mechanisms of the tumor matrisome exists: for example, a role for specific miRNAs, such as hsa-miR-767-5p, hsa-miR-487b-5p, hsa-miR-217, hsa-miR-1-3p and hsa-miR-133b, in regulating matrisome expression in metastatic colorectal carcinoma has been demonstrated and the same mechanism seems to also hold true for fibroblasts in pulmonary fibrosis and for healthy and diseased tendons (Ju et al. 2019; Mullenbrock et al. 2018; Thankam et al. 2016). Similarly, the possible effect of lncRNAs, eventually also differentially methylated, in affecting matrisome expression has been suggested for breast cancer (Heilmann et al. 2017; Li et al. 2018; Tomaru and Hayashizaki 2006), again in line with findings from ageing and non-neoplastic diseases (Mullin et al. 2017; Yang et al. 2016).

Methylation is, of course, another crucial regulator of gene transcription (Siegfried and Simon 2010). Surprisingly, reports about matrisome methylation, especially in cancer, are extremely fragmentary and scant: for example, in hepatocellular carcinoma, dermatopontin (DPT) has been reported to be downregulated by hypermethylation (Fu et al. 2014) and various ECM genes are targeted by methylation in colorectal cancer (Dong et al. 2019), similar to what has been reported for CAFs´ matrisome as well as for ageing (Li et al. 2017; Santi et al. 2018; Zhang et al. 2017).

Additionally, one should consider the potentially enormous role of matrisome mutations and single nucleotide variants (SNVs) in cancer, another topic that is surprisingly neglected and for which a clear definition of the effects is still far to come. Even more, one could consider how all these possible elements of regulation impact on the tumor matrisome: to date, one overarching study has tried to connect these layers together to explain gene expression across the whole transcriptome (Sharma et al. 2018). Rather surprisingly, ECM genes were frequently found among the top 5% of those whose expression cannot be explained by a combination of transcription factors, mutations, CNAs and methylation. The authors concluded that two factors might provide an explanation to the expression of these genes: (1) at least

for some ECM genes, somatic mutations within −10 kb to +1 kb of the TSS (Transcription Start Site, hereby loosely indicating the promoter region) might explain gene variations, though a variable fraction of gene expression remains unexplained, and (2) redundancies in cell pathways might mix up regulatory alterations, creating complex combinatorial patterns that eventually paint the tumor matrisome landscape at a finer scale.

Particularly this latter explanation closes in very well with the inter- and intratumoral heterogeneity of transcription factors and upstream regulatory mechanisms and shows the crucial importance of efforts to catalogue such factors to understand how the tumor matrisome gets shaped. And finally, for those investigators who would like to have more of a hands-on priming experience on this exciting field, we suggest pointing your browsers towards the Cancer Regulome portal (http://www.cancerregulome.org/).

7.3 Analysis of Cancer Matrisome at the Protein Level: Towards the Matrisome "Integrome"

7.3.1 Mass Spectrometry and Proteomics are Powerful Tools to Unveil Protein Composition in Tissue Samples

During the past few years, proteomics has become *the* method of choice to analyze the matrisome at the protein level. In general, the most used proteomic analysis strategy is the so-called discovery or shotgun proteomics which uses bottom-up mass spectrometry, where the proteins are first denatured and digested to peptides by a proteolytic enzyme or a chemical cleavage reagent before being submitted to mass spectrometry (Aebersold and Mann 2003). The most used enzyme is trypsin that cuts the protein after lysine or arginine except when proline immediately follows the cutting site. LysC is often used together with trypsin to complement digestion since it cuts between lysine and any other amino acid, including proline. In addition, unlike trypsin, LysC is active in high concentrations of urea, which is often used as a denaturing agent (Glatter et al. 2012). The digested peptides are then separated by reverse-phase HPLC chromatography and submitted inline to a tandem mass spectrometer (LC-MS/MS), where the peptide ions are first scanned (MS1), then fragmented, and the fragments are scanned (MS2). To identify the proteins, the MS2 spectra are computationally compared to theoretical spectra of peptides obtained by in silico digestion of proteins in a database (Aebersold and Mann 2003).

Beside identification of proteins in the sample, relative quantitation of proteins between samples is often desired, and this can be achieved in settings with or without labeling. SILAC labeling (Ong and Mann 2006) uses amino acids labeled by stable isotopes that are fed to cultured cells or even to whole animals (Krüger et al. 2008) before sample preparation. After labeling, the samples are mixed and submitted to LC-MS/MS. Quantitation is done from MS1 data by comparing the intensities of

different isotopic versions of the same peptide and, with SILAC, it is possible to differentially label up to four samples simultaneously. Another common labeling method uses isobaric tags such as iTRAQ or TMT that are attached to the extracted proteins or digested peptides before mass spectrometry (Ross et al. 2004; Thompson et al. 2003). After labeling, the samples are mixed and run by LC-MS/MS; during the MS1 phase the isobaric labels behave identically but, after fragmentation, the label releases a reporter ion with a specific m/z value for each sample. The intensities of different reporter ions in MS2 spectrum, thus, reflect differences between samples. The most recent commercially available iTRAQ and TMT reagents can be used to label up to eight and sixteen samples, respectively.

As for the label-free quantitation, there are two main approaches here, too: spectral counting and extracted ion chromatogram (XIC) based quantitation. In spectral counting the number of MS2 spectra matching to a certain protein are counted. In XIC based quantitation the ion intensity (usually the chromatographic peak area) of a peptide ion with a given m/z value and elution time is measured (Wang et al. 2008). To compare between different samples, the samples are run sequentially in the same conditions. The sequential runs are then aligned by elution time and m/z value and the ion intensities are compared, and identification from one sample can also be transferred to other samples (Cox et al. 2014). The advantage of XIC-based quantitation with transferred identification is that it can be, in theory, multiplexed to any number of samples.

There are many software tools for quantitation of data. Many of them are commercial, but there are also software such as MaxQuant (Tyanova et al. 2016) and OpenMS (Röst et al. 2016) which are free of charge and can be used for analyzing data from both label-based and label-free experiments.

Studies focused on the matrisome require, furthermore, special methods in sample preparation: to detect as many ECM proteins as possible, in fact, they need to be enriched somehow. The insoluble nature of ECM proteins requires harsh enrichment methods and, for example, decellularization by sodium dodecyl sulphate or other detergent or chaotropic salts have been used for enrichment of ECM in tissue samples—see (Taha and Naba 2019) for review; the outcomes and features of different enrichment strategies have also been systematically compared (Krasny et al. 2016). For cell cultures, also more subtle decellularization methods such as freeze-thaw cycling or trypsinization have been used (Baroncelli et al. 2018; Lee et al. 2019; Mao et al. 2019). Cultured cells can also be removed by hypotonic treatment. For example, we have successfully applied hypotonic lysis for enrichment of extracellular matrix from various 2D and 3D cell cultures of fibroblasts and cancer cells or their co-cultures (Ojalill et al. 2018a, b, c, 2020; Siljamäki et al. 2020). Since enrichment does not, however, remove all cellular proteins the detected proteins need to be filtered for ECM proteins and the usual matrisome lists (Naba et al. 2016) available at http://matrisome.org/ have greatly facilitated filtering the detected proteins for ECM proteins both in our and other studies.

Tissue-derived ECM components can be difficult to dissolve completely in commonly used denaturing agent such as urea or guanidine hydrochloride. However, digestion and subsequent proteomic analysis of a urea soluble fraction of

incompletely dissolved matrix gives very similar results compared to those obtained from digestion of insoluble fraction. In addition, the majority of proteins in an insoluble fraction seems to consist of highly abundant collagen I, III and VI, and their depletion in the soluble fraction can in fact increase the number and diversity of ECM proteins detected (Naba et al. 2017a, b). Thus, at least in quantitative experiments, analyzing the soluble fraction seems to be preferable and, in fact, differential solubility has also been exploited to fractionate ECM proteins (Schiller et al. 2015).

On top of all these difficulties, many ECM proteins are glycosylated, and many prolines and lysines of collagens are hydroxylated. Glycosylation of ECM proteins can decrease solubility and susceptibility to enzymatic digestion. Therefore, many studies have routinely included deglycosylation by PNGaseF treatment before or during enzymatic digestion of ECM proteins—see (Taha and Naba 2019) for review—though it should be kept in mind that PNGaseF causes deamidation of asparagine to aspartate when the N-linked glycosylation is removed. To optimize the coverage of ECM proteins, then, these modifications should be included to the possible amino acid modifications in the search parameters of the identification software.

In addition to hydroxylation and glycosylation, it has also been shown that extracellular proteins can be phosphorylated (Yalak and Vogel 2012) and citrullinated (Sipilä et al. 2017). These modifications may have a significant impact in cancer and other diseases (Yalak and Vogel 2012; Yuzhalin et al. 2018a, b) but, till now, not a large majority of studies on the tumor matrisome have provided these information to the wider audience and it is therefore becoming increasingly important to specifically analyze these modifications in the context of the extracellular matrix.

7.3.2 Proteomics Data Repositories and Portals

Today it is usually mandatory to submit the raw mass spectrometry data along with their metadata to a public repository such as those under the ProteomeXchange consortium (Deutsch et al. 2017) when publishing a study involving proteomics. This has opened the possibility to reanalyze already-published data using different strategies, for example searching for PTMs not addressed in the original publication. To browse repositories within the ProteomeXchange consortium one can, for example, resort to ProteomeCentral (http://proteomecentral.proteomexchange.org). The Clinical Proteomic Tumor Analysis Consortium (CPTAC) has also made, and continues to make, its data publicly available (Edwards et al. 2015) and this is a great opportunity to complement TCGA data at the genomics, epigenomics and transcriptomics level since the proteomics data available from TCGA are importantly limited (*and the bravest of readers might even want to try investigating the genes from our guided walkthrough in this portal!*).

However, analysis of raw (or anyway primary) data is complex and can be especially challenging for researchers not familiar with mass spectrometry. Thus,

there is a pressing need for easily accessible tools that can be used to mine and analyze publicly available data. A great example of this kind of tools is the MatrisomeDB (http://matrisomedb.pepchem.org/) developed by the Naba group, that tries to address this need in the matrisome field (Shao et al. 2020). The database has been constructed by analyzing the raw data of 17 studies where decellularization has been used to enrich ECM before mass spectrometry, and the analysis has been conducted using equivalent parameters for all studies (taking different labeling strategies etc. into account), allowing multiple different PTMs (e.g. hydroxylation, deamidation, phosphorylation, citrullination) and applying mass-tolerant database search (Chick et al. 2015). The database is expandable, and the authors welcome contributions.

7.3.3 Proteomics Reveal the Basic Components of the Tumor Matrisome

Several published studies have proven the principle that mass spectrometry and proteomics are essential methods to unveil novel features in the composition of ECM in tumor stroma. These tools have been widely used to analyze patient-derived tissue samples (Glavey et al. 2017; Gocheva et al. 2017; Naba et al. 2014b), mouse tumors (Barrett et al. 2018), xenografts (Naba et al. 2014a; Tian et al. 2019) as well as in vitro experimental models such as spheroids (Ojalill et al. 2018a, b, c; Siljamäki et al. 2020) and targeted approaches (Tomko et al. 2018).

The analysis of ECM in tumor stroma by mass spectrometry has also revealed some basic facts of matrisome production and accumulation in cancer that challenge common views of both typical ECM sources and primary/metastatic tumors. For example, we now know that despite mesenchymal, fibroblast-like cells are the main source of ECM proteins, also carcinoma cells themselves seem to importantly contribute to the process (Naba et al. 2012, 2014a; Tian et al. 2019). Moreover, the proteins released by cancer cells may further regulate the progression of the disease (Tian et al. 2019). Also, interestingly, the matrisome landscape of liver metastases from colorectal cancer has been shown to more closely resemble the matrisome of primary colorectal tumors, rather than the liver matrisome (Naba et al. 2014b), and this observation also stresses the importance of cancer cells in the regulation of ECM accumulation in tumors.

The proteomics-based analyses of human tumors have often aimed at the identification of novel biomarkers that could be used for prognostication, for the selection of therapies with the highest chance to achieve results or for the identification of novel drug targets. No universal markers of cancer matrisome have been found so far, possibly owing to the redundancy of ECM proteins in tumors already discussed in Paragraph 2.3, but published papers contain several interesting observations that have spurred further studies and findings from proteomics and have led to a more organic understanding of the role of specific matrisome proteins in cancer (Tian et al.

2020). Much alike transcriptomics, several specific expression patterns of ECM proteins have been published for different cancers, with the potential to be useful in the prediction of cancer progression. Still, even though mass spectrometric analysis of cancer matrisome is an important method to analyze human tumors, significantly larger patient numbers should be analyzed before specific ECM signatures could be introduced in clinical practice.

In addition, the development of novel tools in proteomics provides exciting new opportunities for more detailed analysis of cancer matrisome and for the understanding of yet unappreciated elements within. For example, proteolytical processing of ECM components generates biologically active polypeptides—usually referred to as matricryptins—whose eventual roles and importance in tumors are, for the largest part, completely unknown. The best known of them is endostatin, an angiogenesis regulator derived from collagen XVIII (Brassart-Pasco et al. 2020). Recognition of the active proteolytic processes as well as the resulting bioactive degradation products would create new opportunities for prediction of tumor behavior (Rogers and Overall 2013; Zhang et al. 2019). Mass spectrometry is also an optimal method to identify posttranslational modifications (PTMs) in ECM proteins, like collagens (Merl-Pham et al. 2019). for example, the hypoxia-induced expression of lysyl oxidase and the consequently increased hydroxylation of lysine residues in collagens may promote the formation of bone metastasis in breast cancer (Cox et al. 2015). The biological effects of most PTMs, such as citrullination and phosphorylation, in the tumor matrisome are potentially very interesting but, as we already stated, mostly unknown.

The mass spectrometric analysis of tumor tissues has provided valuable information about the composition of ECM in cancer and, recently, complementary proteomics approaches in controlled in vitro microenvironments have been used to dissect the diverse contributions of different cell types to the cancer matrisome. Cell culture models obviously lack e.g. blood vessels and inflammatory cells and it would be therefore important to know how similar or different the ECM composition of a given tumor or tissue (or even a single cell type) is when produced in vitro or in/ex vivo. In our own studies we have tried to answer this question by comparing the ECM extracted from human prostate tumors to the ECM produced in vitro by isolated prostate cancer-derived fibroblasts or by the same fibroblasts co-cultured together with prostate cancer cells in spheroids (Ojalill et al. 2018a, b, c). What we observed was a sharp contrast in between the "core matrisome" (basically, the structural ECM proteins) and the ECM-associated ones. The basic structural proteins in the ECM are, in fact, very much the same in all the setups. When isolated fibroblasts are cultured in monolayers in the presence of ascorbic acid they produce collagen I and many other proteins associated to collagen fibrils and loose connective tissue; these same components could be found in spheroids as well as in ex vivo tumor stroma (Ojalill et al. 2018a, b, c). However, major differences can be observed in ECM-associated proteins for which we could recognize several gene products that modify the ECM in cell cultures, suggesting that the classical cell culture conditions activate ECM renewal whereas, in tissues, the ECM is more stationary. Despite the fact there were differences between fibroblast monolayers and spheroids,

furthermore, we could not conclude that ECM in spheroids would better mimic tissue ECM than the one from monolayers. Notably, many plasma-derived proteins can also associate to the ECM and these were abundant in tissue sample but naturally not present in serum-free cell cultures (Ojalill et al. 2018a, b, c). Cell cultures can, furthermore, be a favorable setup for studying reciprocal cell-to-cell interactions or to dissect the differential contributions of subtle cell types to the tumor matrisome. For example, we have recently analyzed spheroid cultures of skin fibroblasts together with transformed keratinocytes harboring activating RAS mutation and evidenced how the different cell types seem to regulate each other's ECM production, with a particularly important effect on the accumulation of basement membrane proteins, such as laminins (Siljamäki et al. 2020). Another recently published report indicates that subtypes within broadly defined CAFs may exhibit significant differences in the production of ECM components (Brechbuhl et al. 2020). Fibroblasts that are CD146–, in fact, seem to promote the metastatization of human breast cancer cells when compared to CD146+ fibroblasts. Quantitative and qualitative proteomic analyses showed that CD146+ fibroblasts produced, to the vantage of cancer cells, an environment rich in basement membrane proteins, while CD146– fibroblasts exhibited increases in fibronectin, lysyl oxidase, and tenascin C.

7.3.4 Connecting Proteomics to Other Research Methods in Tumor Matrisome

Individual ECM components may play an important role in the progression of cancer. However, the general organization of the ECM and especially its stiffness have been shown to be even more important (Levental et al. 2009). While ECM composition is also a regulator of its own organization, proteomics methods alone are insufficient to give a complete picture of the tumor matrisome. Therefore, the mass spectrometric analysis of the matrisome has also been connected to other methods such as RNA sequencing and microscopy. These approaches have resulted in a more complete image of the complex relationship between ECM composition and organization (Carpino et al. 2019; Mayorca-Guiliani et al. 2017) and helped to develop novel methods in prognostication (Pearce et al. 2018a, b).

In situ decellularization of tissues (ISDoT) is a method that allows both high-resolution imaging and proteomic analysis of native extracellular matrix (Mayorca-Guiliani et al. 2017). In this method the tissue sample, or even a whole mouse organ, is decellularized in a way that leaves the native ECM architecture intact. In the original publication, the authors then analyzed the three-dimensional decellularized tissues using high-resolution fluorescence, second harmonic imaging, and quantitative proteomics. In a mouse model, they followed primary tumor development and progression to metastasis and concluded that cancer-driven ECM remodeling is organ-specific (Mayorca-Guiliani et al. 2017). Combination of histology and mass spectrometry to analyze ECM in the intrahepatic cholangiocarcinoma has also

unveiled the interesting association of collagen III alpha1 chain (COL3A1) expression with high levels of collagen fibers and low abundance of reticular and elastic fibers, suggesting that these biochemical and organizational ECM clues may contribute to the stiffness of the matrix and the promotion of cell migration (Carpino et al. 2019). To an even finer level of detail, Pearce et al. (2018a, b) have studied human metastatic microenvironment using biopsies of high-grade serous ovarian cancer metastases. Using a single sample they were able to combine RNA sequencing, matrisome proteomics, cytokine and chemokine levels, cellularity, ECM organization and biomechanical properties, and the multivariate integration of the different components allowed to recognize profiles that predicted both the extent of disease and the corresponding tissue stiffness (Pearce et al. 2018a, b).

These recent papers have clearly shown the benefits of the concomitant use of distinct methods. It is tempting to predict that, in the future, similar approaches will be used to analyze many different cancer types in parallel as well as larger numbers of patients. These studies may produce a more precise picture of the relationship of ECM composition and organization. Furthermore, we can also expect novel and accurate signatures of matrisome proteins that could be valuable tools in clinical practice.

7.4 Future Perspectives

The composition and functions of the tumor matrisome is diverse and heterogenous, with both predicted local niches of specialization depending on different cell types and functional cell states as well as observed global "themes" and processes that cross (and associate) different tumors together. Mapping the incredible complexity of this system is a challenge that spans all the layers of omics data and impinges on cell and tissue research, metabolic and systemic connections all the way up to entire organisms and species.

The extraction of biochemical, functional and organizational clues from each level of information and the findings that tie these elements together into a common space that reveals the molecular, cellular and supracellular dimensions of the tumor matrisome is propelling us into a new era of research, one in which we will hopefully be able to grasp the global structure of this system and exploit this knowledge to devise new, effective antineoplastic therapies and prognostic/diagnostic methods.

References

Aebersold R, Mann M (2003) Mass spectrometry-based proteomics. Nature 422:198–207
Anastassiou D (2017) Cancer letters. Cancer Lett 393:125–126

Andriani F, Landoni E, Mensah M, Facchinetti F, Miceli R, Tagliabue E, Giussani M, Callari M, De Cecco L, Colombo MP et al (2018) Diagnostic role of circulating extracellular matrix-related proteins in non-small cell lung cancer. BMC Cancer 18:899

Bailey MH, Tokheim C, Porta-Pardo E, Sengupta S, Bertrand D, Weerasinghe A, Colaprico A, Wendl MC, Kim J, Reardon B et al (2018) Comprehensive characterization of Cancer driver genes and mutations. Cell 173:371–385.e18

Bansal AK, Singh LR, Kamli MR (2019) Chapter 9 – Posttranslational modifications associated with Cancer and their therapeutic implications. In: Dar TA, Singh LR (eds) Protein modificomics. Academic Press, Cambridge, pp 203–227

Bao Y, Wang L, Shi L, Yun F, Liu X, Chen Y, Chen C, Ren Y, Jia Y (2019) Transcriptome profiling revealed multiple genes and ECM-receptor interaction pathways that may be associated with breast cancer. Cell Mol Biol Lett 24:38

Baroncelli M, van der Eerden BCJ, Chatterji S, Rull Trinidad E, Kan YY, Koedam M, van Hengel IAJ, Alves RDAM, Fratila-Apachitei LE, Demmers JAA, van de Peppel J, van Leeuwen JPTM (2018) Human osteoblast-derived extracellular matrix with high homology to bone proteome is Osteopromotive. Tissue Eng Part A 24:1377–1389

Barrett AS, Maller O, Pickup MW, Weaver VM, Hansen KC (2018) Compartment resolved proteomics reveals a dynamic matrisome in a biomechanically driven model of pancreatic ductal adenocarcinoma. J Immunol Regen Med 1:67–75

Bartoschek M, Oskolkov N, Bocci M, Lövrot J, Larsson C, Sommarin M, Madsen CD, Lindgren D, Pekar G, Karlsson G et al (2018) Spatially and functionally distinct subclasses of breast cancer-associated fibroblasts revealed by single cell RNA sequencing. Nat Commun 9:1–13

Bergamaschi A, Tagliabue E, Sørlie T, Naume B, Triulzi T, Orlandi R, Russnes HG, Nesland JM, Tammi R, Auvinen P et al (2008) Extracellular matrix signature identifies breast cancer subgroups with different clinical outcome. J Pathol 214:357–367

Bradner JE, Hnisz D, Young RA (2017) Transcriptional addiction in Cancer. Cell 168:629–643

Brandt S, Montagna C, Georgis A, Schüffler PJ, Bühler MM, Seifert B, Thiesler T, Curioni-Fontecedro A, Hegyi I, Dehler S et al (2013) The combined expression of the stromal markers fibronectin and SPARC improves the prediction of survival in diffuse large B-cell lymphoma. Exp Hematol Oncol 2:27

Brassart-Pasco S, Brézillon S, Brassart B, Ramont L, Oudart J, Monboisse JC (2020) Tumor microenvironment: extracellular matrix alterations influence tumor progression. Front Oncol 10:397

Brechbuhl HM, Barrett AS, Kopin E, Hagen JC, Han AL, Gillen AE, Finlay-Schultz J, Cittelly DM, Owens P, Horwitz KB et al (2020) Fibroblast subtypes define a metastatic matrisome in breast cancer. JCI Insight 5(4)

Carpino G, Overi D, Melandro F, Grimaldi A, Cardinale V, Di Matteo S, Mennini G, Rossi M, Alvaro D, Barnaba V, Gaudio E, Mancone C (2019) Matrisome analysis of intrahepatic cholangiocarcinoma unveils a peculiar cancer-associated extracellular matrix structure. Clin Proteomics 16:37

Casamassimi A, Ciccodicola A (2019) Transcriptional regulation: molecules, involved mechanisms, and Misregulation. Int J Mol Sci 20:1281

Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E et al (2012) The cBio Cancer genomics portal: an open platform for exploring multidimensional Cancer genomics data. Cancer Discov 2:401–404

Cerchietti L, Inghirami G, Kotlov N, Svekolkin V, Bagaev A, Frenkel F, Revuelta MV, Phillip JM, Cacciapuoti MT, Rutherford SC, Martin P, Leonard JP (2019) Microenvironmental signatures reveal biological subtypes of diffuse large B-cell lymphoma (DLBCL) distinct from tumor cell molecular profiling. Blood 134:656–656

Chakravarthy A, Khan L, Bensler NP, Bose P, De Carvalho DD (2018) TGF-β-associated extracellular matrix genes link cancer-associated fibroblasts to immune evasion and immunotherapy failure. Nat Commun 9(1). https://doi.org/10.1038/s41467-018-06654-8

Cheon D, Tong Y, Sim M, Dering J, Berel D, Cui X, Lester J, Beach JA, Tighiouart M, Walts AE, Karlan BY, Orsulic S (2014) A collagen-remodeling gene signature regulated by TGF-β signaling is associated with metastasis and poor survival in serous ovarian cancer. Clin Cancer Res 20:711–723

Chick JM, Kolippakkam D, Nusinow DP, Zhai B, Rad R, Huttlin EL, Gygi SP (2015) A mass-tolerant database search identifies a large proportion of unassigned spectra in shotgun proteomics as modified peptides. Nat Biotechnol 33:743–749

Ciasca G, Sassun TE, Minelli E, Antonelli M, Papi M, Santoro A, Giangaspero F, Delfini R, De Spirito M (2016) Nano-mechanical signature of brain tumours. Nanoscale 8:19629–19643

Cox J, Hein MY, Luber CA, Paron I, Nagaraj N, Mann M (2014) Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. Mol Cell Proteomics 13:2513–2526

Cox TR, Rumney RMH, Schoof EM, Perryman L, Høye AM, Agrawal A, Bird D, Latif NA, Forrest H, Evans HR et al (2015) The hypoxic cancer secretome induces pre-metastatic bone lesions through lysyl oxidase. Nature 522:106–110

Deutsch EW, Csordas A, Sun Z, Jarnuczak A, Perez-Riverol Y, Ternent T, Campbell DS, Bernal-Llinares M, Okuda S, Kawano S et al (2017) The ProteomeXchange consortium in 2017: supporting the cultural change in proteomics public data deposition. Nucleic Acids Res 45: D1100–D1106

Dong L, Ma L, Ma GH, Ren H (2019) Genome-wide analysis reveals DNA methylation alterations in obesity associated with high risk of colorectal Cancer. Sci Rep 9:5100

Editorial (2020) The era of massive cancer sequencing projects has reached a turning point. Nature 578:7–8

Edwards NJ, Oberti M, Thangudu RR, Cai S, McGarvey PB, Jacob S, Madhavan S, Ketchum KA (2015) The CPTAC data portal: A resource for Cancer proteomics research. J Proteome Res 14:2707–2713

Finkernagel F, Reinartz S, Lieber S, Adhikary T, Wortmann A, Hoffmann N, Bieringer T, Nist A, Stiewe T, Jansen JM et al (2016) The transcriptional signature of human ovarian carcinoma macrophages is associated with extracellular matrix reorganization. Oncotarget 7:75339–75352

Foroushani A, Agrahari R, Docking R, Chang L, Duns G, Hudoba M, Karsan A, Zare H (2017) Large-scale gene network analysis reveals the significance of extracellular matrix pathway and homeobox genes in acute myeloid leukemia: an introduction to the Pigengene package and its applications. BMC Med Genet 10:16–16

Frezzetti D, De Luca A, Normanno N (2019) Extracellular matrix proteins as circulating biomarkers for the diagnosis of non-small cell lung cancer patients. J Thorac Dis 11:S1252–S1256

Fu Y, Feng M, Yu J, Ma M, Liu X, Li J, Yang X, Wang Y, Zhang Y, Ao J et al (2014) DNA methylation-mediated silencing of matricellular protein dermatopontin promotes hepatocellular carcinoma metastasis by α3β1 integrin-rho GTPase signaling. Oncotarget 5:6701–6715

Gao GF, Parker JS, Reynolds SM, Silva TC, Wang L, Zhou W, Akbani R, Bailey M, Balu S, Berman BP et al (2019) Before and after: comparison of legacy and harmonized TCGA genomic data commons' data. Cell Syst 9:24–34.e10

Ghedira K (2018) Introductory chapter: a brief overview of transcriptional and post-transcriptional regulation. Transcriptional and Post-Transcriptional Regulation

Glatter T, Ludwig C, Ahrné E, Aebersold R, Heck AJR, Schmidt A (2012) Large-scale quantitative assessment of different in-solution protein digestion protocols reveals superior cleavage efficiency of tandem Lys-C/trypsin proteolysis over trypsin digestion. J Proteome Res 11:5145–5156

Glavey SV, Naba A, Manier S, Clauser K, Tahri S, Park J, Reagan MR, Moschetta M, Mishima Y, Gambella M et al (2017) Proteomic characterization of human multiple myeloma bone marrow extracellular matrix. Leukemia 31:2426–2434

Gocheva V, Naba A, Bhutkar A, Guardia T, Miller KM, Li CM, Dayton TL, Sanchez-Rivera FJ, Kim-Kiselak C, Jailkhani N et al (2017) Quantitative proteomics identify tenascin-C as a

promoter of lung cancer progression and contributor to a signature prognostic of patient survival. Proc Natl Acad Sci U S A 114:E5625–E5634

Heilmann K, Toth R, Bossmann C, Klimo K, Plass C, Gerhauser C (2017) Genome-wide screen for differentially methylated long noncoding RNAs identifies Esrp2 and lncRNA Esrp2-as regulated by enhancer DNA methylation with prognostic relevance for human breast cancer. Oncogene 36:6446–6461

Helleman J, Jansen MPHM, Ruigrok-Ritstier K, van Staveren IL, Look MP, Meijer-van Gelder ME, Sieuwerts AM, Klijn JGM, Sleijfer S, Foekens JA, Berns EMJJ (2008) Association of an Extracellular Matrix Gene Cluster with breast Cancer prognosis and endocrine therapy response. Clin Cancer Res 14:5555–5564

Henke E, Nandigama R, Ergün S (2020) Extracellular matrix in the tumor microenvironment and its impact on Cancer therapy. Front Mol Biosci 6:160

Hinshaw DC, Shevde LA (2019) The tumor microenvironment innately modulates Cancer progression. Cancer Res 79:4557–4566

Hoadley KA, Yau C, Wolf DM, Cherniack AD, Tamborero D, Ng S, Leiserson MDM, Niu B, McLellan MD, Uzunangelov V et al (2014) Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. Cell 158:929–944

Hoadley KA, Yau C, Hinoue T, Wolf DM, Lazar AJ, Drill E, Shen R, Taylor AM, Cherniack AD, Thorsson V et al (2018) Cell-of-origin patterns dominate the molecular classification of 10,000 tumors from 33 types of Cancer. Cell 173:291–304.e6

Huang K, Mashl RJ, Wu Y, Ritter DI, Wang J, Oh C, Paczkowska M, Reynolds S, Wyczalkowski MA, Oak N et al (2018) Pathogenic germline variants in 10,389 adult cancers. Cell 173:355–370.e14

Izzi V, Lakkala J, Devarajan R, Ruotsalainen H, Savolainen ER, Koistinen P, Heljasvaara R, Pihlajaniemi T (2017a) An extracellular matrix signature in leukemia precursor cells and acute myeloid leukemia. Haematologica 102:e245–e248

Izzi V, Lakkala J, Devarajan R, Ruotsalainen H, Savolainen ER, Koistinen P, Heljasvaara R, Pihlajaniemi T (2017b) An extracellular matrix signature in leukemia precursor cells and acute myeloid leukemia. Haematologica 102:1807–1809

Izzi V, Lakkala J, Devarajan R, Savolainen ER, Koistinen P, Heljasvaara R, Pihlajaniemi T (2017c) Expression of a specific extracellular matrix signature is a favorable prognostic factor in acute myeloid leukemia. Leuk Res Rep 9:9–13

Izzi V, Lakkala J, Devarajan R, Kääriäinen A, Koivunen J, Heljasvaara R, Pihlajaniemi T (2019) Pan-Cancer analysis of the expression and regulation of matrisome genes across 32 tumor types. Matrix Biology Plus 1:100004

Izzi V, Heljasvaara R, Heikkinen A, Karppinen S, Koivunen J, Pihlajaniemi T (2020) Exploring the roles of MACIT and multiplexin collagens in stem cells and cancer. Semin Cancer Biol 62:134–148

Jansen M, Foekens J, Van Staveren I, Dirkszwager-Kiel M, Ritstier K, Look M, Meijer-Van Gelder M, Portengen H, Dorssers L, Klijn J, Berns E (2004) Molecular classification of tamoxifen resistant breast carcinomas by gene expression profiling. Cancer Res 64:657

Jia D, Liu Z, Deng N, Tan TZ, Huang RY, Taylor-Harding B, Cheon DJ, Lawrenson K, Wiedemeyer WR, Walts AE, Karlan BY, Orsulic S (2016) A COL11A1-correlated pan-cancer gene signature of activated fibroblasts for the prioritization of therapeutic targets. Cancer Lett 382:203–214

Jinka R, Kapoor R, Sistla PG, Raj TA, Pande G (2012) Alterations in cell-extracellular matrix interactions during progression of cancers. Int J Cell Biol 2012:219196

Ju Q, Zhao Y, Dong Y, Cheng C, Zhang S, Yang Y, Li P, Ge D, Sun B (2019) Identification of a miRNA-mRNA network associated with lymph node metastasis in colorectal cancer. Oncol Lett 18:1179–1188

Kim H, Watkinson J, Varadan V, Anastassiou D (2010) Multi-cancer computational analysis reveals invasion-associated variant of desmoplastic reaction involving INHBA, THBS2 and COL11A1. BMC Med Genet 3:51

Kobayashi N, Oda T, Takizawa M, Ishizaki T, Tsukamoto N, Yokohama A, Takei H, Saitoh T, Shimizu H, Honma K et al (2020) Integrin α7 and extracellular matrix laminin 211 interaction promotes proliferation of acute myeloid leukemia cells and is associated with granulocytic sarcoma. Cancers (Basel) 12:363

Krasny L, Paul A, Wai P, Howard BA, Natrajan RC, Huang PH (2016) Comparative proteomic assessment of matrisome enrichment methodologies. Biochem J 473:3979–3995

Krüger M, Moser M, Ussar S, Thievessen I, Luber CA, Forner F, Schmidt S, Zanivan S, Fässler R, Mann M (2008) SILAC mouse for quantitative proteomics uncovers Kindlin-3 as an essential factor for red blood cell function. Cell 134:353–364

