



**HAL**  
open science

# Small ORFs, Big Insights: Drosophila as a Model to Unraveling Microprotein Functions

Hélène Chanut-Delalande, Jennifer Zanet

► **To cite this version:**

Hélène Chanut-Delalande, Jennifer Zanet. Small ORFs, Big Insights: Drosophila as a Model to Unraveling Microprotein Functions. *Cells*, 2024, 13 (19), pp.1645. 10.3390/cells13191645. hal-04748777

**HAL Id: hal-04748777**

**<https://hal.science/hal-04748777v1>**

Submitted on 22 Oct 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Review

# Small ORFs, Big Insights: *Drosophila* as a Model to Unraveling Microprotein Functions

Hélène Chanut-Delalande  and Jennifer Zanet \* 

Unité de Biologie Moléculaire, Cellulaire et du Développement (MCD), UMR 5077, Centre de Biologie Intégrative (CBI), CNRS, UPS, Université de Toulouse, 31062 Toulouse, France; helene.chanut@univ-tlse3.fr

\* Correspondence: jennifer.zanet@univ-tlse3.fr

**Abstract:** Recently developed experimental and computational approaches to identify putative coding small ORFs (smORFs) in genomes have revealed thousands of smORFs localized within coding and non-coding RNAs. They can be translated into smORF peptides or microproteins, which are defined as less than 100 amino acids in length. The identification of such a large number of potential biological regulators represents a major challenge, notably for elucidating the in vivo functions of these microproteins. Since the emergence of this field, *Drosophila* has proved to be a valuable model for studying the biological functions of microproteins in vivo. In this review, we outline how the smORF field emerged and the nomenclature used in this domain. We summarize the technical challenges associated with identifying putative coding smORFs in the genome and the relevant translated microproteins. Finally, recent findings on one of the best studied smORF peptides, Pri, and other microproteins studied so far in *Drosophila* are described. These studies highlight the diverse roles that microproteins can fulfil in the regulation of various molecular targets involved in distinct cellular processes during animal development and physiology. Given the recent emergence of the microprotein field and the associated discoveries, the microproteome represents an exquisite source of potentially bioactive molecules, whose in vivo biological functions can be explored in the *Drosophila* model.

**Keywords:** small ORF; smORF; peptides; microproteins; *pri*; *tal*; *Drosophila*; development



**Citation:** Chanut-Delalande, H.; Zanet, J. Small ORFs, Big Insights: *Drosophila* as a Model to Unraveling Microprotein Functions. *Cells* **2024**, *13*, 1645. <https://doi.org/10.3390/cells13191645>

Academic Editor: Kweon Yu

Received: 13 September 2024

Revised: 27 September 2024

Accepted: 2 October 2024

Published: 2 October 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

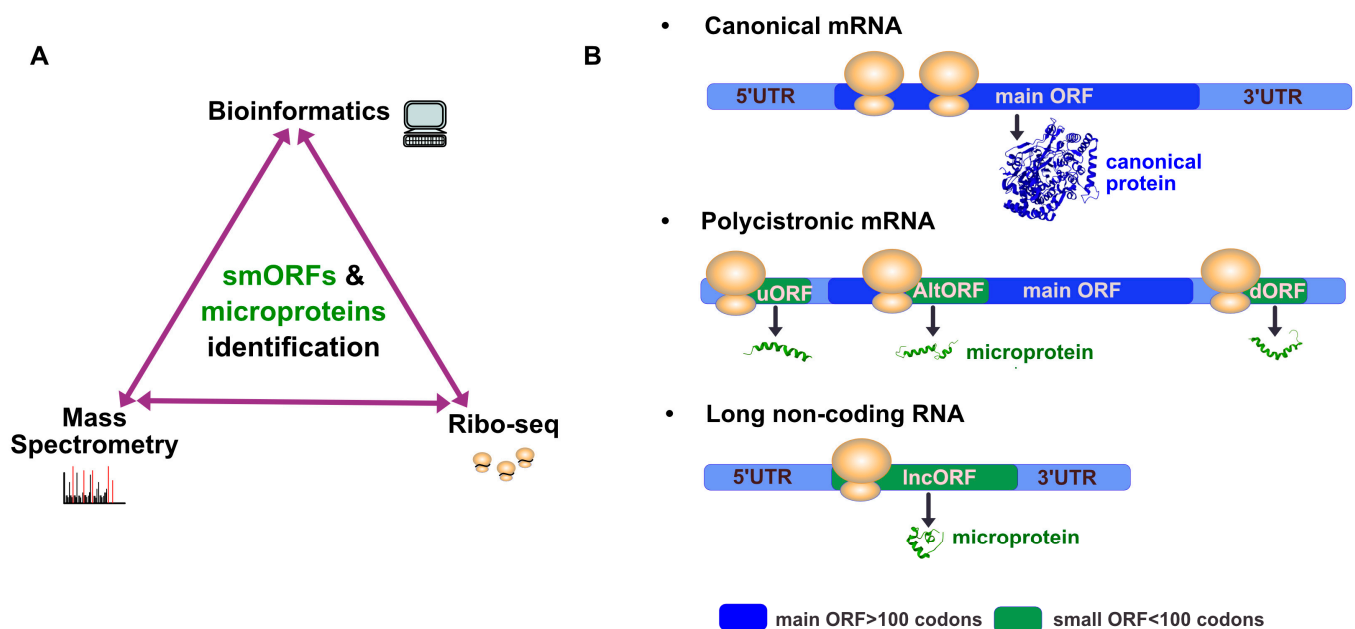
## 1. What Are smORFs and Microproteins?

When genomes were sequenced, specific criteria were established to accurately annotate genes. To prevent the annotation of millions of Open Reading Frames (ORFs), i.e., DNA sequences starting with an ATG and ending with a stop codon, an arbitrary lower limit of 100 codons was set to define an ORF [1–3]. As a result, most small ORF (smORFs), also known as short ORFs (sORFs) or non-canonical ORFs (ncORFs), that do not exceed 100 codons, were not annotated [4]. However, the discovery in the *Drosophila* genome of small ORFs, named *pri*, which were initially annotated as non-coding, are capable of being translated into peptides only 11 amino acids (aa) long, and have crucial developmental roles [5–7], has entirely blurred the frontier between coding and non-coding genomes. This revealed that other as-yet-unannotated smORFs could also be coding and translated, thus pioneering the field of microproteins. Since then, the function of several microproteins has been studied in various organisms, including *Drosophila*, plants, yeast, bacteria, and human cells [8–15]. Microproteins refer to polypeptides translated from small ORFs and come with different names in the literature: micropeptides, microproteins, smORF peptides, SEPs (small ORF encoded peptides), AltORF (Alternative ORF), and small proteins. In addition, the small ORF itself is called ncORF for non-canonical ORF, as opposed to the canonical ORF defined as having a size greater than 100 codons, present in the mRNA and coding for a so-called canonical protein. As the terminology distinguishing a peptide from a protein is based on its length, with the threshold of 50 aa, we now term any product translated from a

smORF (defined as less than 100 codons) a microprotein, which thus includes all polypeptides shorter than 100 aa. Therefore, smORF refers to the DNA/RNA nucleotide sequence, and microprotein to the truly translated polypeptide. This new family of microproteins is now known as the microproteome, smORFome, or non-canonical proteome.

## 2. How to Identify smORFs and Microproteins?

Different approaches have been developed to identify, from millions of potential smORFs, the ones having a true coding potential (Figure 1A). Among them, adapted computational approaches have been crucial to predict putative coding smORFs. The first bioinformatic analyses estimated the existence of several hundreds of smORFs in the *Drosophila* genome [16,17]. However, recently, computational methods progressively integrating machine learning improved their prediction. For instance, MiPepid detects smORFs in the human genome following training on a dataset of high confidence of already proved coding smORF sequences [18], and DeepCPP recognizes nucleotide patterns around start codons, named nucleotide bias, to detect smORF in vertebrate and invertebrate genomes [19]. However, these methods predict different numbers and sets of microproteins, and therefore need further experimental validation.



**Figure 1.** Methods of small ORF identification and their localization in RNA molecules. (A) The putative smORFs present in the genome can be either predicted by dedicated bioinformatic pipelines or identified by Ribo-seq data analysis. Improvements in mass spectrometry approaches allow the identification of novel microproteins. Combining these methodologies enables the identification of smORFs and the microproteins potentially encoded in the genome. (B) The mRNA molecule was classically viewed as monocistronic and encoding one canonical protein from its main ORF (in blue). SmORFs, defined as being under 100 codons (in green), are mainly found in coding RNAs (mRNA), which makes them polycistronic, and non-coding RNAs such as long non-coding RNA (lncORF). In mRNA, smORFs can be nested in the 5'UTR (uORF) and 3'UTR (dORF), and within the main ORF in an alternative frame (AltORF).

To overcome the limitations of the in silico approach and possibly the resulting false-positive prediction of microproteins, an experimental approach, i.e., ribosome profiling (Ribo-seq), can be used to detect precisely the actively translated RNA sequences [20]. This technique has revolutionized the field of microproteins, providing experimental proof of smORF translation. This method consists of isolating and sequencing Ribosome-Protected RNA Fragments (RPFs), corresponding to RNA sequences in the course of

translation, from any biological sample, thus providing an exhaustive view of the translational landscape [20–22]. This approach unveiled that smORFs are not only present in some non-coding RNAs (ncRNAs), but also in some mRNAs, mostly in 5'UTR but also in 3'UTR and within the main—or canonical—ORF in an alternative frame [23,24] (Figure 1B). SmORFs have thus been classified and named according to the type of RNA molecules on which they are localized (coding or non-coding RNA) and their localization on the mRNA molecules (Figure 1) [25]. Upstream ORFs (uORFs), present in the 5'UTR, are found in most eucaryotic mRNAs and regulate the translation of the main downstream ORF, through their RNA sequence itself, which can retain ribosomes to control their access to the main ORF [26]. Thus, uORFs can be translated and represent an important source of microproteins [27,28]. Functional studies on some of them have shown that they can interact with and regulate canonical proteins, either translated from the main ORF found downstream of the uORF [29] or with another canonical protein encoded by a different mRNA [28,30]. Although experimental evidences showed that smORFs present in the 3'UTR (dORFs) can be translated into microproteins, functional data are still scarce, so their role is not yet well understood [27,31]. Besides uORF, lncORF microproteins encoded from long non-coding RNAs (lncRNAs) are a major source of microproteins since 30% of transcribed lncRNAs appear to be translated in human cells and in *Drosophila* [32,33]. Notably, microproteins can also be translated from smORFs present in circular RNAs [34] or pre-miRNAs [35], highlighting that smORFs are hidden throughout the genome. To complicate things a little further, Ribo-seq method revealed non-AUG translation start sites when combined with different drugs (harringtonine or lactimidomycin) that block and accumulate ribosomes at initiation sites [27], thus substantially increasing the repertoire of putative microproteins. Finally, Ribo-seq reveals that translation is pervasive and increases with aging, carcinogenesis, or neurodegeneration [36–38]. However, physiological consequences of the translation landscape reshaping and the functions of resulting microproteins still remain unknown. Collectively, Ribo-seq data, mainly from mammalian cells, uncover the existence of thousands of microproteins whose functions are largely elusive.

