

**Gene expression divergence  
during *Drosophila* head development  
on single nuclei resolution**

Dissertation

for the award of the degree

*"Doctor rerum naturalium" (Dr.rer.nat.)*

of the Georg-August-Universität Göttingen

within the doctoral program

*International Max Planck Research School for*

*Genome Science*

of the Georg-August University School of Science (GAUSS)

submitted by

**Gordon Wiegler**

from Hannover, Germany

Göttingen, September 2022

Thesis Committee

**Dr. Nico Posnien (Supervisor)**

Dept. of Developmental Biology, Johann-Friedrich-Blumenbach-Institute of  
Zoology and Anthropology, Georg-August-University Göttingen

**Prof. Argyris Papantonis**

Institute for Pathology, Universitätsmedizin Göttingen

**Dr. Johannes Söding**

Quantitative and Computational Biology, Max-Planck Institute For  
Multidisciplinary Sciences

Members of the Examination Board

Reviewer: Dr. Nico Posnien

Second Reviewer: Prof. Argyris Papantonis

Further members of the Examination Board:

**Prof. Daniel J. Jackson**

Dept. of Geobiology, GZMB, Georg-August-University Göttingen

**Prof. Gregor Bucher**

Dept. of Evolutionary Developmental Genetics, Geoscience Centre, Georg-  
August-University Göttingen

**Prof. Jochen Rink**

Tissue Dynamics and Regeneration, Max-Planck Institute For Multidisciplinary  
Sciences

Date of the oral examination: 04.11.2022

# Declaration

I herewith declare that I prepared the Dissertation

“Gene expression divergence during *Drosophila* head development on single nuclei resolution”

on my own and with no other sources and aids than quoted.

---

Gordon Wiegler

Göttingen, September 14<sup>th</sup>, 2022

## I. Abbreviations

|          |                                |
|----------|--------------------------------|
| AEL      | After egg laying               |
| DEG      | Differentially expressed gene  |
| EAD      | Eye-antennal imaginal disc     |
| MF       | Morphogenetic furrow           |
| Mel      | <i>Drosophila melanogaster</i> |
| Mau      | <i>Drosophila mauritiana</i>   |
| Sim      | <i>Drosophila simulans</i>     |
| scRNAseq | Single-cell RNA-sequencing     |
| snRNAseq | Single nuclei RNA-sequencing   |
| SNP      | Single nucleotide polymorphism |
| TF       | Transcription factor           |
| QTL      | Quantitative trait locus       |

## II. Table of Contents

|          |  |           |
|----------|--|-----------|
| <b>1</b> | <b>Summary .....</b>   | <b>1</b>  |
| <b>2</b> | <b>Introduction.....</b>   | <b>2</b>  |
| 2.1      | Evolution of organ morphology.....   | 2         |
| 2.2      | The <i>Drosophila</i> compound eye as model to study morphological evolution .....   | 3         |
| 2.2.1    | Compound eye development in <i>D. melanogaster</i> .....   | 3         |
| 2.2.2    | Natural variation in eye size and head morphology .....  | 7         |
| 2.3      | Regulatory processes influencing adult eye size .....  | 8         |
| 2.4      | Single-cell RNA-sequencing as a tool for studying cell population dynamics.....  | 10        |
| 2.5      | Project aims and overview.....   | 12        |
| <b>3</b> | <b>Results.....</b>  | <b>15</b> |
| 3.1      | Single-nuclei RNA-sequencing is a more practical way of gathering expression data from low amounts of input material .....                       | 15        |
| 3.1.1    | Tissue dissociation for scRNAseq with low amount of input material .....   | 15        |
| 3.1.2    | scRNAseq reveals relevant cells and a major impact of heat shock and ribosomal genes .....   | 17        |
| 3.1.3    | Cryo-preservation of imaginal discs for efficient isolation of single nuclei.....  | 19        |
| 3.1.4    | snRNAseq identifies eye-antennal disc gene expression and reduces technical biases...  | 21        |
| 3.1.5    | scRNAseq and snRNAseq capture relevant cell types and show differences cell type composition .....   | 23        |
| 3.2      | Data validation.....   | 25        |
| 3.2.1    | Accuracy of staged EAD .....   | 27        |
| 3.2.2    | Single-Nuclei RNA-sequencing data of 120 h old <i>D. melanogaster</i> EAD is comparable to bulk RNA sequencing data of the same tissue type..... | 29        |
| 3.3      | All major cell types of the <i>Drosophila</i> EAD can be identified using single-nuclei RNA-sequencing .....                                     | 30        |
| 3.3.1    | Temporal ordering of biological samples is faithfully captured by single-nuclei RNA sequencing .....   | 39        |
| 3.4      | Interspecies comparison.....   | 41        |
| 3.4.1    | Divergence in cell type composition between <i>D. melanogaster</i> and <i>D. mauritiana</i> .....  | 42        |
| 3.4.2    | Divergence in cell type composition between <i>D. simulans</i> and <i>D. mauritiana</i> .....  | 58        |

|          |   |           |
|----------|---|-----------|
| <b>4</b> | <b>Discussion .....</b>   | <b>78</b> |
| 4.1      | Single-nuclei sequencing from low amounts of input tissue .....   | 78        |
| 4.2      | Annotation of cell clusters in three <i>Drosophila</i> species .....  | 79        |
| 4.1      | Complex developmental information is preserved in single-nuclei sequencing data .....   | 81        |
| 4.2      | Core cell types of the eye primordium show the highest diversity in cell population sizes within the EAD .....                            | 82        |
| 4.3      | Development of retinal cell types in <i>D. mauritiana</i> is characterized by downregulated genes related to the semaphorin pathway ..... | 83        |
| 4.4      | Development of retinal cell types in <i>D. mauritiana</i> is characterized by upregulated genes related to the ecdysone pathway .....     | 85        |
| 4.5      | Other major genes driving interspecies differences between <i>D. melanogaster</i> and <i>D. mauritiana</i> .....                          | 86        |
| 4.6      | Developmental differences between <i>D. mauritiana</i> and <i>D. simulans</i> .....   | 86        |
| 4.6.1    | Genes upregulated in <i>D. mauritiana</i> .....   | 86        |
| 4.6.2    | Genes upregulated in <i>D. simulans</i> .....   | 87        |
| 4.7      | Temporal snRNAseq datasets of non-model species as a resource for developmental studies .....   | 88        |
| <b>5</b> | <b>Material and Methods .....</b>   | <b>91</b> |
| 5.1      | Experimental Setup .....  | 91        |
| 5.2      | Fly husbandry and larval dissections .....  | 91        |
| 5.3      | Dissociation of larval eye-antennal imaginal discs .....  | 92        |
| 5.3.1    | Dissociation of EAD to obtain single-cell suspensions for scRNAseq .....  | 92        |
| 5.3.2    | Dissociation of EAD to obtain single-nuclei suspensions for snRNAseq .....  | 93        |
| 5.1      | Image based assessment of developmental stages .....  | 94        |
| 5.2      | Library preparation and 10x Genomics sequencing .....   | 95        |
| 5.3      | Genome annotation transfer and read mapping .....   | 96        |
| 5.4      | Data Analysis .....   | 96        |
| 5.4.1    | Preprocessing and general analyses .....  | 98        |
| 5.4.2    | Comparison of scRNAseq and snRNAseq .....   | 99        |
| 5.4.3    | Comparison of snRNAseq data and bulk RNA-sequencing data .....  | 99        |
| 5.4.4    | Differential gene expression analyses .....   | 100       |

|          |  |            |
|----------|--|------------|
| 5.4.5    | Standardization of cluster identity annotations..... | 101        |
| <b>6</b> | <b>Contributions .....</b>                           | <b>105</b> |
| <b>7</b> | <b>Acknowledgments .....</b>                         | <b>106</b> |
| <b>8</b> | <b>Supplementary Material .....</b>                  | <b>107</b> |
| <b>9</b> | <b>References.....</b>                               | <b>123</b> |

# 1 Summary

In this work, I optimized single-nuclei and single-cell RNA sequencing protocols for working with small amounts of input tissue. Using one of these protocols, I generated 14 single-nuclei RNA expression datasets. These data cover up to five different timepoints of *Drosophila* eye-antennal imaginal discs during late larval development. These timepoints span meaningful biological events from cell growth and proliferation to differentiation in *Drosophila* compound eye development. The datasets include data for the closely related non-model species, *D. mauritiana* and *D. simulans*. I show, that temporal information of eye-antennal imaginal disc development is preserved in these datasets. With this, I hope to contribute a valuable resource for further developmental and evolutionary studies of the *Drosophila* head. Using this data, I characterized developmental changes in cell population sizes and gene expression patterns between the different species, which is a crucial time for the development of the compound eye. Beyond general patterns, I find divergence in cell population sizes between *D. mauritiana* and *D. melanogaster* or *D. simulans*, respectively. I aimed to identify core regulating transcription factors within the different cell types of the eye-antennal imaginal disc. Furthermore, I subset the data to explore differences in gene expression between these species within core retinal cell types. By comparing the results of these approaches, I identified several genes which present themselves as promising candidates for explaining the evolutionary changes in the morphology of the eye.

## 2 Introduction

### 2.1 Evolution of organ morphology

Over the course of evolution, multicellular organisms evolved a great variety of structures in different sizes and shapes. These differences are often the result of adaptations to their respective environments. The eye of falconiformes is renowned for its high visual acuity (Fox et al. 1976; Snyder and Miller 1978) and possesses specific adaptations for reducing glare in high lighting conditions (Miller and Stegmann 1991). On the other hand, many animals living in perpetual darkness have only strongly reduced or no eyes at all, such as the cave fish *Astyanax mexicanus* (Jeffery 2001). These adaptations reflect variations in the functionality of the same organ between different animals. Therefore, understanding the processes underlying morphological diversity are highly informative about specific anatomical development and evolutionary processes in general.

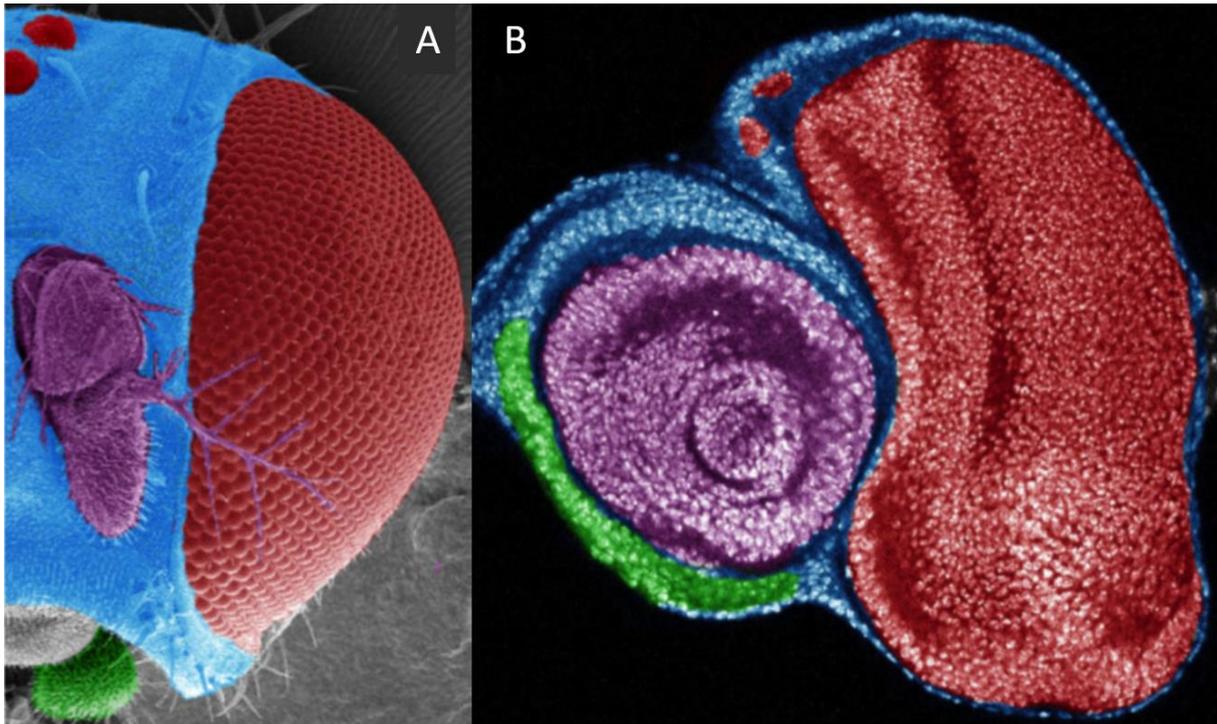
Quantitative genetics and sequencing methods these days allow revealing the genetic and molecular processes underlying morphological diversity. Single homeotic genes, such as *antennapedia*, have been found which can drastically alter the identity of whole body segments (Struhl 1981). However, many morphological traits vary quantitatively, and their genetic basis is complex, often with many genomic loci contributing. In these cases, it is hard to identify individual genes and more importantly, often it is not easy to link the genomic loci to biological processes. To contribute to a better understanding of the mechanisms underlying morphological evolution, I study the molecular and developmental basis of eye size variation in *Drosophila*. The fruit fly *Drosophila melanogaster* is continuing to be studied as a model organism for over a hundred years (Morgan 1910, 1911). This species is very easy to keep in large numbers and it has a very short generational cycle. Due to these advantages, a plethora of molecular tools is established in this organism. Researchers have been taking advantage of this and as such, *D. melanogaster* has been contributing greatly to our understanding of developmental processes in general (Nüsslein-Volhard and Wieschaus 1980; Lewis 2004). In this wake, it has been used to study eye development and its genetic basis for almost as long (Chen 1929). Development and the regulation and expression of genes regulating development are biological processes that have been shown to be highly informative to better characterize morphological diversity. Adult morphology is mostly shaped by these processes and several studies are looking to better understand the connection of development and adult morphology within individual species (Salazar-Ciudad and Jernvall 2010) and between species (Mallarino et al. 2012). Some

developmental processes even continue throughout adult life, such as the continued creation of neurons in adult mammalian brains, which is suggested to be a continuous process from embryogenesis onward (Berg et al. 2019). Gene expression is highly variable over the course of development even within individual tissues (Strober et al. 2019). In *Drosophila*, the genetic origin of many traits could already be linked to genomic loci (Everett et al. 2020).

## 2.2 The *Drosophila* compound eye as model to study morphological evolution

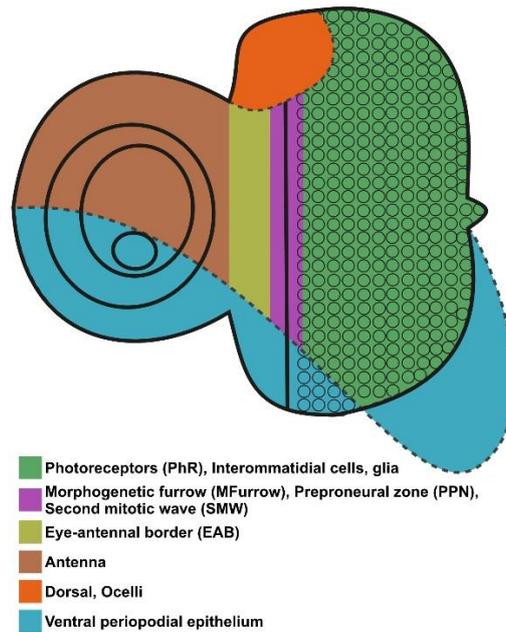
### 2.2.1 Compound eye development in *D. melanogaster*

For many species, the eye is the primary sensory organ. From basic eye spots consisting only of individual or few photosensitive cells, a large number of eye types have evolved, from simple eyes consisting of only single-photoreceptors as in some planarians, to lens eyes in vertebrates, including humans, or compound eyes in insects (Gehring 2014). The compound eye of insects is composed of repeating units of singlet eyes, the so-called ommatidia. These ommatidia are ordered in a repeating hexagonal pattern over the whole eye. Each ommatidium itself consist of eight photoreceptor cells and several supporting cells. These include the cone cells which produce the lens, pigment cells, which serve to isolate the photoreceptors from scattered light and interommatidial cells which form the borders of each ommatidium (Pichaud and Casares 2022). In addition to the primary eyes, insects possess three secondary eyes located centrally on the dorsal head called ocelli. Our current understanding of the developmental and molecular processes involved in the formation of such a sophisticated organ is predominantly based on work done in the fruit fly *Drosophila melanogaster*.



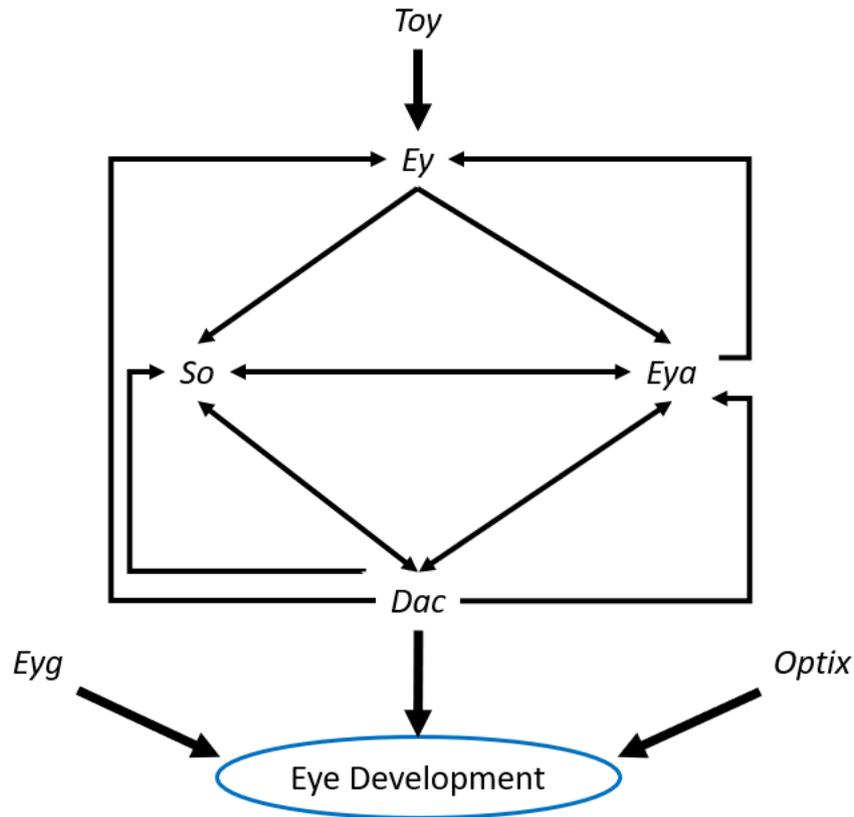
**Figure 1:** The adult head of *Drosophila* and its corresponding precursor regions in the eye-antennal imaginal disc (**B**). Red: Compound eye and ocelli (**A**). Blue: Head capsule (**A**). Purple: Antenna (**A**). Green: Maxillary palp (**A**). The corresponding larval structures are shown in (**B**). Adapted from (Casares and McGregor 2021)

In *D. melanogaster*, as in all holometabolous insects, many external structures of the adult animal, such as legs, wings, halteres, genitals and most of the head including the eyes (Figure 1A), develop from epithelial precursor tissues during larval stages, the so-called imaginal discs (Haynie and Bryant 1986). The imaginal disc that gives rise to many external head structures, most prominently the eyes (Figure 1A), is called the eye-antennal imaginal disc (EAD, Figure 1B). These discs are half-cup shaped and located in the larva between the mouth on the anterior side and the brain on the posterior. The disc is separated into a retinal disc and an antennal disc. The cells of the retinal part ultimately give rise to the compound eye and the ocelli. The antennal disc everts during pupal stages to give rise to the antenna and the maxillary palp. Both parts of the EAD contribute to the head capsule (Figure 1) (Chen, 1929; Haynie, 1986). Eye and head development in *D. melanogaster* is well-studied and accessible for functional assays, making it an excellent system to unravel developmental and molecular processes underlying the regulation of morphological traits.



**Figure 2: Schematic overview of major cell types of the late third instar *D. melanogaster* eye-antennal imaginal disc.** The different areas within the EAD can be distinguished by their own unique gene expression patterns and cell population dynamics. Apart from tissues corresponding to adult structures, cells characterized by specific developmental processes can also be identified, such as cells of the morphogenetic furrow (purple).

The eye-antennal disc itself is a bilayered epithelium which originates from a small group of 20-30 cells set aside during embryogenesis and consists of the disc proper and the peripodial epithelium (Figure 2). During larval development, the cells of the EAD proliferate until it consists of approximately 44,000 cells prior to pupation (Kumar 2018). Already at the L2 larval stage it is possible to visually distinguish the anterior antennal disc from the posterior eye disc. Both discs can also be genetically distinguished by the expression of *cut* (*ct*) and *distalless* (*dll*) in the antennal disc and the expression of *eyeless* (*ey*) and *twin of eyeless* (*toy*) in the eye disc. Growth, patterning and final differentiation are highly coordinated processes in the EAD governed by an intricate gene regulatory network, the so-called retinal determination network (Kumar, 2009; Jang, 2003; Treisman, 2013) (Figure 3). The loss of any of the core genes of this networks causes a severe reduction of retinal tissue. The overexpression of many of these genes can, conversely, cause the formation of ectopic eye tissue.



**Figure 3: Key member genes of the *Drosophila* retinal determination network.** *Toy*: twin of eyeless, *Ey*: eyeless, *So*: sine oculis, *Eya*: eyes absent, *Dac*: dachshund, *Eyg*: Eyegone. This scheme only depicts positive interactions, the entirety of known interactions is more complex.

The main components of the retinal determination network are *twin of eyeless* (*Toy*), *eyeless* (*Ey*), *sine oculis* (*So*), *eyes absent* (*Eya*), and *dachshund* (*Dac*) (Figure 3). Retinal determination begins with the activation of *toy*, which in turn triggers the expression of *ey*. *Ey* establishes the activity of the retinal determination network by activating the expression of its downstream members, *so* and *eya*. These genes directly or indirectly (by causing the expression of *dac*) maintain this network over larval development through a positive feedback loop including *ey*. Other genes, such as *eyegone* (*Eyg*) and *Optix* contribute to eye development independently of this core network.

Once the retinal part of the EAD is determined, differentiation is initiated in the posterior equator region at the L2/L3 stage transition. Differentiation is a dynamic process marked by a visible indentation, the morphogenetic furrow (MF), that moves from the posterior towards the anterior side of the eye primordium. This movement is triggered by the unrepressed expression of *hedgehog* (*hh*), which in turn activates *decapentaplegic* (*dpp*) expression. In the early EAD, *hh* is repressed by the gene *wingless* (*wg*) and it is thought that *hh* expression is permitted when the eye primordium reaches a certain size threshold, since *wg* is expressed and diffuses from

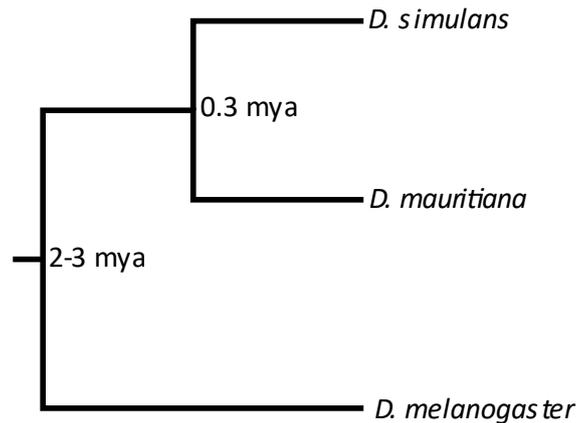
the dorsal-anterior end of the eye primordium. The MF leaves differentiating R8 photoreceptor cells marked by *atonal* (*ato*) expression in its wake. Other photoreceptor cells and support cells that build each ommatidium are subsequently recruited from neighboring cells. Once the first cells are assigned photoreceptor fate, the movement of the MF is propagated by these cells which in turn start expressing *hh* themselves. The movement of the MF comes to a halt once the pool of undifferentiated progenitor cells anterior to the MF is depleted. At the end of larval development, most ommatidial cell clusters are defined, and morphogenesis of the eye and head capsule proceeds during pupal development.

### 2.2.2 Natural variation in eye size and head morphology

Eye morphology in insects may adapt to varying needs of the organism. Dragonflies are known to possess large eyes with more than 30,000 ommatidia. As predators relying on vision, they may require a visual sense with a high resolution. The period of activity on an animal might also influence the morphology of the eye. Finally, some animals do not rely on eyes as a primary sensory organ, and eyes may be reduced in favor of other sensory organs such as antennae. In fact, an extensive study revealed morphological variation in resource allocation of visual and olfactory sensory systems between 62 species of *Drosophila* (Keeseey et al. 2019). Other studies report a general tradeoff between eye- and face size in different *Drosophila* species (Posnien et al. 2012).

Quantitative studies in different *Drosophila* species revealed pervasive natural variation in adult eye size and head morphology, such as in the *virilis* group of *Drosophila* (Reis et al. 2020). These studies demonstrate, that within groups sharing the same basic structure of the eye such as insects, there can be strong variation in eye shape and size proportions.

From preliminary studies it is known that differences in eye and head shape also exist between both *D. melanogaster* as well as *D. simulans* compared to *D. mauritiana* (Posnien et al. 2012; Arif et al. 2013; Gaspar et al. 2020). The most prominent difference is the larger eye size in *D. mauritiana* compared to each of the other two species.



**Figure 4: Phylogeny of selected *Drosophila* Species.** Mya: Million years ago. *D. mauritiana* and *D. simulans* are more closely related to each other than either is to *D. melanogaster*.

*D. simulans* and *D. mauritiana* are more closely related to each other than to *D. melanogaster*, from which they diverged approximately 2 to 3 million years ago (Figure 4). They separated from each other approximately 0.3 million years ago. Both of these species promise to show intriguing insights into the development of morphological diversity as the reason for the eye size variation is different. Between *D. simulans* and *D. mauritiana*, the size difference is explained largely by the larger ommatidia size in *D. mauritiana* (Posnien et al. 2012; Gaspar et al. 2020; Torres-Oliva et al. 2021). The emergent effect of individual larger ommatidia can propagate over the whole eye and result in a major quantitative size difference. Between *D. melanogaster* and *D. mauritiana* on the other hand, the observed eye size difference can be explained mainly by the smaller number of ommatidia in *D. melanogaster*.

In the light of extensive knowledge about eye development and the pervasive natural variation in eye size and shape, *Drosophila* compound eyes are an excellent model to study the molecular and developmental basis of the morphological diversity of complex organs.

### 2.3 Regulatory processes influencing adult eye size

While the retinal determination network contains key regulators which are responsible for eye development and growth, it is largely elusive how subtle morphological differences arise during evolution and are controlled within an organism on a genetic level. To date, a direct translation of the genomic information to a specific phenotype is not possible yet. However, the main stage for evolutionary variation can be on different levels. Comparative interspecies studies have proven to be a powerful tool to gain insight into development and evolution of organ

morphology. When studying the genetic basis of a trait of interest, closely related species which show a divergent morphology in this trait are well suited models. In those organisms, genetic background variation can be expected to be low due to their shared evolutionary history. As such, the likelihood of discovering meaningful genetic regulators for this trait is higher than when comparing distantly related species. Two sister species of *Drosophila melanogaster*, *D. mauritiana* and *D. simulans*, present themselves as suitable models for studying variation in organ morphology in the *Drosophila* group as they are known to show quantitative morphological differences in eye- and head morphology.

Studies have successfully identified genomic regions influencing eye size. For instance, a set of genomic inversions containing hundreds of genes was linked to eye size variation within the *virilis* group of *Drosophila* (Reis et al. 2020). These studies have found genetic origins of causes for known tradeoffs between eye size and interocular distance within different strains of *D. melanogaster* (Norry and Gomez 2017). And similar studies have found quantitative trait loci (QTL) not only within *D. melanogaster* strains, but also in other species such as *D. simulans*. This has revealed that the genetic causes underlying the variation in eye morphology within a species can differ between species (Gaspar et al. 2020).

The genetic basis for differences between two different species, such as *D. simulans* and *D. mauritiana*, appear to be highly complex. By applying QTL mapping, it was found that the loci which are responsible for head trait differences between these species are spread over different genomic regions, and individual loci may influence different facial traits independently (Arif et al. 2013).

Only few studies attempted to link genetic variants or molecular processes to potential developmental processes underlying eye size variation (Posnien et al. 2012). For intra-specific variation in *D. melanogaster*, a single nucleotide polymorphism (SNP) in the gene *cut* (*ct*) was identified that affects the early subdivision of the EAD and is thus functionally linked to variation in eye size via a trade-off with the interocular cuticle (Ramaekers et al. 2019). Natural variation in the expression of *pannier* (*pnr*), a gene that is involved in early dorsal-ventral axis patterning of the EAD (Singh and Choi 2003) and later regulates retinal vs. head cuticle fate (Oros et al. 2010), has been linked to differences in eye size and head morphology between *D. melanogaster* and *D. mauritiana* (Buchberger et al. 2021). Besides these processes, variation in a variety of processes during EAD development might influence adult eye size (reviewed in Casares and McGregor, 2021). The number of progenitor cells of the eye disc as well as their proliferation rate before the onset of the MF could also affect final eye size. The distance the

morphogenetic furrow travels towards the anterior determines the size of the share of the eye primordium that is assigned to retinal fates. The further anterior it travels, the more retinal cells will develop. Additionally the speed of the movement of the MF may influence eye size as the undifferentiated cells anterior to the furrow are in a proliferative state until the morphogenetic furrow passed them (Casares and McGregor 2021).

Despite extensive datasets aiming at revealing the genetic basis of eye size variation in *Drosophila*, we are still lacking a proper mechanistic understanding of the developmental processes involved. One reason for this is that identified positional candidate genes based on quantitative genetics methods may not easily provide hypotheses about their potential expression and function. Comparative genome wide expression studies are an excellent approach to provide potential functional context to positional candidate genes (e.g. (Torres-Oliva et al. 2021)) or serve as technique to identify candidate genes (e.g. Buchberger et al., 2021).

In previous attempts, bulk sequencing of EAD shed first light on gene expression during development in these species (Torres-Oliva et al. 2018). However, these studies have limitations. For instance, transcriptional changes in subpopulations of cells might be invisible in a large bulk data set or might be compensated with opposite transcriptional changes in a different subpopulation of cells. This was shown in divergence in face morphology between *D. melanogaster* and *D. mauritiana* which is influenced by the transcription factor Pannier (Pnr) and its co-factor U-shaped (Ush), specifically their interaction in the dorsal area of the EAD (Buchberger et al. 2021). Differences in cell type proportions and cell population sizes in general can also be expected to be established at the level of transcription, interpreting transcriptomic data in that direction however is difficult.

These processes are highly localized in different regions of the EAD, or possibly within distinct cell types in each ommatidium, and eye morphology could be altered by influencing any of these processes and the behavior of individual cell types. They show that gene expression strongly depends on biological context. The spatial and temporal expression of developmental genes, for example, is tightly regulated throughout development resulting in tissue- and even cell type specific expression profiles. In the light of this context-dependent gene regulation, it is becoming increasingly relevant to study gene expression on a cellular level.

## 2.4 Single-cell RNA-sequencing as a tool for studying cell population dynamics

Nowadays, multiple sequencing technologies are available allowing to quantitatively analyze

the messenger RNA content of single cells (Tang et al. 2009). Single-cell RNA sequencing (scRNAseq) has been proven powerful to reveal the cell type composition of complex tissues or organs in model organisms, such as the fruit fly *Drosophila melanogaster* (Li et al. 2022; Davie et al. 2018), the nematode *Caenorhabditis elegans* (Cao et al. 2017) and mouse (The Tabula Muris Consortium 2020). Also, biological processes, such as development of the optic lobe of the fly brain (Özel et al. 2021), cell-cell communication in tumors (Kumar et al. 2018) and immunity (Stubington et al. 2017; Shalek et al. 2013) have been successfully studied. Since the analysis of scRNAseq data does not require prior knowledge of the tissue of interest, this method is exceptionally well-suited to study the cell type composition of emerging model organisms, such as sponges (Musser et al. 2021), the cnidaria *Nematostella vectensis* (Sebé-Pedrós et al. 2018), *Hydra vulgaris* (Klimovich et al. 2020) and *Clytia hemisphaerica* (Chari et al. 2021), the annelid *Platynereis dumerilii* (Vergara et al. 2017) and the planarian *Schmidtea mediterranea* (Plass et al. 2018; Fincher et al. 2018), the ant *Harpegnathos saltator* (Sheng et al. 2020) and multiple vertebrates (Tosches et al. 2018; Chen et al. 2021). Comparative studies have been performed to reveal divergent and conserved aspects of the motor cortex in human, marmoset, and mouse (Bakken et al. 2021) and during early embryonic development in pigs, humans and cynomolgus monkeys (Liu et al. 2021).

scRNAseq protocols are composed of the following key steps (Kolodziejczyk et al. 2015; Wu et al. 2017): The tissue of interest is dissociated, and individual cells are captured either in microwell plates (Islam et al. 2011) or in micro-droplets (Macosko et al. 2015). Individual captured cells are lysed in the microwell, or droplet and the released polyadenylated RNA (mRNA) is captured using poly-T oligos. The mRNA is reverse transcribed into complementary DNA (cDNA) and cell and molecule specific barcodes are added. Illumina sequencing adapters are ligated to barcoded cDNA molecules, and the cDNA is released and amplified by PCR. The amplified libraries are eventually sequenced using next generation sequencing technologies (e.g. Illumina).

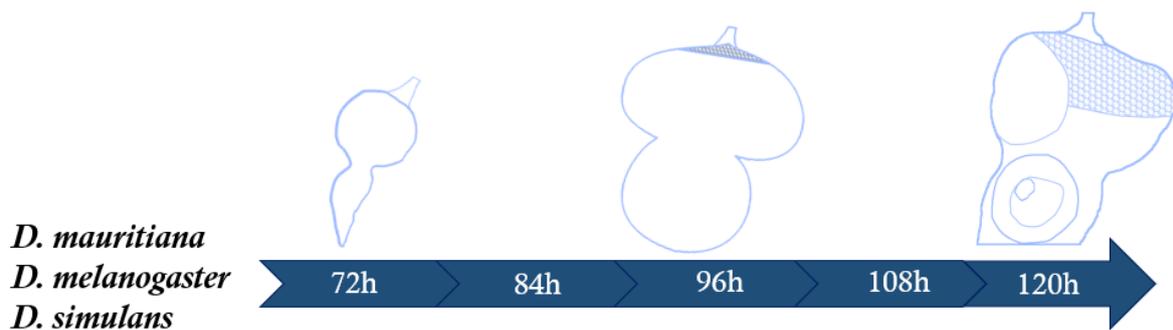
While current scRNAseq technologies allow sequencing up to 10,000 cells in one run (Svensson et al. 2018), many more cells are needed as input material. For instance, mechanical stress during dissociation of complex tissue leads to increased cell death (Khan et al. 2016). Additionally, harsh dissociation conditions using enzymes, such as Trypsin, contribute to cell damage (Hodges et al. 1973; Snow and Allen 1970), altered gene expression (Snow and Allen 1970; Huang et al. 2010) and RNA degradation (Vrtačnik et al. 2014). Due to the high cell loss during dissociation, current scRNAseq methods are limited if small tissue samples are analyzed

because this might require collecting tissue from multiple animals to obtain sufficient starting material. EAD are such small tissue, and even though it is possible to pool several EADs, RNA degradation after dissection limits the amount of input material that can be collected. While gene expression in late eye-antennal disc have been studied at single-cell resolution (Ariss et al. 2018; Bravo González-Blas et al. 2020), earlier stages are less accessible due to low cell numbers.

## 2.5 Project aims and overview

The major aim of my project is to contribute to a better understanding of the molecular and developmental processes underlying morphological evolution. As a model I study compound eye size variation among the three closely related *Drosophila* species *D. melanogaster*, *D. simulans* and *D. mauritiana* for the following reasons: 1) Compound eye development is exceptionally well studied in *D. melanogaster* and therefore it may be possible to link candidate genes identified in my project to known developmental processes during eye-antennal imaginal disc development. 2) Previous work revealed extensive natural variation in eye size and head shape among these three species with *D. mauritiana* having the largest eyes. Eye size differences between *D. mauritiana* and *D. melanogaster* are due to variation in ommatidia number, while *D. mauritiana* and *D. simulans* eye size differences are caused by variation in ommatidia size. 3) Despite the quantitative differences in eye size, the close relationship of the three sister species suggests that general processes involved in eye-antennal disc development should be conserved.

Previous studies based on comparative gene expression during eye-antennal disc development successfully revealed potential candidate genes showing interspecific differences in expression and regulation among these three species. However, as these data are based on sequencing entire eye-antennal discs (i.e. bulk sequencing), it has remained challenging to link candidate genes to specific cellular and developmental processes. Therefore, I aim to gain further insights into cell population dynamics during eye development and to identify potential differences in gene expression and gene regulation that contribute to species specific eye morphology. To this end, I applied comparative single-cell (scRNAseq) and single-nuclei (snRNAseq) RNA sequencing techniques to assess gene expression in developing eye-antennal discs on the level of individual cells. This data was generated for five stages throughout third larval instar development because the main events of eye development, such as disc growth, patterning and photoreceptor cell differentiation happen during this time (Figure 5).



**Figure 5: Scheme of EAD development in *Drosophila* L3 larvae.** The illustrations schematically depict the development with the formation of visible structures of the EAD. The depicted timepoints are timepoints from which I collected pooled EAD samples. Time is measured in hours after egg laying (AEL). I took one sample per each of the three depicted species.

The specific aims of my thesis were:

- I. Single cell RNA sequencing methods generally require large tissue samples to obtain sufficient cells for sequencing. As eye-antennal discs are composed of only 44,000 cells at the late third instar stage and much less during earlier stages targeted here, I first aimed at establishing an efficient protocol for small input samples. To this end, I comprehensively evaluated different dissociation, tissue preservation and sequencing methods. I will present this work in chapter 3.1 and 4.1, which are adaptations from the manuscript “Tissue dissociation for single-cell and single-nuclei RNA sequencing for low amounts of input material” which is currently in revision at *Frontiers in Zoology* (DOI: <https://doi.org/10.21203/rs.3.rs-1623239/v1>).
- II. Based on my established snRNAseq protocol, I generated 14 snRNAseq samples covering five developmental stages for three species. The second aim was to validate these datasets. Specifically, I compared the relative expression of genes within the data to bulk RNA sequencing datasets at 120 h AEL to identify potential biases in read count proportions within the data.
- III. Clustering and annotation of clusters within all datasets to find standardized annotations that can describe the most comparable cell types between the three different species and different timepoints based on published marker genes. I found expected cell populations in expected biological ordering based on the known biology of the EAD in non-model *Drosophila* species.
- IV. Comparison of EAD development within the three species. By gathering data for up to five different timepoints in 12 h steps (Figure 5), I was able to identify developmental trajectories of different compartments of the EAD within the different species.

V. Finally, I aimed to use these datasets to identify developmental dynamics that may drive morphological variation in head- and eye size between different *Drosophila* species on the level of cell populations. I identified relative differences in cell numbers of the most relevant cell populations in the developing *Drosophila* EAD between *D. mauritiana* and *D. melanogaster* or *D. simulans*, each at different timepoints. I performed differential gene expression analyses and explored the most impactful TF behind the gene expression patterns within retinal cell types to identify candidate genes driving these differences and find regulons controlling gene expression on a broader scale.

## 3 Results

### 3.1 Single-nuclei RNA-sequencing is a more practical way of gathering expression data from low amounts of input material

Protocols for scRNAseq require large amounts of input materials to gather only a small amount of cells. Since we are working with very small tissue samples, it was necessary to find a protocol which can yield a representative number of single cells (or nuclei) to gather expression data from. Hence, I tested a number of protocols for their applicability. In the following sections, I show that a combination of mechanical and chemical dissociation works best to obtain sufficient and representative cells for single-cell RNA sequencing (scRNAseq). However, I observed artificial expression of stress related genes, which was most likely due to rather harsh dissociation and cell-sorting conditions. As an alternative, I tested different protocols to isolate single nuclei from fresh and frozen tissue and I show that single-nuclei RNA sequencing (snRNAseq) successfully allowed identifying key cell types without the drawback of stress-response. I find differences in the relative contribution of scRNAseq and snRNAseq to common cell types and I discuss the advantages and disadvantages of both methods. This section provides an overview of different single-cell sequencing approaches when accessibility to tissue samples is limited.

#### 3.1.1 Tissue dissociation for scRNAseq with low amount of input material

For RNA sequencing of single cells heterogenous tissue samples need to be dissociated into live and intact cells. Since about 10,000 cells can be analyzed using the 10x Genomics Chromium System and about 50% of input cells are lost throughout the preparatory steps, I first tested different dissociation protocols to obtain about 20,000 cells from entire larval organs or about 30 eye-antennal imaginal discs at 120 h after egg laying (AEL).

The success of different tissue dissociation protocols was evaluated by estimating the ratio of dead and live cells, as well as the final number of live cells. A dead cell staining with Trypan blue is well-established in homogeneous cell suspensions obtained from cell culture (Pappenheimer 1917; Chan et al. 2015). However, I experienced unreliable dead/live cell ratios with our complex cell suspensions, which was most likely due to Trypan blue positive debris. Therefore, I applied a live-dead assay based on propidium iodide (PI) and calcein green/violet to identify dead and live cells, respectively. This method allows enrichment of live cells via

fluorescence activated cell sorting (FACS), which efficiently also removed debris. Note that the combination of PI and calcein violet resulted in the most efficient separation of live and dead cells due to a lower spectral overlap of both dyes during FACS. Sorted cells were examined by fluorescent microscopy to confirm that they were mostly calcein positive and PI negative.

First, I tested purely enzymatic or mechanical dissociation protocols, respectively. Incubation of eye-antennal discs in 10x TrypLE and 2.5 mg/ml Collagenase even for 2 h did not result in single-cell solutions based on visual assessment. Imaginal discs ground with a Dounce homogenizer showed a high proportion of debris and what appeared to be single-nuclei suspensions. Additionally, different attempts resulted in inconsistent dissociation because the low amount of input tissue was barely visible and due to the manual component, it was difficult to balance complete dissociation with the destruction of cells. Based on these observations I reasoned that efficient tissue dissociation required a combination of enzymatic dissociation with gentle mechanical force.

The basic protocol was based on treatment of the tissue with TrypLE and Collagenase on a shaker at 300 rpm with pipet strokes (1000  $\mu$ l pipet tips) during and after the incubation. I varied the following parameters: enzyme concentration (1x and 10x TrypLE; 2.5 mg/ml and 10 mg/ml Collagenase), incubation time (10 – 60 minutes), incubation temperature (37°C and 30°C), number of pipet strokes (5 strokes during the incubation and 17-20 strokes after the incubation) and filtration of the cell suspension (no filter, 20  $\mu$ m and 35  $\mu$ m filters). 1x TrypLE was insufficient to achieve complete dissociation in a timely manner and the addition of 10 mg/ml Collagenase resulted in an increased yield, as well as fewer cell aggregates (visual assessment). Incubation for up to 60 minutes at 30°C resulted in comparable or slightly more live cells compared to a digestion at 37°C. Filtration with a filter of 35  $\mu$ m mesh size did not drastically reduce the proportion of live cells but decreased the amount of debris. The number of pipet strokes after incubation had the highest impact on cell survival with significantly reduced cell survival after more than 17 strokes. I obtained the best results with 16,208 live cells (58 % survival rate) from 28 eye-antennal discs after 60 minutes incubation at 30°C in 10x TrypLE and 10 mg/ml Collagenase and 5 pipet strokes during and 17 pipet strokes after the incubation. RNA extracted from this sample was of high quality and suitable for 10X Genomics scRNAseq. In summary, for low amount of input material, such as < 50 late L3 eye-antennal discs we propose a protocol that combines enzymatic dissociation in conjunction with slight mechanical disruption.

### 3.1.2 scRNAseq reveals relevant cells and a major impact of heat shock and ribosomal genes

Next, we subjected cells obtained after FACS to a 10X Genomics Chromium run to test if the established dissociation protocol resulted in representative cell types expected in the eye-antennal disc. After droplet-based isolation of RNA from individual cells and subsequent Illumina sequencing, we obtained almost 200 million reads from about 14,500 cells with 13,303 reads and 537 genes per cell (Table 1). 12,000 cells showed less than 10 % mitochondrial gene expression (Figure 6A) confirming that we mostly isolated live cells. Among the top ten genes with most variable expression across cells, we found two heat shock related genes (Hsp23 and lncRNA:Hsromega, Figure 7A) and many reads of the scRNAseq dataset mapped to genes coding for heat shock proteins (Figure 6C), suggesting that the 30°C incubation temperature during dissociation or/and the FACS may impose stress on the cells. The distribution of reads also showed a high expression of cytoplasmic genes, such as eEF1alpha1 and eukaryotic elongation factors. Additionally, a lot of genes coding for ribosomal proteins were expressed in our dataset (Figure 6C). The high content of ribosomal genes is expected for scRNAseq because cytoplasmic mRNA is extracted and ribosomal RNAs are known to be very stable. However, they are often considered uninformative.

**Table 1: Summary statistics for the cell- and nuclei dataset.**

| <b>Dataset</b>  | <b>Cell</b> | <b>Nuclei_MDC</b> |
|---|-------------|-------------------|
| <i>Estimated number of cells (# of discs)</i>                           | 14,487 (28) | 9,048 (41)        |
| <i>Median genes per cell</i>  | 537         | 812               |
| <i>Mean reads per cell</i>  | 13,303      | 13,334            |
| <i>Valid barcodes</i>   | 96.10%      | 97.10%            |
| <i>Number of reads</i>  | 192,731,871 | 120,649,741       |
| <i>Fraction of reads in cells</i>                                       | 38.20%      | 73.50%            |
| <i>Total genes detected</i>   | 11,062      | 12,296            |
| <i>Median UMI counts per cell</i>                                       | 1,249       | 1,383             |
| <i>Reads mapped to genome</i>   | 93.20%      | 85.70%            |
| <i>Reads mapped confidently to genome</i>                               | 86.40%      | 84.50%            |
| <i>Reads mapped confidently to intronic regions</i>                     | 2.10%       | 14.70%            |
| <i>Reads mapped confidently to intergenic regions</i>                   | 9.00%       | 0.80%             |
| <i>Reads mapped confidently to exonic regions</i>                       | 75.20%      | 69.00%            |
| <i>Percentage of cells with high mitochondrial read count (&gt;10%)</i> | 14.00%      | 0.02%             |

We performed an unbiased cluster analysis based on variable gene expression to identify major cell types. Among the top four genes that define a certain cluster, I found well-known genes involved in different processes during eye-antennal disc development. For instance, *cut*, which is expressed in antennal tissue of the disc was strongly expressed in cells of clusters 3 and 5, while *homothorax (hth)*, which is broadly expressed in the eye-antennal disc was found in cells of clusters 3, 5 and 7 (Figure 7B). Clusters 16 and 19 were predominantly defined by expression of members of the enhancer of split gene complex (Figure 7B), which are broadly expressed in the dynamic differentiation wave, the so-called morphogenetic furrow (Wurmbach et al. 1999; Wolff and Ready 1993). In the same clusters I found high expression of *nemo (nmo)* that codes for a Nemo-like kinase (Nlk) is broadly expressed in the retinal region of the eye-antennal disc and has been shown to interact with members of the retinal determination network (Braid and Verheyen 2008). In line with previously reported roles in photoreceptor differentiation (Roark et al. 1995; Luo et al. 1990), I observed expression of *amyloid protein precursor-like (Appl)* and *scratch (scrt)* in clusters 1, 8 and 13. Overall, I found major eye-antennal disc marker genes, suggesting that key cell types were identified in accordance with previous scRNAseq data obtained for late L3 eye-antennal discs (Ariss et al. 2018). Gene ontology (GO) term enrichment analyses for marker genes defining each cell cluster (Supplementary Table S 2) further

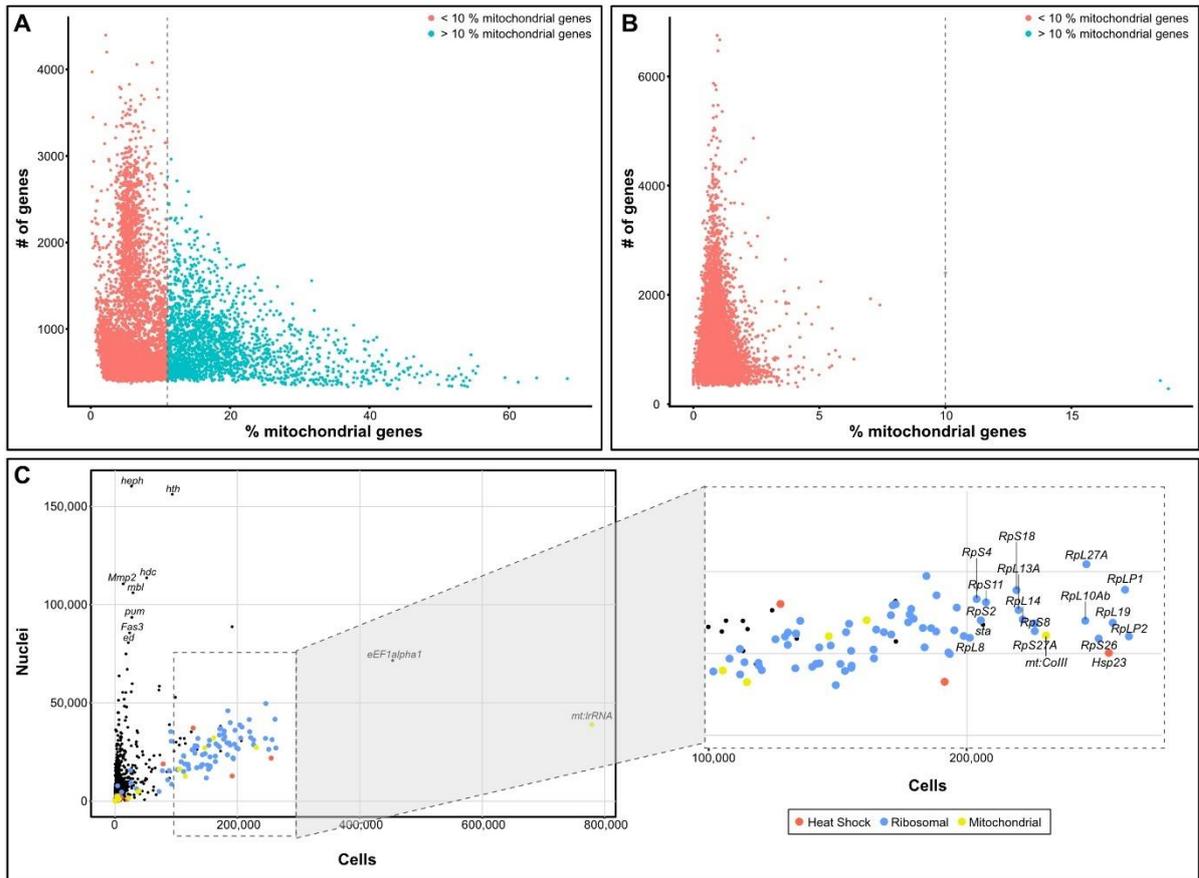
confirmed that cells in each cluster expressed genes involved in relevant biological processes (Supplementary Table S 3).

Besides these relevant biological findings, the potential stress response of the cells was also evident in our cluster analysis because three heat shock genes were among the top four cluster defining genes (*Hsp23*, *Hsp26* and *Hsp68* in clusters 7 and 17; Figure 6B). Those three genes were expressed in most cells of all clusters, and they showed very high expression in cluster 7 (Figure 7B). Moreover, cluster 2 was largely defined by mitochondrial genes (Figure 7B). Note that cluster 0 (Figure 7B) was not clearly defined by a set of marker genes. This could be due to the resolution of the clustering (that was chosen to be comparable to the snRNAseq analysis, see below) as clustering with a lower resolution resulted in clusters with clear marker genes (Supplementary Table S 3).

In summary, our tissue dissociation protocol successfully resulted in a cell suspension containing major cell types of the eye-antennal disc but may pose stress on the cells which is detectable through high expression of heat shock genes. Additionally, the high level of ribosomal genes may introduce a bias during further analysis of such data.