Lavanya V, Jamal S, Ahmed N (2019) Chapter 5 – Ubiquitin mediated posttranslational modification of proteins involved in various signaling diseases. In: Dar TA, Singh LR (eds) Protein modificomics. Academic Press, Cambridge, pp 109–130

Lee KJ, Comerford EJ, Simpson DM, Clegg PD, Canty-Laird EG (2019) Identification and characterization of canine ligament progenitor cells and their extracellular matrix niche. J Proteome Res 18:1328–1339

Lelli KM, Slattery M, Mann RS (2012) Disentangling the many layers of eukaryotic transcriptional regulation. Annu Rev Genet 46:43–68

Levental KR, Yu H, Kass L, Lakins JN, Egeblad M, Erler JT, Fong SFT, Csiszar K, Giaccia A, Weninger W et al (2009) Matrix crosslinking forces tumor progression by enhancing integrin signaling. Cell 139:891–906

Li S, Christiansen L, Christensen K, Kruse TA, Redmond P, Marioni RE, Deary IJ, Tan Q (2017) Identification, replication and characterization of epigenetic remodelling in the aging genome: a cross population analysis. Sci Rep 7:8183

Li J, Wang W, Xia P, Wan L, Zhang L, Yu L, Wang L, Chen X, Xiao Y, Xu C (2018) Identification of a five-lncRNA signature for predicting the risk of tumor recurrence in patients with breast cancer. Int J Cancer 143:2150–2160

Lim SB, Tan SJ, Lim WT, Lim CT (2017) An extracellular matrix-related prognostic and predictive indicator for early-stage non-small cell lung cancer. Nat Commun 8:1734–1736

Lim SB, Tan SJ, Lim W, Lim CT (2018) A merged lung cancer transcriptome dataset for clinical predictive modeling. Scientific Data 5:180136

Lim SB, Yeo T, Lee WD, Bhagat AAS, Tan SJ, Tan DSW, Lim W, Lim CT (2019) Addressing cellular heterogeneity in tumor and circulation for refined prognostication. Proc Natl Acad Sci U S A 116:17957–17962

Malandrino A, Mak M, Kamm RD, Moeendarbary E (2018) Complex mechanics of the heterogeneous extracellular matrix in cancer. Extreme Mech Lett 21:25–34

Malta TM, Sokolov A, Gentles AJ, Burzykowski T, Poisson L, Weinstein JN, Kamińska B, Huelsken J, Omberg L, Gevaert O et al (2018) Machine learning identifies Stemness features associated with oncogenic dedifferentiation. Cell 173:338–354.e15

Mao Y, Block T, Singh-Varma A, Sheldrake A, Leeth R, Griffey S, Kohn J (2019) Extracellular matrix derived from chondrocytes promotes rapid expansion of human primary chondrocytes in vitro with reduced dedifferentiation. Acta Biomater 85:75–83

Mayorca-Guiliani AE, Madsen CD, Cox TR, Horton ER, Venning FA, Erler JT (2017) ISDoT: in situ decellularization of tissues for high-resolution imaging and proteomic analysis of native extracellular matrix. Nat Med 23:890–898

Merl-Pham J, Basak T, Knüppel L, Ramanujam D, Athanason M, Behr J, Engelhardt S, Eickelberg O, Hauck SM, Vanacore R, Staab-Weijnitz CA (2019) Quantitative proteomic profiling of extracellular matrix and site-specific collagen post-translational modifications in an in vitro model of lung fibrosis. Matrix Biology Plus 1:100005

Mullenbrock S, Liu F, Szak S, Hronowski X, Gao B, Juhasz P, Sun C, Liu M, McLaughlin H, Xiao Q, Feghali-Bostwick C, Zheng TS (2018) Systems analysis of transcriptomic and proteomic profiles identifies novel regulation of fibrotic programs by miRNAs in pulmonary fibrosis fibroblasts. Genes (Basel) 9:588

Mullin NK, Mallipeddi NV, Hamburg-Shields E, Ibarra B, Khalil AM, Atit RP (2017) Wnt/β-catenin signaling pathway regulates specific lncRNAs that impact dermal fibroblasts and skin fibrosis. Front Genet 8:183

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

Naba A, Clauser KR, Lamar JM, Carr SA, Hynes RO (2014a) Extracellular matrix signatures of human mammary carcinoma identify novel metastasis promoters. elife 3:e01308

Naba A, Clauser KR, Whittaker CA, Carr SA, Tanabe KK, Hynes RO (2014b) Extracellular matrix signatures of human primary metastatic colon cancers and their metastases to liver. BMC Cancer 14:518–518

Naba A, Clauser KR, Ding H, Whittaker CA, Carr SA, Hynes RO (2016) The extracellular matrix: tools and insights for the "omics" era. Matrix Biol 49:10–24

Naba A, Clauser KR, Mani DR, Carr SA, Hynes RO (2017a) Quantitative proteomic profiling of the extracellular matrix of pancreatic islets during the angiogenic switch and insulinoma progression. Sci Rep 7:40495

Naba A, Pearce OMT, Del Rosario A, Ma D, Ding H, Rajeeve V, Cutillas PR, Balkwill FR, Hynes RO (2017b) Characterization of the extracellular matrix of Normal and diseased tissues using proteomics. J Proteome Res 16:3083–3091

Najafi M, Farhood B, Mortezaee K (2019) Extracellular matrix (ECM) stiffness and degradation as cancer drivers. J Cell Biochem 120:2782–2790

Nath S, Ghatak D, Das P, Roychoudhury S (2015) Transcriptional control of mitosis: deregulation and Cancer. Front Endocrinol 6:60

Nawy T (2018) A pan-cancer atlas. Nat Methods 15:407

Ojalill M, Parikainen M, Rappu P, Aalto E, Jokinen J, Virtanen N, Siljamaki E, Heino J (2018a) Integrin alpha2beta1 decelerates proliferation, but promotes survival and invasion of prostate cancer cells. Oncotarget 9:32435–32447

Ojalill M, Rappu P, Siljamaki E, Taimen P, Bostrom P, Heino J (2018b) The composition of prostate core matrisome in vivo and in vitro unveiled by mass spectrometric analysis. Prostate 78:583–594

Ojalill M, Rappu P, Siljamäki E, Taimen P, Boström P, Heino J (2018c) The composition of prostate core matrisome in vivo and in vitro unveiled by mass spectrometric analysis. Prostate 78:583–594

Ojalill M, Virtanen N, Rappu P, Siljamäki E, Taimen P, Heino J (2020) Interaction between prostate cancer cells and prostate fibroblasts promotes accumulation and proteolytic processing of basement membrane proteins. Prostate 80:715–726

Ong S, Mann M (2006) A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC). Nat Protoc 1:2650–2660

Pang X, Xie R, Zhang Z, Liu Q, Wu S, Cui Y (2019) Identification of SPP1 as an extracellular matrix signature for metastatic castration-resistant prostate Cancer. Front Oncol 9:924

Pearce OMT, Delaine-Smith R, Maniati E, Nichols S, Wang J, Böhm S, Rajeeve V, Ullah D, Chakravarty P, Jones RR et al (2018a) Deconstruction of a metastatic tumor microenvironment reveals a common matrix response in human cancers. Cancer Discov 8:304–319

Pearce OMT, Delaine-Smith R, Maniati E, Nichols S, Wang J, Böhm S, Rajeeve V, Ullah D, Chakravarty P, Jones RR et al (2018b) Deconstruction of a metastatic tumor microenvironment reveals a common matrix response in human cancers. Cancer Discov 8:304–319

Peeney D, Fan Y, Nguyen T, Meerzaman D, Stetler-Stevenson WG (2019) Matrisome-associated gene expression patterns correlating with TIMP2 in Cancer. Sci Rep 9:1–14

Peixoto P, Etcheverry A, Aubry M, Missey A, Lachat C, Perrard J, Hendrick E, Delage-Mourroux-R, Mosser J, Borg C et al (2019) EMT is associated with an epigenetic signature of ECM remodeling genes. Cell Death Dis 10:1–17

Penet M, Kakkad S, Pathak AP, Krishnamachary B, Mironchik Y, Raman V, Solaiyappan M, Bhujwalla ZM (2017) Structure and function of a prostate Cancer dissemination–permissive extracellular matrix. Clin Cancer Res 23:2245–2254

Poltavets V, Kochetkova M, Pitson SM, Samuel MS (2018) The role of the extracellular matrix and its molecular and cellular regulators in Cancer cell plasticity. Front Oncol 8:431

Praktiknjo SD, Obermayer B, Zhu Q, Fang L, Liu H, Quinn H, Stoeckius M, Kocks C, Birchmeier W, Rajewsky N (2020) Tracing tumorigenesis in a solid tumor model at single-cell resolution. Nat Commun 11:1–12

Qin X, Xu H, Gong W, Deng W (2015) The tumor cytosol miRNAs, fluid miRNAs, and exosome miRNAs in lung Cancer. Front Oncol 4:357

Rogers LD, Overall CM (2013) Proteolytic post-translational modification of proteins: proteomic tools and methodology. Mol Cell Proteomics 12:3532–3542

Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, Hattan S, Khainovski N, Pillai S, Dey S, Daniels S et al (2004) Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. Mol Cell Proteomics 3:1154–1169

Röst HL, Sachsenberg T, Aiche S, Bielow C, Weisser H, Aicheler F, Andreotti S, Ehrlich H, Gutenbrunner P, Kenar E et al (2016) OpenMS: a flexible open-source software platform for mass spectrometry data analysis. Nat Methods 13:741–748

Sanchez-Vega F, Mina M, Armenia J, Chatila WK, Luna A, La KC, Dimitriadoy S, Liu DL, Kantheti HS, Saghafinia S et al (2018) Oncogenic signaling pathways in the Cancer genome atlas. Cell 173:321–337.e10

Santi A, Kugeratski FG, Zanivan S (2018) Cancer associated fibroblasts: the architects of stroma remodeling. Proteomics 18:e1700167

Sathe A, Grimes SM, Lau BT, Chen J, Suarez C, Huang RJ, Poultsides G, Ji HP (2020) Single-cell genomic characterization reveals the cellular reprogramming of the gastric tumor microenvironment. Clin Cancer Res 26:2640–2653

Schiller HB, Fernandez IE, Burgstaller G, Schaab C, Scheltema RA, Schwarzmayr T, Strom TM, Eickelberg O, Mann M (2015) Time- and compartment-resolved proteome profiling of the extracellular niche in lung injury and repair. Mol Syst Biol 11:819

Shao X, Taha IN, Clauser KR, Gao Y, Naba A (2020) MatrisomeDB: the ECM-protein knowledge database. Nucleic Acids Res 48:D1136–D1144

Sharma A, Jiang C, De S (2018) Dissecting the sources of gene expression variation in a pan-cancer analysis identifies novel regulatory mutations. Nucleic Acids Res 46:4370–4381

Siegfried Z, Simon I (2010) DNA methylation and gene expression. Wiley Interdiscip Rev Syst Biol Med 2:362–371

Siljamäki E, Rappu P, Riihilä P, Nissinen L, Kähäri V, Heino J (2020) H-Ras activation and fibroblast-induced TGF-β signaling promote laminin-332 accumulation and invasion in cutaneous squamous cell carcinoma. Matrix Biol 87:26–47

Simão D, Silva MM, Terrasso AP, Arez F, Sousa MFQ, Mehrjardi NZ, Šarić T, Gomes-Alves P, Raimundo N, Alves PM, Brito C (2018) Recapitulation of human neural microenvironment signatures in iPSC-derived NPC 3D differentiation. Stem Cell Rep 11:552–564

Sipilä KH, Ranga V, Rappu P, Mali M, Pirilä L, Heino I, Jokinen J, Käpylä J, Johnson MS, Heino J (2017) Joint inflammation related citrullination of functional arginines in extracellular proteins. Sci Rep 7:1–12

Socovich AM, Naba A (2019) The cancer matrisome: from comprehensive characterization to biomarker discovery. Semin Cell Dev Biol 89:157–166

Sood D, Tang-Schomer M, Pouli D, Mizzoni C, Raia N, Tai A, Arkun K, Wu J, Black LD, Scheffler B et al (2019) 3D extracellular matrix microenvironment in bioengineered tissue models of primary pediatric and adult brain tumors. Nat Commun 10:1–14

Taha IN, Naba A (2019) Exploring the extracellular matrix in health and disease using proteomics. Essays Biochem 63:417–432

Tak I, Ali F, Dar JS, Magray AR, Ganai BA, Chishti MZ (2019) Chapter 1 – Posttranslational modifications of proteins and their role in biological processes and associated diseases. In: Dar TA, Singh LR (eds) Protein modificomics. Academic Press, Cambridge, pp 1–35

Thankam FG, Boosani CS, Dilisio MF, Dietz NE, Agrawal DK (2016) MicroRNAs associated with shoulder tendon Matrisome disorganization in Glenohumeral arthritis. PLoS One 11:e0168077

Thompson A, Schäfer J, Kuhn K, Kienle S, Schwarz J, Schmidt G, Neumann T, Hamon C (2003) Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. Anal Chem 75:1895–1904

Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang T, Porta-Pardo E, Gao GF, Plaisier CL, Eddy JA et al (2018) The immune landscape of cancer. Immunity 48:812–830.e14

Tian C, Clauser KR, Öhlund D, Rickelt S, Huang Y, Gupta M, Mani DR, Carr SA, Tuveson DA, Hynes RO (2019) Proteomic analyses of ECM during pancreatic ductal adenocarcinoma progression reveal different contributions by tumor and stromal cells. PNAS 116:19609–19618

Tian C, Öhlund D, Rickelt S, Lidström T, Huang Y, Hao L, Zhao RT, Franklin O, Bhatia SN, Tuveson DA, Hynes RO (2020) Cancer cell-derived Matrisome proteins promote metastasis in pancreatic ductal adenocarcinoma. Cancer Res 80:1461–1474

Ting DT, Wittner BS, Ligorio M, Vincent Jordan N, Shah AM, Miyamoto DT, Aceto N, Bersani F, Brannigan BW, Xega K et al (2014) Single-cell RNA sequencing identifies extracellular matrix gene expression by pancreatic circulating tumor cells. Cell Rep 8:1905–1918

Tomaru Y, Hayashizaki Y (2006) Cancer research with non-coding RNA. Cancer Sci 97:1285–1290

Tomko LA, Hill RC, Barrett A, Szulczewski JM, Conklin MW, Eliceiri KW, Keely PJ, Hansen KC, Ponik SM (2018) Targeted matrisome analysis identifies thrombospondin-2 and tenascin-C in aligned collagen stroma from invasive breast carcinoma. Sci Rep 8:1–11

Triulzi T, Casalini P, Sandri M, Ratti M, Carcangiu ML, Colombo MP, Balsari A, Ménard S, Orlandi R, Tagliabue E (2013) Neoplastic and stromal cells contribute to an extracellular matrix gene expression profile defining a breast Cancer subtype likely to Progress. PLoS One 8:e56761

Tyanova S, Temu T, Cox J (2016) The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. Nat Protoc 11:2301–2319

Walker C, Mojares E, del Río Hernández A (2018) Role of extracellular matrix in development and Cancer progression. Int J Mol Sci 19:3028

Wang M, You J, Bemis KG, Tegeler TJ, Brown DPG (2008) Label-free mass spectrometry-based protein quantification technologies in proteomic analysis. Brief Funct Genomics 7:329–339

Wang M, Zhao J, Zhang L, Wei F, Lian Y, Wu Y, Gong Z, Zhang S, Zhou J, Cao K et al (2017) Role of tumor microenvironment in tumorigenesis. J Cancer 8:761–773

Yalak G, Vogel V (2012) Extracellular phosphorylation and phosphorylated proteins: not just curiosities but physiologically important. Sci Signal 5:re7

Yang Y, Teng Z, Meng S, Yu W (2016) Identification of potential key lncRNAs and genes associated with aging based on microarray data of adipocytes from mice. Res Int 2016:9181702

Yuzhalin AE, Gordon-Weeks AN, Tognoli ML, Jones K, Markelc B, Konietzny R, Fischer R, Muth A, O'Neill E, Thompson PR et al (2018a) Colorectal cancer liver metastatic growth depends on PAD4-driven citrullination of the extracellular matrix. Nat Commun 9:4783

Yuzhalin AE, Urbonas T, Silva MA, Muschel RJ, Gordon-Weeks AN (2018b) A core matrisome gene signature predicts cancer outcome. Br J Cancer 118:435–440

Zhang W, Ota T, Shridhar V, Chien J, Wu B, Kuang R (2013) Network-based survival analysis reveals subnetwork signatures for predicting outcomes of ovarian cancer treatment. PLoS Comput Biol 9:e1002975

Zhang M, Fujiwara K, Che X, Zheng S, Zheng L (2017) DNA methylation in the tumor microenvironment. J Zhejiang Univ Sci B 18:365–372

Zhang HE, Hamson EJ, Koczorowska MM, Tholen S, Chowdhury S, Bailey CG, Lay AJ, Twigg SM, Lee Q, Roediger B et al (2019) Identification of novel natural substrates of fibroblast activation protein-alpha by differential Degradomics and proteomics. Mol Cell Proteomics 18:65–85

Zhu S, Zhang X, Weichert-Leahey N, Dong Z, Zhang C, Lopez G, Tao T, He S, Wood AC, Oldridge D et al (2017) LMO1 synergizes with MYCN to promote neuroblastoma initiation and metastasis. Cancer Cell 32:310–323.e5

Chapter 8
Proteomic and Degradomic Analysis of Body Fluids: Applications, Challenges and Considerations

Konstantinos Kalogeropoulos, Louise Bundgaard, and Ulrich auf dem Keller

Abstract Body fluids are rich sources for proteins originating from organs, tissues, and circulation. They are considered to reflect the state of those tissues and organs and to contain a vast pool of candidate biomarkers for a wide range of conditions. Thus, body fluids are regarded as highly attractive specimens for detecting disease disposition, to monitor markers of progression or to assess the efficacy of treatments. They provide several advantages for screening and clinical testing, such as low cost, minimum to low invasiveness and simple sample collection. However, body fluids are immensely complex mixtures, with most of them containing thousands of proteins and peptides. Therefore, sensitive and accurate high-throughput methods are required for deconvolution of body fluid proteome composition and differential protein abundances. Due to rapid developments in the field of mass spectrometry (MS), MS-based proteomics is deemed to be the most suitable methodology for systematic analysis of body fluid proteomes. In this chapter, we discuss different body fluids and their characteristics, advancements in MS and workflows suitable for body fluid analysis as well as sample preparation and applications in the field of body fluid proteomics and degradomics.

8.1 Body Fluids

Body fluids can be classified as systemic or proximal. Systemic fluids are fluids that represent the overall physiological state of an organism, since they are circulating the whole or most parts of the body. Proximal fluids provide a more specific picture of a tissue, because they are the products of a healthy or diseased organ (Paulo et al. 2011). Body fluids like urine encompass both categories. The direct advantage of analysis of a proximal fluid is the increased chance of detecting specific biomarkers

K. Kalogeropoulos · L. Bundgaard · U. auf dem Keller (✉)
Department of Biotechnology and Biomedicine, Technical University of Denmark, Kongens Lyngby, Denmark
e-mail: uadk@dtu.dk

S. Ricard-Blum (ed.), *Extracellular Matrix Omics*, Biology of Extracellular Matrix 7,
https://doi.org/10.1007/978-3-030-58330-9_8

and downstream products of a disease, while systemic fluids can provide information about the general health state of an organism. Although blood is a systemic fluid with tightly regulated homeostatic mechanisms, it can be a highly valuable sample in cases where proximal fluids cannot be obtained or early diagnosis of a pathological state is the goal (Capelo-Martínez 2019).

8.1.1 Blood

Blood is arguably the most readily accessible biofluid and contains proteins that are shed, leaked or secreted from a multitude of tissues (Ahn and Simpson 2007; Zhao et al. 2018). Plasma is the portion of the blood where the coagulation cascade and clotting mechanisms have been inhibited with anticoagulants. Serum is the blood fraction that is collected after clotting, thereby being free of clotting proteins as well as cells and platelets trapped in the fibrin network (Capelo-Martínez 2019). The proteome composition of plasma and serum has been found to be substantially different (Sapan and Lundblad 2006).

Plasma and serum represent a rich protein source, with numerous of their constituents routinely monitored and quantified in clinical practice. Protein concentration ranges between 60 and 80 mg/ml in specimens from healthy individuals, whereby albumins and globulins account for a large portion of the total protein content. Overall, there is a huge difference in the amounts of individual protein species, which are present in concentrations spanning a range of over ten orders of magnitude (Geyer et al. 2017). About 90% of the total protein amount is represented by the 12–14 most abundant proteins, although it is believed that serum contains up to 10,000 proteins (Kessel et al. 2018). This poses an immense challenge with respect to detection and quantification of lowly abundant proteins in serum that often have high potential to serve as the most indicative biomarkers. Most body fluids exhibit the same problem, and enrichment strategies that partially ameliorate this issue are addressed below. As expected, MS-based proteomics has been extensively applied to investigate plasma proteome composition. The Human Plasma Proteome Project (HPPP) illustrates an effort to document the full protein content of human plasma (Schwenk et al. 2017). Serum proteomics have also been used to examine changes in protein levels in a range of diseases, such as nonalcoholic fatty liver (Younossi et al. 2005) and liver fibrosis (Gressner et al. 2009), rheumatoid arthritis (Park et al. 2015) and cancer (Peng et al. 2018). Several studies and reviews on MS-based proteomics biomarker discovery are available in the literature (Geyer et al. 2016, 2017; Huang et al. 2017), as well as protocols for sample preparation and guidelines for specific analysis workflows (Greco et al. 2017; Lan et al. 2018; Moulder et al. 2018). Peptidomics studies have also been performed for biomarker discovery in serum under normal (Arapidi et al. 2018) and disease conditions (Fan et al. 2012; Yang et al. 2012, 2018; Widlak et al. 2016; Shraibman et al. 2019).

8.1.2 Urine

Urine is another major biofluid that has been thoroughly investigated by MS-based proteomics and can be collected in large volumes in a non-invasive manner (Csősz et al. 2017). Other advantages include minimum cost, the possibility of easy continuous sampling in time-resolved studies and a relatively lower proteome complexity than serum. Urine is a result of blood filtration and contains proteins and peptides that have not been reabsorbed by the organism in this filtration and clearance process (Krochmal et al. 2018). Despite showing a lower protein concentration than serum, it is a body fluid of great interest, especially in the field of biomarker discovery, as it collects components from blood, kidneys and bladder and it has the potential to indicate the natural or pathological conditions of multiple organs (Capelo-Martínez 2019). Urine protein concentration is in the range of 20–100 μg/ml, consisting of mostly low molecular weight proteins (Edwards 2008). As in serum/plasma, albumin is the prevalent protein present in urine. Albumin is followed in abundance by microglobulins and uromodulin, and protein concentration again spans several orders of magnitude (Capelo-Martínez 2019). Urine exhibits high thermodynamic stability, presumably due to its incubation in the bladder, lowering the endogenous proteolytic activity and its low molecular weight protein content (Rai et al. 2005; Tammen et al. 2005). As examples, the urinary proteome or peptidome has been investigated for prostate cancer and rheumatoid arthritis biomarkers (Kang et al. 2014; Jedinak et al. 2018), acute coronary syndromes and kidney diseases (Htun et al. 2017; Sirolli et al. 2019).

8.1.3 Cerebrospinal Fluid

Cerebrospinal fluid (CSF) is an additional body fluid that has attracted attention in proteomics studies. However, invasive methods are required to obtain CSF and the medical condition of the patient has to be taken into consideration before sampling, CSF can be used to detect changes in protein expression in neurological and other disorders and diseases. CSF is a biofluid present in the ventricular system of the central nervous system as well as the surroundings of the spinal cord and brain. CSF is excreted by cells of the choroid plexus and ependymal cells, mediating molecule exchange with blood plasma and brain tissue (Maurer 2008). CSF is fully renewed 3–4 times per day and its composition depends on the sampling location. CSF protein concentration is comparatively low at 0.3–0.7 mg/ml, albeit still suffering from the issue of a few major proteins overshadowing the rest, a characteristic of most biofluids (Davidsson et al. 2002; Hu et al. 2006). CSF proteomics has been employed for the study of spinal cord injury (Streijger et al. 2017), depressive disorder (Al Shweiki et al. 2017), meningitis (Gómez-Baena et al. 2017) and multiple sclerosis (Kroksveen et al. 2015), other than the characterization of the physiological protein composition (Zhang et al. 2005; Barkovits et al. 2018).

8.1.4 Synovial Fluid

Another biofluid of interest is synovial fluid (SF) whose analysis in disease conditions can illuminate pathological states of the joint structure (Driban et al. 2010; Sohn et al. 2012; Kiapour et al. 2019). SF is encapsulated by the joint cavity of the synovial joint. The inner layer of the joint capsule which encircles the joint accommodates a membrane of fibroblast-like cells that produce the viscous SF. SF mediates cell interactions, facilitates molecule transport and lubricates the articular cartilage (Mahendran et al. 2017). Since the synovial membrane has semi-permeable properties, passive diffusion of plasma proteins through the membrane is a natural phenomenon. Excreted proteins from cells in the joint apparatus are also present, making SF a dynamic fluid influenced by a plethora of different factors. Normal human SF protein concentration is in the range of 19–28 mg/ml and increases in pathological states. Also in this body fluid, albumins (approximately 12 mg/ml) and globulins (at 1–3 mg/ml) account for a large portion of the protein content (Hui et al. 2012). Protein size plays an important factor in relative abundances in SF, as it is directly associated with the protein's ability to penetrate the permeable membrane. Hence, large plasma proteins are found at low concentrations in normal SF. On the contrary, pathological SF resembles serum with respect to protein abundance. Synovial fluid proteomic studies are mostly associated with pathologies related to arthritis (Bhattacharjee et al. 2016; Mahendran et al. 2017; Vicenti et al. 2018).

8.1.5 Salivary and Tear Fluid

Salivary fluid and its proteome content has also been investigated, since it as well presents the advantages of non-invasiveness and availability for biomarker discovery (Aqrawi et al. 2017). Saliva is a fluid secreted from the salivary glands and the gingival crevice (Humphrey and Williamson 2001). It contains thousands of proteins (Schulz et al. 2013), of which α-amylase, mucin, cystatins, proline-rich proteins and globulins are most prominent (Hu et al. 2005; Guo et al. 2006; Denny et al. 2008). Saliva is mostly made of water (99%), electrolytes, urea and proteins. Saliva protein concentration is highly variable, but it naturally ranges from 0.7 to 2.4 mg/ml (Lin and Chang 1989; Shaila et al. 2013). Proteomics studies have identified more than a thousand proteins in salivary fluids (https://salivaryproteome.nidcr.nih.gov/). Salivary proteomics or peptidomics have been employed to study differential protein abundance and potential biomarkers in jaw osteonecrosis (Thumbigere-Math et al. 2015), oral lesions (de Jong et al. 2010), periodontitis (Trindade et al. 2015), oral cancer (Gallo et al. 2016; Stuani et al. 2017), as well as systemic conditions such as autoimmune diseases (Ohyama et al. 2015) and diabetes mellitus (Caseiro et al. 2013). The protein content of extracellular vesicles in human saliva in relation to lung cancer has also been investigated (Sun et al. 2018).

Tear is another highly examined body fluid, which can be collected non-invasively. It consists of proteins, lipids and other molecules produced in the lacrimal glands, whereby its normal protein concentration is 5–7 mg/ml (Fullard and Snyder 1990). Tear fluid contains more than 1500 proteins, of which the most prominent ones have been associated with pathogen defense (Zhou et al. 2012). Changes in protein expression levels can reflect inflammatory states in eye-associated or systemic diseases (Li et al. 2008; Le Guezennec et al. 2015). Tear proteomics has also been utilized for the assessment of several ocular-related conditions (Tomosugi et al. 2005; Zhou et al. 2009a,b; Aluru et al. 2012; Leonardi et al. 2014).

8.1.6 Seminal Fluid and Sweat

Semen is a biofluid material, which can be exploited not only in basic research, but also in forensic studies (Merkley et al. 2019). The acellular fraction of semen, or the seminal fluid, accounts for 95% of semen volume (Jodar et al. 2017). More than 6000 proteins have been identified with high confidence in semen proteomic studies investigating the spermatozoal proteome (Amaral et al. 2013). In this, proteins partaking in DNA packaging, RNA metabolism and transport as well as other metabolic processes for energy generation were overrepresented (Amaral et al. 2013; Oliva et al. 2015). In seminal fluid, a little more than 2000 non-redundant proteins have been identified (Gilany et al. 2015), with semenogelins accounting for 80% of the total protein content (Drabovich et al. 2014). About 10% of seminal fluid proteins are encapsulated in extracellular vesicles present in seminal fluid, namely the epididymosomes and the prostasomes whose prominent components are enzymes with GTPase activity and proteins related to phospholipid binding (Jodar et al. 2017). A recent study analyzed the exosomal proteome content in seminal fluid (Yang et al. 2017), complementing results that have been summarized in several review articles on semen proteomics (Gilany et al. 2015; Jodar et al. 2017; Druart and de Graaf 2018).

Sweat, as an excreted fluid, is also important in indicating the physiological state of an organism, and it has been studied with proteomic methods. Similar to salivary fluid, sweat is highly diluted, and the proteins in sweat contribute to defensive mechanisms against pathogens and tissue regeneration following injury (Schittek et al. 2001). Identified proteins are in the range of high hundreds, whereby most abundant proteins include dermicin, clusterin and albumin (Yu et al. 2017). Sweat proteomics studies have explored its role in defense and skin immunity (Csősz et al. 2015; Wu and Liu 2018) as well as sweat protein composition in general (Yu et al. 2017; Na et al. 2019).

8.1.7 Wound Exudate

An additional body fluid that has gained considerable momentum for biomarker discovery and proteomic analysis is wound exudate (Mannello et al. 2014). Wound fluid, which can be classified as proximal fluid, enables investigating the state of the local wound tissue and can serve as an indicator of the wound healing trajectory (Kalkhof et al. 2014; Lindley et al. 2016). Proteins present in wound fluid play an important role in modulating responses to injury and regulating the wound microenvironment (Cavassan et al. 2019). The underlying biological mechanisms of chronic inflammation and non-progressive wounds are still poorly understood. Therefore, wound exudates present an excellent biological matrix for biomarker discovery in chronic, non-healing wounds. Protein concentration depends on sampling method and wound type. Abundant proteins in wound fluids largely overlap with highly abundant serum proteins, of which albumin is the main component. As a result of matrix remodeling and tissue repair and regeneration, extracellular matrix proteins are also overrepresented in these specimens. Fluid from wounds contains inflammatory mediators such as chemokines and growth factors, which are required to orchestrate the distinct phases of the wound healing process. Wound exudate proteins have been found to span at least six orders of magnitude in concentration (Sabino et al. 2015).

8.1.8 Other Body Fluids and Exosomes

Several other types of body fluids such as amniotic fluid and gingival crevicular fluid are available and have been investigated in proteomic studies (Cho et al. 2007; Khurshid et al. 2017; Zhao et al. 2018). Special attention has also to be given to extracellular vesicles (EVs) that are present in basically all body fluids. EVs are a largely unexplored pool of protein transport media that could possess significant relevance for biological questions. Most of the discussed body fluids contain EVs secreted from adjacent or distant cells, and their proteomic content and its changes over time or in different conditions may be of use in biomarker discovery (De Toro et al. 2015; Sódar et al. 2017). Membrane proteins in EVs are involved in cell interactions, adhesion, signaling and ion transport as well as in immune responses (Gutiérrez-Vázquez et al. 2013; Mulcahy et al. 2014; Turturici et al. 2014). Exosomes have been studied in several biofluids, including plasma (Caby et al. 2005), urine (Nilsson et al. 2009), saliva (Ogawa et al. 2011; Sun et al. 2018) and semen (Utleg et al. 2003; Thimon et al. 2008). The diversity of the proteomes discovered in EVs is partly associated with differences in isolation methods (Yáñez-Mó et al. 2015). Almost 35,000 proteoforms have been annotated in EVs and numerous potential markers for different conditions have been identified and validated (Csősz et al. 2017).

8.2 Mass Spectrometry-Based Proteomics

Mass spectrometry (MS)-based technologies have rapidly evolved over the past two decades. For instance, the invention and commercialization of the Orbitrap mass analyzer in MS instruments in 2005 has significantly advanced the field (Eliuk and Makarov 2015). Along with the gradual, steady improvement of MS-Time-of-Flight (ToF) instrumentation, the two technologies have become the gold standard in MS-based experimentation. Further developments in software and data analysis tools as wells as enhancements in instrument speed and sensitivity have made MS-based proteomics the preferred method for interrogation of complex biological samples including body fluids.

Numerous MS-based proteomics workflows have been established, but the method of choice mostly depends on the specific biological question or diagnostic need. In most proteomic analyses of body fluids, a bottom-up methodology is applied, consisting of the following steps: sample preparation, digestion with an endoproteinase for generation of peptides, inline-liquid chromatography (LC) separation and tandem MS analysis. Additional steps are frequently added to this generic workflow, such as a second offline-LC step, desalting, ultrafiltration and protein precipitation. Major differences in workflows depend on, whether a discovery approach will be taken, or if specific analytes are to be monitored and quantified (Schubert et al. 2017).