In *Drosophila melanogaster*, Ribo-seq experiments dedicated to the identification of smORFs have been adapted to sequence RPFs only from the polysomal fraction (Poly-Ribo-seq), i.e., the fraction containing RNA molecules bound to at least two ribosomes, providing evidence of their translation [32,39–41].

Although Ribo-seq appears to be a powerful tool for discovering smORFs, these experiments have limitations. Indeed, the experimental protocol, which is not trivial to set up, requires further improvements for each type of sample and organism to obtain qualitative data [42]. Sequencing depth must be very high, making experimentation expensive, and bioinformatics analyses require the use of different Ribo-seq analysis tools to reliably identify smORFs, which are actually translated [43]. Despite the prominent impact of Ribo-seq on the field of smORF research, the evidence it provides for translation does not prove that a microprotein is produced.

Mass spectrometry (M/S) is the gold standard method for confirming the existence of microproteins. However, due to their low level of expression and their small size, the detection of a large proportion of translated peptides remains difficult [44]. A peptidomic-based strategy thus appears more adapted to the discovery of microproteins since, unlike proteomics, this approach does not require enzymatic digestion and allows the detection of endogenous peptides. Dedicated methods for the extraction and the enrichment of small proteins from samples, combined with the use of a state-of-the-art mass spectrometer and adapted bioinformatics workflows, largely contribute to the identification of novel microproteins [45]. Indeed, the combination of these experimental methods with a custom database, whose peptide sequences are predicted from transcriptomic and Ribo-seq data, allowed the discovery of unannotated microproteins in *Drosophila* [45,46]. Recent peptidomic studies in *Drosophila*, human cell lines, or mice have led to the discovery of several hundred microproteins, which may constitute a set of candidates to be further functionally tested in vivo [8,27,46–48].

To date, the exhaustive number of microproteins in any organisms, whose estimation varies from several hundreds to several thousands, is highly speculative and appears to depend on the method of identification, physiological conditions, and stages of development [33]. Several databases, such as OpentProt [49], SmProt [50], MetamORF [51], and sORFs.org [52], use published experimental and computational data to constitute a bank of predicted microproteins, whose estimated number is highly variable. For instance, SmProt predicts tens of thousands microproteins encoded by the human or the *Drosophila* genomes [50,53]. Recent experimental data and computational analyses allowed the annotation of more than 1000 microproteins in the *Drosophila melanogaster* genome [54]. A large number of studies on human cells or tissues using Ribo-seq and mass spectrometry analyses enabled the annotation of roughly 7000 microproteins in the human genome [25].

The progress of methods to detect microproteins and the attractiveness of this new field should lead to a growing number of confirmed translated microproteins in all organisms. This prospect is incredibly exciting, given the seemingly infinite potential for bioactive peptide discovery [55,56]. Since microproteins have been found recently to play a major role in cancer development [30,57–59], they actually benefit from a particular attention in oncogenesis research due to their potential as therapeutic agents or targets for cancer treatment. However, understanding the function of all these microproteins constitutes a bottleneck and different strategies are implemented to meet this challenge. In the following chapter, we will focus on studies carried out in *Drosophila* to illustrate the functions of the microproteins in vivo that have been studied to date.

### 3. Functions of Microproteins in *Drosophila*

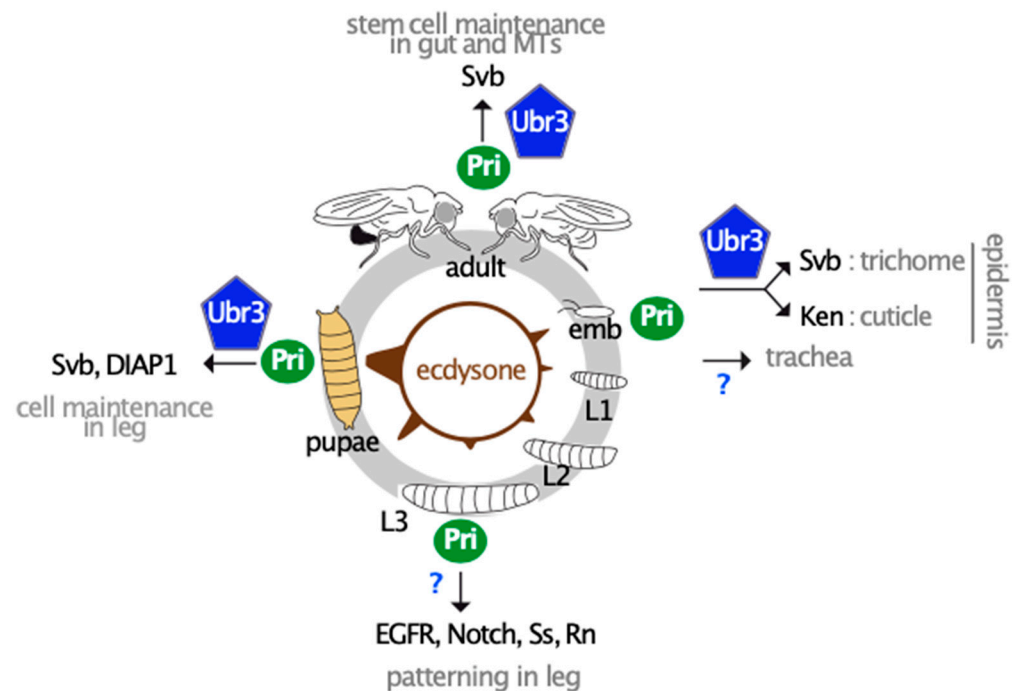
The Pri (Polished rice) peptide, also known as Tal (Tarsal-less) [6,7] or Mlpt (Mille-pattes) in *Tribolium* [5], strongly contributed to the emergence of the microprotein field, demonstrating that smORFs as short as 36 nucleotides could be translated. Pri has been the most extensively studied microprotein, not only in *Drosophila* but also in other arthropods. In this review, we will first depict the various functions of Pri in *Drosophila* and other arthropods and then report on the roles of other microproteins in *Drosophila*.

#### 3.1. pri Peptides as the Pioneers of the Microprotein Field

The *pri* gene was initially identified due to its remarkable dynamic expression pattern during embryonic development [60] and the strong developmental defects provoked by its genetic deletion [6,7]. Pri peptides, 11–32 aa long, are produced from a 1.5 kb long polycistronic mRNA containing four smORFs. The four Pri peptides share the LDPTGQY sequence with no predicted structure, which is conserved among insects and found in crustacean species [6,7]. Hemipteran species additionally contain a 80 aa long peptide, comprising a conserved sequence of 14 aa, but different from LDPTGQY, and whose role is not yet understood [61].

The *pri* gene is highly expressed in specific tissues at developmental transitions (Figure 2), supposing that Pri peptides are produced in large quantities at specific time points. The small size of Pri peptides suggests their fast translation, which favors their rapid function following their expression. Together, these data led researchers to propose that Pri peptides are produced rapidly and in large amounts in a given tissue to trigger their function in a very specific spatiotemporal scale. In addition, genetic experiments performed in epidermal tissues showed a cell non-autonomous function of Pri peptides [7,62–64]. Given their small size, Pri peptides could diffuse passively to neighboring cells through cytoplasmic bridges [65,66]. Although the function of this diffusive property is not yet understood, it might serve to equally distribute peptides in the tissue allowing their synchronous role with adjacent cells.





**Figure 2.** Pri peptides play pleiotropic roles throughout *Drosophila* development. Pri peptides mediate the function of ecdysone throughout the *Drosophila* life cycle to control the timing of morphogenetic processes in diverse tissues but also to protect cells from apoptosis. They execute several functions simultaneously in the same tissue (as shown in embryonic epidermis) but also in two or more tissues at the same time (in epidermis and trachea during embryogenesis). Understanding the molecular functions of Pri peptides allows the comprehension of how they fulfill their pleiotropic roles. Multiple functions of Pri peptides require their partner Ubr3, and together they modify the activity of the Svb and Ken TFs but also DIAP1. However, studies on leg patterning indicate that Pri peptides induce changes in the activity of several signaling factors and TFs (EGFR, Notch, Spineless, and Rotund) independently of Ubr3, suggesting that additional functions remain to be elucidated.

### 3.1.1. Pri Displays Pleiotropic Functions in Diverse Tissues at the Same Time

Various functions of *pri* have been described in insects revealing its pleiotropic roles throughout animal development. In *Tribolium*, *mlpt* has been initially defined as a segmentation gene [5]. In *Drosophila*, *pri* is expressed widely along development as exemplified by its very dynamic expression pattern during embryogenesis in the tracheal system, gut, and epidermis, and also in the adult leg and wing [6,7,67]. These tissues are strongly affected when *pri* function is deregulated, for example, epidermal cells deficient for *pri* function lack cytoplasmic apical structures called trichomes [6,7].

