In-silico definition of the *Drosophila melanogaster* matrisome

Martin N. Davis[a], **Sally Horne-Badovinac**[b] **and Alexandra Naba**[a]

***a* - Department of Physiology and Biophysics,** *University of Illinois at Chicago, 835 S. Wolcott Avenue, Chicago, IL 60612, USA*
***b* - Department of Molecular Genetics and Cell Biology,** *The University of Chicago, 920 East 58th Street, Chicago, IL 60637, USA*

Correspondence to Alexandra Naba: *anaba@uic.edu.*
https://doi.org/10.1016/j.mbplus.2019.100015

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 melanogaster* 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 our pipeline has identified 641 genes constituting the *Drosophila* matrisome. 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 (http://matrisome.org), The FlyBase (https://flybase.org/), and GLAD (https://www.flyrnai.org/tools/glad/web/).

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 –

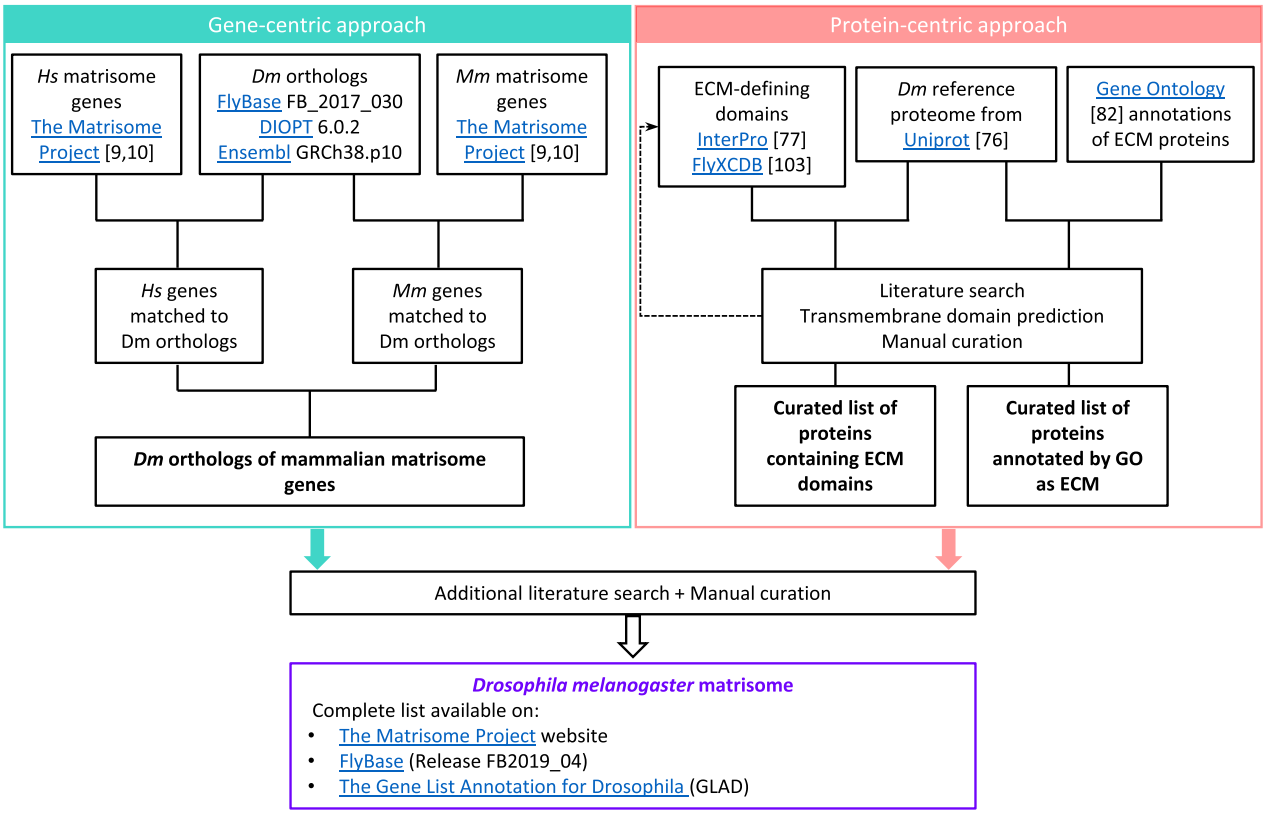

Fig. 1. Bioinformatic workflow to define the *in-silico* matrisome of *Drosophila melanogaster*. The databases FlyBase, DIOPT, and Ensembl were interrogated with the full list of human and mouse matrisome and matrisome-associated gene symbols. Selected InterPro domains, including domains characteristic of collagens, proteoglycans, ECM-affiliated proteins, and cuticle-binding proteins (see Supplementary Table 4) 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. *Hs, Homo sapiens; Mm, Mus musculus; Dm Drosophila melanogaster.*

omic datasets predictive, for example, of cancer patient outcome [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 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 (Fig. 1) and identified 641 genes that we propose to 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.