8.2.1 Discovery Proteomics

Discovery experiments aim at fully deconvoluting the proteomic content of a sample by identifying as many proteins as possible in a wide range of concentrations. These approaches do not require any prior knowledge of the sample composition and can be employed for protein identification and analysis of post-translational modifications. Discovery proteomics workflows can be further subdivided into data-dependent acquisition (DDA) and data-independent acquisition (DIA) methods. In DDA mode, peptides are analyzed at the peptide precursor level (MS1) in predefined packets depending on the instrument cycle time. The most abundant precursor peptides of each packet are selected and fragmented into fragment ions. At the second level of tandem MS, the fragment ions are measured and matched to their precursors. The measurements are then compared to theoretical fragmentation patterns and mass to charge ratios from sequence databases and mapped to the proteins of origin with help of software tools (Meyer 2019). This methodology enables accurate and sensitive identification of thousands of proteins from a single sample in a single experiment. DIA methods differ from DDA by fragmentation of all precursors assigned to distinct windows of mass to charge ratios, which results in a much more complicated data scheme and makes the identification of peptides from mass spectra computationally more challenging. However, a clear advantage of DIA

methods is the unbiased, non-stochastic and universal data acquisition, providing a close to complete overview of the protein constituents in any biological sample (Barkovits et al. 2018). The most popular DIA method, sequential windowed acquisition of all theoretical fragment ion mass spectra (SWATH), has been extensively used in the field of body fluid proteomics and biomarker discovery (Vaswani et al. 2015; Krisp and Molloy 2017; Anjo et al. 2017; Lewandowska et al. 2017; Liao et al. 2017; Miyauchi et al. 2018; Ludwig et al. 2018).

Both DDA and DIA can be applied to gain qualitative or quantitative information (Schubert et al. 2017). Qualitative analyses aim at identifying as many proteins as possible and thereby obtaining the most comprehensive overview of the sample proteome. Typically, these discovery proteomic techniques include supplementary enrichment and/or fractionation steps in an attempt to increase the proteome coverage of the identification procedure. Quantitative proteomics adds an additional dimension by attempting to determine abundances of as many proteins as possible with maximum accuracy. Quantification can be relative by comparing different conditions and/or to control samples or absolute in terms of protein concentration in a sample. Multiple workflows for relative quantification of proteins have been integrated into the general bottom-up proteomics approach and are regularly applied in the analysis of body fluids.

Label-free quantification does not require any additional sample modification and can be performed at the data analysis step. The simplest approach is based on the assumption that the abundance of a protein is related to the number of spectra generated from its peptides and thus determines relative protein quantities by spectral counting. However, this method suffers from the stochastic aspect of mass spectrometric detection in DDA experiments, assay variation and difficulty of quantification at the protein level (Arike and Peil 2014). As a consequence, spectral counting has been shown to be accurate in detecting large changes in protein levels but far less precise for detecting subtle and less significant differences (Liu et al. 2004). This can be overcome by comparing LC-MS peak areas of peptide precursors in MS1 ion current-based quantitative proteomics, a powerful approach for analyzing samples from large cohorts in flexible experimental designs (Wang et al. 2019). A widely applied implementation of this approach is the MaxLFQ algorithm (Cox et al. 2014), which has also been extensively used e.g. in quantitative plasma proteome profiling (Geyer et al. 2016). DIA workflows inherently use label-free quantification but mostly at the MS2 rather than the MS1 level (Pappireddi et al. 2019) with advantages in minimizing the number of missing values and a high quantitation accuracy (Muntel et al. 2019). This has also been extensively exploited in assessing relative protein abundances in body fluid proteomes (Liu et al. 2013, 2015).

In contrast to label-free approaches, label-based relative quantitative proteomics workflows require modifications to the proteome by incorporation of stable isotopes. This can be achieved by metabolic incorporation or chemical labeling. Metabolic incorporation has been implemented with stable isotopic labeling with amino acids in cell culture (SILAC), where natural amino acids in the growth medium are replaced by amino acids with stable isotopes. Cells grown in this culture medium

incorporate the labeled amino acids, allowing to distinguish their proteins from non-labeled control cultures in downstream MS analyses (Kani 2017; Hoedt et al. 2019). SILAC is a consistent, cost-effective and undemanding metabolic labeling method that has been used in a multitude of MS studies (Chen et al. 2015; Wang et al. 2018; Deng et al. 2019). Since SILAC would require incorporation of isotopically labeled amino acids at the organism level, its application is limited in body fluid proteomics. However, SILAC-encoded proteomes can be used as internal standards and have been applied for quantitative comparison of proteins in tissues and blood samples (Zhao et al. 2013; Dittmar and Selbach 2015).

As an alternative to amino acid labeling, chemical labeling utilizes isotopically-encoded chemical tags that are mostly attached to the highly reactive primary amines at lysine side chains and peptide N-termini (Chahrour et al. 2015). Prominent examples are the isotope coded affinity tag (ICAT), isobaric tag for relative and absolute quantitation (iTRAQ) and tandem mass tag (TMT) labels, which are frequently employed to quantitatively compare multiple samples in a single run (Liang et al. 2012; Ren et al. 2017; Moulder et al. 2018; Rao et al. 2019). Recent advancements in the TMT technology allow multiplexed analyses of up to 16 samples in the same MS experiment, significantly reducing measurement time and increasing sample throughput (Thompson et al. 2019). As a consequence, TMT-based quantitative proteomics has advantages in the analysis of samples from larger patient cohorts (Zecha et al. 2019).

8.2.2 Targeted Proteomics

With the rapid progress in identifying comprehensive proteomes by discovery proteomics, targeted approaches have received increasing attention to specifically monitor protein targets such as biomarker candidates. Targeted methods are by nature quantitative and are used when a set of proteins of interest has been defined and needs to be accurately and reproducibly quantified with high sensitivity in many samples. A typical scenario is the identification of biomarker proteins by discovery proteomic analysis of body fluids from a defined group of patients, which are then validated by targeted monitoring in a separate and larger cohort (Altelaar et al. 2013). In this type of MS-based proteomic analyses target proteins or peptides are the focus of the experiment (Sethi et al. 2015). Precursor ions are selectively monitored by the instrument, after specific properties of the precursors such as mass to charge ratios and LC retention times have been predetermined.

Multiple targeted methods exist, but they share the principle of monitoring a specific ion and its subsequent fragmentation products over time. Selected reaction monitoring (SRM) needs triple quadrupole mass spectrometers and is a method that monitors the precursor ion of interest in the first stage of tandem MS and a single specific ion product of a precursor fragmentation reaction in the second MS stage termed a transition. The chromatographic peak of the transition can be used for quantification through numerical integration. Relative quantification of peptides is

performed by comparing chromatographic elution areas recorded for the same transition in samples obtained under different conditions. Multiple reaction monitoring (MRM) is a surrogate method, with the difference of monitoring multiple ion transitions, rendering the analysis and quantification more robust and reliable. Lastly, parallel reaction monitoring (PRM) follows all transitions of a precursor without selection and thus can be performed on the same instruments used for discovery proteomics, such as Orbitrap and ToF analyzers. Thereby, the speed, resolution and sensitivity of current mass spectrometers allow monitoring hundreds of transitions with high precision in a single run. This number can be further increased by spiking in isotopically-labeled internal standard peptides (IS-PRM) to significantly reduce the required measurement time per transition (Gallien et al. 2015). Internal standards also aid in relative quantification and if their concentration is exactly known as for absolute quantitation (AQUA) peptides, they can be used for absolute quantification of peptides and proteins of interest. Concomitantly, reference peptides significantly increase specificity and sensitivity of detection and therefore panels have been developed in particular for the analysis of body fluids (Rice et al. 2019). Targeted methods have been improved over the years, evolving to the method of choice for sensitive, high-precision and high-throughput detection and quantification of target peptides, rendering them especially suitable for biomarker studies (Castro-Gamero et al. 2014). As an example, DIA approaches follow the same principle as PRM but without precursor selection and have the potential to soon enable targeted analysis of complete proteomes.

8.3 Sample Preparation

8.3.1 Sample Handling

Many sample treatments have been developed for MS workflows to improve sample purity, number of protein identifications and quantification accuracy. The following sections give an overview of sample preparation protocols and discuss general considerations related to sample handling. Several articles are readily available and should be consulted for a more comprehensive insight into state-of-the-art sample treatments in body fluid proteomics (Paulo et al. 2011; Hulmes et al. 2004; Kuljanin et al. 2017; Geyer et al. 2019).

As a first measure, it is highly advised to inhibit innate enzymatic activity after sample collection and preprocessing (Hulmes et al. 2004; Rai et al. 2005). Particularly in body fluid samples, endogenous enzymes could still be active, and any activity would alter the proteome, losing its ability to reflect the state of a tissue or organism at time of sampling. Frequently used are protease inhibitors to impair non-physiological protease activities after sample extraction and anticoagulants to inhibit the coagulation cascade in plasma samples. Depending on the specific sample and the biological process in focus, this is extended to phosphatase inhibitors and additional substances interfering with protein-modifying enzymes. With respect to

sample handling and preparation for MS analysis, technical procedures are mostly focused on sample purification and clean-up. For accurate and sensitive analysis of proteome content using MS, samples have to be free of contaminants and biological molecules other than proteins and peptides.

8.3.2 Dynamic Range Reduction

A particular challenge in body fluid proteomics is the enormous dynamic range in protein concentration, spanning up to 12 orders of magnitude, i.e. a ratio of about one copy of a low-abundance protein to one trillion copies of a highly abundant protein (Corthals et al. 2000). In serum, the top 22 most abundant proteins represent 99% of the total protein content, while the remaining 1% comprises thousands of different, lowly abundant proteins. As a result, peptides from the highly abundant proteins tend to overshadow and suppress detection of low-abundance peptides, substantially decreasing protein coverage in MS experiments. Even with the rapid advancements in the field, contemporary mass spectrometers do not match the concentration range of body fluids such as plasma or urine. Moreover, the complexity at the sample level is further increased, if lipids, salts and other metabolites are taken into account and results in a loss of information for lowly abundant proteins, which are frequently the subject of investigation. This dynamic range issue poses a major challenge when sensitivity and reproducibility is a requirement and is the major limiting factor in MS-based proteomics of body fluids. Biomarkers for diseases and pathological conditions are frequently in the lower protein concentration range, making accurate and consistent identification and quantification particularly difficult. To address this problem, MS workflows, especially those optimized for body fluid analysis, include various sample fractionation or enrichment techniques to reduce the dynamic range. This can be achieved either by depletion of high-abundance proteins or by dynamic range compression.

8.3.2.1 Depletion of High-Abundance Proteins

Depletion of high-abundance proteins and the resulting relative increase in low-abundance biomarker candidates is a primary focus in MS-based analysis of body fluids and multiple techniques have been developed (Pisanu et al. 2018). One of the most prevalent depletion methods is column-based immunodepletion, where highly abundant proteins such as albumin and immunoglobulins are captured, thereby lowering their concentration in the sample of interest. General depletion of albumin and IgG proteins can be achieved by dye methods (e.g. Cibacron) and protein A/G based columns, respectively. Monoclonal and polyclonal antibodies or peptide ligand-based columns can be used to remove specific proteins. For instance, there are numerous immunoaffinity commercial kits available for depletion of the most prominent proteins in plasma. Examples include ProteoPrep® (Sigma-Aldrich)

for depletion of albumin and IgG proteins, Seppro® IgY14 (Sigma-Aldrich) for depletion of the 12 most abundant plasma proteins, ProteoPrep 20 (Sigma-Aldrich) for the depletion of the 20 most abundant plasma proteins, Aurum Affi-Gel Blue Mini Columns and Kit (Bio-Rad) for albumin depletion, Pierce™ Top (ThermoFisher Scientific) for depletion of the 12 most abundant plasma proteins, and MARC (Agilent), which are available in multiple configurations. The effectiveness of immunodepletion methods is undisputable. However, apparent disadvantages are the potential loss of low molecular weight proteins that are bound to highly abundant carrier proteins like albumin and the unspecific binding of lowly abundant proteins to the column's affinity ligand.

8.3.2.2 Dynamic Range Compression

An alternative approach to reduce the dynamic range that has attracted attention for its sensitivity and effectiveness is dynamic range compression commercialized as ProteoMiner™ (Bio-Rad). ProteoMiner™ can be more accurately defined as an equalization rather than a depletion method. This technique uses a combinatorial hexapeptide-bead library bound to a chromatographic column that compresses the dynamic range of proteins in plasma samples. The library has a limited but equal binding capacity for each protein, which implies that highly abundant proteins will quickly reach saturation levels, while low-abundance proteins will fully bind to the beads. As the unbound fraction of the sample is washed away, this technique enables concurrent enrichment of the underrepresented, low-concentration proteins and depletion of the highly abundant protein components (Fonslow et al. 2011; Li et al. 2017; Moggridge et al. 2019). Since no proteins are completely depleted from the sample, ligands and other proteins bound to highly abundant carriers will still be represented in the eluted sample. Importantly, relative quantitation of low-abundance biomarker candidates is not impaired, because concentrations of these proteins are generally far below saturation level.

8.4 N-Terminal Enrichment and Degradomics

A growing body of research is devoted to proteolytic events, resulting from enzymatic activity of proteases in a biological system. Proteolysis is a very common post-translational modification, which plays a role in a myriad of biological pathways and processes, such as apoptosis and differentiation (Verhamme et al. 2019; Bond 2019). Body fluids are strongly affected by this activity, since around half of all proteases are secreted and exert their activity in the extracellular space. Typical examples of protease activities in body fluids are the coagulation cascade and the complement activation system. Proteases are also involved in disease and altered levels of proteolytic activity can indicate or cause a pathological state. Degradomics is the

field of research investigating cleavage events in complex samples or systems (Savickas and auf dem Keller 2017; Grozdanić et al. 2019).

Proteolytic cleavage can result in complete degradation or limited processing of a protein substrate. Degraded, unstable proteins are removed from the system and changes in their abundances are indicative for a degradative proteolytic process. However, discovery and analysis of limited proteolytic events by MS-based proteomics requires identification of newly generated protein products. Therefore, newly formed protein N-termini and C-termini are ideal indicators of limited proteolysis (Eckhard et al. 2016). Still, despite high proteolytic activities in body fluids, terminal peptides represent a small portion of the overall peptide content. Since bottom-up proteomics workflows include proteome digestion using trypsin or another suitable endoproteinase, internal protein peptides heavily outnumber terminal peptides in MS-based proteomics experiments, making their detection very difficult. Therefore, several enrichment strategies have been developed to overcome this issue. Because primary amines are more reactive than carboxyl groups, N-terminal enrichment strategies are more prevalent in degradomic studies. Positive enrichment methods selectively enrich for protein termini, while negative enrichment methods aim at depleting internal protein peptides.

8.4.1 Positive Enrichment

Positive enrichment strategies selectively label N-terminal α-amines with chemical affinity tags at the protein level. After digestion, tagged N-terminal peptides can be enriched by affinity purification. Timmer et al. implemented a positive enrichment strategy based on selective guanidation of lysine side chain ε-amines, subsequent protein N-terminal labeling with an amine-reactive biotin tag and streptavidin affinity enrichment of N-terminal peptides after tryptic digest (Timmer et al. 2007). Another positive selection method is based on the enzyme subtiligase, an engineered variant of the protease subtilisin. Subtiligase is able to catalyze the ligation reaction between proteins or peptides. In this method, a biotin-conjugated peptide is ligated selectively to free N-terminal protein α-amines. After digestion, the biotinylated N-termini are affinity purified using streptavidin columns. Finally, the terminal peptides are released by cleavage using tobacco etch virus protease with very distinct specificity for a sequence present in the biotin-labeled peptide tag (Yoshihara et al. 2008). Using this strategy, Wildes et al. recorded the first N-terminome of human blood (Wildes and Wells 2010) and Wiita et al. monitored circulating peptides released from tumors (Wiita et al. 2014). Positive selection methods have proven to produce simplified proteomes for degradomic studies. However, the accurate discrimination between ε- and α-amines is quite difficult, and positively enriched samples do not include natural N-terminal post-translational modifications such as acetylation and cyclization.

8.4.2 Negative Enrichment

Several negative enrichment workflows for protein termini from complex biological samples have been developed and extensively described in the recent literature (Luo et al. 2019; Savickas et al. 2020). Here, we will summarize the principles of the two methods that have been most widely applied in the field of degradomics and body fluid analysis.

8.4.2.1 Combined Fractional Diagonal Chromatography

Combined fractional diagonal chromatography (COFRADIC) is a negative enrichment method that employs two chromatographic techniques for purification of terminal peptides (Staes et al. 2011, 2017). At first, the primary amines are acetylated, the sample is digested, and treated for removal of N-terminal pyroglutamates. Secondly, strong cation exchange chromatography is used for N- and C-terminal peptide enrichment. The more positive tryptic peptides bind to the resin at low pH conditions, leaving the terminal peptides available for collection in the flow-through. The purified peptides are treated with 2,4,6-trinitrobenzenosulfonic acid (TNBS), which increases the hydrophobic properties of C-terminal and internal peptides. Finally, N-termini are recovered by a series of reverse phase liquid chromatography steps. Among many other applications, COFRADIC has been used to discover novel plasma biomarkers for heart failure (Mebazaa et al. 2012).

8.4.2.2 Terminal Amine Isotopic Labeling of Substrates

Terminal amine isotopic labeling of substrates (TAILS) is another method for negative selection of N-termini, which can be multiplexed with the use of TMT/iTRAQ labels (Kleifeld and Doucet 2010; Kleifeld et al. 2011). In this technique, all primary amines are blocked by amine-reactive chemical tags. Following digestion, the peptides are incubated with a high-molecular weight (>100 kDa) aldehyde-derivatized polymer (HPG-ALD), which specifically binds the free primary amines of the internal protein peptides. The polymer bound peptides and the N-terminal peptides are separated by ultrafiltration, keeping the polymer in the retentate and leaving the filtrate enriched with the protein N-termini. Labeling of N-termini with TMT/iTRAQ enables highly multiplexed relative quantification of N-termini and identification of protease cleavage events, which makes TAILS a powerful technique for studying proteolytic landscapes in complex biological matrices. In body fluid degradomics, TAILS has been for example used to characterize proteolysis in human platelets (Prudova et al. 2014) and wound exudates from pigs and patients (Sabino et al. 2015, 2018; Sabino and Hermes 2017).

8.5 Conclusions

Body fluids are important specimens for biomarker discovery, and we have just started to exploit their potential in diagnostics and treatment of disease. While being readily accessible for sampling, their proteomes are highly complex, posing many challenges to their comprehensive analysis. Rapid advancements in MS-based proteomics have helped to overcome many of these issues by development of powerful workflows to reliably assess even low-abundance biomarker candidates with high quantitative accuracy (Table 8.1). This is enabled by combining state-of-the-art instrumentation with advanced sample preparation and approaches to specifically enrich for peptides resulting from post-translational modifications. In particular, proteolysis generates terminal peptides in body fluids that by applying customized degradomics workflows open up an even richer sample space to be explored for devising novel strategies for diagnostics and therapeutic intervention in personalized medicine.

Table 8.1 Overview of original research studies using MS-based proteomics and advanced sample preparation for proteome analysis of body fluids

Body fluid	Methods	Sample preparation	Reference
Plasma/ serum	Discovery, DDA, iTRAQ	Depletion, fractionation	Tremlett et al. (2015)
	Discovery, DDA, label-free	Depletion, fractionation	Zheng et al. (2009)
	Discovery, DDA, TMT	Depletion, fractionation	Zhou et al. (2020)
	Discovery, DDA, TMT	Depletion, fractionation	Mavreli et al. (2020)
	Discovery, DDA, DIA	Precipitation	Lin et al. (2020)
	Discovery, DDA, label-free	Depletion, 2D PAGE separation	Snipsøyr et al. (2020)
	Targeted, MRM	Fractionation	Kim et al. (2014)
	Targeted, MRM	Depletion	Zhao et al. (2014)
	Targeted, MRM	Depletion	Pan et al. (2012)
	Targeted, PRM	Depletion	Kim et al. (2015)
	Discovery, DDA	Fractionation, COFRADIC	Mebazaa et al. (2012)
Urine	Discovery, DDA, label-free	Depletion, SDS-PAGE separation	Kang et al. (2014)
	Discovery, DDA, label-free	Ultracentrifugation	Htun et al. (2017)
	Discovery, DDA, label-free, Targeted (PRM)	Ultracentrifugation	Di Meo et al. (2019)
	Discovery, DIA, DDA, label-free	Ultracentrifugation, fractionation	Jung et al. (2020)

(continued)

Table 8.1 (continued)

Body fluid	Methods	Sample preparation	Reference
CSF	Discovery, DDA, label-free	Fractionation, 2DE separation	Davidsson et al. (2002)
	Discovery, DDA, label-free		Gómez-Baena et al. (2017)
	Discovery, DIA and DDA, label-free	GE separation	Barkovits et al. (2018)
	Discovery, DDA, ICAT	Precipitation	Zhang et al. (2005)
	Targeted, PRM		Streijger et al. (2017)
SF	Discovery, DDA, label-free, targeted (MRM)	Depletion, fractionation	Liao et al. (2004)
	Discovery, DDA, label free	Depletion, SDS-PAGE separation, fractionation	Bhattacharjee et al. (2016)
Saliva	Discovery, DDA, label free	Size exclusion, precipitation	Aqrawi et al. (2017)
	Discovery, DDA, label free	Capillary isoelectric focusing	Guo et al. (2006)
	Discovery, DDA, iTRAQ	Fractionation	de Jong et al. (2010)
	Discovery, DDA, iTRAQ	Fractionation	Caseiro et al. (2013)
	Discovery, DDA, label-free	Depletion, ultracentrifugation	Sun et al. (2018)
	Targeted (MRM)	Depletion	Chi et al. (2020)
Tear	Discovery, DDA, label-free	Fractionation	Zhou et al. (2012)
	Discovery, DDA, iTRAQ	Fractionation	Zhou et al. (2009a)
Semen	Discovery, DDA, label-free	Depletion, ultracentrifugation	Yang et al. (2017)
	Discovery, DDA, label-free		Milardi et al. (2012)
	Discovery, DDA, label-free	Fractionation	Kagedan et al. (2012)
Sweat	Discovery, DDA, label-free	Ultracentrifugation, fractionation	Yu et al. (2017)
	Discovery, DDA, label free, Targeted (MRM)	Precipitation	Csősz et al. (2015)
Wound exudate	Discovery, DDA, label free	SDS-PAGE separation	Kalkhof et al. (2014)
	Discovery, DDA, iTRAQ, Targeted (PRM)	Compression, fractionation, TAILS	Sabino et al. (2015, 2018)

Acknowledgments U.a.d.K. acknowledges supported by a Novo Nordisk Foundation Young Investigator Award (NNF16OC0020670) and by an unrestricted research grant from PAUL HARTMANN AG.

References

Ahn S-M, Simpson RJ (2007) Body fluid proteomics: prospects for biomarker discovery. Proteomics Clin Appl 1:1004–1015. https://doi.org/10.1002/prca.200700217

Al Shweiki MR, Oeckl P, Steinacker P et al (2017) Major depressive disorder: insight into candidate cerebrospinal fluid protein biomarkers from proteomics studies. Expert Rev Proteomics 14:499–514. https://doi.org/10.1080/14789450.2017.1336435

Altelaar AFM, Munoz J, Heck AJR (2013) Next-generation proteomics: towards an integrative view of proteome dynamics. Nat Rev Genet 14:35–48. https://doi.org/10.1038/nrg3356

Aluru SV, Agarwal S, Srinivasan B et al (2012) Lacrimal Proline Rich 4 (LPRR4) protein in the tear fluid is a potential biomarker of dry eye syndrome. PLoS ONE 7:e51979. https://doi.org/10.1371/journal.pone.0051979

Amaral A, Castillo J, Estanyol JM et al (2013) Human sperm tail proteome suggests new endogenous metabolic pathways. Mol Cell Proteomics 12:330–342. https://doi.org/10.1074/mcp.M112.020552

Anjo SI, Santa C, Manadas B (2017) SWATH-MS as a tool for biomarker discovery: from basic research to clinical applications. Proteomics 17:1600278. https://doi.org/10.1002/pmic.201600278

Aqrawi LA, Galtung HK, Vestad B et al (2017) Identification of potential saliva and tear biomarkers in primary Sjögren's syndrome, utilising the extraction of extracellular vesicles and proteomics analysis. Arthritis Res Ther 19:14. https://doi.org/10.1186/s13075-017-1228-x

Arapidi G, Osetrova M, Ivanova O et al (2018) Peptidomics dataset: blood plasma and serum samples of healthy donors fractionated on a set of chromatography sorbents. Data Brief 18:1204–1211. https://doi.org/10.1016/j.dib.2018.04.018

Arike L, Peil L (2014) Spectral counting label-free proteomics. In: Martins-de-Souza D (ed) Shotgun proteomics: methods and protocols. Springer New York, New York, NY, pp 213–222

Barkovits K, Linden A, Galozzi S et al (2018) Characterization of cerebrospinal fluid via data-independent acquisition mass spectrometry. J Proteome Res 17:3418–3430. https://doi.org/10.1021/acs.jproteome.8b00308

Bhattacharjee M, Balakrishnan L, Renuse S et al (2016) Synovial fluid proteome in rheumatoid arthritis. Clin Proteomics 13:12. https://doi.org/10.1186/s12014-016-9113-1

Bond JS (2019) Proteases: history, discovery, and roles in health and disease. J Biol Chem 294:1643–1651. https://doi.org/10.1074/jbc.TM118.004156

Caby M-P, Lankar D, Vincendeau-Scherrer C et al (2005) Exosomal-like vesicles are present in human blood plasma. Int Immunol 17:879–887. https://doi.org/10.1093/intimm/dxh267

Capelo-Martínez J-L (2019) Emerging sample treatments in proteomics

Caseiro A, Ferreira R, Padrão A et al (2013) salivary proteome and peptidome profiling in type 1 diabetes mellitus using a quantitative approach. J Proteome Res 12:1700–1709. https://doi.org/10.1021/pr3010343

Castro-Gamero AM, Izumi C, Rosa JC (2014) Biomarker verification using selected reaction monitoring and shotgun proteomics. In: Martins-de-Souza D (ed) Shotgun proteomics: methods and protocols. Springer New York, New York, NY, pp 295–306

Cavassan NRV, Camargo CC, de Pontes LG et al (2019) Correlation between chronic venous ulcer exudate proteins and clinical profile: a cross-sectional study. J Proteomics 192:280–290. https://doi.org/10.1016/j.jprot.2018.09.009

Chahrour O, Cobice D, Malone J (2015) Stable isotope labelling methods in mass spectrometry-based quantitative proteomics. J Pharm Biomed Anal 113:2–20. https://doi.org/10.1016/j.jpba.2015.04.013

Chen X, Wei S, Ji Y et al (2015) Quantitative proteomics using SILAC: principles, applications, and developments. Proteomics 15:3175–3192. https://doi.org/10.1002/pmic.201500108

Chi L-M, Hsiao Y-C, Chien K-Y et al (2020) Assessment of candidate biomarkers in paired saliva and plasma samples from oral cancer patients by targeted mass spectrometry. J Proteomics 211:103571. https://doi.org/10.1016/j.jprot.2019.103571

Cho C-KJ, Shan SJ, Winsor EJ, Diamandis EP (2007) Proteomics analysis of human amniotic fluid. Mol Cell Proteomics 6(8):1406–1415

Corthals GL, Wasinger VC, Hochstrasser DF, Sanchez JC (2000) The dynamic range of protein expression: a challenge for proteomic research. Electrophoresis 21:1104–1115. https://doi.org/10.1002/(SICI)1522-2683(20000401)21:6<1104::AID-ELPS1104>3.0.CO;2-C

Cox J, Hein MY, Luber CA et al (2014) Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. Mol Cell Proteomics MCP 13:2513–2526. https://doi.org/10.1074/mcp.M113.031591

Csősz É, Emri G, Kalló G et al (2015) Highly abundant defense proteins in human sweat as revealed by targeted proteomics and label-free quantification mass spectrometry. J Eur Acad Dermatol Venereol 29:2024–2031. https://doi.org/10.1111/jdv.13221

Csősz É, Kalló G, Márkus B et al (2017) Quantitative body fluid proteomics in medicine — A focus on minimal invasiveness. J Proteomics 153:30–43. https://doi.org/10.1016/j.jprot.2016.08.009

Davidsson P, Folkesson S, Christiansson M et al (2002) Identification of proteins in human cerebrospinal fluid using liquid-phase isoelectric focusing as a prefractionation step followed by two-dimensional gel electrophoresis and matrix-assisted laser desorption/ionisation mass spectrometry. Rapid Commun Mass Spectrom RCM 16:2083–2088. https://doi.org/10.1002/rcm.834

de Jong EP, Xie H, Onsongo G et al (2010) Quantitative proteomics reveals myosin and actin as promising saliva biomarkers for distinguishing pre-malignant and malignant oral lesions. PLoS ONE 5:e11148. https://doi.org/10.1371/journal.pone.0011148

De Toro J, Herschlik L, Waldner C, Mongini C (2015) Emerging roles of exosomes in normal and pathological conditions: new insights for diagnosis and therapeutic applications. Front Immunol 6. https://doi.org/10.3389/fimmu.2015.00203

Deng J, Erdjument-Bromage H, Neubert TA (2019) Quantitative comparison of proteomes using SILAC. Curr Protoc Protein Sci 95:e74. https://doi.org/10.1002/cpps.74

Denny P, Hagen FK, Hardt M et al (2008) The proteomes of human parotid and submandibular/sublingual gland salivas collected as the ductal secretions. J Proteome Res 7:1994–2006. https://doi.org/10.1021/pr700764j

Di Meo A, Batruch I, Brown MD et al (2019) Identification of prognostic biomarkers in the urinary peptidome of the small renal mass. Am J Pathol 189:2366–2376. https://doi.org/10.1016/j.ajpath.2019.08.015

Dittmar G, Selbach M (2015) SILAC for biomarker discovery. Proteomics Clin Appl 9:301–306. https://doi.org/10.1002/prca.201400112

Drabovich AP, Saraon P, Jarvi K, Diamandis EP (2014) Seminal plasma as a diagnostic fluid for male reproductive system disorders. Nat Rev Urol 11:278–288. https://doi.org/10.1038/nrurol.2014.74

Driban JB, Balasubramanian E, Amin M et al (2010) The potential of multiple synovial-fluid protein-concentration analyses in the assessment of knee osteoarthritis. J Sport Rehabil 19:411–421. https://doi.org/10.1123/jsr.19.4.411

Druart X, de Graaf S (2018) Seminal plasma proteomes and sperm fertility. Anim Reprod Sci 194:33–40. https://doi.org/10.1016/j.anireprosci.2018.04.061

Eckhard U, Marino G, Butler GS, Overall CM (2016) Positional proteomics in the era of the human proteome project on the doorstep of precision medicine. Biochimie 122:110–118. https://doi.org/10.1016/j.biochi.2015.10.018

Edwards DR (ed) (2008) The cancer degradome: proteases and cancer biology. Springer, New York, NY

Eliuk S, Makarov A (2015) Evolution of orbitrap mass spectrometry instrumentation. Annu Rev Anal Chem 8:61–80. https://doi.org/10.1146/annurev-anchem-071114-040325

Fan N-J, Gao C-F, Zhao G et al (2012) Serum peptidome patterns of breast cancer based on magnetic bead separation and mass spectrometry analysis. Diagn Pathol 7:45. https://doi.org/10.1186/1746-1596-7-45

Fonslow BR, Carvalho PC, Academia K et al (2011) Improvements in proteomic metrics of low abundance proteins through proteome equalization using ProteoMiner prior to MudPIT. J Proteome Res 10:3690–3700. https://doi.org/10.1021/pr200304u

Fullard RJ, Snyder C (1990) Protein levels in nonstimulated and stimulated tears of normal human subjects. Invest Ophthalmol Vis Sci 31:1119–1126

Gallien S, Kim SY, Domon B (2015) Large-scale targeted proteomics using internal standard triggered-parallel reaction monitoring (IS-PRM). Mol Cell Proteomics MCP 14:1630–1644. https://doi.org/10.1074/mcp.O114.043968

Gallo C, Ciavarella D, Santarelli A et al (2016) Potential salivary proteomic markers of oral squamous cell carcinoma. Cancer Genomics Proteomics 13:55–62

Geyer PE, Holdt LM, Teupser D, Mann M (2017) Revisiting biomarker discovery by plasma proteomics. Mol Syst Biol 13:942. https://doi.org/10.15252/msb.20156297

Geyer PE, Kulak NA, Pichler G et al (2016) Plasma proteome profiling to assess human health and disease. Cell Syst 2:185–195. https://doi.org/10.1016/j.cels.2016.02.015

Geyer PE, Voytik E, Treit PV et al (2019) Plasma proteome profiling to detect and avoid sample-related biases in biomarker studies. EMBO Mol Med 11. https://doi.org/10.15252/emmm.201910427

Gilany K, Minai-Tehrani A, Savadi-Shiraz E et al (2015) Exploring the human seminal plasma proteome: an unexplored gold mine of biomarker for male Infertility and male reproduction disorder. J Reprod Infertil 16:61–71

Gómez-Baena G, Bennett RJ, Martínez-Rodríguez C et al (2017) Quantitative proteomics of cerebrospinal fluid in paediatric pneumococcal meningitis. Sci Rep 7:7042. https://doi.org/10.1038/s41598-017-07127-6

Greco V, Piras C, Pieroni L, Urbani A (2017) Direct assessment of plasma/serum sample quality for proteomics biomarker investigation. In: Greening DW, Simpson RJ (eds) Serum/plasma proteomics: methods and protocols. Springer New York, New York, NY, pp 3–21

Gressner AM, Gao C-F, Gressner OA (2009) Non-invasive biomarkers for monitoring the fibrogenic process in liver: a short survey. World J Gastroenterol 15:2433. https://doi.org/10.3748/wjg.15.2433

Grozdanić M, Vidmar R, Vizovišek M, Fonović M (2019) Degradomics in biomarker discovery. Proteomics Clin Appl 13:1800138. https://doi.org/10.1002/prca.201800138

Guo T, Rudnick PA, Wang W et al (2006) Characterization of the human salivary proteome by capillary isoelectric focusing/nanoreversed-phase liquid chromatography coupled with ESI-tandem MS. J Proteome Res 5:1469–1478. https://doi.org/10.1021/pr060065m

Gutiérrez-Vázquez C, Villarroya-Beltri C, Mittelbrunn M, Sánchez-Madrid F (2013) Transfer of extracellular vesicles during immune cell-cell interactions. Immunol Rev 251:125–142. https://doi.org/10.1111/imr.12013