### 3.1.3 Cryo-preservation of imaginal discs for efficient isolation of single nuclei

Our scRNAseq data suggested that the applied dissociation conditions were still stressful for the cells. Additionally, the protocol relies on the processing of fresh tissue samples hampering the analysis of even smaller tissue samples. For instance, eye-antennal discs at the late L3 larval stage contain about 60,000 cells, while discs at the transition from the L2 to the L3 stage are only composed of about 5,000 cells. Therefore, about 12 times more discs are needed to obtain sufficiently high cell numbers for scRNAseq applying our single-cell dissociation protocol. As tissue growth is an integral part of developmental processes, more efficient protocols are needed to harness the full potential of single-cellsequencing methods for developmental biology. To this end, I tested two main approaches: First, I evaluated the use of single nuclei for RNA sequencing (i.e. snRNAseq) as snRNAseq has been shown to result in comparable data, especially for tissue samples that are difficult to dissociate into single cells. Second, I tested the effect of cryo-preservation on the subsequent isolation of single nuclei and RNA integrity as this step allows collecting small tissue samples over time.



**Figure 6: Evaluation of read distribution after scRNAseq and snRNAseq.** (A) Total amount of genes (features) over percentage of mitochondrial reads, per cell each. The dashed line indicates a threshold of 10% of reads attributed to mitochondrial genes. In single-cell RNA sequencing data, approximately 14% of cells show a high (>10%) proportion of mitochondrial gene reads on the total number of reads. (B) Total amount of genes (features) over Percentage of mitochondrial reads, per cell each in single-nuclei RNA sequencing data. The dashed line indicates a threshold of 10% of reads attributed to mitochondrial genes. In most nuclei, only a low percentage of reads is attributed to mitochondrial genes. (C) Number of reads found in nuclei (y-axis) over the number of reads of those same genes found in cells. The basis are normalized datasets. Ribosomal genes are highlighted in blue and heat shock genes in red. Ribosomal genes are defined as genes starting RpS- or RpL-. Heat shock genes are defined as genes encoding heat shock proteins, starting Hsp-.

For the isolation of single nuclei, I tested two main protocols: One protocol suggested by 10X Genomics is based on NP40 as detergent and a small number of centrifugation and pipetting steps (10X Genomics 2021). The other protocol had been established for human heart tissue and is based on using Triton X-100 as a detergent and a variety of RNase inhibitors to preserve RNA in single nuclei (Litvinukova et al. 2018). When 30-50 freshly dissected eye-antennal discs at late L3 stage were used for nuclei isolation, both protocols resulted in more than 20,000 nuclei and extracted RNA was of high quality suitable for snRNAseq (Supplementary Table S 4).

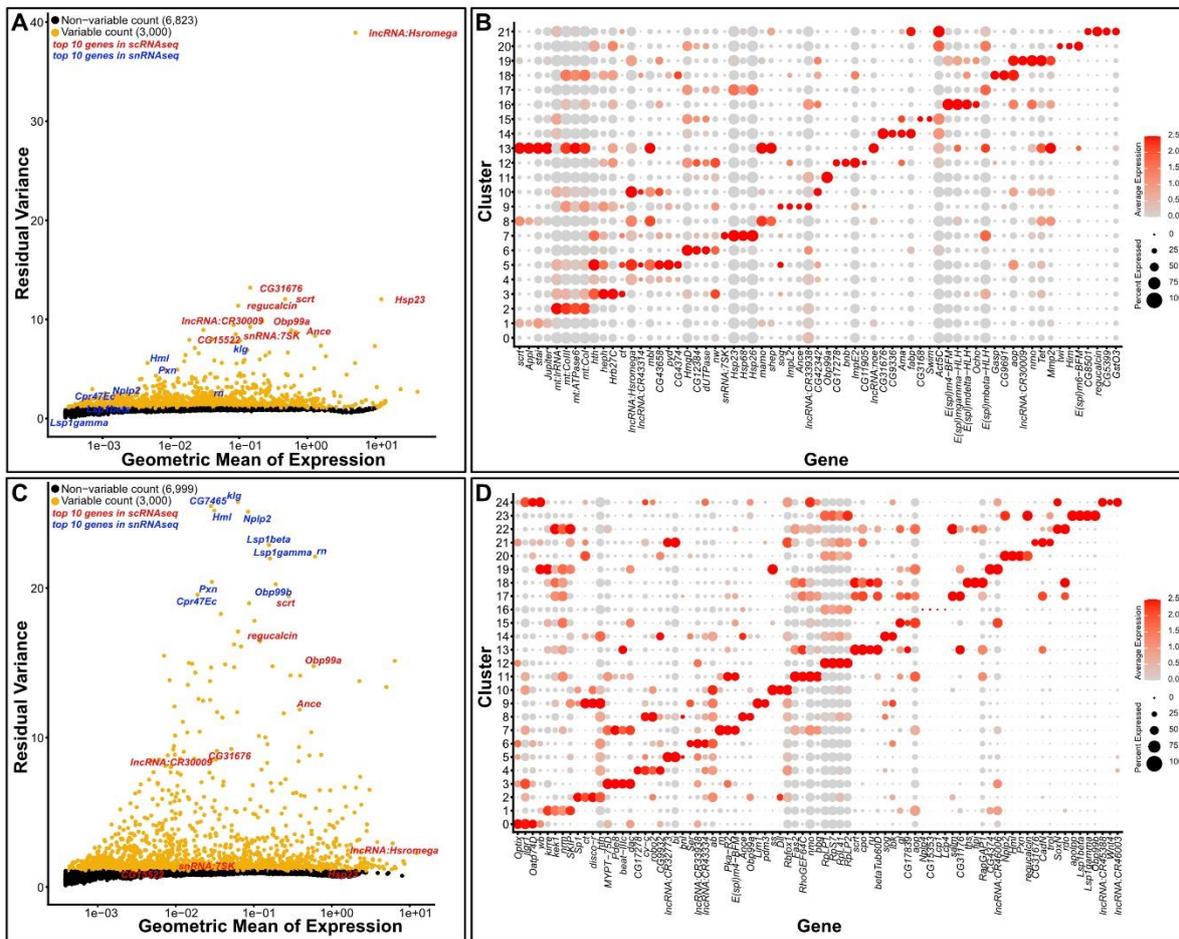
We next dissected imaginal discs, snap-froze them in liquid nitrogen and stored them at  $-80^{\circ}\text{C}$  for at least one day, or up to four weeks. All applied protocols allowed us to isolate more than

20,000 intact nuclei from about 30 cryo-preserved eye-antennal discs. RNA extracted from nuclei isolated with the 10X Genomics protocol resulted in low RNA quality, suggesting a high level of RNA degradation (Supplementary Table S 5, lanes 7 and 8). The addition of citric acid to the dissociation buffer has been shown to preserve RNA integrity in human pancreatic cells. However, the use of citric acid during the 10X Genomics protocol did only marginally improve the quality of RNA extracted from cryo-preserved samples (Supplementary Table S 5, lanes 10 and 11). In contrast, I observed almost no RNA degradation and high RNA quality when I used the protocol that employs RNase inhibitors (Supplementary Table S 5, lanes 1, 2, 4 and 5). RNA integrity was preserved even when eye-antennal discs were thawed for 3.5 h and frozen again prior to nuclei isolation and RNA extraction (Supplementary Table S 4), showing that the use of RNase inhibitors are highly efficient to prevent RNA degradation when processing of cryo-preserved tissue samples. Based on the high yield and the high RNA quality, we conclude that the combination of cryo-preservation and nuclei isolation employing RNase inhibitors is highly efficient to process low input material for snRNAseq.

#### 3.1.4 snRNAseq identifies eye-antennal disc gene expression and reduces technical biases

To test if the nuclei obtained after cryo-preservation are suitable for snRNAseq and represent main cell types of the eye-antennal disc, we subjected the nuclei to a 10X Genomics run to obtain 120 million reads from about 9,000 cells with 13,334 reads and 812 genes per cell (Table 1). For the analysis of the data, we applied the same pipeline and settings as for the scRNAseq dataset, with the exception that intronic reads were included because pre-mRNA is expected in nuclei (Grindberg et al. 2013). Among the 10 genes with most variable expression in the dataset, we found some with known functions and expression in late L3 eye-antennal discs. For instance, the homophilic cell adhesion molecule *Klinglec* (*Klg*) is strongly expressed during R7 photoreceptor development (Butler et al. 1997) and *rotund* (*rn*) that codes for a Kruppel zinc-finger transcription factor is expressed in large parts of the antennal field (St Pierre et al. 2002). Importantly, we did not observe genes associated with heat shock response among the top 10 variable genes (Figure 7C) and no bias of reads originating from heat shock genes was observed (Figure 6C), suggesting that they will not impact subsequent clustering analyses as observed for the scRNAseq data. As expected for snRNAseq data (Wu et al. 2019; Lake et al. 2017; Lake et al. 2018; Bakken et al. 2018), we found only two cells with more than 10 % mitochondrial gene expression (Figure 6B) and fewer reads originating from ribosomal genes (Figure 6C). In fact, we observed 28 % fewer reads of ribosomal genes in the snRNAseq dataset compared to

the scRNAseq data, assuming lower impact on the entire dataset.



**Figure 7: Most variable genes and marker genes for cell clusters after scRNAseq and snRNAseq. (A) Variable feature plot of scRNAseq data of dissociated *Drosophila* EAD cells. The top 10 variable genes of the scRNAseq dataset are labeled in red and the top 10 variable genes of the snRNAseq dataset are labeled in blue. These genes are the ones having the strongest influence on clustering and cell type identification for the respective dataset. The top 3,000 genes (yellow dots) are used for subsequent cluster analyses. (B) Dot plot of the top 4 marker genes (X-axis) for each cell cluster (Y-axis) obtained after scRNAseq. The size of the dots represents the percentage of cells expressing each gene. The color intensity represents the average expression level. (C) Variable feature plot of snRNAseq data of isolated nuclei. The top 10 variable genes of the scRNAseq dataset are labeled in red and the top 10 variable genes of the snRNAseq dataset are labeled in blue. These genes are the ones having the strongest influence on clustering and cell type identification for the respective dataset. The top 3,000 genes (yellow dots) are used for subsequent cluster analyses. (D) Dot plot of the top 4 marker genes (X-axis) for each cell cluster (Y-axis) obtained after snRNAseq. The size of the dots represents the percentage of cells expressing each gene. The color intensity represents the average expression level.**

An unbiased clustering of the snRNAseq data using the same resolution as used for scRNAseq (see above) resulted in 24 unique clusters. A closer examination of the top four genes that define a certain cluster revealed for instance *Sp1* (Estella and Mann 2010), *cut (ct)* (Blochlinger et al. 1993), *disco-related (disco-r)* (Grubbs et al. 2013) and *homothorax (hth)* (Dong et al. 2002; Pichaud and Casares 2000; Pai et al. 1998) in clusters 2 and 9 (Figure 7D). As all four genes

are expressed in the antennal region of the eye-antennal disc and have been implicated in antennal development we conclude that these nuclei originated from antennal tissue. Similarly, we observed expression of *Fasciclin 2 (Fas2)* (Mao and Freeman 2009), *nemo (nmo)* (Choi and Benzer 1994), *big bang (bbg)* (Kim et al. 2006) in cluster 11 (Figure 7D), suggesting that these nuclei contribute to the morphogenetic furrow and differentiating photoreceptors. GO term enrichment analyses for marker genes defining each cell cluster (Supplementary Table S 4) revealed biological processes relevant for eye-antennal disc cells (Supplementary Table S 5). Note that cluster 12 (Figure 7D) was largely defined by ribosomal gene expression. We hypothesize that all nuclei that still have remnants of cytoplasm attached show high rRNA levels compared to pure nuclei and thus lead to the distinct expression profile and clustering result.

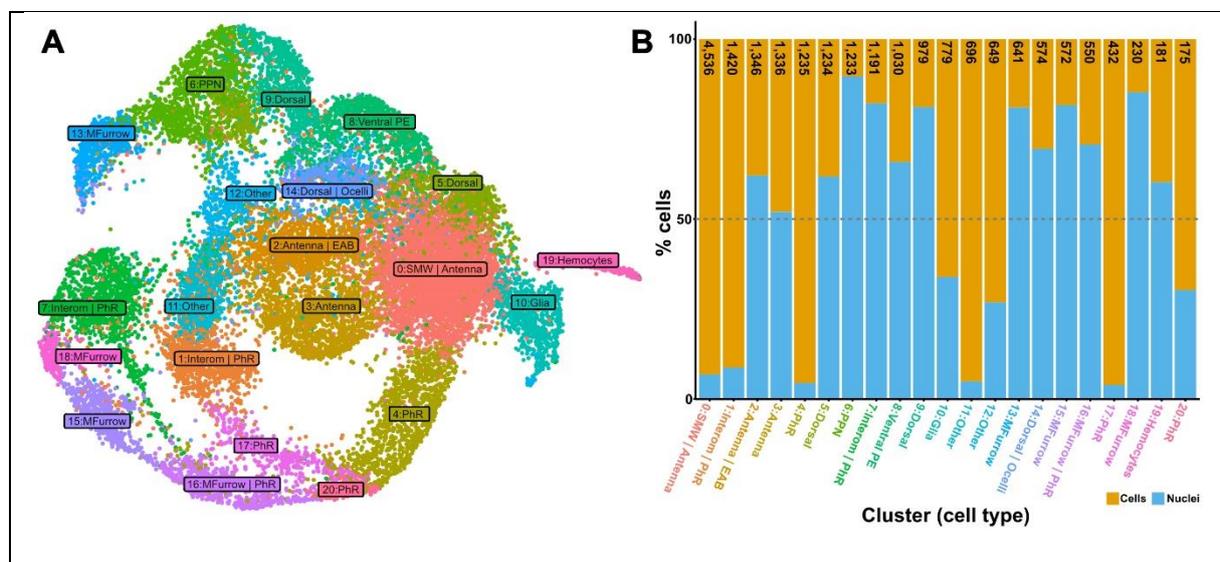
Overall, I conclude that the snRNAseq dataset captured gene expression profiles of eye-antennal disc cells and at the same time, technical artifacts, such as heat shock gene expression and an excess of reads from ribosomal genes were diminished.

### 3.1.5 scRNAseq and snRNAseq capture relevant cell types and show differences cell type composition

Besides the reduced expression of mitochondrial, ribosomal and heat shock genes in the snRNAseq dataset, I wanted to compare the snRNAseq and scRNAseq datasets more thoroughly. As the genes with the most variable expression contain information for the distinction of cell types, I first asked to what extent the most variable genes overlapped in both datasets. Most of the top 10 variable genes in the scRNAseq dataset were also among the most variable genes in the snRNAseq dataset and vice versa (Figure 7A, C; Supplementary Table S 6). Among the top 3,000 most variable genes, 1,480 were present in both datasets, while 1,520 genes were unique for the scRNAseq or the snRNAseq dataset, respectively (Supplementary Table S 6; Supplementary Table S 7). The genes unique for scRNAseq or snRNAseq predominantly represented general cellular and metabolic processes (Supplementary Table S 7A, B; Supplementary Table S 7). Intriguingly, the shared genes were highly enriched for developmental processes relevant for larval eye-antennal discs (Supplementary Table S 7C; Supplementary Table S 7).

To directly compare the snRNAseq and the scRNAseq datasets, I clustered cells and nuclei based on expression profiles and performed dimension reduction using Uniform Manifold Approximation and Projection (UMAP) (McInnes et al. 2020). I obtained 21 clusters which I

annotated based on marker genes used in previous scRNAseq analyses (Ariss et al. 2018; Bravo González-Blas et al. 2020) and based on prior knowledge from the literature (see Materials and Methods for details) (Figure 8A; Supplementary Table S 8). While some cell types could be unequivocally identified, I found signatures for two (e.g. cluster “14:Dorsal | Ocelli” in Figure 8A and Supplementary Table S 8) or more (denoted as “Other” in Figure 8A and Supplementary Table S 8) cell types for some clusters. Based on our cluster (i.e. cell type) annotation, I found that the combined dataset contained all major cell types that have been previously described in scRNAseq data for eye-antennal discs (Ariss et al. 2018; Bravo González-Blas et al. 2020) (Figure 8A, Figure 2).



**Figure 8: Direct comparison of scRNAseq and snRNAseq datasets.** (A) UMAP of integrated single-cell and nuclei RNA expression data. Cells are colored by cluster. (B) Proportion of cell- or nuclei contribution to each cluster by dataset. The total number of cells in each cluster is provided for each cell type at the top of the bar plot.

Finally, I tested whether both datasets contributed equally to the different cell types in the combined analysis. A UMAP with cells color-coded by the dataset suggested that all major cell types were represented in both datasets and that the different datasets contributed unequally to some cell types (compare Figure 8A to Supplementary Table S 9). Therefore, I quantified the number of cells of each cluster originating from scRNAseq and snRNAseq data, respectively (Figure 8B). This analysis confirmed that both datasets contribute to all cell types. However, I observed biases in the data composition as 12 of the 21 clusters were predominantly defined by snRNAseq data, 8 clusters were defined by scRNAseq data and both datasets contributed almost equally to one cluster (Figure 8B). The scRNAseq-biased clusters were those with ambiguous cell type assignment (e.g. “Other” clusters 11 and 12 and “second mitotic wave (SMW) |

Antenna” cluster 0; Figure 8B), glia cells (cluster 10; Figure 8B) and almost all photoreceptor clusters (e.g. clusters 1, 4, 17 and 20; Figure 8B). These cell types are expected in late larval stages when the progression of the morphogenetic furrow slows down and most regions of the disc undergo differentiation (Treisman 2013). Interestingly, the snRNAseq-biased clusters were mostly related to processes taking place earlier in L3 larval development, such as extensive cell proliferation in the proneural zone (“PPN” cluster 6, Figure 8B) and in the eye-antennal border region (“Antenna | EAB” cluster 2; Figure 8B) and progression of the morphogenetic furrow (“MFurrow” clusters 13, 15, 16 and 18; Figure 8B) (Treisman 2013). While previous comparisons of scRNAseq and snRNAseq also showed biases in the relative contribution to different cell types (Wu et al. 2019; Bunis et al. 2020; Denisenko et al. 2020; Ding et al. 2020), the observed discrepancy between our two datasets could partially be explained by experimental difference. Specifically, larvae for scRNAseq were staged based on morphology and behavior (i.e. late wandering larvae at the transition to white prepupae) and discs for snRNAseq were dissected from larvae staged by developmental time (120 h after egg laying). Therefore, I assume that the larvae used for scRNAseq had been slightly older than those obtained for snRNAseq.

In summary, our direct comparison of scRNAseq and snRNAseq data showed that both methods identified all major cell types present in the eye-antennal disc. The observed biases in the contribution to different cell types most likely represent experimental differences of this work, as well as fundamental differences between scRNAseq and snRNAseq.

### 3.2 Data validation

With the optimized protocol, I was able to gather data for 14 unique samples of *Drosophila* EAD. One sample, *D. mauritiana* 96 h, was lost due to a wetting failure during the loading procedure.

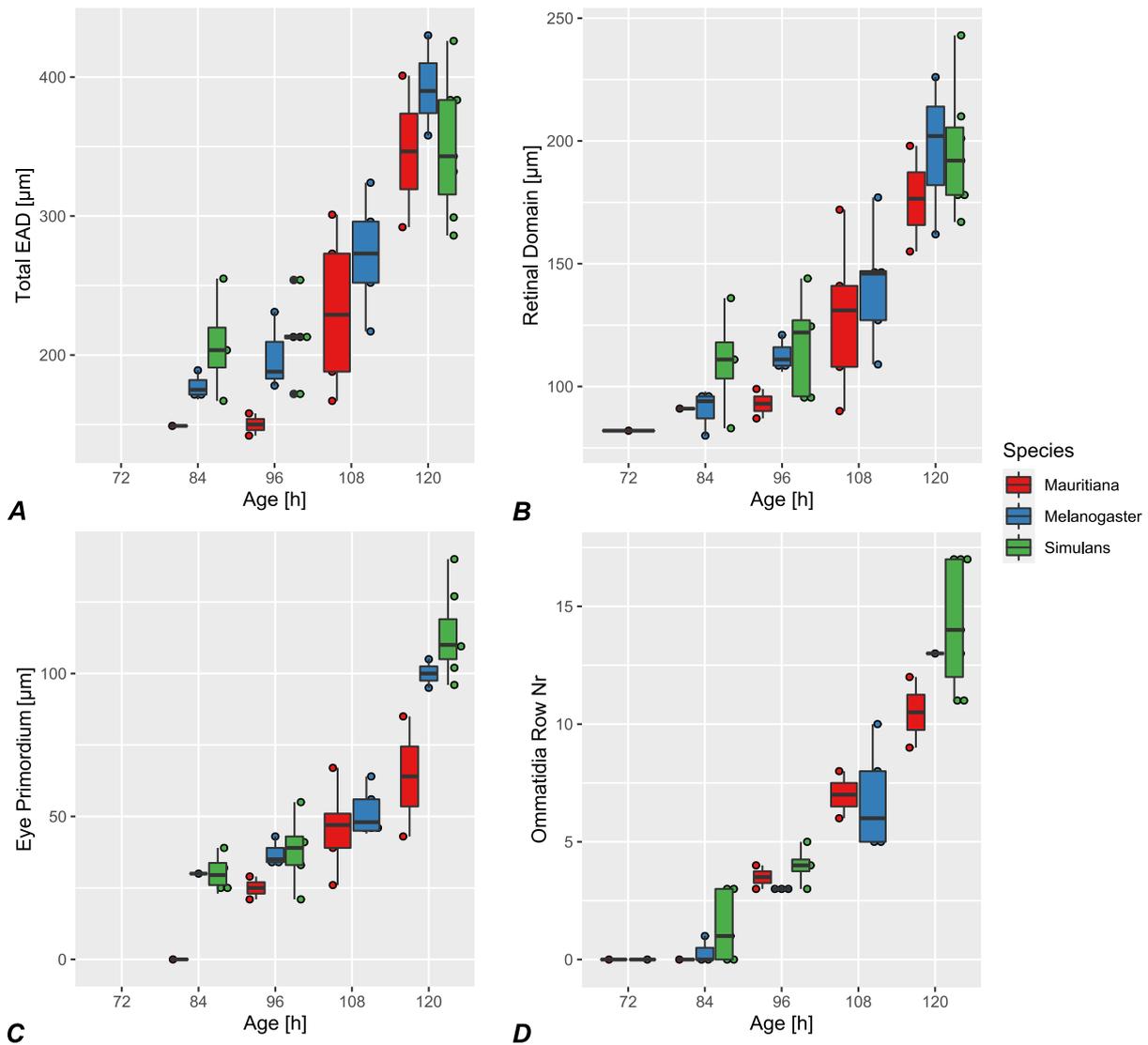
**Table 2: General Cellranger metrics for all 14 datasets.** Mel: *D. melanogaster*, Sim: *D. simulans*, Mau: *D. mauritiana*. Sample “Mel\_120 h” is identical to sample “Nuclei\_MDC” in Table 1.

| Metric   | Mel_120h    | Mel_108h    | Mel_96h     | Mel_84h     | Mel_72h     | Sim_120h    | Sim_108h    | Sim_96h     | Sim_84h     | Sim_72h    | Mau_120h    | Mau_108h    | Mau_84h     | Mau_72h    |
|--|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|------------|-------------|-------------|-------------|------------|
| Estimated Number of Cells                      | 9048        | 4338        | 2314        | 3655        | 2106        | 13569       | 13585       | 5288        | 2832        | 626        | 10988       | 1992        | 2870        | 326        |
| Mean Reads per Cell                            | 13,334      | 24,906      | 100,882     | 65,757      | 51,411      | 18,290      | 17,229      | 37,267      | 43,434      | 73,167     | 23,712      | 98,314      | 42,124      | 166,315    |
| Median Genes per Cell                          | 812         | 915         | 1,112       | 1,215       | 1,294       | 1,178       | 1,045       | 984         | 896         | 1,098      | 1,103       | 1,050       | 1,182       | 1,211      |
| Number of Reads                                | 120,649,741 | 108,045,342 | 233,441,982 | 240,344,643 | 108,272,838 | 248,177,572 | 234,064,778 | 197,069,615 | 123,007,357 | 45,802,824 | 260,552,346 | 195,842,302 | 120,897,087 | 54,218,952 |
| Valid Barcodes                                 | 97.10%      | 96.90%      | 97.80%      | 97.60%      | 97.50%      | 97.00%      | 97.10%      | 97.40%      | 97.60%      | 97.80%     | 97.70%      | 97.20%      | 97.20%      | 97.70%     |
| Sequencing Saturation                          | 67.90%      | 79.30%      | 84.40%      | 71.70%      | 86.30%      | 63.00%      | 55.20%      | 81.90%      | 85.60%      | 88.00%     | 61.10%      | 77.50%      | 83.70%      | 92.80%     |
| Q30 Bases in Barcode                           | 96.60%      | 96.00%      | 95.20%      | 95.30%      | 96.60%      | 94.90%      | 94.70%      | 95.40%      | 96.40%      | 96.30%     | 95.20%      | 95.40%      | 96.50%      | 96.50%     |
| Q30 Bases in RNA Read                          | 95.90%      | 95.20%      | 95.00%      | 94.80%      | 95.70%      | 94.40%      | 94.40%      | 95.40%      | 95.90%      | 95.70%     | 94.70%      | 95.00%      | 95.90%      | 95.90%     |
| Q30 Bases in UMI                               | 96.40%      | 95.70%      | 94.70%      | 94.80%      | 96.30%      | 94.40%      | 94.20%      | 94.90%      | 96.20%      | 96.10%     | 94.60%      | 95.00%      | 96.30%      | 96.30%     |
| Reads Mapped to Genome                         | 85.70%      | 82.60%      | 86.30%      | 89.10%      | 89.50%      | 80.60%      | 82.80%      | 83.90%      | 80.40%      | 71.10%     | 79.80%      | 86.00%      | 84.20%      | 74.60%     |
| Reads Mapped Confidently to Genome             | 84.50%      | 81.60%      | 85.30%      | 87.90%      | 88.50%      | 79.70%      | 82.10%      | 83.20%      | 79.80%      | 70.60%     | 79.50%      | 85.80%      | 83.90%      | 74.40%     |
| Reads Mapped Confidently to Intergenic Regions | 0.80%       | 0.80%       | 0.80%       | 0.60%       | 1.10%       | 2.20%       | 1.70%       | 1.50%       | 1.40%       | 1.40%      | 1.40%       | 1.40%       | 1.80%       | 0.90%      |
| Reads Mapped Confidently to Intronic Regions   | 14.70%      | 16.60%      | 14.90%      | 11.90%      | 12.10%      | 17.00%      | 11.50%      | 10.90%      | 8.70%       | 8.20%      | 12.30%      | 8.20%       | 12.80%      | 6.60%      |
| Reads Mapped Confidently to Exonic Regions     | 69.00%      | 64.10%      | 69.60%      | 75.40%      | 75.30%      | 60.40%      | 68.90%      | 70.80%      | 69.70%      | 61.00%     | 65.80%      | 76.20%      | 69.30%      | 66.90%     |
| Reads Mapped Confidently to Transcriptome      | 75.60%      | 72.00%      | 77.70%      | 79.70%      | 80.70%      | 69.60%      | 73.90%      | 75.10%      | 73.90%      | 65.50%     | 72.40%      | 77.40%      | 73.00%      | 69.30%     |
| Reads Mapped Antisense to Gene                 | 6.50%       | 7.30%       | 5.30%       | 6.10%       | 4.80%       | 6.30%       | 5.10%       | 5.20%       | 3.20%       | 2.30%      | 4.30%       | 5.60%       | 7.40%       | 3.00%      |
| Fraction Reads in Cells                        | 73.50%      | 72.60%      | 60.10%      | 52.40%      | 75.60%      | 83.50%      | 67.50%      | 52.80%      | 61.00%      | 72.50%     | 65.90%      | 29.90%      | 74.90%      | 63.90%     |
| Total Genes Detected                           | 12,296      | 11,517      | 12,292      | 12,512      | 11,970      | 13,312      | 13,109      | 12,140      | 11,553      | 10,583     | 12,729      | 11,414      | 12,351      | 9,611      |
| Median UMI Counts per Cell                     | 1,383       | 1,650       | 2,304       | 2,678       | 2,963       | 2,531       | 2,066       | 2,032       | 1,784       | 2,627      | 2,193       | 2,220       | 2,717       | 3,163      |

In each dataset, I found more than 10,000 unique genes (Table 2). In 120 h old discs, the average number of genes is between 812 in *D. melanogaster* to 1178 in *D. simulans*. The percentages of genes mapped confidently to the genomes ranges from 70% (Sim\_72 h) to 88% (Mel\_72 h). By species, I deem samples gathered from *D. simulans* to be highest in quality, as it shows a high number of captured cells at most stages and information specific to development is easily accessible. Overall, I managed to obtain the maximum numbers of about 10,000 captured cells for 120 h AEL EADs for all three species. For earlier stages, the number of captured cells is generally lower (Table 2). To further evaluate the quality of the obtained data, I tested how comparable the chosen stages were and I compared the snRNAseq data to previously established bulk sequencing data for the same developmental stages.

### 3.2.1 Accuracy of staged EAD

As developmental time can vary between individuals and species, it is important to document such potential differences between samples. During dissection of EADs for snRNAseq, few reference discs were separated and stained using Phalloidin 488 to assess multiple morphological features of EAD development (see Figure 37). The number of ommatidial rows that are present in the eye disc ( $N_{(OM)}$ ) were counted to serve as a measurement for progression of differentiation (i.e. the MF). I also measured the distance from the optic stalk to the MF ( $L_{(RD)}$ ), the central length of the entire eye primordium ( $L_{(EP)}$ ) and the length of the total EAD ( $L_{(EAD)}$ ) to assess the progression of retinal development and the size of the entire disc, respectively.



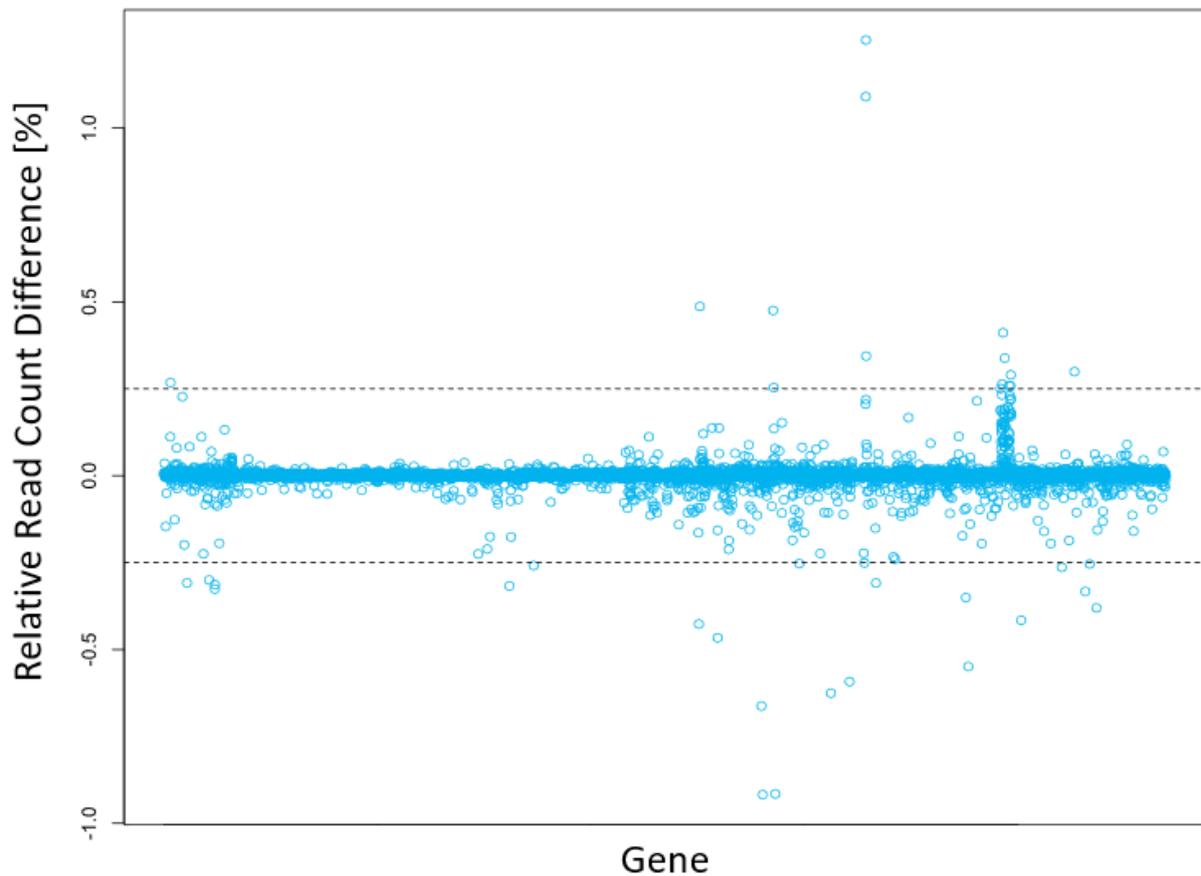
**Figure 9: Metrics for developmental time assessment between different stages and species.** (A) Length of the whole EAD over time. (B) Length of the retinal domain. (C) Length of the eye primordium. (D) Ommatidia row counts over time. Ages are measured in hours after egg laying, lengths are measured in  $\mu\text{m}$ .

I observed a gradual increase of the number for all measured traits showing that successive stages were dissected for all three species (Figure 9). It is important to note that the captured discs were not suited for some measurements for some timepoints leading to missing data. *D. mauritiana* appears to lag behind in the length of the retinal domain (Figure 9B) and average number of ommatidial rows (Figure 9D). This is visible from 96 h onwards, especially in the length of the retinal domain (Figure 9B). However, it shows a similar average EAD length as *D. simulans* at 120 h (Figure 9A). *D. melanogaster* shows the longest EADs at 120 h (Figure 9A). The largest variance between EAD within a species in number of ommatidia rows appears to be in *D. simulans*, both at 84 h and 120 h (Figure 9D).

Generally, earlier stages are comparable between species, showing 0 ommatidia rows at 72 h, and not more than 1 on average in *D. simulans* at 84 h. Differences in later stages may show relevant differences between species in developmental time.

### 3.2.2 Single-Nuclei RNA-sequencing data of 120 h old *D. melanogaster* EAD is comparable to bulk RNA sequencing data of the same tissue type

As described in 3.1, snRNAseq has its special challenges, such as the risk of losing signals of lowly expressed genes, due to the initial low number of transcripts per single nucleus. While the potential loss of lowly expressed genes is an unavoidable drawback of the technology, it is important to estimate the validity of our dataset for averagely- and highly expressed genes. To this end, I compared gene expression of our 120 h AEL *D. melanogaster* snRNAseq dataset to a previously published bulk sequencing dataset (GSE94915 (Torres-Oliva et al. 2018)) of the same stage. In total, 10,796 genes were found in my snRNAseq data as opposed to 13,676 genes in the bulk RNAseq dataset. Out of these genes, 9,388 genes were comparable between both datasets.



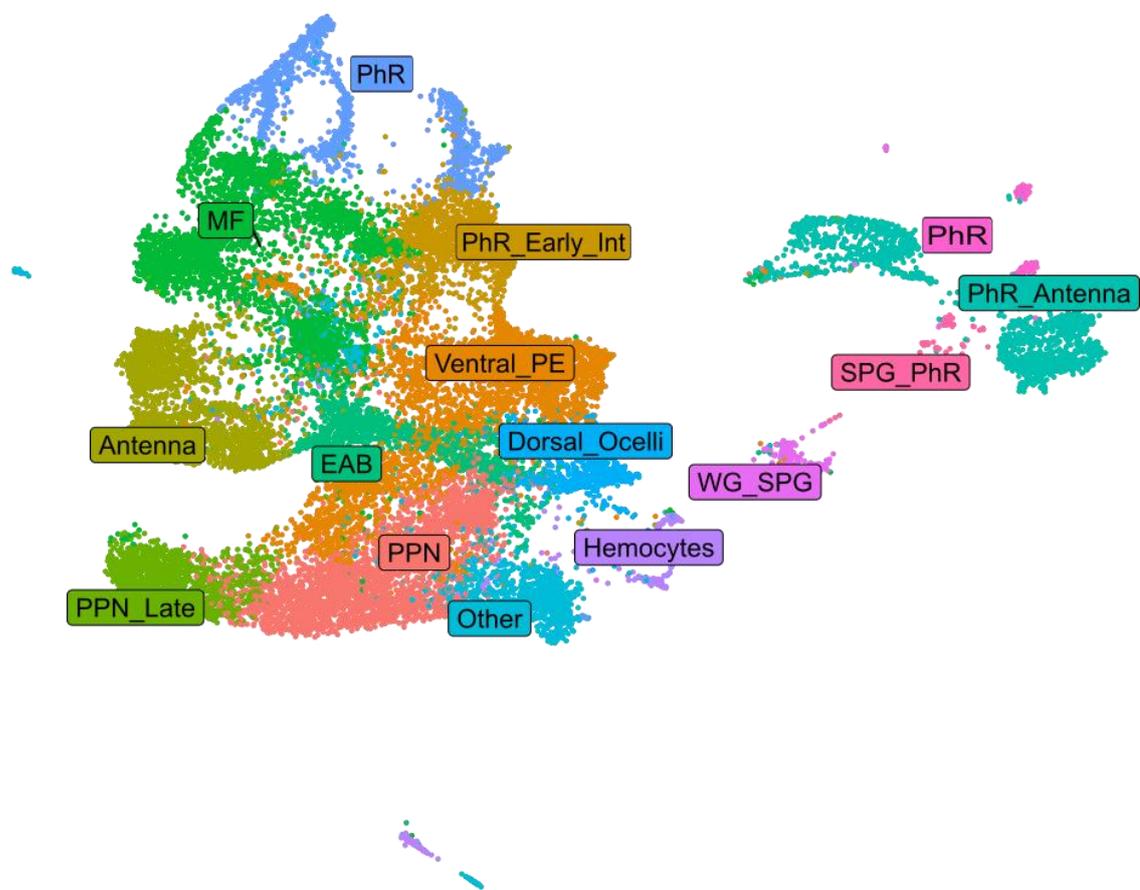
**Figure 10: Relative read count differences between bulk RNA-seq and single-nuclei RNA-seq data of 120 h old *D. melanogaster* EAD.** The dashed lines show a threshold of  $\pm 25\%$ . Most genes show less than 1% relative read count difference between the datasets. Genes are ordered alphabetically. Values  $< 0$  are overrepresented in snRNAseq data and values  $> 0$  are overrepresented in bulk RNA seq data. A collection of ribosomal genes is detected more strongly in bulk RNA seq data.

Out of the shared genes, all genes show less than 1% difference in read counts between the datasets relative to the total reads in each dataset. Less than 25 genes have strongly higher read counts in bulk data ( $> 25\%$  difference) and they are mitochondrial (e.g. *mt:CoI*), heat shock response- (e.g. *Hsp23*) or ribosomal genes (*RpL10*) (Supplementary Table S 10). Thus, overall gene expression in snRNAseq data is comparable to the bulk dataset.

### 3.3 All major cell types of the *Drosophila* EAD can be identified using single-nuclei RNA-sequencing

Single-cell and single-nuclei sequencing methods facilitate analyzing and comparing gene expression in complex tissues on the level of specific cells and cell types. To test, whether all major cell types known to be present in the EAD are captured in my snRNAseq datasets, I first clustered the obtained cells based on gene expression profiles and subsequently applied a cluster

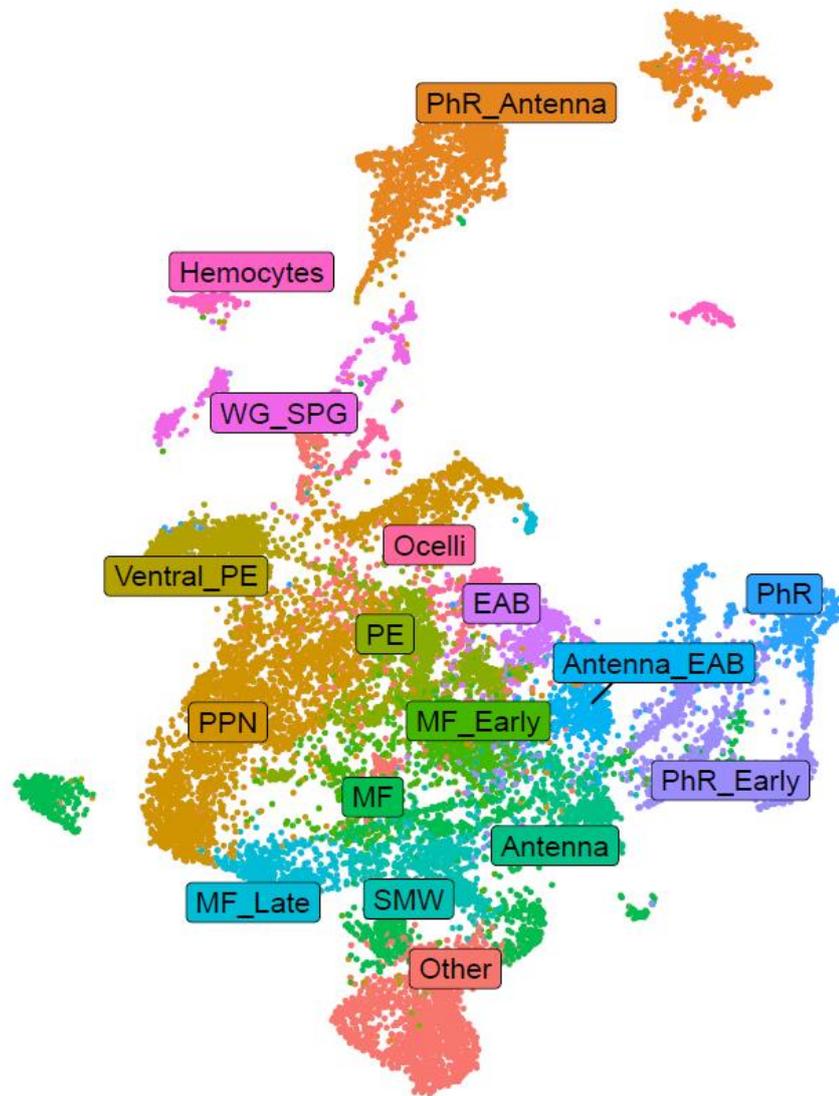
annotation pipeline employing previously published marker genes obtained for 120 h old EAD in *D. melanogaster* (Ariss et al. 2018; Bravo González-Blas et al. 2020). A special emphasis of the clustering and annotation was on comparability between species (see Materials and Methods for details). Clustering results were visualized using Uniform Manifold Approximation and Projection (UMAP), a method to display multi-dimensional data in a two- or three-dimensions. It displays cells as dots and positions them according to their transcriptional similarity to each other (Figure 11). Note that all timepoints were combined for this analysis (see also Figure 18).



**Figure 11: UMAP-plot of manually assigned cell cluster identities based on marker genes in *D. melanogaster* datasets of all timepoints. 15 Clusters were identified and manually assigned cell type identities based on marker score.**

I could identify 15 discrete clusters in *D. melanogaster* (Figure 11) representing all expected cell types (compare to Figure 8). Due to the chosen cluster resolution, some cell types were combined in one larger cluster. For instance, the cluster “PhR\_Early\_Int” included cells expressing early photoreceptor markers, as well as cells expressing interommatidial markers.

Unexpected findings include large clusters expressing photoreceptor marker genes at 72 h (Cluster “PhR\_Antenna”). This early in development, there are no or almost no ommatidial rows expected to be present (Figure 9).



**Figure 12: UMAP-plot of manually assigned cell cluster identities based on marker genes in *D. mauritiana* datasets of all timepoints.** 17 Clusters were identified and manually assigned cell type identities based on marker score.

As in *D. melanogaster*, there is a large cluster expressing photoreceptor and antennal markers in *D. mauritiana*, in this case largely at 84 h (Figure 12). Here, I find a not further defined peripodial epithelium cluster (“PE”). Antennal cells in this UMAP plot are not separated as distinctly from other cells of the EAD as in the other two species. Cluster “Ocelli” contains dorsal cells as well.

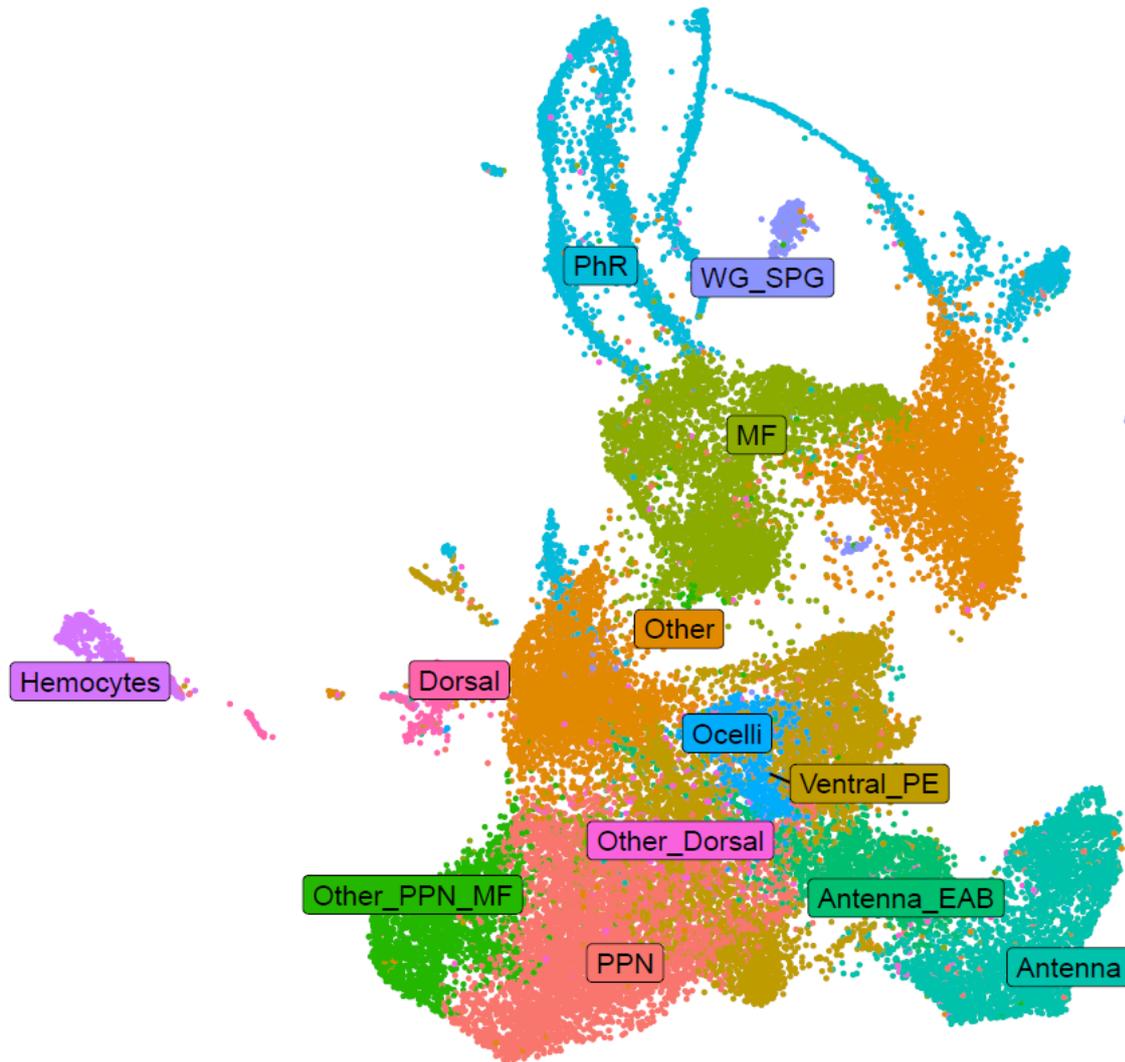
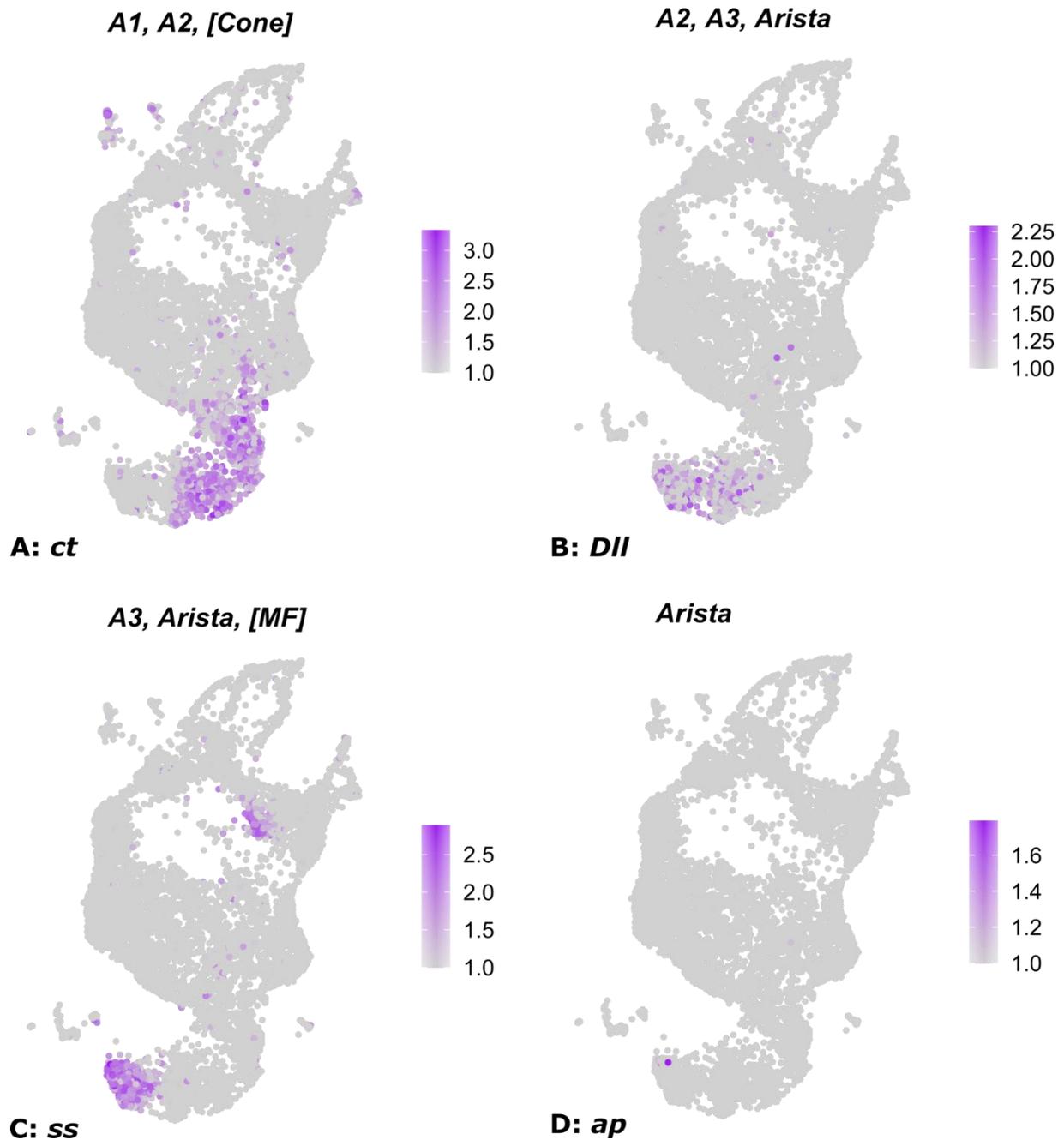


Figure 13: UMAP-plot of manually assigned cell cluster identities based on marker genes in *D. simulans* datasets of all timepoints. 13 Clusters were identified and manually assigned cell type identities based on marker score.

In *D. simulans*, all cell types expected from *D. melanogaster* are found as well. In this data, I find the clearest distinctions between cell types and cells ordered very close to the expectation from the underlying biology. For example, antennal cells clustered best together and were located far away from photoreceptor cells, recapitulating the very different cell types (Figure 13). Each UMAP plot includes the combined data of a single species just for the purpose of giving an overview.