In-silico definition of the *Drosophila melanogaster* matrisome

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 sequence alignment tools built into Flybase (FB2017_03, released June 2017) [71], the ***D**rosophila* RNAi Screening Center's **I**ntegrative **O**rtholog **P**rediction **T**ool (DIOPT, Version 6.0.2, released June 2017) [73,74], 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 previously (Fig. 1) [11]. Although these sequence-alignment meta-algorithms return a confidence score based on the number of algorithms returning a hit, we did not eliminate genes with low-confidence scores at this stage to maximize our potential to identify relevant genes.

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).

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 [77] 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) [78]. Although three ECM domains were initially used to search the UniProt *Drosophila* reference proteome, the domain chitin-binding type R&R consensus (IPR031311), a conserved motif of 35–36 amino acids identified by Rebers and Riddiford (R&R) [79,80], 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 [81]. 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

A.

Matrisome Category	# of Genes
Collagens	4
ECM Glycoproteins	27
Proteoglycans	3
ECM-affiliated	106
ECM Regulators	98
Secreted Factors	75
Apical Matrix	328
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***

Fig. 2. The *Drosophila* matrisome. (A) The *Drosophila* matrisome comprises 641 genes. 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.

49 of the 140 proteins discovered by domains characteristic of ECM proteins (35%) were not previously identified using the orthology approach (Supplementary Table 4C).

Gene-Ontology-based approach to identify additional Drosophila matrisome proteins

The *Drosophila* proteome retrieved from UniProt is also annotated with Gene Ontology (GO; http://geneontology.org) – Cellular Component terms describing the intra- and extracellular localization of proteins [82,83]. 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).

Manual curation of potential Drosophila matrisome genes

The three computational approaches described above identified 1585 genes encoding potential *Drosophila* matrisome proteins. As we reported for other organisms, a purely computational approach is not sufficient to identify ECM genes [9,10,21]. For example, examination of the list revealed the presence of proteins that share structural features with ECM proteins such as transmembrane receptors, anti-microbial peptides or accessory gland proteins, but that are clearly not ECM components. We thus undertook a knowledge-based approach to manually curate the list and either eliminate non-ECM genes or include ECM genes that had been missed by our computational approach. 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 [84]. 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. This step alone excluded 303 genes. We further consulted extensively the literature on the *Drosophila* ECM, and made direct queries of FlyBase [71] and UniProt [76] to examine closely the protein sequences and orthology relationships to mammalian ECM proteins of genes and decide whether there was sufficient evidence to classify these genes as part of the matrisome. 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 for which experimental evidence does not support their classification as matrisome components (see below).

The Drosophila matrisome is composed of 641 genes

Based on the combined result of these analyses, we propose that the *Drosophila* matrisome is composed 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 [22,85]. Below, we describe how these 641 genes have been classified into matrisome categories based on their structure, localization, and/or function.

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 Fig. 2A). Only 1 collagen (pericardin) [86,87] and 6 glycoproteins (artichoke [88], anachronism [89], Defense protein l (2)34Fc, glutactin [90], tiggrin [91], and tenectin [92,93]) were not homologous or orthologous to mammalian core matrisome genes. Since the expansion and complexification of the ECM emerged with the appearance of deuterostomes, a large number of ECM genes or families of genes were not found in *Drosophila*, including fibronectin, matricellular proteins such as thrombospondins and tenascins, and non-basement membrane collagens [23,94–96].

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 Tables 5A and B and Fig. 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 [97]. Our study also identified a total of 24 prolyl-4-hydroxylases (P4Hs). Prolyl-4-hydroxylases catalyze the formation of hydroxyprolines [24,98–100]. 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 [100] and interrogation of The National Human Genome Research Institute **mod**el organism **ENC**yclopedia **O**f **D**NA **E**lement (modENCODE) database [101] 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 [102]. 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 http://prodata.swmed.edu/FlyXCDB/info.list.new21_26.html, [103]). 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).

Classification of Drosophila genes with no mammalian orthologs or homologs

Since the chitin-based ECMs, 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 Fig. 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 (Supplementary Table 5, column C) to reflect their respective proteins' domain structure, enzymatic function, or localization. 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) [104]. 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 [105]. 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 [106,107]. The eggshell superfamily includes two protein classes, corresponding to the vitelline membrane and chorion layers of this ECM, respectively [68,108,109]. 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 [108]. Finally, 11 proteins including new-glue and salivary glue secretion proteins, Eig71Ee [69,110], and the newly identified tandem paralog of Sgs5 (FBgn0038523) [111] were classified as glue proteins.