Hoedt E, Zhang G, Neubert TA (2019) Stable isotope labeling by amino acids in cell culture (SILAC) for quantitative proteomics. In: Woods AG, Darie CC (eds) Advancements of mass spectrometry in biomedical research. Springer International Publishing, Cham, pp 531–539

Htun NM, Magliano DJ, Zhang Z-Y et al (2017) Prediction of acute coronary syndromes by urinary proteome analysis. PLoS ONE 12:e0172036. https://doi.org/10.1371/journal.pone.0172036

Hu S, Loo JA, Wong DT (2006) Human body fluid proteome analysis. Proteomics 6:6326–6353. https://doi.org/10.1002/pmic.200600284

Hu S, Xie Y, Ramachandran P et al (2005) Large-scale identification of proteins in human salivary proteome by liquid chromatography/mass spectrometry and two-dimensional gel

electrophoresis-mass spectrometry. Proteomics 5:1714–1728. https://doi.org/10.1002/pmic.200401037
Huang Z, Ma L, Huang C et al (2017) Proteomic profiling of human plasma for cancer biomarker discovery. Proteomics 17:1600240. https://doi.org/10.1002/pmic.201600240
Hui AY, McCarty WJ, Masuda K et al (2012) A systems biology approach to synovial joint lubrication in health, injury, and disease: a systems biology approach to synovial joint lubrication. Wiley Interdiscip Rev Syst Biol Med 4:15–37. https://doi.org/10.1002/wsbm.157
Hulmes JD, Bethea D, Ho K et al (2004) An investigation of plasma collection, stabilization, and storage procedures for proteomic analysis of clinical samples. Clin Proteomics 1:17–31. https://doi.org/10.1385/CP:1:1:017
Humphrey SP, Williamson RT (2001) A review of saliva: normal composition, flow, and function. J Prosthet Dent 85:162–169. https://doi.org/10.1067/mpr.2001.113778
Jedinak A, Loughlin KR, Moses MA (2018) Approaches to the discovery of non-invasive urinary biomarkers of prostate cancer. Oncotarget 9. https://doi.org/10.18632/oncotarget.25946
Jodar M, Soler-Ventura A, Oliva R (2017) Semen proteomics and male infertility. J Proteomics 162:125–134. https://doi.org/10.1016/j.jprot.2016.08.018
Jung YH, Han D, Shin SH et al (2020) Proteomic identification of early urinary-biomarkers of acute kidney injury in preterm infants. Sci Rep 10:4057. https://doi.org/10.1038/s41598-020-60890-x
Kagedan D, Lecker I, Batruch I et al (2012) Characterization of the seminal plasma proteome in men with prostatitis by mass spectrometry. Clin Proteomics 9:2. https://doi.org/10.1186/1559-0275-9-2
Kalkhof S, Förster Y, Schmidt J et al (2014) Proteomics and metabolomics for in situ monitoring of wound healing. BioMed Res Int 2014:1–12. https://doi.org/10.1155/2014/934848
Kang MJ, Park Y-J, You S et al (2014) Urinary proteome profile predictive of disease activity in rheumatoid arthritis. J Proteome Res 13:5206–5217. https://doi.org/10.1021/pr500467d
Kani K (2017) Quantitative proteomics using SILAC. In: Comai L, Katz JE, Mallick P (eds) Proteomics: methods and protocols. Springer New York, New York, NY, pp 171–184
Kessel C, McArdle A, Verweyen E et al (2018) Proteomics in chronic arthritis—will we finally have useful biomarkers? Curr Rheumatol Rep 20:53. https://doi.org/10.1007/s11926-018-0762-0
Khurshid Z, Mali M, Naseem M et al (2017) Human gingival crevicular fluids (GCF) proteomics: an overview. Dent J 5:12. https://doi.org/10.3390/dj5010012
Kiapour AM, Sieker JT, Proffen BL et al (2019) Synovial fluid proteome changes in ACL injury-induced posttraumatic osteoarthritis: proteomics analysis of porcine knee synovial fluid. PLoS ONE 14:e0212662. https://doi.org/10.1371/journal.pone.0212662
Kim JS, Ahn H-S, Cho SM et al (2014) Detection and quantification of plasma amyloid-β by selected reaction monitoring mass spectrometry. Anal Chim Acta 840:1–9. https://doi.org/10.1016/j.aca.2014.06.024
Kim YJ, Gallien S, El-Khoury V et al (2015) Quantification of SAA1 and SAA2 in lung cancer plasma using the isotype-specific PRM assays. Proteomics 15:3116–3125. https://doi.org/10.1002/pmic.201400382
Kleifeld O, Doucet A, auf dem Keller U, et al (2010) Isotopic labeling of terminal amines in complex samples identifies protein N-termini and protease cleavage products. Nat Biotechnol 28:281–288. https://doi.org/10.1038/nbt.1611
Kleifeld O, Doucet A, Prudova A et al (2011) Identifying and quantifying proteolytic events and the natural N terminome by terminal amine isotopic labeling of substrates. Nat Protoc 6:1578–1611. https://doi.org/10.1038/nprot.2011.382
Krisp C, Molloy MP (2017) SWATH mass spectrometry for proteomics of non-depleted plasma. In: Greening DW, Simpson RJ (eds) Serum/plasma proteomics: methods and protocols. Springer New York, New York, NY, pp 373–383
Krochmal M, Schanstra JP, Mischak H (2018) Urinary peptidomics in kidney disease and drug research. Expert Opin Drug Discov 13:259–268. https://doi.org/10.1080/17460441.2018.1418320

Kroksveen AC, Opsahl JA, Guldbrandsen A et al (2015) Cerebrospinal fluid proteomics in multiple sclerosis. Biochim Biophys Acta Proteins Proteomics 1854:746–756. https://doi.org/10.1016/j.bbapap.2014.12.013

Kuljanin M, Dieters-Castator DZ, Hess DA et al (2017) Comparison of sample preparation techniques for large-scale proteomics. Proteomics 17:1600337. https://doi.org/10.1002/pmic.201600337

Lan J, Núñez Galindo A, Doecke J et al (2018) Systematic evaluation of the use of human plasma and serum for mass-spectrometry-based shotgun proteomics. J Proteome Res 17:1426–1435. https://doi.org/10.1021/acs.jproteome.7b00788

Le Guezennec X, Quah J, Tong L, Kim N (2015) Human tear analysis with miniaturized multiplex cytokine assay on "wall-less" 96-well plate. Mol Vis 21:1151–1161

Leonardi A, Palmigiano A, Mazzola EA et al (2014) Identification of human tear fluid biomarkers in vernal keratoconjunctivitis using iTRAQ quantitative proteomics. Allergy 69:254–260. https://doi.org/10.1111/all.12331

Lewandowska AE, Macur K, Czaplewska P et al (2017) Qualitative and quantitative analysis of proteome and peptidome of human follicular fluid using multiple samples from single donor with LC–MS and SWATH methodology. J Proteome Res 16:3053–3067. https://doi.org/10.1021/acs.jproteome.7b00366

Li S, He Y, Lin Z et al (2017) Digging more missing proteins using an enrichment approach with ProteoMiner. J Proteome Res 16:4330–4339. https://doi.org/10.1021/acs.jproteome.7b00353

Li S, Sack R, Vijmasi T et al (2008) Antibody protein array analysis of the tear film cytokines. Optom Vis Sci 85

Liang S, Xu Z, Xu X et al (2012) Quantitative proteomics for cancer biomarker discovery. Comb Chem High Throughput Screen 15:221–231. https://doi.org/10.2174/138620712799218635

Liao H, Wu J, Kuhn E et al (2004) Use of mass spectrometry to identify protein biomarkers of disease severity in the synovial fluid and serum of patients with rheumatoid arthritis. Arthritis Rheum 50:3792–3803. https://doi.org/10.1002/art.20720

Liao W, Li Z, Li T et al (2017) Proteomic analysis of synovial fluid in osteoarthritis using SWATH-mass spectrometry. Mol Med Rep. https://doi.org/10.3892/mmr.2017.8250

Lin L, Zheng J, Zheng F et al (2020) Advancing serum peptidomic profiling by data-independent acquisition for clear-cell renal cell carcinoma detection and biomarker discovery. J Proteomics 215:103671. https://doi.org/10.1016/j.jprot.2020.103671

Lin LY, Chang CC (1989) Determination of protein concentration in human saliva. Gaoxiong Yi Xue Ke Xue Za Zhi 5:389–397

Lindley LE, Stojadinovic O, Pastar I, Tomic-Canic M (2016) Biology and biomarkers for wound healing. Plast Reconstr Surg 138:18S–28S. https://doi.org/10.1097/PRS.0000000000002682

Liu H, Sadygov RG, Yates JR (2004) A model for random sampling and estimation of relative protein abundance in shotgun proteomics. Anal Chem 76:4193–4201. https://doi.org/10.1021/ac0498563

Liu Y, Buil A, Collins BC et al (2015) Quantitative variability of 342 plasma proteins in a human twin population. Mol Syst Biol 11:786. https://doi.org/10.15252/msb.20145728

Liu Y, Hüttenhain R, Surinova S et al (2013) Quantitative measurements of N-linked glycoproteins in human plasma by SWATH-MS. Proteomics 13:1247–1256. https://doi.org/10.1002/pmic.201200417

Ludwig C, Gillet L, Rosenberger G, et al (2018) Data-independent acquisition-based SWATH MS for quantitative proteomics: a tutorial. Mol Syst Biol 14. https://doi.org/10.15252/msb.20178126

Luo SY, Araya LE, Julien O (2019) Protease substrate identification using N-terminomics. ACS Chem Biol 14:2361–2371. https://doi.org/10.1021/acschembio.9b00398

Mahendran SM, Oikonomopoulou K, Diamandis EP, Chandran V (2017) Synovial fluid proteomics in the pursuit of arthritis mediators: an evolving field of novel biomarker discovery. Crit Rev Clin Lab Sci 54:495–505. https://doi.org/10.1080/10408363.2017.1408561

Mannello F, Ligi D, Canale M, Raffetto JD (2014) Omics profiles in chronic venous ulcer wound fluid: innovative applications for translational medicine. Expert Rev Mol Diagn 14:737–762. https://doi.org/10.1586/14737159.2014.927312

Maurer MH (2008) Proteomics of brain extracellular fluid (ECF) and cerebrospinal fluid (CSF). Mass Spectrom Rev. https://doi.org/10.1002/mas.20213

Mavreli D, Evangelinakis N, Papantoniou N, Kolialexi A (2020) Quantitative comparative proteomics reveals candidate biomarkers for the early prediction of gestational diabetes mellitus: a preliminary study. In Vivo 34:517–525. https://doi.org/10.21873/invivo.11803

Mebazaa A, Vanpoucke G, Thomas G et al (2012) Unbiased plasma proteomics for novel diagnostic biomarkers in cardiovascular disease: identification of quiescin Q6 as a candidate biomarker of acutely decompensated heart failure. Eur Heart J 33:2317–2324. https://doi.org/10.1093/eurheartj/ehs162

Merkley ED, Wunschel DS, Wahl KL, Jarman KH (2019) Applications and challenges of forensic proteomics. Forensic Sci Int 297:350–363. https://doi.org/10.1016/j.forsciint.2019.01.022

Meyer J (2019) Fast proteome identification and quantification from data-dependent acquisition–tandem mass spectrometry (DDA MS/MS) using free software tools. Methods Protoc 2:8. https://doi.org/10.3390/mps2010008

Milardi D, Grande G, Vincenzoni F et al (2012) Proteomic approach in the identification of fertility pattern in seminal plasma of fertile men. Fertil Steril 97:67–73.e1. https://doi.org/10.1016/j.fertnstert.2011.10.013

Miyauchi E, Furuta T, Ohtsuki S et al (2018) Identification of blood biomarkers in glioblastoma by SWATH mass spectrometry and quantitative targeted absolute proteomics. PLoS ONE 13: e0193799. https://doi.org/10.1371/journal.pone.0193799

Moggridge S, Fulton KM, Twine SM (2019) Enriching for low-abundance serum proteins using ProteoMiner™ and protein-level HPLC. In: Fulton KM, Twine SM (eds) Immunoproteomics: methods and protocols. Springer New York, New York, NY, pp 103–117

Moulder R, Bhosale SD, Goodlett DR, Lahesmaa R (2018) Analysis of the plasma proteome using iTRAQ and TMT-based isobaric labeling. Mass Spectrom Rev 37:583–606. https://doi.org/10.1002/mas.21550

Mulcahy LA, Pink RC, Carter DRF (2014) Routes and mechanisms of extracellular vesicle uptake. J Extracell Vesicles 3:24641. https://doi.org/10.3402/jev.v3.24641

Muntel J, Kirkpatrick J, Bruderer R et al (2019) Comparison of protein quantification in a complex background by DIA and TMT workflows with fixed instrument time. J Proteome Res 18:1340–1351. https://doi.org/10.1021/acs.jproteome.8b00898

Na CH, Sharma N, Madugundu AK et al (2019) Integrated transcriptomic and proteomic analysis of human eccrine sweat glands identifies missing and novel proteins. Mol Cell Proteomics 18:1382–1395. https://doi.org/10.1074/mcp.RA118.001101

Nilsson J, Skog J, Nordstrand A et al (2009) Prostate cancer-derived urine exosomes: a novel approach to biomarkers for prostate cancer. Br J Cancer 100:1603–1607. https://doi.org/10.1038/sj.bjc.6605058

Ogawa Y, Miura Y, Harazono A et al (2011) Proteomic analysis of two types of exosomes in human whole saliva. Biol Pharm Bull 34:13–23. https://doi.org/10.1248/bpb.34.13

Ohyama K, Baba M, Tamai M et al (2015) Proteomic profiling of antigens in circulating immune complexes associated with each of seven autoimmune diseases. Clin Biochem 48:181–185. https://doi.org/10.1016/j.clinbiochem.2014.11.008

Oliva R, Castillo J, Estanyol J, Ballescà J (2015) Human sperm chromatin epigenetic potential: genomics, proteomics, and male infertility. Asian J Androl 17:601. https://doi.org/10.4103/1008-682X.153302

Pan S, Chen R, Brand RE et al (2012) Multiplex targeted proteomic assay for biomarker detection in plasma: a pancreatic cancer biomarker case study. J Proteome Res 11:1937–1948. https://doi.org/10.1021/pr201117w

Pappireddi N, Martin L, Wühr M (2019) A review on quantitative multiplexed proteomics. Chembiochem Eur J Chem Biol 20:1210–1224. https://doi.org/10.1002/cbic.201800650

Park Y-J, Chung MK, Hwang D, Kim W-U (2015) Proteomics in rheumatoid arthritis research. Immune Netw 15:177. https://doi.org/10.4110/in.2015.15.4.177

Paulo JA, Vaezzadeh AR, Conwell DL et al (2011) Sample handling of body fluids for proteomics. In: Ivanov AR, Lazarev AV (eds) Sample preparation in biological mass spectrometry. Springer Netherlands, Dordrecht, pp 327–360

Peng L, Cantor DI, Huang C et al (2018) Tissue and plasma proteomics for early stage cancer detection. Mol Omics 14:405–423. https://doi.org/10.1039/C8MO00126J

Pisanu S, Biosa G, Carcangiu L et al (2018) Comparative evaluation of seven commercial products for human serum enrichment/depletion by shotgun proteomics. Talanta 185:213–220. https://doi.org/10.1016/j.talanta.2018.03.086

Prudova A, Serrano K, Eckhard U et al (2014) TAILS N-terminomics of human platelets reveals pervasive metalloproteinase-dependent proteolytic processing in storage. Blood 124:e49–e60. https://doi.org/10.1182/blood-2014-04-569640

Rai AJ, Gelfand CA, Haywood BC et al (2005) HUPO Plasma Proteome Project specimen collection and handling: towards the standardization of parameters for plasma proteome samples. Proteomics 5:3262–3277. https://doi.org/10.1002/pmic.200401245

Rao AA, Mehta K, Gahoi N, Srivastava S (2019) Application of 2D-DIGE and iTRAQ workflows to analyze CSF in gliomas. In: Santamaría E, Fernández-Irigoyen J (eds) Cerebrospinal fluid (CSF) proteomics: methods and protocols. Springer New York, New York, NY, pp 81–110

Ren J, Zhao G, Sun X et al (2017) Identification of plasma biomarkers for distinguishing bipolar depression from major depressive disorder by iTRAQ-coupled LC–MS/MS and bioinformatics analysis. Psychoneuroendocrinology 86:17–24. https://doi.org/10.1016/j.psyneuen.2017.09.005

Rice SJ, Liu X, Zhang J, Belani CP (2019) Absolute quantification of all identified plasma proteins from SWATH data for biomarker discovery. Proteomics 19:e1800135. https://doi.org/10.1002/pmic.201800135

Sabino F, Egli FE, Savickas S et al (2018) Comparative degradomics of porcine and human wound exudates unravels biomarker candidates for assessment of wound healing progression in trauma patients. J Invest Dermatol 138:413–422. https://doi.org/10.1016/j.jid.2017.08.032

Sabino F, Hermes O, auf dem Keller U (2017) Body fluid degradomics and characterization of basic N-terminome. In: Methods in enzymology. Elsevier, pp 177–199

Sabino F, Hermes O, Egli FE et al (2015) In vivo assessment of protease dynamics in cutaneous wound healing by degradomics analysis of porcine wound exudates. Mol Cell Proteomics 14:354–370. https://doi.org/10.1074/mcp.M114.043414

Sapan CV, Lundblad RL (2006) Considerations regarding the use of blood samples in the proteomic identification of biomarkers for cancer diagnosis. Cancer Genom 4

Savickas S, Auf dem Keller U (2017) Targeted degradomics in protein terminomics and protease substrate discovery. Biol Chem 399:47–54. https://doi.org/10.1515/hsz-2017-0187

Savickas S, Kastl P, Auf dem Keller U (2020) Combinatorial degradomics: precision tools to unveil proteolytic processes in biological systems. Biochim Biophys Acta Proteins Proteomics 1868:140392. https://doi.org/10.1016/j.bbapap.2020.140392

Schittek B, Hipfel R, Sauer B et al (2001) Dermcidin: a novel human antibiotic peptide secreted by sweat glands. Nat Immunol 2:1133–1137. https://doi.org/10.1038/ni732

Schubert OT, Röst HL, Collins BC et al (2017) Quantitative proteomics: challenges and opportunities in basic and applied research. Nat Protoc 12:1289–1294. https://doi.org/10.1038/nprot.2017.040

Schulz BL, Cooper-White J, Punyadeera CK (2013) Saliva proteome research: current status and future outlook. Crit Rev Biotechnol 33:246–259. https://doi.org/10.3109/07388551.2012.687361

Schwenk JM, Omenn GS, Sun Z et al (2017) The Human Plasma Proteome Draft of 2017: building on the human plasma peptide atlas from mass spectrometry and complementary assays. J Proteome Res 16:4299–4310. https://doi.org/10.1021/acs.jproteome.7b00467

Sethi S, Chourasia D, Parhar IS (2015) Approaches for targeted proteomics and its potential applications in neuroscience. J Biosci 40:607–627. https://doi.org/10.1007/s12038-015-9537-1

Shaila M, Pai GP, Shetty P (2013) Salivary protein concentration, flow rate, buffer capacity and pH estimation: a comparative study among young and elderly subjects, both normal and with gingivitis and periodontitis. J Indian Soc Periodontol 17:42–46. https://doi.org/10.4103/0972-124X.107473

Shraibman B, Barnea E, Kadosh DM et al (2019) Identification of tumor antigens among the HLA peptidomes of glioblastoma tumors and plasma. Mol Cell Proteomics 18:1255–1268. https://doi.org/10.1074/mcp.RA119.001524

Sirolli V, Pieroni L, Di Liberato L et al (2019) Urinary peptidomic biomarkers in kidney diseases. Int J Mol Sci 21:96. https://doi.org/10.3390/ijms21010096

Snipsøyr MG, Wiggers H, Ludvigsen M et al (2020) Towards identification of novel putative biomarkers for infective endocarditis by serum proteomic analysis. Int J Infect Dis S1201971220300849. https://doi.org/10.1016/j.ijid.2020.02.026

Sódar BW, Kovács Á, Visnovitz T et al (2017) Best practice of identification and proteomic analysis of extracellular vesicles in human health and disease. Expert Rev Proteomics 14:1073–1090. https://doi.org/10.1080/14789450.2017.1392244

Sohn D, Sokolove J, Sharpe O et al (2012) Plasma proteins present in osteoarthritic synovial fluid can stimulate cytokine production via Toll-like receptor 4. Arthritis Res Ther 14:R7. https://doi.org/10.1186/ar3555

Staes A, Impens F, Van Damme P et al (2011) Selecting protein N-terminal peptides by combined fractional diagonal chromatography. Nat Protoc 6:1130–1141. https://doi.org/10.1038/nprot.2011.355

Staes A, Van Damme P, Timmerman E et al (2017) Protease substrate profiling by N-terminal COFRADIC. In: Schilling O (ed) Protein terminal profiling: methods and protocols. Springer New York, New York, NY, pp 51–76

Streijger F, Skinnider MA, Rogalski JC et al (2017) A targeted proteomics analysis of cerebrospinal fluid after acute human spinal cord injury. J Neurotrauma 34:2054–2068. https://doi.org/10.1089/neu.2016.4879

Stuani VT, Rubira CMF, Sant'Ana ACP, Santos PSS (2017) Salivary biomarkers as tools for oral squamous cell carcinoma diagnosis: a systematic review: salivary biomarkers for oral SCC. Head Neck 39:797–811. https://doi.org/10.1002/hed.24650

Sun Y, Huo C, Qiao Z et al (2018) Comparative proteomic analysis of exosomes and microvesicles in human saliva for lung cancer. J Proteome Res 17:1101–1107. https://doi.org/10.1021/acs.jproteome.7b00770

Tammen H, Schulte I, Hess R et al (2005) Peptidomic analysis of human blood specimens: comparison between plasma specimens and serum by differential peptide display. Proteomics 5:3414–3422. https://doi.org/10.1002/pmic.200401219

Thimon V, Frenette G, Saez F et al (2008) Protein composition of human epididymosomes collected during surgical vasectomy reversal: a proteomic and genomic approach. Hum Reprod 23:1698–1707. https://doi.org/10.1093/humrep/den181

Thompson A, Wölmer N, Koncarevic S et al (2019) TMTpro: design, synthesis, and initial evaluation of a proline-based isobaric 16-plex tandem mass tag reagent set. Anal Chem 91:15941–15950. https://doi.org/10.1021/acs.analchem.9b04474

Thumbigere-Math V, Michalowicz B, de Jong E et al (2015) Salivary proteomics in bisphosphonate-related osteonecrosis of the jaw. Oral Dis 21:46–56. https://doi.org/10.1111/odi.12204

Timmer JC, Enoksson M, Wildfang E et al (2007) Profiling constitutive proteolytic events in vivo. Biochem J 407:41–48. https://doi.org/10.1042/BJ20070775

Tomosugi N, Kitagawa K, Takahashi N et al (2005) Diagnostic potential of tear proteomic patterns in Sjögren's syndrome. J Proteome Res 4:820–825. https://doi.org/10.1021/pr0497576

Tremlett H, Dai DLY, Hollander Z et al (2015) Serum proteomics in multiple sclerosis disease progression. J Proteomics 118:2–11. https://doi.org/10.1016/j.jprot.2015.02.018

Trindade F, Amado F, Oliveira-Silva RP et al (2015) Toward the definition of a peptidome signature and protease profile in chronic periodontitis. Proteomics Clin Appl 9:917–927. https://doi.org/10.1002/prca.201400191

Turturici G, Tinnirello R, Sconzo G, Geraci F (2014) Extracellular membrane vesicles as a mechanism of cell-to-cell communication: advantages and disadvantages. Am J Physiol-Cell Physiol 306:C621–C633. https://doi.org/10.1152/ajpcell.00228.2013

Utleg AG, Yi EC, Xie T et al (2003) Proteomic analysis of human prostasomes. The Prostate 56:150–161. https://doi.org/10.1002/pros.10255

Vaswani K, Ashman K, Reed S et al (2015) Applying SWATH mass spectrometry to investigate human cervicovaginal fluid during the menstrual cycle 1. Biol Reprod 93. https://doi.org/10.1095/biolreprod.115.128231

Verhamme IM, Leonard SE, Perkins RC (2019) Proteases: pivot points in functional proteomics. In: Wang X, Kuruc M (eds) Functional proteomics. Springer New York, New York, NY, pp 313–392

Vicenti G, Bizzoca D, Carrozzo M et al (2018) Multi-omics analysis of synovial fluid: a promising approach in the study of osteoarthritis. J Biol Regul Homeost Agents 32:9–13

Wang X, He Y, Ye Y et al (2018) SILAC-based quantitative MS approach for real-time recording protein-mediated cell-cell interactions. Sci Rep 8:8441. https://doi.org/10.1038/s41598-018-26262-2

Wang X, Shen S, Rasam SS, Qu J (2019) MS1 ion current-based quantitative proteomics: a promising solution for reliable analysis of large biological cohorts. Mass Spectrom Rev 38:461–482. https://doi.org/10.1002/mas.21595

Widlak P, Pietrowska M, Polanska J et al (2016) Serum mass profile signature as a biomarker of early lung cancer. Lung Cancer 99:46–52. https://doi.org/10.1016/j.lungcan.2016.06.011

Wiita AP, Hsu GW, Lu CM et al (2014) Circulating proteolytic signatures of chemotherapy-induced cell death in humans discovered by N-terminal labeling. Proc Natl Acad Sci U S A 111:7594–7599. https://doi.org/10.1073/pnas.1405987111

Wildes D, Wells JA (2010) Sampling the N-terminal proteome of human blood. Proc Natl Acad Sci 107:4561–4566. https://doi.org/10.1073/pnas.0914495107

Wu C-X, Liu Z-F (2018) Proteomic profiling of sweat exosome suggests its involvement in skin immunity. J Invest Dermatol 138:89–97. https://doi.org/10.1016/j.jid.2017.05.040

Yáñez-Mó M, Siljander PR-M, Andreu Z et al (2015) Biological properties of extracellular vesicles and their physiological functions. J Extracell Vesicles 4:27066. https://doi.org/10.3402/jev.v4.27066

Yang C, Guo W-B, Zhang W-S et al (2017) Comprehensive proteomics analysis of exosomes derived from human seminal plasma. Andrology 5:1007–1015. https://doi.org/10.1111/andr.12412

Yang J, Chen Y, Xiong X et al (2018) Peptidome analysis reveals novel serum biomarkers for children with autism spectrum disorder in China. Proteomics Clin Appl 12:1700164. https://doi.org/10.1002/prca.201700164

Yang J, Song Y-C, Dang C-X et al (2012) Serum peptidome profiling in patients with gastric cancer. Clin Exp Med 12:79–87. https://doi.org/10.1007/s10238-011-0149-2

Yoshihara HAI, Mahrus S, Wells JA (2008) Tags for labeling protein N-termini with subtiligase for proteomics. Bioorg Med Chem Lett 18:6000–6003. https://doi.org/10.1016/j.bmcl.2008.08.044

Younossi ZM, Baranova A, Ziegler K et al (2005) A genomic and proteomic study of the spectrum of nonalcoholic fatty liver disease. Hepatol Baltim Md 42:665–674. https://doi.org/10.1002/hep.20838

Yu Y, Prassas I, Muytjens CMJ, Diamandis EP (2017) Proteomic and peptidomic analysis of human sweat with emphasis on proteolysis. J Proteomics 155:40–48. https://doi.org/10.1016/j.jprot.2017.01.005

Zecha J, Satpathy S, Kanashova T et al (2019) TMT labeling for the masses: a robust and cost-efficient, in-solution labeling approach. Mol Cell Proteomics MCP 18:1468–1478. https://doi.org/10.1074/mcp.TIR119.001385

Zhang J, Goodlett DR, Peskind ER et al (2005) Quantitative proteomic analysis of age-related changes in human cerebrospinal fluid. Neurobiol Aging 26:207–227. https://doi.org/10.1016/j.neurobiolaging.2004.03.012

Zhao C, Trudeau B, Xie H et al (2014) Epitope mapping and targeted quantitation of the cardiac biomarker troponin by SID-MRM mass spectrometry. Proteomics 14:1311–1321. https://doi.org/10.1002/pmic.201300150

Zhao M, Yang Y, Guo Z et al (2018) A comparative proteomics analysis of five body fluids: plasma, urine, cerebrospinal fluid, amniotic fluid, and saliva. Proteomics Clin Appl 12:1800008. https://doi.org/10.1002/prca.201800008

Zhao S, Li R, Cai X et al (2013) The application of SILAC mouse in human body fluid proteomics analysis reveals protein patterns associated with IgA nephropathy. Evid-Based Complement Altern Med ECAM 2013:275390. https://doi.org/10.1155/2013/275390

Zheng X, Wu S, Hincapie M, Hancock WS (2009) Study of the human plasma proteome of rheumatoid arthritis. J Chromatogr A 1216:3538–3545. https://doi.org/10.1016/j.chroma.2009.01.063

Zhou B, Zhou Z, Chen Y et al (2020) Plasma proteomics-based identification of novel biomarkers in early gastric cancer. Clin Biochem 76:5–10. https://doi.org/10.1016/j.clinbiochem.2019.11.001

Zhou L, Beuerman RW, Chan CM et al (2009a) Identification of tear fluid biomarkers in dry eye syndrome using iTRAQ quantitative proteomics. J Proteome Res 8:4889–4905. https://doi.org/10.1021/pr900686s

Zhou L, Beuerman RW, Chew AP et al (2009b) Quantitative analysis of N-linked glycoproteins in tear fluid of climatic droplet keratopathy by glycopeptide capture and iTRAQ. J Proteome Res 8:1992–2003. https://doi.org/10.1021/pr800962q

Zhou L, Zhao SZ, Koh SK et al (2012) In-depth analysis of the human tear proteome. J Proteomics 75:3877–3885. https://doi.org/10.1016/j.jprot.2012.04.053

Chapter 9
Regulation of Cell-Matrix Adhesion Networks: Insights from Proteomics

Emma S. Koeleman, Alexander Loftus, Athanasia D. Yiapanas, and Adam Byron

Abstract Cell adhesion to the extracellular matrix regulates many fundamental aspects of cellular behaviour, including cell survival, proliferation, differentiation, and migration. Cell-matrix interactions are mediated by integrin-associated adhesion complexes, multiprotein complexes of intracellular scaffolding and signalling proteins that link the cytoskeleton to the extracellular milieu. These linkages enable cells to sense and respond to the biochemical and biomechanical properties of the microenvironment. Advances in proteomic methodologies have enabled the characterisation of adhesion complex composition, interactions, and dynamics, which has led to new insights into the physical and functional properties of cell-matrix adhesion networks. These approaches have broadened our understanding of the molecules that regulate cell adhesion to the extracellular matrix.

9.1 Introduction

Multicellular existence is dependent on interactions between cells and the extracellular matrix (ECM) in which they reside. Cell adhesion to the ECM regulates many central aspects of cellular behaviour, such as cell survival, proliferation, differentiation, and migration (Doyle and Yamada 2016; Geiger and Yamada 2011; Moreno-Layseca and Streuli 2014). Cells interact with the ECM using transmembrane

Emma S. Koeleman and Alexander Loftus contributed equally.

E. S. Koeleman
Cancer Research UK Edinburgh Centre, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK

Leiden University Medical Center, ZC, Leiden, The Netherlands

A. Loftus · A. D. Yiapanas · A. Byron (✉)
Cancer Research UK Edinburgh Centre, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK
e-mail: adam.byron@igmm.ed.ac.uk

S. Ricard-Blum (ed.), *Extracellular Matrix Omics*, Biology of Extracellular Matrix 7,
https://doi.org/10.1007/978-3-030-58330-9_9

adhesion receptors, primarily integrins (Humphries et al. 2006; Kechagia et al. 2019; Michael and Parsons 2020). Upon binding their extracellular ligand, integrins cluster on the cell surface and orchestrate the recruitment of intracellular scaffolding and signalling molecules to sites of cell-matrix contact (Byron et al. 2010; Green and Brown 2019; Wehrle-Haller 2012). These sites, the best characterised of which are focal adhesions, contain multimeric integrin-associated adhesion protein complexes, which mature in response to mechanical cues and link the cytoskeleton and the ECM (Humphries et al. 2015; Jansen et al. 2017; Sun et al. 2016a). Through these cell-matrix adhesion complexes, cells can sense and respond to biochemical and mechanical signals in the extracellular microenvironment, which are key physiological processes. Indeed, many diseases, including cancer, involve either dysregulation of cellular responses to ECM signals or alterations in the mechanical properties of the ECM itself (Cooper and Giancotti 2019; Hamidi and Ivaska 2018; Winograd-Katz et al. 2014).

The components of adhesion complexes form networks of direct and indirect protein associations, which are fundamental to the control of cellular behaviour in the context of the extracellular microenvironment (Byron and Frame 2016; Devreotes and Horwitz 2015; Horton et al. 2016a). Herein, we discuss how such cell-matrix adhesion networks enable cells to respond to extracellular cues through focal adhesion assembly and downstream signalling. In particular, we focus on how proteomic approaches have enabled the characterisation of adhesion complex composition and spatiotemporal dynamics and how these studies have contributed to our overall understanding of cell-matrix adhesion.

9.2 Regulation of Cell-Matrix Adhesions

The interactions of cells with the ECM is mediated by adhesion protein complexes that associate with integrin receptors at the plasma membrane. The assembly and turnover of cell-matrix adhesion complexes are tightly regulated, governed in part by physical and functional connections between adhesion proteins. The large number of proteins that can be recruited to sites of cell adhesion, collectively termed the adhesome (Zaidel-Bar et al. 2007a), are involved in a broad range of signalling pathways, which emphasises the numerous cellular processes in which cell-matrix adhesions play key roles, including cell polarity and migration, cytoskeletal organisation, and regulation of the cell cycle. In this section, the structure, function, and regulation of integrin-associated adhesion complexes are detailed.

