

***In-silico* definition of the *Drosophila melanogaster* matrisome**

Martin N. Davis¹, Sally Horne-Badovinac², Alexandra Naba^{1*}

¹ Department of Physiology and Biophysics, University of Illinois at Chicago, 835 S. Wolcott Avenue, Chicago, IL 60612, U.S.A

² Department of Molecular Genetics and Cell Biology, The University of Chicago, 920 East 58th Street, Chicago, IL 60637, U.S.A

*** Correspondence:**

Dr. Alexandra Naba

Department of Physiology and Biophysics

University of Illinois at Chicago

835 S. Wolcott Avenue

Chicago, IL 60612

U.S.A

Email: anaba@uic.edu

Phone: +1 (312) 355-5417

ABSTRACT

The extracellular matrix (ECM) is an assembly of hundreds of proteins that structurally supports the cells it surrounds and biochemically regulates their functions. *Drosophila* has emerged as a powerful model organism to study fundamental mechanisms underlying ECM protein secretion, ECM assembly, and ECM roles in pathophysiological processes. However, as of today, we do not possess a well-defined list of the components forming the ECM of this organism. We previously reported the development of computational pipelines to define the matrisome - the ensemble of genes encoding ECM and ECM-associated proteins - of humans, mice, zebrafish and *C. elegans*. Using a similar approach, we report here that the *Drosophila* matrisome is composed of 641 genes. We further classify these genes into different structural and functional categories, including an expanded way to classify genes encoding proteins forming apical ECMs. We illustrate how having a comprehensive list of *Drosophila* matrisome proteins can be used to annotate large proteomic datasets and identify unsuspected roles for the ECM in pathophysiological processes. Last, to aid the dissemination and usage of the proposed definition and categorization of the *Drosophila* matrisome by the scientific community, our list has been made available through three public portals: [The Matrisome Project](#), [FlyBase](#), and [GLAD](#).

KEYWORDS

Extracellular Matrix; Basement Membrane; Apical ECM; Cuticle; Eggshell; Glue; Protein Domains

1. Introduction

The extracellular matrix (ECM) is an assembly of hundreds of proteins that structurally supports and biochemically regulates the cells it surrounds [1,2]. The ECM organizes the tissues of all metazoans [3]. It plays a role in a number of biological processes, from development and homeostasis [4–6] to pathological processes including fibrosis and cancer [4,7,8]. With a growing interest from the scientific community in the ECM and the emergence of high-throughput technologies generating large datasets came the realization that a robust definition of the proteins contributing to the formation of the ECM was needed. We thus defined the matrisome of human and mouse [9–11]. This was achieved by developing a computational approach based on protein sequence analysis using key structural features of ECM proteins, including the presence of a signal peptide and specific protein domains found predominantly in ECM and ECM-associated proteins [9,12]. We further proposed to classify the matrisome into the core matrisome, which is the compendium of genes encoding proteins forming the structure of the ECM (collagens, glycoproteins, and proteoglycans), and the matrisome-associated ensemble comprising genes encoding accessory proteins and proteins involved in the remodeling of the ECM [9,10,13]. The adoption of these definitions by the scientific community has allowed the identification of ECM proteins previously unsuspected to play roles in physiological or pathological processes [14–16] and of ECM signatures in –omic datasets predictive, for example, of cancer patient outcomes [7,17–19]. This prompted us and others to further define the matrisome of several model organisms: zebrafish [20], *Caenorhabditis elegans* [21], and planarians [22].

In recent years, there has been a surge of interest in using the genetic tractability of *Drosophila melanogaster* to identify fundamental mechanisms underlying ECM assembly, structure, and function, since several ECM proteins and processes contributing to the formation and assembly of the ECM are conserved between *Drosophila* and other organisms [23–25]. This surge is most evident in studies of basement membrane (BM) biology [26,27]. BM is an ancient and highly conserved ECM that lines the basal surface of epithelial and endothelial tissues and surrounds muscles, adipose tissue, and nerves [26,28,29]. Studies using *Drosophila* have made particularly strong contributions to our understanding of BM secretion and assembly [30–45], and the role BMs play in shaping tissues during development [35,42,46–53]. They have also shown how BMs

heal after injury [54,55] and how they regulate the immune response [56–60]. More recently, work in *Drosophila* has introduced a new role for BM proteins in intercellular adhesion [31].

Although the core BM proteins (type IV collagens, laminins, heparin sulfate proteoglycans, and nidogens) are well known, proteomic studies have revealed that BMs can harbor numerous accessory proteins that vary by tissue [14,61,62]. A comprehensive list of these proteins will provide an important tool for *Drosophila* researchers as they continue to probe the diverse roles BMs play in animal development and physiology.

Drosophila also have ECMs that are unique to arthropods and are therefore not found in any other organism for which the matrisome has been defined. These include: the chitin-based cuticle that forms the animal's exoskeleton and lines the lumens of the foregut and hindgut [63–65]; non-cuticular, chitin-based ECMs that line the lumens of the trachea, salivary glands, and midgut [64,66]; the eggshell that protects the developing embryo [67,68]; and the salivary glue that is produced by the larva to affix the pupa to a surface [69]. Defining the list of proteins that comprise these *Drosophila*-specific ECMs will provide a reference dataset for the arthropod clade and aid with the annotation of large proteomic datasets, including the developmental proteome of *Drosophila* [70]. Moreover, because insects can be both disease vectors and agricultural pests, these data could provide an important source of information to combat these threats to human welfare.

Here, we define the *in-silico* matrisome for *Drosophila melanogaster*. To this end, we developed a computational pipeline that combines orthology comparison, protein sequence analysis, interrogation of experimental proteomic data, and literature search (**Figure 1**) and identified 641 genes that comprise the *Drosophila* matrisome. We further classified these 641 genes into different structural and functional categories based on the model we have proposed for the matrisomes of other organisms [9,20,21]. We then describe the deployment of our list and terminology in the Matrisome Project website (<http://matrisome.org>) and in two databases, FlyBase [71,72] and GLAD [73], broadly used by the *Drosophila* community. Last, we illustrate how this new resource can be used to annotate -omic datasets.

2. *In-silico* definition of the *Drosophila melanogaster* matrisome

2.1. Identification of *Drosophila* orthologs of human and mouse matrisome genes

We first set out to identify the *Drosophila* orthologs of human and mouse matrisome genes. The databases Flybase (FB2017_03, released June 2017) [74], DIOPT (Version 6.0.2, released June 2017) [73], and Ensembl (Ensembl 89, released May 2017) [75] were interrogated with the full list of human and mouse core matrisome and matrisome-associated gene symbols from each of the six categories of ECM components defined by The Matrisome Project (**Figure 1**) [11]. The genes retrieved by each of the three databases (**Supplementary Table 1A and 1B**) were compiled to obtain a list of all predicted *Drosophila* orthologs of human and/or mouse matrisome genes (**Supplementary Table 1C**). The results of this approach led to the identification of 834 putative *Drosophila* matrisome orthologs. Of these genes, 114 were orthologous to a human gene but not a mouse gene, whereas 51 were orthologous to a mouse gene but not a human gene. There were 296 human genes with no *Drosophila* ortholog (**Supplementary Table 2A**) and 340 mouse genes with no *Drosophila* ortholog (**Supplementary Table 2B**).

2.2. Protein-domain-based approach to identify additional *Drosophila* matrisome proteins

Since it is well known that flies also have a large number of ECM proteins that do not have mammalian orthologs (*see Introduction*), we next used the UniProt *Drosophila* reference proteome (downloaded August 10, 2017) [76] to further expand our search for matrisome components (**Supplementary Table 3A**). Taking advantage of the conserved domain-based nature of ECM proteins [12], we selected InterPro domains which were previously used to identify human and mouse matrisome proteins [9,10], including domains characteristic of collagens, proteoglycans, and ECM-affiliated proteins, to search for ECM-domain-containing proteins in the *Drosophila* proteome (**Supplementary Table 4A**). We also included in the search three domains characteristic of proteins involved in the production and maintenance of chitin-based ECMs: insect cuticle protein (IPR000618), chitin-binding domain (IPR002557), and chitin-binding type R&R consensus (IPR031311) [77]. Although three ECM domains were initially used to search the UniProt *Drosophila* reference proteome, the domain chitin-binding type R&R consensus (IPR031311) was found to be redundant with the domain insect cuticle protein (IPR000618) for the identification of the 213 *Drosophila* proteins. (**Supplementary Table 4B**). To complete the

list of proteins composing the *Drosophila* cuticle we further interrogated CuticleDB, a database of structural components of arthropods identified experimentally or through protein sequence analysis [78]. This allowed us to retrieve an additional 7 genes (CG13670, CG7548, CG8541, CG8543, Cpr65Ax1, Edg91, Lcp6) that were added to the class of cuticular proteins.