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

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 FlyBase (FB2019_04, accessible at: https://flybase.org), which is the most comprehensive source of genetic information for this model organism [71,72,112]. 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, accessible at: https://www.flyrnai.org/tools/glad/web/) 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.

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 [113]. 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 [114]. 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.

Conclusion

We propose here an *in-silico* definition of *Drosophila melanogaster* matrisome that comprises 641 genes encoding ECM and ECM-associated proteins. We further propose their comprehensive classification according to structural and/or functional features. Of note, and as it has been the case with the human and mouse matrisomes, this list is meant to evolve as we gain knowledge in the functions of these genes. 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.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.mbplus.2019.100015.

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 lab (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 (C-088), and by grants from the American Cancer Society (RSG-14-176) and National Institutes of Health (R01-GM126047) to SHB.

Received 5 August 2019;
Received in revised form 4 September 2019;
Accepted 5 September 2019
Available online 30 October 2019

Keywords:
Extracellular matrix;
Basement membrane;
Apical ECM;
Cuticle;
Eggshell;
Glue;
Protein domains

References

[1] R.O. Hynes, K.M. Yamada, Extracellular Matrix Biology, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, Cold Spring Harbor Perspectives in Biology, 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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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) https://doi.org/10.1074/mcp.M111.014647(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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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, Hepatol. Baltim. Md. 65 (2017) 969–982, https://doi.org/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) https://doi.org/10.1152/ajpheart.00712.2016 (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. https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/10.1016/bs.ctm.2015.07.001.
[28] R. Jayadev, D.R. Sherwood, Basement membranes, Curr. Biol. CB. 27 (2017) R207–R211, https://doi.org/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, https://doi.org/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) https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/10.1242/dev.044263.

[45] N. Denef, 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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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) https://doi.org/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, https://doi.org/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, https://doi.org/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) https://doi.org/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, https://doi.org/10.1016/j.matbio.2018.07.004.

[56] M. Kiss, A.A. Kiss, M. Radics, N. Popovics, E. Hermesz, K. Csiszár, M. Mink, Drosophila type IV collagen mutation associates with immune system activation and intestinal dysfunction, Matrix Biol. 49 (2016) 120–131, https://doi.org/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. https://doi.org/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. https://doi.org/10.1371/journal.pgen.1004683.

[59] B. Arefin, L. Kucerova, P. Dobes, R. Markus, H. Strnad, Z. Wang, P. Hyrsl, 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, https://doi.org/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 https://doi.org/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) https://doi.org/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, https://doi.org/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, https://doi.org/10.1002/dvdy.24356.

[64] B. Lemaitre, I. Miguel-Aliaga, The digestive tract of Drosophila melanogaster, Annu. Rev. Genet. 47 (2013) 377–404, https://doi.org/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, https://doi.org/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, https://doi.org/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. Beňo, R. Farkas, Fine infrastructure of released and solidified Drosophila larval salivary secretory glue and salivary gland ducts using SEM, Bioinspir. Biomim. (2019) https://doi.org/10.1088/1748-3190/ab2b2b.

[70] N. Casas-Vila, A. Bluhm, S. Sayols, N. Dinges, M. Dejung, T. Altenhein, D. Kappei, B. Altenhein, J.-Y. Roignant, F. Butter, The developmental proteome of Drosophila melanogaster, Genome Res. 27 (2017) 1273–1285, https://doi.org/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, N. Perrimon, S.R. Gelbart, J. Agapite, K. Broll, L. Crosby, G. dos Santos, D. Emmert, L.S. Gramates, K. Falls, V. Jenkins, B. Matthews, C. Sutherland, C. Tabone, P. Zhou, M. Zytkovicz, N. Brown, G. Antonazzo, H. Attrill, P. Garapati, A. Holmes, A. Larkin, S. Marygold, G. Millburn, C. Pilgrim, V. Trovisco, P. Urbano, T. Kaufman, B. Calvi, B. Czoch, J. Goodman, V. Strelets, J. Thurmond, R. Cripps, P. Baker, FlyBase 2.0: the next generation, Nucl. Acids Res. 47 (2019) D759–D765, https://doi.org/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, https://doi.org/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, https://doi.org/10.7150/jgen.12863.

[74] Y. Hu, I. Flockhart, A. Vinayagam, C. Bergwitz, B. Berger, N. Perrimon, S.E. Mohr, An integrative approach to ortholog prediction for disease-focused and other functional studies, BMC Bioinformatics. 12 (2011) 357, https://doi.org/10.1186/1471-2105-12-357.

[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, https://doi.org/10.1093/nar/gkx1098.

[76] The UniProt Consortium, UniProt: the universal protein knowledgebase, Nucleic Acids Res. 45 (2017) D158–D169, https://doi.org/10.1093/nar/gkw1099.