9.2.1 Bidirectional Integrin Signalling

The integrin family of transmembrane adhesion receptors comprises 18 α integrin and 8 β integrin subunits, which can assemble into 24 distinct heterodimeric integrin

receptors (Humphries et al. 2006; Hynes 2002). These receptors can be divided into four classes based on the possible combinations of integrin receptor and extracellular ligand: RGD-binding integrins, which recognise ligands containing an arginine-glycine-aspartate tripeptide motif; LDV-binding integrins, which recognise ligands containing a sequence related to the acidic leucine-aspartate-valine motif; αA domain-containing integrins, which bind collagens and laminins; and non-αA domain-containing laminin-binding integrins (Humphries et al. 2006). The cell- and tissue-specific pattern of integrin expression defines with which ECM molecules a cell can interact, which, in turn, determines the ability of the cell to sense and respond to specific extracellular microenvironments (Fig. 9.1a).

Structurally, integrins are composed of an extracellular "head" domain, at which α and β integrin subunits interface near the ligand-binding domain, long extracellular "thigh" and "calf" domains, a single-pass transmembrane domain, and a short cytoplasmic tail (except for the β4 integrin tail, which is longer) that mediates intracellular protein interactions. The mechanism of integrin activation is mediated by changes in tertiary and quaternary structure, whereby a bent, inactive integrin conformation unfolds stepwise to an extended open conformation, permitting a rapid increase in affinity for adhesive ligands (Askari et al. 2009; Campbell and Humphries 2011; Shattil et al. 2010). However, the structure-function relationships of integrins remain incompletely understood. Indeed, there may be additional pathways for activation of certain integrins, as some integrin heterodimers do not unfold stepwise, instead assuming a constitutively extended conformation (Miyazaki et al. 2018; Wang et al. 2017, 2019).

Receptor activation is tightly regulated by binding to extracellular ligands (outside-in signalling) and intracellular ligands (inside-out signalling), and this enables integrins to signal bidirectionally across the plasma membrane (Ginsberg et al. 2005; Hynes 2002). Outside-in signalling, which is initiated by ligand-bound integrin clustering on the cell surface, triggers integrin heterodimer- and ligand-specific recruitment of adhesion proteins and activation of downstream signalling through, for example, focal adhesion kinase (FAK)–Src, phosphatidylinositol 3-kinase–Akt (also known as protein kinase B), and Ras–mitogen-activated protein kinase pathways. Inside-out signalling regulates binding of the adaptor proteins talin and kindlin to β integrin cytoplasmic tails, thereby inducing conformational changes in integrins that modulate receptor affinity for extracellular ligands (Calderwood et al. 2013; Sun et al. 2019). It is likely that both of these modes of integrin signalling work in dynamic equilibrium to coordinate integrin function, enabling sensing of and response to the microenvironment.

In addition to protein interactions that serve to activate integrins, a number of adhesion proteins have been found to interfere, directly or indirectly, with inside-out integrin activation. The integrin cytoplasmic domain-associated protein 1 (ICAP-1; also known as ITGB1BP1) specifically interacts with the β1 integrin cytoplasmic tail to negatively regulate its function (Bouvard et al. 2003; Degani et al. 2002). SHARPIN negatively regulates integrin activity by binding the α integrin subunit and interfering with the recruitment of talin and kindlin to β integrin cytoplasmic tails (Rantala et al. 2011). The actin-binding protein filamin-A forms a ternary

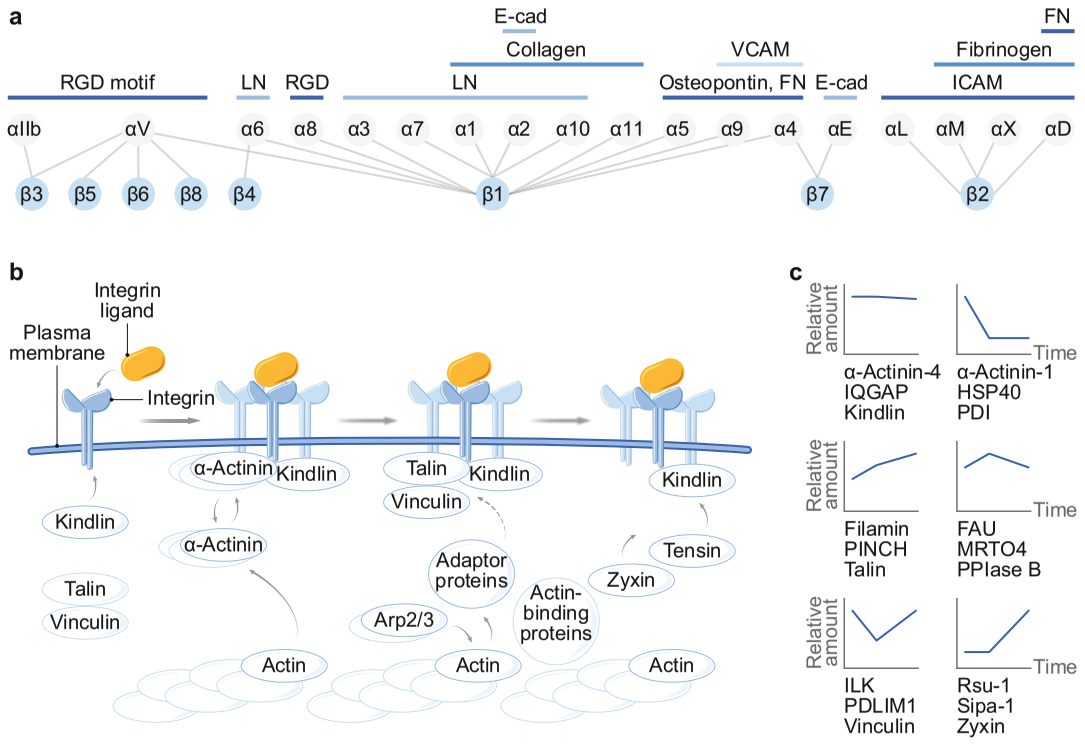

Fig. 9.1 Assembly of integrin-associated adhesion complexes. (**a**) Reported α–β integrin heterodimers and corresponding representative extracellular ligands (Humphries et al. 2006). E-cad, E-cadherin; FN, fibronectin; ICAM, intercellular adhesion molecule; LN, laminin. (**b**) During assembly of cell-matrix adhesions, adhesion proteins are recruited hierarchically to ligand-bound integrin, as determined by fluorescence microscopy studies (Bachir et al. 2014; Hoffmann et al. 2014; Zaidel-Bar et al. 2003). (**c**) Profiling of consensus adhesome protein recruitment during adhesion complex assembly, as determined by proteomic experiments (Horton et al. 2015). Respective examples of consensus adhesome proteins are indicated for each line profile trend. (Parts **b** and **c** adapted by permission from Springer Nature Customer Service Centre GmbH: Springer Nature, Methods in Molecular Biology: Proteomic profiling of integrin adhesion complex assembly by Byron (2018) in Protein Complex Assembly by Marsh (ed) © 2018 Springer Science+Business Media LLC)

complex with αIIbβ3 integrin, which restrains the integrin in a resting state by stabilising a molecular clasp between the α and β integrin cytoplasmic tails (Liu et al. 2015). The actin regulatory SHANK family proteins inhibit integrin activity by sequestering the integrin activator Rap1 to limit G-protein signalling at the plasma membrane (Lilja et al. 2017). The guanosine triphosphatase (GTPase)-activating protein ARAP3, activated by outside-in signalling, functions in a negative feedback loop, inducing negative inside-out signalling to locally inactivate integrins (McCormick et al. 2019). Integrin inactivators are crucial to balance and reset the activation of integrins and therefore strongly influence how cells can interact with the ECM (Bouvard et al. 2013). Inactive and active conformation states of integrins can be monitored experimentally using anti-integrin monoclonal antibodies that detect conformation-dependent epitopes, and many of these antibodies serve as useful tools to regulate integrin conformation and, as a consequence, function (Byron et al. 2009). Both inactivation and constitutive activation of integrins using, for example, monoclonal antibodies can impair cell motility (Huttenlocher et al. 1996), highlighting that maintenance of the dynamic equilibrium between integrin activation states is vital for effective cell movement.

9.2.2 Adhesion Complex Assembly

The dynamic nature of adhesion complex formation, together with the remarkable plasticity displayed by cell-matrix adhesions, reflect the necessity for the cell to respond rapidly to changes in the extracellular microenvironment by regulating cytoskeletal dynamics and cell motility, for example (Vicente-Manzanares and Horwitz 2011; Wolfenson et al. 2013). While the mechanism by which integrin-based adhesions first form is unclear, small, nascent cell-matrix adhesions emerge at the periphery of plasma membrane protrusions following integrin activation and clustering and the recruitment of integrin-associated proteins (Fig. 9.1b). These nascent adhesions either disassemble within approximately 1 minute or mature into larger focal adhesions with longer lifetimes of several minutes. Focal adhesion maturation requires linkage to the actomyosin cytoskeleton (Vicente-Manzanares et al. 2009). Pulling forces generated by active myosin II can change the conformation of several adhesion proteins, such as talin (Dedden et al. 2019; Goult et al. 2013), vinculin (Cohen et al. 2005, 2006), and FAK (Cooper et al. 2003; Lietha et al. 2007), leading to their activation via disruption of autoinhibitory interactions, which can reveal cryptic binding sites for other adhesion proteins (Khan and Goult 2019). The formation of focal adhesions at the leading edge of adherent cells thus enables attachment to the ECM and stabilisation of plasma membrane protrusions, establishing cell polarity that spatially and temporally organises adhesion signalling and facilitates cell migration (Ridley et al. 2003).

The precise spatial associations of adhesion proteins during the assembly of adhesion complexes remain incompletely defined. Studies have used fluorescence microscopy techniques to investigate the sequential recruitment of fluorescently

tagged adhesion proteins to nascent adhesion complexes (Bachir et al. 2014; Zaidel-Bar et al. 2003). Results from these studies suggested a model of hierarchical recruitment of adhesion proteins to ligand-bound integrin, consisting of several successive molecular events (Sun et al. 2014; Zaidel-Bar et al. 2004) (Fig. 9.1b). In addition, several adhesion proteins, including talin and vinculin, appear to associate in pre-complexes prior to accreting in integrin-associated adhesion complexes (Atherton et al. 2020; Bachir et al. 2014; Hoffmann et al. 2014). There is still much to be learned about where and when most adhesion proteins interact during adhesion complex assembly. To complement microscopy-based studies, proteomic approaches can reveal quantitative insights into the compositional dynamics of adhesion complexes during their lifetime (Fig. 9.1c) (see Sect. 9.3).

9.2.3 Focal Adhesion Architecture

The ability of cells to respond to the biophysical properties of the ECM is vital for many cellular processes, including cell adhesion. The linkage between the ECM and the actin cytoskeleton is mediated by mechanosensitive adhesion proteins, which collectively act as a tuneable molecular clutch to transduce mechanical forces across the plasma membrane (Case and Waterman 2015; Sun et al. 2016a). The maturation of cell-matrix adhesions in response to mechanical tension has several important consequences. Enhanced actin polymerisation directly influences the localisation and function of mechanosensitive transcription regulators, such as myocardin-related transcription factor A and Yes-associated protein (Kechagia et al. 2019). In addition, the formation of stress fibres—contractile actomyosin structures—creates a biomechanical connection between the ECM and the nucleus, via integrins and the linker of nucleoskeleton and cytoskeleton complex (Kirby and Lammerding 2018). Through this mechanosensitive connection, mechanical tension can influence various nuclear processes, such as chromatin organisation and gene expression, to regulate an adaptive response to changes in the extracellular environment.

Elegant super-resolution microscopy approaches have shown that focal adhesion proteins are organised into a conserved three-dimensional (3D) nanoscale arrangement. Interferometric analysis of fluorescently tagged adhesion proteins determined that focal adhesions appear to have a multilayered structure, stratified along the axis perpendicular to the plasma membrane (Kanchanawong et al. 2010). An integrin signalling layer within ~30 nm of the plasma membrane contains the adhesion proteins paxillin and FAK, which colocalise with integrin cytoplasmic tails. Actin and the actin-associated proteins α-actinin, vasodilator-stimulated phosphoprotein (VASP), and zyxin localise in an actin regulatory layer more than 50 nm above the plasma membrane. The head domain of the large adaptor protein talin, which binds β integrin cytoplasmic tails, localises in the integrin signalling layer, while its tail domain, which binds actin, localises near the actin regulatory layer. Vinculin resides in an intermediate force transduction layer that spans the region between the integrin signalling and actin regulatory layers (Kanchanawong et al. 2010), but it is initially

recruited proximal to the plasma membrane and then redistributed distally (perpendicular to the plasma membrane) as the focal adhesion matures (Case et al. 2015). Although more difficult to label multiple components of cell-matrix adhesions, cryo-electron tomographic analysis, which enables reconstruction of 3D density maps at subnanometre resolution, revealed the existence of focal adhesion substructures (Patla et al. 2010). These appeared as ring-shaped particles, 20–30 nm in diameter, in the proximity of aligned actin bundles. Intriguingly, the ring-shaped particles were responsive to changes in contractility, suggesting potentially important roles in mechanotransduction (Patla et al. 2010). Together, these high-resolution imaging studies provide clues to the relationships between adhesion complex architecture and the mechanosensitive functions of cell-matrix adhesions at the molecular level (Byron 2011; Morton and Parsons 2011). However, the systems-level organisation of the networks of adhesion complex components and their interactions has yet to be experimentally defined.

9.2.4 Role of ECM in Cell-Matrix Signalling

The ECM is a complex mixture of fibrous proteins (e.g. collagens, elastin), proteoglycans (e.g. perlecan, hyaluronan), and glycoproteins (e.g. laminins, fibronectin). These molecules are organised into multimolecular assemblies, along with soluble growth factors, which are sequestered by components of the ECM. Through its supramolecular structural organisation and interactions with adhesion receptors, the ECM generates vital biophysical and biochemical cues that control cell behaviour. The tissue-specific composition of the networks of ECM molecules tunes the elasticity, viscosity, and tensile strength of the ECM, which confers unique biochemical and topographical features that enable and control cell-matrix adhesion and the functioning of the tissue (Mouw et al. 2014).

Remodelling of ECM, while central to many physiological processes, such as wound healing and angiogenesis, is also associated with pathological processes, such as fibrosis and cancer cell invasion. Indeed, a feature of tumourigenesis is desmoplasia, characterised by altered ECM organisation and increased ECM deposition (Pickup et al. 2014). In addition to serving as a signature of metastatic potential (Naba et al. 2012, 2014), cancer-associated ECM actively influences tumour behaviour by enhancing focal adhesion signalling (Levental et al. 2009). Consequently, it is important to understand how the structure of the ECM, and its regulation and remodelling, influence and are influenced by adhesion signalling networks in health and disease.

The identification and quantification of the components of ECMs represent important steps in understanding how tissue-specific cell-matrix adhesion is regulated. A database of over 1000 ECM and ECM-associated components, termed the matrisome, has been defined by in-silico and manually curated cataloguing (Naba et al. 2012, 2016). Advances in the isolation and proteomic analysis of the ECM have enabled detailed characterisation of ECM composition in various tissues

(Byron et al. 2013; Taha and Naba 2019). For example, in the case of the basement membrane, proteomic studies have enabled the quantification of ECM subproteomes in 39 distinct tissue types, revealing a notable molecular heterogeneity among different basement membranes (Randles et al. 2017). Multiple proteomic analyses of ECM, representing multiple normal and diseased tissue types, have been curated in the MatrisomeDB database (Naba et al. 2016; Shao et al. 2020). In combination with MatrixDB, an interactive database that reports and visualises interactions between ECM components (Clerc et al. 2019), these tools serve as valuable resources for the interrogation of ECM proteomic datasets and will aid in the future contextualisation of cell-matrix adhesion data.

9.3 Proteomic Analysis of Cell-Matrix Adhesion

Although microscopy-based studies have revealed key insights into candidate adhesion proteins and focal adhesion architecture, advances in mass spectrometry-based proteomics enable appreciation of the scale and complexity of the sets of proteins recruited to focal adhesions. The latest literature-curated database of adhesome components, which is derived largely from microscopy-based studies using multiple cell types, catalogues 232 adhesion proteins (Winograd-Katz et al. 2014); an integrated database of adhesion complex proteomes, termed a meta-adhesome, now contains over 2400 proteins (Horton et al. 2015). This marked upward shift in adhesome scale indicates the potential for non-candidate-driven "omics" approaches to characterise more fully the molecules mediating cell adhesion. In this section, proteomic studies that have advanced our understanding of cell-matrix adhesion are discussed.

9.3.1 Isolation of Adhesion Complexes

Integrin-associated adhesion complexes exist as labile multiprotein complexes that link the ECM, via the plasma membrane, to the cytoskeleton. Consequently, adhesion complex purification by conventional biochemical isolation approaches, such as immunoprecipitation, has been confounded by technical limitations. The development of several biochemical methods for the isolation of adhesion complexes has enabled their characterisation by mass spectrometry-based proteomics (Byron 2018; Byron et al. 2011; Geiger and Zaidel-Bar 2012; Jones et al. 2015; Kuo et al. 2012; Li Mow Chee and Byron 2021; Manninen and Varjosalo 2017; Robertson et al. 2017). Initial mass spectrometric analyses of adhesion complex composition using these isolation methods identified substantially more proteins than had previously been associated with focal adhesions, revealing the surprising complexity of adhesion networks (Humphries et al. 2009; Kuo et al. 2011; Schiller et al. 2011). These datasets provided valuable resources, serving as starting points for further research

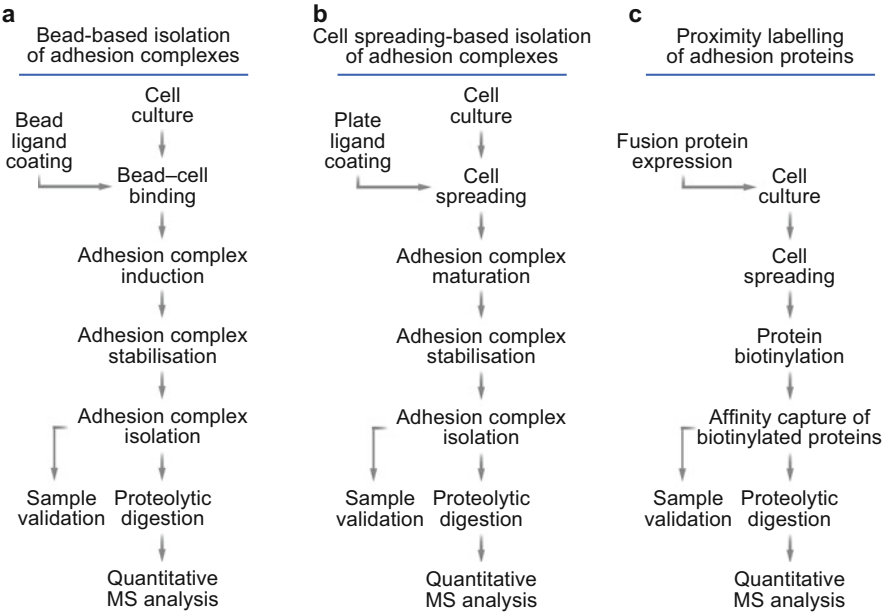

Fig. 9.2 Workflows for the isolation and proteomic analysis of adhesion protein complexes. (**a**) Bead-mediated isolation of integrin-associated adhesion complexes. Cells are incubated with ligand-coated microbeads in suspension to induce adhesion complex formation. (**b**) 2D substrate-mediated isolation of adhesion complexes. Cells are allowed to adhere to ligand-coated dishes to induce adhesion complex maturation. (**c**) Proximity-dependent biotinylation-mediated enrichment of adhesion proteins. The adhesion protein of interest is tagged with, for example, BirA to enable proximity labelling by BioID

to improve understanding of cell-matrix adhesion (Byron et al. 2011; Danen 2009; Gallegos et al. 2011; Schiller and Fässler 2013).

There are three main approaches for the biochemical isolation of adhesion complex-associated proteins. The first approach uses an extracellular integrin ligand (or activating anti-integrin antibody) immobilised on magnetic microbeads, adapted from previous bead-based methods designed to mimic cell-matrix adhesion (Miyamoto et al. 1995; Plopper and Ingber 1993). Adhesion complex formation is induced by binding of the ligand-conjugated beads to cells in suspension, and adhesion complexes are then stabilised using a membrane-permeable, cleavable crosslinker and purified using a combination of carefully optimised sonication and nonionic detergent extraction (Fig. 9.2a). This ligand affinity purification method enables gentle and rapid isolation of labile integrin-associated adhesion complexes (Bass et al. 2011; Byron et al. 2012a, 2015; Humphries et al. 2009), and it can be modified to capture different stages of adhesion complex assembly (Horton et al. 2015).

The second approach allows cells to spread on a defined ECM substrate to induce formation of cell-matrix adhesions. These sites can be stabilised using a crosslinker,

if necessary, after which plasma membranes are burst or solubilised using a hypotonic lysis buffer or a detergent-containing lysis buffer, respectively, and hydrodynamic force (e.g. high-shear flow washing) is applied to remove cell bodies (Fig. 9.2b). This set of related cell spreading-based methods enables isolation of cell-matrix adhesions from adherent cells in culture (Ajeian et al. 2016; Atkinson et al. 2018; Horton et al. 2015; Kuo et al. 2011; Lock et al. 2018, Ng et al. 2014; Robertson et al. 2015; Schiller et al. 2011, 2013).

The third approach uses a proximity-dependent biotinylation technique, such as BioID (Roux et al. 2018). This strategy provides an alternative to ligand affinity purification of adhesion complexes, instead labelling and enriching for proteins that associate in the vicinity of the tagged adhesion protein of interest (Fig. 9.2c). Most BioID protocols require a biotin labelling period of many hours, so will necessarily capture protein associations that occur during that extended timeframe, whereas more efficient proximity labelling methods, such as TurboID (Branon et al. 2018), offer the opportunity for higher temporal resolution. In addition to capturing strong protein-protein interactions, this approach permits the identification of weak and transient protein associations that are a feature of labile adhesion protein complexes (Chastney et al. 2020; Dong et al. 2016; Hennigan et al. 2019; Myllymäki et al. 2019; Rahikainen et al. 2019; Van Itallie et al. 2014).

9.3.2 Proteomic Characterisation of Adhesion Complexes

The isolation of subcellular fractions enriched for adhesion complexes permits the unbiased characterisation of their composition using global analytical approaches such as mass spectrometry. Techniques for mass spectrometric quantification include label-based methods, such as metabolic labelling (Kani 2017) or chemical isobaric labelling (Gritsenko et al. 2016; Núñez et al. 2017; Zhang and Elias 2017), and label-free methods (Arike and Peil 2014; Helm and Baginsky 2018; Moulder et al. 2016). Label-free quantification, in particular, is broadly applicable to most cell systems and is straightforward to implement, so is commonly used for the analysis of adhesion complex proteomes.

Quantitative proteomic analysis has been used to interrogate the composition of adhesion complexes under various experimental conditions, including those induced by ligand engagement of different integrin heterodimers. Affinity purification of adhesion complexes induced by the integrin ligands fibronectin and vascular cell adhesion molecule 1 (VCAM-1) were enriched for α5β1 and α4β1 integrins, respectively (Humphries et al. 2009). Proteomic analysis identified overlapping (core) and receptor-specific subnetworks of adhesion proteins, with a more expansive network induced by fibronectin than by VCAM-1, perhaps reflecting the requirement for greater robustness in fibronectin-based adhesions as compared to less extensive VCAM-1-based interactions involved in, for example, motile behaviour of leukocytes at sites of inflammation (Byron et al. 2011; Humphries et al. 2009). Proteomic analysis of adhesion complexes isolated from cells expressing chimeric α4β1

integrins, in which the cytoplasmic tail of α4 integrin was replaced with that of α5 integrin, enabled the examination of α integrin subunit-dependent adhesion protein recruitment, all triggered by the same engagement of VCAM-1 with the α4 integrin extracellular domain (Byron et al. 2012a). Isolation of fibronectin-induced adhesion complexes from pan-integrin-null fibroblasts that were reconstituted with β1, αV, or β1 and αV integrins enabled assessment of the specificity of adhesion complex composition of these integrin classes (Schiller et al. 2013). Proteomic analysis implicated specific and synergistic roles for β1 integrins (through a RhoA–ROCK–myosin II pathway) and αV integrins (through a GEF-H1–RhoA–mDia1 pathway), with the two integrin classes cooperating to sense the rigidity of the ECM (Schiller and Fässler 2013; Schiller et al. 2013).

Myosin II-mediated tension regulates recruitment of adhesion proteins such as vinculin and FAK to nascent adhesions to reinforce the linkage of the cytoskeleton to the ECM (Pasapera et al. 2010). To gain insights into focal adhesion maturation, proteomic analyses of adhesion complexes isolated from cells treated with a myosin II inhibitor revealed systems-level changes in the integrin adhesome upon tension release and focal adhesion turnover (Horton et al. 2015; Kuo et al. 2011; Schiller et al. 2011). Recruitment of β-PIX (also known as ARHGEF7), a Rac1 and Cdc42 guanine nucleotide exchange factor, to disassembling adhesion complexes was identified upon myosin II inhibition with blebbistatin, suggesting a negative regulatory role in focal adhesion maturation (Kuo et al. 2011). β-PIX silencing impaired adhesion complex disassembly and promoted the formation of medium-to-large focal adhesions in migrating cells (Hiroyasu et al. 2017; Kuo et al. 2011). In contrast, the recruitment to adhesion complexes of proteins containing LIN-11, Isl1, and MEC-3 (LIM) domains, such as lipoma-preferred partner, paxillin, PINCH, and zyxin, was impaired in the presence of blebbistatin (Horton et al. 2015; Schiller et al. 2011). Some of these LIM domain-containing proteins are also recruited to stress fibres in response to cell stretching (Kim-Kaneyama et al. 2005; Yoshigi et al. 2005), with a growing body of evidence implicating LIM domain-containing proteins in mediating mechanotransduction events (Smith et al. 2014).

A key advance in the understanding of the molecular complexity and diversity of focal adhesions was attained from the integration of multiple fibronectin-based cell-matrix adhesion subproteomes from different cell types, resulting in the development of a meta-adhesome in silico (Horton et al. 2016a). Analysis and refinement of the 2412-protein meta-adhesome enabled the construction of a consensus adhesome, which defines a core subset of 60 frequently identified adhesion proteins (Horton et al. 2015). Based on known protein interactions and functions, the consensus adhesome was organised into four distinct axes, centred around integrin-linked kinase (ILK)–PINCH–kindlin, FAK–paxillin, talin–vinculin, and α-actinin–zyxin–VASP protein subnetworks. Moreover, this study moved beyond the analysis of steady-state adhesion complexes to assess the dynamics of adhesome assembly and turnover (Fig. 9.3a). In time-course experiments that (1) allowed cells to engage with fibronectin to promote adhesion complex assembly or (2) treated cells with nocodazole to trigger adhesion complex disassembly upon nocodazole removal, time-resolved adhesion complexes were isolated. While this approach—as for all

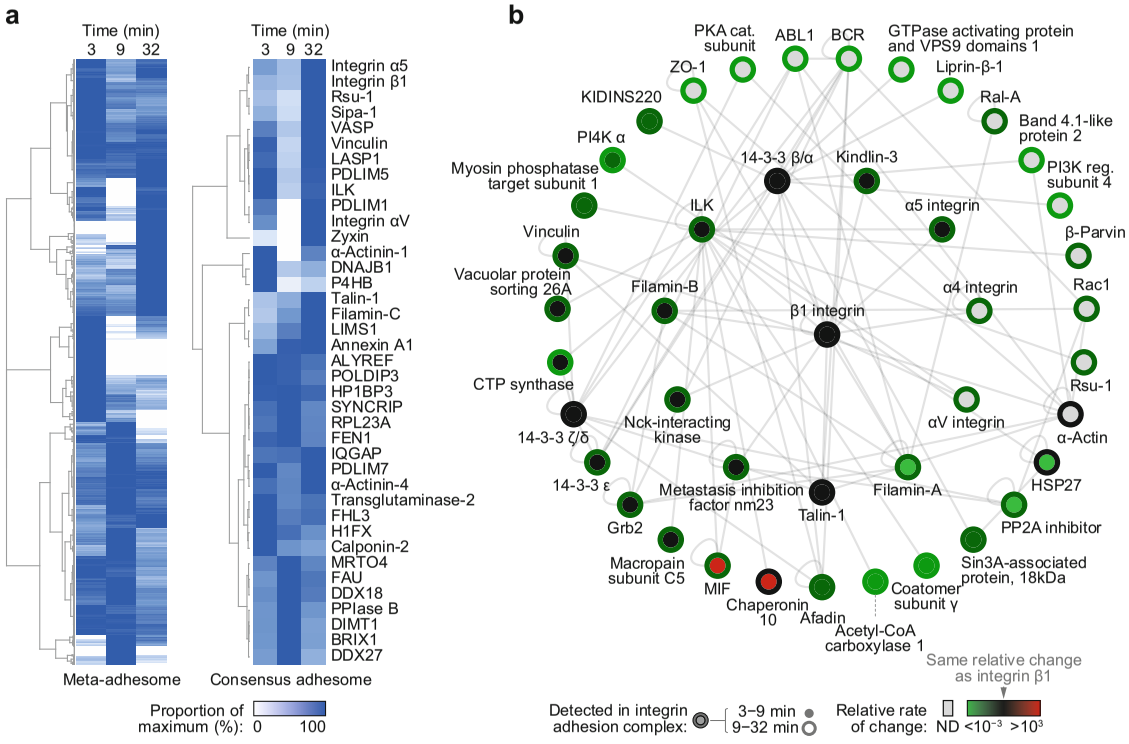

Fig. 9.3 Proteomic analysis of the dynamics of adhesion protein recruitment. (**a**) Analysis of fibronectin-induced adhesion complex formation using hierarchical clustering (Horton et al. 2015). Quantitative mass spectrometry data were expressed as the proportion of maximum relative abundance in the time-course for each identified meta-adhesome protein (left) and consensus adhesome protein (right). (**b**) Interaction network analysis of fibronectin-induced adhesion complex formation (Byron 2018). Quantitative mass spectrometry data were expressed as the rate of change of abundance between adjacent time points relative to the rate of change of β1 integrin abundance, which was used to colour the nodes (proteins) of the network. Reported protein interactions are represented as edges (grey connecting lines). The two-hop β1 integrin neighbourhood (proteins within two putative direct interactions of β1 integrin) is shown. (Reprinted by permission from Springer Nature Customer Service Centre GmbH: Springer Nature, Methods in Molecular Biology: Proteomic profiling of integrin adhesion complex assembly by Byron (2018) in Protein Complex Assembly by Marsh (ed) © 2018 Springer Science+Business Media LLC)

current approaches for the biochemical isolation of integrin-associated adhesion complexes—yields a heterogeneous population of adhesion complex structures, proteomic analysis of these pools of complexes nonetheless enabled the temporal profiling of compositional snapshots of consensus adhesome components (Byron 2018; Horton et al. 2015). This analysis revealed distinct dynamics of adhesion protein recruitment, with many core adhesion proteins quantified in higher abundance at more advanced stages of adhesion complex assembly (Horton et al. 2015). In addition to providing a comprehensive resource describing focal adhesion composition, which can be used as a framework for further interrogation of adhesion-associated proteins and integration with other datasets, these findings underscore the utility of quantitative proteomic approaches in characterising temporally resolved adhesion complex composition.

9.3.3 Proximal Adhesion Protein Associations

Proximity biotinylation approaches, such as BioID, provide an alternative approach for the labelling and enrichment of proteins that associate in the vicinity of a tagged adhesion protein of interest. This is achieved by fusing a mutated biotin ligase, BirA*, to the adhesion protein of interest to promiscuously biotinylate proximal untagged proteins. Biotin-labelled proteins are then affinity purified from a cell lysate and analysed by quantitative mass spectrometry. While proximity biotinylation does not distinguish between direct binding, indirect interaction, or vicinal association, it does permit the identification of weak and transient protein associations, which is particularly relevant for labile integrin-associated adhesion complexes. Furthermore, as the biotin ligase activity of BirA* is spatially constrained by the localisation of its fusion protein, this approach can identify proteins within 15 nm of the adhesion protein of interest (Kim et al. 2014), which, although providing neither temporal resolution nor the subcellular spatial information afforded by subcellular fractionation or light microscopy, offers intermolecular spatial resolution comparable to that of super-resolution microscopy (Kanchanawong et al. 2010).

Proximal interactomes of several adhesion proteins have been determined using BioID. Proximity labelling identified distinct subsets of proteins closely associated with paxillin and kindlin-2, including those that were specifically related to the additional function of kindlin-2 at cell-cell junctions (Dong et al. 2016). Furthermore, overlapping proteins proximal to both paxillin and kindlin-2 were identified, such as Kank2, which also binds directly to talin and was identified in a talin BioID dataset (Dong et al. 2016; Rahikainen et al. 2019; Sun et al. 2016b). Kank2 was also associated with BirA-tagged Merlin, an ezrin-radixin-moesin family protein that was identified in the paxillin and kindlin-2 BioID datasets (Dong et al. 2016; Hennigan et al. 2019). Indeed, several adhesome proteins, including α-actinin, talin, and vinculin, were identified as proximal to Merlin, in addition to an enrichment of

cell-cell adhesion proteins, suggesting roles for Merlin in cell junctional complexes (Hennigan et al. 2019).

BirA tagging of the N- and C-termini (extracellular and cytoplasmic portions, respectively) of β4 integrin was used to interrogate proteins associated with this hemidesmosomal integrin (Myllymäki et al. 2019). In addition to the β4 integrin heterodimeric partner α6 integrin, BirA fusion to the β4 integrin cytoplasmic tail enabled mass spectrometric identification of cell-cell junction scaffold proteins, such as AHNAK and utrophin, but revealed limited overlap with the consensus adhesome, in keeping with the role of α6β4 integrin as a core component of hemidesmosomes, which are structurally distinct from focal adhesions. However, the extracellularly tagged β4 integrin was not efficiently expressed on the cell surface, possibly as its folding or ability to heterodimerise with α6 integrin were impaired (Myllymäki et al. 2019), emphasising the importance of considering the potential structural or steric effects of relatively large protein tags on adhesion proteins, many of which are allosterically regulated.