The proximity of the cell types in the UMAP plots approximate their order in the physical tissue. Within all species plots, I observed a general layout with antennal nuclei on one side and photoreceptors on the other. Within the photoreceptors (“PhR”), I observed distinct stripe-like patterns, which are not found for other cell types. Directly adjacent to the photoreceptors are

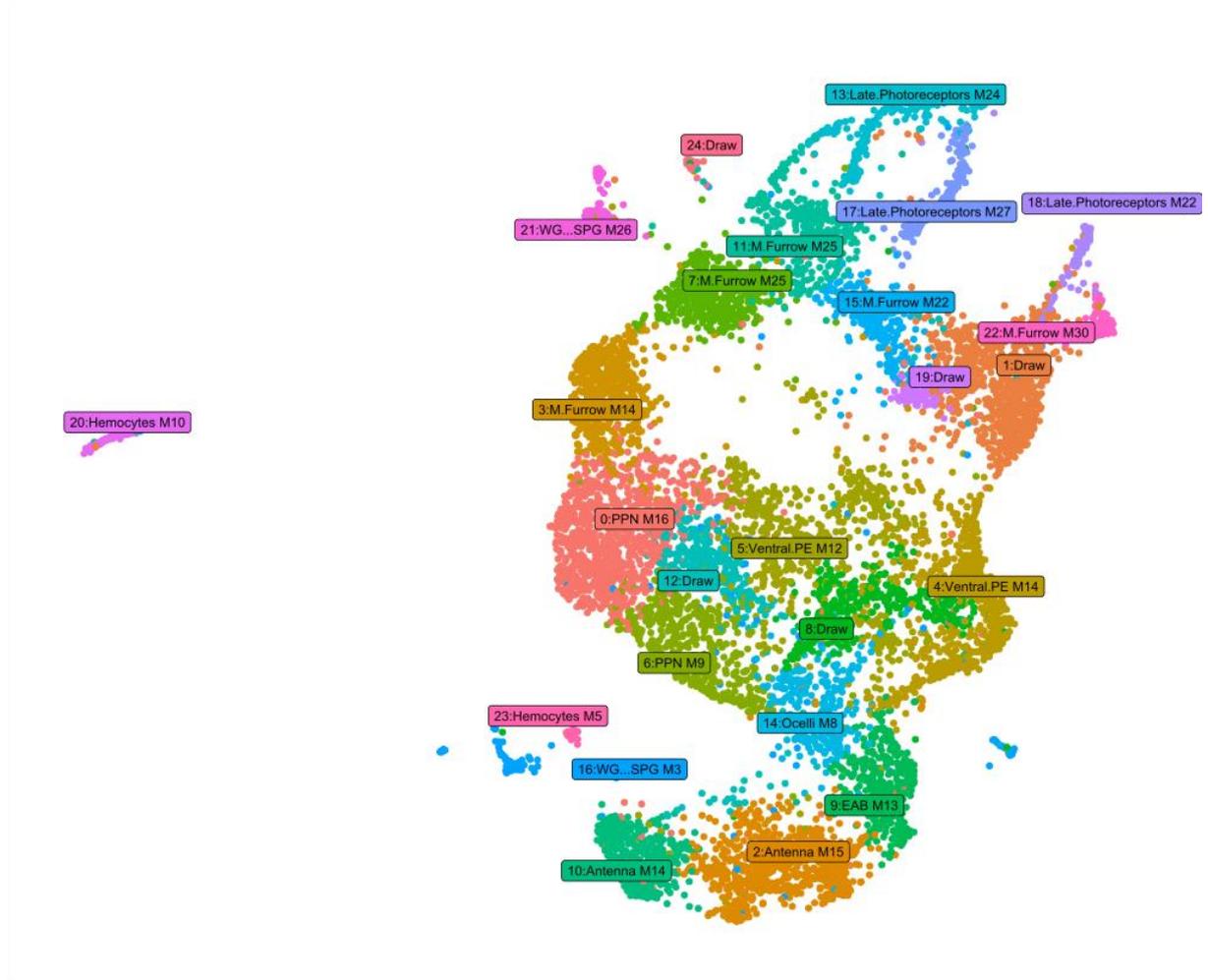
cells of the morphogenetic furrow (“MF”). One of the clusters at the border between photoreceptors and MF cells is usually identifiable as cells of the second mitotic wave. Between the MF and the antennal clusters I found pre-proneural cells (“PPN”), cells of the peripodial epithelium (“Ventral”-“PE”), dorsal- and ocelli cells at varying positions. At the chosen resolutions, there were usually two antennal clusters, one of which represented eye-antennal border cells, which are located at the interface of the antennal and the retinal part of the EAD. Cells which are not part of the EAD proper are immediately recognizable as their clusters were located far away from those of the EAD. Those are Hemocytes and glia cells (in “WG\_SPG”), the latter for which migrate into the EAD from the brain (Silies et al. 2007). Overall, I find the same major cell types expected from *D. melanogaster* in data of the non-model species *D. mauritiana* and *D. simulans*.



**Figure 14: Marker gene expression for Antennal segments in 120 h AEL *D. melanogaster* EAD.** (A) *ct*: Marks the outermost antennal segments A1 and A2, and is also expressed in cone cells (top). (B) *Dll* marks antennal segments A2 and A3 and the arista. (C) *ss* marks antennal segment A3 and the arista. (D) *ap* marks the innermost Arista.

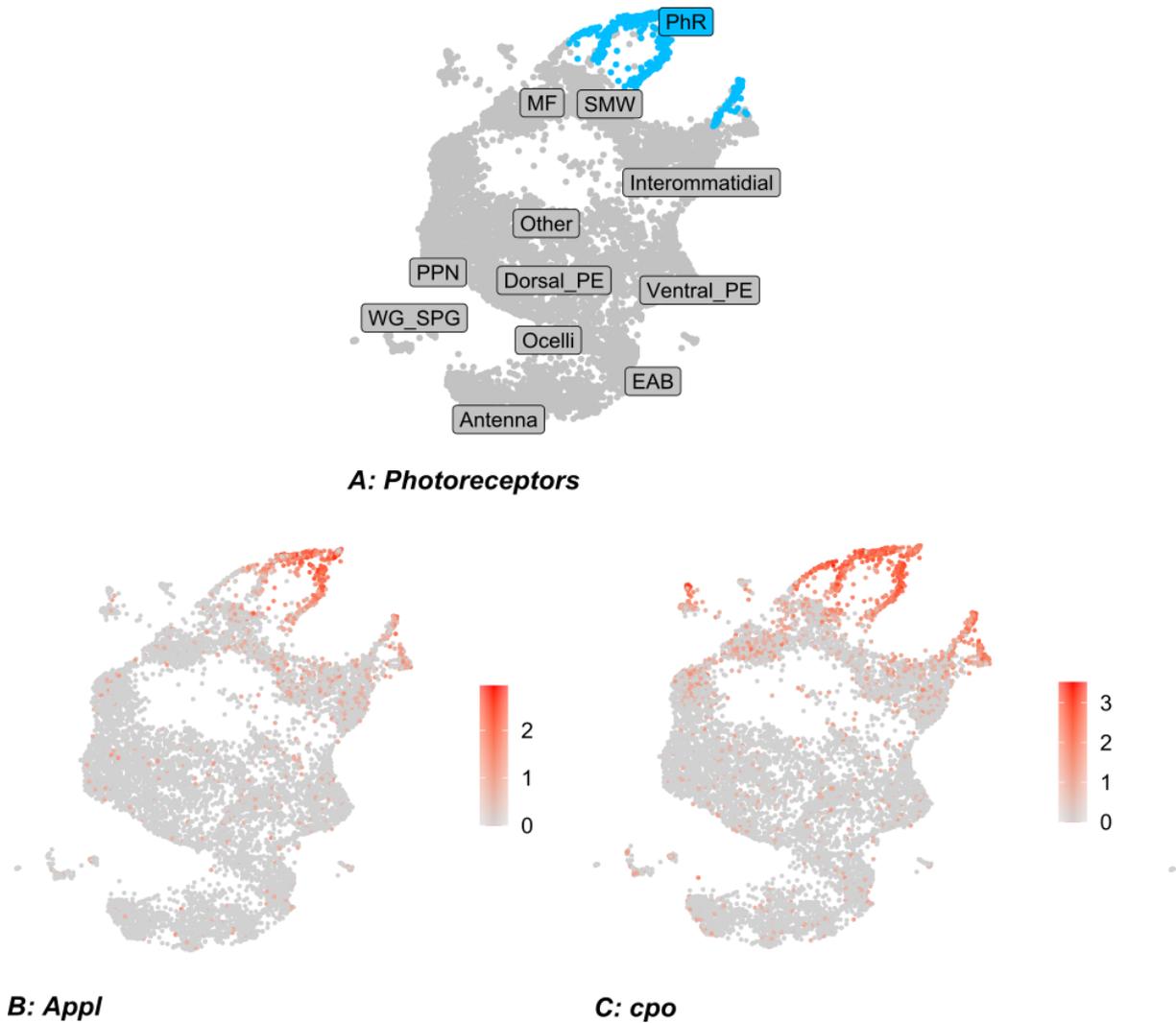
Since most cell types are usually represented with several different subclusters in many datasets (for instance, I find three different clusters related to antennal cell types in 120 h AEL *D. melanogaster* data, cluster 2, 10 and 9 in Figure 15), I investigated their identity further by analyzing the expression of known marker genes. Within the antennal clusters, I found a subset of nuclei expressing *cut* (*ct*) (Figure 14A) which has been shown to be expressed in the outermost antennal segments up to the EAB (Emerald et al. 2003) (see Figure 15). A small

subset of nuclei expressed *apterous* (*ap*) (Figure 14D), which is known to drive arista development (Roignant et al. 2010; Cohen et al. 1992) and *Distalless* (*Dll*) was broadly expressed in the antennal clusters 2 and 10 (Figure 14B). The expression patterns of these genes are in accordance with the order of the different antennal segments (A1, A2, A3, Arista) in the developing antenna (Emerald et al. 2003; Bravo González-Blas et al. 2020).



**Figure 15: Unbiased cluster results annotated by automatic annotation in 120 h AEL *D. melanogaster* EAD.** Photoreceptor clusters (here 13,17 and 18, automatically annotated as “Late.Photoreceptors”) form stripe like structures in UMAP plots in 108 h and 120 h old datasets.

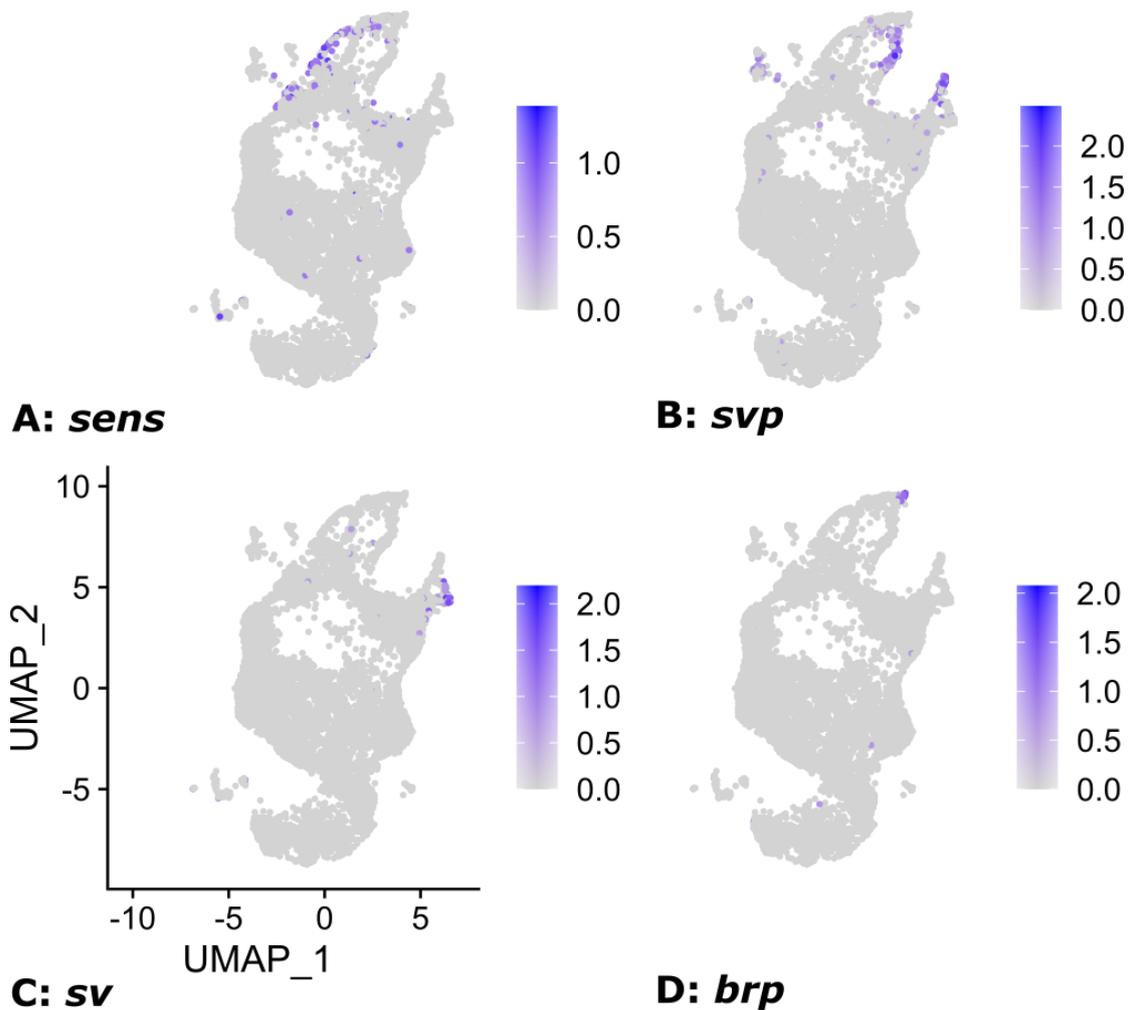
The observation in antennal clusters suggested that similar trends may be observable for the different stripe-like photoreceptor structures which often form distinct clusters (Cluster 13, 17 and 18 in Figure 15).



**Figure 16: Photoreceptor cell characterization in snRNAseq data of 120 h AEL *D. melanogaster* EADs.** (A) Location of photoreceptor cells (PhR) in a UMAP representation of the data (blue). (B) Expression of the top PhR DEG, *Appl*, in the data. (C) expression of the second top DEG, *cpo*, in the data. Hemocyte cluster was removed for clarity.

To test if the stripes represent different photoreceptors, I analyzed the expression of known marker genes in these PhR clusters. Photoreceptor cells seem to stretch along a path which is characterized by an expression gradient of genes such as  $\beta$  amyloid protein precursor-like (*Appl*) and couch potato (*cpo*) (Figure 16A and B). They have been detected previously in photoreceptor cells (Torroja et al. 1996; Ariss et al. 2018). The photoreceptor clusters appear to converge in cells expressing *bruchpilot* (*brp*) (Figure 17D), whose gene product has been detected in neuromuscular junctions (Graf et al. 2009; Oswald et al. 2010) including adult photoreceptor synapses (Hamanaka and Meinertzhagen 2010). Since these groups appear in every dataset, this might indicate interesting, distinct gene expression patterns within them.

Therefore, I first aimed to examine, whether these stripes correspond to specific photoreceptor types using marker genes previously established by (Bravo González-Blas et al. 2020).



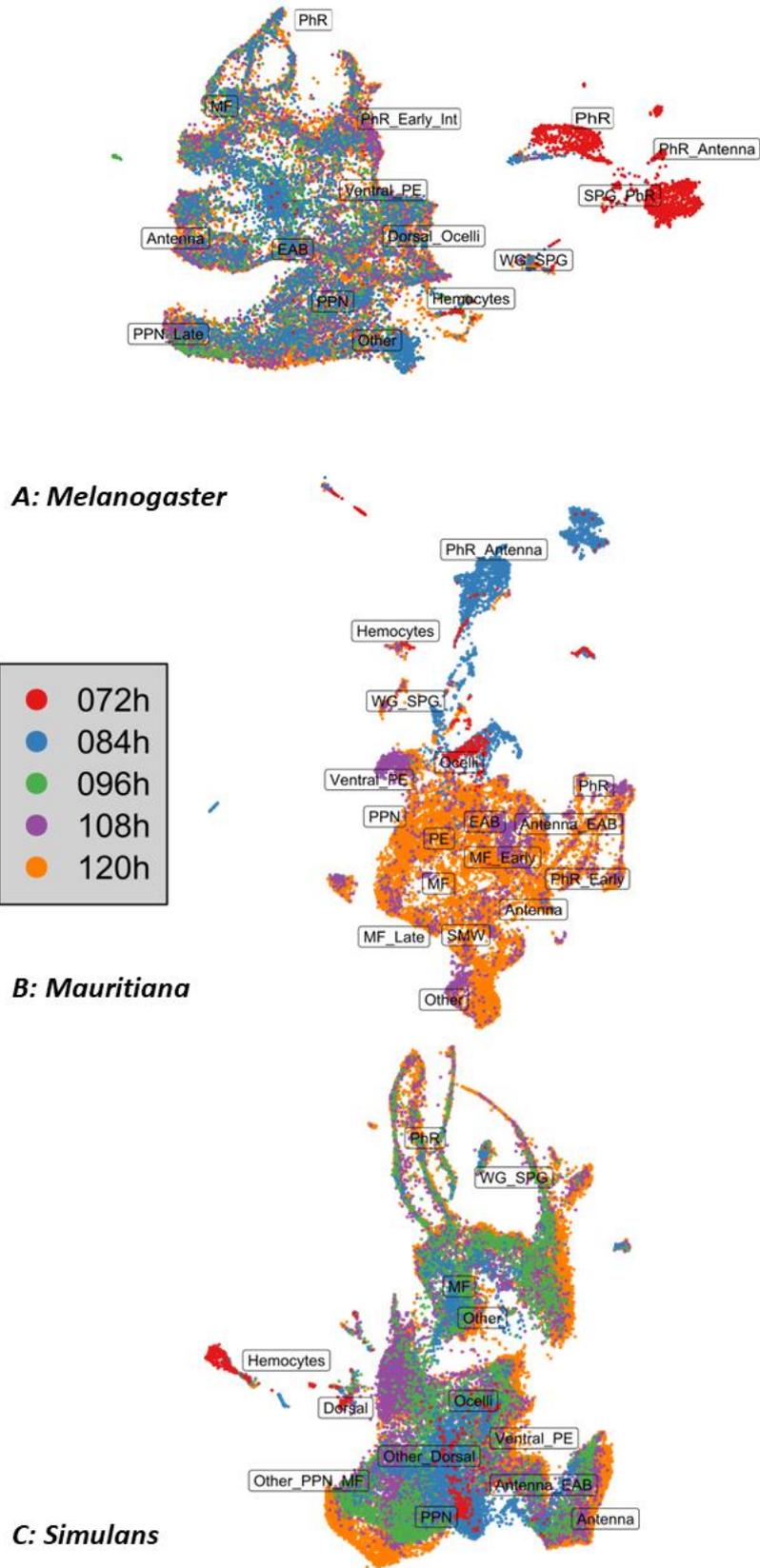
**Figure 17: Highlighted marker scores for specific photoreceptor subtypes 120 h *D. melanogaster* data.** (A) *senseless* marks R8 photoreceptors, the first photoreceptors to form. (B) *seven up* (*svp*) marks R3-R4. (C) *shaven* (*sv*) marks R7 Photoreceptor cells as well as cone cells. (D) *bruchpilot* (*brp*) is characteristic for photoreceptor synapses.

*Senseless* (*sens*) is expressed in R8 photoreceptor cells (Nolo et al. 2000; Fu and Baker 2003; Frankfort et al. 2001) (Figure 17A). Intermediate born photoreceptors R3 and R4 can be identified by the expression of *seven up* (*svp*). In my data, these photoreceptor types do not cluster within a single “stripe” or cluster (Figure 17B). The expression of *Shaven* (*sv*) is restricted to a single cluster (Cluster 18 in Figure 15) but is characteristic for both R7 photoreceptors as well as cone cells (Charlton-Perkins et al. 2011) (Figure 17C). In my dataset, I could not find that genes in specific stretches of the PhR clusters can be identified as single

distinct photoreceptor types, with the exception that R8 cells marked by *sens* showed a very distinct distribution.

### 3.3.1 Temporal ordering of biological samples is faithfully captured by single-nuclei RNA sequencing

The dynamics in gene expression between different compartments of the EAD over the course of development are largely elusive. Not much is known yet about the behavior of different cell lineages within the EAD, and it is not known whether the timepoint specific- or the cell type specific gene expression patterns have the biggest influence on clustering. Since I collected data from up to five timepoints of larval development per species, this allows the analysis of developmental dynamics of the EAD within a species.



**Figure 18: UMAP-plots of all timepoints by species.** Each dot is colored by the age of the origin sample in hours after egg laying and represents a single nucleus. (A) Data of *D. melanogaster*. (B) Data of *D. mauritiana*. In this dataset, the 96 h timepoint is missing due to a wetting failure. (C) Data of *D. simulans*. This dataset shows the clearest temporal ordering of nuclei for the youngest to the oldest timepoint.

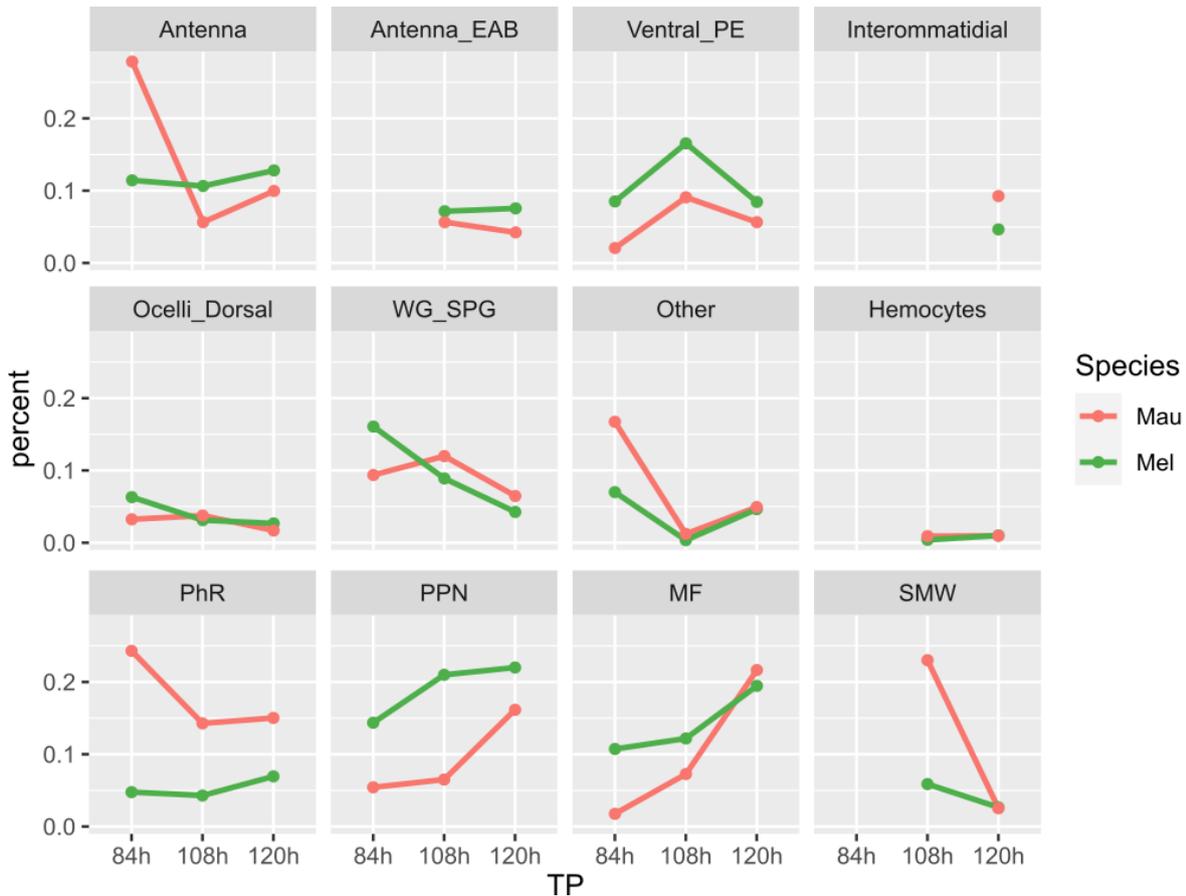
To compare the timepoints, for each species all timepoints were integrated using Seurat. To compare the timepoints, for each species all timepoints were integrated and visualized using Seurat as well. The UMAP plots show concentric or linear structures. This order is most pronounced in the central cell types of the eye primordium, especially in *D. simulans* near the PPN (Figure 18 C). In case of differentiated cell types such as photoreceptors, there is expectedly no such gradient visible, the respective clusters are instead composed of cells from several timepoints without a specific order (Figure 18 A, B, C). Especially in *D. simulans* (Figure 18C) and to a lesser degree in *D. melanogaster* (Figure 18B), there is also an internal temporal gradient visible within the antennal clusters. To detect this gradient, it is necessary to use a large number of principle components during preprocessing which indicates that cluster- and cell type specific transcription outweighs global temporal transcriptional differences. I see outlier clusters in all species, especially in younger timepoints. In *D. melanogaster*, there are very large, disconnected clusters of 72 h old cells (Figure 18A). This is also visible in *D. simulans* data, but these clusters are visibly smaller (Figure 18C). In *D. mauritiana*, there are outlier clusters of 84 h old cells (Figure 18B). In *D. simulans* data, I see an accumulation of MF cells in 108 h (Figure 18C). Please note that the 96 h AEL data is missing for *D. mauritiana* what may disrupt the temporal ordering. In summary, I observed that developmental information is retained in all species' datasets with earlier stages being transcriptionally most different from later stages. Depending on the cluster, both the cell type specific signals (e. g. in photoreceptor cells) or the timepoint specific gene expression (e.g. in antennal cells and around the PPN cells) has the largest impact on clustering.

### 3.4 Interspecies comparison

Multiple studies have described differences in eye- and head morphology between adults of *D. melanogaster*, *D. mauritiana* and *D. simulans*. As stated earlier, the best described morphological differences are those between *D. mauritiana* and *D. simulans*, which differ in eye size due to larger ommatidia in *D. mauritiana*. It is also known that there are eye size differences between *D. melanogaster* and *D. mauritiana* which can be explained mainly by a larger number of ommatidia in *D. mauritiana*. Since many cell populations of the EAD proliferate and differentiate during late larval development, I expected to see differences in cell population compositions between these species. Therefore, I explored the general trends of cell population dynamics and differences in cell-type specific gene expression and regulation in the two pairwise species-comparisons.

### 3.4.1 Divergence in cell type composition between *D. melanogaster* and *D. mauritiana*

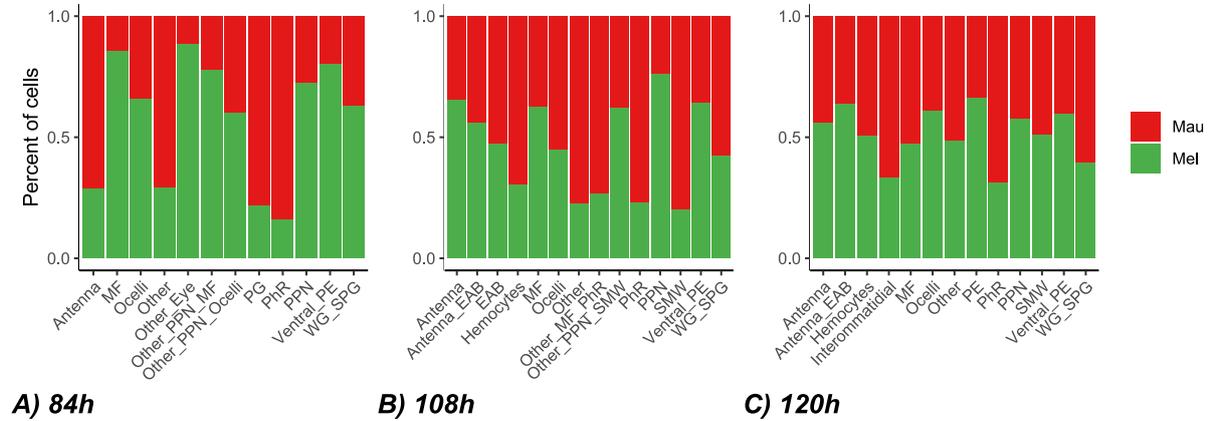
The larger eye size of *D. mauritiana* compared to *D. melanogaster* is due to a higher number of ommatidia in *D. mauritiana*. To gain insights into the potential cell types driving this difference, I first compared the relative cell population size for each cell type and species.



**Figure 19: Progression of cell type proportions within *D. melanogaster* and *D. mauritiana*.** Points in this plot depict the share of cells within one species, which are identified as a specific cell type. This plot includes only cell types which I could unequivocally identify in both species.

I observed the most pronounced differences in cell type proportions between *D. melanogaster* and *D. mauritiana* in retinal cell types, especially PhR, PPN and MF (Figure 19, bottom row). The share of cells within the data that are PhR cells interestingly appears to decrease over time in both species, as does the number of PPN cells. The share of MF cells shows an increasing direction in both species. The share of MF cells is initially larger in *D. melanogaster*, but peaks at similar levels in both species at 120 h. There appears to be a constantly larger share of PPN cells in *D. melanogaster*, and a constantly larger share of PhR cells in *D. mauritiana*. Further differences are found in the share of antennal cells which is larger in *D. melanogaster* with an

exception at 84 h. Another visible difference is in the larger proportion of cells in “Ventral\_PE” which appears to be biased towards *D. melanogaster* at every timepoint.

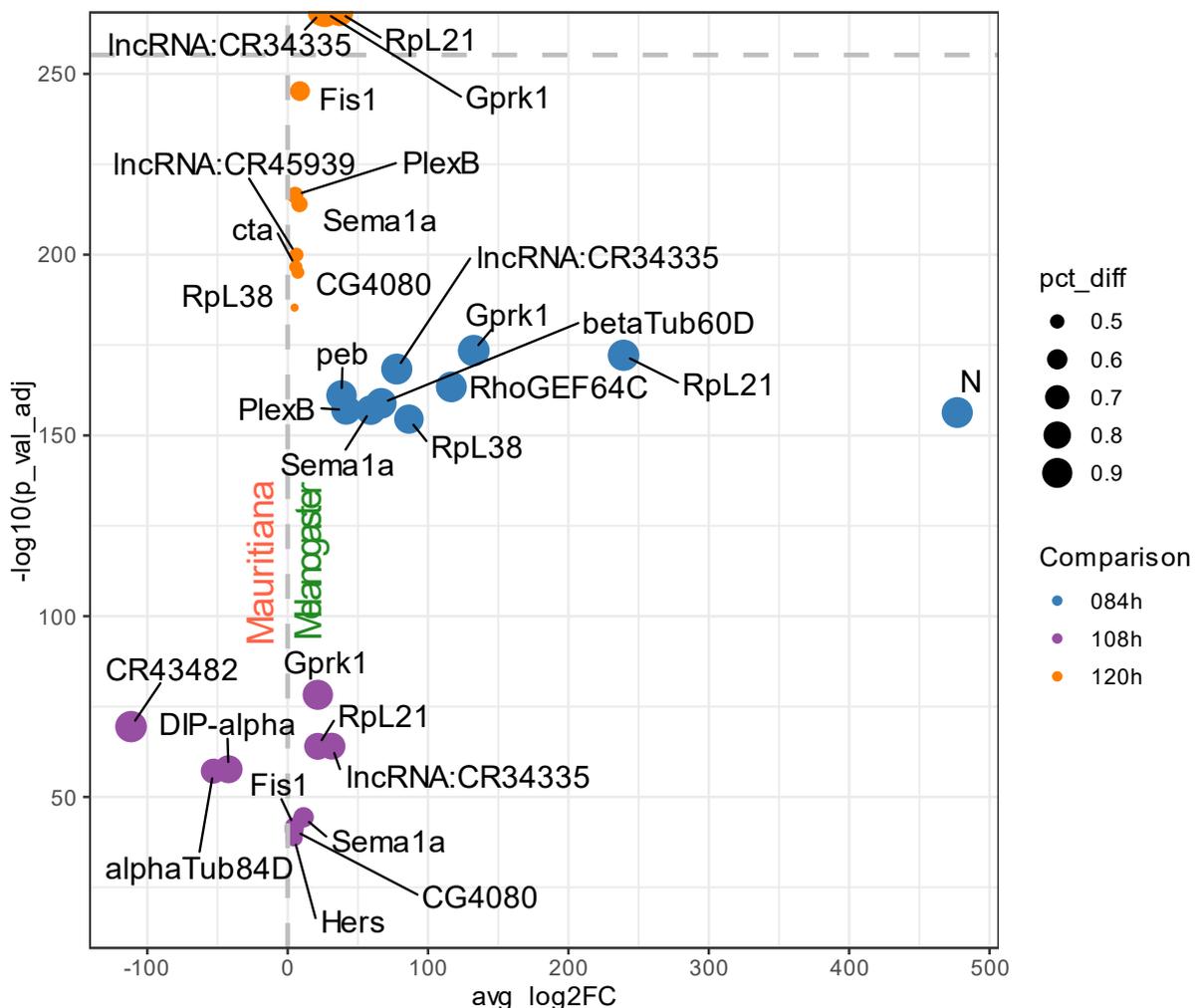


**Figure 20: Cell type composition in the EAD of *D. melanogaster* and *D. mauritiana*.** This plot depicts the composition of cell populations by species at three different timepoints, (A) 84 h AEL, (B) 108 h AEL, (C) 120 h AEL. The contributing numbers to each cluster of cells is corrected by the proportion of the total number of cells per species compared to each other. Some cells, such as interommatidial cells, could not be identified in all timepoints.

Most of the observed trends could be confirmed when pairwise integrated and cell number corrected data was analyzed. I found a disproportionately large number of cells expressing a photoreceptor or interommatidial marker signature in *D. mauritiana* as well (Figure 20). For photoreceptor cells, this bias is found at all stages (Figure 20A, B and C). At 120 h, also Interommatidial cells can be distinguished from other retinal cell types. Interestingly, the share of glia cells is also larger here. As is consistent with Figure 19, *D. melanogaster* appears to have a larger share of peripodial epithelium cells at all timepoints (“Ventral\_PE” Figure 20A, B and C, and also “PE” in Figure 20C). Both species have a near even share of antennal cells at 120 h (Figure 20C) and 108 h (Figure 20B), but I find a larger share of antennal cells in *D. mauritiana* at 84 h (Figure 20A). Consistent with Figure 19, cells in the second mitotic wave show a strong bias towards a larger share in *D. mauritiana* at 108 h (Figure 20B), which equalizes between the species at 120 h (Figure 20C). In summary, the most consistent differences in were observed for PhR, PPN and MF cell types. To understand the causes of this variation, I next analyzed the gene expression dynamics within the most variable cell types of the retina, photoreceptor cells, cells of the morphogenetic furrow, and pre-proneural cells.

### 3.4.1.1 Photoreceptor cell characterization between *D. melanogaster* and *D. mauritiana*.

Photoreceptor cells are the amongst the most characteristic cell type in the eye and enable the organs main function. Between *D. melanogaster* and *D. mauritiana* morphological variation can largely explained by ommatidia number, I expect the genes diverging between these species to have a lower impact since photoreceptor cells represent cells near completed differentiation. However, traces of global alterations to gene expression in the EAD might still be detectable here. A differential expression analysis specifically for the photoreceptor cells revealed 1,056 genes upregulated in *D. mauritiana* photoreceptors and 2,849 genes upregulated in *D. melanogaster* photoreceptors over all timepoints (Figure 21, extensive list in Supplementary Table S 11).



**Figure 21: Selected top-DEGs between photoreceptor cells of *D. mauritiana* and *D. melanogaster*.** Depicted are the top 10 significant ( $p < 0.05$ ) DEGs which show an expression percentage difference of over 20% and a log2 fold change of over 1. Heat shock response genes were excluded from this list. *D. mauritiana* is depicted with a negative- and *D. melanogaster* with a positive average log2 fold change. Pct\_diff: Percentage difference in cells expressing the gene between the two species.  $-\log_{10}(P\text{-values})$  above the horizontal dashed line are infinite. The complete list is in full list in Supplementary Table S 11.

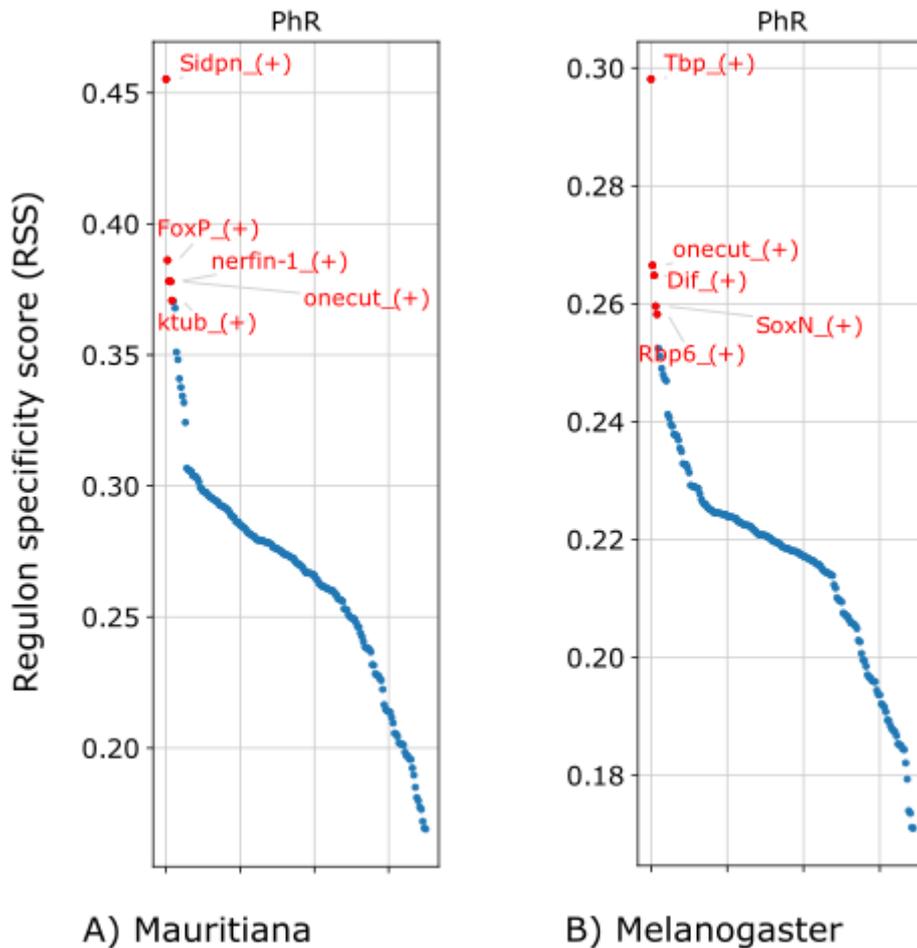
A GO term enrichment analysis for each stage showed top GO-terms related to tissue or organ development enriched in each species. More specific terms enriched in only one of the species include “Regulation of cell shape” and “Regulation of cell migration” enriched in *D. melanogaster* at 108 h (Supplementary Table S 11D), and “Semaphorin-plexin pathway involved in axon guidance” at 120 h (Supplementary Table S 11F). In *D. mauritiana*, I find “Animal organ morphogenesis” and “Positive regulation of cellular process” at 120 h (Supplementary Table S 11E). The highest number of significantly enriched GO-terms are visible at 84 h (Supplementary Table S 11A and B).

**Table 3: Photoreceptor specific DEGS between *D. melanogaster* and *D. mauritiana*.** Depicted are the top 10 genes by p-value, the total number of DEGs in each group is written in brackets.

| <b>Mau_84 h<br/>(103)</b> | <b>Mau_108 h<br/>(292)</b> | <b>Mau_120 h<br/>(225)</b> | <b>Mel_84 h<br/>(2410)</b> | <b>Mel_108 h (38)</b> | <b>Mel_120 h</b> |
|---------------------------|----------------------------|----------------------------|----------------------------|-----------------------|------------------|
| Rgk3                      | CR43482                    | dgo                        | peb                        | cathD                 | Ubi-p63E         |
| robo3                     | DIP-alpha                  | CG42339                    | N                          | obst-B                | RNASEK           |
| CG10019                   | alphaTub84D                | RpS13                      | bbg                        | alphaTub84B           | cathD            |
| Dscam3                    | CG1909                     | UQCR-C1                    | CG31176                    | CG31997               | HIP-R            |
| CG17193                   | lncRNA:noe                 | Pcmt                       | eya                        | AdamTS-B              | obst-B           |
| dpr15                     | Rbp6                       | RpS3                       | kst                        | CG10465               | CG9691           |
| Tbh                       | ImpE2                      | Vta1                       | hh                         | lncRNA:CR44024        | CG17839          |
| NPFR                      | Gs2                        | Tis11                      | crb                        | Ogre                  |                  |
| CG14024                   | Hr4                        | goe                        | mamo                       | Blot                  |                  |
| Dh31-R                    | Cadps                      | Rpl118                     | stg                        | Hh                    |                  |

The DEGs I find are dominated by a species signal, covering cell type specific interspecies DEGs.

Table 3 shows genes which are species- specifically upregulated only in photoreceptor cells. This includes, for instance, *ImpE*-genes, such as *ImpE2* which appeared here upregulated in 108 h old *D. mauritiana* photoreceptors. They are predicted to be involved in morphogenesis of the EAD epithelium (Paine-Saunders et al. 1990).



**Figure 22: Regulons in 120 h AEL photoreceptor cells in (A) *D. mauritiana* and (B) *D. melanogaster*.** Regulons are ordered by descending regulon specificity score. The top 5 regulons are highlighted in red. Extensive lists of regulons can be found in Supplementary Table S 18.

To understand how genes are regulated in a specific cluster of cells, SCENIC was performed which identifies regulatory units called regulons within cell clusters. This analysis is species specific. Figure 22 shows an example output in *D. mauritiana* (A) and *D. melanogaster* (B) at 120 h. The top regulons found in both species include genes such as *onecut*, which is a transcription factor known to play a role in photoreceptor differentiation. It encodes a transcription factor that binds two different DNA motifs (*Cut* and *Hox*). It is mainly expressed in mature neurons and maintains neuronal identity, and expressed specifically in photoreceptor cells in third instar larvae (Nguyen et al. 2000).

**Table 4: Top 10 regulons in photoreceptor cells which are shared between *D. melanogaster* and *D. mauritiana* and regulons which are species specific by descending regulon specificity score. Regulons highlighted in green appear in multiple cell populations. A (+) Indicates that the genes characteristic for this cell type are positively correlated with the listed TF. Red: genes which are found among the top 30 DEGs in the respective species in this cell type as well.**

| Age   | Shared_MelMau   | Mel                   | Mau          |
|-------|-----------------|-----------------------|--------------|
| 120 h | Tbp_(+)         | Dif_(+)               | Sidpn_(+)    |
|       | onecut_(+)      | Rbp6_(+)              | FoxP_(+)     |
|       | SoxN_(+)        | CG9727_(+)            | nerfin-1_(+) |
|       | Atf3_(+)        | run_(+)               | ktub_(+)     |
|       | REPTOR-BP_(+)   | MTF-1_(+)             | CG18599_(+)  |
|       | sv_(+)          | salm_(+)              | bru3_(+)     |
|       | l_(3)neo38_(+)  | Abl_(+)               | fd59A_(+)    |
|       | gl_(+)          | E_(spl)mbeta-HLH_(+)  | svp_(+)      |
|       | Jra_(+)         | ewg_(+)               | Hr4_(+)      |
|       | pnt_(+)         | E_(spl)m5-HLH_(+)     | bsh_(+)      |
| 108 h | ERR_(+)         | ase_(+)               | nerfin-1_(+) |
|       | run_(+)         | gcm2_(+)              | Rbp6_(+)     |
|       | Dif_(+)         | dmrt99B_(+)           | CG9650_(+)   |
|       | onecut_(+)      | Myb_(+)               | so_(+)       |
|       | sv_(+)          | E_(spl)mdelta-HLH_(+) | dl_(+)       |
|       | crc_(+)         | bru3_(+)              | CG4730_(+)   |
|       | l_(3)neo38_(+)  | Atf3_(+)              | Drgx_(+)     |
|       | gl_(+)          | pros_(+)              | Pbp95_(+)    |
|       | pnt_(+)         | SoxN_(+)              | Argk_(+)     |
|       | sr_(+)          | vnd_(+)               | crp_(+)      |
| 84 h  | Hr38_(+)        | Rbp6_(+)              | pdm2_(+)     |
|       | lab_(+)         | Dif_(+)               | Hr3_(+)      |
|       | Nfl_(+)         | sug_(+)               | dimm_(+)     |
|       | vri_(+)         | gl_(+)                | en_(+)       |
|       | E5_(+)          | Eip75B_(+)            | CG3065_(+)   |
|       | sr_(+)          | zfh1_(+)              | eve_(+)      |
|       | Mef2_(+)        | mor_(+)               | bru3_(+)     |
|       | Pdp1_(+)        | l(3)neo38_(+)         | tj_(+)       |
|       | Usf_(+)         | klu_(+)               | svp_(+)      |
|       | CHES-1-like_(+) | Sirt6_(+)             | slou_(+)     |

To understand which genes are conserved and diverging between species, I separated the identified regulons of photoreceptors by species. In the top 10 regulons shared between both species, ordered by regulon specificity score, I again find *onecut* amongst the top regulators at both 108 h and 120 h. Other genes shared between these stages as well as species are *pointed* (*pnt*), *shaven* (*sv*), genes with known roles in photoreceptor differentiation (Rawlins et al. 2003; O'Neill et al. 1994; Charlton-Perkins et al. 2011) and *glass* (*gl*), which is expressed in all cells

in the MF and posterior to the MF (Ellis et al. 1993). Interestingly, *gl* is found to be specific to *D. melanogaster* at 84 h.

Genes with no reported specific functions in photoreceptor development at 120 h include *REPTOR-binding partner (REPTOR-BP)* and *Activating transcription factor 3 (Atf3)*, which is best known for its role in the immune system (Rynes et al. 2012) but also has a reported function in the development of the abdomen (Sekyrova et al. 2010). *Atf3* appears as *D. melanogaster* specific at 108 h. At this timepoint, it is also among the top 30 photoreceptor specific DEGs upregulated in *D. melanogaster*. The top hit at 120 h, TATA binding protein (*Tbp*) encodes a basal transcription factor required at most RNA Pol I and Pol II-transcribed genes. This gene has no reported function related to photoreceptor cells.

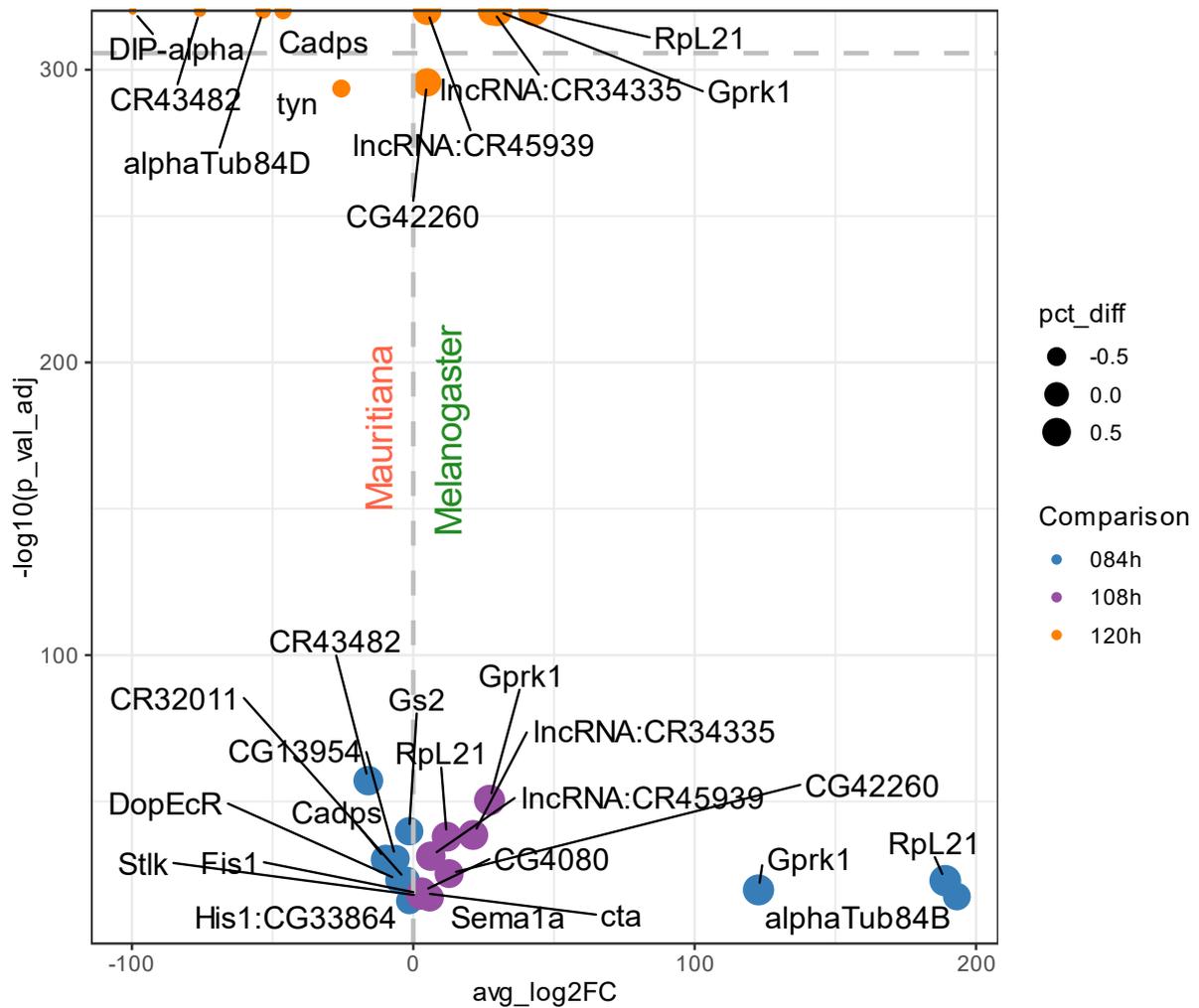
As the top *D. melanogaster* specific hits, I find *Dorsal-related immunity factor (Dif)* and *RNA-binding protein 6 (Rbp6)*. While there is no reported function for *Dif* in the EAD, *Rbp6* known to be expressed posterior to the MF (Siddall et al. 2012). It is an RNA binding protein inferred to be involved in the regulation of translation (Gaudet et al. 2011). In the larva, it is known to be expressed in the brain and the ventral nerve cord (Siddall et al. 2012). Both of these genes are also specific to *D. melanogaster* at 84 h, but *Rbp6* is found *D. mauritiana* specific at 108 h. Here, it is unique to *D. mauritiana* and found as a DEG as well (Table 3).

As a photoreceptor specific hit, *spalt major (salm)* encodes a zinc finger transcriptional repressor. It mediates most *dpp* functions during development of the central part of the wing through regulation of the products of *kni* and *ara*. The product of *salm* is required for cell specification during the development of the nervous system, muscle and trachea and inferred to be involved in photoreceptor cell differentiation (Mollereau et al. 2001; Domingos et al. 2004). Interestingly, at both 120 h and 108 h, *enhancers of split (E(spl))* genes appear as top hits in *D. melanogaster*. These genes are more characteristic for the MF. The top hit for *D. mauritiana* is *Similar to deadpan (Sidpn)*. This gene is reportedly expressed in the eye in adult *D. melanogaster* (Xu et al. 2004), but its functions are so far only predicted, for instance to enable DNA-binding transcription factor activity and to be involved in anterior-posterior pattern specification as well as regulating neurogenesis (Gaudet et al. 2011). At both 108 h and 84 h, I find *stripe (sr)* among the highest shared regulators in photoreceptor cells. This gene has no function reported connected to photoreceptors specifically. The top shared regulator at 84 h is *Hr38*, which is a hormone receptor, which, among other functions, is part of the ecdysone pathway (Kozlova et al. 1998). The least specific regulon of the shared top 10 is *Checkpoint suppressor 1-like (CHES-1 like)*, which I otherwise find as a major regulon in PPN (see Table

8 and Table 14). *Hormone receptor 4 (Hr4)* is found in the top 10 *D. mauritiana* specific regulons at 120 h. At 108 h it is both in the top 30 *D. mauritiana* specific regulons (Supplementary Table S 18), and also found as an upregulated DEG (Table 3, red). This gene is a central component of the ecdysone pathway (King-Jones et al. 2005).