[77] A.L. Mitchell, T.K. Attwood, P.C. Babbitt, M. Blum, P. Bork, A. Bridge, S.D. Brown, H.-Y. Chang, S. El-Gebali, M.I. Fraser, J. Gough, D.R. Haft, H. Huang, I. Letunic, R. Lopez, A. Luciani, F. Madeira, A. Marchler-Bauer, H. Mi, D.A. Natale, M. Necci, G. Nuka, C. Orengo, A.P. Pandurangan, T. Paysan-Lafosse, S. Pesseat, S.C. Potter, M.A. Qureshi, N.D. Rawlings, N. Redaschi, L.J. Richardson, C. Rivoire, G. A. Salazar, A. Sangrador-Vegas, C.J.A. Sigrist, I. Sillitoe, G. G. Sutton, N. Thanki, P.D. Thomas, S.C.E. Tosatto, S.-Y. Yong, R.D. Finn, InterPro in 2019: improving coverage, classification and access to protein sequence annotations, Nucleic Acids Res. 47 (2019) D351–D360, https://doi.org/10.1093/nar/gky1100.

[78] 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, https://doi.org/10.1016/j.ibmb.2007.03.007.

[79] J.E. Rebers, L.M. Riddiford, Structure and expression of a Manduca sexta larval cuticle gene homologous to Drosophila cuticle genes, J. Mol. Biol. 203 (1988) 411–423, https://doi.org/10.1016/0022-2836(88)90009-5.

[80] J.E. Rebers, J.H. Willis, A conserved domain in arthropod cuticular proteins binds chitin, Insect Biochem. Mol. Biol. 31

(2001) 1083–1093, https://doi.org/10.1016/S0965-1748(01)00056-X.
[81] 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, https://doi.org/10.1186/1471-2105-5-138.
[82] 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, https://doi.org/10.1038/75556.
[83] The Gene Ontology Consortium, Expansion of the Gene Ontology knowledgebase and resources, Nucleic Acids Res. 45 (2017) D331–D338, https://doi.org/10.1093/nar/gkw1108.
[84] 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–W432, https://doi.org/10.1093/nar/gkm256.
[85] 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, https://doi.org/10.1038/nature25473.
[86] 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.
[87] 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) https://doi.org/10.1242/bio.030361.
[88] 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, https://doi.org/10.1534/genetics.113.156323.
[89] 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.
[90] 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.
[91] 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.
[92] 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, https://doi.org/10.1016/j.modgep.2006.01.007.
[93] 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, https://doi.org/10.1016/j.ydbio.2010.02.008.
[94] R.O. Hynes, The evolution of metazoan extracellular matrix, J. Cell Biol. 196 (2012) 671–679, https://doi.org/10.1083/jcb.201109041.
[95] D.F. Mosher, J.C. Adams, Adhesion-modulating/matricellular ECM protein families: a structural, functional and evolutionary appraisal, Matrix Biol. 31 (2012) 155–161, https://doi.org/10.1016/j.matbio.2012.01.003.
[96] 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, https://doi.org/10.1091/mbc.E10-03-0251.
[97] 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, https://doi.org/10.1126/science.1217817.
[98] J. Myllyharju, Prolyl 4-hydroxylases, the key enzymes of collagen biosynthesis, Matrix Biol. 22 (2003) 15–24.
[99] 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, https://doi.org/10.1007/BF00202644.
[100] 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.
[101] 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, https://doi.org/10.1038/459927a.
[102] R.O. Hynes, The extracellular matrix: not just pretty fibrils, Science. 326 (2009) 1216–1219, https://doi.org/10.1126/science.1176009.
[103] 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, https://doi.org/10.1016/j.jmb.2018.06.002.
[104] 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) https://doi.org/10.1371/journal.pone.0012536.
[105] 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, https://doi.org/10.1016/j.tcb.2010.06.002.
[106] 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, https://doi.org/10.1016/j.ibmb.2010.02.001.
[107] 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, https://doi.org/10.1111/j.1365-2583.2009.00902.x.
[108] 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, https://doi.org/10.1016/j.ydbio.2006.01.028.
[109] V.E. Alatortsev, New genes for vitelline membrane proteins in Drosophila, Mol. Biol. 40 (2006) 330–332, https://doi.org/10.1134/S002689330602021X.
[110] 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, https://doi.org/10.1016/j.ibmb.2004.09.001.

[111] 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, https://doi.org/10.1186/s12862-019-1364-9.

[112] 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, https://doi.org/10.1007/978-1-4939-7737-6_16.

[113] 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, https://doi.org/10.1021/pr200004t.

[114] 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, https://doi.org/10.1016/j.matbio.2016.10.008.