Using this method, we identified 353 *Drosophila* proteins with ECM domains: 140 using domains previously used to identify mammalian matrisome proteins and an additional 213 using domains characteristic of *Drosophila* proteins (**Supplementary Table 4B**). We compared the list of proteins identified with human matrisome domains to the proteins identified via gene orthology and found that 49 of the 140 proteins discovered by domains characteristic of ECM proteins (35%) were not previously identified using the orthology approach (**Supplementary Table 4C**).

2.3. Gene-Ontology-based approach to identify additional *Drosophila* matrisome proteins

The *Drosophila* proteome retrieved from UniProt is also annotated with Gene Ontology (GO) – Cellular Component terms describing the intra- and extracellular localization of proteins [79,80]. The Gene Ontology terms extracellular matrix (GO:0031012), extracellular region (GO:0005576), extracellular space (GO:0005615), basement membrane (GO:0005604), and proteinaceous extracellular matrix (GO:0005578) were used to identify ECM components. The term proteinaceous extracellular matrix was found to be redundant with the term extracellular matrix, but the other four terms made significant contributions to the breadth of the search, which retrieved 1308 proteins from the *Drosophila* proteome (**Supplemental Table 3B**). As GO annotations have been found previously to lack specificity to define ECM components [10], analysis of all proteins identified by GO annotation was performed using the Phobius signal peptide predictor [81]. Proteins that lack a signal peptide and did not exhibit other significant ECM characteristics were excluded along with proteins predicted to be cytoplasmic, proteins with multiple transmembrane domains, and proteins with contradictory GO annotation such as cytosolic (GO:0005829) or lysosomal (GO:0005764) localization.

2.4. The *Drosophila* matrisome is composed of 641 genes

The three computational approaches described above identified 1,585 genes encoding potential *Drosophila* matrisome proteins. We then consulted selected published papers using proteomic and

bioinformatic methods, read reviews on *Drosophila* ECM, and made direct queries of FlyBase [71]. This final curation step allowed us to identify some genes encoding ECM proteins that were missed by our computational screen and to eliminate some genes identified computationally for which experimental evidence does not support their classification as matrisome components (see below). The combined result of these analyses is the generation of the *Drosophila* matrisome, comprised of 641 genes (**Supplemental Table 5**). Interestingly, this number represents 4% of the 15,500 protein-coding genes in the *Drosophila* genome, which is comparable to the percentage of the genome encoding ECM proteins in humans, mice, zebrafish, and *C. elegans* [9,20,21] and is likely to be similar to the proportion of matrisome genes in the planarian genome [22,82]. Below, we describe how these 641 genes have been classified into matrisome categories based on their structure, localization, and/or function.

2.4.1. Classification of Drosophila genes orthologous or homologous to mammalian matrisome genes

Genes with orthology or homology to human genes were categorized based on the previously proposed mammalian matrisome divisions (core matrisome or matrisome-associated) and categories (collagens, glycoproteins and proteoglycans for the core matrisome, and ECM-affiliated proteins, ECM regulators and secreted factors for matrisome-associated components) [9].

The *Drosophila* matrisome contains 34 core matrisome genes: 4 collagens, 27 glycoproteins, and 3 proteoglycans, the majority of which are orthologous to mammalian core matrisome genes (**Supplementary Table 5A and 5B and Figure 2A**). Only 1 collagen (pericardin) [83,84] and 6 glycoproteins (artichoke [85], anachronism [86], Defense protein l(2)34Fc, glutactin [87], tigrin [88], and tenectin [89,90]) were *Drosophila*-specific.

In addition to core matrisome genes, we predict that the *Drosophila* genome encodes 279 matrisome-associated genes, including 219 that are orthologous or homologous to mammalian genes (**Supplementary Table 5A and 5B and Figure 2A**).

We previously defined ECM-affiliated proteins as proteins either somewhat structurally related to core ECM proteins or that have been found experimentally to be associated with the ECM in detergent-insoluble fractions of tissue lysates by proteomics [9,11]. Our computational approach

predicts that 106 *Drosophila* genes encode ECM-affiliated proteins. Among these are galectins, C-type lectins (structurally characterized by three InterPro domains, IPR001304, IPR016186, and IPR016187), mucins, and semaphorins, some of which are orthologous or homologous to mammalian genes (**Supplementary Table 5A**). In addition, we classified under this category 6 collagen-triple-helix repeat-containing proteins and 9 fibrinogen-domain-containing proteins with no clear mammalian orthologs.

The ECM regulators category groups enzymes participating in the synthesis or remodeling of the ECM together with the regulators of these enzymes (including inhibitors). We identified 98 ECM regulators (**Supplementary Table 5**), including matrix metalloproteinases [25], cathepsins, ADAMs, and two orthologs of the recently identified serine/threonine kinase family Fam20 [91]. Our study also identified a total of 24 prolyl-4-hydroxylases (P4Hs). Prolyl-4-hydroxylases catalyze the formation of hydroxyprolines [24,92–94]. The most well-recognized role of this post-translational modification is to stabilize collagen triple-helical structures. Interestingly, and as previously noted [24], the human genome encodes 44 collagen genes and 3 P4Hs, whereas the *Drosophila* genome encodes only 4 collagen genes, 6 collagen-triple-helix repeat-containing proteins and yet 24 P4Hs. Both previous work [94] and interrogation of The National Human Genome Research Institute **model** organism **ENCyclopedia Of DNA Element** (modENCODE) database [95] indicate that the P4Hs are expressed in a tissue-specific manner and at different developmental stages. Whether P4Hs have additional substrates in *Drosophila* remains to be determined.

Last, we previously included secreted factors in our definition of the matrisome, since the ECM is recognized as a reservoir of growth factors and other soluble factors [96]. These 75 proteins (**Supplementary Table 5**) were defined using a combination of orthology or homology annotations, GO terms, literature references, and the presence of characteristic domains not previously used to define secreted factors but identified from the examination of the *Drosophila melanogaster* extracellular domain database (FlyXCDB, [97]). These domains are the PDGF/VEGF domain (IPR000072), the Spaetzle domain (IPR032104), the von Willebrand factor type C (IPR029277), the insulin-like domain (IPR016179), the eclosion hormone domain (IPR006825), and the interleukin-17 family domain (IPR010345).

2.4.2. Classification of *Drosophila* genes with no mammalian orthologs or homologs

Since the chitin-based ECM, eggshell, and salivary glue are all secreted from the apical side of epithelial tissues, we classified both the structural and regulatory proteins associated with these ECMs under a category termed “Apical ECM” (**Supplementary Table 5 and Figure 2B**). As a group, these proteins comprise nearly 50% of the *Drosophila* matrisome. To further subdivide this diverse group of proteins, an additional level of classification was created to reflect their respective proteins’ domain structure, enzymatic function, or localization (**Supplementary Table 5, column C**). The chitin-binding-domain-containing proteins and R&R chitin-binding-domain-containing proteins families refer to proteins containing InterPro domains IPR002557 and IPR031311, respectively. The Chitinase and Chitinase-like families also have a group of defining domains: the chitinase II domain (IPR011583) and three glycoside hydrolase domains (IPR029070, IPR001223, and IPR017853). The Tweedle family represents the only proteins with the domain DUF243 (IPR004145) [98]. Chitin deacetylases were identified based on the presence of a glycoside hydrolase/deacetylase domain (IPR011330). A group of 11 zona-pellucida-domain-containing proteins was also identified [99]. These proteins have a shared structural attribute, the zona pellucida domain (IPR001507), which we originally used to identify core components of the mammalian matrisome. However, since zona-pellucida-domain-containing proteins do not present clear orthology or homology with mammalian proteins, we classified them apart.