#### 3.4.1.2 Morphogenetic Furrow cell characterization between *D. melanogaster* and *D. mauritiana*

The largest differences in cell population sizes are visible at 84 h, (Figure 19) with a proportional bias towards *D. melanogaster*. (Figure 20). Due to the decisive role of the MF in the determination of ommatidial cell fates, this cell type may influence the further development of its successor cells. If it does so in species specific manner, this difference could be visible in its gene expression patterns. For *D. mauritiana*, I find 624 upregulated DEGs in this cluster over all timepoints. For *D. melanogaster*, I find 205 upregulated DEGs in MF cells over all timepoints. (Figure 23, extensive list in Supplementary Table S 12).



**Figure 23: Selected top-DEGs between morphogenetic furrow cells of *D. mauritiana* and *D. melanogaster*.** Depicted are the top 10 significant ( $p < 0.05$ ) DEGs which show an expression percentage difference of over 20% and a log2 fold change of over 1. Heat shock response genes were excluded from this list. Mau: *D. mauritiana*. Mel: *D. melanogaster*. Pct\_diff: Percentage difference in cells expressing the gene between the two species. The complete list is in full list in Supplementary Table S 12.

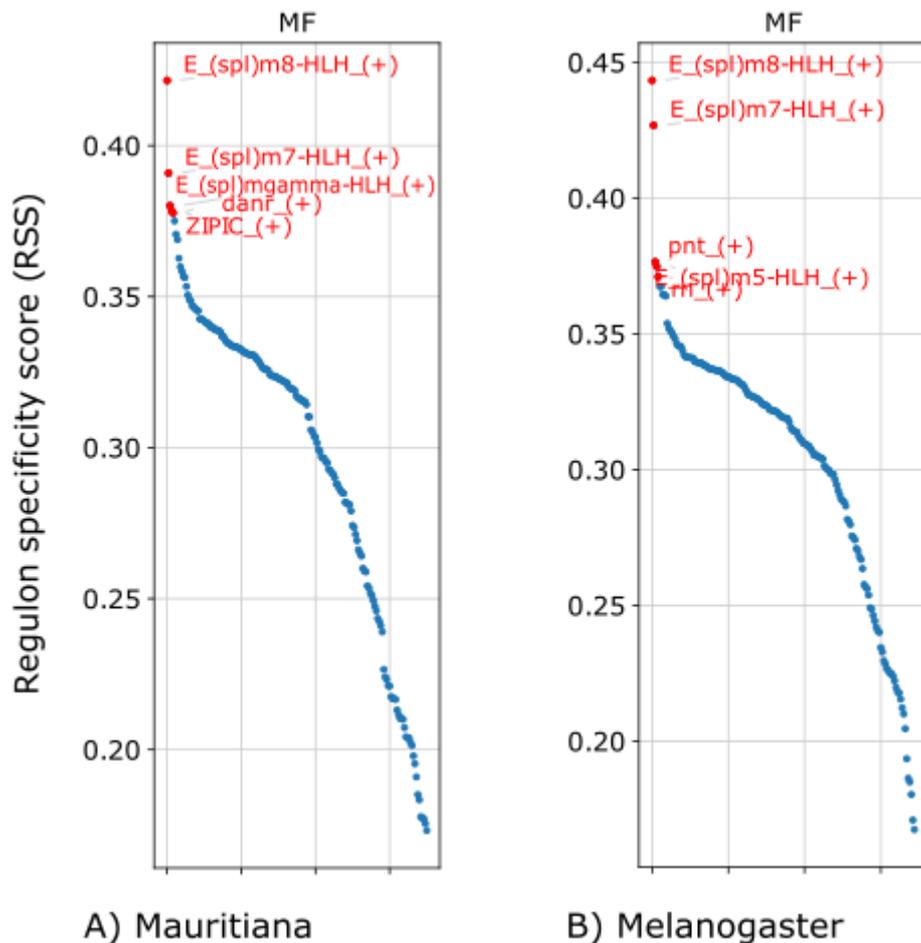
Outstanding genes among the DEGs between *D. mauritiana* and *D. melanogaster* are most importantly *G protein-coupled receptor kinase 1 (Gprk1)*. While it is known to be involved in the transduction of visible light (Lee et al. 2004), there is no role described in EAD development. I find it to be upregulated in *D. melanogaster* compared to *D. mauritiana* in all three timepoints (Figure 23). I find a large number of significantly enriched GO-terms from these genes related to developmental processes and tissue differentiation at 120 h in both *D. melanogaster* and *D. mauritiana* (Supplementary Table S 12E and F). In *D. mauritiana*, “neurogenesis” and “epithelium development” appear as the most specific developmental terms (Supplementary Table S 12E). At 84 h I find “cytoplasmic translation” as most significantly enriched. in *D. melanogaster* (Supplementary Table S 12B). Beyond that, there are several

differentially expressed pseudogenes such as *CR32011*, *CR43482* upregulated in *D. mauritiana* at 84 h, several long non-coding RNA genes such as lncRNA:*CR34335* and sparsely described genes such as *Ste20-like kinase (Stlk)* which are upregulated in *D. melanogaster* at 108 h.

**Table 5: MF specific DEGS between *D. melanogaster* and *D. mauritiana*.** Depicted are the top 10 genes by p-value, the total number of DEGs in each group is written in brackets.

| Mau_84 h | Mau_108 h | Mau_120 h | Mel_84 h<br>(20) | Mel_108 h | Mel_120 h   |
|----------|-----------|-----------|------------------|-----------|-------------|
| CG13954  |           |           | nmo              | unc-13    | alphaTub84B |
| DopEcR   |           |           | ctp              | Trn       |             |
| Alg-2    |           |           | stai             |           |             |
| CG30460  |           |           | Rpl8             |           |             |
| CG5888   |           |           | Calr             |           |             |
| CG11700  |           |           | Hsc70-4          |           |             |
| Msr-110  |           |           | Fkbp12           |           |             |

Looking at the genes which out of all three cell types are found as differentially expressed only in MF cells, there are no genes found uniquely upregulated in stages older than 84 h in *D. mauritiana* (Table 5). For *D. melanogaster*, there are a few genes upregulated also at 108 h and 120 h. At 108 h these genes are *tartan (trn)* and *unc-13*. *Trn* is a gene reportedly expressed in several sensory organ precursors, including the EAD (Chang et al. 1993). *Unc-13* is a gene known to be expressed in the adult retina (Xu et al. 1998). The only gene specifically upregulated at 120 h is  *$\alpha$ -Tubulin at 84B (alphaTub84B)*, encoding tubulin, a rather ubiquitously expressed gene as it provides a component of microtubules (Matthews et al. 1989). At 84 h, I find *nmo* to be specific to *D. melanogaster* in this cluster, which, as described earlier (3.1.2 and 3.1.4), interacts with members of the retinal determination network.



**Figure 24: Regulons in 120 h AEL morphogenetic furrow cells in (A) *D. mauritiana* and (B) *D. melanogaster*.** Regulons are ordered by descending regulon specificity score. The top 5 regulons are highlighted in red. Extensive lists of regulons can be found in Supplementary Table S 18.

Applying SCENIC in cells of the MF, I expectedly find genes of the *Enhancer of split* family at 120 h as top regulons in both *D. melanogaster* and *D. mauritiana* (Figure 24). Beyond these, I find *pointed* (*pnt*), a gene known to be involved in photoreceptor development (O'Neill et al. 1994).

**Table 6: Top 10 regulons in morphogenetic furrow cells which are shared between *D. melanogaster* and *D. mauritiana* and Regulons which are species specific by descending regulon specificity score.** Regulons highlighted in green appear in multiple cell populations. A (+) Indicates that the genes characteristic for this cell type are positively correlated with the listed TF. Red: genes which are found among the top 30 DEGs in the respective species in this cell type as well.

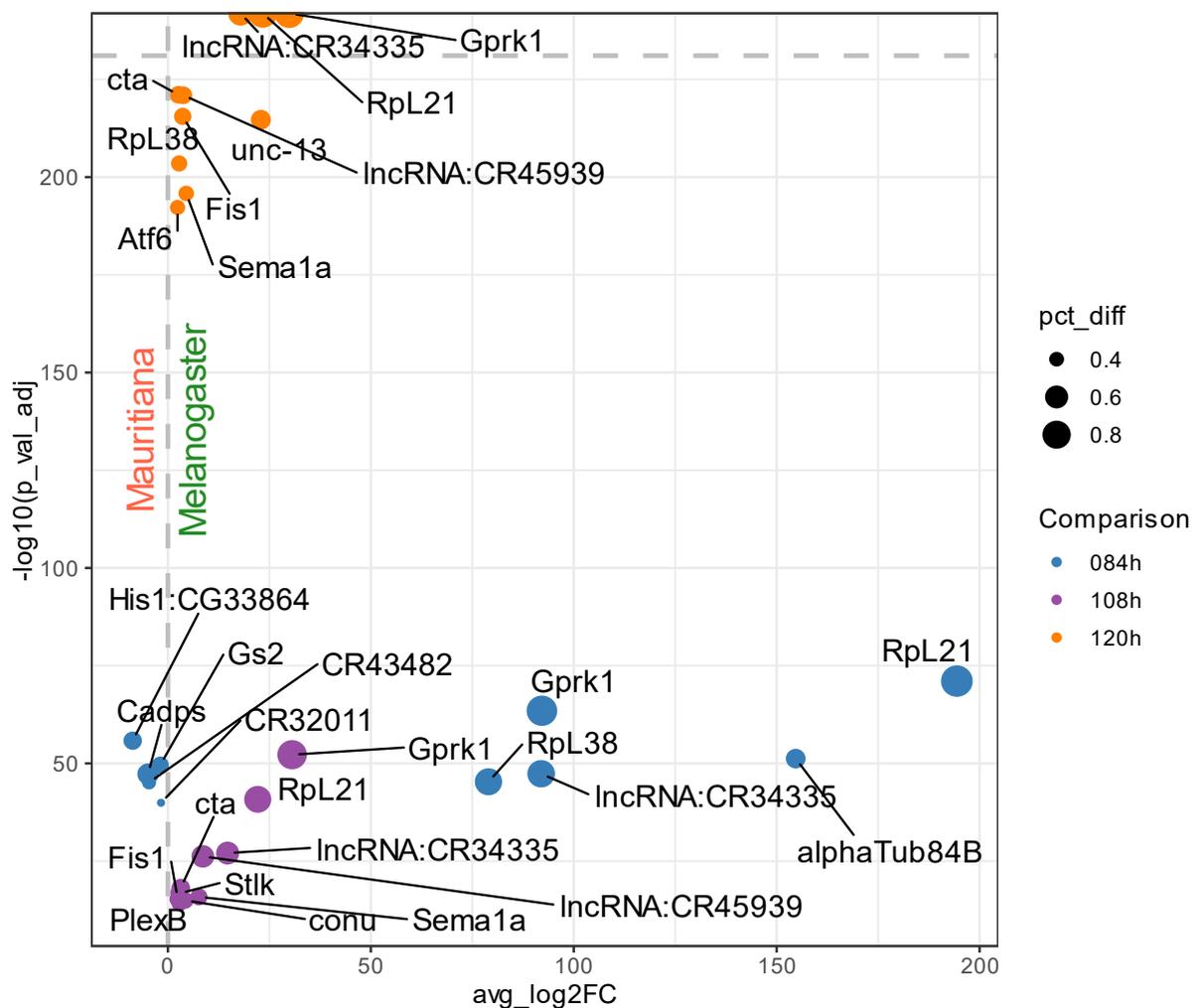
| Age   | Shared_MelMau     | Mel                   | Mau                   |
|-------|-------------------|-----------------------|-----------------------|
| 120 h | E_(spl)m8-HLH_(+) | E_(spl)m5-HLH_(+)     | E_(spl)mgamma-HLH_(+) |
|       | E_(spl)m7-HLH_(+) | E_(spl)m3-HLH_(+)     | danr_(+)              |
|       | pnt_(+)           | sage_(+)              | vri_(+)               |
|       | rn_(+)            | Hr78_(+)              | gcm_(+)               |
|       | Trf2_(+)          | Rbp6_(+)              | maf-S_(+)             |
|       | Nf-YC_(+)         | Cnx99A_(+)            | en_(+)                |
|       | foxo_(+)          | Irp-1B_(+)            | br_(+)                |
|       | Nf-YB_(+)         | tra2_(+)              | TfIIB_(+)             |
|       | aop_(+)           | Hr39_(+)              | cyc_(+)               |
|       | Pur-alpha_(+)     | Abl_(+)               | CG10348_(+)           |
| 108 h | run_(+)           | ase_(+)               | so_(+)                |
|       | Brf_(+)           | E_(spl)mdelta-HLH_(+) | E_(spl)mgamma-HLH_(+) |
|       | gl_(+)            | orb_(+)               | gsb-n_(+)             |
|       | E_(spl)m7-HLH_(+) | B-H1_(+)              | dl_(+)                |
|       | E_(spl)m8-HLH_(+) | E_(spl)m5-HLH_(+)     | esg_(+)               |
|       | pnt_(+)           | CG7101_(+)            | bigmax_(+)            |
|       | aop_(+)           | Atf3_(+)              | Sox14_(+)             |
|       | jumu_(+)          | bs_(+)                | Lim1_(+)              |
|       | Dif_(+)           | gcm2_(+)              | crol_(+)              |
|       | rn_(+)            | pros_(+)              | Kdm4B_(+)             |
| 84 h  | pnt_(+)           | E(spl)m8-HLH_(+)      | drm_(+)               |
|       | so_(+)            | E(spl)m7-HLH_(+)      | sens_(+)              |
|       | aop_(+)           | orb_(+)               | Ets65A_(+)            |
|       | Dll_(+)           | sc_(+)                | E_(spl)m3-HLH_(+)     |
|       | dl_(+)            | gl_(+)                | h_(+)                 |
|       | foxo_(+)          | Pur-alpha_(+)         | E_(spl)m7-HLH_(+)     |
|       | ey_(+)            | ftz-f1_(+)            | al_(+)                |
|       | Mitf_(+)          | l(3)neo38_(+)         | bon_(+)               |
|       | grh_(+)           | klu_(+)               | lz_(+)                |
|       | Stat92E_(+)       | slp1_(+)              | knrl_(+)              |

Regulons explaining gene expression in the MF in both species include several neuronal markers such as *pnt*, which was discussed above in photoreceptor cells. As expected, I also find several genes of the enhancer of split family. Interestingly, two genes of this family are consistently species-specific top regulons at 108 h as well as 120 h. These are *E(spl)m5* in *D. melanogaster* and *E(spl)mgamma-HLH* in *D. mauritiana*. The only gene found in the top 30 regulons in the MF using SCENIC which is also a species specific DEG in the MF is *Atf6*. It is upregulated in *D. melanogaster* in all three different cell types, (see for instance, PPN, Table

8). It is a known transcription factor, but no function in eye development or expression in the EAD is described yet. An important regulator I found among the top 5 in *D. mauritiana* at 120 h (Figure 24 and Table 6) is the gene *distal antenna-related (danr)*, a gene known to be involved with the retinal determination network (Suzanne et al. 2003).

### 3.4.1.3 Pre-proneural cell characterization between *D. melanogaster* and *D. mauritiana*

Pre-proneural cells show a similar trend between *D. mauritiana* and *D. melanogaster*. Nevertheless, within *D. melanogaster*, a larger part of all cells is identified as pre-proneural cells at every timepoint (Figure 19). For *D. mauritiana*, I find a total of 987 upregulated DEGs in this cluster over all timepoints. For *D. melanogaster*, I find a total of 318 upregulated DEGs in PPN cells over all timepoints (extensive list in Supplementary Table S 13).



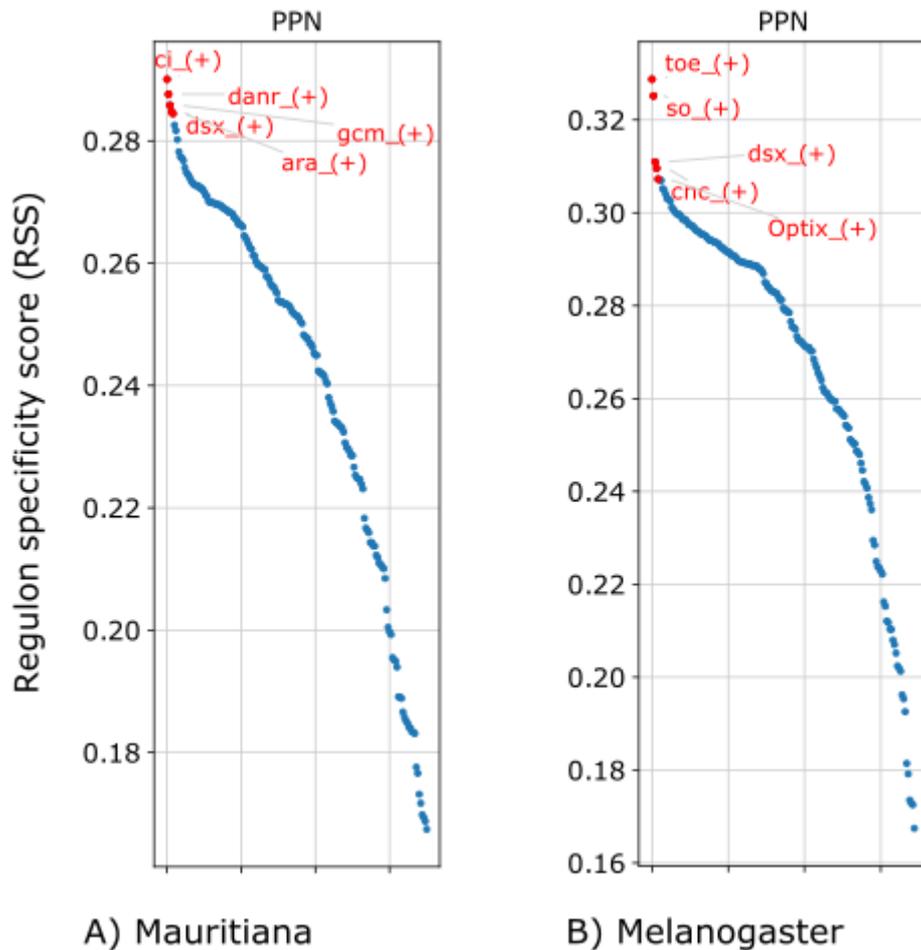
**Figure 25: Selected top-DEGs between pre-proneural cells of *D. mauritiana* and *D. melanogaster*.** Depicted are the top 10 significant ( $p < 0.05$ ) DEGs which show an expression percentage difference of over 20% and a log2 fold change of over 1. Heat shock response genes were excluded from this list. Mau: *D. mauritiana*. Mel: *D. melanogaster*. Pct\_diff: Percentage difference in cells expressing the gene between the two species. The complete list is in full list in Supplementary Table S 12.

*Semaphorin 1a (Sema1a)* is in the top 10 DEGs in the PPN between *D. melanogaster* and *D. mauritiana* (Figure 25) and indeed GO-Term enrichment analysis of the shows a significant enrichment of genes related to the semaphorin-plexin pathway at 84 h, 108 h, and 120 h in *D. melanogaster* (Supplementary Table S 13B, D, and F). In *D. mauritiana*, I again find an enrichment of neurogenesis related GO-terms (Supplementary Table S 13C and E). The overall composition of the top 10 DEGs in each timepoint is similar to those within the MF (Figure 24). Unique to the PPN cells in the top 10 here is *Plexin B (PlexB)*, another member of the Semaphorin-plexin pathway (Gaudet et al. 2011).

**Table 7: Pre-proneural cell specific DEGS between *D. melanogaster* and *D. mauritiana*.** Depicted are the top 10 genes by p-value, the total number of DEGs in each group is written in brackets.

| <b>Mau_84 h (24)</b> | <b>Mau_108 h (145)</b> | <b>Mau_120 h (140)</b> | <b>Mel_84 h (16)</b> | <b>Mel_108 h</b> | <b>Mel_120 h</b> |
|----------------------|------------------------|------------------------|----------------------|------------------|------------------|
| obst-A               | Dlg5                   | icl n                  | shi                  | CG6739           | CG6739           |
| His2A:CG33865        | RhoGAP93B              | unk                    | side-IV              | Paics            | CG14502          |
| DIP-alpha            | east                   | CG16791                | dnc                  | mRpS5            | sxc              |
| CR40461              | CG43861                | uif                    | CG44422              | Oatp74D          | rgn              |
| cac                  | dac                    | <b>klu</b>             | Fer1HCH              |                  |                  |
| His4:CG33905         | aPKC                   | tinc                   | caps                 |                  |                  |
| IA-2                 | jigr1                  | Myo10A                 | Rab11                |                  |                  |
| lncRNA:cherub        | Mical                  | Cp1                    | CG4502               |                  |                  |
| Obp44a               | CG3226                 | Nlp                    | tna                  |                  |                  |
| <b>bru3</b>          | CG42327                | CG43736                | p120ctn              |                  |                  |

In the PPN cells, I find a large number of genes upregulated specifically in *D. mauritiana* at all timepoints (Table 7). As found in other comparisons, I find a member of the ecdysone pathway, *Oatp74D*, upregulated in *D. melanogaster*. Other genes upregulated in *D. melanogaster* PPN have no reported function in eye development. The gene *CG6739* is upregulated at 108 h as well as 120 h and has only predicted functions (Hammonds et al. 2013). The genes *dachshund* (*dac*, 108 h) and *klumpfuss* (*klu*, 120 h), which are upregulated in *D. mauritiana*, play known and important roles in eye development (Anderson et al. 2017) (See also 2.2.1). I also find *klu* among the top 30 regulons in PPN at 108 h.



**Figure 26: Regulons in 120 h AEL pre-proneural cells in (A) *D. mauritiana* and (B) *D. melanogaster*.** Regulons are ordered by descending regulon specificity score, the top 5 Regulons are highlighted in red. Extensive lists of regulons can be found in Supplementary Table S 18.

To understand how genes are regulated in pre-proneural cells, SCENIC was performed which identifies regulatory units called regulons within cell clusters. This analysis was performed in each species separately. Figure 26 shows the output in *D. mauritiana* (A) and *D. melanogaster* (B) at 120 h. In both species, this analysis reveals for instance, *doublesex* (*dsx*) amongst the top regulons in 120 h (Figure 26) and 108 h (Table 8, first column) old PPN cells. This gene is known to play a major role in sexual differentiation of somatic cells and is reportedly expressed in neural cells in *D. melanogaster* larvae (Lee et al. 2002; Rideout et al. 2010). Apart from this,

**Table 8: Top 10 regulons in pre-proneural cells which are shared between *D. melanogaster* and *D. mauritiana* and regulons which are species specific by descending regulon specificity score.** Regulons highlighted in green appear in multiple cell populations. A (+) Indicates that the genes characteristic for this cell type are positively correlated with the listed TF. Red: genes which are found among the top 30 DEGs in the respective species in this cell type as well.

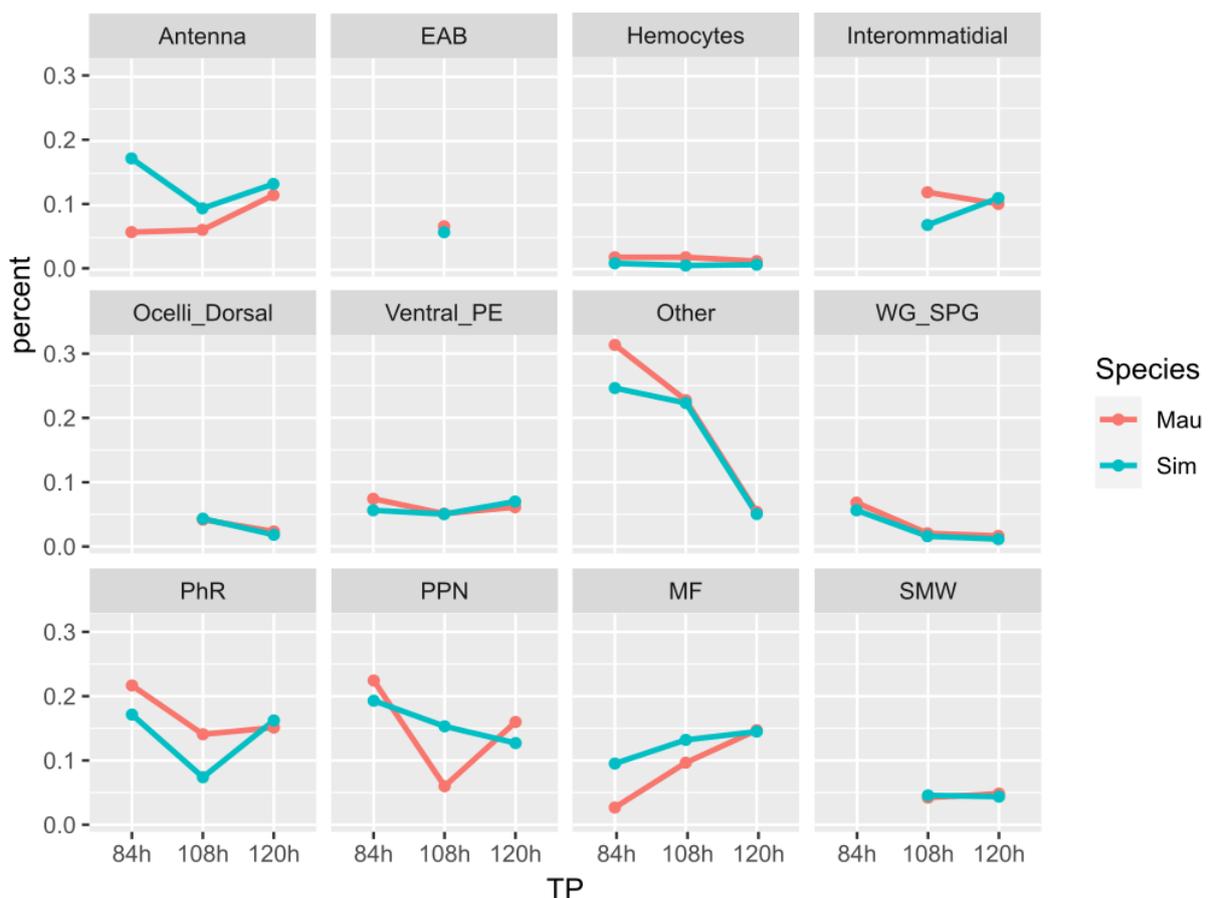
| Age    | Shared_MelMau   | Mel               | Mau                   |
|--------|-----------------|-------------------|-----------------------|
| 120 h  | so_(+)          | toe_(+)           | ci_(+)                |
|        | dsx_(+)         | Optix_(+)         | danr_(+)              |
|        | cnc_(+)         | Atf6_(+)          | gcm_(+)               |
|        | CHES-1-like_(+) | trx_(+)           | ara_(+)               |
|        | kni_(+)         | her_(+)           | bowl_(+)              |
|        | Dp_(+)          | Kdm2_(+)          | caup_(+)              |
|        | E2f1_(+)        | Cnx99A_(+)        | vri_(+)               |
|        | Hcf_(+)         | Irp-1B_(+)        | Sin3A_(+)             |
|        | crol_(+)        | sd_(+)            | br_(+)                |
|        | foxo_(+)        | ewg_(+)           | CG10348_(+)           |
| 108 h  | CHES-1-like_(+) | ara_(+)           | ey_(+)                |
|        | Pdp1_(+)        | caup_(+)          | Scr_(+)               |
|        | dsx_(+)         | toe_(+)           | Optix_(+)             |
|        | Dp_(+)          | ac_(+)            | CG10321_(+)           |
|        | jim_(+)         | vri_(+)           | ci_(+)                |
|        | nej_(+)         | Myc_(+)           | Hr96_(+)              |
|        | E2f1_(+)        | eyg_(+)           | Hr3_(+)               |
|        | EcR_(+)         | Atf6_(+)          | crol_(+)              |
|        | cnc_(+)         | ham_(+)           | Trl_(+)               |
|        | E_(bx)_(+)      | Cnx99A_(+)        | dar1_(+)              |
| Shared | Mel             | Mau               |                       |
| 84 h   | so_(+)          | toy_(+)           | h_(+)                 |
|        | ey_(+)          | ara_(+)           | Rbf2_(+)              |
|        | Mitf_(+)        | Hr4_(+)           | al_(+)                |
|        | dl_(+)          | slp1_(+)          | sd_(+)                |
|        | grh_(+)         | Syp_(+)           | bon_(+)               |
|        | aop_(+)         | Eip74EF_(+)       | Ets65A_(+)            |
|        | Spps_(+)        | E(bx)_(+)         | shn_(+)               |
|        | foxo_(+)        | pan_(+)           | mirr_(+)              |
|        | pnt_(+)         | Hcf_(+)           | E_(spl)m3-<br>HLH_(+) |
|        | Chd1_(+)        | Pur-<br>alpha_(+) | sob_(+)               |

In *D. melanogaster*, the gene *twin of eyegone* (*toe*) is worth mentioning. As the name suggests, this gene is involved in eye development where it acts as a transcriptional repressor (Czerny et al. 1999) and is known to be expressed in the eye disc (Yao et al. 2008; Jang et al. 2003). The gene *araucan* (*ara*) is interestingly found in the top 10 regulons in *D. melanogaster* at 108 h and in *D. mauritiana* at 120 h. This gene is a negative regulator of cell growth (Barrios et al.

2015) and involved in anterior-posterior patterning (Barrios et al. 2015) and inferred to be involved in compound eye development (Tomlinson 2003). In the adult, it is expressed in R7 and R8 photoreceptors (Mazzoni et al. 2008). *Checkpoint suppressor 1-like (CHES-1-like)* encodes a transcription factor that regulates cell division (Ahmad et al. 2012). It is known to upregulate *dpp* expression in some tissues (Yu et al. 2016). *Atf6* is upregulated in the PPN of *D. melanogaster* at 108 h and 120 h, where it is among the top 10 DEGs (Figure 25). It is described in 3.4.1.2.

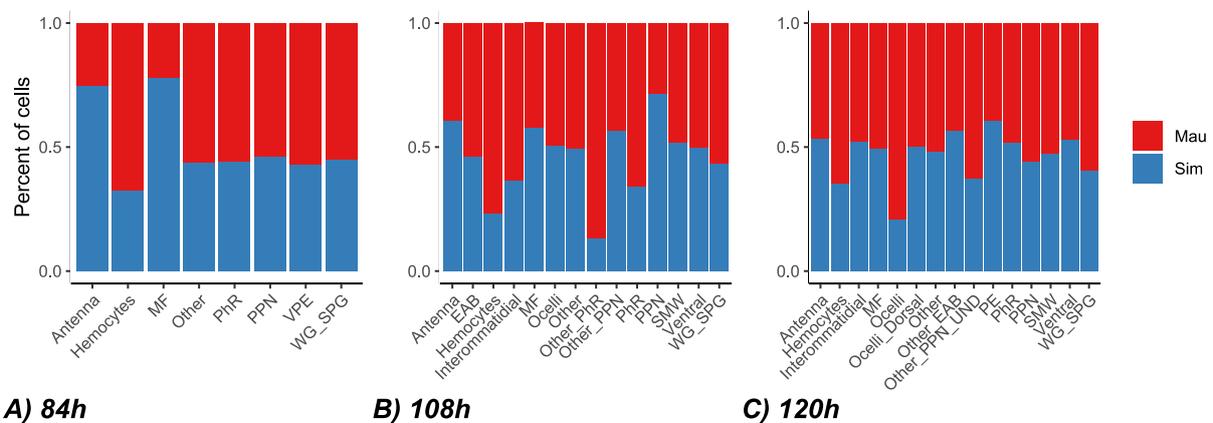
### 3.4.2 Divergence in cell type composition between *D. simulans* and *D. mauritiana*

The larger eye size of *D. mauritiana* compared to *D. simulans* is known to be explainable mostly by the larger size of ommatidia in *D. mauritiana*. Therefore, I expect few differences in cell numbers, and expect differences to lie mainly in gene expression levels.



**Figure 27: Progression of cell type proportions within *D. simulans* and *D. mauritiana*.** Points in this plot depict the share of cells within one species which are identified as a specific cell type. This plot includes only cell types which I could identify in both species.

Over time, most non-retinal cell types show similar tendencies of expansion and contraction between the two species (e.g. “Ventral\_PE”, “WG\_SPG” and “Antenna” from 108 h on, Figure 27). All retinal cell types reach similar proportions in both species at 120 h. The share of photoreceptor cells on the whole disc is highly divergent with a bias towards a larger proportion in *D. mauritiana* until 120 h, where both species end up with a similar proportion. A slightly increasing trend can be seen in the morphogenetic furrow again with a bias towards a larger proportion within *D. simulans*. Pre-proneural cells show a roughly decreasing trend in both species with a bias towards a larger proportion within *D. simulans*.

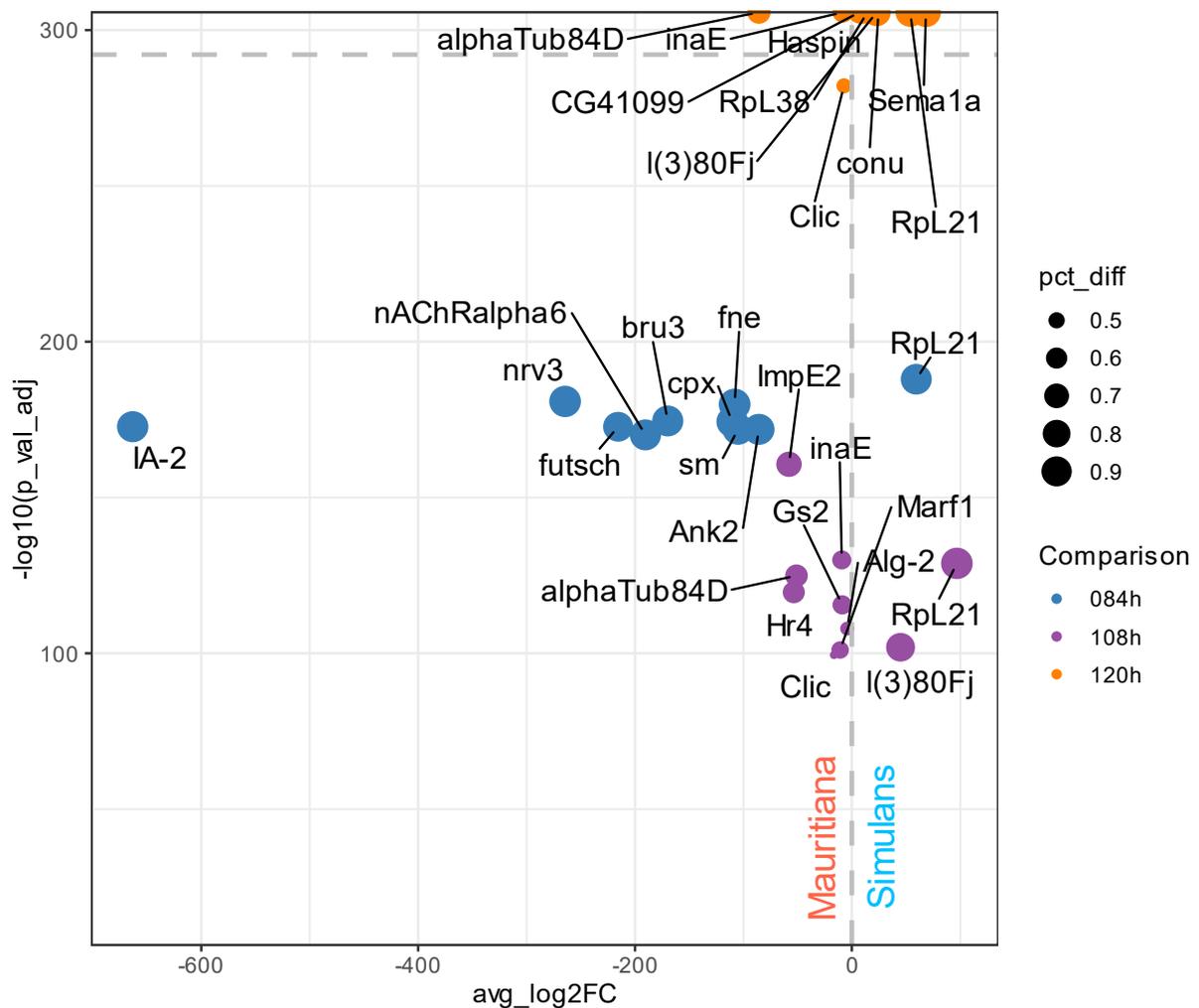


**Figure 28: Cell type composition in the EAD of *D. simulans* and *D. mauritiana*.** This plot depicts the composition of cell populations by species at three different timepoints, (A) 84 h AEL, (B) 108 h AEL, (C) 120 h AEL. The contributing numbers to each cluster of cells is corrected by the proportion of the total number of cells per species compared to each other. Some cells, such as interommatidial cells, could not be identified in all timepoints.

At 84 h AEL, most cell types are evenly represented between *D. simulans* and *D. mauritiana* (Figure 28A). Exceptions to this are a larger number of cells in antennal cells and cells of the morphogenetic furrow. In *D. simulans*, this bias however is absent from 108 h on (Figure 28B). At 108 h, I observe the strongest cell type biases between the two species. At this point, there is a larger number of PPN cells in *D. simulans* as well which is fainting at 120 h. The proportion of photoreceptor cells is slightly biased towards *D. mauritiana* at 84 h and 108 h, but evens out at 120 h. The number of ocelli cells ends up with a bias towards *D. mauritiana* at 120 h (Figure 28C). At this timepoint, most cell types are evenly represented between these two species. There is a slightly higher share in peripodial epithelium cells in *D. simulans* at 108 h, and a lower number of hemocytes, which are not part of the EAD proper.

### 3.4.2.1 Photoreceptor cell characterization between *D. simulans* and *D. mauritiana*

As presented above, I find similar numbers of cells expressing a photoreceptor marker gene signature between *D. simulans* and *D. mauritiana* in integrated data (Figure 28). Since I know of differences in ommatidia size between adult *D. mauritiana* and *D. simulans*, the causes for these differences likely lie in differential gene expression within cells making up the eye. In a differential expression analysis within the photoreceptor cells, I found a total of 1,087 genes upregulated in *D. mauritiana* photoreceptors and 302 genes upregulated in *D. melanogaster* photoreceptors over all timepoints (Figure 29, extensive list in Supplementary Table S 14).



**Figure 29: Selected top-DEGs between photoreceptor cells of *D. mauritiana* and *D. simulans*.** Depicted are the top 10 significant ( $p < 0.05$ ) DEGs which show an expression percentage difference of over 20% and a log2 fold change of over 1. Heat shock response genes were excluded from this list. *D. mauritiana* is depicted with a negative- and *D. simulans* with a positive average log2 fold change. Pct\_diff: Percentage difference in cells expressing the gene between the two species.  $-\log_{10}(P\text{-values})$  above the horizontal dashed line are infinite. The complete list is in full list in Supplementary Table S 14.

At all timepoints, for each species I find enriched GO-terms related to tissue development, indicating developmentally relevant differences between the species, from general terms such as “anatomical structure morphogenesis” to specific terms such as “photoreceptor cell axon guidance” (Supplementary Table S 14). One of the genes which is contributing to this is upregulated in *D. simulans* at late stages, *Sema1a* (Figure 29, orange). This gene is known to encode a transmembrane protein (Gaudet et al. 2011). It contributes most to the biological processes found enriched in genes upregulated in *D. simulans* both at 120 h and 108 h. Several related genes are also upregulated (e.g. *Sema2a*) or downregulated (e.g. *Sema5c*,  $p\_val\_adj = 5.82E-08$ ,  $avg\_log2FC = 4.0$ , *Sema2b*,  $avg\_log2FC -19$ ,  $p > 0.05$ ) in *D. simulans* compared to *D. mauritiana* (Supplementary Table S 14).

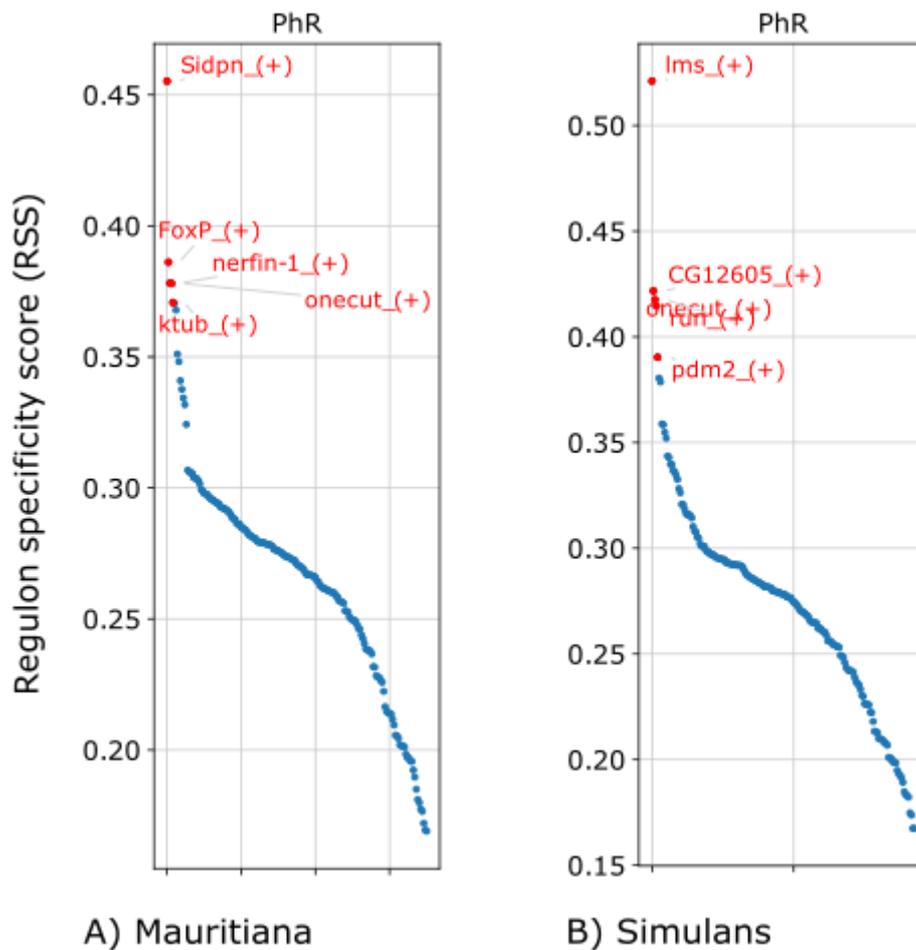
Further biological differences in *D. simulans* appear to be driven by the gene *Haspin*, which is involved in regulation of chromosome organization during mitosis. In the data, there is a read bias towards *D. mauritiana*, which may cause the GO results for genes upregulated in *D. simulans* to be less significant than genes upregulated in *D. mauritiana*. For instance, 173 genes of the input query map to the term “animal organ development” with the top genes including *MKP3* and *Nckx30C* (Supplementary Table S 14).

**Table 9: Photoreceptor specific DEGS between *D. simulans* and *D. mauritiana*.** Depicted are the top 10 genes by p-value, the total number of DEGs in each group is written in brackets.

| Mau_84 h<br>(643) | Mau_108 h<br>(23) | Mau_120 h<br>(15) | Sim_84 h<br>(83) | Sim_108 h<br>(53) | Sim_120 h<br>(13) |
|-------------------|-------------------|-------------------|------------------|-------------------|-------------------|
| Cam               | Eip93F            | Ca-alpha1T        | fru              | CG32982           | lncRNA:CR44268    |
| nAChRalpha7       | Nckx30C           | CG9170            | nw               | CG9691            | PlexB             |
| CG2269            | Hou               | Hou               | N                | CG46385           | vtd               |
| mei-P26           | kek3              | AMPdeam           | Fas3             | vtd               | CG30428           |
| CG42540           | Pde1c             | Nckx30C           | dpy              | CG6040            | MED21             |
| Hk                | Rbp6              | Rip11             | stg              | CG7432            | Yp3               |
| CG44422           | jeb               | ScpX              | grh              | PlexB             | CG7194            |
| Eno               | GlcAT-P           | His2A:CG33865     | rols             | CG30428           | lncRNA:CR46483    |
| pan               | CG42337           | sn                | ds               | exp               | Cht2              |
| igl               | chp               | pHCl-1            | uif              | KrT95D            | dpr8              |
| Stacl             | Ca-beta           | GCS2beta          | eya              | CG40191           | DIP1              |
| Gapdh2            | Dscam2            | GlcAT-P           | baz              | Akr1B             | sfl               |
| NaCP60E           | Ace               | qua               | fj               | su(f)             | Pdk1              |

*Nckx30C* is also amongst the top cell type exclusive DEGs in *D. mauritiana* at both 108 h as well as 120 h (Table 9). Its product is a calcium-potassium exchanger which is known to be expressed in the adult eye (Haug-Collet et al. 1999) and only predicted to play a role in

compound eye development (Webel et al. 2002). Furthermore, there is a large number of sparsely described genes upregulated especially in *D. simulans* at different stages *CG32982*, *CG9691*, *CG46385*, *CG6040*, *CG7432*, *CG30428*, *CG40191*, *CG7194*). One of which, *CG30428*, is upregulated at both 108 h and 120 h. The same is true for *verthandi* (*vtd*), a member of the cohesin complex, which is known to play a role in gene expression (Hallson et al. 2008), but not yet specifically in eye development.



**Figure 30: Regulons in 120 h AEL photoreceptor cells in (A) *D. mauritiana* and (B) *D. simulans*.** Regulons are ordered by descending regulon specificity score. The top 5 regulons are highlighted in red. Extensive lists of regulons can be found in Supplementary Table S 18.

To understand how genes are regulated in a specific cluster of cells, SCENIC was performed which identifies regulatory units called regulons within cell clusters. This analysis is species specific. Figure 30 shows an example output in *D. mauritiana* (A) and *D. simulans* (B) at 120 h. Among the top common regulators is, as was the case between *D. melanogaster* and *D.*

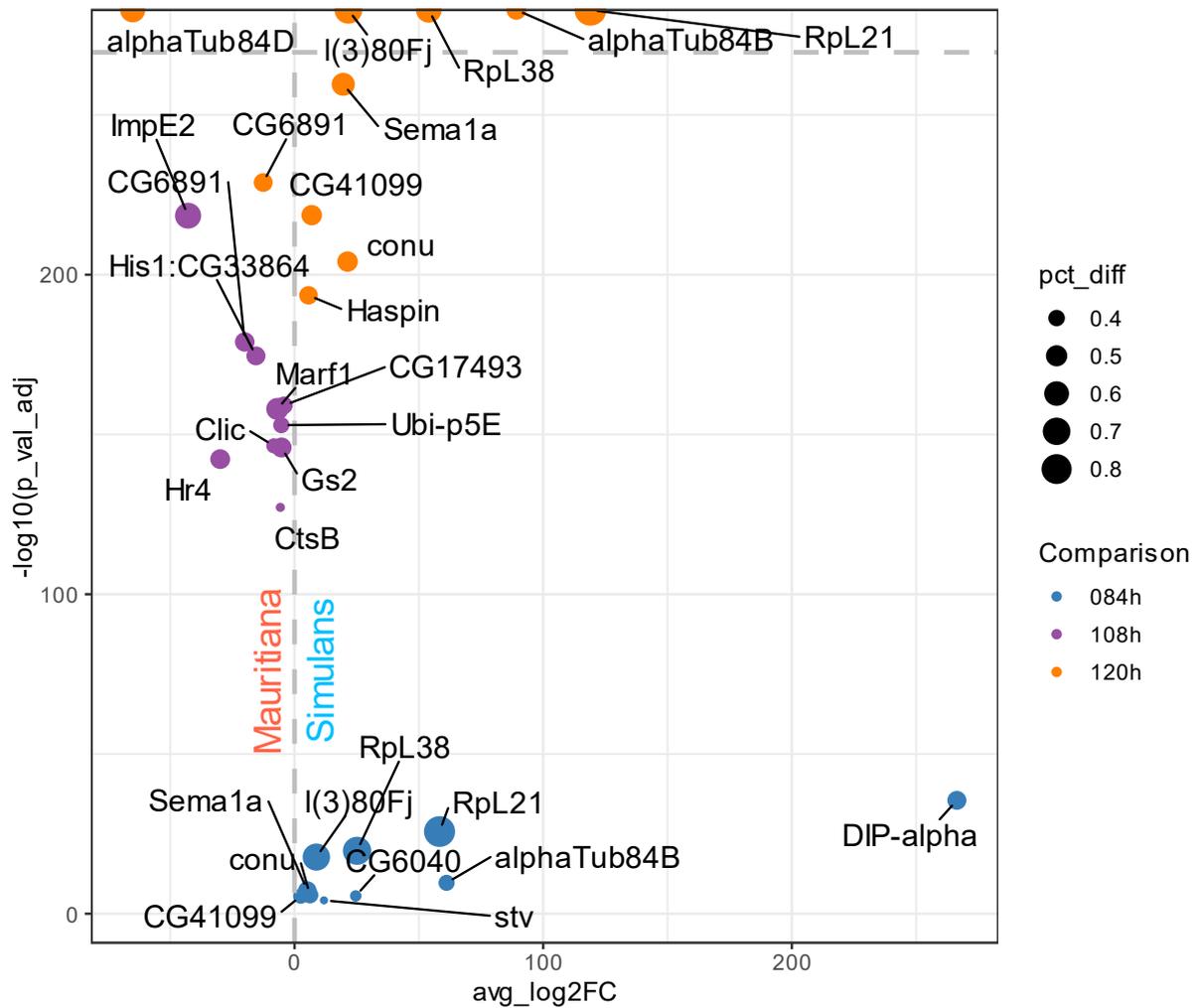
*mauritiana*, the gene *onecut*. While their specific ranks may differ, similar highly specific regulons are found in photoreceptors between different species.

**Table 10: Top 10 regulons in photoreceptor cells which are shared between *D. simulans* and *D. mauritiana* and Regulons which are species specific by descending regulon specificity score.** Regulons highlighted in green appear in multiple cell populations. A (+) Indicates that the genes characteristic for this cell type are positively correlated with the listed TF. Red: genes which are found among the top 30 DEGs in the respective species in this cell type as well.

| Age   | Shared_SimMau                          | Sim                                    | Mau                  |
|-------|--|--|----------------------|
| 120 h | <i>onecut</i> _(+)                     | <i>lms</i> _(+)                        | <i>Sidpn</i> _(+)    |
|       | CG18599_(+)                            | <i>CG12605</i> _(+)                    | <i>FoxP</i> _(+)     |
|       | <i>Atf3</i> _(+)                       | <i>run</i> _(+)                        | <i>nerfin-1</i> _(+) |
|       | <i>gl</i> _(+)                         | <i>pdm2</i> _(+)                       | <i>ktub</i> _(+)     |
|       | <i>SoxN</i> _(+)                       | <i>Dif</i> _(+)                        | <i>bru3</i> _(+)     |
|       | <i>svp</i> _(+)                        | <i>Nfl</i> _(+)                        | <i>fd59A</i> _(+)    |
|       | <i>Jra</i> _(+)                        | <i>gce</i> _(+)                        | <i>Hr4</i> _(+)      |
|       | <i>crp</i> _(+)                        | <i>ase</i> _(+)                        | <i>bsh</i> _(+)      |
|       | <i>l</i> _(3) <i>neo38</i> _(+)        | <i>vtd</i> _(+)                        | <i>Ets65A</i> _(+)   |
|       | <i>Spps</i> _(+)                       | <i>Kdm4B</i> _(+)                      | <i>TfIIB</i> _(+)    |
| 108 h | <i>onecut</i> _(+)                     | <i>vnd</i> _(+)                        | <i>nerfin-1</i> _(+) |
|       | <i>run</i> _(+)                        | <i>SoxN</i> _(+)                       | <i>Rbp6</i> _(+)     |
|       | <i>ERR</i> _(+)                        | <i>E</i> _(spl) <i>mdelta-HLH</i> _(+) | <i>CG9650</i> _(+)   |
|       | <i>Brf</i> _(+)                        | <i>CG12605</i> _(+)                    | <i>Dif</i> _(+)      |
|       | <i>l</i> _(3) <i>neo38</i> _(+)        | <i>dmrt99B</i> _(+)                    | <i>Mondo</i> _(+)    |
|       | <i>sv</i> _(+)                         | <i>bru3</i> _(+)                       | <i>CG4730</i> _(+)   |
|       | <i>gl</i> _(+)                         | <i>Atf3</i> _(+)                       | <i>Drgx</i> _(+)     |
|       | <i>pnt</i> _(+)                        | <i>E</i> _(spl) <i>m5-HLH</i> _(+)     | <i>Pbp95</i> _(+)    |
|       | <i>E</i> _(spl) <i>mgamma-HLH</i> _(+) | <i>CG43347</i> _(+)                    | <i>Argk</i> _(+)     |
|       | <i>crc</i> _(+)                        | <i>ap</i> _(+)                         | <i>Kdm4B</i> _(+)    |
| 84 h  | <i>Nfl</i> _(+)                        | <i>Nf-YA</i> _(+)                      | <i>Hr38</i> _(+)     |
|       | <i>en</i> _(+)                         | <i>Brf</i> _(+)                        | <i>pdm2</i> _(+)     |
|       | <i>E5</i> _(+)                         | <i>unc-4</i> _(+)                      | <i>Hr3</i> _(+)      |
|       | <i>sr</i> _(+)                         | <i>Rbp6</i> _(+)                       | <i>lab</i> _(+)      |
|       | <i>CG3065</i> _(+)                     | <i>Dif</i> _(+)                        | <i>dimm</i> _(+)     |
|       | <i>Mef2</i> _(+)                       | <i>tgo</i> _(+)                        | <i>vri</i> _(+)      |
|       | <i>svp</i> _(+)                        | <i>Smox</i> _(+)                       | <i>eve</i> _(+)      |
|       | <i>gt</i> _(+)                         | <i>sug</i> _(+)                        | <i>bru3</i> _(+)     |
|       | <i>B-H1</i> _(+)                       | <i>gl</i> _(+)                         | <i>tj</i> _(+)       |
|       | <i>grn</i> _(+)                        | <i>ftz-f1</i> _(+)                     | <i>Pdp1</i> _(+)     |

*Vtd*, differentially upregulated in *D. simulans* photoreceptor cells, is also found as a *D. simulans* specific top regulator at 120 h by SCENIC (Table 10, highlighted red). Apart from *onecut I* again find *sv* and *gl* (see 3.4.1.1). As was the case between *D. melanogaster* and *D. mauritiana* as well, the top regulator for *D. mauritiana* at 120 h is *Sidpn* (Figure 30 A, see also 3.4.1.1). Other genes, such as *Sp1-like factor for pairing sensitive-silencing (Spps)*, were not found in previous comparisons as shared regulons and have no described function related to eye- and head development yet. This also includes non-shared regulons such as *Glial precursor HLH protein (HLH)*. Genes shared by these species *ATF3* is known for a role in abdominal morphogenesis and upregulated in *D. simulans* as well (Sekyrova et al. 2010). A gene upregulated in *D. simulans* is *Smad on X (Smox)* at 84 h (Wells et al. 2017). This gene has a described roll in R8 photoreceptor differentiation but is found in this data as highly specific in photoreceptors uniquely in this species.