Recently, a multiplexed BioID approach, analysing the proximal relationships of 16 adhesion proteins, has been taken to begin to define a proximity-dependent adhesome (Chastney et al. 2020). The selected adhesion protein baits of interest, including ILK, PINCH, kindlin-2, FAK, paxillin, vinculin, and zyxin, spanned all four axes of the consensus adhesome (Horton et al. 2015), capturing proximal interactions across the breadth of the adhesome. Bioinformatic analysis of the proteomic datasets identified five clusters of bait adhesion proteins, which were enriched for associated proteins with roles in cell-matrix adhesion but also had functional distinctions (Chastney et al. 2020). The clusters of bait proteins broadly represented the four axes of the consensus adhesome (Horton et al. 2015), providing further evidence of putative functional modules in adhesome networks. Furthermore, the proximal protein associations that contributed to these clusters correlated well with the stratified layers of focal adhesion organisation reported by super-resolution microscopy (Kanchanawong et al. 2010), supporting the hypothesis that different nanoscale zones of focal adhesions have distinct roles in cell-matrix adhesion. Indeed, mapping of bait-dependent biotinylation sites on the large adaptor protein talin revealed that adhesion protein baits preferentially bound to either its N- or C-terminus (Chastney et al. 2020), reinforcing the model of focal adhesion architecture in which talin participates in interactions across multiple layers of the focal adhesion substructure (Kanchanawong et al. 2010). Thus, by extending beyond the repertoires of individual adhesion protein-proximal interactomes to large-scale interrogation of spatially defined adhesion protein associations, new insights into adhesome network interactions and focal adhesion architecture can be inferred.

9.3.4 Phosphoproteomic Analysis of Adhesion Signalling

Focal adhesions are characterised by enriched zones of tyrosine phosphorylation (Ballestrem et al. 2006; Iyer et al. 2005; Kirchner et al. 2003), and protein

phosphorylation plays a central role in the spatiotemporal regulation of adhesion signalling (Lopez-Sanchez et al. 2015; Pasapera et al. 2015; Qu et al. 2014; Wu et al. 2015; Zaidel-Bar et al. 2007b). For example, following cell-matrix contact, the tyrosine kinase FAK is recruited, via paxillin and talin, to nascent adhesions. There, upon membrane-induced release of autoinhibition, FAK dimers trans-autophosphorylate residue tyrosine-397 (Acebrón et al. 2020). Src, another tyrosine kinase, then binds the phosphorylated tyrosine-397 of FAK and phosphorylates the activation loop of the FAK kinase domain to induce FAK catalytic activation (Lietha et al. 2007). Src then autophosphorylates residue tyrosine-416, thereby enhancing its own kinase activity, and proceeds to phosphorylate proximal adhesion proteins to initiate multiple downstream signalling pathways.

Proteomic analysis of adhesion complexes isolated in the presence or absence of FAK and Src inhibitors revealed that adhesive signalling through FAK and Src appears to be uncoupled from focal adhesion composition and maturation, suggesting that a proportion of phosphotyrosine signalling by focal adhesions is independent of their molecular composition (Horton et al. 2016b). While FAK and Src are central tyrosine kinases in adhesion signalling, 95 protein kinases (approximately 18% of the human kinome) have been identified in adhesome databases in total (Schoenherr et al. 2018), suggesting the occurrence of diverse phosphoprotein signalling events at cell-matrix adhesions.

Understanding how kinase signalling contributes to adhesion complex dynamics is complicated by a number of challenges in analysing endogenous protein phosphorylation, including the low abundance of phosphoproteins, the transient nature and low stoichiometry of protein phosphorylation, and the hydrophilicity of resultant phosphopeptides (following proteolytic digestion of phosphoproteins). Together, these can limit the sensitivity of mass spectrometric analyses (Steen et al. 2006). To overcome these challenges, selective phosphopeptide (or phosphoprotein) enrichment is important to permit increased coverage of the phosphoproteome (Leitner 2016); without it, very few phosphorylation events are detected in isolated adhesion complexes (Robertson et al. 2017).

To understand the range of phosphorylation events in adhesion proteins, titanium dioxide beads were used to enrich for phosphopeptides from proteolytically digested adhesion complex isolations, and approximately half of adhesion proteins identified by mass spectrometry were found to be phosphorylated (Robertson et al. 2015). These data also showed that phosphorylation events in adhesion complexes can arise from adhesion-induced phosphorylation of resident adhesion complex proteins or from recruitment of constitutively phosphorylated proteins to adhesion complexes (Robertson et al. 2015). Kinase prediction analysis of the phospho-adhesome identified Cdk1 as a likely kinase involved in the phosphorylation of adhesion proteins. Indeed, inhibition of Cdk1 resulted in loss of paxillin-containing focal adhesions and partial depolymerisation of the actin cytoskeleton, demonstrating the role of Cdk1 in the formation of cell-matrix adhesions (Robertson et al. 2015). Further study has implicated Cdk1 in the regulation of cell adhesion during the cell cycle, allowing adherent cells to enter mitosis through the disassembly of adhesion complexes via Cdk1 inhibition (Jones et al. 2018). Phosphoproteomics has thus revealed novel

indications of the pathways through which cell adhesion proteins are linked to other cellular processes, such as the cell cycle. Further analysis of the range of post-translational modifications that influence adhesion protein function could provide tremendous insights into cell-matrix adhesion signalling.

9.4 Perspectives and Future Directions

The identification and quantification of many proteins by mass spectrometric analysis of integrin-associated adhesion complexes results in high-dimensional datasets that can be challenging to interpret, especially when testing multiple experimental conditions in parallel. Using databases of reported and predicted protein interactions, interaction network analysis has been used extensively to add biological context to adhesion complex proteomes. Experimentally defined adhesomes can be mapped and visualised as graphs of proteins (represented by nodes of the graph) and partitioned into subnetworks or communities of associated proteins, driven by algorithms that compute clusters of proteins based on their protein-protein connections (represented by edges that connect nodes) (Li Mow Chee and Byron 2021). Networks derived from the meta-adhesome and consensus adhesome can provide more coherent representations of potential associations between adhesion complex components and thus serve as useful base graphs—starting points from which to threshold and probe mapped adhesion complex proteomic data (Fig. 9.3b). As the size, complexity, and resolution of characterised adhesome networks continue to increase, examination of the underlying network structure and prediction of novel biological outcomes from complex adhesome network architecture will benefit from machine learning approaches to extract new functional properties of cell-matrix adhesion networks.

Current biochemical and proteomic approaches for the analysis of integrin-associated adhesion complexes are limited to compositional snapshots of heterogeneous adhesion complex structures from a cell population. In the future, techniques such as imaging mass spectrometry may permit the analysis of specific cell-matrix adhesion sites to refine our understanding of native adhesion complex composition. In addition, while cell-matrix adhesions have been observed in cells migrating in a 3D ECM (Kubow and Horwitz 2011), ex vivo (van Geemen et al. 2014), and in vivo (Gunawan et al. 2019), adhesome characterisation has thus far been limited to cells cultured on 2D substrates. Technologies such as proximity labelling, which can be applied in vivo (Branon et al. 2018), may be able to overcome this challenge.

Many of the proteins in the literature-curated adhesome or the consensus adhesome have well-documented roles at cell-matrix adhesions. Under certain experimental conditions, however, some adhesion proteins have been described to localise and function at sites other than focal adhesions. Noncanonical roles for adhesion proteins are emerging as important ancillary (or sometimes primary) functions for adhesion proteins. For example, using a proteomic approach, regulator of chromosome condensation 2 (RCC2; also known as TD-60) was found to reside

in specific integrin ligand-induced adhesion complexes, enabling it to control directional cell migration via restriction of Rac1 and Arf6 GTPase activity (Byron et al. 2012a; Humphries et al. 2009). Yet the canonical role of RCC2 is as a component of the mitotic machinery, from where it regulates prometaphase-to-metaphase progression (Mollinari et al. 2003). Indeed, recent evidence indicates that unconventional adhesion complexes, with distinct protein composition from canonical cell-matrix adhesions, mediate cell attachment during mitosis (Dix et al. 2018; Jones et al. 2018; Lock et al. 2018). This example is from a growing body of evidence that suggests that many proteins associated with cell adhesion possess multifunctionality, especially in the context of specific subcellular locations (Byron and Frame 2016; Byron et al. 2012b; Hervy et al. 2006; Kleinschmidt and Schlaepfer 2017), potentially acting as moonlighting proteins that influence multiple physiological processes (Byron et al. 2012b; Jeffery 2019).

Several adhesion proteins, such as FAK, paxillin, TRIP6, and zyxin, have been reported to translocate to the nucleus (Byron and Frame 2016; Hervy et al. 2006; Kleinschmidt and Schlaepfer 2017; Wang and Gilmore 2003). Conditions associated with cellular stress, such as oxidative stress or oncogenic stress, appear to promote nuclear shuttling of a number of adhesion proteins. For example, FAK, which can localise to and function in the nucleus (Canel et al. 2017; Golubovskaya et al. 2005; Lim et al. 2008, 2012; Serrels et al. 2015, 2017), accumulates in the nucleus of squamous cell carcinoma cells, but it is not detectable in the nucleus of normal counterpart keratinocytes (Serrels et al. 2015). In the nucleus, FAK scaffolds transcription regulators to modulate cytokine expression, which influences tumour immune evasion and tumour growth (Serrels et al. 2015, 2017). Shuttling of adhesion proteins between cell-matrix adhesions and the nucleus could provide a mechanism for transcriptional response to, and feedback with, the extracellular microenvironment. For many adhesome components, however, their functions in the nucleus and at other subcellular locales, and how these relate to their functions at cell-matrix adhesions, are poorly recognised. Understanding the spatiotemporal regulation of these noncanonical adhesion protein processes remains an area of intense scrutiny.

Summary

Our understanding of the molecules that mediate cell adhesion to the ECM has been advanced by the development of proteomic methodologies for the quantification of integrin-associated adhesion complexes. Within the framework of hypothesis-driven research, such approaches provide valuable opportunities for the comprehensive characterisation of cell-matrix adhesion networks and for the interrogation of adhesion signalling pathways. In addition, these unbiased analyses offer the prospect of discovering new cellular roles for adhesion proteins and systems-level modelling of cell adhesion.

Acknowledgements A.B. was funded by Cancer Research UK. A.D.Y. is supported by a Medical Research Council studentship. A.L. is supported by a Cancer Research UK studentship.

References

Acebrón I, Righetto RD, Schoenherr C, de Buhr S, Redondo P, Culley J, Rodríguez CF, Daday C, Biyani N, Llorca O, Byron A, Chami M, Gräter F, Boskovic J, Frame MC, Stahlberg H, Lietha D (2020) Structural basis of Focal Adhesion Kinase activation on lipid membranes. EMBO J e104743. https://doi.org/10.15252/embj.2020104743

Ajeian JN, Horton ER, Astudillo P, Byron A, Askari JA, Millon-Frémillon A, Knight D, Kimber SJ, Humphries MJ, Humphries JD (2016) Proteomic analysis of integrin-associated complexes from mesenchymal stem cells. Proteomics Clin Appl 10:51–57. https://doi.org/10.1002/prca.201500033

Arike L, Peil L (2014) Spectral counting label-free proteomics. Methods Mol Biol 1156:213–222. https://doi.org/10.1007/978-1-4939-0685-7_14

Askari JA, Buckley PA, Mould AP, Humphries MJ (2009) Linking integrin conformation to function. J Cell Sci 122:165–170. https://doi.org/10.1242/jcs.018556

Atherton P, Lausecker F, Carisey A, Gilmore A, Critchley D, Barsukov I, Ballestrem C (2020) Relief of talin autoinhibition triggers a force-independent association with vinculin. J Cell Biol 219:e201903134. https://doi.org/10.1083/jcb.201903134

Atkinson SJ, Gontarczyk AM, Alghamdi AA, Ellison TS, Johnson RT, Fowler WJ, Kirkup BM, Silva BC, Harry BE, Schneider JG, Weilbaecher KN, Mogensen MM, Bass MD, Parsons M, Edwards DR, Robinson SD (2018) The β3-integrin endothelial adhesome regulates microtubule-dependent cell migration. EMBO Rep 19:e44578. https://doi.org/10.15252/embr.201744578

Bachir AI, Zareno J, Moissoglu K, Plow EF, Gratton E, Horwitz AR (2014) Integrin-associated complexes form hierarchically with variable stoichiometry in nascent adhesions. Curr Biol 24:1845–1853. https://doi.org/10.1016/j.cub.2014.07.011

Ballestrem C, Erez N, Kirchner J, Kam Z, Bershadsky A, Geiger B (2006) Molecular mapping of tyrosine-phosphorylated proteins in focal adhesions using fluorescence resonance energy transfer. J Cell Sci 119:866–875. https://doi.org/10.1242/jcs.02794

Bass MD, Williamson RC, Nunan RD, Humphries JD, Byron A, Morgan MR, Martin P, Humphries MJ (2011) A syndecan-4 hair trigger initiates wound healing through caveolin- and RhoG-regulated integrin endocytosis. Dev Cell 21:681–693. https://doi.org/10.1016/j.devcel.2011.08.007

Bouvard D, Vignoud L, Dupé-Manet S, Abed N, Fournier HN, Vincent-Monegat C, Retta SF, Fassler R, Block MR (2003) Disruption of focal adhesions by integrin cytoplasmic domain-associated protein-1 alpha. J Biol Chem 278:6567–6574. https://doi.org/10.1074/jbc.M211258200

Bouvard D, Pouwels J, De Franceschi N, Ivaska J (2013) Integrin inactivators: balancing cellular functions in vitro and in vivo. Nat Rev Mol Cell Biol 14:430–442. https://doi.org/10.1038/nrm3599

Branon TC, Bosch JA, Sanchez AD, Udeshi ND, Svinkina T, Carr SA, Feldman JL, Perrimon N, Ting AY (2018) Efficient proximity labeling in living cells and organisms with TurboID. Nat Biotechnol 36:880–887. https://doi.org/10.1038/nbt.4201

Byron A (2011) Analyzing the anatomy of integrin adhesions. Sci Signal 4:jc3. https://doi.org/10.1126/scisignal.2001896

Byron A (2018) Proteomic profiling of integrin adhesion complex assembly. Methods Mol Biol 1764:193–236. https://doi.org/10.1007/978-1-4939-7759-8_13

Byron A, Frame MC (2016) Adhesion protein networks reveal functions proximal and distal to cell-matrix contacts. Curr Opin Cell Biol 39:93–100. https://doi.org/10.1016/j.ceb.2016.02.013

Byron A, Humphries JD, Askari JA, Craig SE, Mould AP, Humphries MJ (2009) Anti-integrin monoclonal antibodies. J Cell Sci 122:4009–4011. https://doi.org/10.1242/jcs.056770

Byron A, Morgan MR, Humphries MJ (2010) Adhesion signalling complexes. Curr Biol 20: R1063–R1067. https://doi.org/10.1016/j.cub.2010.10.059

Byron A, Humphries JD, Bass MD, Knight D, Humphries MJ (2011) Proteomic analysis of integrin adhesion complexes. Sci Signal 4:pt2. https://doi.org/10.1126/scisignal.2001827

Byron A, Humphries JD, Craig SE, Knight D, Humphries MJ (2012a) Proteomic analysis of α4β1 integrin adhesion complexes reveals α-subunit-dependent protein recruitment. Proteomics 12:2107–2114. https://doi.org/10.1002/pmic.20110048

Byron A, Humphries JD, Humphries MJ (2012b) Alternative cellular roles for proteins identified using proteomics. J Proteome 75:4184–4185. https://doi.org/10.1016/j.jprot.2012.04.049

Byron A, Humphries JD, Humphries MJ (2013) Defining the extracellular matrix using proteomics. Int J Exp Pathol 94:75–92. https://doi.org/10.1111/iep.12011

Byron A, Askari JA, Humphries JD, Jacquemet G, Koper EJ, Warwood S, Choi CK, Stroud MJ, Chen CS, Knight D, Humphries MJ (2015) A proteomic approach reveals integrin activation state-dependent control of microtubule cortical targeting. Nat Commun 6:6135. https://doi.org/10.1038/ncomms7135

Calderwood DA, Campbell ID, Critchley DR (2013) Talins and kindlins: partners in integrin-mediated adhesion. Nat Rev Mol Cell Biol 14:503–517. https://doi.org/10.1038/nrm3624

Campbell ID, Humphries MJ (2011) Integrin structure, activation, and interactions. Cold Spring Harb Perspect Biol 3:a004994. https://doi.org/10.1101/cshperspect.a004994

Canel M, Byron A, Sims AH, Cartier J, Patel H, Frame MC, Brunton VG, Serrels B, Serrels A (2017) Nuclear FAK and Runx1 cooperate to regulate IGFBP3, cell-cycle progression, and tumor growth. Cancer Res 77:5301–5312. https://doi.org/10.1158/0008-5472.CAN-17-0418

Case LB, Waterman CM (2015) Integration of actin dynamics and cell adhesion by a three-dimensional, mechanosensitive molecular clutch. Nat Cell Biol 17:955–963. https://doi.org/10.1038/ncb3191

Case LB, Baird MA, Shtengel G, Campbell SL, Hess HF, Davidson MW, Waterman CM (2015) Molecular mechanism of vinculin activation and nanoscale spatial organization in focal adhesions. Nat Cell Biol 17:880–892. https://doi.org/10.1038/ncb3180

Chastney MR, Lawless C, Humphries JD, Warwood S, Jones MC, Knight D, Jorgensen C, Humphries MJ (2020) Topological features of integrin adhesion complexes revealed by multiplexed proximity biotinylation. J Cell Biol 219:e202003038. https://doi.org/10.1083/jcb.202003038

Clerc O, Deniaud M, Vallet SD, Naba A, Rivet A, Perez S, Thierry-Mieg N, Ricard-Blum S (2019) MatrixDB: integration of new data with a focus on glycosaminoglycan interactions. Nucleic Acids Res 47:D376–D381. https://doi.org/10.1093/nar/gky1035

Cohen DM, Chen H, Johnson RP, Choudhury B, Craig SW (2005) Two distinct head-tail interfaces cooperate to suppress activation of vinculin by talin. J Biol Chem 280:17109–17117. https://doi.org/10.1074/jbc.M414704200

Cohen DM, Kutscher B, Chen H, Murphy DB, Craig SW (2006) A conformational switch in vinculin drives formation and dynamics of a talin-vinculin complex at focal adhesions. J Biol Chem 281:16006–16015. https://doi.org/10.1074/jbc.M600738200

Cooper J, Giancotti FG (2019) Integrin signaling in cancer: mechanotransduction, stemness, epithelial plasticity, and therapeutic resistance. Cancer Cell 35:347–367. https://doi.org/10.1016/j.ccell.2019.01.007

Cooper LA, Shen TL, Guan JL (2003) Regulation of focal adhesion kinase by its amino-terminal domain through an autoinhibitory interaction. Mol Cell Biol 23:8030–8041. https://doi.org/10.1128/mcb.23.22.8030-8041.2003

Danen EH (2009) Integrin proteomes reveal a new guide for cell motility. Sci Signal 2:pe58. https://doi.org/10.1126/scisignal.289pe58

Dedden D, Schumacher S, Kelley CF, Zacharias M, Biertümpfel C, Fässler R, Mizuno N (2019) The architecture of talin1 reveals an autoinhibition mechanism. Cell 179:120–131.e13. https://doi.org/10.1016/j.cell.2019.08.034

Degani S, Balzac F, Brancaccio M, Guazzone S, Retta SF, Silengo L, Eva A, Tarone G (2002) The integrin cytoplasmic domain-associated protein ICAP-1 binds and regulates rho family GTPases during cell spreading. J Cell Biol 156:377–387. https://doi.org/10.1083/jcb.200108030

Devreotes P, Horwitz AR (2015) Signaling networks that regulate cell migration. Cold Spring Harb Perspect Biol 7:a005959. https://doi.org/10.1101/cshperspect.a005959

Dix CL, Matthews HK, Uroz M, McLaren S, Wolf L, Heatley N, Win Z, Almada P, Henriques R, Boutros M, Trepat X, Baum B (2018) The role of mitotic cell-substrate adhesion re-modeling in animal cell division. Dev Cell 45:132–145.e3. https://doi.org/10.1016/j.devcel.2018.03.009

Dong JM, Tay FP, Swa HL, Gunaratne J, Leung T, Burke B, Manser E (2016) Proximity biotinylation provides insight into the molecular composition of focal adhesions at the nanometer scale. Sci Signal 9:rs4. https://doi.org/10.1126/scisignal.aaf3572

Doyle AD, Yamada KM (2016) Mechanosensing via cell-matrix adhesions in 3D microenvironments. Exp Cell Res 343:60–66. https://doi.org/10.1016/j.yexcr.2015.10.033

Gallegos L, Ng MR, Brugge JS (2011) The myosin-II-responsive focal adhesion proteome: a tour de force? Nat Cell Biol 13:344–346. https://doi.org/10.1038/ncb0411-344

Geiger B, Yamada KM (2011) Molecular architecture and function of matrix adhesions. Cold Spring Harb Perspect Biol 3:a005033. https://doi.org/10.1101/cshperspect.a005033

Geiger T, Zaidel-Bar R (2012) Opening the floodgates: proteomics and the integrin adhesome. Curr Opin Cell Biol 24:562–568. https://doi.org/10.1016/j.ceb.2012.05.004

Ginsberg MH, Partridge A, Shattil SJ (2005) Integrin regulation. Curr Opin Cell Biol 17:509–516. https://doi.org/10.1016/j.ceb.2005.08.010

Golubovskaya VM, Finch R, Cance WG (2005) Direct interaction of the N-terminal domain of focal adhesion kinase with the N-terminal transactivation domain of p53. J Biol Chem 280:25008–25021. https://doi.org/10.1074/jbc.M414172200

Goult BT, Xu XP, Gingras AR, Swift M, Patel B, Bate N, Kopp PM, Barsukov IL, Critchley DR, Volkmann N, Hanein D (2013) Structural studies on full-length talin1 reveal a compact autoinhibited dimer: implications for talin activation. J Struct Biol 184:21–32. https://doi.org/10.1016/j.jsb.2013.05.014

Green HJ, Brown NH (2019) Integrin intracellular machinery in action. Exp Cell Res 378:226–231. https://doi.org/10.1016/j.yexcr.2019.03.011

Gritsenko MA, Xu Z, Liu T, Smith RD (2016) Large-scale and deep quantitative proteome profiling using isobaric labeling coupled with two-dimensional LC-MS/MS. Methods Mol Biol 1410:237–247. https://doi.org/10.1007/978-1-4939-3524-6_14

Gunawan F, Gentile A, Fukuda R, Tsedeke AT, Jiménez-Amilburu V, Ramadass R, Iida A, Sehara-Fujisawa A, Stainier DYR (2019) Focal adhesions are essential to drive zebrafish heart valve morphogenesis. J Cell Biol 218:1039–1054. https://doi.org/10.1083/jcb.201807175

Hamidi H, Ivaska J (2018) Every step of the way: integrins in cancer progression and metastasis. Nat Rev Cancer 18:533–548. https://doi.org/10.1038/s41568-018-0038-z

Helm S, Baginsky S (2018) MSE for label-free absolute protein quantification in complex proteomes. Methods Mol Biol 1696:235–247. https://doi.org/10.1007/978-1-4939-7411-5_16

Hennigan RF, Fletcher JS, Guard S, Ratner N (2019) Proximity biotinylation identifies a set of conformation-specific interactions between Merlin and cell junction proteins. Sci Signal 12:eaau8749. https://doi.org/10.1126/scisignal.aau8749

Hervy M, Hoffman L, Beckerle MC (2006) From the membrane to the nucleus and back again: bifunctional focal adhesion proteins. Curr Opin Cell Biol 18:524–532. https://doi.org/10.1016/j.ceb.2006.08.006

Hiroyasu S, Stimac GP, Hopkinson SB, Jones JCR (2017) Loss of β-PIX inhibits focal adhesion disassembly and promotes keratinocyte motility via myosin light chain activation. J Cell Sci 130:2329–2343. https://doi.org/10.1242/jcs.196147

Hoffmann JE, Fermin Y, Stricker RL, Ickstadt K, Zamir E (2014) Symmetric exchange of multi-protein building blocks between stationary focal adhesions and the cytosol. eLife 3:e02257. https://doi.org/10.7554/eLife.02257

Horton ER, Byron A, Askari JA, Ng DHJ, Millon-Frémillon A, Robertson J, Koper EJ, Paul NR, Warwood S, Knight D, Humphries JD, Humphries MJ (2015) Definition of a consensus integrin adhesome and its dynamics during adhesion complex assembly and disassembly. Nat Cell Biol 17:1577–1587. https://doi.org/10.1038/ncb3257

Horton ER, Humphries JD, James J, Jones MC, Askari JA, Humphries MJ (2016a) The integrin adhesome network at a glance. J Cell Sci 129:4159–4163. https://doi.org/10.1242/jcs.192054

Horton ER, Humphries JD, Stutchbury B, Jacquemet G, Ballestrem C, Barry ST, Humphries MJ (2016b) Modulation of FAK and Src adhesion signaling occurs independently of adhesion complex composition. J Cell Biol 212:349–364. https://doi.org/10.1083/jcb.201508080

Humphries JD, Byron A, Humphries MJ (2006) Integrin ligands at a glance. J Cell Sci 119:3901–3903. https://doi.org/10.1242/jcs.03098

Humphries JD, Byron A, Bass MD, Craig SE, Pinney JW, Knight D, Humphries MJ (2009) Proteomic analysis of integrin-associated complexes identifies RCC2 as a dual regulator of Rac1 and Arf6. Sci Signal 2:ra51. https://doi.org/10.1126/scisignal.2000396

Humphries JD, Paul NR, Humphries MJ, Morgan MR (2015) Emerging properties of adhesion complexes: what are they and what do they do? Trends Cell Biol 25:388–397. https://doi.org/10.1016/j.tcb.2015.02.008

Huttenlocher A, Ginsberg MH, Horwitz AF (1996) Modulation of cell migration by integrin-mediated cytoskeletal linkages and ligand-binding affinity. J Cell Biol 134:1551–1562. https://doi.org/10.1083/jcb.134.6.1551

Hynes RO (2002) Integrins: bidirectional, allosteric signaling machines. Cell 110:673–687. https://doi.org/10.1016/s0092-8674(02)00971-6

Iyer VV, Ballestrem C, Kirchner J, Geiger B, Schaller MD (2005) Measurement of protein tyrosine phosphorylation in cell adhesion. Methods Mol Biol 294:289–302. https://doi.org/10.1385/1-59259-860-9:289

Jansen KA, Atherton P, Ballestrem C (2017) Mechanotransduction at the cell-matrix interface. Semin Cell Dev Biol 71:75–83. https://doi.org/10.1016/j.semcdb.2017.07.027

Jeffery CJ (2019) Multitalented actors inside and outside the cell: recent discoveries add to the number of moonlighting proteins. Biochem Soc Trans 47:1941–1948. https://doi.org/10.1042/BST20190798

Jones MC, Humphries JD, Byron A, Millon-Frémillon A, Robertson J, Paul NR, Ng DH, Askari JA, Humphries MJ (2015) Isolation of integrin-based adhesion complexes. Curr Protoc Cell Biol 66:9.8.1–9.8.15. https://doi.org/10.1002/0471143030.cb0908s66

Jones MC, Askari JA, Humphries JD, Humphries MJ (2018) Cell adhesion is regulated by CDK1 during the cell cycle. J Cell Biol 217:3203–3218. https://doi.org/10.1083/jcb.201802088

Kanchanawong P, Shtengel G, Pasapera AM, Ramko EB, Davidson MW, Hess HF, Waterman CM (2010) Nanoscale architecture of integrin-based cell adhesions. Nature 468:580–584. https://doi.org/10.1038/nature09621

Kani K (2017) Quantitative proteomics using SILAC. Methods Mol Biol 1550:171–184. https://doi.org/10.1007/978-1-4939-6747-6_13

Kechagia JZ, Ivaska J, Roca-Cusachs P (2019) Integrins as biomechanical sensors of the microenvironment. Nat Rev Mol Cell Biol 20:457–473. https://doi.org/10.1038/s41580-019-0134-2

Khan RB, Goult BT (2019) Adhesions assemble!–autoinhibition as a major regulatory mechanism of integrin-mediated adhesion. Front Mol Biosci 6:144. https://doi.org/10.3389/fmolb.2019.00144

Kim DI, Birendra KC, Zhu W, Motamedchaboki K, Doye V, Roux KJ (2014) Probing nuclear pore complex architecture with proximity-dependent biotinylation. Proc Natl Acad Sci U S A 111:E2453–E2461. https://doi.org/10.1073/pnas.1406459111

Kim-Kaneyama JR, Suzuki W, Ichikawa K, Ohki T, Kohno Y, Sata M, Nose K, Shibanuma M (2005) Uni-axial stretching regulates intracellular localization of Hic-5 expressed in smooth-muscle cells in vivo. J Cell Sci 118:937–949. https://doi.org/10.1242/jcs.01683

Kirby TJ, Lammerding J (2018) Emerging views of the nucleus as a cellular mechanosensor. Nat Cell Biol 20:373–381. https://doi.org/10.1038/s41556-018-0038-y

Kirchner J, Kam Z, Tzur G, Bershadsky AD, Geiger B (2003) Live-cell monitoring of tyrosine phosphorylation in focal adhesions following microtubule disruption. J Cell Sci 116:975–986. https://doi.org/10.1242/jcs.00284

Kleinschmidt EG, Schlaepfer DD (2017) Focal adhesion kinase signaling in unexpected places. Curr Opin Cell Biol 45:24–30. https://doi.org/10.1016/j.ceb.2017.01.003

Kubow KE, Horwitz AR (2011) Reducing background fluorescence reveals adhesions in 3D matrices. Nat Cell Biol 13:3–5. https://doi.org/10.1038/ncb0111-3

Kuo JC, Han X, Hsiao CT, Yates JR 3rd, Waterman CM (2011) Analysis of the myosin-II-responsive focal adhesion proteome reveals a role for β-pix in negative regulation of focal adhesion maturation. Nat Cell Biol 13:383–393. https://doi.org/10.1038/ncb2216

Kuo JC, Han X, Yates JR 3rd, Waterman CM (2012) Isolation of focal adhesion proteins for biochemical and proteomic analysis. Methods Mol Biol 757:297–323. https://doi.org/10.1007/978-1-61779-166-6_19

Leitner A (2016) Enrichment strategies in phosphoproteomics. Methods Mol Biol 1355:105–121. https://doi.org/10.1007/978-1-4939-3049-4_7

Levental KR, Yu H, Kass L, Lakins JN, Egeblad M, Erler JT, Fong SF, Csiszar K, Giaccia A, Weninger W, Yamauchi M, Gasser DL, Weaver VM (2009) Matrix crosslinking forces tumor progression by enhancing integrin signaling. Cell 139:891–906. https://doi.org/10.1016/j.cell.2009.10.027

Li Mow Chee F, Byron A (2021) Network analysis of integrin adhesion complexes. Methods Mol Biol 2217. [in press]

Lietha D, Cai X, Ceccarelli DF, Li Y, Schaller MD, Eck MJ (2007) Structural basis for the autoinhibition of focal adhesion kinase. Cell 129:1177–1187. https://doi.org/10.1016/j.cell.2007.05.041

Lilja J, Zacharchenko T, Georgiadou M, Jacquemet G, De Franceschi N, Peuhu E, Hamidi H, Pouwels J, Martens V, Nia FH, Beifuss M, Boeckers T, Kreienkamp HJ, Barsukov IL, Ivaska J (2017) SHANK proteins limit integrin activation by directly interacting with Rap1 and R-Ras. Nat Cell Biol 19:292–305. https://doi.org/10.1038/ncb3487

Lim ST, Chen XL, Lim Y, Hanson DA, Vo TT, Howerton K, Larocque N, Fisher SJ, Schlaepfer DD, Ilic D (2008) Nuclear FAK promotes cell proliferation and survival through FERM-enhanced p53 degradation. Mol Cell 29:9–22. https://doi.org/10.1016/j.molcel.2007.11.031

Lim ST, Miller NL, Chen XL, Tancioni I, Walsh CT, Lawson C, Uryu S, Weis SM, Cheresh DA, Schlaepfer DD (2012) Nuclear-localized focal adhesion kinase regulates inflammatory VCAM-1 expression. J Cell Biol 197:907–919. https://doi.org/10.1083/jcb.201109067

Liu J, Das M, Yang J, Ithychanda SS, Yakubenko VP, Plow EF, Qin J (2015) Structural mechanism of integrin inactivation by filamin. Nat Struct Mol Biol 22:383–389. https://doi.org/10.1038/nsmb.2999

Lock JG, Jones MC, Askari JA, Gong X, Oddone A, Olofsson H, Göransson S, Lakadamyali M, Humphries MJ, Strömblad S (2018) Reticular adhesions are a distinct class of cell-matrix adhesions that mediate attachment during mitosis. Nat Cell Biol 20:1290–1302. https://doi.org/10.1038/s41556-018-0220-2

Lopez-Sanchez I, Kalogriopoulos N, Lo IC, Kabir F, Midde KK, Wang H, Ghosh P (2015) Focal adhesions are foci for tyrosine-based signal transduction via GIV/Girdin and G proteins. Mol Biol Cell 26:4313–4324. https://doi.org/10.1091/mbc.E15-07-0496

Manninen A, Varjosalo M (2017) A proteomics view on integrin-mediated adhesions. Proteomics 17:1600022. https://doi.org/10.1002/pmic.201600022