Groups without clear structural similarities were classified by other means. Proteins of the cuticle that did not meet the definitions above were classified by their shared GO term, chitin-based cuticle development (GO:0040003). Included in this class were also a number of genes reported to be cuticle proteins of low complexity [100,101]. The eggshell superfamily includes two protein classes, corresponding to the vitelline membrane and chorion layers of this ECM, respectively [68,102,103]. The vitelline membrane proteins were defined by GO term or literature search. The chorion proteins had all previously been assigned chorion-related GO terms and chorion-related protein names, except Cp38 which has chorion in the name and is cited [102]. Finally, 11 proteins including new-glue and salivary glue secretion proteins, Eig71Ee [69,104], and the newly identified tandem paralog of Sgs5 (FBgn0038523) [105] were classified as glue proteins.

3. Accessing the *Drosophila* matrisome and utilizing it to annotate large datasets

3.1. The *Drosophila* matrisome is available from three sources

To facilitate the use of our definition and categorization of the *Drosophila* matrisome by the scientific community, the list devised here has been made available through three public platforms. Similar to the matrisome lists of human, mouse, zebrafish and *C. elegans*, the *Drosophila* matrisome list can be found on the Matrisome Project website (<http://matrisome.org>) [11]. Moreover, it has been implemented in two databases widely used by the *Drosophila* community. The *Drosophila* matrisome is available within the “Gene Groups” section of the August 2019 release of [FlyBase](#) (FB2019_04), which is the most comprehensive source of genetic information for this model organism [71,72,106]. In addition, as a result of the Matrisome analysis, two new terms were added to the Gene Ontology Cellular Component aspect: chitin-based extracellular matrix (GO:0062129) and adhesive extracellular matrix (GO:0062130), allowing more precise GO annotation of the constituents of these specific types of ECM. All *Drosophila* cuticle proteins and glue genes have now been annotated with these respective terms in FlyBase.

The *Drosophila* matrisome is also available in the [Gene List Annotation for *Drosophila*](#) (GLAD) database, which is maintained by the Perrimon laboratory to enhance the utility of the cell-based RNAi screening (DRSC) and *in vivo* fly RNAi (TRiP) collections for the community [73]. For consistency with the current GLAD nomenclature, the matrisome forms a new gene list/group; the matrisome divisions are listed as sub-groups, the categories as sub-sub-groups, and the families are listed under comments.

3.2 The *Drosophila* matrisome provides a powerful tool to annotate large datasets

One powerful application of the matrisome list for any species is in the annotation of large -omic datasets [11]. Thus, as a proof of principle, we used the newly defined *Drosophila* matrisome to re-evaluate two recently published datasets that focus heavily on ECM-associated proteins. In the first study, Baycin-Hizal and colleagues identified 399 N-glycosylated proteins of the *Drosophila* head region using solid phase extraction of N-linked glycopeptides coupled to LC-MS/MS [107]. They reported that 4.5% of the proteins identified experimentally in their study were part of the ECM. We found, however, that 13% of the proteins they identified (which included 8 of the 26 glycoproteins and 2 of the 3 proteoglycans we have predicted) are in fact matrisome proteins, more than double the original number. In the second study, Sessions and colleagues reported changes in

the abundance of ECM proteins in the *Drosophila* heart during aging [108]. 104 of the proteins detected were identified as ECM proteins using the Software Tool for Rapid Annotation and Differential Comparison of Protein Post-Translational Modifications (STRAP PTM) developed by Spender and colleagues. Of these 104 proteins, 27 are part of the matrisome, whereas 77 are not. Examination of these 77 proteins revealed that most are in fact localized intracellularly, with little evidence to support that they are ECM components. We retrieved the raw mass spectrometry data from the ProteomeXchange repository (PXD006120) and reannotated the data using the matrisome list. We identified a total of 46 matrisome proteins, finding 19 additional proteins not originally annotated as belonging to the ECM. Together, these two examples demonstrate the power of our matrisome list to comprehensively annotate large experimental datasets. We thus propose that the use of our annotations and nomenclature would assist in the comprehensive identification of ECM signatures contributing to cellular, physiological and pathological phenotypes.

4. Conclusion

We defined here the *in-silico Drosophila melanogaster* matrisome. In addition to reporting the identification of 641 genes encoding ECM and ECM-associated proteins, we further propose their comprehensive classification according to structural and/or functional features. We hope that this list and nomenclature will aid with the annotations of large datasets, and thus further our understanding of the roles of the ECM in fundamental biological processes and pathophysiology.

Acknowledgements

We would like to thank the members of the Naba Lab for critical reading of the manuscript. We would also like to thank our long-time collaborator Karl Clauser for his help analyzing the raw proteomic data from the Sessions et al., 2016 study.

Last, we would like to thank Dr. Helen Attrill (University of Cambridge), curator of FlyBase, for her help with the curation of the matrisome list and for its implementation to FlyBase, and Dr. Claire Yanhui Hu from the *Drosophila* RNAi Screening Center (DRSC) and Transgenic RNAi Project (TRiP) in Norbert Perrimon 's laboratory (Harvard Medical School) for implementing our list to GLAD.

Funding

This work was supported by a start-up fund from the Department of Physiology and Biophysics at UIC and a Catalyst Award from the Chicago Biomedical Consortium with support from the Searle Funds at the Chicago Community Trust to AN, and by grants from the American Cancer Society (RSG-14-176) and National Institutes of Health (R01-GM126047) to SHB.

REFERENCES

- [1] R.O. Hynes, K.M. Yamada, *Extracellular Matrix Biology.*, Cold Spring Harbor Perspectives in Biology, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2012. http://cshperspectives.cshlp.org/site/misc/extracellular_matrix_biology.xhtml.
- [2] J.H. Fessler, L.I. Fessler, *Drosophila extracellular matrix*, *Annu. Rev. Cell Biol.* 5 (1989) 309–339. doi:10.1146/annurev.cb.05.110189.001521.
- [3] S. Özbek, P.G. Balasubramanian, R. Chiquet-Ehrismann, R.P. Tucker, J.C. Adams, *The Evolution of Extracellular Matrix*, *Mol. Biol. Cell.* 21 (2010) 4300–4305. doi:10.1091/mbc.E10-03-0251.
- [4] C. Bonnans, J. Chou, Z. Werb, *Remodelling the extracellular matrix in development and disease*, *Nat. Rev. Mol. Cell Biol.* 15 (2014) 786–801. doi:10.1038/nrm3904.
- [5] T. Rozario, D.W. DeSimone, *The extracellular matrix in development and morphogenesis: a dynamic view*, *Dev. Biol.* 341 (2010) 126–140. doi:10.1016/j.ydbio.2009.10.026.
- [6] B.J. Dzamba, D.W. DeSimone, *Extracellular Matrix (ECM) and the Sculpting of Embryonic Tissues*, *Curr. Top. Dev. Biol.* 130 (2018) 245–274. doi:10.1016/bs.ctdb.2018.03.006.
- [7] A.M. Socovich, A. Naba, *The cancer matrisome: From comprehensive characterization to biomarker discovery*, *Semin. Cell Dev. Biol.* 89 (2019) 157–166. doi:10.1016/j.semcdb.2018.06.005.
- [8] P. Pakshir, B. Hinz, *The big five in fibrosis: Macrophages, myofibroblasts, matrix, mechanics, and miscommunication*, *Matrix Biol.* 68–69 (2018) 81–93. doi:10.1016/j.matbio.2018.01.019.
- [9] A. Naba, K.R. Clauser, S. Hoersch, H. Liu, S.A. Carr, R.O. Hynes, *The matrisome: in silico definition and in vivo characterization by proteomics of normal and tumor extracellular matrices*, *Mol. Cell. Proteomics.* 11 (2012) M111.014647. doi:10.1074/mcp.M111.014647.
- [10] A. Naba, S. Hoersch, R.O. Hynes, *Towards definition of an ECM parts list: an advance on GO categories*, *Matrix Biol.* 31 (2012) 371–372. doi:10.1016/j.matbio.2012.11.008.
- [11] A. Naba, K.R. Clauser, H. Ding, C.A. Whittaker, S.A. Carr, R.O. Hynes, *The extracellular matrix: Tools and insights for the “omics” era*, *Matrix Biol.* 49 (2016) 10–24. doi:10.1016/j.matbio.2015.06.003.
- [12] E. Hohenester, J. Engel, *Domain structure and organisation in extracellular matrix proteins*, *Matrix Biol.* 21 (2002) 115–128. doi:10.1016/S0945-053X(01)00191-3.
- [13] R.O. Hynes, A. Naba, *Overview of the Matrisome--An Inventory of Extracellular Matrix Constituents and Functions*, *Cold Spring Harb. Perspect. Biol.* 4 (2012) a004903–a004903. doi:10.1101/cshperspect.a004903.
- [14] R. Lennon, A. Byron, J.D. Humphries, M.J. Randles, A. Carisey, S. Murphy, D. Knight, P.E. Brenchley, R. Zent, M.J. Humphries, *Global analysis reveals the complexity of the human glomerular extracellular matrix*, *J. Am. Soc. Nephrol.* 25 (2014) 939–951. doi:10.1681/ASN.2013030233.
- [15] V.L. Massey, C.E. Dolin, L.G. Poole, S.V. Hudson, D.L. Siow, G.N. Brock, M.L. Merchant, D.W. Wilkey, G.E. Arteel, *The hepatic “matrisome” responds dynamically to injury*:

- Characterization of transitional changes to the extracellular matrix in mice, *Hepatology*. Baltim. Md. 65 (2017) 969–982. doi:10.1002/hep.28918.
- [16] M.C. Staiculescu, J. Kim, R.P. Mecham, J. Wagenseil, Mechanical behavior and matrisome gene expression in aneurysm-prone thoracic aorta of newborn lysyl oxidase knockout mice, *Am. J. Physiol. Heart Circ. Physiol.* (2017) ajpheart.00712.2016. doi:10.1152/ajpheart.00712.2016.
- [17] V. Izzi, J. Lakkala, R. Devarajan, A. Kääriäinen, J. Koivunen, R. Heljasvaara, T. Pihlajaniemi, Pan-Cancer analysis of the expression and regulation of matrisome genes across 32 tumor types, *Matrix Biol. Plus.* (2019) 100004. doi:10.1016/j.mbplus.2019.04.001.
- [18] A.E. Yuzhalin, T. Urbonas, M.A. Silva, R.J. Muschel, A.N. Gordon-Weeks, A core matrisome gene signature predicts cancer outcome, *Br. J. Cancer.* 118 (2018) 435–440. doi:10.1038/bjc.2017.458.
- [19] O.M.T. Pearce, R.M. Delaine-Smith, E. Maniati, S. Nichols, J. Wang, S. Böhm, V. Rajeeve, D. Ullah, P. Chakravarty, R.R. Jones, A. Montfort, T. Dowe, J. Gribben, J.L. Jones, H.M. Kocher, J.S. Serody, B.G. Vincent, J. Connelly, J.D. Brenton, C. Chelala, P.R. Cutillas, M. Lockley, C. Bessant, M.M. Knight, F.R. Balkwill, Deconstruction of a Metastatic Tumor Microenvironment Reveals a Common Matrix Response in Human Cancers, *Cancer Discov.* 8 (2018) 304–319. doi:10.1158/2159-8290.CD-17-0284.
- [20] P. Nauroy, S. Hughes, A. Naba, F. Ruggiero, The in-silico zebrafish matrisome: A new tool to study extracellular matrix gene and protein functions, *Matrix Biol.* 65 (2018) 5–13. doi:10.1016/j.matbio.2017.07.001.
- [21] A.C. Teuscher, E. Jongsma, M.N. Davis, C. Statzer, J.M. Gebauer, A. Naba, C.Y. Ewald, The in-silico characterization of the *Caenorhabditis elegans* matrisome and proposal of a novel collagen classification, *Matrix Biol. Plus.* 1 (2019) 100001. doi:10.1016/j.mbplus.2018.11.001.
- [22] L.E. Cote, E. Simental, P.W. Reddien, Muscle functions as a connective tissue and source of extracellular matrix in planarians, *Nat. Commun.* 10 (2019) 1592. doi:10.1038/s41467-019-09539-6.
- [23] J.C. Adams, Matricellular Proteins: Functional Insights From Non-mammalian Animal Models, *Curr. Top. Dev. Biol.* 130 (2018) 39–105. doi:10.1016/bs.ctdb.2018.02.003.
- [24] J. Myllyharju, K.I. Kivirikko, Collagens, modifying enzymes and their mutations in humans, flies and worms, *Trends Genet. TIG.* 20 (2004) 33–43. doi:10.1016/j.tig.2003.11.004.
- [25] A. Page-McCaw, Remodeling the model organism: matrix metalloproteinase functions in invertebrates, *Semin. Cell Dev. Biol.* 19 (2008) 14–23. doi:10.1016/j.semcdb.2007.06.004.
- [26] W. Ramos-Lewis, A. Page-McCaw, Basement membrane mechanics shape development: Lessons from the fly, *Matrix Biol.* 75–76 (2019) 72–81. doi:10.1016/j.matbio.2018.04.004.

- [27] A.J. Isabella, S. Horne-Badovinac, Building from the Ground up: Basement Membranes in *Drosophila* Development, *Curr. Top. Membr.* 76 (2015) 305–336. doi:10.1016/bs.ctm.2015.07.001.
- [28] R. Jayadev, D.R. Sherwood, Basement membranes, *Curr. Biol. CB.* 27 (2017) R207–R211. doi:10.1016/j.cub.2017.02.006.
- [29] A. Pozzi, P.D. Yurchenco, R.V. Iozzo, The nature and biology of basement membranes, *Matrix Biol.* 57–58 (2017) 1–11. doi:10.1016/j.matbio.2016.12.009.
- [30] G. Wolfstetter, I. Dahlitz, K. Pfeifer, U. Töpfer, J.A. Alt, D.C. Pfeifer, R. Lakes-Harlan, S. Baumgartner, R.H. Palmer, A. Holz, Characterization of *Drosophila* Nidogen/entactin reveals roles in basement membrane stability, barrier function and nervous system patterning, *Dev. Camb. Engl.* 146 (2019). doi:10.1242/dev.168948.
- [31] J. Dai, M. Ma, Z. Feng, J.C. Pastor-Pareja, Inter-adipocyte Adhesion and Signaling by Collagen IV Intercellular Concentrations in *Drosophila*, *Curr. Biol. CB.* 27 (2017) 2729–2740.e4. doi:10.1016/j.cub.2017.08.002.
- [32] K. Itoh, Y. Akimoto, S. Kondo, T. Ichimiya, K. Aoki, M. Tiemeyer, S. Nishihara, Glucuronylated core 1 glycans are required for precise localization of neuromuscular junctions and normal formation of basement membranes on *Drosophila* muscles, *Dev. Biol.* 436 (2018) 108–124. doi:10.1016/j.ydbio.2018.02.017.
- [33] Y. Matsubayashi, A. Louani, A. Dragu, B.J. Sánchez-Sánchez, E. Serna-Morales, L. Yolland, A. Gyoergy, G. Vizcay, R.A. Fleck, J.M. Heddleston, T.-L. Chew, D.E. Siekhaus, B.M. Stramer, A Moving Source of Matrix Components Is Essential for De Novo Basement Membrane Formation, *Curr. Biol. CB.* 27 (2017) 3526–3534.e4. doi:10.1016/j.cub.2017.10.001.
- [34] O. Devergne, G.H. Sun, T. Schüpbach, Stratum, a Homolog of the Human GEF Mss4, Partnered with Rab8, Controls the Basal Restriction of Basement Membrane Proteins in Epithelial Cells, *Cell Rep.* 18 (2017) 1831–1839. doi:10.1016/j.celrep.2017.02.002.
- [35] A.J. Isabella, S. Horne-Badovinac, Rab10-Mediated Secretion Synergizes with Tissue Movement to Build a Polarized Basement Membrane Architecture for Organ Morphogenesis, *Dev. Cell.* 38 (2016) 47–60. doi:10.1016/j.devcel.2016.06.009.
- [36] C.F. Cummings, V. Pedchenko, K.L. Brown, S. Colon, M. Rafi, C. Jones-Paris, E. Pokydeshava, M. Liu, J.C. Pastor-Pareja, C. Stothers, I.A. Ero-Tolliver, A.S. McCall, R. Vanacore, G. Bhave, S. Santoro, T.S. Blackwell, R. Zent, A. Pozzi, B.G. Hudson, Extracellular chloride signals collagen IV network assembly during basement membrane formation, *J. Cell Biol.* 213 (2016) 479–494. doi:10.1083/jcb.201510065.
- [37] J. Shahab, C. Baratta, B. Scuric, D. Godt, K.J.T. Venken, M.J. Ringuette, Loss of SPARC dysregulates basal lamina assembly to disrupt larval fat body homeostasis in *Drosophila melanogaster*, *Dev. Dyn. Off. Publ. Am. Assoc. Anat.* 244 (2015) 540–552. doi:10.1002/dvdy.24243.
- [38] D. Hollfelder, M. Frasch, I. Reim, Distinct functions of the laminin β LN domain and collagen IV during cardiac extracellular matrix formation and stabilization of alary muscle