3.4.2.2 Morphogenetic Furrow cell characterization between *D. simulans* and *D. mauritiana*  
In *D. simulans* and *D. mauritiana*, the share of cells of the MF within each species follows a similar trajectory as between *D. melanogaster* and *D. mauritiana*, with proportions evening out at 120 h (Figure 27 and Figure 28). Here, there is just a slight bias towards a higher share of MF cells in *D. simulans* before 120 h (Figure 27). I investigate the possible divergence in gene expression between these species due to the central role of the MF in eye development. For *D. mauritiana*, I find 162 upregulated DEGs in this cluster over all timepoints. For *D. simulans*, I find 91 upregulated DEGs in MF cells over all timepoints. (Figure 31, extensive list in Supplementary Table S 15).



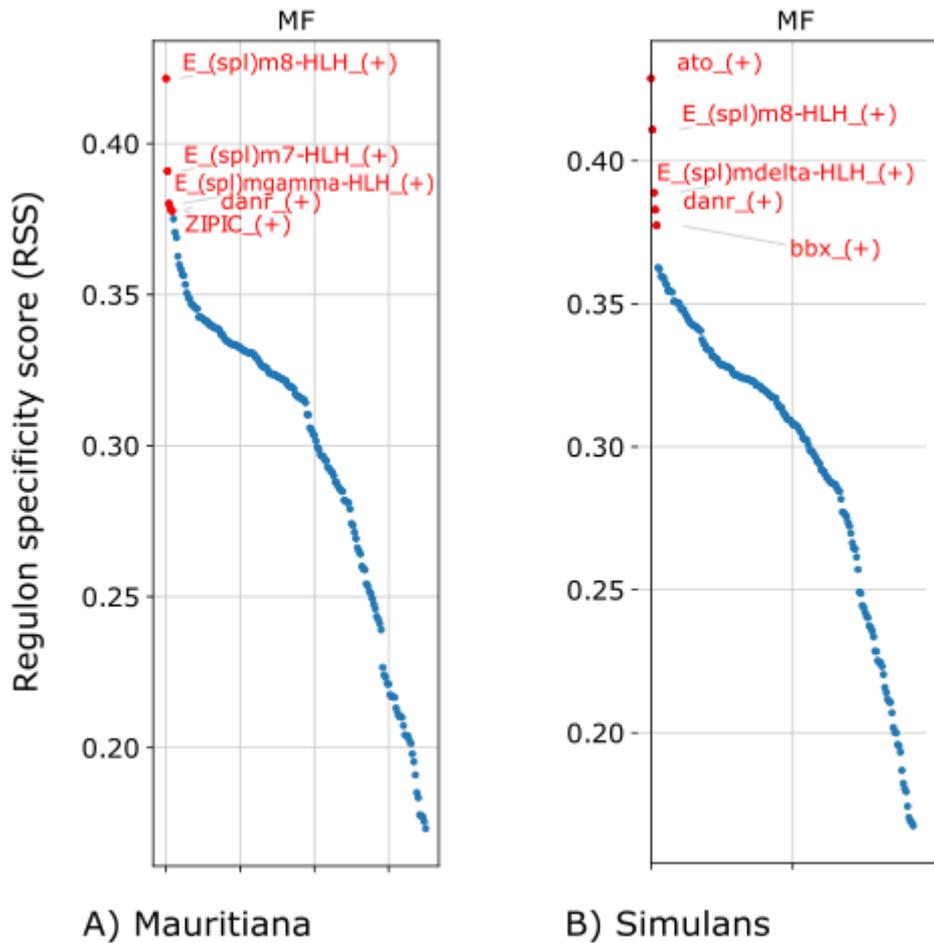
**Figure 31: Selected top-DEGs between morphogenetic furrow cells of *D. mauritiana* and *D. simulans*.** Depicted are the top 10 significant ( $p < 0.05$ ) DEGs which show an expression percentage difference of over 20% and a log2 fold change of over 1. Heat shock response genes were excluded from this list. *D. mauritiana* is depicted with a negative- and *D. simulans* with a positive average log2 fold change. Pct\_diff: Percentage difference in cells expressing the gene between the two species.  $-\log_{10}(P\text{-values})$  above the horizontal dashed line are infinite. The complete list is in Supplementary Table S 15.

The top species DEGs in the MF contain genes associated with a diverse set of processes including the semaphorin- (*Sema1a*) and ecdysone (*Hr4*) pathways. Consequently, I also only found few enriched GO-terms. These are in *D. mauritiana* at 108 h (“negative regulation of R7 cell differentiation”) and 120 h (e.g. “cellular component assembly”) (Supplementary Table S 15B and D).

**Table 11: Morphogenetic furrow specific DEGS between *D. simulans* and *D. mauritiana*.** Depicted are the top 10 genes by p-value, the total number of DEGs in each group is written in brackets.

| <b>Mau_84 h</b> | <b>Mau_108 h<br/>(26)</b> | <b>Mau_120 h</b> | <b>Sim_84 h</b> | <b>Sim_108 h</b> | <b>Sim_120 h</b> |
|-----------------|---------------------------|------------------|-----------------|------------------|------------------|
|                 | His2A:CG33865             |                  | Ubi-p63E        | Oatp74D          | SP1029           |
|                 | nrm                       |                  |                 | Imp              | NUCB1            |
|                 | frm                       |                  |                 |                  | tral             |
|                 | CG15628                   |                  |                 |                  |                  |
|                 | hwt                       |                  |                 |                  |                  |
|                 | luna                      |                  |                 |                  |                  |
|                 | CG31176                   |                  |                 |                  |                  |
|                 | upSET                     |                  |                 |                  |                  |
|                 | CG42340                   |                  |                 |                  |                  |
|                 | CG32645                   |                  |                 |                  |                  |

There are only few (32) interspecies DEGs uniquely differentially expressed in the MF (Table 11). In *D. mauritiana*, 108 h is the only timepoint for which I find upregulated DEGs in the MF. It includes several sparsely described genes like CG15628, CG31176, CG42340 and CG32645. The top hit here is a histone gene, *His2A:CG33865*. The first more specific hit is *neuromusculin (nrm)*, which is presumed to be expressed in photoreceptor neurons (Kania et al. 1993). In *D. simulans* at 84 h I find *Ubi-p63E* upregulated which is involved in protein modifications in a variety of processes. At 108 h, I find *Oatp74D* and *Imp*. *Oatp74D* is a part of the ecdysone signaling pathway which plays an important role in larval development. It is, however, expressed in the adult head as well (Aradska et al. 2015). *Imp* is a gene which, among other processes, is involved in stem cell maturation and neuronal remodeling and was found to be expressed broadly in the adult *D. melanogaster* (Li et al. 2022). At 120 h, I find three genes upregulated in *D. simulans*. *SP1029* is expressed in several tissues in the embryo including the head epidermis and maxillary sensory complex (Tomancak et al. 2002; Tomancak et al. 2007; Hammonds et al. 2013) and has been shown to be expressed in the adult head as well (Aradska et al. 2015). For *NUCB1*, there is now development specific function described, it is however broadly expressed in the embryo and in the adult head (Aradska et al. 2015; Otte et al. 1999). *Trailer hitch (Tral)* so far is only known for being involved in dorsoventral patterning (Wilhelm et al. 2005).



**Figure 32: Regulons in 120 h AEL morphogenetic furrow cells in (A) *D. mauritiana* and (B) *D. simulans*.** Regulons are ordered by descending regulon specificity score. The top 5 regulons are highlighted in red. Extensive lists of regulons can be found in Supplementary Table S 18.

To understand how genes are regulated in a specific cluster of cells, SCENIC was performed which identifies regulatory units called regulons within cell clusters. This analysis is species specific. Figure 32 shows an example output in MF cells of *D. mauritiana* (A) and *D. simulans* (B) at 120 h. In both species, I find the earlier described MF marker genes of the group of the enhancers of split to be among the highly specific regulons in this cell type.

**Table 12: Top 10 regulons in morphogenetic furrow cells which are shared between *D. simulans* and *D. mauritiana* and Regulons which are species specific by descending regulon specificity score.** Regulons highlighted in green appear in multiple cell populations. A (+) Indicates that the genes characteristic for this cell type are positively correlated with the listed TF. Red: genes which are found among the top 30 DEGs in the respective species in this cell type as well.

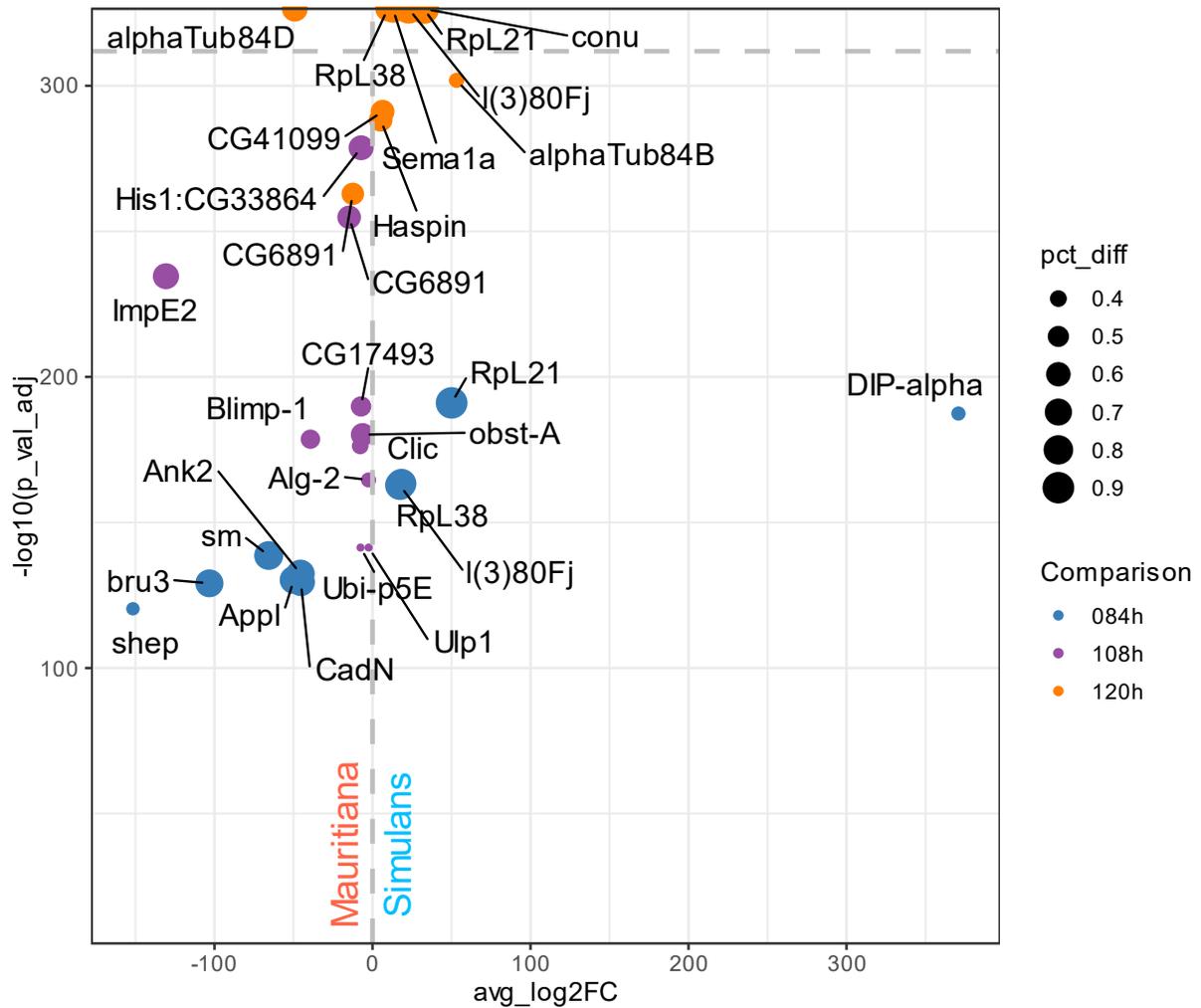
| Age                      | Shared_SimMau            | Sim                          | Mau                          |                          |
|--------------------------|--------------------------|------------------------------|------------------------------|--------------------------|
| 120 h                    | <i>E_(spl)m8-HLH_(+)</i> | <i>ato_(+)</i>               | <i>E_(spl)mgamma-HLH_(+)</i> |                          |
|                          | <i>danr_(+)</i>          | <i>E_(spl)mdelta-HLH_(+)</i> | <i>ZIPIC_(+)</i>             |                          |
|                          | <i>SREBP_(+)</i>         | <i>bbx_(+)</i>               | <i>vri_(+)</i>               |                          |
|                          | <i>ac_(+)</i>            | <i>ftz-f1_(+)</i>            | <i>gcm_(+)</i>               |                          |
|                          | <i>Trf2_(+)</i>          | <i>ey_(+)</i>                | <i>br_(+)</i>                |                          |
|                          | <i>BEAF-32_(+)</i>       | <i>E_(spl)m5-HLH_(+)</i>     | <i>CG10431_(+)</i>           |                          |
|                          | <i>maf-S_(+)</i>         | <i>Zif_(+)</i>               | <i>TfIIIB_(+)</i>            |                          |
|                          | <i>rn_(+)</i>            | <i>Sry-delta_(+)</i>         | <i>CG10348_(+)</i>           |                          |
|                          | <i>Dref_(+)</i>          | <i>Cnot4_(+)</i>             | <i>bsh_(+)</i>               |                          |
|                          | <i>RplI215_(+)</i>       | <i>Hr3_(+)</i>               | <i>ktub_(+)</i>              |                          |
|                          | 108 h                    | <i>en_(+)</i>                | <i>vnd_(+)</i>               | <i>gsb-n_(+)</i>         |
|                          |                          | <i>run_(+)</i>               | <i>B-H1_(+)</i>              | <i>esg_(+)</i>           |
|                          |                          | <i>E_(spl)mgamma-HLH_(+)</i> | <i>E_(spl)mdelta-HLH_(+)</i> | <i>Sox14_(+)</i>         |
|                          |                          | <i>gl_(+)</i>                | <i>Usf_(+)</i>               | <i>Mondo_(+)</i>         |
| <i>E_(spl)m7-HLH_(+)</i> |                          | <i>B-H2_(+)</i>              | <i>Lim1_(+)</i>              |                          |
| <i>E_(spl)m8-HLH_(+)</i> |                          | <i>Atf3_(+)</i>              | <i>Dif_(+)</i>               |                          |
| <i>pnt_(+)</i>           |                          | <i>orb_(+)</i>               | <i>Kdm4B_(+)</i>             |                          |
| <i>tgo_(+)</i>           |                          | <i>E_(spl)m5-HLH_(+)</i>     | <i>Optix_(+)</i>             |                          |
| <i>Spps_(+)</i>          |                          | <i>br_(+)</i>                | <i>Mad_(+)</i>               |                          |
| <i>E_(spl)m3-HLH_(+)</i> |                          | <i>Doc3_(+)</i>              | <i>ci_(+)</i>                |                          |
| 84 h                     |                          | <i>pnt_(+)</i>               | <i>E_(spl)mgamma-HLH_(+)</i> | <i>drm_(+)</i>           |
|                          |                          | <i>so_(+)</i>                | <i>ato_(+)</i>               | <i>sens_(+)</i>          |
|                          |                          | <i>aop_(+)</i>               | <i>gl_(+)</i>                | <i>Ets65A_(+)</i>        |
|                          |                          | <i>Dll_(+)</i>               | <i>sima_(+)</i>              | <i>E_(spl)m3-HLH_(+)</i> |
|                          | <i>al_(+)</i>            | <i>Rbp6_(+)</i>              | <i>h_(+)</i>                 |                          |
|                          | <i>dl_(+)</i>            | <i>toe_(+)</i>               | <i>E_(spl)m7-HLH_(+)</i>     |                          |
|                          | <i>bon_(+)</i>           | <i>Brf_(+)</i>               | <i>Mitf_(+)</i>              |                          |
|                          | <i>foxo_(+)</i>          | <i>CG5846_(+)</i>            | <i>knrl_(+)</i>              |                          |
|                          | <i>ey_(+)</i>            | <i>Pur-alpha_(+)</i>         | <i>sob_(+)</i>               |                          |
|                          | <i>lz_(+)</i>            | <i>mor_(+)</i>               | <i>Chd1_(+)</i>              |                          |

Between *D. mauritiana* and *D. simulans*, the top regulons I find as specific contain enhancers of split such as *E(spl)m8-HLH* and *E(spl)mgamma-HLH*. No genes of this family are identified as top 10 shared regulons at 84 h, but interestingly a few of them are exclusive to either species at different timepoints. At 108 h, for instance, enhancers of split are only identified as top regulons in *D. simulans*, but none are exclusively highly specific in *D. mauritiana*. Another

shared gene known to be expressed in the MF and involved in photoreceptor development is pointed (*pnt*) (Brunner et al. 1994; Rawlins et al. 2003). I find *ey*, a core member of the retinal determination network (see 2.2.1), as a shared regulon at 84 h and specific to *D. simulans* at 120 h. *Dorsocross3* (*Doc3*) and *bobby sox* (*bbx*) are genes with no reported role in eye development that I find as top regulons in *D. simulans* at 108 h and 120 h respectively. *Bbx* specifically was not found as a significant gene in any other comparison in this data yet. I found none of the top 30 regulons as differentially expressed genes in the MF between *D. mauritiana* and *D. simulans*.

#### 3.4.2.3 Pre-proneural cell characterization between *D. simulans* and *D. mauritiana*

The share of PPN cells declines in both species as the EAD is growing and the pool of progenitor cells is shrinking over the course of development (Figure 27). *D. simulans* has a larger share of these cells at early stages (Figure 28), although a bias in pairwise data is only visible at 108 h (Figure 28B). For *D. mauritiana*, I find 531 upregulated DEGs in this cluster over all timepoints. For *D. simulans*, I find 191 upregulated DEGs in PPN cells over all timepoints (Figure 33, extensive list in Supplementary Table S 16).



**Figure 33: Selected top-DEGs between pre-proneural cells of *D. mauritiana* and *D. simulans*.** Depicted are the top 10 significant ( $p < 0.05$ ) DEGs which show an expression percentage difference of over 20% and a log2 fold change of over 1. Heat shock response genes were excluded from this list. *D. mauritiana* is depicted with a negative- and *D. simulans* with a positive average log2 fold change. Pct\_diff: Percentage difference in cells expressing the gene between the two species.  $-\log_{10}(\text{P-values})$  above the horizontal dashed line are infinite. The complete list is in Supplementary Table S 16.

I see similar globally differentially expressed gene between *D. simulans* and *D. mauritiana* in PPN as in other cell types (e.g. *Sema1a*, *conu*, *alphaTub84D*, Figure 33, see also 3.4.1.2). Relevant and more specific enriched GO-terms for differentially expressed genes in this cell type include “axon midline choice recognition” in *D. simulans* at 120 h.

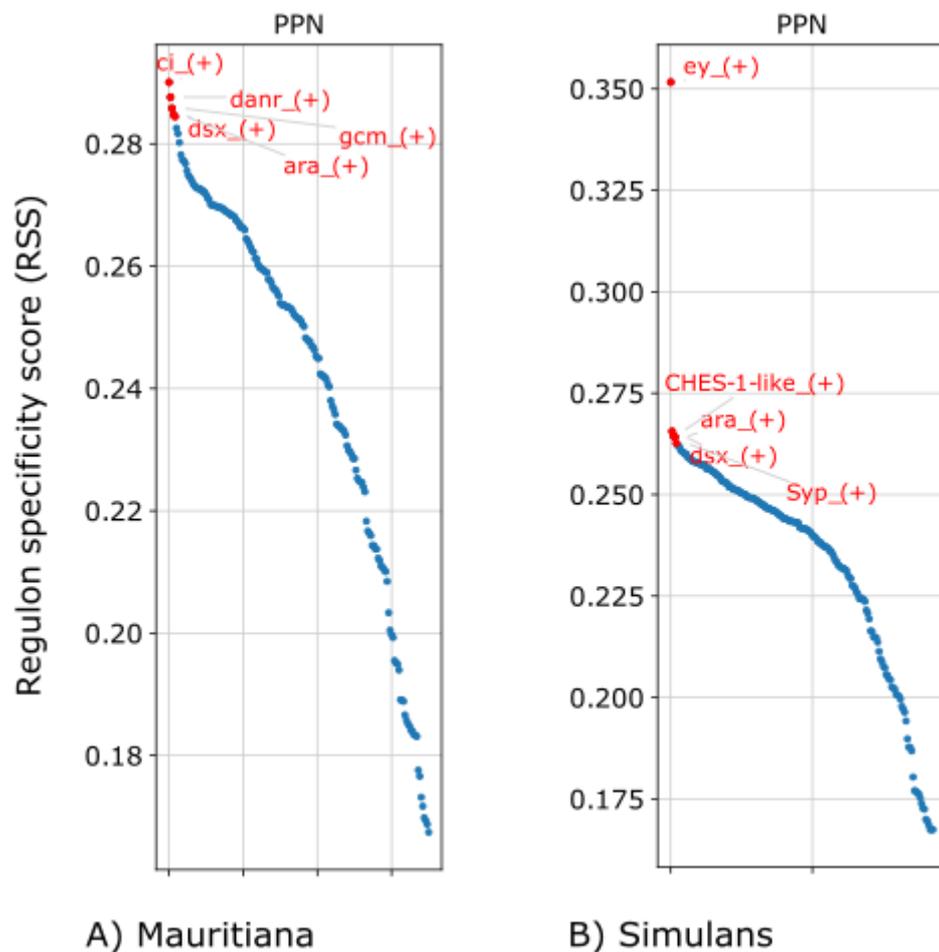
**Table 13: Pre-proneural cell specific DEGS between *D. simulans* and *D. mauritiana*.** Depicted are the top 10 genes by p-value, the total number of DEGs in each group is written in brackets.

| Mau_84 h<br>(38) | Mau_108 h<br>(58) | Mau_120 h | Sim_84 h<br>(33) | Sim_108 h | Sim_120 h |
|------------------|-------------------|-----------|------------------|-----------|-----------|
| Tet              | CG4374            | rn        | CG42673          | CG6739    | CG15414   |
| Ca-alpha1D       | Mur2B             | upSET     | ASPP             | Mob2      | sm        |
| Nrt              | ImpE3             | Mp        | Stlk             | NijA      | drl       |
| stan             | stw               | Oatp74D   | Set1             | CG8547    | Hs6st     |
| lncRNA:noe       | ImpE1             | zormin    | Mob2             | Best2     | Kr-h1     |
| dati             | neo               |           | Wdr62            | Arc1      | CG31998   |
| vvl              | GstD3             |           | nuf              |           | hid       |
| Fim              | Cht10             |           | Impl2            |           | stv       |
| gogo             | CG43980           |           | CG33298          |           | CG1909    |
| p130CAS          | CG31191           |           | RpS25            |           |           |

Genes which appear to be upregulated in different cell types in *D. mauritiana* compared to *D. simulans* include the *ecdysone-inducible genes*, *ImpE1* and *ImpE2* (Table 13). These genes are hypothesized to be involved in EAD morphogenesis (Andres et al. 1993). Transcripts of *ImpE2* are detected at high levels in prepupal stages, and at low levels in larvae of *D. melanogaster*. I find them upregulated in *D. mauritiana* as well at 108 h in cells of the MF (see 3.4.2.2). Another top DEG in *D. mauritiana* is *rotund (Rn)* which has been associated with compound eye development through a mutant phenotype, which includes a change of eye size (St Pierre et al. 2002; Del Alamo and Mlodzik 2008). I again find the gene *Organic anion transporting polypeptide 74D (Oatp74D)*, a member of ecdysone pathway upregulated here in *D. mauritiana* (see 3.4.2.2). *Zormin* is a gene uniquely upregulated in this comparison in *D. mauritiana*. It is described to be expressed in the adult head (Aradska et al. 2015), however, there are no reported functions connected to the EAD. *Ten-Eleven Translocation (TET) family protein (Tet)* is upregulated in *D. mauritiana* at 84 h and found as a regulon as well (Supplementary Table S 16). It is reportedly expressed in the eye disc, but its function there is unknown (Grubbs et al. 2013). It is, however, involved in neuronal morphology and development in the brain (Yao et al. 2018). Of special interest is the gene *neyo (neo)* due to its reported function in regulating embryonic cell shape (Fernandes et al. 2010). Several other specific genes such as *Cht1* and *Mp* have no reported function related to eye and head development yet.

In *D. simulans*, the most significant upregulated gene I find at 120 h with a reported role to eye or head development is the gene *derailed (drl)*. This gene is reported to be involved in the axonal growth in photoreceptor cells (Grillenzoni et al. 2007). *Head involution defective (Hid)* is involved in apoptosis and reportedly deregulated in *rbf* mutants in the eye (Tanaka-Matakatsu

et al. 2009). Among the upregulated genes in this cell types are several genes with reported roles in eye development which were not found as significant markers previously. These include *Krüppel homolog 1 (Kr-h1)* which is known to be involved in photoreceptor maturation (Fichelson et al. 2012) as well as the gene *Mob2*, which is upregulated at both 84 h and 108 h in *D. simulans* compared to *D. mauritiana* and involved in rhabdomere development (Liu et al. 2009).



**Figure 34: Regulons in 120 h AEL pre-proneural cells in (A) *D. mauritiana* and (B) *D. simulans*.** Regulons are ordered by descending regulon specificity score. The top 5 regulons are highlighted in red. Extensive lists of regulons can be found in Supplementary Table S 18.

To understand how genes are regulated in a specific cluster of cells, SCENIC was performed which identifies regulatory units called regulons within cell clusters. This analysis is species specific. Figure 34 shows an example output in *D. mauritiana* (A) and *D. simulans* (B) at 120 h. In both species here, I find *doublesex (dsx)* as a common regulon. This gene is best known for its role in sexual differentiation, but has no function reported in eye development. In third

instar larvae, it is known to be expressed in the central nervous system (Lee et al. 2002). Also, I find the gene *araucan* (*ara*) in both species again as was the case in *D. melanogaster* as well (see 3.4.1.3).

**Table 14: Top 10 regulons in pre-proneural cells which are shared between *D. simulans* and *D. mauritiana* and Regulons which are species specific by descending regulon specificity score.** Regulons highlighted in green appear in multiple cell populations. A (+) Indicates that the genes characteristic for this cell type are positively correlated with the listed TF. Red: genes which are found among the top 30 DEGs in the respective species in this cell type as well.

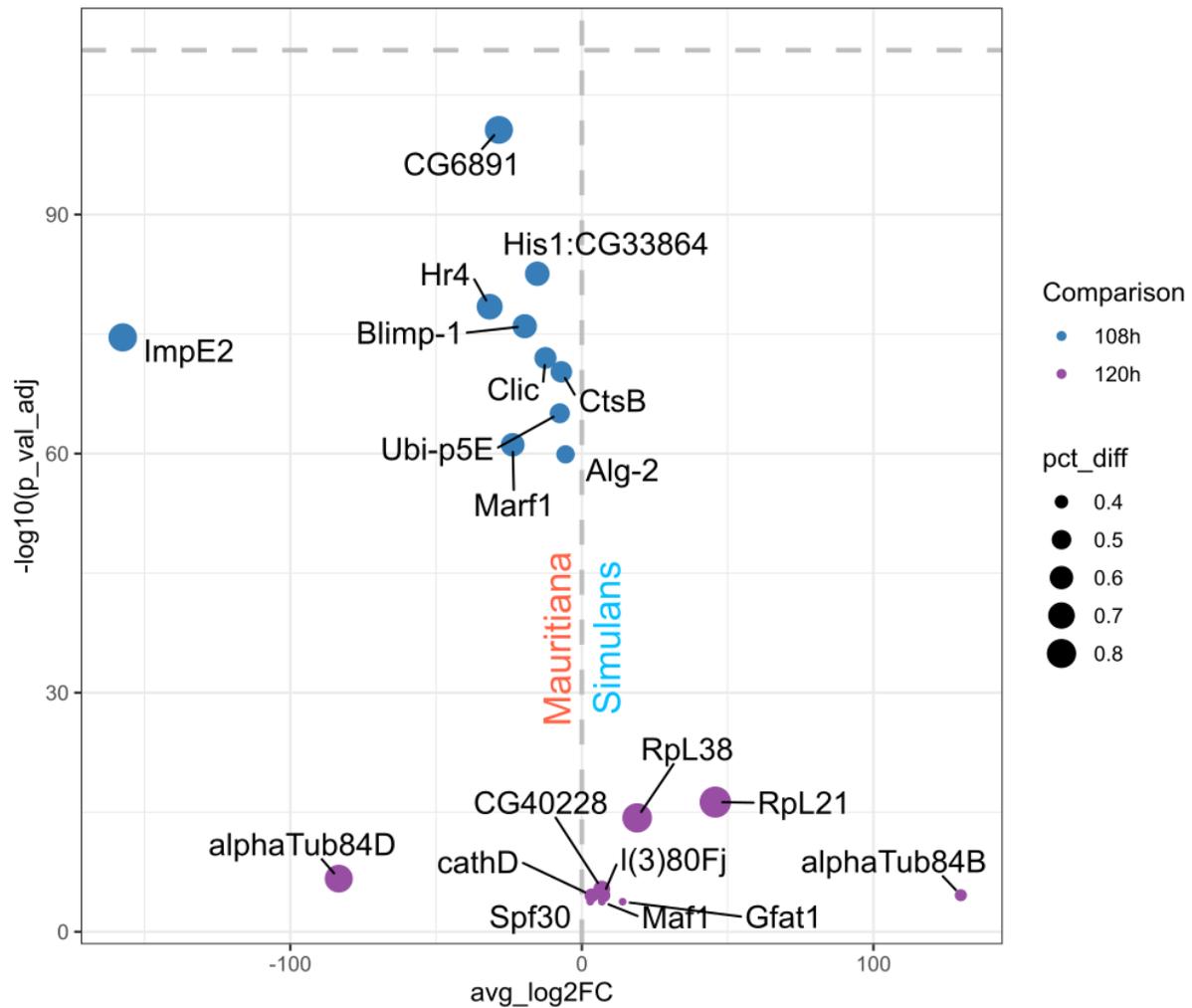
| Age        | Shared_SimMau   | Sim               | Mau                   |           |
|------------|-----------------|-------------------|-----------------------|-----------|
| 120 h      | CHES-1-like_(+) | ey_(+)            | gcm_(+)               |           |
|            | dsx_(+)         | Syp_(+)           | bowl_(+)              |           |
|            | ara_(+)         | Scr_(+)           | caup_(+)              |           |
|            | danr_(+)        | Hnf4_(+)          | so_(+)                |           |
|            | crol_(+)        | sd_(+)            | vri_(+)               |           |
|            | mirr_(+)        | cwo_(+)           | knrl_(+)              |           |
|            | Trf2_(+)        | sima_(+)          | Sin3A_(+)             |           |
|            | jumu_(+)        | Ets98B_(+)        | br_(+)                |           |
|            | EcR_(+)         | Trl_(+)           | CG10431_(+)           |           |
|            | ci_(+)          | ken_(+)           | CG10348_(+)           |           |
|            | 108 h           | dsx_(+)           | ara_(+)               | ey_(+)    |
|            |                 | CHES-1-like_(+)   | caup_(+)              | Scr_(+)   |
|            |                 | crol_(+)          | eyg_(+)               | Pdp1_(+)  |
|            |                 | jim_(+)           | Max_(+)               | Optix_(+) |
| E2f1_(+)   |                 | trx_(+)           | CG10321_(+)           |           |
| Dp_(+)     |                 | Myc_(+)           | ci_(+)                |           |
| EcR_(+)    |                 | schlank_(+)       | sd_(+)                |           |
| nej_(+)    |                 | Atf6_(+)          | Hr96_(+)              |           |
| Mef2_(+)   |                 | Cnx99A_(+)        | Hr3_(+)               |           |
| E_(bx)_(+) |                 | CG43347_(+)       | dar1_(+)              |           |
| 84 h       | so_(+)          | ovo_(+)           | h_(+)                 |           |
|            | ey_(+)          | Optix_(+)         | Mitf_(+)              |           |
|            | dl_(+)          | fru_(+)           | Rbf2_(+)              |           |
|            | grh_(+)         | trx_(+)           | Chd1_(+)              |           |
|            | aop_(+)         | Hr39_(+)          | Ets65A_(+)            |           |
|            | al_(+)          | Hcf_(+)           | shn_(+)               |           |
|            | sd_(+)          | toe_(+)           | EcR_(+)               |           |
|            | bon_(+)         | pan_(+)           | E_(spl)m3-<br>HLH_(+) |           |
|            | Spps_(+)        | Hnf4_(+)          | sob_(+)               |           |
|            | foxo_(+)        | Pur-<br>alpha_(+) | drm_(+)               |           |

In shared and species-specific genes, I find several genes which are members of Iroquois complex, such as *ara* and *caup*, which are shared at 120 h, only in *simulans* at 108 h and *mirr* shared at 120 h. These genes are usually expected in the dorsal region of the EAD for which I used them as markers in this work. As was the case between *D. melanogaster* and *D. mauritiana*, I find *CHES-1-like* again as a major regulator for gene expression in PPN cells (Table 14).

For *D. simulans*, I find *eyegone* (*eyg*) and *Atf6* upregulated at 108 h (Table 14, red). *Eyg* is a very impactful gene in eye development as a member of the retinal determination network (Figure 3). I already found *Atf6* as an upregulated, differentially expressed regulon in all examined cell types in *D. melanogaster* (3.4.1, description in 3.4.1.2). Here, it is upregulated and differentially expressed only in the PPN at 108 h. I find two differentially expressed genes in *D. mauritiana* which are also regulons, *Tet* (see Table 13 and description below the table), and *Abl tyrosine kinase* (*Abl*), a gene required for photoreceptor morphogenesis (Xiong and Rebay 2011). I also found this gene as upregulated in *D. melanogaster* PhR and MF at 120 h, however, it is not found in the top 30 regulons there. Further species-specific upregulated regulons include *Optix* and *ovo* in *D. simulans* at 84 h (Table 14, red). *Optix* is another member of the Retinal determination network (see Figure 3). *Ovo* is involved in several processes in the germline (Mével-Ninio et al. 1995). Among the top 30, I additionally find *nervous fingers 1* (*nerfin-1*) which a zinc finger transcription factor that regulates early axon guidance (Kuzin et al. 2005) and *Ecdysone-induced protein 78C* (*Eip78C*), another member of the ecdysone pathway.

#### 3.4.2.4 Ocelli

In ocelli I found a bias in cell population composition between *D. simulans* and *D. mauritiana* at 120 h (Figure 28C). Ocelli clusters were identified up from 108 h AEL, so this analysis includes the timepoints 108 h and 120 h AEL. In total I found 294 upregulated DEGs in this cluster in *D. mauritiana* in total and 50 upregulated DEGs in *D. simulans* (Figure 35, extensive list in Supplementary Table S 17). The top 10 genes at 108 h are exclusively upregulated in *D. mauritiana*. I also find no significantly enriched GO-terms in *D. simulans* (Supplementary Table S 17). Top GO-terms in *D. mauritiana* are largely related to development. The only significantly enriched GO terms are in *D. mauritiana* at 108 h and related to pupal development and “regulation of biological quality” at 120 h (Supplementary Table S 17).



**Figure 35: Selected top-DEGs between ocelli cells of *D. mauritiana* and *D. simulans*.** Depicted are the top 10 significant ( $p < 0.05$ ) DEGs which show an expression percentage difference of over 20% and a  $\log_2$  fold change of over 1. Heat shock response genes were excluded from this list. *D. mauritiana* is depicted with a negative- and *D. simulans* with a positive average  $\log_2$  fold change.  $pct\_diff$ : Percentage difference in cells expressing the gene between the two species.  $-\log_{10}(P\text{-values})$  above the horizontal dashed line are infinite. The complete list is in Supplementary Table S 17.

Again, I find a member of the *ImpE* genes upregulated in *D. mauritiana* at 108 h in ocelli, as well as *Hr4*, which, as described earlier are involved in the ecdysone pathway. I find similar DEGs as in previous comparisons including several ribosomal genes (*RpL21*, *RpL38*).

**Table 15 Top 10 regulons in ocelli cells which are shared between *D. simulans* and *D. mauritiana* and Regulons which are species specific by descending regulon specificity score.** Regulons highlighted in green appear in multiple cell populations. A (+) Indicates that the genes characteristic for this cell type are positively correlated with the listed TF. Red: genes which are found among the top 30 DEGs in the respective species in this cell type as well.

| Age   | Shared_SimMau        | Sim                   | Mau         |
|-------|----------------------|-----------------------|-------------|
| 120 h | C15_(+)              | Doc3_(+)              | NK7.1_(+)   |
|       | cato_(+)             | E_(spl)m3-<br>HLH_(+) | caup_(+)    |
|       | gsb_(+)              | ey_(+)                | bowl_(+)    |
|       | slp2_(+)             | sc_(+)                | gcm_(+)     |
|       | eg_(+)               | Dad_(+)               | CG10348_(+) |
|       | drm_(+)              | Ets98B_(+)            | Sin3A_(+)   |
|       | Rel_(+)              | al_(+)                | CG10431_(+) |
|       | Doc2_(+)             | brk_(+)               | br_(+)      |
|       | Pdp1_(+)             | cwo_(+)               | klu_(+)     |
|       | gsb-n_(+)            | Trl_(+)               | gcm2_(+)    |
| 108 h | Awh_(+)              | eg_(+)                | tup_(+)     |
|       | Sox102F_(+)          | sob_(+)               | D_(+)       |
|       | en_(+)               | drm_(+)               | pdm3_(+)    |
|       | pnr_(+)              | eyg_(+)               | dar1_(+)    |
|       | slp2_(+)             | pb_(+)                | odd_(+)     |
|       | al_(+)               | Myc_(+)               | Eip74EF_(+) |
|       | E_(spl)mbeta-HLH_(+) | Cnx99A_(+)            | Pdp1_(+)    |
|       | kni_(+)              | Got1_(+)              | inv_(+)     |
|       | gsb_(+)              | B-H2_(+)              | Mitf_(+)    |
|       | Brf_(+)              | caup_(+)              | disco-r_(+) |

Among the top regulon are genes expected to be expressed in the dorsal region, such as *pannier* (*pnr*), which is shared between both species at 108 h. The top shared regulon at 120 h is *C15*, which is associated with dorsal tissue in the embryo (Hammonds et al. 2013; Stathopoulos et al. 2002; Tomancak et al. 2002; Tomancak et al. 2007). At 120 h, one of the top genes uniquely upregulated in the ommatidial cluster and only in *D. simulans* is *Daughters against dpp* (*Dad*). This genes is known to be involved in imaginal disc development (Stultz et al. 2006) and developmental patterning (McClure and Schubiger 2005). The top hit in *D. simulans* at 108 h is *eagle* (*eg*), which is a gene found in the top 10 regulons in both species at 120 h and, among other processes, is involved in axon pathfinding (Dittrich et al. 1997). 120 h, *Dorsocross3* (*Doc3*) is expressed in the dorsal embryo and visual anlagen (Hammonds et al. 2013; Tomancak et al. 2002; Tomancak et al. 2007). The top hit in *D. mauritiana* at 120 h, *NK7.1* does not have a known function in eye development, but suspected to be expressed in the eye disc as deduced from reporter fly lines (Bloomington Drosophila Stock Center 2019). *Brother of odd with*

*entrails limited* (*Bowl*) is a known patterning gene which is expressed, for instance, in the antennal disc (Brás-Pereira and Casares 2008) but with no known function in the eye. This gene is part of several which are uniquely upregulated in this cell type in a species-specific manner with potentially undescribed functions in the regulation of ommatidial development.

## 4 Discussion

I present single-nuclei RNA-sequencing datasets for three different species, including the non-model species *D. simulans* and *D. mauritiana* of up to five different timepoints. I show, that these data contain timepoint specific, ordered developmental information. I can meaningfully quantify relative differences in cell type population sizes between species. I find known key-regulators of eye development characterizing gene expression of specific cell types, as well as novel candidates putatively contributing to the formation of morphological variation in the *Drosophila* head.

### 4.1 Single-nuclei sequencing from low amounts of input tissue

Assessing genome wide gene expression for individual cells has proven powerful to describe the heterogeneity of complex tissues, identify novel cell types and to study biological processes, such as immunity and cell-cell interactions at unprecedented detail. Despite the technological advances, single-cell RNA sequencing (scRNAseq) methods still require many cells as starting material. Therefore, I evaluated different dissociation protocols and compared scRNAseq to single-nuclei RNA sequencing (snRNAseq) with special emphasis on low-input material. Based on data obtained for eye-antennal imaginal discs of *Drosophila melanogaster*, I found snRNAseq superior to scRNAseq for the following reasons: 1) The isolation of nuclei requires fewer experimental steps compared to tissue dissociation into live cells, increasing reproducibility across experiments. This feature is especially relevant if gene expression comparisons are needed on the level of individual cells, for example to assess the effect of experimental manipulations, to study different developmental stages or to compare species/populations. 2) In our evaluation experiment, I observed significantly reduced stress-related expression responses and reduced ribosomal gene expression in snRNAseq data, suggesting that more informative reads contribute to biological insights. 3) We showed highly efficient nuclei isolation and high-quality RNA extraction from frozen tissue (Krishnaswami et al. 2016). It is a major advantage to have the opportunity to collect tissue over time and process samples simultaneously, especially for low-input material. 4) In line with previous reports (Wu et al. 2019; Lake et al. 2017; Grindberg et al. 2013; Lake et al. 2018; Bakken et al. 2018; Habib et al. 2016; Ding et al. 2020; Slyper et al. 2020), our snRNAseq dataset contained sufficient expression information to unravel major cell types expected in eye-antennal imaginal discs. 5) While scRNAseq has been shown to result in biased cell composition, due to different cell sizes,

shapes, and survival rate upon dissociation (Darmanis et al. 2015), the more streamlined nuclei isolation procedure should ensure a more representative assessment for snRNAseq, especially for complex organs, such as nervous tissue (Yim et al. 2022). For instance, I found indications that snRNAseq may be more efficient in capturing rather complicated cell types, such as the large polyploid cells of the peripodial epithelium (“Ventral PE” cluster 8; Figure 8B).

It is important to consider major differences in the analysis and interpretation of scRNAseq and snRNAseq data. For instance, snRNAseq data contains intronic reads originating from immature nuclear RNA (Grindberg et al. 2013). Accordingly, well-annotated genome resources are advantageous and analyses pipelines need to be adjusted to also include reads mapped to introns in subsequent read quantification. snRNAseq data captures only rather transient nuclear RNA, while scRNAseq also includes cytoplasmic mature mRNA. Hence, gene regulation events acting on the level of nuclear export (Palazzo and Lee 2018; Wickramasinghe and Laskey 2015), splicing (Gehring and Roignant 2021) or mRNA maturation (Galloway and Cowling 2019; Mittleman et al. 2020) may contribute to differences in expression information derived from nuclei and cells, respectively. If cytoplasmic RNA molecules are of special interest and thus single-cells need to be isolated, I strongly suggest a dissociation protocol combining chemical and mechanical treatment of tissue samples in conjunction with FACS-aided live cell selection based on fluorescent live-dead cell staining. Our direct comparison of scRNAseq and snRNAseq data showed that the different datasets contributed differently to the obtained cell types. While some of these differences can be explained by slightly different larval staging procedures in our experiments, it remains to be established, which exact cellular or molecular features influence the more efficient recovery of certain cell types in sc/snRNAseq. In the light of our findings, it will be important to focus on either scRNAseq or snRNAseq if comparative questions are tackled.

In summary, based on a thorough evaluation of different dissociation and sequencing protocols I suggest a highly efficient snRNAseq procedure to obtain high-quality expression data for individual nuclei. Our procedure is specifically tested for low-input material and will therefore be perfectly suited for future studies with limited access to tissue samples.

#### 4.2 Annotation of cell clusters in three *Drosophila* species

The project I present comes with a number of peculiar challenges for a single-nuclei sequencing project. I identify cell type specific candidate genes for divergence in eye morphology between pairs of closely related *Drosophila* species. This requires marker gene lists for single-cell data

which at the time stem only from late L3 EAD of *D. melanogaster* (Ariss et al. 2018). Although I expect most genes to be expressed in similar patterns in closely related species, there was a specific challenge in standardizing cluster annotations when integrating data from different species. Most expected cell types could immediately be identified in late L3 EAD of *D. mauritiana* as well as *D. simulans* by applying an automated method of cluster annotation. By the time the analysis was performed, this was most often done manually. Automatic methods such as machine learning based annotation of cell clusters usually rely on a sufficient number of well-curated reference datasets, of which at the time only two were available, which also slightly differed in annotation (e.g. naming antennal rings). Furthermore, this data was only covering late-stage *D. melanogaster* EAD.

In most datasets, almost all clusters could be assigned to known cell types. A few clusters in several datasets remain where a clear identity could not be assigned. This often concerns the first one or two clusters. Applying less strict criteria or random-sampled resolution changes did generally not resolve this. On the one hand, it could speak for stochastic RNA content or loss of RNA and might indicate these cells should be filtered out. On the other hand, this might be caused by their underlying biology. Firstly, it could mean that these cells show a very average gene expression profile. This might be due to two different reasons. Firstly, it could be undifferentiated cells. Although Ariss et al. identified undifferentiated cells of the area in the eye proper by the characteristic expression of both *hth* and *toy*, which, in our data, are usually part of the “PPN” clusters, these cells could be general imaginal disc or EAD progenitor cells in early stages. Some of these clusters do however express DEGs, but these might be ribosomal genes. They are not markers for any specific clusters and filtered out in many cases. Ribosomal genes could be leftovers from improperly dissociated cells as the proportion of ribosomal genes is a lot higher in cells (see Figure 6). I chose to keep ribosomal genes in my analyses, as they might have relevant influences on developmental processes. This is shown, for instance, for *RpS12*, which influences several cellular processes including cellular competition through the regulation of *Xrp1* (Ji et al. 2019) or the knockout of *RpS6*, which causes tissue overgrowth in *D. melanogaster* larvae (Stewart and Denell 1993). In single-cell sequencing data, I also found clusters differentially expressing heat shock genes, which may indicate a stress response due to the dissociation procedure, however, they are usually largely absent in single-nuclei sequencing data.

In case of cells expressing markers for different clusters, these cells might be in the process of differentiation or a mixture between two different cell types that could not be separated

properly. In these cases, the selected resolution might not be high enough to resolve these clusters.

Generally, a quality cutoff is performed automatically by the Cellranger software during the preprocessing of snRNAseq data. This can cut off the number of sequenced cells drastically. Cutoff values can be changed to include more cells; however, the cells excluded by default are often of poor quality. In one case there is a larger number of poor-quality cells (*D. mauritiana* 108 h AEL). Since I aimed to gather samples from the same parent populations, the time for gathering EAD was also limited to the reproductive time of one generation of flies, otherwise the number of captured EAD could be increased further, also for younger timepoints by freezing additional EAD.

#### 4.1 Complex developmental information is preserved in single-nuclei sequencing data

Gathering sufficiently large amounts of tissue from small organs for snRNAseq is a time-consuming procedure. The sets of snRNAseq data that I generated span up to five stages of eye-antennal disc development, which can provide a source for identifying novel genes involved in EAD morphogenesis for a variety of processes. For the data to be useful, they need to preserve not only tissue- but also temporal information. Generally, I observed that the developmental order that is assumed from staging the larvae is preserved when the timepoints were analyzed together. The eye discs at each stage also contain cells at different developmental stages. For instance, PPN cells anterior to the MF represent undifferentiated proliferating cells, while cells behind the furrow start undergoing differentiation. Observing this disc- and stage-intrinsic structure is possible, for instance, just by examining the data from 120 h old discs. While the developmental signal was best preserved in *D. simulans*, it was less obvious in *D. melanogaster* data. It may be possible that additional developmental signal is hidden in higher-order principal components (PCs 40-50), which were excluded in my pipeline because I focused predominantly on species-comparisons. For future developmental studies, it could be advantageous to use a far larger number of principle components or skip dimensionality reduction entirely. Downstream nearest-neighbor and clustering steps should still be able to identify structures in the data, if they exist, provided sufficient computational power is available to perform analyses on very large data.

As we have developmental data for three species, it is expected that genes and central regulators found in all species would represent important factors for EAD development. Indeed, I found

common regulators preserved between two or three species. For instance, the gene *Checkpoint suppressor 1-like* (*CHES-1-like*) is a top regulon in all species in the PPN (Table 8 and Table 14), and in one case in PhR (Table 4). *CHES-1-like* is a transcription factor and has been shown to upregulate *dpp* expression, which is required for the initiation and progression of the MF. In cardiac progenitor cells it regulates cell division (Ahmad et al. 2012), and as such may be involved in tissue growth within the EAD.

Overall, my finding indicates that the genetic control of development between EADs of different stages is complex and my very preliminary analysis of developmental aspects strongly suggest that the datasets are well-suited to gain exciting novel insights into eye-antennal disc development. A difficulty when approaching this problem is the fact that cells in all intermediate steps of cellular maturation are present at most stages simultaneously. At any given timepoint, one would expect to find undifferentiated and proliferating cells, cells undergoing differentiation (within the MF/SMW and exiting it) and photoreceptor cells (with the exception of 72 h). I find surprisingly strong expression of photoreceptor marker genes in stages earlier than expected. Photoreceptor markers genes may have different functions in early development. Cells that develop independently from the eye primordium, such as antennal cells, show their very own gradients, that are partially visible in the data. Capturing these features allows to use subsets of the larger datasets to study cell type specific developmental processes on a cellular level while also allowing to view them in the broader context of the tissue.