McCormick B, Craig HE, Chu JY, Carlin LM, Canel M, Wollweber F, Toivakka M, Michael M, Astier AL, Norton L, Lilja J, Felton JM, Sasaki T, Ivaska J, Hers I, Dransfield I, Rossi AG,

Vermeren S (2019) A negative feedback loop regulates integrin inactivation and promotes neutrophil recruitment to inflammatory sites. J Immunol 203:1579–1588. https://doi.org/10.4049/jimmunol.1900443

Michael M, Parsons M (2020) New perspectives on integrin-dependent adhesions. Curr Opin Cell Biol 63:31–37. https://doi.org/10.1016/j.ceb.2019.12.008

Miyamoto S, Akiyama SK, Yamada KM (1995) Synergistic roles for receptor occupancy and aggregation in integrin transmembrane function. Science 267:883–885. https://doi.org/10.1126/science.7846531

Miyazaki N, Iwasaki K, Takagi J (2018) A systematic survey of conformational states in β1 and β4 integrins using negative-stain electron microscopy. J Cell Sci 131:jcs216754. https://doi.org/10.1242/jcs.216754

Mollinari C, Reynaud C, Martineau-Thuillier S, Monier S, Kieffer S, Garin J, Andreassen PR, Boulet A, Goud B, Kleman JP, Margolis RL (2003) The mammalian passenger protein TD-60 is an RCC1 family member with an essential role in prometaphase to metaphase progression. Dev Cell 5:295–307. https://doi.org/10.1016/s1534-5807(03)00205-3

Moreno-Layseca P, Streuli CH (2014) Signalling pathways linking integrins with cell cycle progression. Matrix Biol 34:144–153. https://doi.org/10.1016/j.matbio.2013.10.011

Morton PE, Parsons M (2011) Dissecting cell adhesion architecture using advanced imaging techniques. Cell Adhes Migr 5:351–359. https://doi.org/10.4161/cam.5.4.16915

Moulder R, Goo YA, Goodlett DR (2016) Label-free quantitation for clinical proteomics. Methods Mol Biol 1410:65–76. https://doi.org/10.1007/978-1-4939-3524-6_4

Mouw JK, Ou G, Weaver VM (2014) Extracellular matrix assembly: a multiscale deconstruction. Nat Rev Mol Cell Biol 15:771–785. https://doi.org/10.1038/nrm3902

Myllymäki SM, Kämäräinen UR, Liu X, Cruz SP, Miettinen S, Vuorela M, Varjosalo M, Manninen A (2019) Assembly of the β4-integrin interactome based on proximal biotinylation in the presence and absence of heterodimerization. Mol Cell Proteomics 18:277–293. https://doi.org/10.1074/mcp.RA118.001095

Naba A, Clauser KR, Hoersch S, Liu H, Carr SA, Hynes RO (2012) The matrisome: in silico definition and in vivo characterization by proteomics of normal and tumor extracellular matrices. Mol Cell Proteomics 11:M111.014647. https://doi.org/10.1074/mcp.M111.014647

Naba A, Clauser KR, Lamar JM, Carr SA, Hynes RO (2014) Extracellular matrix signatures of human mammary carcinoma identify novel metastasis promoters. eLife 3:e01308. https://doi.org/10.7554/eLife.01308

Naba A, Clauser KR, Ding H, Whittaker CA, Carr SA, Hynes RO (2016) The extracellular matrix: tools and insights for the "omics" era. Matrix Biol 49:10–24. https://doi.org/10.1016/j.matbio.2015.06.003

Ng DH, Humphries JD, Byron A, Millon-Frémillon A, Humphries MJ (2014) Microtubule-dependent modulation of adhesion complex composition. PLoS One 9:e115213. https://doi.org/10.1371/journal.pone.0115213

Núñez EV, Domont GB, Nogueira FC (2017) iTRAQ-based shotgun proteomics approach for relative protein quantification. Methods Mol Biol 1546:267–274. https://doi.org/10.1007/978-1-4939-6730-8_23

Pasapera AM, Schneider IC, Rericha E, Schlaepfer DD, Waterman CM (2010) Myosin II activity regulates vinculin recruitment to focal adhesions through FAK-mediated paxillin phosphorylation. J Cell Biol 188:877–890. https://doi.org/10.1083/jcb.200906012

Pasapera AM, Plotnikov SV, Fischer RS, Case LB, Egelhoff TT, Waterman CM (2015) Rac1-dependent phosphorylation and focal adhesion recruitment of myosin IIA regulates migration and mechanosensing. Curr Biol 25:175–186. https://doi.org/10.1016/j.cub.2014.11.043

Patla I, Volberg T, Elad N, Hirschfeld-Warneken V, Grashoff C, Fässler R, Spatz JP, Geiger B, Medalia O (2010) Dissecting the molecular architecture of integrin adhesion sites by cryo-electron tomography. Nat Cell Biol 12:909–915. https://doi.org/10.1038/ncb2095

Pickup MW, Mouw JK, Weaver VM (2014) The extracellular matrix modulates the hallmarks of cancer. EMBO Rep 15:1243–1253. https://doi.org/10.15252/embr.201439246

Plopper G, Ingber DE (1993) Rapid induction and isolation of focal adhesion complexes. Biochem Biophys Res Commun 193:571–578. https://doi.org/10.1006/bbrc.1993.1662

Qu H, Tu Y, Guan JL, Xiao G, Wu C (2014) Kindlin-2 tyrosine phosphorylation and interaction with Src serve as a regulatable switch in the integrin outside-in signaling circuit. J Biol Chem 289:31001–31013. https://doi.org/10.1074/jbc.M114.580811

Rahikainen R, Öhman T, Turkki P, Varjosalo M, Hytönen VP (2019) Talin-mediated force transmission and talin rod domain unfolding independently regulate adhesion signaling. J Cell Sci 132:jcs226514. https://doi.org/10.1242/jcs.226514

Randles MJ, Humphries MJ, Lennon R (2017) Proteomic definitions of basement membrane composition in health and disease. Matrix Biol 57–58:12–28. https://doi.org/10.1016/j.matbio.2016.08.006

Rantala JK, Pouwels J, Pellinen T, Veltel S, Laasola P, Mattila E, Potter CS, Duffy T, Sundberg JP, Kallioniemi O, Askari JA, Humphries MJ, Parsons M, Salmi M, Ivaska J (2011) SHARPIN is an endogenous inhibitor of β1-integrin activation. Nat Cell Biol 13:1315–1324. https://doi.org/10.1038/ncb2340

Ridley AJ, Schwartz MA, Burridge K, Firtel RA, Ginsberg MH, Borisy G, Parsons JT, Horwitz AR (2003) Cell migration: integrating signals from front to back. Science 302:1704–1709. https://doi.org/10.1126/science.1092053

Robertson J, Jacquemet G, Byron A, Jones MC, Warwood S, Selley JN, Knight D, Humphries JD, Humphries MJ (2015) Defining the phospho-adhesome through the phosphoproteomic analysis of integrin signalling. Nat Commun 6:6265. https://doi.org/10.1038/ncomms7265

Robertson J, Humphries JD, Paul NR, Warwood S, Knight D, Byron A, Humphries MJ (2017) Characterization of the phospho-adhesome by mass spectrometry-based proteomics. Methods Mol Biol 1636:235–251. https://doi.org/10.1007/978-1-4939-7154-1_15

Roux KJ, Kim DI, Burke B, May DG (2018) BioID: a screen for protein-protein interactions. Curr Protoc Protein Sci 91:19.23.1–19.23.15. https://doi.org/10.1002/cpps.51

Schiller HB, Fässler R (2013) Mechanosensitivity and compositional dynamics of cell-matrix adhesions. EMBO Rep 14:509–519. https://doi.org/10.1038/embor.2013.49

Schiller HB, Friedel CC, Boulegue C, Fässler R (2011) Quantitative proteomics of the integrin adhesome show a myosin II-dependent recruitment of LIM domain proteins. EMBO Rep 12:259–266. https://doi.org/10.1038/embor.2011.5

Schiller HB, Hermann MR, Polleux J, Vignaud T, Zanivan S, Friedel CC, Sun Z, Raducanu A, Gottschalk KE, Théry M, Mann M, Fässler R (2013) β1- and αv-class integrins cooperate to regulate myosin II during rigidity sensing of fibronectin-based microenvironments. Nat Cell Biol 15:625–636. https://doi.org/10.1038/ncb2747

Schoenherr C, Frame MC, Byron A (2018) Trafficking of adhesion and growth factor receptors and their effector kinases. Annu Rev Cell Dev Biol 34:29–58. https://doi.org/10.1146/annurev-cellbio-100617-062559

Serrels A, Lund T, Serrels B, Byron A, McPherson RC, von Kriegsheim A, Gómez-Cuadrado L, Canel M, Muir M, Ring JE, Maniati E, Sims AH, Pachter JA, Brunton VG, Gilbert N, Anderton SM, Nibbs RJ, Frame MC (2015) Nuclear FAK controls chemokine transcription, Tregs, and evasion of anti-tumor immunity. Cell 163:160–173. https://doi.org/10.1016/j.cell.2015.09.001

Serrels B, McGivern N, Canel M, Byron A, Johnson SC, McSorley HJ, Quinn N, Taggart D, von Kreigsheim A, Anderton SM, Serrels A, Frame MC (2017) IL-33 and ST2 mediate FAK-dependent antitumor immune evasion through transcriptional networks. Sci Signal 10:eaan8355. https://doi.org/10.1126/scisignal.aan8355

Shao X, Taha IN, Clauser KR, Gao YT, Naba A (2020) MatrisomeDB: the ECM-protein knowledge database. Nucleic Acids Res 48:D1136–D1144. https://doi.org/10.1093/nar/gkz849

Shattil SJ, Kim C, Ginsberg MH (2010) The final steps of integrin activation: the end game. Nat Rev Mol Cell Biol 11:288–300. https://doi.org/10.1038/nrm2871

Smith MA, Hoffman LM, Beckerle MC (2014) LIM proteins in actin cytoskeleton mechanoresponse. Trends Cell Biol 24:575–583. https://doi.org/10.1016/j.tcb.2014.04.009

Steen H, Jebanathirajah JA, Rush J, Morrice N, Kirschner MW (2006) Phosphorylation analysis by mass spectrometry: myths, facts, and the consequences for qualitative and quantitative measurements. Mol Cell Proteomics 5:172–181. https://doi.org/10.1074/mcp.M500135-MCP200

Sun Z, Lambacher A, Fässler R (2014) Nascent adhesions: from fluctuations to a hierarchical organization. Curr Biol 24:R801–R803. https://doi.org/10.1016/j.cub.2014.07.061

Sun Z, Guo SS, Fässler R (2016a) Integrin-mediated mechanotransduction. J Cell Biol 215:445–456. https://doi.org/10.1083/jcb.201609037

Sun Z, Tseng HY, Tan S, Senger F, Kurzawa L, Dedden D, Mizuno N, Wasik AA, Thery M, Dunn AR, Fässler R (2016b) Kank2 activates talin, reduces force transduction across integrins and induces central adhesion formation. Nat Cell Biol 18:941–953. https://doi.org/10.1038/ncb3402

Sun Z, Costell M, Fässler R (2019) Integrin activation by talin, kindlin and mechanical forces. Nat Cell Biol 21:25–31. https://doi.org/10.1038/s41556-018-0234-9

Taha IN, Naba A (2019) Exploring the extracellular matrix in health and disease using proteomics. Essays Biochem 63:417–432. https://doi.org/10.1042/EBC20190001

van Geemen D, Smeets MW, van Stalborch AM, Woerdeman LA, Daemen MJ, Hordijk PL, Huveneers S (2014) F-actin-anchored focal adhesions distinguish endothelial phenotypes of human arteries and veins. Arterioscler Thromb Vasc Biol 34:2059–2067. https://doi.org/10.1161/ATVBAHA.114.304180

Van Itallie CM, Tietgens AJ, Aponte A, Fredriksson K, Fanning AS, Gucek M, Anderson JM (2014) Biotin ligase tagging identifies proteins proximal to E-cadherin, including lipoma preferred partner, a regulator of epithelial cell-cell and cell-substrate adhesion. J Cell Sci 127:885–895. https://doi.org/10.1242/jcs.140475

Vicente-Manzanares M, Horwitz AR (2011) Adhesion dynamics at a glance. J Cell Sci 124:3923–3927. https://doi.org/10.1242/jcs.095653

Vicente-Manzanares M, Ma X, Adelstein RS, Horwitz AR (2009) Non-muscle myosin II takes centre stage in cell adhesion and migration. Nat Rev Mol Cell Biol 10:778–790. https://doi.org/10.1038/nrm2786

Wang Y, Gilmore TD (2003) Zyxin and paxillin proteins: focal adhesion plaque LIM domain proteins go nuclear. Biochim Biophys Acta 1593:115–120. https://doi.org/10.1016/s0167-4889(02)00349-x

Wang J, Dong X, Zhao B, Li J, Lu C, Springer TA (2017) Atypical interactions of integrin αVβ8 with pro-TGF-β1. Proc Natl Acad Sci U S A 114:E4168–E4174. https://doi.org/10.1073/pnas.1705129114

Wang J, Su Y, Iacob RE, Engen JR, Springer TA (2019) General structural features that regulate integrin affinity revealed by atypical αVβ8. Nat Commun 10:5481. https://doi.org/10.1038/s41467-019-13248-5

Wehrle-Haller B (2012) Structure and function of focal adhesions. Curr Opin Cell Biol 24:116–124. https://doi.org/10.1016/j.ceb.2011.11.001

Winograd-Katz SE, Fässler R, Geiger B, Legate KR (2014) The integrin adhesome: from genes and proteins to human disease. Nat Rev Mol Cell Biol 15:273–288. https://doi.org/10.1038/nrm3769

Wolfenson H, Lavelin I, Geiger B (2013) Dynamic regulation of the structure and functions of integrin adhesions. Dev Cell 24:447–458. https://doi.org/10.1016/j.devcel.2013.02.012

Wu JC, Chen YC, Kuo CT, Wenshin Yu H, Chen YQ, Chiou A, Kuo JC (2015) Focal adhesion kinase-dependent focal adhesion recruitment of SH2 domains directs SRC into focal adhesions to regulate cell adhesion and migration. Sci Rep 5:18476. https://doi.org/10.1038/srep18476

Yoshigi M, Hoffman LM, Jensen CC, Yost HJ, Beckerle MC (2005) Mechanical force mobilizes zyxin from focal adhesions to actin filaments and regulates cytoskeletal reinforcement. J Cell Biol 171:209–215. https://doi.org/10.1083/jcb.200505018

Zaidel-Bar R, Ballestrem C, Kam Z, Geiger B (2003) Early molecular events in the assembly of matrix adhesions at the leading edge of migrating cells. J Cell Sci 116:4605–4613. https://doi.org/10.1242/jcs.00792

Zaidel-Bar R, Cohen M, Addadi L, Geiger B (2004) Hierarchical assembly of cell-matrix adhesion complexes. Biochem Soc Trans 32:416–420. https://doi.org/10.1042/BST0320416

Zaidel-Bar R, Itzkovitz S, Ma'ayan A, Iyengar R, Geiger B (2007a) Functional atlas of the integrin adhesome. Nat Cell Biol 9:858–867. https://doi.org/10.1038/ncb0807-858

Zaidel-Bar R, Milo R, Kam Z, Geiger B (2007b) A paxillin tyrosine phosphorylation switch regulates the assembly and form of cell-matrix adhesions. J Cell Sci 120:137–148. https://doi.org/10.1242/jcs.03314

Zhang L, Elias JE (2017) Relative protein quantification using tandem mass tag mass spectrometry. Methods Mol Biol 1550:185–198. https://doi.org/10.1007/978-1-4939-6747-6_14

Chapter 10
Integrative Models for TGF-β Signaling and Extracellular Matrix

Nathalie Théret, Jérôme Feret, Arran Hodgkinson, Pierre Boutillier, Pierre Vignet, and Ovidiu Radulescu

Abstract The extracellular matrix (ECM) is the most important regulator of cell-cell communication within tissues. ECM is a complex structure, made up of a wide variety of molecules including proteins, proteglycans and glycoaminoglycans. It contributes to cell signaling through the action of both its constituents and their proteolytic cleaved fragments called matricryptins. In addition, ECM acts as a "reservoir" of growth factors and cytokines and regulates their bioavailability at the cell surface. By controlling cell signaling inputs, ECM plays a key role in regulating cell phenotype (differentiation, proliferation, migration, etc.).

In this context, signaling networks associated with the polypeptide transforming growth factor TGF-β are unique since their activation are controlled by ECM and TGF-β is a major regulator of ECM remodeling in return.

10.1 TGF-β and Extracellular Matrix, a Win-Win Relationship

TGF-β is a prototype of a large family of growth factors that play an essential role in essential biological processes, such as tissue morphogenesis and homeostasis, but also in numerous diseases such as fibrosis and cancer (Tian et al. 2011). Initially

N. Théret (✉) · P. Vignet
University Rennes, Inserm, EHESP, Rennes, France

University Rennes, Inria, CNRS, IRISA, Rennes, France
e-mail: nathalie.theret@univ-rennes1.fr

J. Feret
Inria Paris, Paris, France

DI-ENS, ENS, CNRS, PSL University, Paris, France

A. Hodgkinson · O. Radulescu
LPHI, CNRS, University of Montpellier, Montpellier, France

P. Boutillier
Harvard Medical School, Boston, USA

S. Ricard-Blum (ed.), *Extracellular Matrix Omics*, Biology of Extracellular Matrix 7,
https://doi.org/10.1007/978-3-030-58330-9_10

identified as a promoter of fibroblast growth and transformer of cell phenotype, TGF-β has been rapidly designed as a *bifunctional regulator of cellular growth* (Roberts et al. 1985) depending on the environmental context. Beside its role in cell proliferation, TGF-β is implicated in numerous biological functions including cell differentiation, migration, chemotaxis and ECM production and remodeling.

TGF-β is synthesized as an inactive homodimeric large precursor molecule consisting of a self-inhibiting propeptide, the latency-associated protein (LAP), in addition to the covalently linked active form of TGF-β. Pro-TGF-β is then intracellularly cleaved by furin-type enzymes to generate mature TGF-β, which remains non-covalently associated with LAP as the small latent complex (SLC) and the LAP dimer is covalently bound by a latent TGF-β-binding protein (LTBPs) to form the large latent complex (LLC). LLCs are sequestered in the extracellular matrix (ECM) that forms complex molecular networks (Hynes and Naba, Cold Spring Harb Perspect Biol 4(1):a004903, 2012; Ricard-Blum and Vallet, Matrix Biol 75:170–189, 2019) where LTBP interacts with several components, including fibronectin and fibrilin. The activation process of TGF-β requires dissociation of TGF-β from the ECM-bound LLC and implicates protease—and/or non protease—dependent mechanisms, which differ according to the cell microenvironment (Lodyga and Hinz 2019; Robertson and Rifkin 2016) (see Fig. 10.1). Mechanisms of activation mainly include mechanical interactions (Hinz 2015) involving integrins such as αvβ6 and αvβ8 integrins (Brown and Marshall 2019), chemical interaction involving proteases, such as matrix metalloproteases MMP-2 and MMP-9, and thrombospondin-1 (Murphy-Ullrich and Suto 2018) and physical stress such as heat or reactive oxidative species (Annes et al. 2003). Together, all molecules involved in the dynamic storage and destocking of TGF-β form a protein network in which the role of each one in TGF-β signaling is obviously part of the sum.

When activated, TGF-β binds to specific receptors to induce a variety of signaling pathways depending on the cell and the micro environmental context. TGF-β receptors are trans membrane serine/threonine kinases, that include type I (TGFBR1) and type II (TGFBR2) receptors. The canonical pathway involves a Smad-dependent cascade which induces nuclear signaling to regulate transcription of target genes. Receptor regulated Smads (or R-Smads) are transcription factors initially anchored to the cell membrane by SARA proteins. Following their phosphorylation by TGFBR1, Smads are detached and shuttled to the nucleus, where they activate gene transcription. Numerous Smad-binding partners and transcriptional coactivators and cosrepressors for Smads have been reported as leading to a wide variety of TGF-β-dependent transcriptional signatures (Feng and Derynck 2005). Additionally, cross talk between the TGF-β/Smad pathway and other signaling pathways such as Wnt and Hyppo (Luo 2017; Piersma et al. 2015) complicates the Smad-dependent signature. Otherwise, TGF-β induces non-Smad pathways through binding to TGFBR2, leading to activation of mitogen-activated protein kinase (MAPK), Rho-like GTPase signaling pathways and phosphatidylinositol-3-kinase/AKT pathways (Zhang 2017). Because all of these pathways are also activated by many other extracellular factors and matrix components, the expression of TGF-β-target genes is highly modulated by the cell environment.

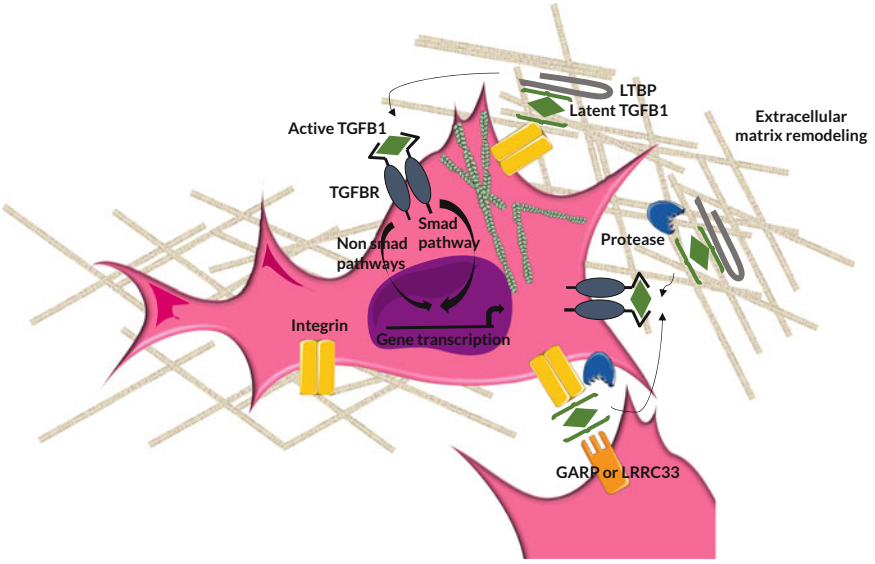

Fig. 10.1 Schematic representation of TGF-β activation and signaling (adapted from (Lodyga and Hinz 2019)). Latent form of TGF-β is sequestered within ECM as a large latent complex associated with LTBP that binds to ECM. Release of the active peptide of TGF-β from this large latent complex involves mechanical and non-mechanical mechanisms depending or not upon protease activities. Integrins or other cell surface receptors such as GARP and LRRC33 bind latent-TGF-β. Strength constraints between cytoskeleton-linked integrins and LTBP-linked ECM induce release of active TGF-β peptide. Protease activities are involved in activation of latent-TGF-β bound to cell surface receptors and contribute to release of TGF-β during extracellular matrix remodeling. Active TGF-β signals through TGF-β receptors (TGFBR) and activation of smad- and non smad-dependent pathways leading to the transcriptional regulation of TGF-β target genes. Most of them are genes coding for ECM compounds and proteases that contribute to ECM remodeling and regulation of TGF-β signal in return

Together the TGF-β-related signal behaves as a system with numerous competitive pathways and regulatory loops, allowing a fine tuning of cell response to various conditions. Understanding how ECM and TGF-β work together to maintain tissue homeostasis, and how alteration of this equilibrium is affected in various pathologies, requires integrative and modeling approaches. Such models aim to predict cell responses to a "TGF-β dependent signal" and, ultimately, identify putative targets suitable for future therapy.

10.2 Modeling Approaches for TGF-β Signaling

Numerous models have been developed to describe the behaviour of the canonical Smad-pathway (Clarke et al. 2006b; Zi et al. 2012). These models, using chemical reaction networks (CRN) and ordinary differential equations (ODEs) focused on

Smad phosphorylation (Clarke et al. 2006b); receptor trafficking (Vilar et al. 2006b); Smad nucleocytoplasmic shuttling (Melke et al. 2006b, Schmierer et al. 2008a); and Smad oligodimerization (Nakabayashi and Sasaki 2009a) that allow an understanding of the dynamics and flexibility of the Smad-dependent pathway.

Importantly, models for receptor trafficking that control the transient or permanent TGF-β-dependent response, enriched the behaviours of TGF-β dependent phenotypes (Vilar et al. 2006b) and integrative models have now coupled receptor trafficking to Smad pathways (Chung et al. 2009; Zi et al. 2011a; Wegner et al. 2012, Nicklas and Saiz 2013; Shankaran and Wiley 2008). The general picture is that the interaction between various Smad channels is a major determinant in shaping the distinct responses to single and multiple ligand stimulation for different cell types (Nicklas and Saiz 2013). The amount of Smad shuttled to the nucleus seems, for most of these models, to depend in a graded, linear manner on the concentration of ligands, while remaining able to be temporally modulated in a transient or oscillatory manner (Cellière et al. 2011). The Smad pathway is also able to encode the speed of variation of the input signal into the shape, transient or permanent (Vilar et al. 2006a), or the amplitude (Sorre et al. 2014) of the output signal, with possible important consequences for morphogen readout and patterning in developmental biology (Sorre et al. 2014). Parametric sensitivity analysis of these models emphasized the importance of various processes for the Smad response (Clarke et al. 2006b). Recently, we have developed an alternative analysis method, based on tropical geometry, that extends steady state calculation to calculation of metastable (long live transient) states (Samal et al. 2016). This method detected two classes of metastable states with antagonistic low and high TGFR1 and TGFR2 values, suggesting that important signal processing leading to flexible response is already performed at the level of the receptor system. These states were given phenotypic interpretations. Using this approach, analysis of proteomic data from NCI-60 cancer cell lines associated non-aggressive and agressive lines to low and high expression TGFBR2 states, respectively (Samal et al. 2016).

Taking advantages of such ODE-based models, we have developed our own models to study the role of the tumor biomarker ADAM12 (Gruel et al. 2009) and the tumor suppressor TIF1γ (Andrieux et al. 2012), thereby demonstrating that small numerical differential models may be useful tools to investigate the role of new regulatory components of the canonical TGF-β signaling pathway. Some of these models are available on the BioModels database (Malik-Sheriff et al. 2019), see Table 10.1. Although most of these models did not integrate the events that take place within the extracellular space and only consider cell surface receptors and free ligands as inputs, a few models include ECM variables providing crude descriptions of the coupling interactions between ECM and TGF-β signaling.

Combinatorial explosion of variables and parameters prevents the use of ODE approaches to integrate all TGF-β dependent pathways including Smad, non Smad-dependent pathways and cross-talk with other pathways. To overcome this limitation, we previously developed a discrete formalism for a large-scale model for TGF-β dependent signaling (Andrieux et al. 2014). In this formalism molecular species and complexes are represented as boolean variables placed in the nodes of a

Table 10.1 Models of TGF-β signaling

Refs.	Pubmed/Biomodels Id	# vars	Method	ECM
Andrieux et al. (2012)	22,461,896	21	ODE	No
Andrieux et al. (2014)	24,618,419	9000	BAN	Yes
Ascolani and Liò (2014)	24,586,338	13	DDE	No
Cellière et al. (2011)	22,051,045/BIOMD0000000600	18	ODE	No
Chung et al. (2009)	19,254,534	17	ODE	No
Clarke et al. (2006a)	17,186,703/BIOMD0000000112	10	ODE	No
Khatibi et al. (2017)	28,407,804	11	DDE	No
Li et al. (2017)	29,322,934	14	ODE	Yes
Lucarelli et al. (2018)	29,248,373	13	ODE	No
Melke et al. (2006a)	17,012,329	17	ODE	No
Musters and van Riel (2004)	17,270,884	5	PL	Yes
Nakabayashi and Sasaki (2009b)	19,358,856	7	ODE	No
Nicklas and Saiz (2013)	23,804,438	38	ODE	No
Proctor and Gartland (2016)	27,379,013/BIOMD0000000612	37	ODE	No
Schmierer et al. (2008b)	18,443,295/BIOMD0000000173	26	ODE	No
Shankaran and Wiley (2008)	18,780,891	∼15	ODE	No
Steinway et al. (2014)	25,189,528	65	BAN	No
Steinway et al. (2015)	28,725,463	69	BAN	No
Strasen et al. (2018)	29,371,237	26	ODE	No
Tortolina et al. (2015)	25,671,297/ MODEL1601250000	460	ODE	No
Venkatraman et al. (2012)	23,009,856/BIOMD0000000447	13	ODE	Yes
Vilar et al. (2006a)	16,446,785/BIOMD0000000101	6	ODE	No
Vilar and Saiz (2011)	22,098,729	12	ODE	No
Vizán et al. (2013)	24,327,760/BIOMD0000000499	26	ODE	No
Wang et al. (2014)	24,901,250	27	ODE	No
Warsinske et al. (2015)	26,384,829	10	ODE	Yes
Wegner et al. (2012)	22,284,904/BIOMD0000000410	53	ODE	No
Zhang et al. (2018)	29,872,541	7	ODE	No
Zi and Klipp (2007)	17,895,977/BIOMD0000000163	16	ODE	No
Zi et al. (2011b)	21,613,981/BIOMD0000000342	21	ODE	No

ODE ordinary differential equations, *DDE* delay differential equations, *BAN* Boolean automaton networks, *PL* hybrid, piecewise linear ODEs

network and connected by "guarded transitions", i.e. monomolecular transformations taking place if logical conditions on regulators and events defining the order of firing are satisfied. Due to the events, both synchronous and asynchronous network dynamics can be simulated. In this case, a trajectory of the network represents a sequence of such transitions. Based on this formalism, we generated a TGF-β network composed of more than 9000 nodes extracted from the Pathway Interaction Database (now available at https://www.pathwaycommons.org), including ECM biochemical interactions, and that allowed us to explore 15934 trajectories involving 145 TGF-β-target genes (Andrieux et al. 2014). A guarded transition-based model,

however, is not appropriate to describe the dynamics of extracellular networks that regulate TGF-β activation. A discrete dynamic modeling approach was also used to model TGF-β-driven epithelial-mesenchymal transition in hepatocellular carcinoma (Steinway et al. 2014, 2015). The authors focused on the dynamics of cross-talks between TGF-β signaling and other signaling pathways but did not integrate extracellular matrix regulation. To take into account the complexity of extracellular matrix dynamics regulating TGF-β signaling, we develop new approaches that are described in the two next parts. The first one uses the rule based formalism Kappa (Danos and Laneve 2004) that allows us to describe the extracellular interaction networks. The second uses mesoscopic PDEs over time, space and structure dimension to integrate multi-scale and multi-physical parameters.

10.3 Kappa, a Formalism Adapted to Model the Biological Component Networks of the Extracellular Matrix

Modeling the ECM can hardly be done by traditional techniques, because it involves the formation of large compounds of proteins. We use the Kappa modeling environment (Boutillier et al. 2018b) to summarize the knowledge that is available in the literature about the molecular interactions surrounding the activation of TGF-β in the ECM.

Complex systems of interaction between molecules are difficult to model for several reasons. Firstly, due to many potential bindings between proteins and numerous potential post-translational changes of conformation, there exists a large (if not infinite, in the case of polymers) number of different kinds of molecular complexes. It is often even impossible to enumerate them. Secondly, the dynamics of these systems is usually triggered by concentration- and time-scale separation; competition against shared-resources; complex causality chains; and nonlinear feedback loops. As a consequence, taking a biochemical approach, which consists in summarising reactions between molecules as generic, local patterns of interactions, seems to be the only viable alternative for modeling these systems and understanding how the dynamics of their populations of molecules may emerge from individual interactions at the microscopic level.

Kappa (Danos and Laneve 2004) is a site-graph rewriting formalism, that is freely inspired by reaction schema encountered in organic chemistry. The main idea is to describe each instance of protein as a node in a graph. Each kind of protein has some interaction sites which can bind pair-wise. The interactions between molecules are formalised by the means of rewrite rules. The rules either stand for interactions that are detailed in the literature or for some fictitious interactions that exist to make assumptions about the information that is missing or to roughly simplify some parts that we do not want to detail too much. The use of rules eases frequent updates of the models, which enables the modeler to test numerous scenarios or to modify the environment of the model. A set of rules can be interpreted as a dynamical system

which de-scribes the evolution of a soup of molecules. There are several choices: when the number of different kinds of molecular complexes is not too great, the set of rules may be translated into ODEs (Camporesi et al. 2017). Each set of rules also induces a continuous time Markov chain the execution traces of which can be sampled by simulation (Danos et al. 2007b). Thanks to the use of specific data-structures (Danos et al. 2007b; Boutillier et al. 2017), the computation cost of such simulation does not depend on the number of kinds of molecular complexes, which may even be infinite.

Kappa ecosystem (Boutillier et al. 2018b) offers several tools to assist the modeler during her task. Static analysis (Danos et al. 2008; Feret and Lý 2018; Boutillier et al. 2018a) may be used to curate models. Canonical and secondary pathways may be extracted thanks to causality analysis (Danos et al. 2007a, 2012). Formal methods can also be used to identify the key elements in information propagation. The result is a model reduction which never loses any information about the quantities that are observed in the model (Feret et al. 2009; Danos et al. 2010; Camporesi et al. 2013).

We wrote a model for the influence of the ECM on TGF-β signal, including numerous extracellular interactions that are documented in the literature, and some fictitious rules to stub gene activity and its interaction with TGF-β. The model is made of around 300 interaction rules which are freely available on the web (Théret et al. 2020b). Each rule is parameterised by a kinetic rate. Some of the rates are deduced from precise information about the concentration of proteins at stationary distribution and their half-time periods. Some others are chosen approximately, in order to best model what is known about the time scales of each interaction. Our model comprises around 30 kinds of proteins. The potential bonds between these proteins are summarised in Fig. 10.2, which provides a convenient snapshot of the model, while not detailing every rule. Selected portions of the models are depicted and explained intuitively in (Théret et al. 2020a).