### 3.1.2. Pri Peptides Belong to a Tripartite Pri/Ubr3/Svb Module Controlling Epidermal Differentiation

When the role of *pri* in trichome formation was discovered, the transcription factor (TF) Shavenbaby (Svb) was the most downstream player known to control this process [68]. Svb directly triggers the expression of effector genes that participate in the building and the maintenance of the trichome [69–71]. Deciphering the molecular interaction between Svb and Pri peptides in the control of trichome formation greatly contributed to the understanding of the molecular mode of action of Pri peptides [72,73]. The combination of in vivo analyses and biochemical experiments in *Drosophila* S2 cells (genome wide screen) brought a precise understanding of the molecular function of Pri peptides during epidermal differentiation [72,73]. Pri peptides interacts with the E3 ubiquitin ligase Ubr3, a large molecule which is ubiquitously expressed in the organism and triggers the ubiquitination of specific proteins to allow their degradation via the proteasome [74]. The binding of Pri peptides to Ubr3 modifies the repertoire of its targets and Ubr3 acquires the potential to recognize

additional targets, including Svb [73]. The Svb TF is initially produced as a long protein that represses the transcription of its target genes. Its N-terminal part contains a degron domain, i.e., a sequence that address the protein to the proteasome. This degron is recognized by the Pri/Ubr3 complex and ubiquitinated by Ubr3. This post-translational modification leads to the partial degradation of Svb through the proteasome and the resulting processed form, which has lost its repressor domain, activates the transcription of effector genes [72,73]. Therefore, the Pri/Ubr3 complex controls the conversion of Svb from a repressor into an activator of transcription. Adult epidermal cells are similarly producing trichomes using the same molecular mechanism showing a reiteration of the Pri/Svb/Ubr3 module [73]. Overall, these molecular studies revealed how very small microproteins can modify the activity of a larger protein, here Ubr3, with ubiquitous roles, to enlarge the spectrum of its biological functions for specific roles during development [73].

The processing of Svb to trigger trichome formation induced through a two-step process raised the question of the role of such a sophisticated mechanism controlled by the Pri peptides. Understanding how *pri* expression is regulated brought a larger vision of the function of Pri peptides. Based on a genetic screen, the expression of *pri* was shown to be directly regulated by the steroid hormone ecdysone, which controls the timing of important developmental transitions and metamorphosis during insect development [63,75]. The receptor of ecdysone, EcR, directly binds to the genomic region of *pri* composed of a large number of enhancers spread over at least 50 kb, each of them carrying specific spatiotemporal activities [75]. Through its cis-regulatory sequences, *pri* integrates the systemic signal of ecdysone and converts it into specific responses in tissues. *pri* responds to ecdysone in a large variety of cell types, illustrating its function as a wide responsive gene of ecdysone [63,75,76]. While other responsive genes of ecdysone encode nuclear receptors regulating a temporal cascade of transcriptional networks, the Pri peptides offer an alternative mode of action of ecdysone in tissues, directly regulating the processes of differentiation. As Pri peptides modify the activity of a large ubiquitous enzyme, their mechanistic role might facilitate the acquisition of novel functions of ecdysone.

### 3.1.3. The Functions of the Pri/Ubr3/Svb Module Are Conserved across Insect Species

The *pri* gene is present in the large phylum of arthropods [5,6] and functional studies on Pri peptides have now been carried out on a broad array of insect species, i.e., the coleoptera *Tribolium castaneum*, the hymenoptera *Nasonia vitripennis*, and three hemipteran species *Gerris buenoi*, *Oncopeltus fasciatus* [77], and *Rhodnius prolixus* [61]. These studies established that the roles of Pri peptides are conserved in epidermal derivatives and leg formation, but also in the segmentation of insects. These roles are shared with the E3 ubiquitin ligase Ubr3 and the TF Svb, highlighting the conservation of the functional Pri/Ubr3/Svb module and its roles in embryonic/postembryonic development of insects [77]. However, the *Drosophila melanogaster* dipteran species diverges from other species since the function of the Pri/Ubr3/Svb module in segmentation has been lost due to the restriction of *svb* expression in early *Drosophila* embryos [77]. A similar absence of *pri* function in segmentation has been described for another dipteran species, *Clogmia albipunctata*, which might also relate to *svb* expression changes [78].

### 3.1.4. Pri Peptides Display Pleiotropic Functions Throughout Tissue Differentiation

Detailed analyses of the functional role of *pri* in the respiratory system, embryonic epidermis, and leg development sustained that Pri peptides play a variety of roles throughout tissue development, using different molecular partners, which are partially identified. In the respiratory system, *pri* is expressed throughout the formation of the tubular network [6,7]. In this organ, Pri peptides control both the fate decision of cells involved in the formation of specific branches and morphogenesis of tubes through actin cytoskeleton remodeling, and deposition of the extracellular matrix in the lumen of the tubes [7,76,79]. However, Pri molecular partners involved in the tracheal system have not yet been identified. In embryonic epidermis, *pri* plays an additional role to trichome formation since the

cuticle, which is the exoskeleton that entirely covers the animal, is dramatically affected in the absence of *pri* [80]. Under the control of ecdysone, Pri peptides regulate the production of a proper embryonic cuticle by inhibiting the expression of a large group of cuticle genes, which are normally expressed at later stages [80]. This function is executed by Pri peptides and Ubr3, which together control the activity of the Ken transcription factor. Therefore, in embryonic epidermis, Pri peptides, with their Ubr3 partner, coordinate two differentiation programs governed by two distinct transcription factors, Svb and Ken, for proper epidermis function [80]. Finally, during leg development, *pri* is expressed at two specific time points in a spatially restricted manner corresponding to two main functions during leg patterning and later on for cell survival [64]. While Pri peptides act on multiple factors independently of Ubr3 during leg patterning at larval stage, i.e., tarsal transcriptional programs (Rotund and Spineless TFs) and Notch and EGFR signaling pathways [62,64], Pri peptides and Ubr3 trigger Svb processing to protect cells from apoptosis during metamorphosis [64]. Identifying Pri peptide partners other than Ubr3 would constitute an entry point to decipher novel Pri functions.

### 3.1.5. Reiteration of the Pri/Ubr3/Svb Module Acting in Other Tissues

Studying the functions of *pri* or *svb* at post-embryonic stages in *Drosophila* showed that the Pri/Ubr3/Svb module is reused in other tissues than the epidermis and leg to control tissue integrity during adulthood. *svb* and *pri* genes are expressed in stem cells of the digestive system (Malpighian tubules and intestine) where the Svb protein is processed in its transcriptional activate form to protect stem cells from apoptosis during adulthood [81,82]. Interestingly, *svb*, but not *pri*, is also expressed in the enterocytes, the intestinal differentiated cells, where the repressive form of Svb is required to trigger and maintain cell differentiation [82]. Post-embryonic roles of *pri* have also been described in the kissing bug *Rhodnius prolixus*, an important vector of the *Trypanosoma cruzi* parasite, which causes the Chagas disease, showing its function in digestive physiology and molting [83].

In summary, since their discovery, the pioneer Pri peptides have been extensively studied for more than 15 years, which make them one of the best characterized microproteins. Together, these studies reveal that Pri peptides have multiple features, i.e., dynamic expression patterns, diverse molecular partners, and a large number of targets, collectively conferring pleiotropic functions throughout development and adult life.

### 3.2. Microproteome Is a Pool of Regulatory Molecules

Given the tremendous number of uncharacterized microproteins, functional screens have been conducted to investigate their biological functions. Screens in human cell lines, aiming at identifying microproteins involved in cell growth and defining their interactomes, revealed their involvement in diverse cellular processes, such as translational regulation and endocytosis [27,28].

In *Drosophila*, functional screens have been conducted to explore the roles of the annotated microproteins in vivo by inducing their loss of function through RNAi or CRISPR techniques [64,84]. The specific microprotein depletion in developing legs showed that a substantial proportion of the tested microproteins (23/93) regulates leg morphogenesis [64]. Observations of defects in tissue growth, cell survival, patterning, and cuticle formation suggest that microproteins play regulatory roles in diverse cellular processes that drive the distinct morphogenetic events essential for leg formation [64]. Using bioinformatic analyses, Bosch et al. identified 298 microproteins conserved between human and *Drosophila* amongst the 1000 annotated microproteins in the *Drosophila* genome [84]. The knock-out (KO) of 115 of those conserved genes by CRISPR/Cas9 revealed that 14 are essential genes. In the same study, KO mutant flies for 25 other microproteins that were never characterized in other organisms were assessed for viability under normal and stressful conditions (here food-induced metabolic stress). Nearly all KO flies were viable and fertile in normal conditions, but several KO flies exhibited phenotypes under stressful conditions such as developmental delays, low viability, or lethality [84]. In *Drosophila*,



microprotein sequence analyses showed that microproteins are enriched in transmembrane domains and mitochondrial-associated peptide motifs [64,84], as also observed in other species, suggesting a role of microproteins in membrane biology and mitochondrial metabolism [85,86]. Interestingly, the analysis of genes encoding mitochondrial microproteins revealed distinct dynamic expression patterns during embryogenesis, suggesting that microproteins contribute to the cell and tissue specificities of mitochondrial activity over time [84].