- attachments revealed by EMS mutagenesis in *Drosophila*, *BMC Dev. Biol.* 14 (2014) 26. doi:10.1186/1471-213X-14-26.
- [39] A.S. McCall, C.F. Cummings, G. Bhave, R. Vanacore, A. Page-McCaw, B.G. Hudson, Bromine is an essential trace element for assembly of collagen IV scaffolds in tissue development and architecture, *Cell.* 157 (2014) 1380–1392. doi:10.1016/j.cell.2014.05.009.
- [40] O. Devergne, K. Tsung, G. Barcelo, T. Schüpbach, Polarized deposition of basement membrane proteins depends on Phosphatidylinositol synthase and the levels of Phosphatidylinositol 4,5-bisphosphate, *Proc. Natl. Acad. Sci. U. S. A.* 111 (2014) 7689–7694. doi:10.1073/pnas.1407351111.
- [41] D.W. Lerner, D. McCoy, A.J. Isabella, A.P. Mahowald, G.F. Gerlach, T.A. Chaudhry, S. Horne-Badovinac, A Rab10-dependent mechanism for polarized basement membrane secretion during organ morphogenesis, *Dev. Cell.* 24 (2013) 159–168. doi:10.1016/j.devcel.2012.12.005.
- [42] J.C. Pastor-Pareja, T. Xu, Shaping Cells and Organs in *Drosophila* by Opposing Roles of Fat Body-Secreted Collagen IV and Perlecan, *Dev. Cell.* 21 (2011) 245–256. doi:10.1016/j.devcel.2011.06.026.
- [43] G. Sorrosal, L. Pérez, H. Herranz, M. Milán, Scarface, a secreted serine protease-like protein, regulates polarized localization of laminin A at the basement membrane of the *Drosophila* embryo, *EMBO Rep.* 11 (2010) 373–379. doi:10.1038/embor.2010.43.
- [44] J.M. Urbano, C.N. Torgler, C. Molnar, U. Tepass, A. López-Varea, N.H. Brown, J.F. de Celis, M.D. Martín-Bermudo, *Drosophila* laminins act as key regulators of basement membrane assembly and morphogenesis, *Dev. Camb. Engl.* 136 (2009) 4165–4176. doi:10.1242/dev.044263.
- [45] N. Deneff, Y. Chen, S.D. Weeks, G. Barcelo, T. Schüpbach, Crag regulates epithelial architecture and polarized deposition of basement membrane proteins in *Drosophila*, *Dev. Cell.* 14 (2008) 354–364. doi:10.1016/j.devcel.2007.12.012.
- [46] J. Wittes, T. Schüpbach, A Gene Expression Screen in *Drosophila melanogaster* Identifies Novel JAK/STAT and EGFR Targets During Oogenesis, *G3 Bethesda Md.* 9 (2019) 47–60. doi:10.1534/g3.118.200786.
- [47] J. Chlasta, P. Milani, G. Runel, J.-L. Duteyrat, L. Arias, L.-A. Lamiré, A. Boudaoud, M. Grammont, Variations in basement membrane mechanics are linked to epithelial morphogenesis, *Dev. Camb. Engl.* 144 (2017) 4350–4362. doi:10.1242/dev.152652.
- [48] J.B. Skeath, B.A. Wilson, S.E. Romero, M.J. Snee, Y. Zhu, H. Lacin, The extracellular metalloprotease AdamTS-A anchors neural lineages in place within and preserves the architecture of the central nervous system, *Dev. Camb. Engl.* 144 (2017) 3102–3113. doi:10.1242/dev.145854.
- [49] M. Ma, X. Cao, J. Dai, J.C. Pastor-Pareja, Basement Membrane Manipulation in *Drosophila* Wing Discs Affects Dpp Retention but Not Growth Mechanoregulation, *Dev. Cell.* 42 (2017) 97–106.e4. doi:10.1016/j.devcel.2017.06.004.

- [50] M.C. Díaz de la Loza, A. Díaz-Torres, F. Zurita, A.E. Rosales-Nieves, E. Moeendarbary, K. Franze, M.D. Martín-Bermudo, A. González-Reyes, Laminin Levels Regulate Tissue Migration and Anterior-Posterior Polarity during Egg Morphogenesis in *Drosophila*, *Cell Rep.* 20 (2017) 211–223. doi:10.1016/j.celrep.2017.06.031.
- [51] J. Crest, A. Diz-Muñoz, D.-Y. Chen, D.A. Fletcher, D. Bilder, Organ sculpting by patterned extracellular matrix stiffness, *ELife.* 6 (2017). doi:10.7554/eLife.24958.
- [52] A.J. Isabella, S. Horne-Badovinac, Dynamic regulation of basement membrane protein levels promotes egg chamber elongation in *Drosophila*, *Dev. Biol.* 406 (2015) 212–221. doi:10.1016/j.ydbio.2015.08.018.
- [53] S.L. Haigo, D. Bilder, Global tissue revolutions in a morphogenetic movement controlling elongation, *Science.* 331 (2011) 1071–1074. doi:10.1126/science.1199424.
- [54] A.M. Howard, K.S. LaFever, A.M. Fenix, C.R. Scurrah, K.S. Lau, D.T. Burnette, G. Bhave, N. Ferrell, A. Page-McCaw, DSS-induced damage to basement membranes is repaired by matrix replacement and crosslinking, *J. Cell Sci.* 132 (2019). doi:10.1242/jcs.226860.
- [55] W. Ramos-Lewis, K.S. LaFever, A. Page-McCaw, A scar-like lesion is apparent in basement membrane after wound repair in vivo, *Matrix Biol.* 74 (2018) 101–120. doi:10.1016/j.matbio.2018.07.004.
- [56] M. Kiss, A.A. Kiss, M. Radics, N. Popovics, E. Hermes, K. Csiszár, M. Mink, *Drosophila* type IV collagen mutation associates with immune system activation and intestinal dysfunction, *Matrix Biol.* 49 (2016) 120–131. doi:10.1016/j.matbio.2015.09.002.
- [57] Y. Zang, M. Wan, M. Liu, H. Ke, S. Ma, L.-P. Liu, J.-Q. Ni, J.C. Pastor-Pareja, Plasma membrane overgrowth causes fibrotic collagen accumulation and immune activation in *Drosophila* adipocytes, *ELife.* 4 (2015) e07187. doi:10.7554/eLife.07187.
- [58] M.J. Kim, K.-M. Choe, Basement membrane and cell integrity of self-tissues in maintaining *Drosophila* immunological tolerance, *PLoS Genet.* 10 (2014) e1004683. doi:10.1371/journal.pgen.1004683.
- [59] B. Arefin, L. Kucerova, P. Dobes, R. Markus, H. Strnad, Z. Wang, P. Hyrs, M. Zurovec, U. Theopold, Genome-wide transcriptional analysis of *Drosophila* larvae infected by entomopathogenic nematodes shows involvement of complement, recognition and extracellular matrix proteins, *J. Innate Immun.* 6 (2014) 192–204. doi:10.1159/000353734.
- [60] J.C. Pastor-Pareja, M. Wu, T. Xu, An innate immune response of blood cells to tumors and tissue damage in *Drosophila*, *Dis. Model. Mech.* 1 (2008) 144–154; discussion 153. doi:10.1242/dmm.000950.
- [61] G. Uechi, Z. Sun, E.M. Schreiber, W. Halfter, M. Balasubramani, Proteomic view of basement membranes from human retinal blood vessels, inner limiting membranes, and lens capsules, *J. Proteome Res.* (2014). doi:10.1021/pr5002065.
- [62] M.J. Randles, M.J. Humphries, R. Lennon, Proteomic definitions of basement membrane composition in health and disease, *Matrix Biol.* 57–58 (2017) 12–28. doi:10.1016/j.matbio.2016.08.006.