Our goal, when designing this model, is three-fold. Firstly, the interaction rules are written to organize what is known in the literature and to let us make some assumptions about what is not known. It is a way to make knowledge about the models and its different variants navigable. Secondly, the different semantics of Kappa allow the execution of the rules. This makes knowledge executable. The last, longer term objective, is to understand how the macroscopic behaviour may emerge from the interactions between the individual instances of proteins. In the end, the semantics of Kappa make it possible to better approach the dynamics of the multiple molecular interactions that contribute to the activation of TGF-β. Using parameters specific to the pathological context, this model can allow us to identify the key events that regulate this activation and consequently potential therapeutic targets.

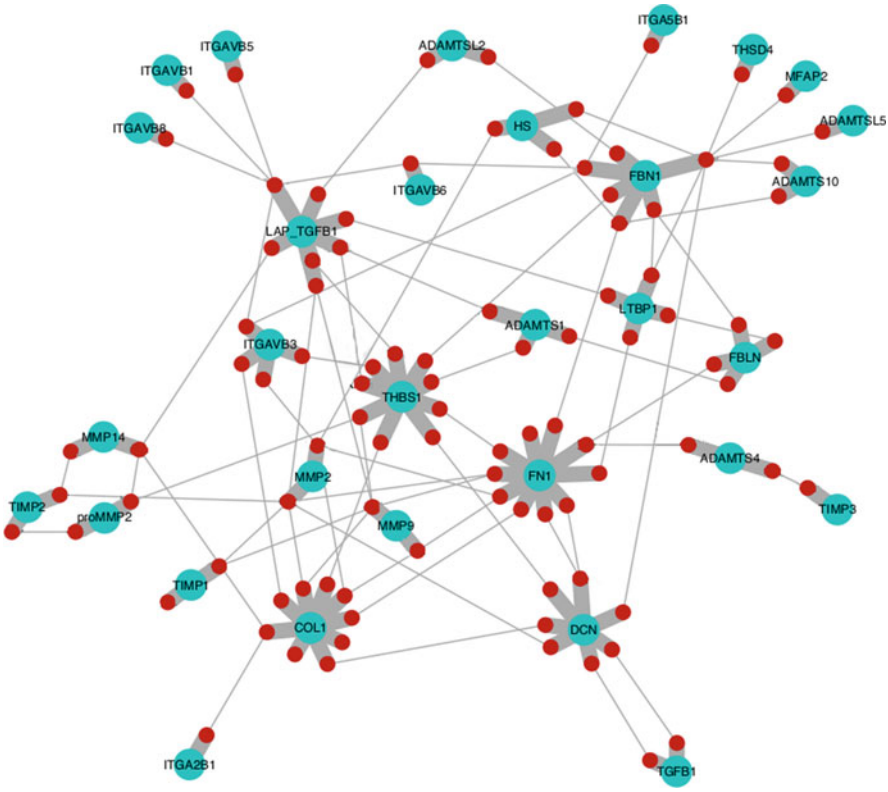

Fig. 10.2 Contact map of the Kappa model for TGF-β activation: Projection of model describing the molecule interaction networks. Proteins and glycosaminoglycans are represented by turquoise nodes, binding sites are represented by red nodes (if the sequence involved is not known, the site is designated by a letter, x, y, z etc.). The lines between the sites illustrate potential links involving these binding sites. *ITGA-x-B-y* Integrin alpha-x Beta-y, *LAP-TGFB1* latent TGFB1, *THBS1* Thrombospondin, *HS* Heparan Sulfate, *FBN1* Fibrillin 1, *FN1* Fibronectin, *FBLN* Fibulin, *THSD4* ADAMTSL-6, *MFAP2* Microfibril Associated Protein 2, *LTBP1* Latent Transforming Growth Factor Beta Binding Protein 1, *MMP* Matrix Metalloprotease, *TIMP* Tissue inhibitor of MMP, *COL1* Type 1 collagen, *DCN* Decorin

10.4 Mesoscale and Multi-Scale Tissue Models Integrating TGF-β Signaling and its Interaction with the ECM

The coupling of TGF-β with the ECM is a multi-scale and multi-physical problem. It involves chemical kinetics of intracellular signaling and of ECM bio-chemical processes, but also more complex physico-chemical processes such as polymerisation and viscoelastic dynamics of collagen and fibrin fibres of the ECM, as well as population dynamics of various cell types.

A simple, but rather limited, solution to modelling such a complex situation is model merging. Models representing several levels of organisation can be merged together to cope with the coupling between scales. Merging, however, is not straightforward even when models of different levels are of the same type (PDEs, ODEs or Markov processes). For instance, it is relatively easy to couple ODE models of intracellular pathways with models of the same type of the extracellular matrix, by using the standard technique of compartments (Venkatraman et al. 2012; Li et al. 2017). It is much more difficult to couple single cell dynamics with the population dynamics because the variables of these models cannot be simply juxtaposed with different spatial locations.

Single scale, ODE population dynamics models were used to study the role of TGF-β in immunotherapy and wound healing (Waugh and Sherratt 2006; Wilson and Levy 2012; Hu et al. 2019; Arciero et al. 2004; Bianchi et al. 2015). In these models the ECM variables are implicit in the cell-cell interactions but do not follow dynamical equations. These simplifications are extreme and could lead to inaccurate conclusions for processes depending critically on the dynamical structuring of the ECM, such as in wound repair for instance.

Hybrid approaches combining discrete cell positions with continuous description of collagen matrix and other ECM components have been used to study the role of the TGF-β/ECM interactions in wound repair (Dallon et al. 2001; Wang et al. 2019; Cumming et al. 2009). In these models, fibroblasts and immune cell motility and proliferation are affected by TGF-β, and cells interact one with another or with the collagen chemically or by direct, physical contact. A similar model was used to couple TGF-β signaling with the micro-environment in a preliminary study of tumor-stroma interactions (Morshed et al. 2018). Furthermore, there is a need for models including mechanical stresses known to be generated in ECM by cell traction and vessel growth.

Agent-based modeling enabling cells to have individual behaviours including division and motility has been used for models of epidermis in the context of wound healing (Wang et al. 2009; Stern et al. 2012; Sun et al. 2009; Adra et al. 2010). However, this solution is computationally expensive and has limitations in terms of biochemical details that can be used to model or parameterise the cell behaviour; its interaction with the micro- environment; or the number of cells and, moreover, is entirely based on numerical simulation.

Continuous modelling using partial differential equations (PDEs) can be justified by coarse graining (homogenisation) when the spatial scale of interest is much larger than the cell size and the typical dimensions of fibers. This modelling can take into account spatially inhomogeneous densities of various cell types and ligands, as well as collagen fibers and other ECM constituents. Directed, un-directed, and chemically mediated cell mobility are taken into account in Keller-Segel PDE systems or in similar systems used in oncology and likewise in the context of chondrogenesis or fibrosis (Bitsouni et al. 2017; Kim and Othmer 2013; Friedman and Hao 2017; Chen et al. 2018). Although quite flexible, easy to simulate in 2D and 3D, and sometimes leading to analytic results, these models consider intracellular dynamics as instantaneous and do not handle mechanical interaction.

While informative, all these studies integrated the extracellular world as a very simplified input that did not capture the extracellular dynamics of TGF-β life in its full complexity. For instance, the kinetics of production and degradation of TGF-β including its latent form bound to ECM and its active soluble form was considered in the study of TGF-β interaction with chondrocyte and mesenchymal stem cells (Chen et al. 2018). The underlying biochemical networks that regulate such dynamics, however, remained unexplored. Importantly, the ECM is a complex molecular network combining proteins, proteolglycans and glycoaminoglycans that is constantly remodeled through modification of components' synthesis and their degradation by proteases. This specific tissue microenvironment is disrupted in pathological processes and directly affects the TGF-β-mediated signal by modifying its storage and release.

Because of slow intracellular dynamics, differences occur not only between cells of different types and genotypes but also between clonal cells. Non-genetic sources of variability are particularly important in the development of resistance to drug treatment of tumors or, more generally, in the adaptation of cell populations to stresses. We have recently introduced a new approach to cope with this variability while remaining in a continuous frame-work that is convenient for simulation and analysis. This approach uses mesoscopic PDEs over temporal, spatial, and structural dimensions (Hodgkinson et al. 2018, 2019). Mesoscale models are obtained from the Liouville continuity equation. For illustration, let us consider that there are n types of cells. In this model cells are distinguished by two types of variables, a discrete one representing the type $i \in \{1,\ldots, n\}$ and a continuous one $y \in \mathrm{R}^m$ representing the internal state (the vector of concentrations of m biochemical species). Then $c = (c_{1,\ldots}, \ldots, c_n)$ represents a vector of cell distributions satisfying the equation

$$\frac{\partial c(x,y,t)}{\partial t} = -\nabla_x F_x(c,x,y,t) - \nabla_y F_y(c,x,y,t) + S(c,x,y,t), \tag{10.1}$$

where x is the spatial position; y is the cell's internal state (structure variable); F_x is the spatial flux; F_y is the structural flux; and S is the source term. If the cell's internal state follows ODEs $dy/dt = \Phi(y)$, then the structural flux is advective $F_y = c\Phi$. The spatial flux function contains terms related to cell motility; undirected (diffusion) or directed (chemotaxis, haptotaxis). The source term integrates cell proliferation, death and transformation from one cell type to another.

The mesoscale formalism can also integrate mechanical stresses by addition of constitutive equations coupled to cell densities and biochemistry. Biochemistry has been coupled to stress by using microscopic Brownian dynamics of ECM remodeling (Malandrino et al. 2019) or Kramer's formula for chemical reaction rates with a free energy dependent on pulling force of actin filaments (Cockerill et al. 2015). Another option would be to incorporate slower structural changes in ECM fibre density or mechanical stress into a temporally, spatially, and structurally distributed ECM population, possessing its own dynamical PDE. Existing models are oversimplified and incomplete and there is strong need for a general continuous

mechanical theory relating structure, deformation, cell population dynamics and biochemistry in the ECM.

10.5 Conclusion

Modelling TGF-β signaling integrating extracellular activation processes and intracellular pathways raises several open challenges.

An important challenge is to simulate processes that occur at highly separated time-scales as well as the spatial organization of ECM interactions. Deriving a model that would scale up to the size of these systems, and to the long periods which have to be simulated to observe the phenomena of interest, requires precise abstractions of populations of cells. These abstractions consist in changing the grain of description of the behaviours of these cells and can be formalised in many ways thanks to mathematical tools such as closure operators, Galois connections, ideals, changes of variables (Cousot and Cousot 1977). Yet, it is important to keep the information that mainly drives the dynamics of the whole system. Indeed the diversity of behaviours—even among identical cells—may constitute an important part of the signal that is computed by the interactions between the proteins and that which controls the behaviour of the system at the macroscopic level. In our opinion, neither non-deterministic approximations nor homogeneous abstractions would likely offer satisfying solutions to solve this issue. Non-deterministic systems are systems in which, at each moment of the execution, the immediate future has to be chosen among the elements of a set of potential behaviours, but where the choice of element is not specified, as opposed to deterministic systems for which there always exists a unique potential future; reactive systems for which the choice of the next event is triggered by an interaction with an external environment; and stochastic systems for which a distribution of probabilities defines the likelihood of each potential immediate future behaviour. Non-deterministic abstractions flatly over-approximate all the potential behaviours of each cell without providing any information about their probability distribution. They can hardly be used in composite models, since many potential behaviours would have to be considered for each cell, and, thus, it would require the consideration of too many cases across the population of cells. Homogeneous abstractions consist in abstracting away the diversity of behaviours of the population of cells, and to replace them with several copies of a unique system, which behaves in the same way. Abstractions with too great a homogeneity would keep only most probable behaviours for the cells, whilst ignoring others. As a consequence, such a system would not faithfully model the diversity of behaviours among the population of cells, which may be a key ingredient in explaining the overall behaviour of the composite model. Stochastic models keep enough information about the distribution of the different behaviours within a population of cell. Moreover, they can be easily integrated within a multi-scale model. Nevertheless, they come with an important combinatorial cost and can hardly be simplified.

Mesoscale PDE models represent a promising direction for modeling the ECM. They are well suited for implementing middle-out modelling strategies, in which several levels of organisation are treated together, but with just enough details to render the essence of the overall organisation. Mesoscale models can be built from scratch, but can also include already available models, or parts of them, after model reduction. The model reduction procedure, based on time-scale separation and singular perturbations, averaging or homogeneization, is not yet well established and is the subject of active research (Radulescu et al. 2012). Although deterministic, these models can render the stochastic behaviour of "microscopic" variables, responsible, among other factors, for the heterogeneity of cellular decision. In this approach, stochastic simulations are replaced by calculation of probability distributions of microscopic variables, that follow PDEs in mesoscopic descriptions. In spite of recent progress, the important question of the coupling between mechanical stress, biochemistry and cell behaviour has been treated only superficially. The ECM is a complex medium, including insoluble fibers-forming molecules (collagen, fibronectin, elastin) and soluble molecules such as proteoglycans and glycoaminoglycans which constitute a hydrated gel, of relatively high viscosity, and confer elastic properties to the ECM and glycoproteins characterized by their adhesive properties (laminin, fibronectin, tenascin, etc.). The viscoelastic properties of this medium, in particular its capacity to transmit mechanical cues that further influence cellular processes, such as differentiation, proliferation, survival and migration, could be instrumental for tissue remodelling. Constitutive equations, eventually inspired from the physics of polymers and gels (Larson 2013; Prost et al. 2015), should be able to provide the continuous mechanics theoretical framework needed for relating stress, strain, signaling and cell decisions in tissue models.

References

Adra S, Sun T, MacNeil S, Holcombe M, Small-Wood, R. (2010) Development of a three dimensional multiscale computational model of the human epidermis. PLoS One 5(1):e8511

Andrieux G, Fattet L, Le Borgne M, Rimokh R, Théret N (2012) Dynamic regulation of tgf-b signaling by tif1γ: a computational approach. PLoS One 7(3):e33761

Andrieux G, Le Borgne M, Théret N (2014) An integrative modeling framework reveals plasticity of tgf-β signaling. BMC Syst Biol 8(1):30

Annes JP, Munger JS, Rifkin DB (2003) Making sense of latent tgfβ activation. J Cell Sci 116 (2):217–224

Arciero J, Jackson T, Kirschner D (2004) A mathematical model of tumor-immune evasion and sirna treatment. Discrete and Continuous Dynamical Systems Series B 4(1):39–58

Ascolani G, Liò P (2014) Modeling tgf-β in early stages of cancer tissue dynamics. PLoS One 9(2): e88533

Bianchi A, Painter KJ, Sherratt JA (2015) A mathematical model for lymphangiogenesis in normal and diabetic wounds. J Theor Biol 383:61–86

Bitsouni V, Chaplain MA, Eftimie R (2017) Mathematical modelling of cancer invasion: the multiple roles of tgf-β path- way on tumour proliferation and cell adhesion. Math Models Methods Appl Sci 27(10):1929–1962

Boutillier P, Ehrhard T, Krivine J (2017) Incremental update for graph rewriting. In: Yang H (ed) Programming languages and systems – 26th European Symposium on Programming, ESOP 2017, held as part of the European joint conferences on theory and practice of software, ETAPS 2017, Uppsala, Sweden, April 22–29, 2017, Proceedings, volume 10201 of Lecture notes in computer science. Springer, Berlin, pp 201–228.

Boutillier P, Camporesi F, Coquet J, Feret J, Lý KQ, Théret N, Vignet P (2018a) Kasa: a static analyzer for kappa. In: Ceska M, Safránek D (eds) Computational methods in systems biology – 16th International Conference, CMSB 2018, Brno, Czech Republic, September 12–14, 2018, Proceedings, volume 11095 of Lecture notes in computer science. Springer, Berlin, pp 285–291

Boutillier P, Maasha M, Li X, Medina-Abarca HF, Krivine J, Feret J, Cristescu I, Forbes AG, Fontana W (2018b) The kappa platform for rule-based modeling. Bioinformatics 34(13):i583–i592

Brown NF, Marshall JF (2019) Integrin-mediated tgfβ activation modulates the tumour microenvironment. Cancers 11(9):1221

Camporesi F, Feret J, Hayman JM (2013) Context-sensitive flow analyses: a hierarchy of model reductions. In: Gupta A, Henzinger TA (eds) Computational methods in systems biology – 11th International Conference, CMSB 2013, Klosterneuburg, Austria, September 22–24, 2013. Proceedings, volume 8130 of Lecture notes in computer science. Springer, Berlin, pp 220–233

Camporesi F, Feret J, Lý KQ (2017) Kade: a tool to compile kappa rules into (reduced) ODE models. In: Feret J, Koeppl H (eds) Computational methods in systems biology – 15th International Conference, CMSB 2017, Darmstadt, Germany, September 27–29, 2017, Proceedings, volume 10545 of Lecture notes in computer science. Springer, Berlin, pp 291–299

Cellière G, Fengos G, Hervé M, Iber D (2011) The plasticity of tgf-β signaling. BMC Syst Biol 5 (1):184

Chen MJ, Whiteley JP, Please CP, Schwab A, Ehlicke F, Waters SL, Byrne HM (2018) Inducing chondrogenesis in msc/chondrocyte co-cultures using exogenous tgf-β: a mathematical model. J Theor Biol 439:1–13

Chung S-W, Miles FL, Sikes RA, Cooper CR, Farach-Carson MC, Ogunnaike BA (2009) Quantitative modeling and analysis of the transforming growth factor beta signaling pathway. Biophys J 96(5):1733–1750

Clarke DC, Betterton M, Liu X (2006a) Systems theory of smad signaling. IEE Proc Syst Biol 153 (6):412–424

Clarke DC, Betterton MD, Liu X (2006b) Systems theory of Smad signaling. Syst Biol (Stevenage) 153(6):412–424

Cockerill M, Rigozzi MK, Terentjev EM (2015) Mechanosensitivity of the 2nd kind: Tgf-β mechanism of cell sensing the substrate stiffness. PLoS One 10(10):e0139959

Cousot P, Cousot R (1977) Abstract interpretation: a unified lattice model for static analysis of programs by construction or approximation of fixpoints. In: Graham RM, Harrison MA, Sethi R (eds) Conference record of the fourth ACM symposium on principles of programming languages, Los Angeles, California, USA, January 1977. ACM, California, pp 238–252

Cumming BD, McElwain D, Upton Z (2009) A mathematical model of wound healing and subsequent scarring. J R Soc Interface 7(42):19–34

Dallon JC, Sherratt JA, Maini PK (2001) Modeling the effects of transforming growth factor-beta on extracellular matrix alignment in dermal wound repair. Wound Repair Regen 9(4):278–286

Danos V, Laneve C (2004) Formal molecular biology. Theor Comput Sci 325(1):69–110

Danos V, Feret J, Fontana W, Harmer R, Krivine J (2007a) Rule-based modelling of cellular signaling. In: Caires L, Vasconcelos VT (eds) CONCUR 2007 – Concurrency theory, 18th International Conference, CONCUR 2007, Lisbon, Portugal, September 3–8, 2007, Proceedings, volume 4703 of Lecture notes in computer science. Springer, Berlin, pp 17–41

Danos V, Feret J, Fontana W, Krivine J (2007b) Scalable simulation of cellular signaling networks. In: Shao Z (ed) Programming languages and systems, 5th Asian Symposium, APLAS 2007, Singapore, November 29–December 1, 2007, Proceedings, volume 4807 of Lecture notes in computer science. Springer, Berlin, pp 139–157

Danos V, Feret J, Fontana W, Krivine J (2008) Abstract interpretation of cellular signaling networks. In: Logozzo F, Peled DA, Zuck LD (eds) Verification, model checking, and abstract interpretation, 9th International Conference, VMCAI 2008, San Francisco, USA, January 7–9, 2008, Proceedings, volume 4905 of Lecture notes in computer science. Springer, Berlin, pp 83–97

Danos V, Feret J, Fontana W, Harmer R, Krivine J (2010) Abstracting the differential semantics of rule-based models: exact and automated model reduction. In: Proceedings of the 25th Annual IEEE Symposium on Logic in Computer Science, LICS 2010, 11–14 July 2010, Edinburgh, United Kingdom. IEEE Computer Society, pp 362–381

Danos V, Feret J, Fontana W, Harmer R, Hayman JM, Krivine J, Thompson-Walsh CD, Winskel G (2012) Graphs, rewriting and pathway reconstruction for rule-based models. In: D'Souza D, Kavitha T, Radhakrishnan J (eds) IARCS annual Conference on Foundations of Software Technology and Theoretical Computer Science, FSTTCS 2012, December 15–17, 2012, Hyderabad, India, volume 18 of LIPIcs. Schloss Dagstuhl – Leibniz-Zentrum fuer Informatik, pp 276–288

Feng X-H, Derynck R (2005) Specificity and versatility in tgf-beta signaling through Smads. Annu Rev Cell Dev Biol 21:659–693

Feret J, Lý KQ (2018) Reachability analysis via orthogonal sets of patterns. Electr Notes Theor Comput Sci 335:27–48

Feret J, Danos V, Krivine J, Harmer R, Fontana W (2009) Internal coarse-graining of molecular systems. Proc Natl Acad Sci U S A 106:6453–6458

Friedman A, Hao W (2017) Mathematical modeling of liver fibrosis. Math Biosci Eng 14 (1):143–164

Gruel J, Leborgne M, LeMeur N, Théret N (2009) In silico investigation of ADAM12 effect on TGF-beta receptors trafficking. BMC Res Notes 2:193

Hinz B (2015) The extracellular matrix and transforming growth factor-β1: tale of a strained relationship. Matrix Biol 47:54–65

Hodgkinson A, Uzé G, Radulescu O, Trucu D (2018) Signal propagation in sensing and reciprocating cellular systems with spatial and structural heterogeneity. Bull Math Biol 80 (7):1900–1936

Hodgkinson A, Le Cam L, Trucu D, Radulescu O (2019) Spatiogenetic and phenotypic modelling elucidates resistance and re-sensitisation to treatment in heterogeneous melanoma. J Theor Biol 466:84–105

Hu X, Ke G, Jang SR-J (2019) Modeling pancreatic cancer dynamics with immunotherapy. Bull Math Biol 81(6):1885–1915

Hynes RO, Naba A (2012) Overview of the matrisome—an inventory of extracellular matrix constituents and functions. Cold Spring Harb Perspect Biol 4(1):a004903

Khatibi S, Zhu H-J, Wagner J, Tan CW, Manton JH, Burgess AW (2017) Mathematical model of tgf-β signaling: feedback coupling is consistent with signal switching. BMC Syst Biol 11(1):48

Kim Y, Othmer HG (2013) A hybrid model of tumor–stromal interactions in breast cancer. Bull Math Biol 75(8):1304–1350

Larson RG (2013) Constitutive equations for polymer melts and solutions: Butterworths series in chemical engineering. Butterworth-Heinemann, Oxford

Li H, Venkatraman L, Narmada BC, White JK, Yu H, Tucker-Kellogg L (2017) Computational analysis reveals the coupling between bistability and the sign of a feedback loop in a TGF-β1 activation model. BMC Syst Biol 11(Suppl 7):136

Lodyga M, Hinz B (2019) Tgf-β1–a truly transforming growth factor in fibrosis and immunity. In: Seminars in cell & developmental biology, vol 101. Elsevier, Amsterdam, pp 123–139

Lucarelli P, Schilling M, Kreutz C, Vlasov A, Boehm ME, Iwamoto N, Steiert B, Lattermann S, Wäsch M, Stepath M, Matter MS, Heikenwälder M, Hoffmann K, Deharde D, Damm G, Seehofer D, Muciek M, Gretz N, Lehmann WD, Timmer J, Klingmüller U (2018) Resolving the combinatorial complexity of smad protein complex formation and its link to gene expression. Cell Syst 6(1):75–89.e11

Luo K (2017) Signaling cross talk between tgf-β/smad and other signaling pathways. Cold Spring Harb Perspect Biol 9(1):a022137

Malandrino A, Trepat X, Kamm RD, Mak M (2019) Dynamic filopodial forces induce accumulation, damage, and plastic remodeling of 3d extracellular matrices. PLoS Comput Biol 15(4): e1006684

Malik-Sheriff RS, Glont M, Nguyen TVN, Tiwari K, Roberts MG, Xavier A, Vu MT, Men J, Maire M, Kananathan S, Fairbanks EL, Meyer JP, Arankalle C, Varusai TM, Knight-Schrijver V, Li L, Dueñas-Roca C, Dass G, Keating SM, Park YM, Buso N, Rodriguez N, Hucka M, Hermjakob H (2019) BioModels–15 years of sharing computational models in life science. Nucleic Acids Res 48(D1):D407–D415. https://doi.org/10.1093/nar/gkz1055

Melke P, Jönsson H, Pardali E, ten Dijke P, Peterson, C. (2006a) A rate equation approach to elucidate the kinetics and robustness of the tgf-β pathway. Biophys J 91(12):4368–4380

Melke P, Jönsson H, Pardali E, ten Dijke P, Peterson C (2006b) A rate equation approach to elucidate the kinetics and robustness of the TGF-beta pathway. Biophys J 91(12):4368–4380

Morshed A, Dutta P, Dillon RH (2018) Mathematical modeling and numerical simulation of the tgf-β/smad signaling pathway in tumor microenvironments. Appl Numer Math 133:41–51

Murphy-Ullrich JE, Suto MJ (2018) Thrombospondin-1 regulation of latent tgf-β activation: a therapeutic target for fibrotic disease. Matrix Biol 68:28–43

Musters MWJM, van Riel NA (2004) Analysis of the transforming growth factor-beta/sub 1/pathway and extracellular matrix formation as a hybrid system. Conf Proc IEEE Eng Med Biol Soc 4:2901–2904

Nakabayashi J, Sasaki A (2009a) A mathematical model of the stoichiometric control of Smad complex formation in TGF-beta signal transduction pathway. J Theor Biol 259(2):389–403

Nakabayashi J, Sasaki A (2009b) A mathematical model of the stoichiometric control of smad complex formation in tgf-β signal transduction pathway. J Theor Biol 259(2):389–403

Nicklas D, Saiz L (2013) Computational modelling of smad-mediated negative feedback and crosstalk in the tgf-β superfamily network. J R Soc Interface 10(86):20130363

Piersma B, Bank RA, Boersema M (2015) Signaling in fibrosis: TGF-β, WNT, and YAP/TAZ converge. Front Med (Lausanne) 2:59

Proctor CJ, Gartland A (2016) Simulated interventions to ameliorate age-related bone loss indicate the importance of timing. Front Endocrinol 7:61

Prost J, Jülicher F, Joanny J-F (2015) Active gel physics. Nat Phys 11(2):111–117

Radulescu O, Gorban AN, Zinovyev A, Noel V (2012) Reduction of dynamical biochemical reactions networks in computational biology. Front Genet 3:131

Ricard-Blum S, Vallet SD (2019) Fragments generated upon extracellular matrix remodeling: biological regulators and potential drugs. Matrix Biol 75:170–189

Roberts AB, Anzano MA, Wakefield LM, Roche NS, Stern DF, Sporn MB (1985) Type beta transforming growth factor: a bifunctional regulator of cellular growth. Proc Natl Acad Sci 82 (1):119–123

Robertson IB, Rifkin DB (2016) Regulation of the bioavailability of tgf-β and tgf-β-related proteins. Cold Spring Harb Perspect Biol 8(6):a021907

Samal SS, Naldi A, Grigoriev D, Weber A, Théret N, Radulescu O (2016) Geometric analysis of pathways dynamics: application to versatility of tgf-β receptors. Biosystems 149:3–14

Schmierer B, Tournier AL, Bates PA, Hill CS (2008a) Mathematical modeling identifies Smad nucleocytoplasmic shuttling as a dynamic signal-interpreting system. Proc Natl Acad Sci U S A 105(18):6608–6613

Schmierer B, Tournier AL, Bates PA, Hill S (2008b) Mathematical modeling identifies smad nucleocytoplasmic shuttling as a dynamic signal-interpreting system. Proc Natl Acad Sci 105 (18):6608–6613

Shankaran H, Wiley HS (2008) Smad signaling dynamics: insights from a parsimonious model. Sci Signal 1(36):pe41

Sorre B, Warmflash A, Brivanlou AH, Siggia ED (2014) Encoding of temporal signals by the tgf-β pathway and implications for embryonic patterning. Dev Cell 30(3):334–342

Steinway SN, Zañudo JGT, Ding W, Rountree CB, Feith DJ, Loughran TP, Albert R (2014) Network modeling of TGFβ signaling in hepatocellular carcinoma epithelial-to-mesenchymal transition reveals joint sonic hedgehog and Wnt pathway activation. Cancer Res 74 (21):5963–5977

Steinway SN, Zañudo JGT, Michel PJ, Feith J, Loughran TP, Albert R (2015) Combinatorial interventions inhibit tgf-β-driven epithelial-to-mesenchymal transition and support hybrid cellular phenotypes. NPJ Syst Biol Appl 1:15014

Stern JR, Christley S, Zaborina O, Alverdy JC, An G (2012) Integration of tgf-β-and egfr-based signaling pathways using an agent-based model of epithelial restitution. Wound Repair Regen 20(6):862–863

Strasen J, Sarma U, Jentsch M, Bohn S, Sheng C, Horbelt D, Knaus P, Legewie S, Loewer A (2018) Cell-specific responses to the cytokine TGFβ are determined by variability in protein levels. Mol Syst Biol 14(1):e7733

Sun T, Adra S, Smallwood R, Holcombe M, Mac-Neil, S. (2009) Exploring hypotheses of the actions of tgf-β1 in epidermal wound healing using a 3d computational multiscale model of the human epidermis. PLoS One 4(12):e8515

Théret N, Feret J, Boutillier P, Vignet P (2020a) Selected parts of the kappa model for the TGF-beta extracellular matrix. https://github.com/feret/TGF-Kappa/tree/master/doc/ selected_parts_of_the_model.pdf

Théret N, Feret J, Cocquet J, Vignet P, Boutillier P, Camporesi F (2020b). https://github.com/feret/TGF-Kappa/tree/master/model

Tian M, Neil JR, Schiemann WP (2011) Transforming growth factor-β and the hallmarks of cancer. Cell Signal 23(6):951–962

Tortolina L, Duffy DJ, Maffei M, Castagnino N, Carmody AM, Kolch W, Kholodenko BN, De Ambrosi C, Barla A, Biganzoli EM et al (2015) Advances in dynamic modeling of colorectal cancer signaling-network regions, a path toward targeted therapies. Oncotarget 6(7):5041

Venkatraman L, Chia S-M, Narmada BC, White JK, Bhowmick SS, Dewey CF Jr, So PT, Tucker-Kellogg L, Yu H (2012) Plasmin triggers a switch-like decrease in thrombospondin-dependent activation of tgf-β1. Biophys J 103(5):1060–1068

Vilar JM, Saiz L (2011) Trafficking coordinate description of intracellular transport control of signaling networks. Biophys J 101(10):2315–2323

Vilar JM, Jansen R, Sander C (2006a) Signal processing in the tgf-β superfamily ligand-receptor network. PLoS Comp Biol 2(1):e3

Vilar JMG, Jansen R, Sander C (2006b) Signal processing in the TGF-beta superfamily ligand-receptor network. PLoS Comput Biol 2(1):e3

Vizán P, Miller DSJ, Gori I, Das D, Schmierer B, Hill CS (2013) Controlling long-term signaling: receptor dynamics determine attenuation and refractory behavior of the TGF-β pathway. Sci Signal 6(305):ra106

Wang Z, Birch CM, Sagotsky J, Deisboeck TS (2009) Cross-scale, cross-pathway evaluation using an agent-based non- small cell lung cancer model. Bioinformatics 25(18):2389–2396

Wang J, Tucker-Kellogg L, Ng IC, Jia R, Thiagarajan P, White JK, Yu H (2014) The self-limiting dynamics of tgf-β signaling in silico and in vitro, with negative feedback through ppm1a up-regulation. PLoS Comput Biol 10(6):e1003573

Wang Y, Guerrero-Juarez CF, Qiu Y, Du H, Chen W, Figueroa S, Plikus MV, Nie Q (2019) A multiscale hybrid mathematical model of epidermal-dermal interactions during skin wound healing. Exp Dermatol 28(4):493–502

Warsinske HC, Ashley SL, Linderman JJ, Moore BB, Kirschner DE (2015) Identifying mechanisms of homeostatic signaling in fibroblast differentiation. Bull Math Biol 77(8):1556–1582

Waugh HV, Sherratt JA (2006) Macrophage dynamics in diabetic wound dealing. Bull Math Biol 68(1):197–207

Wegner K, Bachmann A, Schad J-U, Lucarelli P, Sahle S, Nickel P, Meyer C, Klingmüller U, Dooley S, Kummer U (2012) Dynamics and feedback loops in the transforming growth factor β signaling pathway. Biophys Chem 162:22–34

Wilson S, Levy D (2012) A mathematical model of the enhancement of tumor vaccine efficacy by immunotherapy. Bull Math Biol 74(7):1485–1500

Zhang YE (2017) Non-Smad signaling pathways of the TGF-β family. Cold Spring Harb Perspect Biol 9(2)

Zhang J, Tian X-J, Chen Y-J, Wang W, Watkins S, Xing J (2018) Pathway crosstalk enables cells to interpret tgf-β duration. NPJ Syst Biol Appl 4(1):18

Zi Z, Klipp E (2007) Constraint-based modeling and kinetic analysis of the smad dependent tgf-β signaling pathway. PLoS One 2(9):e936

Zi Z, Feng Z, Chapnick DA, Dahl M, Deng D, Klipp E, Moustakas A, Liu X (2011a) Quantitative analysis of transient and sustained transforming growth factor-β signaling dynamics. Mol Syst Biol 7:492

Zi Z, Feng Z, Chapnick DA, Dahl M, Deng D, Klipp E, Moustakas A, Liu X (2011b) Quantitative analysis of transient and sustained transforming growth factor-β signaling dynamics. Mol Syst Biol 7(1):492

Zi Z, Chapnick DA, Liu X (2012) Dynamics of TGF-β/Smad signaling. FEBS Lett 586 (14):1921–1928