### 3.3. Microproteins and de novo Genes

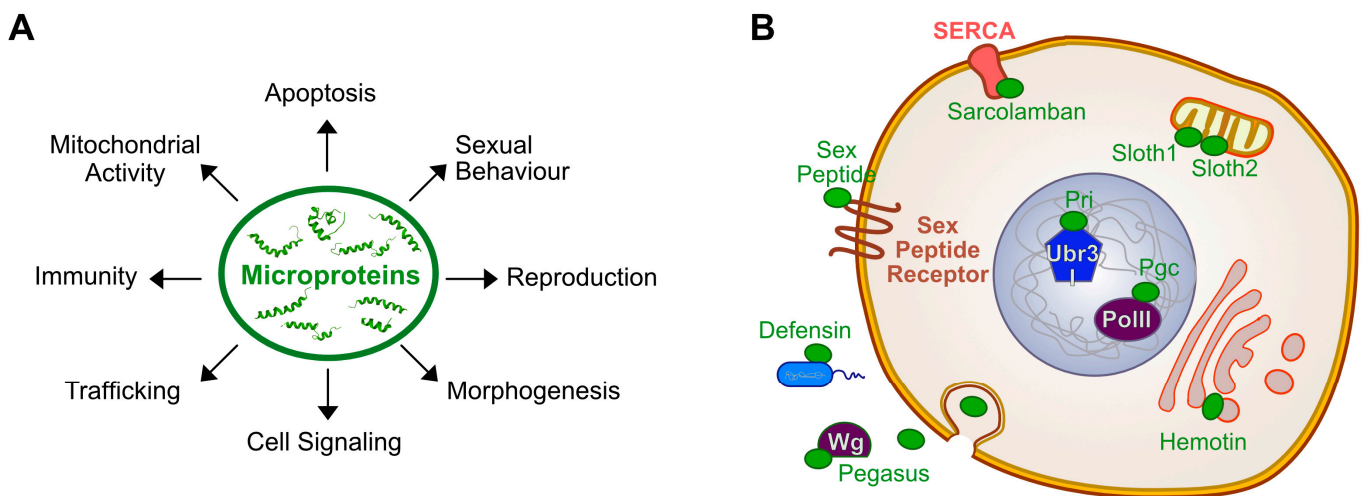
De novo genes are new protein-coding genes that recently appeared and are specific to a clade or to a few species. They emerge from non-coding regions, as well as from lncRNAs, and can evolve rapidly [87]. Despite their young age and their lack of conservation, the new genes may be essential for *Drosophila* development [87,88]. Since microproteins might have emerged from lncRNAs through the acquisition of robust translation during evolution, they could represent an important source of de novo proteins [28,32,87,89]. Interestingly, among 555 de novo genes recently identified in *Drosophila melanogaster* [90], we observed that half of them (225) code for microproteins, showing that genes encoding for microproteins strongly contribute to evolution. It has been proposed that the acquisition of new coding gene status involves several sequential steps: initial transcription, followed by ribosome binding, translation, and finally, the sustained maintenance of robust translation [32]. The maintenance of this translation may be enhanced by the functional properties of the nascent microprotein, which improve the organism fitness [32]. The pervasive ribosome binding could represent a mechanism for the emergence of new protein products that could confer an adaptive advantage to the organism. Although no de novo microprotein has been investigated to date for its contribution to *Drosophila* fitness, several newly born genes have been shown to be essential for *Drosophila* [87,88], which may presage that some de novo microproteins may confer an advantage for animal well-being.

### 3.4. Microproteins Represent an Additional Layer of Regulators for Biological Processes and Protein Activity

Since the first studies in *Drosophila* (reviewed in [91]), a significant number of microproteins has been described, revealing their various roles in vivo. Approaches to investigate microprotein functions rely on gain- and loss-of-function and phenotypic analyses, specifically examining the physiology and/or morphology of the tissue or organ in which the microprotein is expressed. The field greatly benefits from CRISPR/Cas9 genetic tools, which enable, for example, the generation of precise mutations at endogenous loci to disrupt smORF translation (codon start or smORF sequence deletion) or to tag the microprotein [80,92,93]. Finally, the molecular target of the microprotein can be identified by proteomic approaches. In this section, we review recent findings on the roles of microproteins in *Drosophila* development and physiology, illustrating their extensive functions across tissues over time (Figure 3).

In addition to Pri peptides and their essential roles in morphogenesis and cell survival during development, loss of function of other microproteins leads to developmental defects of varying degrees of severity. For instance, Pegasus, a 80 aa microprotein carrying a signal peptide, has been shown to increase lethality in *Drosophila* when deleted [94]. Its function has been studied in the wing imaginal disc, where it is expressed in the region that gives rise to the adult wing. Secreted Pegasus is essential for the diffusion of Wingless and the establishment of its gradient, which is necessary for patterning the number of mechanosensory bristles at the wing margin [94]. Another example is *wurmchen 2* (*wrm2*), a 57 aa transmembrane microprotein, whose depletion results in larval lethality. Although it localizes to the apical extracellular matrix of the embryonic tracheal tubes, no morphogenetic defect was observed, suggesting its essential role in later developmental processes [93]. Notably, *wrm2* is encoded by a bicistronic gene (*wurmchen1/2*), which transcribes an RNA molecule also encoding *wurmchen 1*, a small transmembrane protein of 162 aa, essential

for the septate junction complex in the embryonic tracheal tube [93]. Unlike *pri* [6,7] or *sarcolamban* [12] genes, which are poly/bicistronic encoding microproteins with redundant biological functions, the *wurmchen1/2* gene encodes two microproteins localized within the same protein complex but displaying distinct biological roles [93]. Similarly, the two Sloth1 and Sloth2 microproteins, encoded by the bicistronic gene *sloth1/2*, are part of the same protein complex III in the mitochondrial inner membrane, but each of them has a unique role since they are not functionally redundant [92]. Sloth1 and Sloth2 are essential microproteins involved in mitochondrial activity in neurons [92]. Together, these studies support the existence of polycistronic genes in eukaryotes, as observed previously in human cells [23,27]. They are highly represented among smORF genes since 32 smORFs are encoded by polycistronic transcripts among the 298 smORFs conserved between *Drosophila* and human [84].



**Figure 3.** Microproteins regulate multiple physiological, developmental, and cellular functions in *Drosophila*. (A) Studies investigating the roles of microproteins in *Drosophila* show that they regulate a wide range of biological functions, as illustrated here. (B) Schematic representation of microproteins and their molecular targets identified so far in *Drosophila*. Microproteins regulate different types of canonical proteins localized in all cellular compartments, such as in nuclei with Pgc (Polar granule component) binding to Polymerase II and Pri peptides to Ubr3. Microproteins can also localize to membranes, like Hemotin in endosomes, or Sloth1/2 in mitochondrial membrane. They can also be secreted: Pegasus interacts with Wg (Wingless) in the extracellular space, Sex Peptide binds to its GPCR (G Protein-Coupled Receptors) receptor (Sex Peptide Receptor) and the AMP Defensin attaches to bacteria cell wall (in blue).

Another specific feature of microproteins is their significant enrichment in transmembrane motifs [85] and, in *Drosophila*, two microproteins were characterized as important regulators of cell trafficking [95,96]. Hemotin is a 88 aa transmembrane microprotein localized in the membrane of early endosomes in *Drosophila* macrophages, required for endosome biogenesis [95]. Its depletion results in abnormal, enlarged endosomes that are less efficient at degrading bacteria and pathogens, leading to a reduced resistance to infection and a diminished immunity. Noteworthy, the vertebrate homolog of Hemotin is the 88 aa microprotein Stannin, and both share a well-conserved transmembrane domain and molecular properties [95]. More recently, the microprotein Purriato was characterized as a transmembrane protein important for regulating autophagy in *Drosophila* muscles and thus preserving muscle fiber integrity [96]. These data support that microproteins are a valuable source of new players in membrane biology.

The family of Antimicrobial Peptides (AMPs) participates in the *Drosophila* innate immune system and is largely composed of microproteins since 23 of the 32 AMPs are less than 100 aa [97]. Their primary mode of action is to disrupt bacteria and fungi membranes

through their insertion into the microbial membrane. Each AMP family appears to fight a specific group of pathogens, which is important for maintaining a balanced microbiome while effectively fighting infection. However, these AMPs appear to play other roles apart from their primary function in immunity. For example, the AMP Defensin, known to inhibit the growth of Gram-negative bacteria, induces tumor cell death [98]. Indeed, the presence of a tumor increases the expression of both Tumor Necrosis Factor (TNF) and Defensin. TNF alters the charge of tumor cell membranes, making them more sensitive to Defensin, which in turn induces cell death and tumor regression [98]. Thus, the function of microproteins could be co-opted in other biological processes, similarly to canonical proteins.

Finally, microproteins exhibit important roles in reproduction and sexual behavior. The iconic Sex Peptide (SP, 55 aa), discovered by biochemical approach before the birth of the microprotein field, is one of the best-characterized microprotein in *Drosophila* as SP is a crucial regulator of the Post-Mating Response in females [99]. Recently, the smallest microprotein MSAmiP (9 aa), found to be encoded by the *Drosophila* genome, is localized in the alleged *msa* lncRNA [100]. MSAmiP is expressed in the male accessory glands, which are the analog of the human prostate in *Drosophila*, and is required for sperm competition [100]. Furthermore, the translation of MSAmiP from the *msa* lncRNA, already characterized as a pre-miRNA, reveals the bimodal functions (non-coding and coding) of this gene. Another microprotein encoded from a previously annotated lncRNA is the microprotein IBIN (for Induced By INfection), first described as a lncRNA involved in immunity [101]. Surprisingly, IBIN is overexpressed in the eyes of flies after sighting a wasp and is required to accelerate mating behavior, thus coupling vision, changes in gene expression, and sexual behavior [102].

These recent findings have demonstrated that microproteins in *Drosophila* play crucial roles in a large variety of cellular and physiological processes across different cell types during development, highlighting their importance as a biological additional regulatory layer (Figure 3).