- [63] A. Öztürk-Çolak, B. Moussian, S.J. Araújo, *Drosophila* chitinous aECM and its cellular interactions during tracheal development, *Dev. Dyn. Off. Publ. Am. Assoc. Anat.* 245 (2016) 259–267. doi:10.1002/dvdy.24356.
- [64] B. Lemaitre, I. Miguel-Aliaga, The digestive tract of *Drosophila melanogaster*, *Annu. Rev. Genet.* 47 (2013) 377–404. doi:10.1146/annurev-genet-111212-133343.
- [65] A.L. Stahl, M. Charlton-Perkins, E.K. Buschbeck, T.A. Cook, The cuticular nature of corneal lenses in *Drosophila melanogaster*, *Dev. Genes Evol.* 227 (2017) 271–278. doi:10.1007/s00427-017-0582-7.
- [66] S. Luschnig, A. Uv, Luminal matrices: an inside view on organ morphogenesis, *Exp. Cell Res.* 321 (2014) 64–70. doi:10.1016/j.yexcr.2013.09.010.
- [67] G.L. Waring, Morphogenesis of the eggshell in *Drosophila*, *Int. Rev. Cytol.* 198 (2000) 67–108.
- [68] L.H. Margaritis, F.C. Kafatos, W.H. Petri, The eggshell of *Drosophila melanogaster*. I. Fine structure of the layers and regions of the wild-type eggshell, *J. Cell Sci.* 43 (1980) 1–35.
- [69] D. Benova-Liszekova, M. Beño, R. Farkas, Fine infrastructure of released and solidified *Drosophila* larval salivary secretory glue and salivary gland ducts using SEM, *Bioinspir. Biomim.* (2019). doi:10.1088/1748-3190/ab2b2b.
- [70] N. Casas-Vila, A. Bluhm, S. Sayols, N. Dinges, M. Dejung, T. Altenhein, D. Kappeli, B. Altenhein, J.-Y. Roignant, F. Butter, The developmental proteome of *Drosophila melanogaster*, *Genome Res.* 27 (2017) 1273–1285. doi:10.1101/gr.213694.116.
- [71] J. Thurmond, J.L. Goodman, V.B. Strelets, H. Attrill, L.S. Gramates, S.J. Marygold, B.B. Matthews, G. Millburn, G. Antonazzo, V. Trovisco, T.C. Kaufman, B.R. Calvi, FlyBase Consortium, FlyBase 2.0: the next generation, *Nucleic Acids Res.* 47 (2019) D759–D765. doi:10.1093/nar/gky1003.
- [72] H. Attrill, K. Falls, J.L. Goodman, G.H. Millburn, G. Antonazzo, A.J. Rey, S.J. Marygold, FlyBase: establishing a Gene Group resource for *Drosophila melanogaster*, *Nucleic Acids Res.* 44 (2016) D786–D792. doi:10.1093/nar/gkv1046.
- [73] Y. Hu, A. Comjean, L.A. Perkins, N. Perrimon, S.E. Mohr, GLAD: an Online Database of Gene List Annotation for *Drosophila*, *J. Genomics.* 3 (2015) 75–81. doi:10.7150/jgen.12863.
- [74] L.S. Gramates, S.J. Marygold, G.D. Santos, J.-M. Urbano, G. Antonazzo, B.B. Matthews, A.J. Rey, C.J. Tabone, M.A. Crosby, D.B. Emmert, K. Falls, J.L. Goodman, Y. Hu, L. Ponting, A.J. Schroeder, V.B. Strelets, J. Thurmond, P. Zhou, the FlyBase Consortium, FlyBase at 25: looking to the future, *Nucleic Acids Res.* 45 (2017) D663–D671. doi:10.1093/nar/gkw1016.
- [75] D.R. Zerbino, P. Achuthan, W. Akanni, M.R. Amode, D. Barrell, J. Bhai, K. Billis, C. Cummins, A. Gall, C.G. Girón, L. Gil, L. Gordon, L. Haggerty, E. Haskell, T. Hourlier, O.G. Izuogu, S.H. Janacek, T. Juettemann, J.K. To, M.R. Laird, I. Lavidas, Z. Liu, J.E. Loveland, T. Maurel, W. McLaren, B. Moore, J. Mudge, D.N. Murphy, V. Newman, M. Nuhn, D. Ogeh, C.K. Ong, A. Parker, M. Patricio, H.S. Riat, H. Schuilenburg, D. Sheppard,

- H. Sparrow, K. Taylor, A. Thormann, A. Vullo, B. Walts, A. Zadissa, A. Frankish, S.E. Hunt, M. Kostadima, N. Langridge, F.J. Martin, M. Muffato, E. Perry, M. Ruffier, D.M. Staines, S.J. Trevanion, B.L. Aken, F. Cunningham, A. Yates, P. Flicek, *Ensembl* 2018, *Nucleic Acids Res.* 46 (2018) D754–D761. doi:10.1093/nar/gkx1098.
- [76] The UniProt Consortium, UniProt: the universal protein knowledgebase, *Nucleic Acids Res.* 45 (2017) D158–D169. doi:10.1093/nar/gkw1099.
- [77] M.V. Karouzou, Y. Spyropoulos, V.A. Iconomidou, R.S. Cornman, S.J. Hamodrakas, J.H. Willis, *Drosophila* cuticular proteins with the R&R Consensus: annotation and classification with a new tool for discriminating RR-1 and RR-2 sequences, *Insect Biochem. Mol. Biol.* 37 (2007) 754–760. doi:10.1016/j.ibmb.2007.03.007.
- [78] C.K. Magkrioti, I.C. Spyropoulos, V.A. Iconomidou, J.H. Willis, S.J. Hamodrakas, *cuticleDB: a relational database of Arthropod cuticular proteins*, *BMC Bioinformatics.* 5 (2004) 138. doi:10.1186/1471-2105-5-138.
- [79] M. Ashburner, C.A. Ball, J.A. Blake, D. Botstein, H. Butler, J.M. Cherry, A.P. Davis, K. Dolinski, S.S. Dwight, J.T. Eppig, M.A. Harris, D.P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J.C. Matese, J.E. Richardson, M. Ringwald, G.M. Rubin, G. Sherlock, *Gene ontology: tool for the unification of biology*. The Gene Ontology Consortium, *Nat. Genet.* 25 (2000) 25–29. doi:10.1038/75556.
- [80] The Gene Ontology Consortium, *Expansion of the Gene Ontology knowledgebase and resources*, *Nucleic Acids Res.* 45 (2017) D331–D338. doi:10.1093/nar/gkw1108.
- [81] L. Käll, A. Krogh, E.L.L. Sonnhammer, *Advantages of combined transmembrane topology and signal peptide prediction--the Phobius web server*, *Nucleic Acids Res.* 35 (2007) W429–432. doi:10.1093/nar/gkm256.
- [82] M.A. Grohme, S. Schloissnig, A. Rozanski, M. Pippel, G.R. Young, S. Winkler, H. Brandl, I. Henry, A. Dahl, S. Powell, M. Hiller, E. Myers, J.C. Rink, *The genome of Schmidtea mediterranea and the evolution of core cellular mechanisms*, *Nature.* 554 (2018) 56–61. doi:10.1038/nature25473.
- [83] A. Chartier, S. Zaffran, M. Astier, M. Sémériva, D. Gratecos, Pericardin, a *Drosophila* type IV collagen-like protein is involved in the morphogenesis and maintenance of the heart epithelium during dorsal ectoderm closure, *Dev. Camb. Engl.* 129 (2002) 3241–3253.
- [84] A.C. Wilmes, N. Klinke, B. Rotstein, H. Meyer, A. Paululat, *Biosynthesis and assembly of the Collagen IV-like protein Pericardin in Drosophila melanogaster*, *Biol. Open.* 7 (2018). doi:10.1242/bio.030361.
- [85] M. Andrés, E. Turiégano, M.C. Göpfert, I. Canal, L. Torroja, *The extracellular matrix protein artichoke is required for integrity of ciliated mechanosensory and chemosensory organs in Drosophila embryos*, *Genetics.* 196 (2014) 1091–1102. doi:10.1534/genetics.113.156323.
- [86] A.J. Ebens, H. Garren, B.N. Cheyette, S.L. Zipursky, *The Drosophila anachronism locus: a glycoprotein secreted by glia inhibits neuroblast proliferation*, *Cell.* 74 (1993) 15–27.