#### 4.2 Core cell types of the eye primordium show the highest diversity in cell population sizes within the EAD

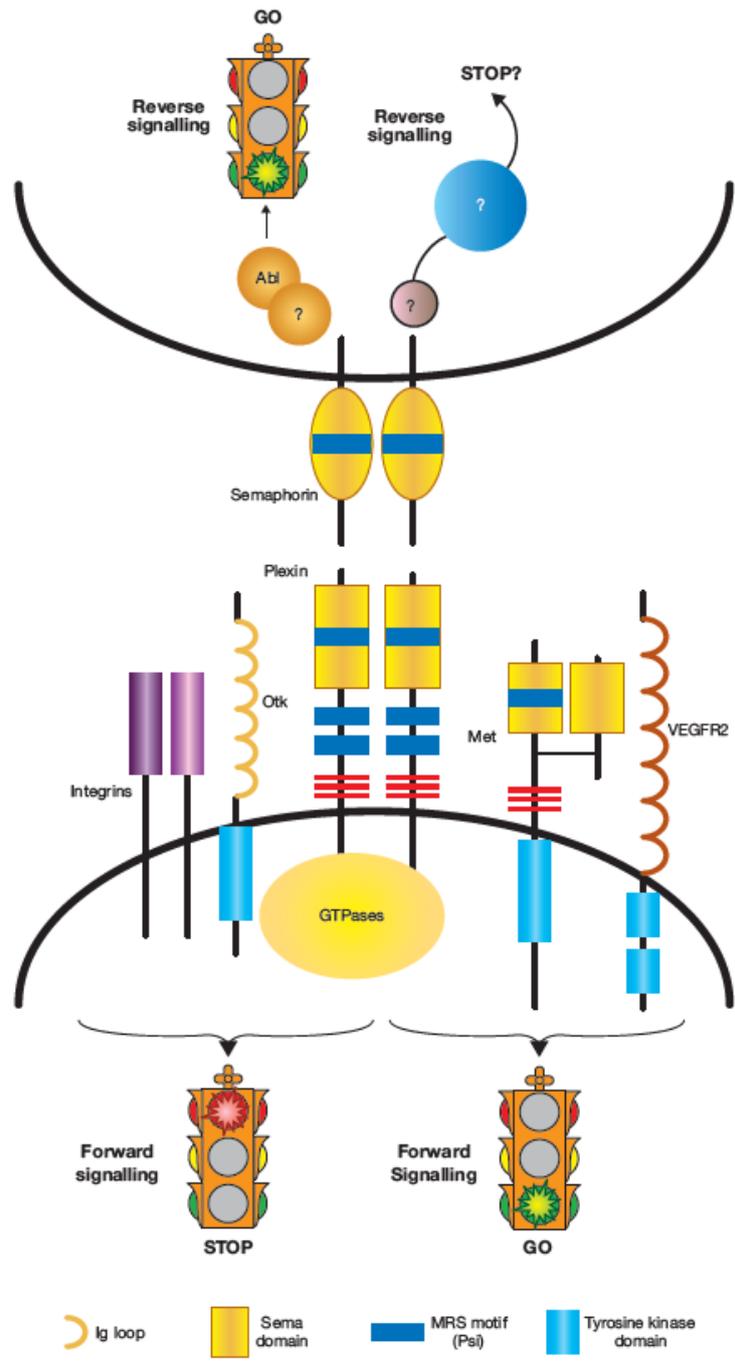
In pairwise comparisons between the three examined species, we observe differences in cell type composition in several cell populations. However, the annotation of cell types within EADs relies largely on late-stage data. For solid developmental studies, it would be helpful to make an effort to carefully describe cell types in early stages. There are several challenges in this attempt as the cells are expected to be less differentiated in general. In the growing number of available snRNAseq and scRNAseq datasets, our data contributes to a gaining better understanding also of early and difficult to access stages of EAD development. We show that we can reliably identify the most prominent cell types in our datasets in non-model species as well as in *D. melanogaster* and that our data is also comparable to preliminary

bulk-sequencing results. To achieve this, we are pooling data from around 30 EAD per sample in single samples per timepoint. To quantify differences in cell type composition more carefully, an increased number of samples would produce more accurate cell population estimations. This can be supported by applying methods such as single-cell differential composition analysis, which allow estimation of the robustness of cell number variations in comparable clusters between different datasets (Cao et al. 2019).

By plotting the proportions of cell populations in the total datasets (Figure 19 and Figure 27), I find core retinal cell types to be the most dynamic within the EAD. Photoreceptor cell clusters also present themselves as the most diverse group of cells in terms of gene expression patterns. We find surprisingly strong expression of photoreceptor marker genes in stages earlier than expected, but also later stages show several different photoreceptor clusters which may be worth examining further. Since photoreceptor types are known to be defined sequentially (Treisman 2013), my presumption was that the photoreceptor types might line up in a developmental trajectory in the UMAP plot. This was not the case. They rather arrange along a gradient of *Appl* and *cpo* expression, with cells expressing *brp* at the very tip of the converging clusters. What drives the apparent dissimilarity between these individual striped clusters remains to be examined and could potentially give insight into development and heterogeneity of different photoreceptor types.

#### 4.3 Development of retinal cell types in *D. mauritiana* is characterized by downregulated genes related to the semaphorin pathway

Noticeable interspecies differences in these cells lie here in genes related to the semaphorin-plexin pathway. Both *D. simulans* and *D. melanogaster* show an upregulation in genes involved in this pathway compared to *D. mauritiana*. Semaphorins are a family of proteins which are involved in morphological regulation within several tissues as well as guiding axonal growth and as such highly interesting for further investigation (Alto and Terman 2017).



**Figure 36: Components of the semaphorin-plexin signaling pathway.** OTK: Off track kinase, VEGFR2: vascular endothelial growth factor receptor type 2. This figure from (Comoglio et al. 2004) illustrates the pathways components involved in forward and reverse signaling.

In this pathway, signals are mediated by semaphorins and their ligands, the plexins (Jongbloets and Pasterkamp 2014). In contrast to many other signaling pathways, it is bidirectional and can deliver a “stop” or a “go” signal (Figure 36). This unique attribute allows its application in axon guidance. How the reversal of signaling in this pathway works is not exactly known. However, the involvement of *Abl kinase* likely plays a role in this process (Toyofuku et al. 2004) which I find this gene to be upregulated in PhR cells of *D. mauritiana* compared to *D. simulans* at 84 h

and also found as a unique regulon in PhR cells of *D. melanogaster* compared to *D. mauritiana* at 120 h (Table 4). The semaphorin/plexin pathway can influence tissue morphology mainly by regulating cytoskeletal reconfiguration which is described for several tissues. Furthermore, there is data suggesting its relevance for eye development specifically. In the human retina, it is known that point mutations in *Sema4A* can cause retinal degeneration (Nojima et al. 2013). A function this gene has in the human retina is the support of photoreceptor cell survival by endosomal sorting in the retinal pigment epithelium (Toyofuku et al. 2012). A similar role is known for Semaphorin4D which is involved in the diurnal clearance phagocytosis in those same cells, also maintaining photoreceptor survival (Bullock et al. 2018). In *D. melanogaster*, it is known that the interaction between *Sema1a* and *PlexA* coordinates axon guidance as well (Yu et al. 2010). *Sema1a* is perhaps the most relevant diverging gene of *D. mauritiana* in my comparisons. It is among the top 10 upregulated DEGs in *D. simulans* at 84 h in PhR, MF and PPN (Figure 29, Figure 31 and Figure 33). Furthermore, it is upregulated in *D. melanogaster* PhR as well at all time points (Figure 21) as well as in the PPN at 84 h and 108 h (Figure 25). Here, a semaphorin ligand, PlexB, is upregulated as well (Figure 25). These findings indicate a downregulation of the semaphorin-plexin pathway in *D. mauritiana* compared to both of its sister species. Axon guidance is the most prominent function of this pathway which does not necessarily influence the visible phenotype of an organism. However, it also influences the migration of myocardial cells and as such has a morphology shaping function there (Toyofuku et al. 2004), which makes this pathway an intriguing target for further morphological studies between these *Drosophila* species.

#### 4.4 Development of retinal cell types in *D. mauritiana* is characterized by upregulated genes related to the ecdysone pathway

I regularly find genes involved in the ecdysone pathway to be differentially expressed between species. Ecdysone is a hormone which is present in many animals. An important function during development of *Drosophila* larvae is the control of the molting cycles, and the initiation of pupariation. As a hormone, it has a global effect on organ growth (Herboso et al. 2015). Among the genes I find upregulated and related to the ecdysone pathway are *DopEcR*, upregulated in *D. mauritiana* at 84 h (Table 5), and in several comparisons, *Hormone receptor 4 (Hr4)*. Other genes from this pathway that I find here include *Hr38*, *Hr78*, *Eip74EF* and *Eip75B*. In proneuronal cells I find the gene *Oatp74D* diverging in different comparisons. It is among the top upregulated genes in *D. mauritiana* compared to *D. simulans* at 120 h but also in *D.*

*melanogaster* at 108 h compared to *D. mauritiana*. It is known to be expressed in the adult head and involved in the ecdysone pathway which is tightly connected to organismal development and control of pupariation. It has been shown that the ecdysone pathway is involved in controlling proliferation of eye progenitor cells via the synergy of *hth* and *tsh* (Neto et al. 2017).

#### 4.5 Other major genes driving interspecies differences between *D. melanogaster* and *D. mauritiana*

Most information that is available about functions of genes was gathered from *D. melanogaster* due to its long history as a model organism for genetic research. It is reported, for instance, that the expression of the developmental patterning gene *bowl* extends within the antennal disc to the border of the eye disc in *D. melanogaster* (Brás-Pereira and Casares 2008). In *D. mauritiana* I found it to be a species-specific top regulon in PPN and Ocelli compared to Both *D. melanogaster* and *D. simulans*. In the latter comparison, I also found it as a regulon in Ocelli. In *D. mauritiana* I observe significantly upregulated genes encoding histone subunits, when comparing multiple datasets. This finding is made for different timepoints. For instance, *His2A:CG33865* as upregulated in *D. mauritiana* at 84 h, 108 h and 120 h in different cell types compared to one of the other species. There can be various reasons for variation in histone configurations. In the simplest case an enrichment of histones on the DNA can indicate higher mitotic activity. However, it is known that some histone variations are cell type specific, and as such could play a role in development and cell type specification or maturation as well. Major known regulators of eye development such as *eyg* likely influence eye development on many levels in a variety of cell types in the EAD. One of the most important genes specific to *D. mauritiana* is *Rbp6*. Its function within the EAD is not known, however, its paralog *musashi* (*msi*) is inferred to be involved in photoreceptor cell fate commitment (Hirota et al. 1999). Both of these genes are also specific to *D. melanogaster* at 84 h, but *Rbp6* is found *D. mauritiana* specific at 108 h. Here, it is unique to *D. mauritiana* and found as a DEG as well (Table 3).

#### 4.6 Developmental differences between *D. mauritiana* and *D. simulans*

##### 4.6.1 Genes upregulated in *D. mauritiana*

The genes of highest importance for developmental processes are likely those which are at the same time upregulated in a specific cell type in a species as well as found as an important regulon, as these genes can influence the transcription of many other target genes. One such gene expressed in *D. mauritiana* is *PAR-domain protein 1* (*Pdp1*). It is upregulated at 84 h in

*D. mauritiana* photoreceptor cells compared to *D. simulans* (Table 10). Its main function is in maintaining the circadian clock in neuronal cells (Cyran et al. 2003). Interestingly, *Pdp1* null mutant larvae are reported to show significant growth delays. Its upregulation in *D. mauritiana* might thus have a growth promoting effect, potentially contributing to the enlarged eyes in *D. mauritiana*. There are several genes upregulated in *D. mauritiana* which are related to tissue or cell growth. one such gene is *nerfin 1*. The upregulation in *D. mauritiana* is surprising since *nerfin-1* expression is known to restrict the activity of the transcriptional coactivator *Yorkie* (*Yki*), negatively regulating tissue growth (Guo et al. 2019). As a member of the Hippo pathway, which controls organ size, this gene is a highly interesting gene for future research. This gene further motivates to compare gene expression between the last pair, *D. melanogaster* and *D. simulans* as it was found as highly specific to *D. mauritiana* in both interspecies comparisons performed here.

*Ten-Eleven Translocation (TET) family protein (Tet)* is upregulated in *D. mauritiana* at 84 h and found in the top 30 regulons at that timepoint in the PPN (Supplementary Table S 16). Despite its function in the EAD being unknown, it is, involved in neuronal morphology and development in the brain, and defects in this gene cause abnormal locomotion in *Drosophila* larvae. (Yao et al. 2018). In the zebrafish *Danio rerio*, *Tet* is involved in the morphogenesis of differentiated retinal neurons (Seritrakul and Gross 2017) which raises the possibility of a similar function in *D. mauritiana*.

#### 4.6.2 Genes upregulated in *D. simulans*

The most relevant upregulated gene in *D. simulans* might be *verthandi* (*vtd*). It is a top regulon in 120 h old photoreceptor cells (Table 10), and uniquely upregulated in photoreceptors and no other cell type at both 108 h and 120 h (Table 9). *Vtd* mutants show a variety of phenotypes, including defects in fertility, cuticle, bristles, wings, and importantly eyes (Schulze et al. 2001). It is a member of the cohesin complex, encoding specifically the cohesin Drad21 (Hallson et al. 2008). Its relation to eye development lies in it being a *hedgehog* (*hh*) regulator (Hallson et al. 2008). *Hh*, as explained earlier, is involved in triggering the progression of the morphogenetic furrow. *Vtd* is known to be expressed in the EAD (Warren et al. 2000), specifically, it shows high levels of expression in S-phase cells of proliferating imaginal tissues, and in the early endocycling cells of the embryonic gut. Another intriguing candidate gene is *Mob2*. as stated, it is involved in rhabdomere development (Liu et al. 2009). However, it is also

known to be involved in cell morphogenesis (He et al. 2005), potentially contributing to differences we see in ommatidia morphology between the two species.

In pre-proneural cells, further species-specific upregulated regulons in *D. simulans* include *Optix* and *ovo* in *D. simulans* at 84 h (Table 14). Again, I find *Optix*, a member of the retinal determination network (see Figure 3). *Ovo* is a transcriptional activator (Bohère et al. 2018) involved in several processes in the germline (Mével-Ninio et al. 1995). More importantly, it is known for a mutation, *shavenbaby* (Delon et al. 2003). The reversion of *Svb* and another mutation, *lzl*, can result in phenotypes that affect cuticle and eye development (Mével-Ninio et al. 1991). Furthermore, in planarians *ovo* is one of if not the only known planarian transcription factor with expression exclusively in the eye and eye progenitors and the only TF together with *six-1/2* and *eya*, that is expressed in all cells of the eye (Lapan and Reddien 2012). Most importantly, this gene can influence epidermal cell shape. With this function, it is possible that this gene contributes to the smaller ommatidia in the eyes of *D. simulans* by cytoskeletal remodeling (Delon et al. 2003; Payre et al. 1999).

*Ecdysone-induced protein 78C (Eip78C)*, another member of the ecdysone pathway which induces pupariation (Stone and Thummel 1993). In the ocelli, the top hit in *D. simulans* at 108 h is *eagle (eg)*, which is a gene found in the top 10 regulons in both species at 120 h and, among other processes, is involved in axon pathfinding (Dittrich et al. 1997). 120 h, *Dorsocross3 (Doc3)* is expressed in the dorsal embryo and visual anlagen (Hammonds et al. 2013; Tomancak et al. 2002; Tomancak et al. 2007).

Overall, I have narrowed down species- and cell type specific differentially expressed to about ten genes per comparison. These genes are the most promising candidates for explaining differences in eye morphology between *D. mauritiana* and *D. simulans* using this approach. The applied techniques give an outlook to further research and investigation of these candidate genes using molecular biological approaches, especially manipulating their function using, for instance, RNA interference or CRISPR/Cas9 mediated knockouts.

#### 4.7 Temporal snRNAseq datasets of non-model species as a resource for developmental studies

My complex snRNAseq dataset covering three closely related *Drosophila* species and up to five developmental timepoints opens the possibility for a number of follow up studies. As most genetic and developmental studies deal with late L3 stage EADs, our current knowledge of earlier developmental stages remains limited. My data is covering successive stages including

early L3 stages at 72 h and 84 h AEL should be thoroughly re-analyzed to identify novel cell types at early stages that may be defined by so far unknown central regulators.

In interspecies analyses discussed in several previous sections, the majority of highly significant DEGs or the most specific regulons that diverge between species are highly timepoint specific in individual tissues. There are, however, notable exceptions such as *Gprk1*, upregulated in *D. melanogaster* over three timepoints (Figure 23) in MF cells as well as *Rbp6* which varies in species specificity over time (discussed in 4.5) while still showing a cell type specific expression pattern as visible through applying snRNAseq. Genes such as these, showing peculiar expression patterns should, in the future, be functionally examined, and their spatial expression should be confirmed applying *in situ* hybridization techniques, such as Hybridization Chain Reaction (HCR) (Evanko 2004). In this work, I investigated putative candidate genes mainly with these applications in mind. Due to this, the focus here lies on positive cluster markers which are upregulated in a specific cell type, species or time point. Though they are less readily accessible, other meaningful differences might be explained, or more easily identified, by genes specifically downregulated in the aforementioned groups.

Despite the focus in pairwise species comparisons, potentially important genes with a regulatory function that are shared between species can also be approached with our data. For instance, the gene *nejire* (*Nej*) is found as a specific regulator in PPN cells, but furthermore it appears to be highly specific for these cells at 108 h AEL in all three species (Table 8 and Table 14, column 1). This gene has a described role in compound eye development in general and photoreceptor differentiation in particular (Kumar et al. 2004). Potential challenges investigating EAD development using our data lie in developmentally relevant cell communication pathways with non-EAD cells, although global signaling such as through the ecdysone pathway (see 4.4) can potentially still be detected.

Since the EAD is a complex tissue which unites cells in different states of development within a single tissue, new tools to trace cellular developmental lineages in sc/sn RNAseq data can be applied in the future. Many of these tools are limited to applications viewing development in a linear way, for instance the transition of healthy- to diseased tissue, or diseased to cured. Data from several different timepoints can potentially help disentangling organismal- from cell type specific development by including only the specific cell types of interest that belong to the same developmental lineage. One could, for example, subset the datasets to include only PPN, MF and PhR cells to apply tools for developmental studies without distorting the individual analyses with data from tissues following different developmental lineages, such as the developing

antenna. For this kind of application, our data provides unique opportunities of combining different developmental and temporal analyses. While potentially allowing to use the aforementioned tools, this procedure requires careful selection of input cell groups, with a specific consideration for the tissue of interest. Applying such analyses could for instance allow to study photoreceptors and their precursors, cells in different antennal segments, or potentially dedicated cells forming the head cuticle or the maxillary palp in detail. Eventually, this information would add to the effort of creating organism-wide cell atlases, such as the fly cell atlas (Li et al. 2022).

The data I gathered stem from around 30 pooled EAD per dataset and likely includes samples from different sexes. Sex determination in *D. melanogaster* happens on a cell-by-cell level (Salz and Erickson 2010) and many of the genes which maintain sex specific gene expression, such as genes of the *sex lethal* family are known. One exciting application for these datasets could be the investigation of sexual dimorphisms in the eye on a cellular level. It is known that female *Drosophila* have bigger eyes than males in many species (Hilbrant et al. 2014). Furthermore, male flies often show a sex specific enlargement of the dorsal eye, the so-called love spot. Understanding how these differences come about has the potential to shed light on the gene regulatory network and the cellular mechanisms that integrate the sex determination information in a developing organ.

SnRNAseq adds a new global dimension to gene expression data. In addition to gene expression level, or log<sub>2</sub> fold change between bulk RNA sequenced samples, it is possible to assess the “spread” of a gene over all- or a subset of cells and compare that to other subsets or datasets. In bulk sequencing of 72 h, 96 h and 120 h old *D. melanogaster* EAD, (Torres-Oliva et al. 2018) identified 11 clusters describe the behavior of the expression levels of genes they include over three different timepoints. It would be intriguing to see if these genes change their “spread” in a similar manner, and if their behavior is limited to specific cell types. It is possible that individual genes strongly increase or decrease in only, for instance, a single subpopulation of cells. This could happen, for instance, if the per-cell expression levels remain similar, but the number of cells expressing a gene on a medium level changes over time. This behavior would be expected from the increasing number of photoreceptors in the EAD as the MF progresses, or continued migration of glial cells onto the EAD.

## 5 Material and Methods

### 5.1 Experimental Setup

For the experiments, the following strains were used. *Drosophila melanogaster*: Oregon R (OreR), *D. mauritiana*: TAM16 and *D. simulans*: *yellow vermillion forked* (YVF). For each species, 5 timepoints were examined, which were set as hours after egg laying (AEL): 72 h, 84 h, 96 h, 108 h and 120 h, resulting in a total of 15 samples.

For tissue dissociation, the samples were shuffled to prevent batch effect. *Drosophila* species have preferred times for egg laying, with most eggs being laid in the hours around noon. Due to this and logistical timing restrictions, egg laying and larval dissections needed to be performed partly in batches.

### 5.2 Fly husbandry and larval dissections

Flies were kept in fly food vials (400 g of malt extract, 400 g of corn flour, 50 g of soy flour, 110 g of sugar beet syrup, 51 g of agar, 90 g of yeast extract, 31.5 ml of propionic acid, and 7.5 g of Nipagin dissolved in 40 ml of Ethanol, water up to 5 L) in incubators at 25°C prior to the experiment for at least one generation. The incubators maintain 12 h light and dark cycles. For single-cell RNA sequencing (scRNAseq), eye-antennal discs were dissected from late wandering L3 larvae. For single-nuclei RNA sequencing (snRNAseq), eye-antennal discs were dissected from late L3 larvae at 120 h after egg laying. To control for larval density, eggs were deposited on yeast-coated apple agar plates for 1-2 hours and incubated for 24 h before 40 first instar larvae were transferred to food vials for further incubation.

For each dissociation experiment, at least 15 larvae were dissected for no more than 1 h in 400 µl ice-cold 1x PBS. When larval tissue was used for tests, the larvae were everted, and a mix of inner organs (e.g. imaginal discs, gut, brain etc.) was isolated. When eye-antennal discs were used, around 30 eye-antennal discs were dissected (generally from about 15 larvae). All organs were transferred into a microcentrifuge tube containing Storage Buffer (4% BSA, 0.2U Protector RNase inhibitor (Merck; 3335399001) in PBS). If the sample was to be frozen for later nuclei extraction, the tube was submerged in liquid nitrogen for 2 minutes and stored at -80 °C until further processing. For each time point around 5 more larvae were dissected to be used for imaging as a reference for developmental stage/biological age. They were fixed in 4%

PFA for 20 minutes and washed 3x20 minutes with PBT, both on a shaker at 4°C. For imaging, they were stained with Phalloidin 488 for 20 min.

### 5.3 Dissociation of larval eye-antennal imaginal discs

#### 5.3.1 Dissociation of EAD to obtain single-cell suspensions for scRNAseq

In the following, the dissociation protocol is described that was used to obtain the live cells used for scRNAseq. Supplementary Table S 1 contains detailed information about the protocol steps that had been varied and tested to achieve efficient dissociation.

~30 eye-antennal discs were dissociated in 10x TrypLE (Thermo Fisher Scientific; A1217701) containing 2.5mg/ml Collagenase (Invitrogen; 17100017) for 30 minutes on the shaker at 30°C and 300 rpm. Every 15 minutes, or, if the digestion time was only 15 min, once after 7.5 min, the discs were pipetted up-and-down 5 times using a 1,000 µl pipet tip to dissociate cell clumps efficiently. The reaction was stopped using Schneider's supplemented medium (SSM, 0.02 mg/ml Insulin in Schneiders Medium (ThermoFisher/Gibco; 21720-024)). The suspension was gently pipetted up and down around 17x with a 1,000 µl pipet tip and passed through a 35 µm cell strainer (Corning; 352235). The suspension was centrifuged for 5 minutes at 1,000 ref. For low amounts of tissue, the pellet might be small and barely visible on the side wall of the tube. Therefore, it is advantageous to use a swing bucket centrifuge to ensure that the pellet accumulates in the center at the bottom of the tube. The supernatant was removed, and the pellet was resuspended in 1x PBS. The suspension was centrifuged again (see above), the supernatant was removed, and the pellet was resuspended in 0.04% BSA (Invitrogen; 17100017) and 0.2 U/µl Protector RNase Inhibitor (Sigma-Aldrich; 3335402001) in 1x PBS.

The cells were stained depending on the application: For testing non-fluorescent live-dead assays, 10 µl of a cell suspension were mixed with 10 µl Trypan Blue (Invitrogen; 15250061). 10 µl of this solution were transferred onto a counting chamber and cells were counted using a Zeiss Telaval 31. For fluorescence-based assays, Calcein-AM (green: Sigma-Aldrich; 56496 or violet: Sigma-Aldrich; ThermoFisher Scientific; C34858) was used to stain live cells at a final concentration of 0.5 µg/ml. The suspension was incubated for 30 minutes to 1 h in the dark at room temperature on a shaker. Either DAPI or Propidium Iodide were used to stain nuclei at a final concentration of 1 µg/ml each and incubated for 10-30 min. The cell suspension was then immediately processed by Fluorescence Activated Cell Sorting (FACS) at the Universitätsmedizin Göttingen or at the Center for Molecular and Cellular Bioengineering Dresden using a Becton Dickinson BD FACSAria™ II Cell Sorter or BD FACSAria™ III Cell

Sorter. In consecutive gating steps, living cells were selected out from debris, damaged cells, and doublets. Events which were positive for calcein, as well as negative for Propidium Iodide were interpreted as live, undamaged cells. After FACS, cells were visually inspected under the microscope, counted and the volume of the suspension was adjusted with PBS and 0.04% BSA to achieve a concentration of ~1,000 cells per  $\mu\text{l}$  to match the optimal requirements for 10X Genomics scRNAseq.

### 5.3.2 Dissociation of EAD to obtain single-nuclei suspensions for snRNAseq

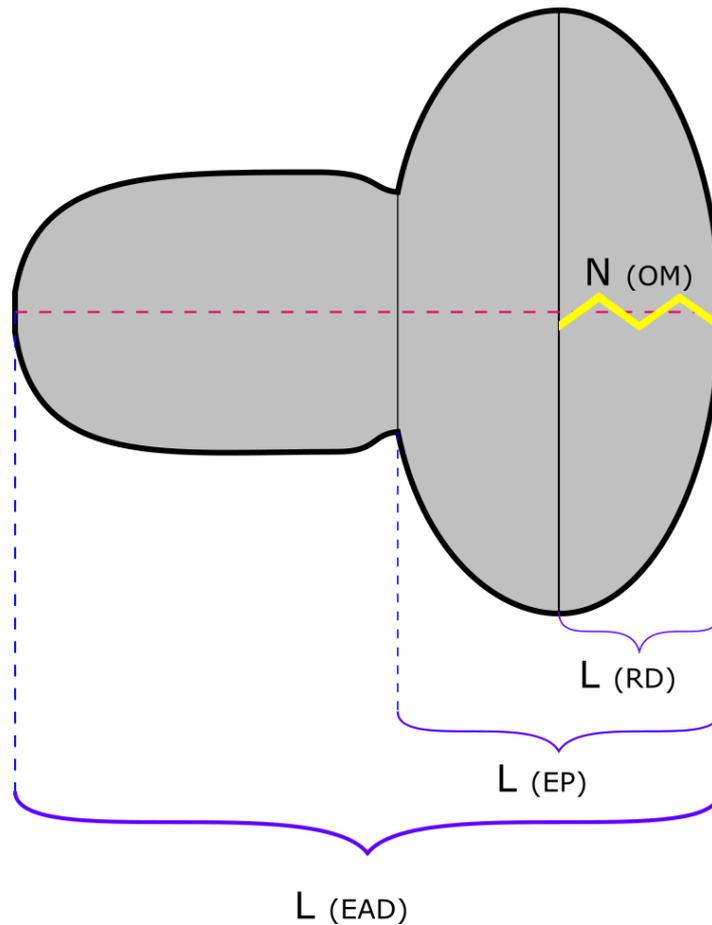
Frozen tissue was thawed at 4 °C and kept on ice for the following steps unless specified otherwise. The tissue was transferred into a precooled Dounce homogenizer (2 ml) and 500  $\mu\text{l}$  of Homogenization Buffer (HB) (0.4U/ $\mu\text{l}$  RiboLock RNase Inhibitor (ThermoFisher Scientific; EO0381), 0.2U/ $\mu\text{l}$  SUPERase In™ RNase Inhibitor (ThermoFisher Scientific; AM2694), 0.10% (v/v) Triton X-100 in NIM2; Nuclei isolation buffer 2 (NIM2): 1  $\mu\text{M}$  DTT, 1x Protease Inhibitor (Promega; G6521) in NIM1; Nuclei isolation buffer 1 (NIM1): 250 mM Sucrose, 25 mM KCl, 5mM MgCl<sub>2</sub>, 10 mM Tris HCl, pH 8 in nuclease free water) was added. The tissue was homogenized with 8 strokes of the tight pestle and kept on ice whenever possible. If the homogenization seemed incomplete after visual inspection, 1 stroke was added at a time up to a maximum of 11 strokes. The homogenized tissue was filtered through a 30  $\mu\text{m}$  MACS SmartStrainer (Miltenyi; 130-098-458) to exclude larger debris. The homogenizer was furthermore washed with 2x 500  $\mu\text{l}$  of HB to transfer as much of the tissue as possible to the cell strainer. The nuclei suspension was centrifuged at 500 g for 5 minutes at 4 °C in a swing bucket centrifuge to obtain a nuclei pellet. The supernatant was removed, and the pellet was resuspended in 500  $\mu\text{l}$  Storage Buffer.

For the following steps, the samples were transported to Dresden Concept Genome Center on dry ice. Samples were thawed on ice at room temperature and then kept on ice. For subsequent FACS either 5  $\mu\text{l}$  of a 100  $\mu\text{g}/\text{ml}$  DAPI solution (Carl Roth; 6335.1) or 1 drop of NucBlue™ (Hoechst 33342; Invitrogen Live ReadyProbes™; R37605) was added and the nuclei were incubated for 10-20 minutes for the staining to occur. During the exposure time of the staining, the sample was immediately transferred to FACS (Becton Dickinson (BD™) FACS Aria III Flow Cytometry Cell Sorter) to collect intact nuclei into a 1.5 ml microcentrifuge tube pre-coated with 1% BSA containing 0.04% BSA in 5  $\mu\text{l}$  PBS. The concentration should be around 1,000 nuclei per  $\mu\text{l}$  to match the optimal requirements for 10X Genomics snRNAseq. The gates were set to select for DAPI positive nuclei. Particles smaller than 1  $\mu\text{m}$  were excluded to remove

small debris and damaged nuclei. Doublets and irregular shaped debris were also filtered out through gating as much as possible. Nozzle size was 100  $\mu\text{m}$ . FACS was performed at the Universitätsmedizin Göttingen or at the Dresden Concept Genome center using a Becton Dickinson BD FACSAria™ II Cell Sorter or BD FACSAria™ III Cell Sorter. The settings were adjusted using unstained and stained samples.

### 5.1 Image based assessment of developmental stages

Developmental time might differ slightly between different species. Therefore, images of eye-antennal imaginal discs from the same batches used for snRNAseq were taken to assess the progress of disc development in each species for each time point. The images were taken on an LSM (Confocal Microscope Zeiss LSM980) at 40x resolution. As the larvae age, ommatidia start differentiating which is a visible indicator for development. For each time point, around five discs were taken as a reference for developmental age from the same batch that was used for nuclei isolation.



**Figure 37: Measurement Scheme for developmental metrics.** Left: Anterior. Right: Posterior. Red dashed line: Midline set for reference between the anterior center of the antennal disc and the center of the optic stalk. N(OM): Counting pattern of the number of ommatidial rows. The number of ommatidia was counted around the midline in a zig-zag pattern due to the offset of the hexagonal ommatidia (counted along the yellow path). L(EP): Length of the eye primordium. L(RD): Length of the head domain. L(EAD): Total length of the eye-antennal imaginal disc, measured between the center of the anterior end of the antennal disc and the optic stalk.

The row number of ommatidia was counted (Figure 37). Additional metrics were taken for the assessment of the general size of the EAD such as the length of the eye primordium and the total disc length. Due to time constraints, only few discs could be taken and analyzed. Image analysis was performed using Fiji (Schindelin et al. 2012). As a measure for the developmental progression of the discs, the distance from the Optic stalk to the anterior most point of the antennal disc was measured.

## 5.2 Library preparation and 10x Genomics sequencing

scRNAseq and snRNAseq were performed at the Dresden Concept Genome Center on a 10x Genomics Chromium sequencing system. The viability of the sorted cells or quality of nuclei were visually inspected under a light microscope (with 200x magnification) from a small aliquot of cells or nuclei stained with Trypan blue.

Up to 20,000 cells/nuclei were carefully mixed with reverse transcription mix using the Chromium Single Cell 3' Library & Gel beads chemistry v3 (10X Genomics, PN 1000075) and loaded into a Chromium Single Cell B Chip (10X Genomics, PN 1000073) on the 10X Genomics Chromium system (Zheng et al. 2017).

Following the guidelines of the 10X Genomics user manual, the droplets were directly subjected to reverse transcription, the emulsion was broken, and cDNA was purified using Dynabeads MyOne Silane (10X Genomics). After cDNA amplification (11 cycles for cells, 12 cycles for nuclei), the sample was purified and underwent a quality control check on the Fragment Analyzer.

Preparation of single-cell or -nuclei RNA-seq libraries (fragmentation, dA-Tailing, adapter ligation and an indexing PCR step with 12 cycles (cells) or 15 cycles (nuclei)) followed the manufacturer's recommendations. After quantification, the libraries were sequenced on an Illumina NextSeq 500 using a high-output flowcell in PE mode (R1: 28 cycles; I1: 8 cycles; R2: 56 cycles) or on the Illumina Novaseq 6000 system with a S2 flowcell in PE mode (R1: 28 cycles; I1: 8 cycles; R2: 94 cycles). An average of 13,000 fragments per cell were sequenced.

### 5.3 Genome annotation transfer and read mapping

Genome annotation was transferred from *Drosophila melanogaster*, annotation r6.37 to a genome sequence from OreR (REF) using LiftOff (Shumate and Salzberg 2020). The obtained sequencing data from scRNAseq/snRNAseq of *D. melanogaster* EAD were mapped to a genome of the strain Oregon-R (OreR) (FBsn0000276) (Buchberger et al. 2021; Torres-Oliva et al. 2016). The five datasets of *D. simulans* were mapped to a genome of the strain *yellow vermillion forked* (YVF)(DSSC, University of California, San Diego, Stock no.14021-0251.146) and the four datasets of *D. mauritiana* were mapped to a genome of (TAM16) (Nolte et al. 2013). scRNAseq/snRNAseq reads were counted with 10x Genomics Cellranger 5 using default settings for mapping single-cell data. For mapping single-nuclei data, the option "--include-introns" was added.

### 5.4 Data Analysis

The following chapters describe the general workflow for snRNAseq dataset that is shared between most of them.

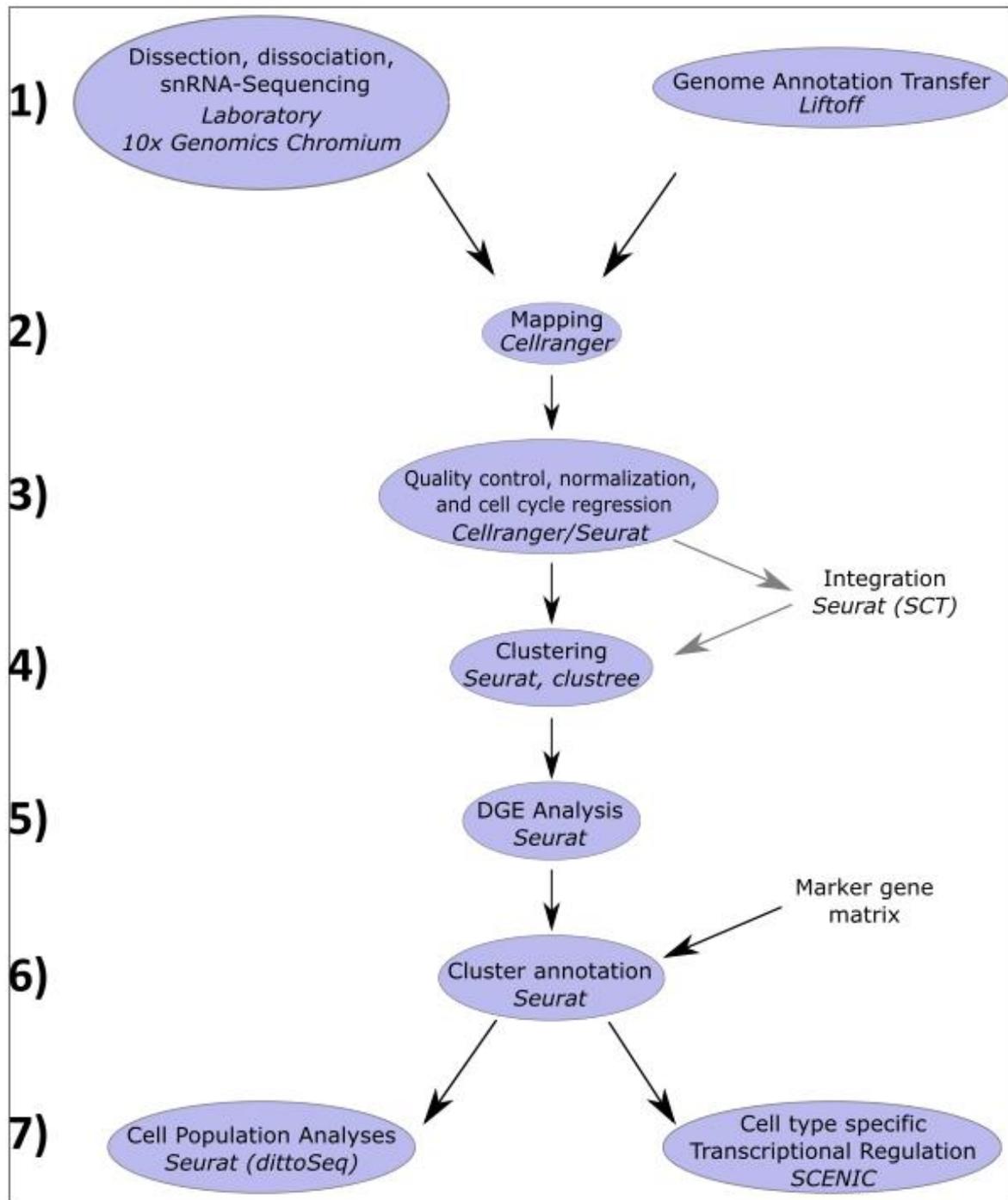


Figure 38: Overview over the general analysis workflow for each dataset. Grey arrows: Integration was performed only to create pairwise comparisons (Interspecies and single cells to single nuclei).

All mapped datasets are subjected to quality control and normalization first. For pairwise dataset comparisons, data integration was performed by using the SCT workflow of Seurat. Afterwards, clustering is performed, and clusters were annotated using first an automatized algorithm followed by manual inspection. Then, DEGs were identified between meaningful clusters. From here on, the following analyses differ between datasets and scientific questions.

#### 5.4.1 Preprocessing and general analyses

General data analyses were performed using R version 4.1.1 (2021-08-10). Specifically, the package Seurat (Hao et al. 2021) was used for single-cell specific applications. This includes quality control steps, such as calculating the percentage of mitochondrial, ribosomal and heat shock related genes and removing doublets and cells or nuclei of poor quality. Cells of poor quality were defined as expressing more than 3,000 or less than 300 genes. Nuclei of poor quality were defined as the top 1 % of nuclei expressing the highest number of genes or genes in cells, or the top 1 % of nuclei expressing the highest number of genes or less than 300 genes. These numbers were set for the purpose of comparability. Genes were kept if they were expressed in at least 5 cells (for scRNAseq) or 3 nuclei (for snRNAseq). Normalization across all cells which passed quality cutoffs within a dataset was performed using the SCTransform method (Hafemeister und Satija 2019).

Certain biological signals can interfere with the clustering in snRNAseq data. For instance, nuclei might cluster based on their cell cycle phase instead of their cell type specific signals. To reduce noise introduced by different cell cycle phases, the difference between the G2M and S phase scores was regressed out. This way, the information about the specific cell cycle phase of each nucleus was excluded, while the difference between proliferating and non-proliferating cells was retained. This is important to distinguish differentiated tissue, such as photoreceptors from tissue which is still proliferating. Since some cell types might be characterized by their cell cycle phase, such as nuclei undergoing the second mitotic wave, the information of the whether a cell is mitotically active is retained. The regression is based on a list of *Drosophila* cell cycle markers created by Harvard Chan Bioinformatics Core based on orthology search of a list of human cell cycle markers (Tirosch et al. 2016) (Supplementary Table S 9).

After quality control and normalization of the data, a principal component analysis is performed on the data to reduce dimensionality and thus speed up computing as well as to remove noise. 50 Principal components were chosen for downstream analysis for every dataset.

To identify major cell types, unbiased clustering was performed using a nearest-neighbor clustering algorithm (Waltman and van Eck 2013). For clustering, it is necessary to pick a resolution at which the data is clustered. This influences the level at which the data is examined. With increasing resolution, it is possible to find cell types which are present only in ever smaller numbers in the tissue, or which are transcriptionally similar to other cell types. The drawback of this procedure is that some of these some might be less informative, and with higher resolutions computational demands increase as well. To find a suitable resolution, an additional

package called `clustree` (Zappia and Oshlack 2018) was used. This package allows visually estimating the confidence of different clustering resolution in relation to each other. The clustering was chosen also at the minimum resolution where the expected stable cell types (see example in Supplementary Table S 10) could be found. If not stated otherwise, the `ggplot2` package was used to create scientific plots (Hadley 2016). Inkscape was used to annotate and combine the plot (The Inkscape Team 2019).

#### 5.4.2 Comparison of scRNAseq and snRNAseq

To compare the most variable genes between scRNAseq and snRNAseq, the top 3,000 variable genes for both datasets were obtained based on the differential expression analysis (see above). Both lists were compared to identify those genes that were unique for each dataset and those that were shared. A gene ontology enrichment analysis was performed for the two lists of unique genes and the list of shared genes, respectively, using Gene Ontology. Gene ontology enrichment plots were created using ShinyGO (version 0.76) (Ge et al. 2020).

For the analysis of integrated data from both single cells and single nuclei, I applied the standard workflow in Seurat (Stuart et al. 2019). I used 3,000 integration features and used `FindIntegrationAnchors` and `IntegrateData` adapted for datasets normalized using `SCtransform`. Barplots of dataset-specific cell type proportions were created using the R package `dittoSeq` (Bunis et al. 2020). The percentages of cells or nuclei in these plots were corrected for absolute number by multiplying the percentages in cells by 0.74, the ratio of filtered nuclei to filtered cells. UMAP plots were created using Seurat. Each dot represents a single cell or nucleus. They are positioned based on their relative transcriptional similarity to each other.

#### 5.4.3 Comparison of snRNAseq data and bulk RNA-sequencing data

To have a reference for the accuracy of captured snRNAseq data, gene expression of our 120 h AEL *D. melanogaster* dataset were compared to a previously published bulk sequencing dataset (GSE94915 (Torres-Oliva et al. 2018) ). This dataset contains three replicates of RNA read counts in 120 h old *D. melanogaster* EAD. The datasets were made comparable by pooling the reads of each gene over all cells in single nuclei sequencing data and using the median read count for each gene of the three bulk sequencing replicates. FlyBase (Gramates et al. 2022) identifiers of this dataset were validated and converted to gene symbols using the FlyBase website ID validator.

#### 5.4.4 Differential gene expression analyses

Differential gene expression (DGE) analysis is a standard step in single-cell analysis used to distinguish cell clusters based on their characteristic gene expression. Clusters can be determined by clustering analysis or predetermined such as when combining different time points or species.

##### 5.4.4.1 Differential gene expression analysis for cluster annotation

I performed DGE analysis to assign and annotate cell clusters identified by unbiased clustering in my data based on the expression of marker genes when compared to markers from the literature, most importantly previous work done by Ariss et al. 2018. These marker genes enriched in each cell cluster were identified by differential expression analyses (i.e. genes expressed in a cluster vs. all other clusters) followed by a cutoff of  $\log_2$ fold-change  $> 0.25$  and an adjusted p-value  $< 0.05$ . Differentially expressed genes for each cell cluster were used to test for enrichment of gene ontology (GO) terms (i.e. Biological Process) using the R package gprofiler2 (GO enrichment for scRNAseq: Supplementary Table S 3; GO enrichment for snRNAseq: Supplementary Table S 5) using the g:SCS algorithm for correcting for multiple testing. As input, the respective genes lists were submitted with the genes ranked from lowest to highest p-value unless stated otherwise. The top four genes defining each cell cluster were chosen by the lowest adjusted p-value.

##### 5.4.4.2 DGE for timepoint specific gene expression comparisons

For timepoint specific DGE, the raw reads were used to identify DEGs. This was done because normalization can possibly remove biologically relevant variation between datasets. Heatmaps were created using the R package pheatmap (Raivo Kolde 2019).

##### 5.4.4.3 DGE for interspecies cell type comparisons

For cell type specific interspecies comparisons, the raw reads were used to identify DEGs. This was done because normalization can possibly remove biologically relevant variation between datasets. I furthermore applied a cutoff of a percentage expression change of 20% or more, i.e. the percentage of cells of an examined cell population expressing a DEG changes by at least 20%. Differential gene expression analyses were also performed between species and timepoints.

#### 5.4.4.4 SCENIC analysis for identifying regulons underlying cell type specific gene expression patterns

To explain the gene expression patterns seen in predetermined cell types, single-cell regulatory network inference and clustering (SCENIC) was used for identifying transcription factors showing high specificity (Aibar et al. 2017). This method is applied because these TFs are often not directly detectable since they often show only relatively low expression levels. Differences in these transcription factors between species or stages may explain broader phenotypic differences in adult flies and eye development in general. In contrast, common regulons might be crucial TFs for the development or maintenance of a specific cell type. For SCENIC analysis, a Seurat object was converted to a loom file using the R package SeuratDisk (Hoffman 2021). SCENIC analysis was performed in Python using pySCENIC (version 0.12.1). Extensive lists of regulons can be found in Supplementary Table S 18.

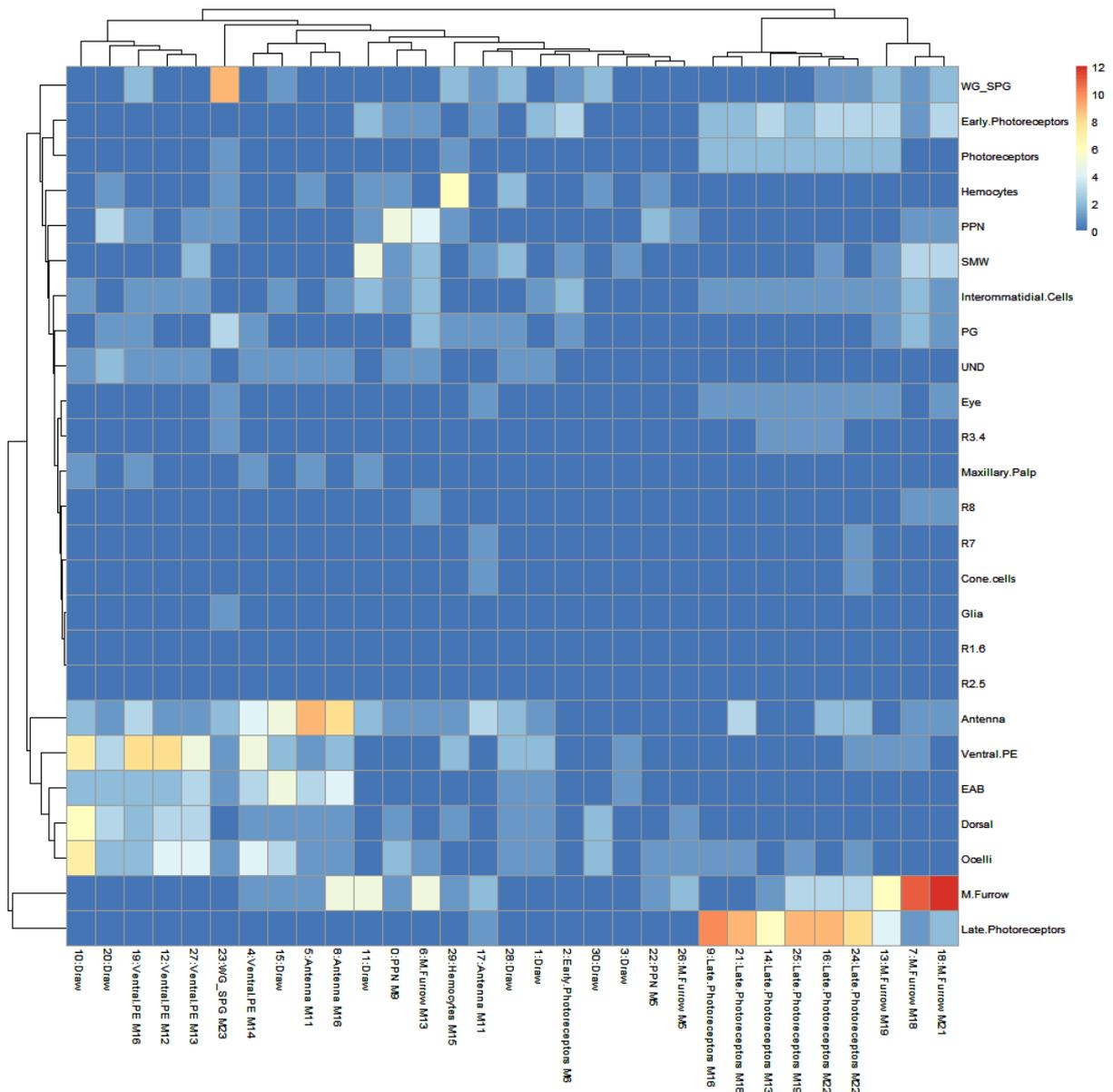
#### 5.4.5 Standardization of cluster identity annotations

Cell types, or cell states, can be identified by performing differential gene expression analysis between each cluster. For my work, I used a simple algorithm to assign a cell type identity to each cluster. I gathered published markers from previously published single-cell RNA sequencing datasets as well as other literature (list of cell cluster markers for scRNAseq: Supplementary Table S 2; list of cell cluster markers for snRNAseq: Supplementary Table S 4). These markers were used to assemble a marker gene matrix which show genes over cell types. If a gene is a marker for one or more cell types, it got assigned a 1 for each cell type. This matrix was used to assign cell identity scores to each cluster within a dataset based on the DEG identified as explained in 5.4.4.1. Then, the cluster gets assigned the identity of the highest value. If there are two identities with an equal score, they are assigned as “Other”, if no markers were found it is assigned “No Markers”. Based on the differentially expressed marker genes I am able to identify most of the expected cell types in the EAD.

**Table 16: Overview over the most common cell types and their abbreviations used in this work.**

|     | <b>Name</b>                            | <b>Abbreviation</b> |
|-----|--|---------------------|
| 1)  | Pre-Proneural cells                    | PPN                 |
| 2)  | (Ventral) Peripodial Epithelium cells  | Ventral_PE          |
| 3)  | Dorsal cells                           | Dorsal              |
| 4)  | Ocelli                                 | Ocelli              |
| 5)  | Morphogenetic Furrow cells             | MF                  |
| 6)  | Second-Mitotic Wave cells              | SMW                 |
| 7)  | PhotoReceptors                         | PhR                 |
| 8)  | Interommatidial cells                  | Interommatidial     |
| 9)  | Eye-antennal border cells              | EAB                 |
| 10) | Antennal cells                         | Antenna             |
| 11) | Wrapping- and Subperineural Glia cells | WG_SPG              |
| 12) | Hemocytes                              | Hemocytes           |

This procedure is used to be able to quickly assign cell type identities to all clusters in all datasets. However, this results in inconsistent annotations when comparing different datasets. Some cell types were not found because the markers of one cell type shrouded another (e.g. the “Ocelli” cluster was often hidden in dorsal- and PE clusters). Therefore, it was necessary to come up with standardized cell types to identify in our data. To do this, markers gene scores were first plotted in a heatmap for each dataset using the R package pheatmap (Raivo Kolde 2019). This allows to see clusters which can be coalesced and clusters which are hidden inside of others.



**Figure 39:**Heatmap of cell type scores in 120 h old EAD of *D. simulans*. Displayed are identified cell clusters (X) and a list of possible identities (Y). In this plot, each cluster number (0-30) has automatically been assigned a cell identity. Each identity in the cluster names is followed by MX, where X is the marker score for the assigned (and therefore highest) identity. This plot is an aid for manual cell type assignment and to resolve clusters with a complex gene expression pattern such as “Draw” clusters.