#### 4. Conclusions

Since the discovery of the microprotein field, both experimental and computational biology have made significant advances leading to the identification of thousands of microproteins encoded by small Open Reading Frames (smORFs) in the genomes of all living organisms. Important recent contributions are coming from vertebrates studies, especially of humans [27,28,30,48,103,104]. Indeed, microproteins have been shown to play important roles in human diseases, especially in cancer where they act as tumor suppressors or oncogenic factors (for review, see [30,57]). Moreover, microproteins represent a source of cancer neoantigens, thus opening up great possibilities in terms of therapeutic approaches, e.g., as targets of vaccines [59]. Despite these important advances, only a small proportion of microproteins have been investigated for their biologic functions, particularly at the organismal level, indicating that numerous exciting discoveries remain to be made given the large number of microproteins whose functions are unknown. If we consider the possibility that many of the canonical proteins are regulated by one or more microproteins, as exemplified with the mammalian SERCA pump regulated by two microproteins [105], the elucidation of the functional repertoire of microproteins will require a substantial body of research. Large-scale approaches using proteomic, genetic, or computational methods, which proved their effectiveness [27,28,46,64,84,106,107], remain techniques of choice for undertaking functional studies of microproteins. Up to now, the evolutive conservation of microproteins constitutes an important requisite to define the importance of microproteins. However, investigations of the interactomes of young human microproteins reveal their ability to interact with essential canonical proteins via short amino acid motifs mediating protein–protein interactions, despite their lack of conservation [28]. This result suggests that sequence conservation is not a compulsory criterion of functionality since these novel microproteins, specific to the human clade, can interact with canonical and conserved proteins through these peptide motifs [28]. Instead, these data indicate that peptide motifs

in de novo microproteins could mediate interactions with other proteins to modulate their activity. As illustrated by the Sarcolamban/Phospholamban microproteins regulating the SERCA muscle pump in vertebrates and invertebrates, non-conserved microproteins from one species could interact with canonical proteins from other species, which, despite their lack of sequence conservation, can be functionally swapped due to their conservation in the 3D structure [12]. Altogether, the use of the *Drosophila* model has proven to be highly relevant for elucidating the mechanisms regulating human physiology and pathology, thus highlighting the importance of performing functional studies of microproteins in this species [108]. Finally, the original and eclectic roles of microproteins discovered in the fruit fly underscore the fundamental role of this genetic model for understanding how microproteins influence animal development, physiology, and behavior.

**Author Contributions:** Conceptualization, H.C.-D. and J.Z., writing—original draft preparation, H.C.-D. and J.Z.; writing—review and editing, H.C.-D. and J.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Acknowledgments:** The authors thanks Clément Immarigeon, David Ohayon, Aurore Pelletier, Simon Marques-Prieto and Philippe Valenti for critical reading.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

- Dinger, M.E.; Pang, K.C.; Mercer, T.R.; Mattick, J.S. Differentiating Protein-Coding and Noncoding RNA: Challenges and Ambiguities. *PLoS Comput. Biol.* **2008**, *4*, e1000176. [[CrossRef](#)] [[PubMed](#)]
- Misra, S.; Crosby, M.A.; Mungall, C.J.; Matthews, B.B.; Campbell, K.S.; Hradecky, P.; Huang, Y.; Kaminker, J.S.; Millburn, G.H.; Prochnik, S.E.; et al. Annotation of the *Drosophila melanogaster* euchromatic Genome: A Systematic Review. *Genome Biol.* **2002**, *3*, 1–22. [[CrossRef](#)] [[PubMed](#)]
- Oliver, S.G.; van der Aart, Q.J.M.; Agostoni-Carbone, M.L.; Aigle, M.; Alberghina, L.; Alexandraki, D.; Antoine, G.; Anwar, R.; Ballesta, J.P.G.; Benit, P.; et al. The Complete DNA Sequence of Yeast Chromosome III. *Nature* **1992**, *357*, 38–46. [[CrossRef](#)]
- Basrai, M.A.; Hieter, P.; Boeke, J.D. Small Open Reading Frames: Beautiful Needles in the Haystack. *Genome Res.* **1997**, *7*, 768–771. [[CrossRef](#)]
- Savard, J.; Marques-Souza, H.; Aranda, M.; Tautz, D. A Segmentation Gene in *Tribolium* Produces a Polycistronic mRNA That Codes for Multiple Conserved Peptides. *Cell* **2006**, *126*, 559–569. [[CrossRef](#)]
- Galindo, M.I.; Pueyo, J.I.; Fouix, S.; Bishop, S.A.; Couso, J.P. Peptides Encoded by Short ORFs Control Development and Define a New Eukaryotic Gene Family. *PLoS Biol.* **2007**, *5*, e106. [[CrossRef](#)] [[PubMed](#)]
- Kondo, T.; Hashimoto, Y.; Kato, K.; Inagaki, S.; Hayashi, S.; Kageyama, Y. Small Peptide Regulators of Actin-Based Cell Morphogenesis Encoded by a Polycistronic mRNA. *Nat. Cell Biol.* **2007**, *9*, 660–665. [[CrossRef](#)]
- Slavoff, S.A.; Mitchell, A.J.; Schwaid, A.G.; Cabili, M.N.; Ma, J.; Levin, J.Z.; Karger, A.D.; Budnik, B.A.; Rinn, J.L.; Saghatelian, A. Peptidomic Discovery of Short Open Reading Frame–Encoded Peptides in Human Cells. *Nat. Chem. Biol.* **2013**, *9*, 59–64. [[CrossRef](#)]
- Kastenmayer, J.P.; Ni, L.; Chu, A.; Kitchen, L.E.; Au, W.-C.; Yang, H.; Carter, C.D.; Wheeler, D.; Davis, R.W.; Boeke, J.D.; et al. Functional Genomics of Genes with Small Open Reading Frames (sORFs) in *S. Cerevisiae*. *Genome Res.* **2006**, *16*, 365–373. [[CrossRef](#)]
- Hanada, K.; Higuchi-Takeuchi, M.; Okamoto, M.; Yoshizumi, T.; Shimizu, M.; Nakaminami, K.; Nishi, R.; Ohashi, C.; Iida, K.; Tanaka, M.; et al. Small Open Reading Frames Associated with Morphogenesis Are Hidden in Plant Genomes. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 2395–2400. [[CrossRef](#)]
- Chng, S.C.; Ho, L.; Tian, J.; Reversade, B. ELABELA: A Hormone Essential for Heart Development Signals via the Apelin Receptor. *Dev. Cell* **2013**, *27*, 672–680. [[CrossRef](#)] [[PubMed](#)]
- Magny, E.G.; Pueyo, J.I.; Pearl, F.M.G.; Cespedes, M.A.; Niven, J.E.; Bishop, S.A.; Couso, J.P. Conserved Regulation of Cardiac Calcium Uptake by Peptides Encoded in Small Open Reading Frames. *Science* **2013**, *341*, 1116–1120. [[CrossRef](#)]
- Pauli, A.; Norris, M.L.; Valen, E.; Chew, G.-L.; Gagnon, J.A.; Zimmerman, S.; Mitchell, A.; Ma, J.; Dubrulle, J.; Reyon, D.; et al. Toddler: An Embryonic Signal That Promotes Cell Movement via Apelin Receptors. *Science* **2014**, *343*, 1248636. [[CrossRef](#)] [[PubMed](#)]
- Storz, G.; Wolf, Y.I.; Ramamurthi, K.S. Small Proteins Can No Longer Be Ignored. *Annu. Rev. Biochem.* **2014**, *83*, 753–777. [[CrossRef](#)] [[PubMed](#)]
- Wenkel, S.; Emery, J.; Hou, B.-H.; Evans, M.M.S.; Barton, M.K. A Feedback Regulatory Module Formed by LITTLE ZIPPER and HD-ZIPIII Genes. *Plant Cell* **2007**, *19*, 3379–3390. [[CrossRef](#)]