- [87] P.F. Olson, L.I. Fessler, R.E. Nelson, R.E. Sterne, A.G. Campbell, J.H. Fessler, Glutactin, a novel *Drosophila* basement membrane-related glycoprotein with sequence similarity to serine esterases, *EMBO J.* 9 (1990) 1219–1227.
- [88] F.J. Fogerty, L.I. Fessler, T.A. Bunch, Y. Yaron, C.G. Parker, R.E. Nelson, D.L. Brower, D. Gullberg, J.H. Fessler, Tiggrin, a novel *Drosophila* extracellular matrix protein that functions as a ligand for *Drosophila* alpha PS2 beta PS integrins, *Dev. Camb. Engl.* 120 (1994) 1747–1758.
- [89] S. Fraichard, A.-L. Bouge, I. Chauvel, H. Bouhin, Tenectin, a novel extracellular matrix protein expressed during *Drosophila melanogaster* embryonic development, *Gene Expr. Patterns GEP.* 6 (2006) 772–776. doi:10.1016/j.modgep.2006.01.007.
- [90] S. Fraichard, A.-L. Bougé, T. Kendall, I. Chauvel, H. Bouhin, T.A. Bunch, Tenectin is a novel alphaPS2betaPS integrin ligand required for wing morphogenesis and male genital looping in *Drosophila*, *Dev. Biol.* 340 (2010) 504–517. doi:10.1016/j.ydbio.2010.02.008.
- [91] V.S. Tagliabracci, J.L. Engel, J. Wen, S.E. Wiley, C.A. Worby, L.N. Kinch, J. Xiao, N.V. Grishin, J.E. Dixon, Secreted kinase phosphorylates extracellular proteins that regulate biomineralization, *Science.* 336 (2012) 1150–1153. doi:10.1126/science.1217817.
- [92] J. Myllyharju, Prolyl 4-hydroxylases, the key enzymes of collagen biosynthesis, *Matrix Biol.* 22 (2003) 15–24.
- [93] E. Freydl, F. Meins, T. Boller, J.-M. Neuhaus, Kinetics of prolyl hydroxylation, intracellular transport and C-terminal processing of the tobacco vacuolar chitinase, *Planta.* 197 (1995) 250–256. doi:10.1007/BF00202644.
- [94] E.W. Abrams, D.J. Andrew, Prolyl 4-hydroxylase alpha-related proteins in *Drosophila melanogaster*: tissue-specific embryonic expression of the 99F8-9 cluster, *Mech. Dev.* 112 (2002) 165–171.
- [95] S.E. Celniker, L.A.L. Dillon, M.B. Gerstein, K.C. Gunsalus, S. Henikoff, G.H. Karpen, M. Kellis, E.C. Lai, J.D. Lieb, D.M. MacAlpine, G. Micklem, F. Piano, M. Snyder, L. Stein, K.P. White, R.H. Waterston, modENCODE Consortium, Unlocking the secrets of the genome, *Nature.* 459 (2009) 927–930. doi:10.1038/459927a.
- [96] R.O. Hynes, The extracellular matrix: not just pretty fibrils, *Science.* 326 (2009) 1216–9. doi:10.1126/science.1176009.
- [97] J. Pei, L.N. Kinch, N.V. Grishin, FlyXCDB—A Resource for *Drosophila* Cell Surface and Secreted Proteins and Their Extracellular Domains, *J. Mol. Biol.* 430 (2018) 3353–3411. doi:10.1016/j.jmb.2018.06.002.
- [98] R.S. Cornman, The distribution of GYR- and YLP-like motifs in *Drosophila* suggests a general role in cuticle assembly and other protein-protein interactions, *PloS One.* 5 (2010). doi:10.1371/journal.pone.0012536.
- [99] S. Plaza, H. Chanut-Delalande, I. Fernandes, P.M. Wassarman, F. Payre, From A to Z: apical structures and zona pellucida-domain proteins, *Trends Cell Biol.* 20 (2010) 524–532. doi:10.1016/j.tcb.2010.06.002.

- [100] J.H. Willis, Structural cuticular proteins from arthropods: annotation, nomenclature, and sequence characteristics in the genomics era, *Insect Biochem. Mol. Biol.* 40 (2010) 189–204. doi:10.1016/j.ibmb.2010.02.001.
- [101] R.S. Cornman, J.H. Willis, Annotation and analysis of low-complexity protein families of *Anopheles gambiae* that are associated with cuticle., *Insect Mol. Biol.* 18 (2009) 607–622. doi:10.1111/j.1365-2583.2009.00902.x.
- [102] M. Fakhouri, M. Elalayli, D. Sherling, J.D. Hall, E. Miller, X. Sun, L. Wells, E.K. LeMosy, Minor proteins and enzymes of the *Drosophila* eggshell matrix, *Dev. Biol.* 293 (2006) 127–141. doi:10.1016/j.ydbio.2006.01.028.
- [103] V.E. Alatorsev, New genes for vitelline membrane proteins in *Drosophila*, *Mol. Biol.* 40 (2006) 330–332. doi:10.1134/S002689330602021X.
- [104] A.M. Korayem, M. Fabbri, K. Takahashi, C. Scherfer, M. Lindgren, O. Schmidt, R. Ueda, M.S. Dushay, U. Theopold, A *Drosophila* salivary gland mucin is also expressed in immune tissues: evidence for a function in coagulation and the entrapment of bacteria, *Insect Biochem. Mol. Biol.* 34 (2004) 1297–1304. doi:10.1016/j.ibmb.2004.09.001.
- [105] J.-L. Da Lage, G.W.C. Thomas, M. Bonneau, V. Courtier-Orgogozo, Evolution of salivary glue genes in *Drosophila* species, *BMC Evol. Biol.* 19 (2019) 36. doi:10.1186/s12862-019-1364-9.
- [106] A.J. Rey, H. Attrill, S.J. Marygold, Using FlyBase to Find Functionally Related *Drosophila* Genes, *Methods Mol. Biol.* Clifton NJ. 1757 (2018) 493–512. doi:10.1007/978-1-4939-7737-6_16.
- [107] D. Baycin-Hizal, Y. Tian, I. Akan, E. Jacobson, D. Clark, J. Chu, K. Palter, H. Zhang, M.J. Betenbaugh, GlycoFly: a database of *Drosophila* N-linked glycoproteins identified using SPEG--MS techniques, *J. Proteome Res.* 10 (2011) 2777–2784. doi:10.1021/pr200004t.
- [108] A.O. Sessions, G. Kaushik, S. Parker, K. Raedschelders, R. Bodmer, J.E. Van Eyk, A.J. Engler, Extracellular matrix downregulation in the *Drosophila* heart preserves contractile function and improves lifespan, *Matrix Biol.* 62 (2017) 15–27. doi:10.1016/j.matbio.2016.10.008.

FIGURE LEGENDS

Figure 1. Bioinformatic workflow to define the *in-silico* matrisome of *Drosophila melanogaster*.

The databases Flybase, DIOPT, and Ensembl were interrogated with the full list of human and mouse matrisome and matrisome-associated gene symbols. Selected InterPro domains, including domains characteristic of collagens, proteoglycans, ECM-affiliated proteins, and cuticle-binding proteins (see Supplementary Table S4) were used to identify ECM-domain-containing proteins in the reference proteome. The Gene Ontology annotations related to the ECM were then used to identify previously-annotated ECM components. Finally, selected published literature using proteomic and/or bioinformatic methods, as well as reviews on the subject, were searched to identify ECM proteins not identified by the orthology-based or protein-sequencing methods. These data were combined and manually curated to generate the first complete *Drosophila* matrisome.

Figure 2. The *Drosophila* matrisome.

(A) The *Drosophila* matrisome is made up of 641 genes. Of these, 27 are homologs/orthologs to mammalian core matrisome genes, 219 are homologs/orthologs to mammalian matrisome-associated genes, and the remaining 395 are specific to *Drosophila*. These genes are then divided into either categories which we have previously defined, or the newly proposed apical matrix category.