Table 16 includes cell types I can confidently identify in (almost) all datasets. Some clusters could not unequivocally be assigned to cell types (see “Draw” clusters in Figure 39). In contrast to the identities defined by Ariss et al., less defined identities, such as “Early.Photoreceptors” are grouped together (in this case with “Late.Photoreceptors”) as “PhR” in my manual annotation (e.g. cluster 2 in Figure 39). These clusters often contain heterogeneous groups of cells. An often occurring example are combinations of an antennal- and EAB cells (denoted as “Antennal\_EAB” cluster) or dorsal and ocelli cells (denoted as “Dorsal\_Ocelli” cluster). Interommatidial cells and cells of the SMW also often do not form individual clusters (see

cluster 11 in Figure 39), especially at younger stages. Beyond that, I can often identify sub types, such as different photoreceptor types and antennal segments (A1, A2, A3, Arista). Glia cells mostly group together within WG\_SPG. For each dataset new names for each cluster were then assigned as listed above.

It is to note that the resulting cell types represent those which I have a high confidence in. More specific clusters, such as different photoreceptor clusters which may represent different photoreceptor types, are not immediately visible using this annotation. Furthermore, some cell clusters either show ambiguous DEG profiles, or I could not find any markers at all. These clusters are labeled as “Other”.

With the assigned cell types, clusters were compared using different tools. The R package dittoSeq was used to directly compare cluster compositions in pairwise integrated data. To this end, cell numbers in these clusters were corrected by multiplying the absolute cell numbers with a factor correcting for the proportions of the total number of cells of the pairwise compared species.

## 6 Contributions

This project was conceptualized by Dr. Nico Posnien. Gordon Wiegleb performed the experiments and data analysis. snRNAsequencing and library preparation was done in collaboration with the Dresden Concept Genome Center. The Project was financially supported by the International Max-Planck Research School for Genome Science and the Göttingen Graduate Center for Neurosciences, Biophysics, and Molecular Biosciences.

## 7 Acknowledgments

I am thankful that I could be in these research schools and in the same wake would like to especially thank Henriette Irmer and Frauke Bergmann for their always friendly support and enthusiasm for the IMPRS-GS. I would like to thank the CMCB Flow Cytometry Core Facility in Dresden and the Core Facility Cell-Sorting at the UMG in Göttingen for their support with FACS sorting. Thanks to the group of Holger Bastians for providing us with equipment for flow cytometry. I also want to thank Xingbo Xu at the UMG, as well as Jordi Solana and Ana Veloso for advice on single-nuclei RNA sequencing. I would like to thank Johannes Söding and Argyris Papantonis for being lively members of my TAC and giving valuable input on my whole journey. I would like to thank Gregor Bucher, Daniel Jackson and Jochen Rink for agreeing to be members of my examination board.

Most importantly, I want to thank Nico Posnien, for being a great supervisor, and a great person. I thank you so much for all the chances you have given me, and even if I go, I hope you will have a successful way ahead! Of course, I want to thank all former members of our group as well. Just as Nico is a great supervisor, I also appreciate that I could share the group with such friendly colleagues.

Special thanks go to Julia, Carmela, and Noel and everybody who does not believe in taking busses at night, you made this a most pleasant journey since I met you.

I will not close this chapter without thanking 猪猪侠. I am incredibly happy we met, and that you were with me in this intense time, bringing joy no matter what.

Vielen Dank an meine Eltern und meine Familie. Ihr wisst, wie sehr ich euch brauche, und ich möchte, dass ihr wisst, wie viel mir das bedeutet. Ohne ich euch hätte ich den Weg bis hier hin nicht geschafft, und vermutlich auch schon viel früher nicht mehr.

## 8 Supplementary Material

### **This Section contains:**

Supplementary Table S 1

Supplementary Table S 1 – Supplementary Table S 17

**The following Supplementary Tables are available separately as files under DOI:  
<https://doi.org/10.25625/IZOLWH>**

Supplementary Table S 2: List of marker genes for each cluster in scRNAseq analysis.

Supplementary Table S 3: GO analysis of cell clusters identified in scRNAseq.

Supplementary Table S 4: List of marker genes for each cluster in snRNAseq analysis.

Supplementary Table S 5: GO analysis of cell clusters identified in snRNAseq.

Supplementary Table S 6: List of top 3,000 variable genes for scRNAseq and snRNAseq data, respectively.

Supplementary Table S 7: Top 3,000 variable genes and GO enrichment results for scRNAseq, snRNAseq and shared.

Supplementary Table S 8: Score Matrix used to annotate cell types and list of references for individual marker genes used for cluster/cell type annotation.

Supplementary Table S 9: Cell Cycle Marker Genes used to identify the cell cycle phases of nuclei.

Supplementary Table S 10: Read Count differences between bulk- and snRNAseq data of 120 h old *D. melanogaster* EAD.

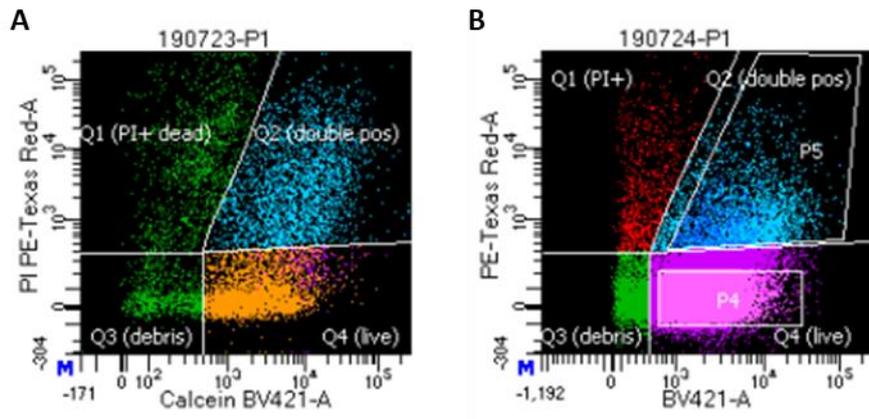
Supplementary Table S 13- Supplementary Table S 17:

Differentially expressed gene lists for pairwise cell type comparisons between species.

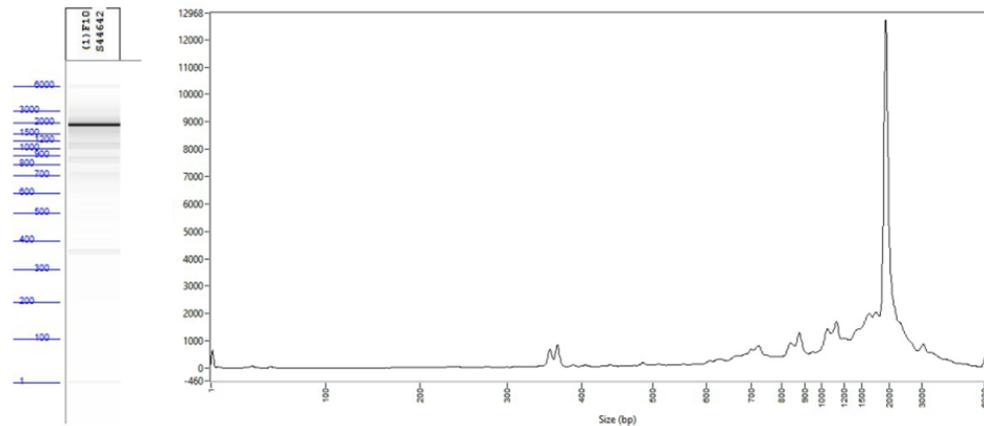
Supplementary Table S 18: Table of regulon specificity scores for all cell types by species.

**Supplementary Table S 1: Overview of different dissociation conditions.** Samples within blocks (highlighted in grey and white) were prepared in parallel. The Flow Cytometer only provides percentages of survival because it stops after a defined number of events (i.e. ~50,000 cells) and therefore absolute Numbers are not meaningful. “Pipetting” refers to the number of strokes during and after incubation. The cells obtained by experiment/block 12 were subjected to 10x Genomics scRNAseq.

| Block | Enzyme during homogenization      | Tissue        | # EAD | Filtration   | Pipet strokes, # | Temperature, °C | Incubation time, min | Shaking, rpm | Survivalrate, % | cell number | Stain              | Device         |
|-------|-----------------------------------|---------------|-------|--------------|------------------|-----------------|----------------------|--------------|-----------------|-------------|--------------------|----------------|
| 1     | 1x TrypLE                         | Larval organs | NA    | NA           | 5 + 20           | 30              | 15                   | 300          | 16              | NA          | CaGreen, PI        | Flow Cytometer |
|       | 1x TrypLE                         | Larval organs | NA    | NA           | 5 + 20           | 37              | 15                   | 300          | 23              | NA          | CaGreen, PI        | Flow Cytometer |
|       | 10x TrypLE                        | Larval organs | NA    | NA           | 5 + 20           | 30              | 15                   | 300          | 29              | NA          | CaGreen, PI        | Flow Cytometer |
|       | 10x TrypLE                        | Larval organs | NA    | NA           | 5 + 20           | 37              | 15                   | 300          | 27              | NA          | CaGreen, PI        | Flow Cytometer |
| 2     | 10x TrypLE                        | Larval organs | NA    | NA           | 5 + 20           | 30              | 15                   | 300          | 25              | NA          | CaGreen, PI        | Flow Cytometer |
|       | 10x TrypLE                        | Larval organs | NA    | NA           | 5 + 20           | 37              | 15                   | 300          | 29              | NA          | CaGreen, PI        | Flow Cytometer |
|       | 10x TrypLE, 2.5 mg/ml Collagenase | Larval organs | NA    | NA           | 5 + 20           | 30              | 15                   | 300          | 36              | NA          | CaGreen, PI        | Flow Cytometer |
| 3     | 10x TrypLE, 2.5 mg/ml Collagenase | Larval organs | NA    | NA           | 5 + 20           | 37              | 15                   | 300          | 28              | NA          | CaGreen, PI        | Flow Cytometer |
|       | 10x TrypLE                        | Larval organs | NA    | 35 µm        | 5 + 20           | 30              | 10                   | 300          | 41              | NA          | CaGreen, PI        | Flow Cytometer |
|       | 10x TrypLE                        | Larval organs | NA    | 35 µm        | 5 + 20           | 30              | 20                   | 300          | 31              | NA          | CaGreen, PI        | Flow Cytometer |
|       | 10x TrypLE                        | Larval organs | NA    | 35 µm, 20 µm | 5 + 20           | 30              | 20                   | 300          | 30              | NA          | CaGreen, PI        | Flow Cytometer |
| 4     | 10x TrypLE                        | Larval organs | NA    | 35 µm        | 5 + 20           | 30              | 10                   | 300          | 36              | NA          | CaGreen, PI        | Flow Cytometer |
|       | 10x TrypLE                        | Larval organs | NA    | 35 µm        | 5 + 20           | 30              | 15                   | 300          | 35              | NA          | CaGreen, PI        | Flow Cytometer |
| 5     | 10x TrypLE                        | Larval organs | NA    | 35 µm        | 5 + 20           | 30              | 20                   | 300          | 39              | NA          | CaGreen, PI        | Flow Cytometer |
|       | 10x TrypLE, 2.5mg/ml Collagenase  | EAD           | 22    | 35 µm        | 5 + 20           | 30              | 15                   | 300          | 1               | 394         | CaGreen, PI, DAPI  | FACS           |
|       | 10x TrypLE, 2.5 mg/ml Collagenase | EAD           | 32    | 35 µm        | 5 + 18           | 30              | 10                   | 300          | 14              | 16,875      | CaGreen, PI, DAPI  | FACS           |
|       | 10x TrypLE, 2.5 mg/ml Collagenase | EAD           | 31    | 35 µm        | 5 + 18           | 30              | 30                   | 300          | 13              | 16,382      | CaGreen, PI, DAPI  | FACS           |
|       | 10x TrypLE, 2.5 mg/ml Collagenase | EAD           | 31    | 35 µm        | 5 + 17           | 30              | 30                   | 300          | 5               | 8,613       | CaGreen, PI, DAPI  | FACS           |
|       | 10x TrypLE, 2.5 mg/ml Collagenase | EAD           | 30    | 35 µm        | 5 + 17           | 30              | 30                   | 300          | NA              | 25          | CaGreen, PI, DAPI  | FACS           |
|       | 10x TrypLE, 10 mg/ml Collagenase  | EAD           | 30    | 35 µm        | 5 + 17           | 30              | 30                   | 300          | NA              | 350         | CaViolet, PI, DAPI | FACS           |
|       | 10x TrypLE, 10 mg/ml Collagenase  | EAD           | 28    | 35 µm        | 5 + 17           | 30              | 30                   | 300          | 41              | 2,985       | CaViolet, PI, DAPI | FACS           |
| 11    | 10x TrypLE, 10 mg/ml Collagenase  | EAD           | 28    | 35 µm        | 5 + 17           | 30              | 60                   | 300          | 58              | 16,208      | CaViolet, PI, DAPI | FACS           |
|       | 10x TrypLE, 10 mg/ml Collagenase  | EAD           | 28    | 35 µm        | 5 + 17           | 30              | 60                   | 300          | 58              | 16,208      | CaViolet, PI, DAPI | FACS           |

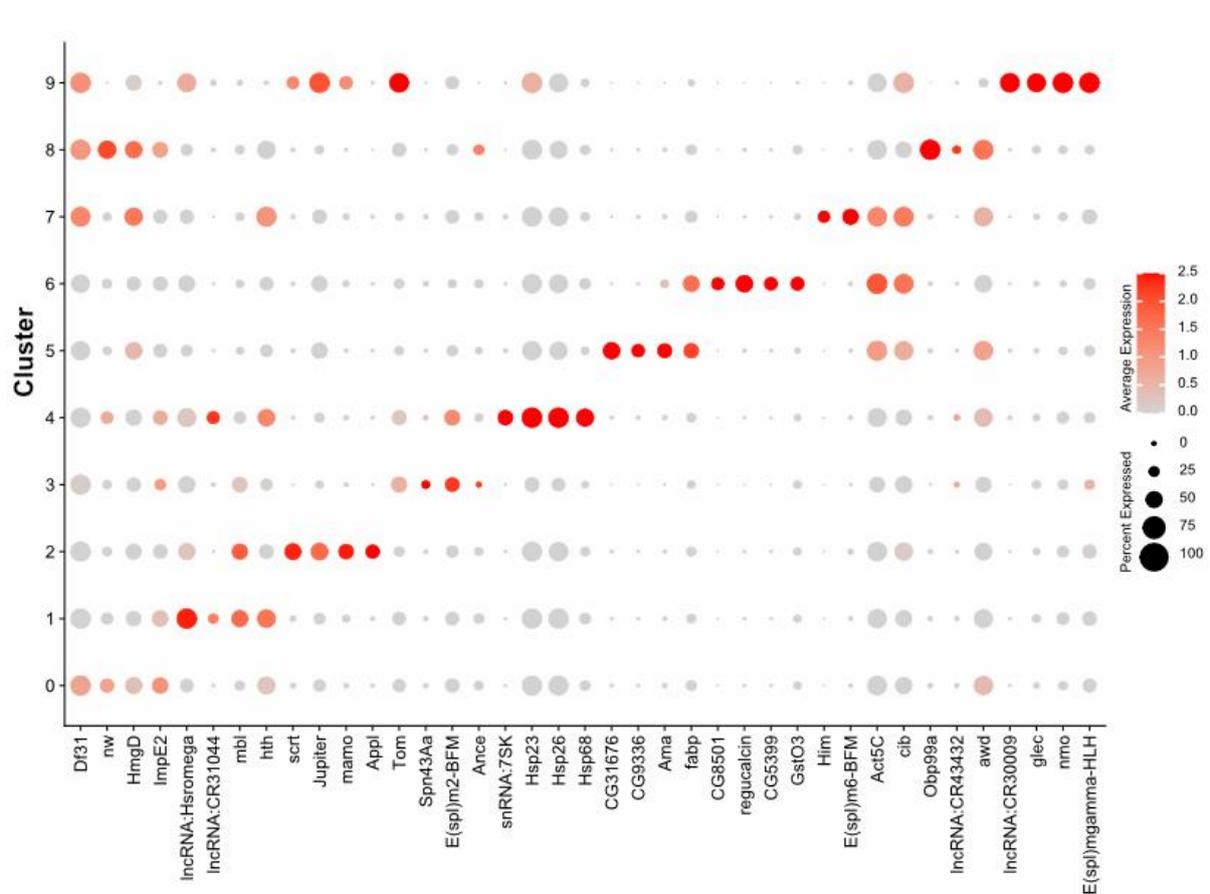


**Supplementary Table S 1: FACS-plot of *D. melanogaster* eye-antennal disc cells after live/dead cell staining.** (A) Counterstaining of propidium iodide to label dead cells (y-axis, Q1) and Calcein violet to label live cells (x-axis, Q4). Double positive signals might indicate dying cells or incompletely separated cells (Q2). This method allows removing debris (Q3) efficiently. (B) The cell population P4 (i.e. 16,208 living cells) were isolated and used for the scRNAseq run using 10x Genomics.

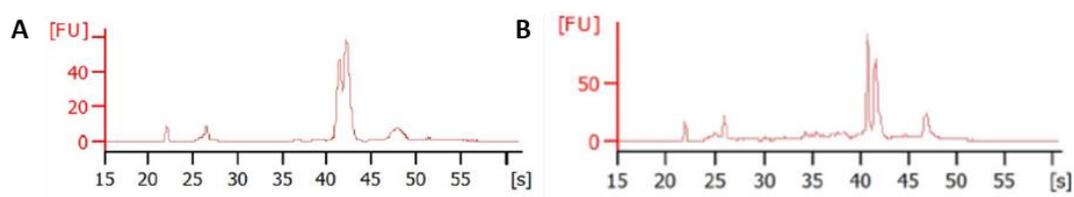


| Sample ID | Sample   | Range         | ng/uL | % Total | nmole/L | Avg. Size | CV   |
|-----------|----------|---------------|-------|---------|---------|-----------|------|
| S44642    | 30 cells | 100 - 5500 bp | 15.9  | 99.2    | 17.7    | 1479      | 56.3 |

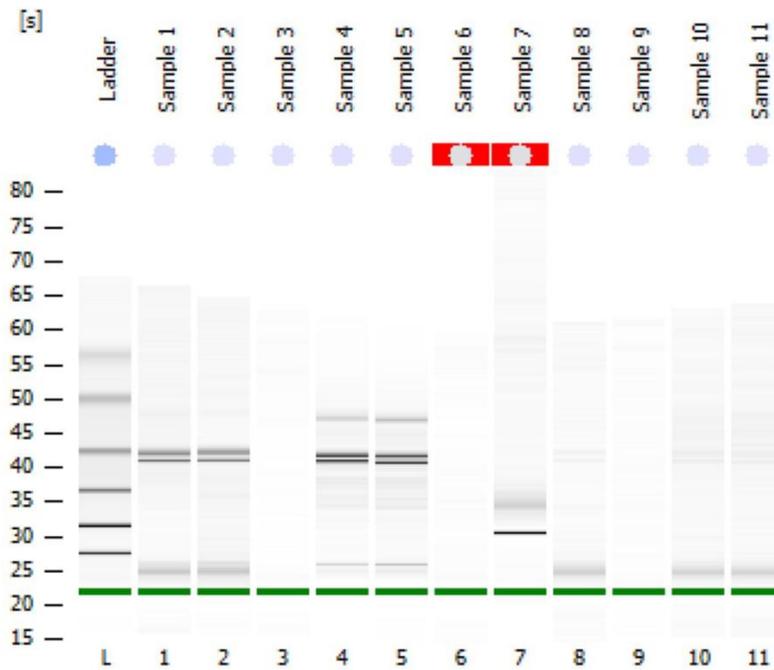
**Supplementary Table S 2: Quality and quantity of cDNA after reverse transcription of mRNA fraction (polyA-based enrichment) and full-length cDNA amplification from cell lysate of 30 cells sorted from cell suspension of eye-antennal discs run on Fragment Analyzer (Agilent).** Size distribution of all fragments show little impact on degradation (almost no cDNA detectable below 400 bp).



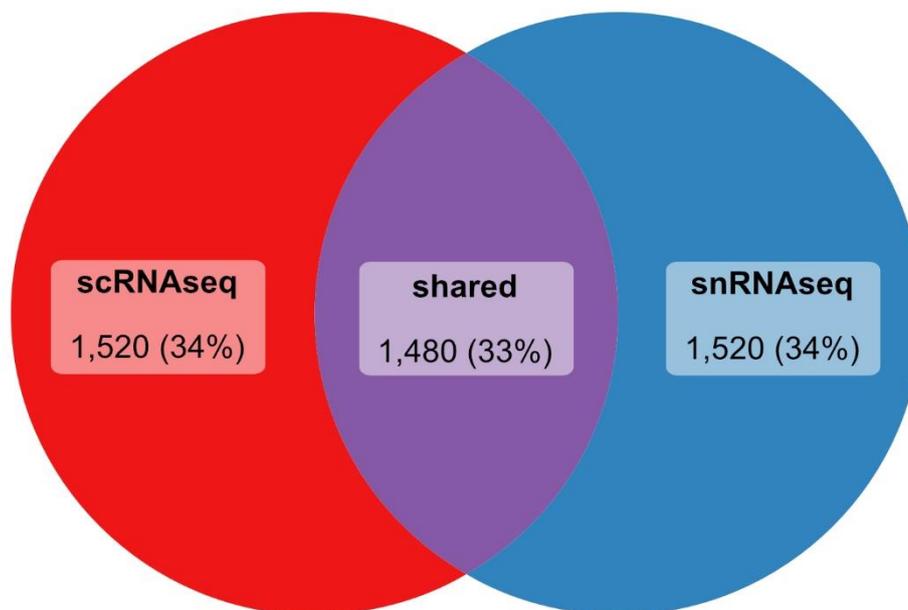
**Supplementary Table S 3: Dotplot of the top four overexpressed genes in each cluster in scRNAseq data.** Dots are colored by the expression level of each gene and scaled by the percentage of cells within a cluster which express the respective gene. Clustering was performed with Seurat at a resolution of 0.2.



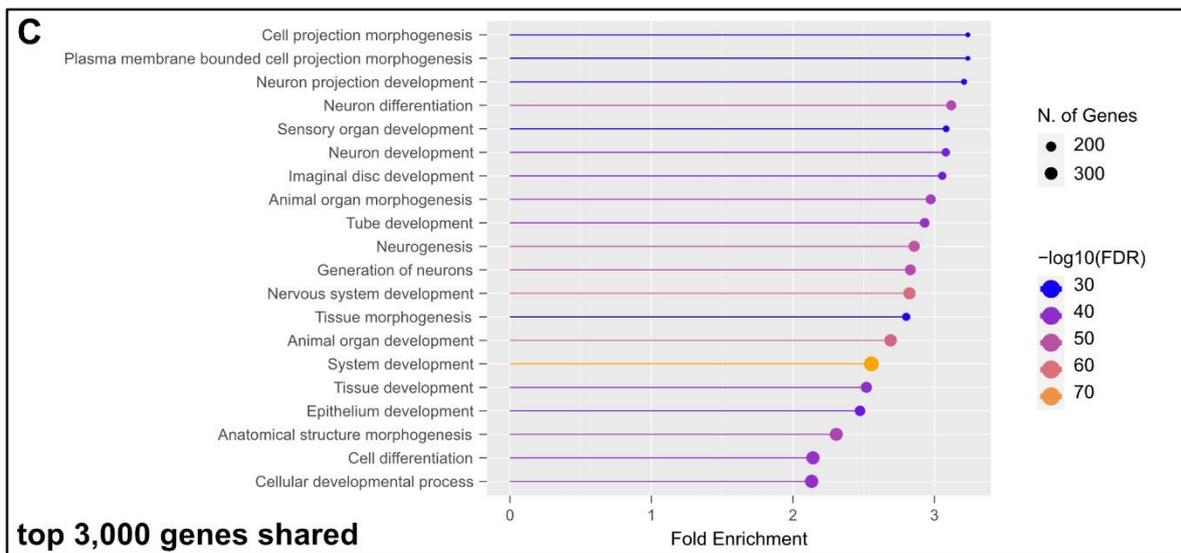
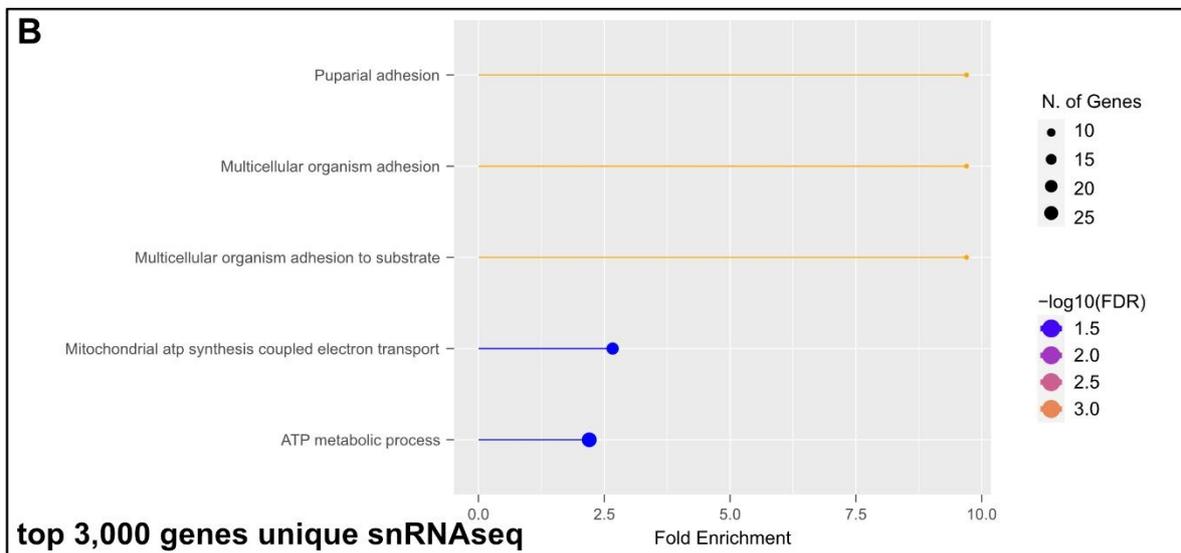
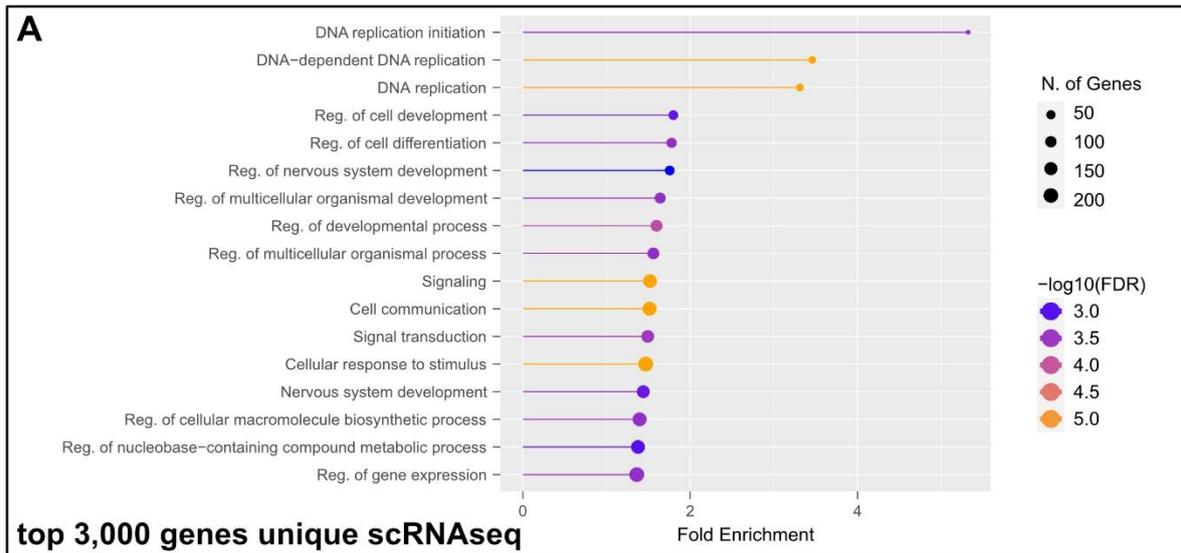
**Supplementary Table S 4: Fluorescence intensity curves from Bioanalyzer for fresh- and cryopreserved nuclei obtained by different nuclei extraction protocols.** (A) The RNA was extracted directly from a fresh sample (36 eye-antennal discs), which was dissociated using the 10x Genomics protocol with 0.1% IGEPAL as a detergent. (B) RNA isolated from a cryopreserved sample, which was dissociated using a protocol based on Triton X-100 as a detergent and a variety of RNase inhibitors. Note that the sample was thawed for 3.5h before being frozen again. Both curves are close to the expectation of RNA isolated from *D. melanogaster*.



**Supplementary Table S 5: BioAnalyzer results comparing different nuclei isolation protocols for frozen samples.** Samples 1, 2, 4 and 5 were dissociated using the protocol based on Triton X-100 as a detergent and a variety of RNase inhibitors (Litvinukova et al. 2018). Samples 1 and 2 were dissociated by pipetting up and down and samples 4 and 5 were dissociated using a Dounce homogenizer. Samples 7 and 8 were dissociated using the protocol “10x Genomics® Isolation of Nuclei for Single Cell RNA Sequencing” (10X Genomics 2021). Samples 7 and 8 were dissociated using only citric acid buffer and samples 10 and 11 were dissociated using only a detergent. Note that for *D. melanogaster*, intense bands are expected at about 40s and weaker bands at 25s and 45-50s (Jeffrey A. Fabrick, J. A. and Hull 2017). Each run was repeated once.

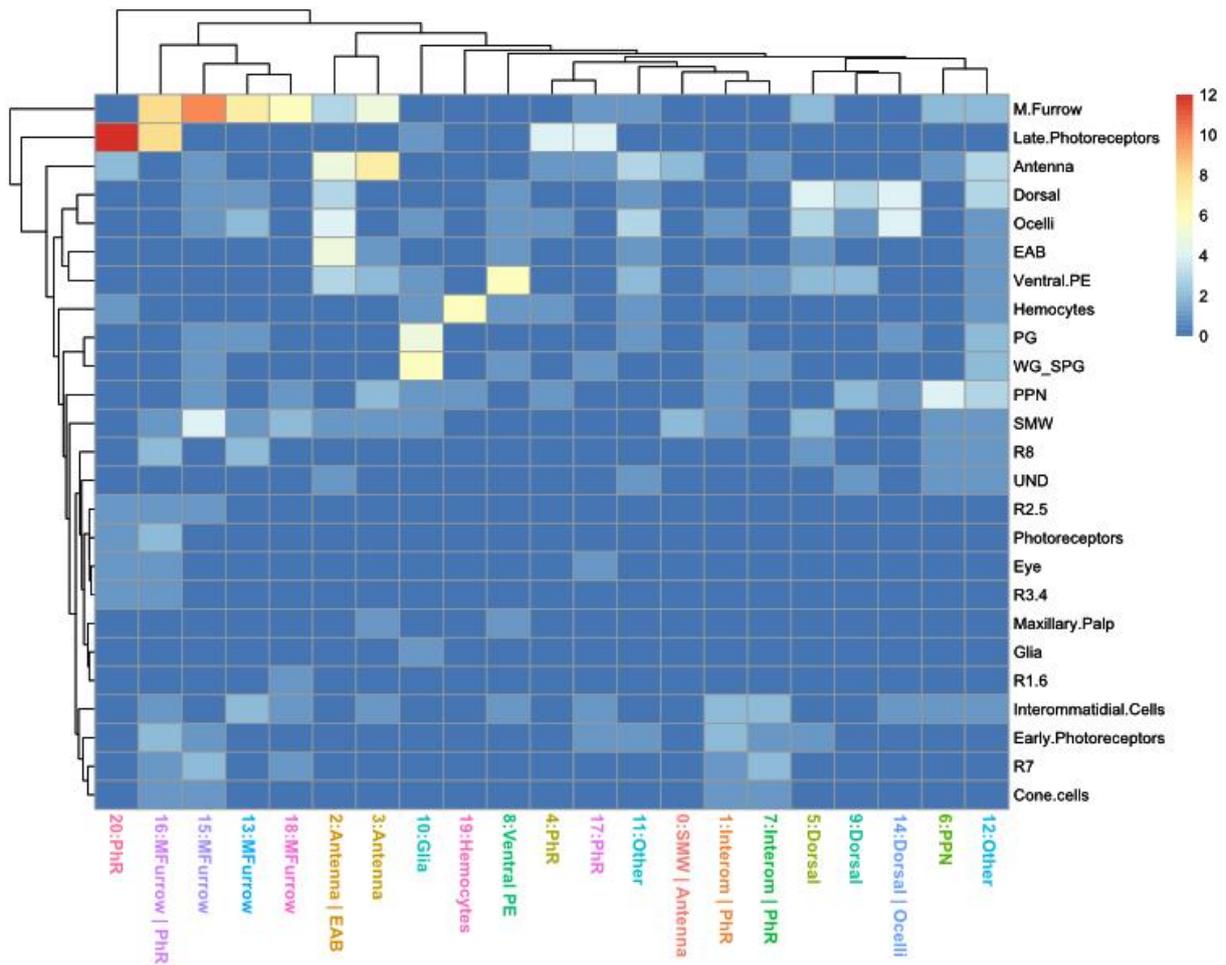


**Supplementary Table S 6: Venn-diagram of the top 3,000 variable genes in scRNAseq, snRNAseq and the overlap of both datasets.** See also Supplementary Table S 7 for a full list of genes in each category.

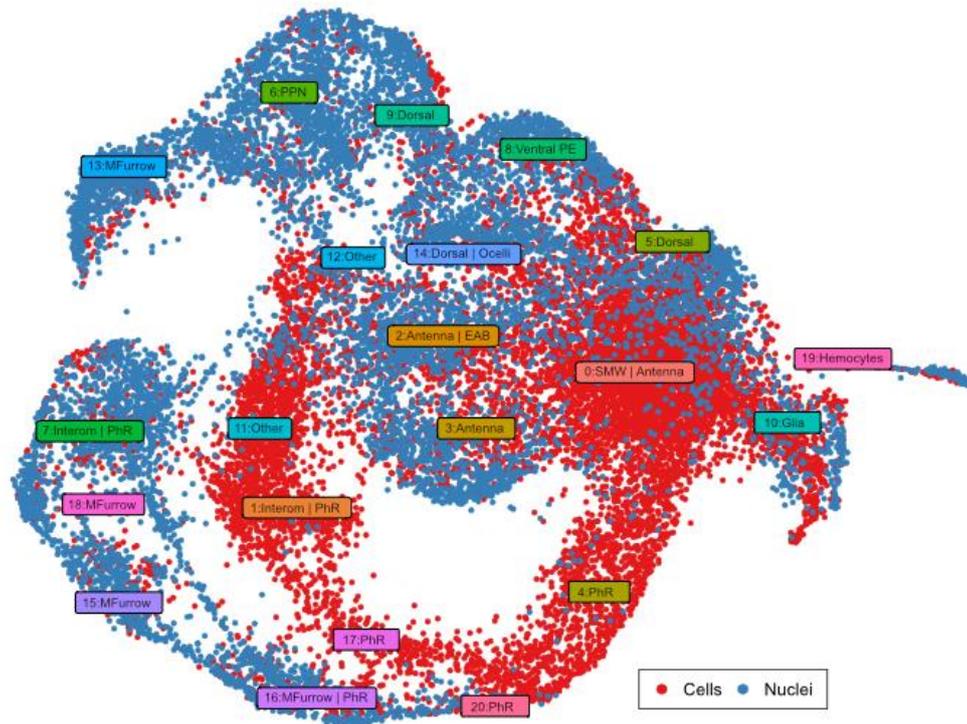


**Supplementary Table S 7: Gene ontology enrichment analysis for genes with most variable expression.** (A) Top 3,000 genes unique to scRNAseq (i.e. 1,520 genes). (B) Top 3,000 genes unique to snRNAseq (i.e.

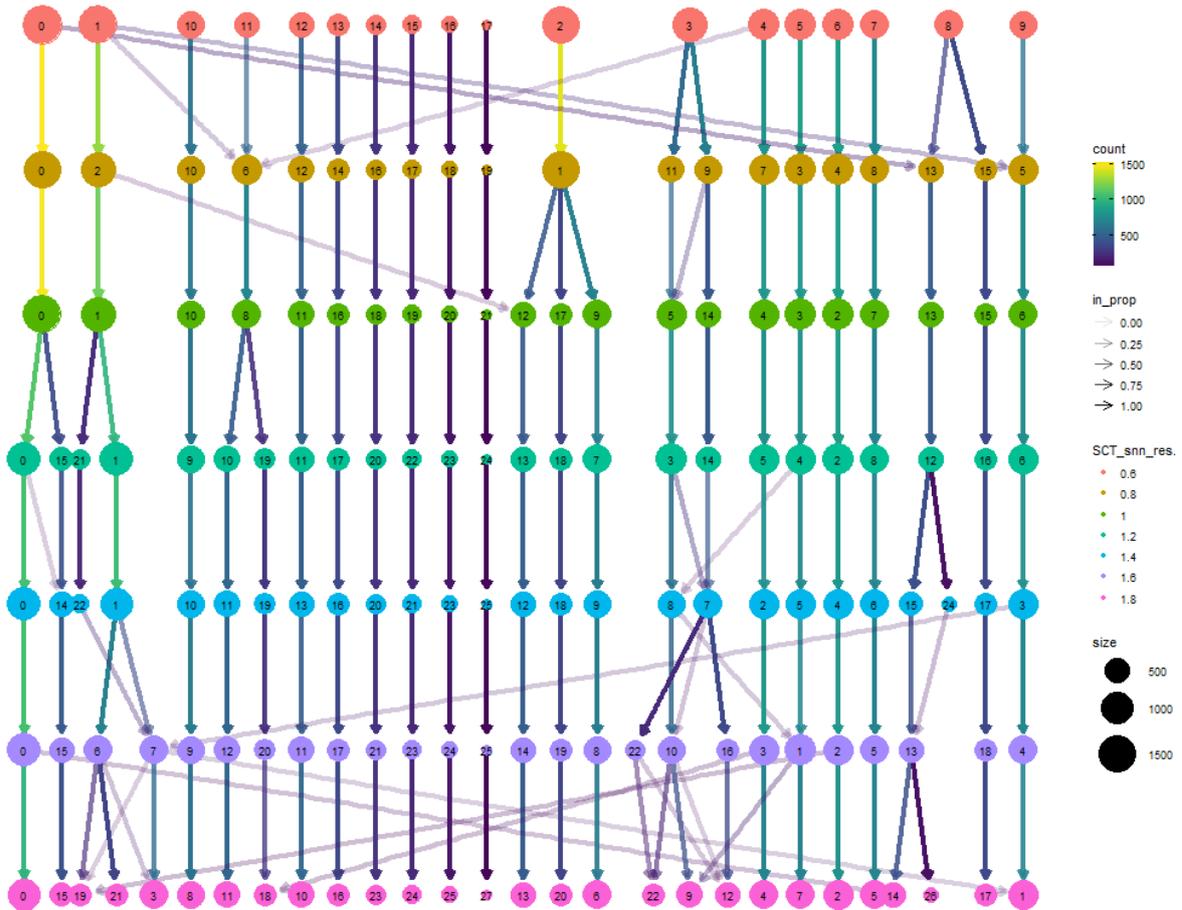
1,520 genes). (C) Top 3,000 genes shared between scRNAseq and snRNAseq (i.e. 1,480 genes). See also Supplementary Table S 7 for a full list of enriched GO terms.



**Supplementary Table S 8: Heatmap of marker gene score for cell clusters obtained by integrated analysis of scRNAseq and snRNAseq.** This heatmap shows the score for each potential cell type (Y-axis) in each cluster (X-axis). The cell types are annotated based on the highest scoring identity in the heatmap. The clusters are grouped based on their transcriptional similarity to each other. For clusters which express an equal number of marker genes for two different identities both identities were assigned. Clusters with unresolved identities (i.e. more than two equal assignments) are called “Other”. The colors of the cluster names correspond to the colors in UMAP and the bar plot in Figure 4 in the main text. The marker score is calculated using a matrix of published marker genes.



**Supplementary Table S 9: UMAP of integrated scRNAseq and snRNAseq data.** This UMAP is identical to the UMAP in Figure 4A, but the cells are labeled based on the dataset of origin (i.e. scRNAseq: red; snRNAseq: blue). The cluster numbers, names and color code are the same as in Figure 8A.



**Supplementary Table S 10: Clustering tree produced by the R package clustree for snRNAseq data of 120 h old *D. simulans* EAD.** Each level (colored dots) shows the number of clusters found using a specific resolution shown on the right (“SCT\_snn\_res”). With increasing resolution, the number of clusters increase. Arrows describe the number of cells switching identities between resolutions (“count”). A high fluctuation of cells indicates unstable clustering, meaning clusters with only few meaningful differences (see resolution 1.6 to 1.8). In this dataset, stable clustering was found at resolutions between 1 and 1.2, the latter was chosen for downstream analyses.

**A: Mauritiana, 84h**

| id | source | term_id    | term_name   | term_size | intersection_size | p_value |
|----|--------|------------|---|-----------|-------------------|---------|
| 1  | GO BP  | GO:0007218 | neuroepitaxial signaling pathway                                      | 78        | 4                 | 4.7e-03 |
| 2  | GO BP  | GO:0007188 | G protein-coupled receptor signaling pathway                          | 241       | 5                 | 2.2e-02 |
| 3  | GO BP  | GO:0007168 | receptor guanine nucleotide signaling pathway                         | 9         | 2                 | 3.4e-02 |
| 4  | GO BP  | GO:1901841 | regulation of high voltage-gated calcium channel activity             | 1         | 1                 | 7.2e-02 |
| 5  | GO BP  | GO:1901842 | negative regulation of high voltage-gated calcium channel activity    | 1         | 1                 | 7.7e-02 |
| 6  | GO BP  | GO:0042330 | taxi  | 355       | 7                 | 9.3e-02 |
| 7  | GO BP  | GO:0011934 | cGMP-mediated signaling   | 15        | 2                 | 1.8e-01 |
| 8  | GO BP  | GO:1901386 | negative regulation of voltage-gated calcium channel activity         | 3         | 1                 | 2.3e-01 |
| 9  | GO BP  | GO:0006935 | chemotaxis  | 295       | 6                 | 2.6e-01 |
| 10 | GO BP  | GO:2001258 | negative regulation of calcium channel activity                       | 4         | 1                 | 3.1e-01 |
| 11 | GO BP  | GO:1901020 | negative regulation of calcium ion transmembrane transporter activity | 4         | 1                 | 3.1e-01 |
| 12 | GO BP  | GO:1903170 | negative regulation of calcium ion transmembrane transport            | 5         | 1                 | 3.9e-01 |
| 13 | GO BP  | GO:0009038 | neuron recognition  | 117       | 4                 | 3.9e-01 |
| 14 | GO BP  | GO:0016201 | synaptic target inhibition  | 15        | 2                 | 4.1e-01 |
| 15 | GO BP  | GO:0008039 | synaptic target recognition   | 50        | 3                 | 4.3e-01 |

[gProfiler \(bit.cs.ut.ee/gprofiler\)](#)

**B: Melanogaster, 84h**

| id | source | term_id    | term_name                                     | term_size | intersection_size | p_value |
|----|--------|------------|---|-----------|-------------------|---------|
| 1  | GO BP  | GO:0004229 | epithelium development                        | 966       | 390               | 8.1e-06 |
| 2  | GO BP  | GO:0009888 | tissue development                            | 1033      | 411               | 1.3e-05 |
| 3  | GO BP  | GO:0049513 | animal organ development                      | 1258      | 507               | 1.5e-02 |
| 4  | GO BP  | GO:0071840 | cellular component organization or biogenesis | 3301      | 1934              | 8.6e-01 |
| 5  | GO BP  | GO:0009653 | anatomical structure morphogenesis            | 1584      | 254               | 1.4e-04 |
| 6  | GO BP  | GO:0002009 | morphogenesis of an epithelium                | 595       | 266               | 3.1e-07 |
| 7  | GO BP  | GO:0010467 | gene expression                               | 2596      | 841               | 6.3e-07 |
| 8  | GO BP  | GO:0048729 | tissue morphogenesis                          | 614       | 270               | 2.7e-06 |
| 9  | GO BP  | GO:0035295 | tube development                              | 719       | 300               | 7.5e-06 |
| 10 | GO BP  | GO:0009887 | animal organ morphogenesis                    | 761       | 165               | 9.9e-06 |
| 11 | GO BP  | GO:0071444 | imaginal disc development                     | 373       | 295               | 3.1e-05 |
| 12 | GO BP  | GO:0048556 | anatomical structure development              | 2937      | 254               | 3.3e-05 |
| 13 | GO BP  | GO:0032502 | developmental process                         | 3013      | 368               | 2.1e-04 |
| 14 | GO BP  | GO:0048731 | system development                            | 1534      | 242               | 2.6e-04 |
| 15 | GO BP  | GO:0007275 | multicellular organism development            | 2444      | 315               | 3.1e-01 |

[gProfiler \(bit.cs.ut.ee/gprofiler\)](#)

**C: Mauritiana, 108h**

| id | source | term_id    | term_name                                      | term_size | intersection_size | p_value |
|----|--------|------------|--|-----------|-------------------|---------|
| 1  | GO BP  | GO:0098916 | anterograde transsynaptic signaling            | 321       | 20                | 3.6e-11 |
| 2  | GO BP  | GO:0007288 | chemical synaptic transmission                 | 321       | 20                | 3.6e-11 |
| 3  | GO BP  | GO:0099537 | trans-synaptic signaling                       | 324       | 20                | 4.3e-11 |
| 4  | GO BP  | GO:0099535 | synaptic signaling                             | 332       | 20                | 6.3e-11 |
| 5  | GO BP  | GO:007287  | cell-cell signaling                            | 532       | 24                | 2.3e-09 |
| 6  | GO BP  | GO:0091777 | regulation of trans-synaptic signaling         | 150       | 12                | 3.8e-07 |
| 7  | GO BP  | GO:0050804 | modulation of chemical synaptic transmission   | 150       | 12                | 3.8e-07 |
| 8  | GO BP  | GO:0001505 | regulation of neurotransmitter levels          | 122       | 10                | 9.2e-06 |
| 9  | GO BP  | GO:0007274 | neuromuscular synaptic transmission            | 82        | 9                 | 1.5e-05 |
| 10 | GO BP  | GO:0017156 | calcium-ion regulated exocytosis               | 13        | 5                 | 1.5e-05 |
| 11 | GO BP  | GO:0017158 | regulation of calcium ion-dependent exocytosis | 9         | 4                 | 3.3e-05 |
| 12 | GO BP  | GO:0099483 | signal release from synapse                    | 104       | 9                 | 3.3e-05 |
| 13 | GO BP  | GO:0007289 | neurotransmitter secretion                     | 104       | 9                 | 3.3e-05 |
| 14 | GO BP  | GO:0043289 | regulation of ion transport                    | 92        | 9                 | 4.2e-05 |
| 15 | GO BP  | GO:0023061 | signal release                                 | 145       | 10                | 4.9e-05 |

[gProfiler \(bit.cs.ut.ee/gprofiler\)](#)

**D: Melanogaster, 108h**

| id | source | term_id    | term_name   | term_size | intersection_size | p_value |
|----|--------|------------|---|-----------|-------------------|---------|
| 1  | GO BP  | GO:0030334 | regulation of cell migration                        | 88        | 6                 | 2.5e-03 |
| 2  | GO BP  | GO:2000145 | regulation of cell motility                         | 92        | 6                 | 3.2e-03 |
| 3  | GO BP  | GO:0071526 | semaphorin-plexin signaling pathway                 | 10        | 2                 | 1.8e-02 |
| 4  | GO BP  | GO:001270  | regulation of cellular component movement           | 134       | 6                 | 1.9e-02 |
| 5  | GO BP  | GO:0040512 | regulation of locomotion                            | 124       | 6                 | 1.9e-02 |
| 6  | GO BP  | GO:0004439 | trachea morphogenesis                               | 11        | 3                 | 2.1e-02 |
| 7  | GO BP  | GO:0002803 | regulation of anatomical structure morphogenesis    | 187       | 4                 | 2.7e-02 |
| 8  | GO BP  | GO:0005541 | respiratory system development                      | 259       | 8                 | 5.2e-02 |
| 9  | GO BP  | GO:0009653 | anatomical structure morphogenesis                  | 1584      | 20                | 8.4e-02 |
| 10 | GO BP  | GO:0009605 | response to external stimulus                       | 1071      | 16                | 9.5e-02 |
| 11 | GO BP  | GO:0035217 | labial disc development                             | 4         | 2                 | 1.0e-01 |
| 12 | GO BP  | GO:0008996 | response to stimulus                                | 3201      | 31                | 1.1e-01 |
| 13 | GO BP  | GO:0031551 | regulation of tube size: open tracheal system       | 50        | 4                 | 1.3e-01 |
| 14 | GO BP  | GO:1901073 | flurazepam-containing compound biosynthetic process | 4         | 2                 | 1.5e-01 |
| 15 | GO BP  | GO:0006031 | chitin biosynthetic process                         | 4         | 2                 | 1.5e-01 |

[gProfiler \(bit.cs.ut.ee/gprofiler\)](#)

**E: Mauritiana, 120h**

| id | source | term_id    | term_name                                 | term_size | intersection_size | p_value |
|----|--------|------------|---|-----------|-------------------|---------|
| 1  | GO BP  | GO:0009887 | animal organ morphogenesis                | 781       | 41                | 1.8e-06 |
| 2  | GO BP  | GO:0048522 | positive regulation of cellular process   | 1724      | 75                | 2.4e-06 |
| 3  | GO BP  | GO:0007089 | regulation of biological process          | 4401      | 147               | 2.5e-06 |
| 4  | GO BP  | GO:0050794 | regulation of cellular process            | 4110      | 140               | 2.5e-06 |
| 5  | GO BP  | GO:0048518 | positive regulation of biological process | 1974      | 75                | 7.2e-06 |
| 6  | GO BP  | GO:0048731 | system development                        | 1534      | 60                | 1.0e-05 |
| 7  | GO BP  | GO:0048513 | animal organ development                  | 1253      | 54                | 1.4e-05 |
| 8  | GO BP  | GO:0006007 | biological regulation                     | 4887      | 155               | 3.0e-05 |
| 9  | GO BP  | GO:0009653 | anatomical structure morphogenesis        | 1584      | 61                | 9.3e-05 |
| 10 | GO BP  | GO:0007275 | multicellular organism development        | 2444      | 82                | 9.3e-05 |
| 11 | GO BP  | GO:0032502 | developmental process                     | 3013      | 95                | 9.7e-05 |
| 12 | GO BP  | GO:0030403 | regulation of organelle organization      | 354       | 21                | 1.6e-04 |
| 13 | GO BP  | GO:0035295 | tube development                          | 719       | 37                | 1.6e-04 |
| 14 | GO BP  | GO:0022009 | neurogenesis                              | 847       | 41                | 1.6e-04 |
| 15 | GO BP  | GO:0048556 | anatomical structure development          | 2937      | 90                | 2.1e-04 |

[gProfiler \(bit.cs.ut.ee/gprofiler\)](#)

**F: Melanogaster, 120h**

| id | source | term_id    | term_name   | term_size | intersection_size | p_value |
|----|--------|------------|---|-----------|-------------------|---------|
| 1  | GO BP  | GO:0002803 | regulation of anatomical structure morphogenesis  | 187       | 4                 | 3.9e-03 |
| 2  | GO BP  | GO:0071526 | semaphorin-plexin signaling pathway   | 10        | 2                 | 9.2e-03 |
| 3  | GO BP  | GO:0005770 | regulation of axonogenesis  | 51        | 2                 | 2.6e-01 |
| 4  | GO BP  | GO:0008545 | motor neuron axon guidance  | 77        | 2                 | 5.9e-01 |
| 5  | GO BP  | GO:0007169 | regulation of developmental process   | 869       | 4                 | 6.4e-01 |
| 6  | GO BP  | GO:0043553 | mitochondrial fragmentation involved in apoptotic process                                 | 2         | 1                 | 6.4e-01 |
| 7  | GO BP  | GO:2000208 | regulation of multicellular organismal development  | 281       | 3                 | 6.5e-01 |
| 8  | GO BP  | GO:0040002 | antigen processing and presentation of peptide antigen                                    | 1         | 1                 | 7.2e-01 |
| 9  | GO BP  | GO:0002495 | antigen processing and presentation of peptide antigen via MHC class II                   | 1         | 1                 | 7.2e-01 |
| 10 | GO BP  | GO:0002504 | antigen processing and presentation of peptide or polysaccharide antigen via MHC class II | 1         | 1                 | 7.2e-01 |
| 11 | GO BP  | GO:0051128 | regulation of cellular component organization   | 713       | 4                 | 7.3e-01 |
| 12 | GO BP  | GO:0003334 | regulation of cell migration  | 88        | 2                 | 7.7e-01 |
| 13 | GO BP  | GO:0008360 | regulation of cell shape  | 66        | 2                 | 7.9e-01 |
| 14 | GO BP  | GO:1902285 | semaphorin-plexin signaling pathway involved in neuron projection guidance                | 2         | 1                 | 8.0e-01 |
| 15 | GO BP  | GO:1902287 | semaphorin-plexin signaling pathway involved in axon guidance                             | 2         | 1                 | 8.0e-01 |

[gProfiler \(bit.cs.ut.ee/gprofiler\)](#)

**Supplementary Table S 11: GO-Term enrichment analysis within PhR cells of DEGs between *D. melanogaster* and *D. mauritiana*.** This figure depicts up to the top 15 enriched GO-terms by p-value. Species comparisons were performed pairwise by timepoint.