16. Ladoukakis, E.; Pereira, V.; Magny, E.G.; Eyre-Walker, A.; Couso, J. Hundreds of Putatively Functional Small Open Reading Frames in *Drosophila*. *Genome Biol.* **2011**, *12*, R118. [[CrossRef](#)]
17. Mackowiak, S.D.; Zauber, H.; Bielow, C.; Thiel, D.; Kutz, K.; Calviello, L.; Mastrobuoni, G.; Rajewsky, N.; Kempa, S.; Selbach, M.; et al. Extensive Identification and Analysis of Conserved Small ORFs in Animals. *Genome Biol.* **2015**, *16*, 179. [[CrossRef](#)]
18. Zhu, M.; Gribskov, M. MiPepid: MicroPeptide Identification Tool Using Machine Learning. *BMC Bioinform.* **2019**, *20*, 559. [[CrossRef](#)] [[PubMed](#)]
19. Zhang, Y.; Jia, C.; Fullwood, M.J.; Kwoh, C.K. DeepCPP: A Deep Neural Network Based on Nucleotide Bias Information and Minimum Distribution Similarity Feature Selection for RNA Coding Potential Prediction. *Brief. Bioinform.* **2021**, *22*, 2073–2084. [[CrossRef](#)]
20. Ingolia, N.T.; Ghaemmaghami, S.; Newman, J.R.S.; Weissman, J.S. Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling. *Science* **2009**, *324*, 218–223. [[CrossRef](#)]
21. Ingolia, N.T.; Lareau, L.F.; Weissman, J.S. Ribosome Profiling of Mouse Embryonic Stem Cells Reveals the Complexity and Dynamics of Mammalian Proteomes. *Cell* **2011**, *147*, 789–802. [[CrossRef](#)] [[PubMed](#)]
22. Ingolia, N.T.; Brar, G.A.; Rouskin, S.; McGeachy, A.M.; Weissman, J.S. The Ribosome Profiling Strategy for Monitoring Translation in Vivo by Deep Sequencing of Ribosome-Protected mRNA Fragments. *Nat. Protoc.* **2012**, *7*, 1534–1550. [[CrossRef](#)]
23. Moulleron, H.; Delcourt, V.; Roucou, X. Death of a Dogma: Eukaryotic mRNAs Can Code for More than One Protein. *Nucleic Acids Res.* **2016**, *44*, 14–23. [[CrossRef](#)] [[PubMed](#)]
24. Ingolia, N.T.; Brar, G.A.; Stern-Ginossar, N.; Harris, M.S.; Talhouarne, G.J.S.; Jackson, S.E.; Wills, M.R.; Weissman, J.S. Ribosome Profiling Reveals Pervasive Translation Outside of Annotated Protein-Coding Genes. *Cell Rep.* **2014**, *8*, 1365–1379. [[CrossRef](#)]
25. Mudge, J.M.; Ruiz-Orera, J.; Prensner, J.R.; Brunet, M.A.; Calvet, F.; Jungreis, I.; Gonzalez, J.M.; Magrane, M.; Martinez, T.F.; Schulz, J.F.; et al. Standardized Annotation of Translated Open Reading Frames. *Nat. Biotechnol.* **2022**, *40*, 994–999. [[CrossRef](#)]
26. Renz, P.F.; Valdivia-Francia, F.; Sendoel, A. Some like It Translated: Small ORFs in the 5'UTR. *Exp. Cell Res.* **2020**, *396*, 112229. [[CrossRef](#)] [[PubMed](#)]
27. Chen, J.; Brunner, A.-D.; Cogan, J.Z.; Nuñez, J.K.; Fields, A.P.; Adamson, B.; Itzhak, D.N.; Li, J.Y.; Mann, M.; Leonetti, M.D.; et al. Pervasive Functional Translation of Noncanonical Human Open Reading Frames. *Science* **2020**, *367*, 1140–1146. [[CrossRef](#)]
28. Sandmann, C.-L.; Schulz, J.F.; Ruiz-Orera, J.; Kirchner, M.; Ziehm, M.; Adami, E.; Marczenke, M.; Christ, A.; Liebe, N.; Greiner, J.; et al. Evolutionary Origins and Interactomes of Human, Young Microproteins and Small Peptides Translated from Short Open Reading Frames. *Mol. Cell* **2023**, *83*, 994–1011.e18. [[CrossRef](#)]
29. Jayaram, D.R.; Frost, S.; Argov, C.; Liju, V.B.; Anto, N.P.; Muraleedharan, A.; Ben-Ari, A.; Sinay, R.; Smoly, I.; Novoplansky, O.; et al. Unraveling the Hidden Role of a uORF-Encoded Peptide as a Kinase Inhibitor of PKCs. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2018899118. [[CrossRef](#)]
30. Hofman, D.A.; Ruiz-Orera, J.; Yannuzzi, I.; Murugesan, R.; Brown, A.; Clauser, K.R.; Condurat, A.L.; Van Dinter, J.T.; Engels, S.A.G.; Goodale, A.; et al. Translation of Non-Canonical Open Reading Frames as a Cancer Cell Survival Mechanism in Childhood Medulloblastoma. *Mol. Cell* **2024**, *84*, 261–276.e18. [[CrossRef](#)]
31. Wu, Q.; Wright, M.; Gogol, M.M.; Bradford, W.D.; Zhang, N.; Bazzini, A.A. Translation of Small Downstream ORFs Enhances Translation of Canonical Main Open Reading Frames. *EMBO J.* **2020**, *39*, e104763. [[CrossRef](#)] [[PubMed](#)]
32. Patraquim, P.; Magny, E.G.; Pueyo, J.I.; Platero, A.I.; Couso, J.P. Translation and Natural Selection of Micropeptides from Long Non-Canonical RNAs. *Nat. Commun.* **2022**, *13*, 6515. [[CrossRef](#)]
33. Prensner, J.R.; Abelin, J.G.; Kok, L.W.; Clauser, K.R.; Mudge, J.M.; Ruiz-Orera, J.; Bassani-Sternberg, M.; Moritz, R.L.; Deutsch, E.W.; van Heesch, S. What Can Ribo-Seq, Immunopeptidomics, and Proteomics Tell Us About the Noncanonical Proteome? *Mol. Cell. Proteom.* **2023**, *22*, 100631. [[CrossRef](#)]
34. Zhang, M.; Zhao, K.; Xu, X.; Yang, Y.; Yan, S.; Wei, P.; Liu, H.; Xu, J.; Xiao, F.; Zhou, H.; et al. A Peptide Encoded by Circular Form of LINC-PINT Suppresses Oncogenic Transcriptional Elongation in Glioblastoma. *Nat. Commun.* **2018**, *9*, 4475. [[CrossRef](#)]
35. Montigny, A.; Tavormina, P.; Duboe, C.; San Clémente, H.; Aguilar, M.; Valenti, P.; Laressergues, D.; Combier, J.-P.; Plaza, S. *Drosophila* Primary microRNA-8 Encodes a microRNA-Encoded Peptide Acting in Parallel of miR-8. *Genome Biol.* **2021**, *22*, 118. [[CrossRef](#)]
36. Sudmant, P.H.; Lee, H.; Dominguez, D.; Heiman, M.; Burge, C.B. Widespread Accumulation of Ribosome-Associated Isolated 3' UTRs in Neuronal Cell Populations of the Aging Brain. *Cell Rep.* **2018**, *25*, 2447–2456.e4. [[CrossRef](#)]
37. Kesner, J.S.; Chen, Z.; Shi, P.; Aparicio, A.O.; Murphy, M.R.; Guo, Y.; Trehan, A.; Lipponen, J.E.; Recinos, Y.; Myeku, N.; et al. Noncoding Translation Mitigation. *Nature* **2023**, *617*, 395–402. [[CrossRef](#)] [[PubMed](#)]
38. Laumont, C.M.; Vincent, K.; Hesnard, L.; Audemard, É.; Bonneil, É.; Laverdure, J.-P.; Gendron, P.; Courcelles, M.; Hardy, M.-P.; Côté, C.; et al. Noncoding Regions Are the Main Source of Targetable Tumor-Specific Antigens. *Sci. Transl. Med.* **2018**, *10*, eaau5516. [[CrossRef](#)] [[PubMed](#)]
39. Aspden, J.L.; Eyre-Walker, Y.C.; Phillips, R.J.; Amin, U.; Mumtaz, M.A.S.; Brocard, M.; Couso, J.-P. Extensive Translation of Small Open Reading Frames Revealed by Poly-Ribo-Seq. *eLife* **2014**, *3*, e03528. [[CrossRef](#)]
40. Li, H.; Hu, C.; Bai, L.; Li, H.; Li, M.; Zhao, X.; Czajkowsky, D.M.; Shao, Z. Ultra-Deep Sequencing of Ribosome-Associated Poly-Adenylated RNA in Early *Drosophila* Embryos Reveals Hundreds of Conserved Translated sORFs. *DNA Res.* **2016**, *23*, 571–580. [[CrossRef](#)]