(B) The genes that encode proteins that make up the apical matrix of *Drosophila* were further divided into classes and sub-classes.

SUPPLEMENTARY TABLE LEGENDS

Supplementary Table 1. *Drosophila* orthologs to human and mouse matrisome genes

A. and B. List of *Drosophila* genes identified by orthology to (A) human and (B) mouse matrisome genes (column A) and the matrisome category of their human or mouse orthologs (columns B-G). 783 orthologs to human matrisome genes and 720 orthologs to mouse genes were discovered.

C. List of all *Drosophila* orthologs to human or mouse matrisome genes (column A). Columns B and C indicate whether they were orthologous to human genes, mouse genes, or both. A total of 834 *Drosophila* orthologs to human and mouse matrisome genes were discovered.

Supplementary Table 2. Human and mouse genes with no *Drosophila* ortholog

A. Human genes with no *Drosophila* orthologs predicted.

B. Mouse genes with no *Drosophila* orthologs predicted.

Supplementary Table 3. UniProt reference proteome

A. The UniProt *Drosophila* reference proteome was retrieved August 10th, 2017. To allow for cross-referencing with our gene-specific list of orthologs each gene in the reference proteome was assigned a primary gene name (column E) based on the first gene name listed under column F.

B. The reference proteome was interrogated with three Gene Ontology terms, extracellular matrix (GO:0031012), extracellular region (GO:0005576), and basement membrane (GO:0005604), to obtain a set of 1308 putative matrisome proteins.

Supplementary Table 4. InterPro domains

A. List of ECM domains (column C) used to detect *Drosophila* matrisome proteins and their matrisome categories (columns A and B).

B. All *Drosophila* proteins retrieved from the UniProt *Drosophila* reference proteome and the InterPro domain families (column A) used to retrieve them.

C. Comparing the proteins identified by the protein-domain-based approach to the list of proteins previously generated by the orthology-based approach demonstrated that 49 novel proteins were identified.

Supplementary Table 5. The complete annotated *Drosophila melanogaster* matrisome

A. *Drosophila* matrisome genes are organized by matrisome division, category, and class (columns A-C). Each gene is associated with a unique FlyBase gene ID (columns D and E). Also provided are all UniProt IDs associated with the gene and a protein name pulled from the UniProt *Drosophila* reference proteome (columns F-H). InterPro domains, Gene Ontology terms, and alternative gene names were also obtained from InterPro (columns I-K, M). Genes added to the matrisome via the gene-centric, orthology-based approach have their human orthologs or homologs listed (column L).

B. The number of genes in each matrisome division and category. All combinations of classes of the apical matrix category of genes are also shown, as well as the same genes divided into their most granular (listed first in A, column C) classes or families (see also Figure 2).

Figure 1. Bioinformatic workflow to define the *in-silico* matrisome of *Drosophila melanogaster*

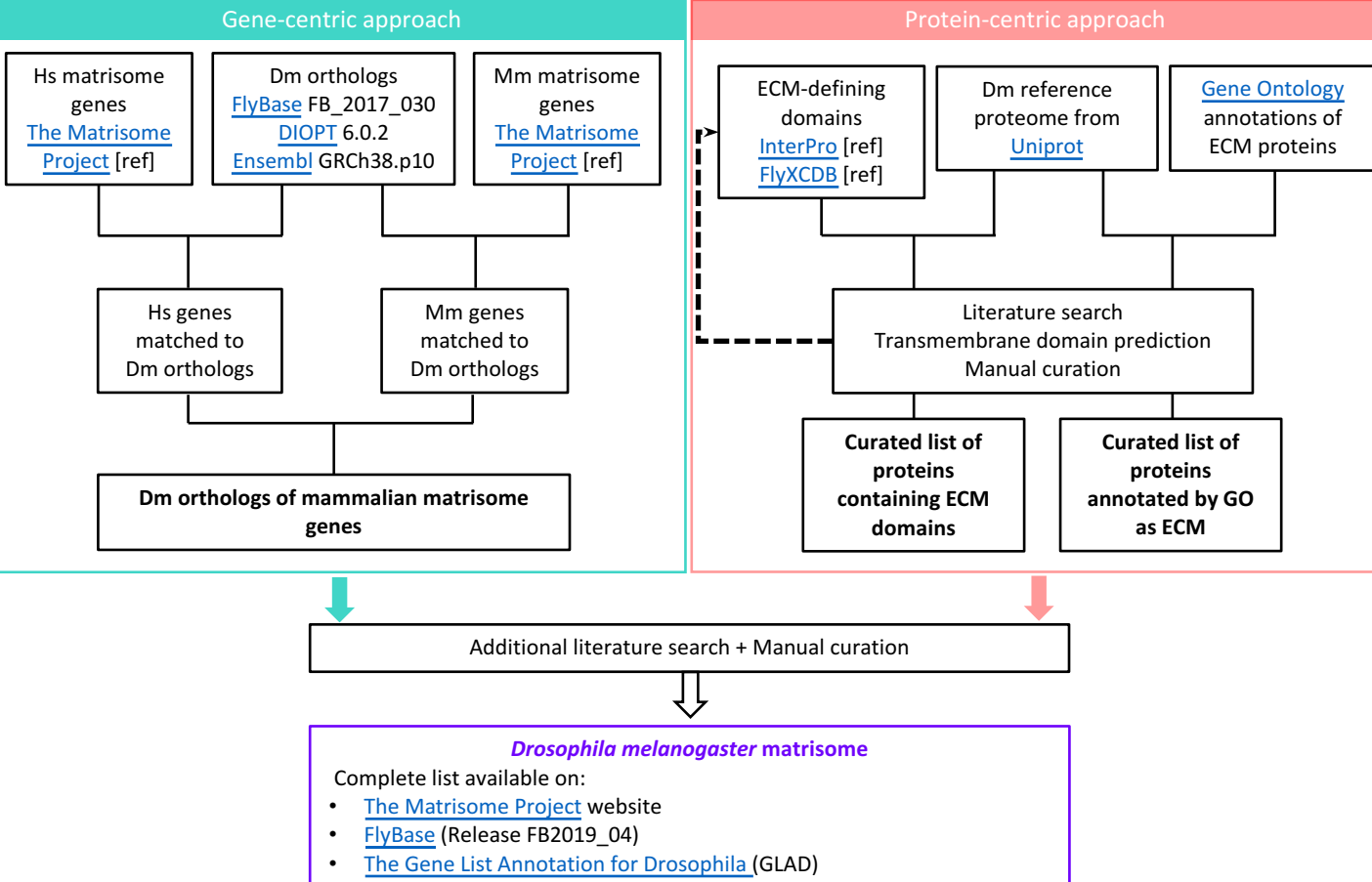


Figure 2. The *Drosophila* matrisome.

A

Matrisome Division	Matrisome Category	# of Genes
Homologs/Orthologs to Mammalian Core Matrisome Genes	Collagens	3
	ECM Glycoproteins	21
	Proteoglycans	3
	Total	27
Homologs/Orthologs to Mammalian Matrisome-Associated Genes	ECM-affiliated	83
	ECM Regulators	97
	Secreted Factors	39
	Total	219
<i>Drosophila</i> Matrisome Genes	Collagens	1
	ECM Glycoproteins	6
	ECM-affiliated	23
	ECM Regulators	1
	Secreted Factors	36
	Apical Matrix	328
	Total	395
TOTAL	641	

B

Apical Matrix Classes and Sub-classes	# of Genes
Cuticle	56
Cuticle; Chitin-binding-domain-containing Proteins	2
Cuticle; Chitin-binding-domain-containing Proteins; Chitinase	5
Cuticle; Chitinase	2
Chitinase	2
Cuticle; Chitinase-like	4
Chitinase-like	2
Cuticle; Tweedle	26
Cuticle; R&R Chitin-binding-domain-containing Proteins	79
R&R Chitin-binding-domain-containing Proteins	2
Chitin-binding-domain-containing Proteins	85
Chitin-binding-domain-containing Proteins; Chitin Deacetylase	5
Chitin Deacetylase	1
Chitin-binding-domain-containing Proteins; Mucin	9
Glue	11
Zona Pellucida-domain-containing Proteins	11
Eggshell; Vitelline Membrane	14
Eggshell; Chorion	12
TOTAL	328