41. Patraquim, P.; Mumtaz, M.A.S.; Pueyo, J.I.; Aspden, J.L.; Couso, J.-P. Developmental Regulation of Canonical and Small ORF Translation from mRNAs. *Genome Biol.* **2020**, *21*, 128. [[CrossRef](#)] [[PubMed](#)]
42. Douka, K.; Agapiou, M.; Birds, I.; Aspden, J.L. Optimization of Ribosome Footprinting Conditions for Ribo-Seq in Human and *Drosophila melanogaster* Tissue Culture Cells. *Front. Mol. Biosci.* **2022**, *8*, 791455. [[CrossRef](#)]
43. Tong, G.; Hah, N.; Martinez, T.F. Comparison of Software Packages for Detecting Unannotated Translated Small Open Reading Frames by Ribo-Seq. *Brief. Bioinform.* **2024**, *25*, bbae268. [[CrossRef](#)] [[PubMed](#)]
44. Wacholder, A.; Carvunis, A.-R. Biological Factors and Statistical Limitations Prevent Detection of Most Noncanonical Proteins by Mass Spectrometry. *PLoS Biol.* **2023**, *21*, e3002409. [[CrossRef](#)]
45. Fabre, B.; Combier, J.-P.; Plaza, S. Recent Advances in Mass Spectrometry-Based Peptidomics Workflows to Identify Short-Open-Reading-Frame-Encoded Peptides and Explore Their Functions. *Curr. Opin. Chem. Biol.* **2021**, *60*, 122–130. [[CrossRef](#)]
46. Fabre, B.; Choteau, S.A.; Duboé, C.; Pichereaux, C.; Montigny, A.; Korona, D.; Deery, M.J.; Camus, M.; Brun, C.; Burlet-Schiltz, O.; et al. In Depth Exploration of the Alternative Proteome of *Drosophila melanogaster*. *Front. Cell Dev. Biol.* **2022**, *10*, 901351. [[CrossRef](#)] [[PubMed](#)]
47. Wang, Z.; Pan, N.; Yan, J.; Wan, J.; Wan, C. Systematic Identification of Microproteins during the Development of *Drosophila melanogaster*. *J. Proteome Res.* **2022**, *21*, 1114–1123. [[CrossRef](#)] [[PubMed](#)]
48. Martinez, T.F.; Lyons-Abbott, S.; Bookout, A.L.; De Souza, E.V.; Donaldson, C.; Vaughan, J.M.; Lau, C.; Abramov, A.; Baquero, A.F.; Baquero, K.; et al. Profiling Mouse Brown and White Adipocytes to Identify Metabolically Relevant Small ORFs and Functional Microproteins. *Cell Metab.* **2023**, *35*, 166–183.e11. [[CrossRef](#)]
49. Brunet, M.A.; Lucier, J.-F.; Levesque, M.; Leblanc, S.; Jacques, J.-F.; Al-Saedi, H.R.H.; Guilloy, N.; Grenier, F.; Avino, M.; Fournier, I.; et al. OpenProt 2021: Deeper Functional Annotation of the Coding Potential of Eukaryotic Genomes. *Nucleic Acids Res.* **2021**, *49*, D380–D388. [[CrossRef](#)]
50. Li, Y.; Zhou, H.; Chen, X.; Zheng, Y.; Kang, Q.; Hao, D.; Zhang, L.; Song, T.; Luo, H.; Hao, Y.; et al. SmProt: A Reliable Repository with Comprehensive Annotation of Small Proteins Identified from Ribosome Profiling. *Genom. Proteom. Bioinform.* **2021**, *19*, 602–610. [[CrossRef](#)]
51. Choteau, S.A.; Wagner, A.; Pierre, P.; Spinelli, L.; Brun, C. MetamORF: A Repository of Unique Short Open Reading Frames Identified by Both Experimental and Computational Approaches for Gene and Metagene Analyses. *Database* **2021**, *2021*, baab032. [[CrossRef](#)] [[PubMed](#)]
52. Olexiouk, V.; Crappé, J.; Verbruggen, S.; Verhegen, K.; Martens, L.; Menschaert, G. sORFs.Org: A Repository of Small ORFs Identified by Ribosome Profiling. *Nucleic Acids Res.* **2016**, *44*, D324–D329. [[CrossRef](#)] [[PubMed](#)]
53. Hao, Y.; Zhang, L.; Niu, Y.; Cai, T.; Luo, J.; He, S.; Zhang, B.; Zhang, D.; Qin, Y.; Yang, F.; et al. SmProt: A Database of Small Proteins Encoded by Annotated Coding and Non-Coding RNA Loci. *Brief. Bioinform.* **2018**, *19*, 636–643. [[CrossRef](#)] [[PubMed](#)]
54. Hoskins, R.A.; Carlson, J.W.; Wan, K.H.; Park, S.; Mendez, I.; Galle, S.E.; Booth, B.W.; Pfeiffer, B.D.; George, R.A.; Svirskas, R.; et al. The Release 6 Reference Sequence of the *Drosophila melanogaster* Genome. *Genome Res.* **2015**, *25*, 445–458. [[CrossRef](#)] [[PubMed](#)]
55. Torres, M.D.T.; Brooks, E.F.; Cesaro, A.; Sberro, H.; Gill, M.O.; Nicolaou, C.; Bhatt, A.S.; De La Fuente-Nunez, C. Mining Human Microbiomes Reveals an Untapped Source of Peptide Antibiotics. *Cell* **2024**, *187*, 5453–5467.e15. [[CrossRef](#)] [[PubMed](#)]
56. Torres, J.P.; Lin, Z.; Watkins, M.; Salcedo, P.F.; Baskin, R.P.; Elhabian, S.; Safavi-Hemami, H.; Taylor, D.; Tun, J.; Concepcion, G.P.; et al. Small-Molecule Mimicry Hunting Strategy in the Imperial Cone Snail, *Conus imperialis*. *Sci. Adv.* **2021**, *7*, eabf2704. [[CrossRef](#)]
57. Wright, B.W.; Yi, Z.; Weissman, J.S.; Chen, J. The Dark Proteome: Translation from Noncanonical Open Reading Frames. *Trends Cell Biol.* **2022**, *32*, 243–258. [[CrossRef](#)]
58. Merino-Valverde, I.; Greco, E.; Abad, M. The Microproteome of Cancer: From Invisibility to Relevance. *Exp. Cell Res.* **2020**, *392*, 111997. [[CrossRef](#)]
59. Camarena, M.E.; Theunissen, P.; Ruiz, M.; Ruiz-Orera, J.; Calvo-Serra, B.; Castelo, R.; Castro, C.; Sarobe, P.; Fortes, P.; Perera-Bel, J.; et al. Microproteins Encoded by Noncanonical ORFs Are a Major Source of Tumor-Specific Antigens in a Liver Cancer Patient Meta-Cohort. *Sci. Adv.* **2024**, *10*, eadn3628. [[CrossRef](#)]
60. Inagaki, S.; Numata, K.; Kondo, T.; Tomita, M.; Yasuda, K.; Kanai, A.; Kageyama, Y. Identification and Expression Analysis of Putative mRNA-like Non-Coding RNA in *Drosophila*: mRNA-like Non-Coding RNAs in *Drosophila*. *Genes Cells* **2005**, *10*, 1163–1173. [[CrossRef](#)]
61. Tobias-Santos, V.; Guerra-Almeida, D.; Mury, F.; Ribeiro, L.; Berni, M.; Araujo, H.; Logullo, C.; Feitosa, N.M.; de Souza-Menezes, J.; Pessoa Costa, E.; et al. Multiple Roles of the Polycistronic Gene Tarsal-Less/Mille-Pattes/Polished-Rice During Embryogenesis of the Kissing Bug *Rhodnius prolixus*. *Front. Ecol. Evol.* **2019**, *7*, 379. [[CrossRef](#)]
62. Pueyo, J.I.; Couso, J.P. The 11-Aminoacid Long Tarsal-Less Peptides Trigger a Cell Signal in *Drosophila* Leg Development. *Dev. Biol.* **2008**, *324*, 192–201. [[CrossRef](#)] [[PubMed](#)]
63. Chanut-Delalande, H.; Hashimoto, Y.; Pelissier-Monier, A.; Spokony, R.; Dib, A.; Kondo, T.; Bohère, J.; Niimi, K.; Latapie, Y.; Inagaki, S.; et al. Pri Peptides Are Mediators of Ecdysone for the Temporal Control of Development. *Nat. Cell Biol.* **2014**, *16*, 1035–1044. [[CrossRef](#)]
64. Markus, D.; Pelletier, A.; Boube, M.; Port, F.; Boutros, M.; Payre, F.; Obermayer, B.; Zanet, J. The Pleiotropic Functions of Pri smORF Peptides Synchronize Leg Development Regulators. *PLoS Genet.* **2023**, *19*, e1011004. [[CrossRef](#)] [[PubMed](#)]

65. Poodry, C.A.; Schneiderman, H.A. The Ultrastructure of the Developing Leg of *Drosophila melanogaster*. *Wilhelm Roux' Archiv Entwicklungsmechanik Org.* **1970**, *166*, 1–44. [[CrossRef](#)]
66. McLean, P.F.; Cooley, L. Protein Equilibration Through Somatic Ring Canals in *Drosophila*. *Science* **2013**, *340*, 1445–1447. [[CrossRef](#)]
67. Pi, H.; Huang, Y.-C.; Chen, I.-C.; Lin, C.-D.; Yeh, H.-F.; Pai, L.-M. Identification of 11-Amino Acid Peptides That Disrupt Notch-Mediated Processes in *Drosophila*. *J. Biomed. Sci.* **2011**, *18*, 42. [[CrossRef](#)]
68. Payre, F.; Vincent, A.; Carreno, S. Ovo/Svb Integrates Wingless and DER Pathways to Control Epidermis Differentiation. *Nature* **1999**, *400*, 271–275. [[CrossRef](#)]
69. Chanut-Delalande, H.; Fernandes, I.; Roch, F.; Payre, F.; Plaza, S. Shavenbaby Couples Patterning to Epidermal Cell Shape Control. *PLoS Biol.* **2006**, *4*, e290. [[CrossRef](#)]
70. Fernandes, I.; Chanut-Delalande, H.; Ferrer, P.; Latapie, Y.; Waltzer, L.; Affolter, M.; Payre, F.; Plaza, S. Zona Pellucida Domain Proteins Remodel the Apical Compartment for Localized Cell Shape Changes. *Dev. Cell* **2010**, *18*, 64–76. [[CrossRef](#)]
71. Menoret, D.; Santolini, M.; Fernandes, I.; Spokony, R.; Zanet, J.; Gonzalez, I.; Latapie, Y.; Ferrer, P.; Rouault, H.; White, K.P.; et al. Genome-Wide Analyses of Shavenbaby Target Genes Reveals Distinct Features of Enhancer Organization. *Genome Biol.* **2013**, *14*, R86. [[CrossRef](#)] [[PubMed](#)]
72. Kondo, T.; Plaza, S.; Zanet, J.; Benrabah, E.; Valenti, P.; Hashimoto, Y.; Kobayashi, S.; Payre, F.; Kageyama, Y. Small Peptides Switch the Transcriptional Activity of Shavenbaby During *Drosophila* Embryogenesis. *Science* **2010**, *329*, 336–339. [[CrossRef](#)]
73. Zanet, J.; Benrabah, E.; Li, T.; Pélissier-Monier, A.; Chanut-Delalande, H.; Ronsin, B.; Bellen, H.J.; Payre, F.; Plaza, S. Pri sORF Peptides Induce Selective Proteasome-Mediated Protein Processing. *Science* **2015**, *349*, 1356–1358. [[CrossRef](#)] [[PubMed](#)]
74. Tasaki, T.; Sohr, R.; Xia, Z.; Hellweg, R.; Hörtnagl, H.; Varshavsky, A.; Kwon, Y.T. Biochemical and Genetic Studies of UBR3, a Ubiquitin Ligase with a Function in Olfactory and Other Sensory Systems. *J. Biol. Chem.* **2007**, *282*, 18510–18520. [[CrossRef](#)]
75. Dib, A.; Zanet, J.; Mancheno-Ferris, A.; Gallois, M.; Markus, D.; Valenti, P.; Marques-Prieto, S.; Plaza, S.; Kageyama, Y.; Chanut-Delalande, H.; et al. Pri smORF Peptides Are Wide Mediators of Ecdysone Signaling, Contributing to Shape Spatiotemporal Responses. *Front. Genet.* **2021**, *12*, 714152. [[CrossRef](#)] [[PubMed](#)]
76. Taira, Y.; Wada, H.; Hayashi, S.; Kageyama, Y. Polished Rice Mediates Ecdysone-Dependent Control of *Drosophila* Embryonic Organogenesis. *Genes Cells* **2021**, *26*, 269–281. [[CrossRef](#)]
77. Ray, S.; Rosenberg, M.I.; Chanut-Delalande, H.; Decaras, A.; Schwertner, B.; Toubiana, W.; Auman, T.; Schnellhammer, I.; Teuscher, M.; Valenti, P.; et al. The Mlpt/Ubr3/Svb Module Comprises an Ancient Developmental Switch for Embryonic Patterning. *eLife* **2019**, *8*, e39748. [[CrossRef](#)]
78. Jiménez-Guri, E.; Wotton, K.R.; Jaeger, J. Tarsal-Less Is Expressed as a Gap Gene but Has No Gap Gene Phenotype in the Moth Midge *Clogmia Albipunctata*. *R. Soc. Open Sci.* **2018**, *5*, 180458. [[CrossRef](#)]
79. Öztürk-Çolak, A.; Stephan-Otto Attolini, C.; Casanova, J.; Araújo, S.J. Blimp-1 Mediates Tracheal Lumen Maturation in *Drosophila melanogaster*. *Genetics* **2018**, *210*, 653–663. [[CrossRef](#)]
80. Gallois, M.; Menoret, D.; Marques-Prieto, S.; Montigny, A.; Valenti, P.; Moussian, B.; Plaza, S.; Payre, F.; Chanut-Delalande, H. Pri Peptides Temporally Coordinate Transcriptional Programs during Epidermal Differentiation. *Sci. Adv.* **2024**, *10*, eadg8816. [[CrossRef](#)]
81. Bohère, J.; Mancheno-Ferris, A.; Al Hayek, S.; Zanet, J.; Valenti, P.; Akino, K.; Yamabe, Y.; Inagaki, S.; Chanut-Delalande, H.; Plaza, S.; et al. Shavenbaby and Yorkie Mediate Hippo Signaling to Protect Adult Stem Cells from Apoptosis. *Nat. Commun.* **2018**, *9*, 5123. [[CrossRef](#)] [[PubMed](#)]
82. Al Hayek, S.; Alsawadi, A.; Kambris, Z.; Boquete, J.; Bohère, J.; Immarigeon, C.; Ronsin, B.; Plaza, S.; Lemaitre, B.; Payre, F.; et al. Steroid-dependent Switch of OvoL/Shavenbaby Controls Self-renewal versus Differentiation of Intestinal Stem Cells. *EMBO J.* **2021**, *40*, e104347. [[CrossRef](#)] [[PubMed](#)]
83. Silva, C.A.O.; da Silveira Alves, S.; da Costa Rodrigues, B.; Fraga Egidio, J.A.; Ribeiro, L.; Logullo, C.; Mury, F.B.; das Graças Santos, D.; Portal, T.; Monteiro-de-Barros, C.; et al. The Mlpt smORF Gene Is Essential for Digestive Physiology and Molting during Nymphal Stages in the Kissing Bug *Rhodnius prolixus*. *Insect Biochem. Mol. Biol.* **2024**, *172*, 104154. [[CrossRef](#)]
84. Bosch, J.A.; Keith, N.; Escobedo, F.; Fisher, W.W.; LaGraff, J.T.; Rabasco, J.; Wan, K.H.; Weiszmann, R.; Wu, Y.; Hu, Y.; et al. Molecular and Functional Characterization of the *Drosophila melanogaster* Conserved smORFome. *Cell Rep.* **2023**, *42*, 113311. [[CrossRef](#)]
85. Makarewich, C.A. The Hidden World of Membrane Microproteins. *Exp. Cell Res.* **2020**, *388*, 111853. [[CrossRef](#)]
86. Kamradt, M.L.; Makarewich, C.A. Mitochondrial Microproteins: Critical Regulators of Protein Import, Energy Production, Stress Response Pathways, and Programmed Cell Death. *Am. J. Physiol. Cell Physiol.* **2023**, *325*, C807–C816. [[CrossRef](#)]
87. Reinhardt, J.A.; Wanjiru, B.M.; Brant, A.T.; Saelao, P.; Begun, D.J.; Jones, C.D. De Novo ORFs in *Drosophila* Are Important to Organismal Fitness and Evolved Rapidly from Previously Non-Coding Sequences. *PLoS Genet.* **2013**, *9*, e1003860. [[CrossRef](#)]
88. Chen, S.; Zhang, Y.E.; Long, M. New Genes in *Drosophila* Quickly Become Essential. *Science* **2010**, *330*, 1682–1685. [[CrossRef](#)] [[PubMed](#)]
89. Ruiz-Orera, J.; Villanueva-Cañas, J.L.; Albà, M.M. Evolution of New Proteins from Translated sORFs in Long Non-Coding RNAs. *Exp. Cell Res.* **2020**, *391*, 111940. [[CrossRef](#)]
90. Peng, J.; Zhao, L. The Origin and Structural Evolution of de Novo Genes in *Drosophila*. *Nat. Commun.* **2024**, *15*, 810. [[CrossRef](#)]
91. Zanet, J.; Chanut-Delalande, H.; Plaza, S.; Payre, F. Small Peptides as Newcomers in the Control of *Drosophila* Development. In *Current Topics in Developmental Biology*; Elsevier: Amsterdam, The Netherlands, 2016; Volume 117, pp. 199–219. ISBN 978-0-12-801382-3.

92. Bosch, J.A.; Ugur, B.; Pichardo-Casas, I.; Rabasco, J.; Escobedo, F.; Zuo, Z.; Brown, B.; Celniker, S.; Sinclair, D.A.; Bellen, H.J.; et al. Two Neuronal Peptides Encoded from a Single Transcript Regulate Mitochondrial Complex III in *Drosophila*. *eLife* **2022**, *11*, e82709. [[CrossRef](#)]
93. Königsmann, T.; Parfentev, I.; Urlaub, H.; Riedel, D.; Schuh, R. The Bicistronic Gene Würmchen Encodes Two Essential Components for Epithelial Development in *Drosophila*. *Dev. Biol.* **2020**, *463*, 53–62. [[CrossRef](#)]
94. Magny, E.G.; Platero, A.I.; Bishop, S.A.; Pueyo, J.I.; Aguilar-Hidalgo, D.; Couso, J.P. Pegasus, a Small Extracellular Peptide Enhancing Short-Range Diffusion of Wingless. *Nat. Commun.* **2021**, *12*, 5660. [[CrossRef](#)] [[PubMed](#)]
95. Pueyo, J.I.; Magny, E.G.; Sampson, C.J.; Amin, U.; Evans, I.R.; Bishop, S.A.; Couso, J.P. Hemotin, a Regulator of Phagocytosis Encoded by a Small ORF and Conserved across Metazoans. *PLoS Biol.* **2016**, *14*, e1002395. [[CrossRef](#)] [[PubMed](#)]
96. Pueyo, J.I.; Salazar, J.; Grincho, C.; Berni, J.; Towler, B.P.; Newbury, S.F. Purriato Is a Conserved Small Open Reading Frame Gene That Interacts with the CASA Pathway to Regulate Muscle Homeostasis and Epithelial Tissue Growth in *Drosophila*. *Front. Cell Dev. Biol.* **2023**, *11*, 1117454. [[CrossRef](#)] [[PubMed](#)]
97. Hanson, M.A.; Lemaitre, B. New Insights on *Drosophila* Antimicrobial Peptide Function in Host Defense and Beyond. *Curr. Opin. Immunol.* **2020**, *62*, 22–30. [[CrossRef](#)]
98. Parvy, J.-P.; Yu, Y.; Dostalova, A.; Kondo, S.; Kurjan, A.; Bulet, P.; Lemaitre, B.; Vidal, M.; Cordero, J.B. The Antimicrobial Peptide Defensin Cooperates with Tumour Necrosis Factor to Drive Tumour Cell Death in *Drosophila*. *eLife* **2019**, *8*, e45061. [[CrossRef](#)]
99. Chen, P.S.; Stumm-Zollinger, E.; Aigaki, T.; Balmer, J.; Bienz, M.; Böhlen, P. A Male Accessory Gland Peptide That Regulates Reproductive Behavior of Female *D. melanogaster*. *Cell* **1988**, *54*, 291–298. [[CrossRef](#)]
100. Immarigeon, C.; Frei, Y.; Delbare, S.Y.N.; Gligorov, D.; Machado Almeida, P.; Grey, J.; Fabbro, L.; Nagoshi, E.; Billeter, J.-C.; Wolfner, M.F.; et al. Identification of a Micropeptide and Multiple Secondary Cell Genes That Modulate *Drosophila* Male Reproductive Success. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2001897118. [[CrossRef](#)]
101. Valanne, S.; Salminen, T.S.; Järvelä-Stölting, M.; Vesala, L.; Rämetsä, M. Immune-Inducible Non-Coding RNA Molecule lincRNA-IBIN Connects Immunity and Metabolism in *Drosophila melanogaster*. *PLoS Pathog.* **2019**, *15*, e1007504. [[CrossRef](#)]
102. Ebrahim, S.A.M.; Talross, G.J.S.; Carlson, J.R. Sight of Parasitoid Wasps Accelerates Sexual Behavior and Upregulates a Micropeptide Gene in *Drosophila*. *Nat. Commun.* **2021**, *12*, 2453. [[CrossRef](#)] [[PubMed](#)]
103. Na, Z.; Dai, X.; Zheng, S.-J.; Bryant, C.J.; Loh, K.H.; Su, H.; Luo, Y.; Buhagiar, A.F.; Cao, X.; Baserga, S.J.; et al. Mapping Subcellular Localizations of Unannotated Microproteins and Alternative Proteins with MicroID. *Mol. Cell* **2022**, *82*, 2900–2911.e7. [[CrossRef](#)] [[PubMed](#)]
104. Vieira de Souza, E.; Bookout, A.L.; Barnes, C.A.; Miller, B.; Machado, P.; Basso, L.A.; Bizarro, C.V.; Saghatelian, A. Rp3: Ribosome Profiling-Assisted Proteogenomics Improves Coverage and Confidence during Microprotein Discovery. *Nat. Commun.* **2024**, *15*, 6839. [[CrossRef](#)] [[PubMed](#)]
105. Nelson, B.R.; Makarewich, C.A.; Anderson, D.M.; Winders, B.R.; Troupes, C.D.; Wu, F.; Reese, A.L.; McAnally, J.R.; Chen, X.; Kavalali, E.T.; et al. A Peptide Encoded by a Transcript Annotated as Long Noncoding RNA Enhances SERCA Activity in Muscle. *Science* **2016**, *351*, 271–275. [[CrossRef](#)]
106. Ma, J.; Ward, C.C.; Jungreis, I.; Slavoff, S.A.; Schwaid, A.G.; Neveu, J.; Budnik, B.A.; Kellis, M.; Saghatelian, A. Discovery of Human sORF-Encoded Polypeptides (SEPs) in Cell Lines and Tissue. *J. Proteome Res.* **2014**, *13*, 1757–1765. [[CrossRef](#)]
107. Fesenko, I.; Shabalina, S.A.; Mamaeva, A.; Knyazev, A.; Glushkevich, A.; Lyapina, I.; Ziganshin, R.; Kovalchuk, S.; Kharlampieva, D.; Lazarev, V.; et al. A Vast Pool of Lineage-Specific Microproteins Encoded by Long Non-Coding RNAs in Plants. *Nucleic Acids Res.* **2021**, *49*, 10328–10346. [[CrossRef](#)]
108. Verheyen, E.M. The Power of *Drosophila* in Modeling Human Disease Mechanisms. *Dis. Models Mech.* **2022**, *15*, dmm049549. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.