**A: Mauritiana, 84h**

| id | source | term_id    | term_name   | term_size | intersection_size | p_value |
|----|--------|------------|---|-----------|-------------------|---------|
| 1  | GO:BP  | GO:000642  | glutamine biosynthetic process                          | 2         | 1                 | 2.5e-01 |
| 2  | GO:BP  | GO:0045956 | positive regulation of calcium ion-dependent exocytosis | 1         | 1                 | 2.5e-01 |
| 3  | GO:BP  | GO:0045213 | neurotransmitter receptor metabolic process             | 3         | 1                 | 3.7e-01 |
| 4  | GO:BP  | GO:1903307 | positive regulation of regulated secretory pathway      | 2         | 1                 | 4.9e-01 |
| 5  | GO:BP  | GO:1909504 | dense core granule exocytosis                           | 2         | 1                 | 4.9e-01 |
| 6  | GO:BP  | GO:0006538 | glutamate catabolic process                             | 5         | 1                 | 6.2e-01 |
| 7  | GO:BP  | GO:0043410 | positive regulation of MAPK cascade                     | 71        | 2                 | 7.1e-01 |
| 8  | GO:BP  | GO:0043649 | dicarboxylic acid catabolic process                     | 8         | 1                 | 9.9e-01 |
| 9  | GO:BP  | GO:0048149 | behavioral response to ethanol                          | 55        | 1                 | 1.0e+00 |
| 10 | GO:BP  | GO:0046700 | heterocycle catabolic process                           | 200       | 1                 | 1.0e+00 |
| 11 | GO:BP  | GO:0046986 | cell development  | 1453      | 2                 | 1.0e+00 |
| 12 | GO:BP  | GO:0046958 | neuroscience learning                                   | 15        | 1                 | 1.0e+00 |
| 13 | GO:BP  | GO:0046903 | secretion   | 254       | 1                 | 1.0e+00 |
| 14 | GO:BP  | GO:0048513 | animal organ development                                | 1253      | 1                 | 1.0e+00 |
| 15 | GO:BP  | GO:0048515 | spermatid differentiation                               | 132       | 1                 | 1.0e+00 |

gProfiler (ht.cx.cs.ut.ee.gprofiler)

**B: Melanogaster, 84h**

| id | source | term_id    | term_name                                    | term_size | intersection_size | p_value |
|----|--------|------------|--|-----------|-------------------|---------|
| 1  | GO:BP  | GO:0021811 | cytoplasmic translation                      | 150       | 42                | 1.7e-27 |
| 2  | GO:BP  | GO:0006412 | translation                                  | 474       | 35                | 3.7e-14 |
| 3  | GO:BP  | GO:0049343 | peptide biosynthetic process                 | 322       | 35                | 2.9e-13 |
| 4  | GO:BP  | GO:0006518 | peptide metabolic process                    | 624       | 35                | 2.2e-12 |
| 5  | GO:BP  | GO:0043604 | amide biosynthetic process                   | 573       | 36                | 5.3e-12 |
| 6  | GO:BP  | GO:0034545 | cellular macromolecule biosynthetic process  | 691       | 39                | 1.1e-11 |
| 7  | GO:BP  | GO:0043603 | cellular amide metabolic process             | 721       | 60                | 6.9e-11 |
| 8  | GO:BP  | GO:0044495 | cellular component biogenesis                | 1703      | 155               | 9.1e-11 |
| 9  | GO:BP  | GO:0048522 | positive regulation of cellular process      | 1724      | 151               | 1.2e-09 |
| 10 | GO:BP  | GO:0022907 | cellular component assembly                  | 1509      | 137               | 5.2e-09 |
| 11 | GO:BP  | GO:0048516 | positive regulation of biological process    | 1674      | 157               | 1.3e-08 |
| 12 | GO:BP  | GO:0007275 | multicellular organism development           | 2444      | 192               | 2.3e-08 |
| 13 | GO:BP  | GO:0044287 | cellular protein metabolic process           | 2982      | 181               | 2.8e-08 |
| 14 | GO:BP  | GO:0048731 | system development                           | 1534      | 134               | 3.4e-08 |
| 15 | GO:BP  | GO:1901566 | organonitrogen compound biosynthetic process | 1067      | 43                | 3.6e-08 |

gProfiler (ht.cx.cs.ut.ee.gprofiler)

**C: Mauritiana, 108h**

| id | source | term_id    | term_name  | term_size | intersection_size | p_value |
|----|--------|------------|--|-----------|-------------------|---------|
| 1  | GO:BP  | GO:0091163 | establishment of epithelial cell planar polarity | 5         | 2                 | 1.4e-02 |
| 2  | GO:BP  | GO:0091162 | establishment of epithelial cell polarity        | 14        | 2                 | 1.3e-01 |
| 3  | GO:BP  | GO:0060429 | epithelium development                           | 966       | 7                 | 1.9e-01 |
| 4  | GO:BP  | GO:0035233 | enriched lateral tubule pattern formation        | 19        | 2                 | 2.4e-01 |
| 5  | GO:BP  | GO:0026988 | tissue development                               | 1053      | 7                 | 3.2e-01 |
| 6  | GO:BP  | GO:0064689 | negative regulation of protein kinase activity   | 24        | 2                 | 3.8e-01 |
| 7  | GO:BP  | GO:0033673 | negative regulation of kinase activity           | 25        | 2                 | 4.1e-01 |
| 8  | GO:BP  | GO:0051348 | negative regulation of transferase activity      | 28        | 2                 | 5.2e-01 |
| 9  | GO:BP  | GO:0045496 | male analla development                          | 2         | 1                 | 5.5e-01 |
| 10 | GO:BP  | GO:0045497 | female analla development                        | 2         | 1                 | 5.5e-01 |
| 11 | GO:BP  | GO:007444  | imaginal disc development                        | 573       | 5                 | 6.1e-01 |
| 12 | GO:BP  | GO:0060241 | regulation of cellular localization              | 132       | 3                 | 6.9e-01 |
| 13 | GO:BP  | GO:0098107 | stade morphogenesis                              | 42        | 2                 | 7.7e-01 |
| 14 | GO:BP  | GO:0035017 | cuticle pattern formation                        | 34        | 2                 | 7.7e-01 |
| 15 | GO:BP  | GO:0032879 | regulation of localization                       | 569       | 5                 | 8.0e-01 |

gProfiler (ht.cx.cs.ut.ee.gprofiler)

**D: Melanogaster, 108h**

| id | source | term_id    | term_name   | term_size | intersection_size | p_value |
|----|--------|------------|---|-----------|-------------------|---------|
| 1  | GO:BP  | GO:0022603 | regulation of anatomical structure morphogenesis  | 187       | 4                 | 3.3e-02 |
| 2  | GO:BP  | GO:0071526 | semaphorin-plexin signaling pathway   | 10        | 2                 | 4.2e-02 |
| 3  | GO:BP  | GO:0018382 | synaptic vesicle priming  | 16        | 2                 | 3.3e-01 |
| 4  | GO:BP  | GO:0018380 | neurotrophin inactivation   | 6         | 1                 | 3.5e-01 |
| 5  | GO:BP  | GO:0023252 | signaling   | 2042      | 11                | 3.8e-01 |
| 6  | GO:BP  | GO:0048002 | antigen processing and presentation of peptide antigen                                    | 1         | 1                 | 4.3e-01 |
| 7  | GO:BP  | GO:0025204 | antigen processing and presentation of peptide or polysaccharide antigen via MHC class II | 1         | 1                 | 4.3e-01 |
| 8  | GO:BP  | GO:002495  | antigen processing and presentation of peptide antigen via MHC class I                    | 1         | 1                 | 4.3e-01 |
| 9  | GO:BP  | GO:0071165 | signal transduction   | 1813      | 10                | 4.3e-01 |
| 10 | GO:BP  | GO:0071154 | cell communication  | 2097      | 11                | 4.6e-01 |
| 11 | GO:BP  | GO:0008345 | motor neuron axon guidance  | 77        | 3                 | 4.7e-01 |
| 12 | GO:BP  | GO:0018381 | synaptic vesicle docking  | 21        | 2                 | 5.7e-01 |
| 13 | GO:BP  | GO:0018382 | antigen processing and presentation   | 2         | 1                 | 6.5e-01 |
| 14 | GO:BP  | GO:0090969 | synaptic vesicle tethering involved in synaptic-vesicle exocytosis                        | 1         | 1                 | 6.9e-01 |
| 15 | GO:BP  | GO:0099011 | neuronal dense core vesicle exocytosis  | 1         | 1                 | 6.9e-01 |

gProfiler (ht.cx.cs.ut.ee.gprofiler)

**E: Mauritiana, 120h**

| id | source | term_id    | term_name                          | term_size | intersection_size | p_value |
|----|--------|------------|------------------------------------|-----------|-------------------|---------|
| 1  | GO:BP  | GO:0048731 | system development                 | 1534      | 578               | 3.8e-91 |
| 2  | GO:BP  | GO:0048513 | animal organ development           | 1253      | 504               | 7.8e-89 |
| 3  | GO:BP  | GO:0009653 | anatomical structure morphogenesis | 1584      | 584               | 8.5e-87 |
| 4  | GO:BP  | GO:0048856 | anatomical structure development   | 2837      | 863               | 4.1e-86 |
| 5  | GO:BP  | GO:0007275 | multicellular organism development | 2444      | 776               | 7.8e-85 |
| 6  | GO:BP  | GO:0032502 | developmental process              | 3013      | 894               | 1.3e-83 |
| 7  | GO:BP  | GO:0030154 | cell differentiation               | 1790      | 613               | 3.8e-77 |
| 8  | GO:BP  | GO:0048869 | cellular developmental process     | 1799      | 615               | 4.0e-77 |
| 9  | GO:BP  | GO:0009887 | animal organ morphogenesis         | 761       | 345               | 8.3e-74 |
| 10 | GO:BP  | GO:0007389 | nervous system development         | 1170      | 453               | 1.0e-72 |
| 11 | GO:BP  | GO:0048468 | cell development                   | 1453      | 516               | 9.7e-69 |
| 12 | GO:BP  | GO:0022008 | neurogenesis                       | 947       | 386               | 7.7e-68 |
| 13 | GO:BP  | GO:0048999 | generation of neurons              | 883       | 367               | 3.0e-66 |
| 14 | GO:BP  | GO:0060429 | epithelium development             | 966       | 388               | 1.7e-65 |
| 15 | GO:BP  | GO:0050789 | regulation of biological process   | 4401      | 1124              | 1.7e-65 |

gProfiler (ht.cx.cs.ut.ee.gprofiler)

**F: Melanogaster, 120h**

| id | source | term_id    | term_name                                     | term_size | intersection_size | p_value |
|----|--------|------------|---|-----------|-------------------|---------|
| 1  | GO:BP  | GO:0009888 | tissue development                            | 1053      | 79                | 1.3e-10 |
| 2  | GO:BP  | GO:0030154 | cell differentiation                          | 1790      | 116               | 1.3e-10 |
| 3  | GO:BP  | GO:0071840 | cellular component organization or biogenesis | 3301      | 194               | 1.5e-10 |
| 4  | GO:BP  | GO:0048869 | cellular developmental process                | 1799      | 116               | 1.9e-10 |
| 5  | GO:BP  | GO:0048868 | cell development                              | 1453      | 89                | 6.3e-10 |
| 6  | GO:BP  | GO:0015043 | cellular component organization               | 3120      | 181               | 7.3e-09 |
| 7  | GO:BP  | GO:0032502 | developmental process                         | 3013      | 176               | 1.2e-08 |
| 8  | GO:BP  | GO:0048522 | positive regulation of cellular process       | 1724      | 115               | 6.3e-08 |
| 9  | GO:BP  | GO:0004529 | epithelium development                        | 966       | 69                | 6.6e-08 |
| 10 | GO:BP  | GO:0048518 | positive regulation of biological process     | 1674      | 121               | 1.4e-07 |
| 11 | GO:BP  | GO:0048513 | animal organ development                      | 1253      | 84                | 2.4e-07 |
| 12 | GO:BP  | GO:0048856 | anatomical structure development              | 2837      | 149               | 4.4e-07 |
| 13 | GO:BP  | GO:0009887 | cellular process                              | 8771      | 306               | 7.2e-07 |
| 14 | GO:BP  | GO:0007444 | imaginal disc development                     | 573       | 46                | 6.3e-07 |
| 15 | GO:BP  | GO:0009653 | anatomical structure morphogenesis            | 1584      | 96                | 6.9e-07 |

gProfiler (ht.cx.cs.ut.ee.gprofiler)

**Supplementary Table S 12: Enriched GO-terms in DEGs within MF cells between *D. melanogaster* and *D. mauritiana*.** This figure depicts up to the top 15 enriched GO-terms by p-value. Species comparisons were performed pairwise by timepoint.

**A: Mauritiana, 84h**

| id | source | term_id    | term_name   | term_size | intersection_size | p_value |
|----|--------|------------|---|-----------|-------------------|---------|
| 1  | GO:BP  | GO:007268  | chemical synaptic transmission                          | 321       | 6                 | 1.1e-01 |
| 2  | GO:BP  | GO:0098916 | anterograde trans-synaptic signaling                    | 321       | 6                 | 1.1e-01 |
| 3  | GO:BP  | GO:0099537 | trans-synaptic signaling                                | 324       | 6                 | 1.2e-01 |
| 4  | GO:BP  | GO:0099536 | synaptic signaling                                      | 532       | 6                 | 1.4e-01 |
| 5  | GO:BP  | GO:0051046 | regulation of secretion                                 | 104       | 3                 | 1.5e-01 |
| 6  | GO:BP  | GO:0049956 | positive regulation of calcium ion-dependent exocytosis | 1         | 1                 | 2.2e-01 |
| 7  | GO:BP  | GO:0062334 | nucleosome assembly                                     | 130       | 3                 | 2.3e-01 |
| 8  | GO:BP  | GO:0085642 | glutamine biosynthetic process                          | 2         | 1                 | 2.9e-01 |
| 9  | GO:BP  | GO:0055075 | potassium ion homeostasis                               | 13        | 2                 | 4.1e-01 |
| 10 | GO:BP  | GO:0045213 | neurotransmitter receptor metabolic process             | 3         | 1                 | 4.3e-01 |
| 11 | GO:BP  | GO:1990504 | dense core granule exocytosis                           | 2         | 1                 | 4.3e-01 |
| 12 | GO:BP  | GO:1903307 | positive regulation of regulated secretory pathway      | 2         | 1                 | 4.3e-01 |
| 13 | GO:BP  | GO:0010715 | regulation of extracellular matrix disassembly          | 1         | 1                 | 4.3e-01 |
| 14 | GO:BP  | GO:0010716 | negative regulation of extracellular matrix disassembly | 1         | 1                 | 4.3e-01 |
| 15 | GO:BP  | GO:0034728 | rucleosome organization                                 | 166       | 3                 | 4.7e-01 |

[g:Profiler \(bit.cs.ut.ee/goprofiler\)](http://bit.cs.ut.ee/goprofiler)

**B: Melanogaster, 84h**

| id | source | term_id    | term_name   | term_size | intersection_size | p_value |
|----|--------|------------|---|-----------|-------------------|---------|
| 1  | GO:BP  | GO:007424  | open tracheal system development                      | 250       | 26                | 3.6e-08 |
| 2  | GO:BP  | GO:0060541 | respiratory system development                        | 259       | 26                | 7.9e-08 |
| 3  | GO:BP  | GO:0035152 | regulation of tube architecture, open tracheal system | 92        | 15                | 1.2e-06 |
| 4  | GO:BP  | GO:0042116 | cell-cell junction organization                       | 82        | 12                | 4.0e-05 |
| 5  | GO:BP  | GO:0035151 | regulation of tube size, open tracheal system         | 90        | 9                 | 3.4e-04 |
| 6  | GO:BP  | GO:0071526 | semaphorinplexin signaling pathway                    | 10        | 3                 | 4.5e-04 |
| 7  | GO:BP  | GO:0060936 | regulation of anatomical structure size               | 244       | 11                | 6.9e-04 |
| 8  | GO:BP  | GO:0035150 | regulation of tube size                               | 54        | 9                 | 6.9e-04 |
| 9  | GO:BP  | GO:0069888 | tissue development                                    | 1053      | 61                | 8.5e-04 |
| 10 | GO:BP  | GO:0050996 | response to stimulus                                  | 3201      | 76                | 1.2e-03 |
| 11 | GO:BP  | GO:0034332 | adherens junction organization                        | 35        | 7                 | 1.3e-03 |
| 12 | GO:BP  | GO:0016043 | cellular component organization                       | 3120      | 81                | 1.8e-03 |
| 13 | GO:BP  | GO:0048731 | system development                                    | 1534      | 49                | 1.9e-03 |
| 14 | GO:BP  | GO:0007155 | cell adhesion   | 242       | 16                | 2.2e-03 |
| 15 | GO:BP  | GO:0035159 | regulation of tube length, open tracheal system       | 36        | 7                 | 4.2e-03 |

[g:Profiler \(bit.cs.ut.ee/goprofiler\)](http://bit.cs.ut.ee/goprofiler)

**C: Mauritiana, 108h**

| id | source | term_id    | term_name   | term_size | intersection_size | p_value |
|----|--------|------------|---|-----------|-------------------|---------|
| 1  | GO:BP  | GO:0050794 | regulation of cellular process                          | 4110      | 90                | 6.0e-08 |
| 2  | GO:BP  | GO:0048699 | generation of neurons                                   | 883       | 31                | 1.9e-07 |
| 3  | GO:BP  | GO:0007399 | nervous system development                              | 1170      | 41                | 3.3e-07 |
| 4  | GO:BP  | GO:0048731 | system development                                      | 1534      | 47                | 1.0e-06 |
| 5  | GO:BP  | GO:0022008 | neurogenesis  | 947       | 31                | 1.1e-06 |
| 6  | GO:BP  | GO:0050789 | regulation of biological process                        | 4401      | 91                | 1.3e-06 |
| 7  | GO:BP  | GO:0030182 | neuron differentiation                                  | 810       | 30                | 2.2e-06 |
| 8  | GO:BP  | GO:0065007 | biological regulation                                   | 4887      | 93                | 1.1e-05 |
| 9  | GO:BP  | GO:0009802 | cell morphogenesis                                      | 574       | 22                | 2.9e-05 |
| 10 | GO:BP  | GO:0048689 | cellular developmental process                          | 1799      | 41                | 5.4e-05 |
| 11 | GO:BP  | GO:0051240 | positive regulation of multicellular organismal process | 202       | 14                | 5.6e-05 |
| 12 | GO:BP  | GO:0097485 | neuron projection guidance                              | 281       | 16                | 6.7e-05 |
| 13 | GO:BP  | GO:0048523 | negative regulation of cellular process                 | 1415      | 39                | 7.7e-05 |
| 14 | GO:BP  | GO:0040011 | locomotion  | 704       | 24                | 1.1e-04 |
| 15 | GO:BP  | GO:0006835 | chemotaxis  | 295       | 16                | 1.3e-04 |

[g:Profiler \(bit.cs.ut.ee/goprofiler\)](http://bit.cs.ut.ee/goprofiler)

**D: Melanogaster, 108h**

| id | source | term_id    | term_name  | term_size | intersection_size | p_value |
|----|--------|------------|--|-----------|-------------------|---------|
| 1  | GO:BP  | GO:0028003 | regulation of anatomical structure morphogenesis | 187       | 4                 | 1.1e-02 |
| 2  | GO:BP  | GO:0071526 | semaphorinplexin signaling pathway               | 10        | 2                 | 2.7e-02 |
| 3  | GO:BP  | GO:0051128 | regulation of cellular component organization    | 713       | 5                 | 1.4e-01 |
| 4  | GO:BP  | GO:0016060 | metarhodopsin inactivation                       | 5         | 1                 | 3.9e-01 |
| 5  | GO:BP  | GO:0050770 | regulation of axonogenesis                       | 51        | 2                 | 7.4e-01 |
| 6  | GO:BP  | GO:0048285 | organelle fission                                | 340       | 1                 | 1.0e+00 |
| 7  | GO:BP  | GO:0048383 | mesectoderm development                          | 6         | 1                 | 1.0e+00 |
| 8  | GO:BP  | GO:0048468 | cell development                                 | 1453      | 3                 | 1.0e+00 |
| 9  | GO:BP  | GO:0048489 | synaptic vesicle transport                       | 24        | 1                 | 1.0e+00 |
| 10 | GO:BP  | GO:0048499 | synaptic vesicle membrane organization           | 14        | 1                 | 1.0e+00 |
| 11 | GO:BP  | GO:0048511 | rhythmic process                                 | 224       | 1                 | 1.0e+00 |
| 12 | GO:BP  | GO:0048512 | circadian behavior                               | 112       | 1                 | 1.0e+00 |
| 13 | GO:BP  | GO:0048513 | animal organ development                         | 1253      | 1                 | 1.0e+00 |
| 14 | GO:BP  | GO:0048518 | positive regulation of biological process        | 1874      | 6                 | 1.0e+00 |
| 15 | GO:BP  | GO:0048519 | negative regulation of biological process        | 1659      | 5                 | 1.0e+00 |

[g:Profiler \(bit.cs.ut.ee/goprofiler\)](http://bit.cs.ut.ee/goprofiler)

**E: Mauritiana, 120h**

| id | source | term_id    | term_name                           | term_size | intersection_size | p_value |
|----|--------|------------|-------------------------------------|-----------|-------------------|---------|
| 1  | GO:BP  | GO:009887  | animal organ morphogenesis          | 761       | 40                | 6.7e-09 |
| 2  | GO:BP  | GO:0048731 | system development                  | 1534      | 57                | 1.3e-07 |
| 3  | GO:BP  | GO:0022008 | neurogenesis                        | 947       | 42                | 4.0e-07 |
| 4  | GO:BP  | GO:0048859 | formation of anatomical boundary    | 53        | 11                | 4.0e-07 |
| 5  | GO:BP  | GO:009853  | anatomical structure morphogenesis  | 1584      | 57                | 4.7e-07 |
| 6  | GO:BP  | GO:0048645 | animal organ formation              | 54        | 11                | 5.0e-07 |
| 7  | GO:BP  | GO:0048513 | animal organ development            | 1253      | 49                | 7.5e-07 |
| 8  | GO:BP  | GO:0048689 | cellular developmental process      | 1799      | 61                | 9.7e-07 |
| 9  | GO:BP  | GO:0010160 | formation of animal organ boundary  | 44        | 10                | 9.9e-07 |
| 10 | GO:BP  | GO:0050793 | regulation of developmental process | 689       | 34                | 1.7e-06 |
| 11 | GO:BP  | GO:0048699 | generation of neurons               | 883       | 39                | 2.2e-06 |
| 12 | GO:BP  | GO:0030154 | cell differentiation                | 1790      | 60                | 2.2e-06 |
| 13 | GO:BP  | GO:0030182 | neuron differentiation              | 810       | 37                | 2.5e-06 |
| 14 | GO:BP  | GO:0048729 | tissue morphogenesis                | 614       | 31                | 5.9e-06 |
| 15 | GO:BP  | GO:0032302 | developmental process               | 3013      | 83                | 6.5e-06 |

[g:Profiler \(bit.cs.ut.ee/goprofiler\)](http://bit.cs.ut.ee/goprofiler)

**F: Melanogaster, 120h**

| id | source | term_id    | term_name  | term_size | intersection_size | p_value |
|----|--------|------------|--|-----------|-------------------|---------|
| 1  | GO:BP  | GO:0028003 | regulation of anatomical structure morphogenesis                   | 187       | 4                 | 2.6e-02 |
| 2  | GO:BP  | GO:0071526 | semaphorinplexin signaling pathway                                 | 10        | 2                 | 4.0e-02 |
| 3  | GO:BP  | GO:0016060 | metarhodopsin inactivation   | 5         | 1                 | 4.0e-01 |
| 4  | GO:BP  | GO:0016082 | synaptic vesicle priming   | 18        | 2                 | 5.2e-01 |
| 5  | GO:BP  | GO:0098525 | presynaptic dense core vesicle exocytosis                          | 1         | 1                 | 5.6e-01 |
| 6  | GO:BP  | GO:0096068 | synaptic vesicle tethering involved in synaptic vesicle exocytosis | 1         | 1                 | 5.6e-01 |
| 7  | GO:BP  | GO:0090111 | neuronal dense core vesicle exocytosis                             | 1         | 1                 | 5.6e-01 |
| 8  | GO:BP  | GO:0061789 | dense core granule priming   | 1         | 1                 | 5.6e-01 |
| 9  | GO:BP  | GO:0051128 | regulation of cellular component organization                      | 713       | 5                 | 6.5e-01 |
| 10 | GO:BP  | GO:0016081 | synaptic vesicle docking   | 21        | 2                 | 9.1e-01 |
| 11 | GO:BP  | GO:0043653 | mitochondrial fragmentation involved in apoptotic process          | 2         | 1                 | 9.6e-01 |
| 12 | GO:BP  | GO:0002029 | protein polyubiquitination   | 101       | 1                 | 1.0e+00 |
| 13 | GO:BP  | GO:0048285 | organelle fission  | 340       | 1                 | 1.0e+00 |
| 14 | GO:BP  | GO:0048489 | synaptic vesicle transport   | 24        | 1                 | 1.0e+00 |
| 15 | GO:BP  | GO:0048468 | cell development   | 1453      | 3                 | 1.0e+00 |

[g:Profiler \(bit.cs.ut.ee/goprofiler\)](http://bit.cs.ut.ee/goprofiler)

**Supplementary Table S 13: Enriched GO-terms in DEGs within PPN cells between *D. melanogaster* and *D. mauritiana*.** This figure depicts up to the top 15 enriched GO-terms by p-value. Species comparisons were performed pairwise by timepoint.

**A: Mauritiana, 84h**

| id | source | term_id    | term_name                            | term_size | intersection_size | p_value |
|----|--------|------------|--------------------------------------|-----------|-------------------|---------|
| 1  | GO:BP  | GO:009537  | trans-synaptic signaling             | 324       | 105               | 7.2e-07 |
| 2  | GO:BP  | GO:0098916 | anterograde trans-synaptic signaling | 321       | 104               | 2.3e-06 |
| 3  | GO:BP  | GO:007288  | chemical synaptic transmission       | 321       | 104               | 2.3e-06 |
| 4  | GO:BP  | GO:009536  | synaptic signaling                   | 332       | 105               | 1.1e-05 |
| 5  | GO:BP  | GO:007267  | cell-cell signaling                  | 532       | 124               | 5.0e-00 |
| 6  | GO:BP  | GO:007154  | cell communication                   | 2087      | 356               | 2.9e-48 |
| 7  | GO:BP  | GO:002302  | signaling                            | 2342      | 350               | 1.0e-47 |
| 8  | GO:BP  | GO:0050808 | synapse organization                 | 310       | 110               | 2.2e-41 |
| 9  | GO:BP  | GO:0034330 | cell junction organization           | 381       | 121               | 4.7e-40 |
| 10 | GO:BP  | GO:007399  | nervous system development           | 1170      | 202               | 1.2e-36 |
| 11 | GO:BP  | GO:0046731 | system development                   | 1534      | 252               | 3.1e-34 |
| 12 | GO:BP  | GO:007010  | behavior                             | 586       | 136               | 1.9e-33 |
| 13 | GO:BP  | GO:0085007 | biological regulation                | 4887      | 546               | 2.9e-33 |
| 14 | GO:BP  | GO:0065008 | regulation of biological quality     | 1383      | 240               | 6.1e-33 |
| 15 | GO:BP  | GO:0042330 | axis                                 | 355       | 100               | 5.1e-29 |

g:Profiler (bit.cs.ut.ee/gprofiler)

**C: Mauritiana, 108h**

| id | source | term_id    | term_name                        | term_size | intersection_size | p_value |
|----|--------|------------|----------------------------------|-----------|-------------------|---------|
| 1  | GO:BP  | GO:0048731 | system development               | 1534      | 22                | 2.0e-03 |
| 2  | GO:BP  | GO:0097305 | response to alcohol              | 121       | 6                 | 3.8e-03 |
| 3  | GO:BP  | GO:0050808 | synapse organization             | 310       | 8                 | 6.9e-03 |
| 4  | GO:BP  | GO:0048686 | neuron development               | 634       | 13                | 1.1e-02 |
| 5  | GO:BP  | GO:0048686 | anatomical structure development | 2037      | 20                | 2.7e-02 |
| 6  | GO:BP  | GO:0042221 | response to chemical             | 1322      | 10                | 3.2e-02 |
| 7  | GO:BP  | GO:007399  | nervous system development       | 1170      | 17                | 3.7e-02 |
| 8  | GO:BP  | GO:0034330 | cell junction organization       | 381       | 8                 | 5.9e-02 |

g:Profiler (bit.cs.ut.ee/gprofiler)

**E: Mauritiana, 120h**

| id | source | term_id    | term_name   | term_size | intersection_size | p_value |
|----|--------|------------|---|-----------|-------------------|---------|
| 1  | GO:BP  | GO:0065008 | regulation of biological quality                                    | 1383      | 6                 | 1.3e-03 |
| 2  | GO:BP  | GO:0032528 | microvillus organization  | 7         | 2                 | 2.0e-03 |
| 3  | GO:BP  | GO:0120036 | plasma membrane bounded cell projection organization                | 714       | 4                 | 1.5e-02 |
| 4  | GO:BP  | GO:0002251 | organ or tissue specific immune response                            | 12        | 2                 | 1.6e-02 |
| 5  | GO:BP  | GO:0002385 | mucosal immune response   | 12        | 2                 | 1.6e-02 |
| 6  | GO:BP  | GO:0030030 | cell projection organization  | 729       | 4                 | 1.6e-02 |
| 7  | GO:BP  | GO:0007281 | germ cell development   | 754       | 4                 | 1.6e-02 |
| 8  | GO:BP  | GO:0030100 | regulation of endocytosis   | 62        | 3                 | 3.2e-02 |
| 9  | GO:BP  | GO:0022412 | cellular process involved in reproduction in multicellular organism | 869       | 4                 | 3.5e-02 |
| 10 | GO:BP  | GO:0003006 | developmental process involved in reproduction                      | 937       | 4                 | 4.4e-02 |
| 11 | GO:BP  | GO:0002501 | multicellular organismal process                                    | 3830      | 9                 | 4.6e-02 |
| 12 | GO:BP  | GO:0007276 | gamete generation   | 957       | 4                 | 4.9e-02 |

g:Profiler (bit.cs.ut.ee/gprofiler)

**B: Simulans, 84h**

| id | source | term_id    | term_name                          | term_size | intersection_size | p_value |
|----|--------|------------|------------------------------------|-----------|-------------------|---------|
| 1  | GO:BP  | GO:0009653 | anatomical structure morphogenesis | 1584      | 76                | 1.9e-22 |
| 2  | GO:BP  | GO:0048729 | tissue morphogenesis               | 614       | 48                | 6.4e-21 |
| 3  | GO:BP  | GO:0009688 | tissue development                 | 1033      | 61                | 6.3e-21 |
| 4  | GO:BP  | GO:0020209 | morphogenesis of an epithelium     | 595       | 48                | 8.9e-20 |
| 5  | GO:BP  | GO:0004028 | epithelium development             | 966       | 55                | 6.4e-18 |
| 6  | GO:BP  | GO:0035295 | tube development                   | 719       | 48                | 2.3e-16 |
| 7  | GO:BP  | GO:0048513 | animal organ development           | 1283      | 60                | 2.8e-16 |
| 8  | GO:BP  | GO:0009887 | animal organ morphogenesis         | 761       | 47                | 3.5e-16 |
| 9  | GO:BP  | GO:0035239 | tube morphogenesis                 | 509       | 39                | 4.9e-16 |
| 10 | GO:BP  | GO:0005262 | epithelial tube morphogenesis      | 466       | 36                | 1.0e-14 |
| 11 | GO:BP  | GO:0046731 | system development                 | 1534      | 93                | 8.4e-14 |
| 12 | GO:BP  | GO:0048686 | anatomical structure development   | 2037      | 98                | 9.9e-14 |
| 13 | GO:BP  | GO:0005241 | respiratory system development     | 239       | 22                | 1.4e-13 |
| 14 | GO:BP  | GO:0007444 | imaginal disc development          | 573       | 33                | 3.1e-13 |
| 15 | GO:BP  | GO:0007424 | open tracheal system development   | 250       | 21                | 6.7e-13 |

g:Profiler (bit.cs.ut.ee/gprofiler)

**D: Simulans, 108h**

| id | source | term_id    | term_name   | term_size | intersection_size | p_value |
|----|--------|------------|---|-----------|-------------------|---------|
| 1  | GO:BP  | GO:0009653 | anatomical structure morphogenesis                      | 1584      | 23                | 9.6e-02 |
| 2  | GO:BP  | GO:0048047 | mating behavior, sex discrimination                     | 9         | 2                 | 1.7e-01 |
| 3  | GO:BP  | GO:0051056 | regulation of small GTPase mediated signal transduction | 88        | 2                 | 1.8e-01 |
| 4  | GO:BP  | GO:0016199 | axon innervation choice point recognition               | 28        | 2                 | 2.3e-01 |
| 5  | GO:BP  | GO:0009888 | tissue development                                      | 1033      | 17                | 3.0e-01 |
| 6  | GO:BP  | GO:0016198 | axon choice point recognition                           | 28        | 2                 | 3.3e-01 |
| 7  | GO:BP  | GO:0048707 | instar larval or pupal morphogenesis                    | 444       | 9                 | 4.9e-01 |
| 8  | GO:BP  | GO:0009569 | macromolecule biosynthetic process                      | 2013      | 10                | 4.9e-01 |
| 9  | GO:BP  | GO:0009887 | animal organ morphogenesis                              | 761       | 12                | 5.1e-01 |
| 10 | GO:BP  | GO:0048518 | positive regulation of biological process               | 1874      | 7                 | 9.3e-01 |
| 11 | GO:BP  | GO:0034260 | negative regulation of GTPase activity                  | 2         | 1                 | 9.7e-01 |
| 12 | GO:BP  | GO:0005021 | protein stabilization                                   | 41        | 2                 | 9.8e-01 |
| 13 | GO:BP  | GO:0007090 | regulation of catalytic activity                        | 611       | 3                 | 6.9e-01 |
| 14 | GO:BP  | GO:0007552 | metamorphosis   | 470       | 10                | 5.1e-01 |
| 15 | GO:BP  | GO:0009886 | post-embryonic animal morphogenesis                     | 459       | 9                 | 6.2e-01 |

g:Profiler (bit.cs.ut.ee/gprofiler)

**F: Simulans, 120h**

| id | source | term_id    | term_name  | term_size | intersection_size | p_value |
|----|--------|------------|--|-----------|-------------------|---------|
| 1  | GO:BP  | GO:2000035 | semaphorin-plexin signaling pathway involved in regulation of photoreceptor cell axon guidance | 2         | 1                 | 1.5e-01 |
| 2  | GO:BP  | GO:0068886 | cellular response to virus   | 11        | 2                 | 3.7e-01 |
| 3  | GO:BP  | GO:0071526 | semaphorin-plexin signaling pathway  | 10        | 2                 | 3.3e-01 |
| 4  | GO:BP  | GO:0014033 | neural crest cell differentiation  | 5         | 1                 | 3.9e-01 |
| 5  | GO:BP  | GO:2000038 | regulation of photoreceptor cell axon guidance   | 5         | 1                 | 3.9e-01 |
| 6  | GO:BP  | GO:0014032 | neural crest cell development  | 5         | 1                 | 2.0e-01 |
| 7  | GO:BP  | GO:0009497 | mesenchymal cell migration   | 5         | 1                 | 3.9e-01 |
| 8  | GO:BP  | GO:0001755 | neural crest cell migration  | 5         | 1                 | 3.9e-01 |
| 9  | GO:BP  | GO:0048843 | negative regulation of axon extension involved in axon guidance                                | 6         | 1                 | 4.6e-01 |
| 10 | GO:BP  | GO:0048884 | stem cell development  | 6         | 1                 | 4.6e-01 |
| 11 | GO:BP  | GO:0005922 | negative regulation of chemotaxis  | 7         | 1                 | 5.4e-01 |
| 12 | GO:BP  | GO:0048762 | mesenchymal cell differentiation   | 7         | 1                 | 5.4e-01 |
| 13 | GO:BP  | GO:0018107 | peptidylserine phosphorylation   | 19        | 2                 | 6.6e-01 |
| 14 | GO:BP  | GO:0008039 | synaptic target recognition  | 50        | 2                 | 6.7e-01 |
| 15 | GO:BP  | GO:0018210 | peptidylserine modification  | 21        | 2                 | 6.9e-01 |

g:Profiler (bit.cs.ut.ee/gprofiler)

**Supplementary Table S 14: GO-terms enriched in DEGs in PhR cells between *D. mauritiana* and *D. simulans*.** This figure depicts up to the top 15 enriched GO-terms by p-value. Species comparisons were performed pairwise by timepoint.

**A: Simulans, 84h**

| id | source | term_id    | term_name   | term_size | intersection_size | p_value |
|----|--------|------------|---|-----------|-------------------|---------|
| 1  | GOBP   | GO:006412  | translation   | 474       | 3                 | 2.2e-01 |
| 2  | GOBP   | GO:0043043 | peptide biosynthetic process  | 522       | 3                 | 3.0e-01 |
| 3  | GOBP   | GO:0043604 | amide biosynthetic process  | 573       | 3                 | 3.9e-01 |
| 4  | GOBP   | GO:0002181 | cytoplasmic translation   | 150       | 2                 | 4.2e-01 |
| 5  | GOBP   | GO:0005518 | peptide metabolic process   | 624       | 3                 | 5.0e-01 |
| 6  | GOBP   | GO:0034645 | cellular macromolecule biosynthetic process                             | 691       | 3                 | 5.9e-01 |
| 7  | GOBP   | GO:0043603 | cellular amide metabolic process  | 721       | 3                 | 7.7e-01 |
| 8  | GOBP   | GO:0070602 | regulation of centromeric sister chromatid cohesion                     | 1         | 1                 | 8.2e-01 |
| 9  | GOBP   | GO:0034091 | regulation of maintenance of sister chromatid cohesion                  | 1         | 1                 | 8.2e-01 |
| 10 | GOBP   | GO:0034093 | positive regulation of maintenance of sister chromatid cohesion         | 1         | 1                 | 8.2e-01 |
| 11 | GOBP   | GO:0034184 | positive regulation of maintenance of mitotic sister chromatid cohesion | 1         | 1                 | 8.2e-01 |
| 12 | GOBP   | GO:0072355 | histone H3-T3 phosphorylation   | 1         | 1                 | 8.2e-01 |
| 13 | GOBP   | GO:0071962 | mitotic sister chromatid cohesion, centromeric                          | 1         | 1                 | 8.2e-01 |
| 14 | GOBP   | GO:0071963 | maintenance of mitotic sister chromatid cohesion, centromeric           | 1         | 1                 | 8.2e-01 |
| 15 | GOBP   | GO:0034182 | regulation of maintenance of mitotic sister chromatid cohesion          | 1         | 1                 | 8.2e-01 |

[p-Profiler \(DR.cs.uk.ac.gene@dr\)](#)

**B: Mauritiana, 108h**

| id | source | term_id    | term_name  | term_size | intersection_size | p_value |
|----|--------|------------|--|-----------|-------------------|---------|
| 1  | GOBP   | GO:0045677 | negative regulation of RT cell differentiation                         | 7         | 3                 | 8.0e-03 |
| 2  | GOBP   | GO:0110118 | negative regulation of compound eye photoreceptor cell differentiation | 8         | 3                 | 1.6e-02 |
| 3  | GOBP   | GO:0045676 | regulation of RT cell differentiation                                  | 24        | 4                 | 1.6e-02 |
| 4  | GOBP   | GO:0002185 | inlar larval or pupal development                                      | 561       | 14                | 1.6e-02 |
| 5  | GOBP   | GO:0110116 | regulation of compound eye photoreceptor cell differentiation          | 25        | 4                 | 1.9e-02 |
| 6  | GOBP   | GO:0007191 | post-embryonic development   | 658       | 15                | 2.2e-02 |
| 7  | GOBP   | GO:0097336 | response to alcohol  | 121       | 5                 | 2.2e-02 |
| 8  | GOBP   | GO:0007275 | multicellular organism development                                     | 2444      | 32                | 2.7e-02 |
| 9  | GOBP   | GO:0048856 | anatomical structure development                                       | 2837      | 34                | 3.1e-02 |
| 10 | GOBP   | GO:0045500 | severless signaling pathway  | 30        | 4                 | 4.0e-02 |
| 11 | GOBP   | GO:0046533 | negative regulation of photoreceptor cell differentiation              | 12        | 3                 | 5.0e-02 |
| 12 | GOBP   | GO:0045665 | negative regulation of neuron differentiation                          | 12        | 3                 | 8.0e-02 |
| 13 | GOBP   | GO:0045501 | regulation of severless signaling pathway                              | 13        | 3                 | 7.2e-02 |
| 14 | GOBP   | GO:0038127 | ERBB signaling pathway   | 88        | 5                 | 7.3e-02 |
| 15 | GOBP   | GO:0007173 | epidermal growth factor receptor signaling pathway                     | 88        | 5                 | 7.3e-02 |

[p-Profiler \(DR.cs.uk.ac.gene@dr\)](#)

**D: Mauritiana, 120h**

| id | source | term_id    | term_name   | term_size | intersection_size | p_value |
|----|--------|------------|---|-----------|-------------------|---------|
| 1  | GOBP   | GO:0065008 | regulation of biological quality                          | 1363      | 10                | 3.1e-03 |
| 2  | GOBP   | GO:0022607 | cellular component assembly                               | 1509      | 7                 | 1.3e-02 |
| 3  | GOBP   | GO:0044085 | cellular component biogenesis                             | 1703      | 7                 | 3.0e-02 |
| 4  | GOBP   | GO:0002251 | organ or tissue specific immune response                  | 12        | 2                 | 1.0e-01 |
| 5  | GOBP   | GO:0002385 | mucosal immune response                                   | 12        | 2                 | 1.0e-01 |
| 6  | GOBP   | GO:0031400 | negative regulation of protein modification process       | 85        | 3                 | 1.9e-01 |
| 7  | GOBP   | GO:0051128 | regulation of cellular component organization             | 713       | 8                 | 1.9e-01 |
| 8  | GOBP   | GO:0000066 | regulation of anatomical structure size                   | 244       | 4                 | 2.2e-01 |
| 9  | GOBP   | GO:0051130 | positive regulation of cellular component organization    | 279       | 4                 | 3.6e-01 |
| 10 | GOBP   | GO:0044087 | regulation of cellular component biogenesis               | 326       | 4                 | 5.0e-01 |
| 11 | GOBP   | GO:0032289 | negative regulation of cellular protein metabolic process | 286       | 4                 | 5.1e-01 |
| 12 | GOBP   | GO:0051248 | negative regulation of protein metabolic process          | 287       | 4                 | 5.2e-01 |
| 13 | GOBP   | GO:0010715 | regulation of extracellular matrix disassembly            | 1         | 1                 | 8.2e-01 |
| 14 | GOBP   | GO:0010716 | negative regulation of extracellular matrix disassembly   | 1         | 1                 | 8.2e-01 |
| 15 | GOBP   | GO:0071920 | endocannabinoid signaling pathway                         | 1         | 1                 | 8.9e-01 |

[p-Profiler \(DR.cs.uk.ac.gene@dr\)](#)

**C: Simulans, 108h**

| id | source | term_id    | term_name   | term_size | intersection_size | p_value |
|----|--------|------------|---|-----------|-------------------|---------|
| 1  | GOBP   | GO:0006412 | translation   | 474       | 3                 | 5.9e-02 |
| 2  | GOBP   | GO:0043043 | peptide biosynthetic process  | 522       | 3                 | 7.0e-02 |
| 3  | GOBP   | GO:0043604 | amide biosynthetic process  | 573       | 3                 | 1.0e-01 |
| 4  | GOBP   | GO:0005518 | peptide metabolic process   | 624       | 3                 | 1.4e-01 |
| 5  | GOBP   | GO:0034645 | cellular macromolecule biosynthetic process                             | 691       | 3                 | 1.8e-01 |
| 6  | GOBP   | GO:0043603 | cellular amide metabolic process  | 721       | 3                 | 2.1e-01 |
| 7  | GOBP   | GO:0002181 | cytoplasmic translation   | 150       | 2                 | 4.4e-01 |
| 8  | GOBP   | GO:0034091 | regulation of maintenance of sister chromatid cohesion                  | 1         | 1                 | 5.2e-01 |
| 9  | GOBP   | GO:0034093 | positive regulation of maintenance of sister chromatid cohesion         | 1         | 1                 | 5.2e-01 |
| 10 | GOBP   | GO:0034182 | regulation of maintenance of mitotic sister chromatid cohesion          | 1         | 1                 | 5.2e-01 |
| 11 | GOBP   | GO:0034184 | positive regulation of maintenance of mitotic sister chromatid cohesion | 1         | 1                 | 5.2e-01 |
| 12 | GOBP   | GO:0072355 | histone H3-T3 phosphorylation   | 1         | 1                 | 5.2e-01 |
| 13 | GOBP   | GO:0045876 | positive regulation of sister chromatid cohesion                        | 1         | 1                 | 5.2e-01 |
| 14 | GOBP   | GO:0071962 | mitotic sister chromatid cohesion, centromeric                          | 1         | 1                 | 5.2e-01 |
| 15 | GOBP   | GO:0070602 | regulation of centromeric sister chromatid cohesion                     | 1         | 1                 | 5.2e-01 |

[p-Profiler \(DR.cs.uk.ac.gene@dr\)](#)

**E: Simulans, 120h**

| id | source | term_id    | term_name  | term_size | intersection_size | p_value |
|----|--------|------------|--|-----------|-------------------|---------|
| 1  | GOBP   | GO:0006412 | translation  | 474       | 3                 | 2.2e-01 |
| 2  | GOBP   | GO:0043043 | peptide biosynthetic process   | 522       | 3                 | 2.9e-01 |
| 3  | GOBP   | GO:0043604 | amide biosynthetic process   | 573       | 3                 | 3.8e-01 |
| 4  | GOBP   | GO:0002181 | cytoplasmic translation  | 150       | 2                 | 4.2e-01 |
| 5  | GOBP   | GO:0005518 | peptide metabolic process  | 624       | 3                 | 4.9e-01 |
| 6  | GOBP   | GO:0034091 | positive regulation of sister chromatid cohesion                                     | 1         | 1                 | 5.9e-01 |
| 7  | GOBP   | GO:0007200 | positive regulation of maintenance of mitotic sister chromatid cohesion, centromeric | 1         | 1                 | 5.9e-01 |
| 8  | GOBP   | GO:0007191 | regulation of maintenance of mitotic sister chromatid cohesion, centromeric          | 1         | 1                 | 5.9e-01 |
| 9  | GOBP   | GO:0070602 | regulation of centromeric sister chromatid cohesion                                  | 1         | 1                 | 5.9e-01 |
| 10 | GOBP   | GO:0071960 | maintenance of mitotic sister chromatid cohesion, centromeric                        | 1         | 1                 | 5.9e-01 |
| 11 | GOBP   | GO:0071962 | mitotic sister chromatid cohesion, centromeric                                       | 1         | 1                 | 5.9e-01 |
| 12 | GOBP   | GO:0072355 | histone H3-T3 phosphorylation  | 1         | 1                 | 5.9e-01 |
| 13 | GOBP   | GO:0034184 | positive regulation of maintenance of mitotic sister chromatid cohesion              | 1         | 1                 | 5.9e-01 |
| 14 | GOBP   | GO:0034182 | regulation of maintenance of mitotic sister chromatid cohesion                       | 1         | 1                 | 5.9e-01 |
| 15 | GOBP   | GO:0034093 | positive regulation of maintenance of sister chromatid cohesion                      | 1         | 1                 | 5.9e-01 |

[p-Profiler \(DR.cs.uk.ac.gene@dr\)](#)

**Supplementary Table S 15: GO-terms enriched in DEGS in MF cells between *D. mauritiana* and *D. simulans*.** This figure depicts up to the top 15 enriched GO-terms by p-value. Species comparisons were performed pairwise by timepoint. No GO-terms were found enriched for *D. mauritiana* at 84 h.

**A: Mauritiana, 84h**

| id | source | term_id    | term_name                            | term_size | intersection_size | p_value |
|----|--------|------------|--------------------------------------|-----------|-------------------|---------|
| 1  | GO:BP  | GO:0007267 | cell-cell signaling                  | 532       | 78                | 5.0e-41 |
| 2  | GO:BP  | GO:0098537 | trans-synaptic signaling             | 324       | 63                | 8.9e-40 |
| 3  | GO:BP  | GO:0098536 | synaptic signaling                   | 332       | 63                | 4.2e-39 |
| 4  | GO:BP  | GO:0098916 | anterograde trans-synaptic signaling | 321       | 62                | 6.9e-39 |
| 5  | GO:BP  | GO:0007268 | chemical synaptic transmission       | 321       | 62                | 6.9e-39 |
| 6  | GO:BP  | GO:0007399 | nervous system development           | 1170      | 129               | 1.9e-35 |
| 7  | GO:BP  | GO:0030162 | neuron differentiation               | 810       | 98                | 1.1e-28 |
| 8  | GO:BP  | GO:0007154 | cell communication                   | 2087      | 164               | 1.7e-28 |
| 9  | GO:BP  | GO:0048699 | generation of neurons                | 883       | 102               | 2.5e-28 |
| 10 | GO:BP  | GO:0048666 | neuron development                   | 634       | 86                | 3.0e-28 |
| 11 | GO:BP  | GO:0023052 | signaling                            | 2042      | 157               | 4.4e-28 |
| 12 | GO:BP  | GO:0022008 | neurogenesis                         | 947       | 105               | 8.9e-28 |
| 13 | GO:BP  | GO:0048731 | system development                   | 1634      | 136               | 2.0e-27 |
| 14 | GO:BP  | GO:0050808 | synapse organization                 | 310       | 57                | 4.1e-26 |
| 15 | GO:BP  | GO:0061564 | axon development                     | 349       | 61                | 4.7e-25 |

[gProfiler \(bit.cs.ut.ee/gprofiler\)](http://bit.cs.ut.ee/gprofiler)

**B: Simulans, 84h**

| id | source | term_id    | term_name                                     | term_size | intersection_size | p_value |
|----|--------|------------|---|-----------|-------------------|---------|
| 1  | GO:BP  | GO:0035114 | imaginal disc-derived appendage morphogenesis | 332       | 17                | 1.0e-10 |
| 2  | GO:BP  | GO:0035107 | appendage morphogenesis                       | 334       | 17                | 1.1e-10 |
| 3  | GO:BP  | GO:0009887 | animal organ morphogenesis                    | 761       | 23                | 1.3e-10 |
| 4  | GO:BP  | GO:0048737 | imaginal disc-derived appendage development   | 339       | 17                | 1.5e-10 |
| 5  | GO:BP  | GO:0048736 | appendage development                         | 342       | 17                | 1.7e-10 |
| 6  | GO:BP  | GO:0009653 | anatomical structure morphogenesis            | 1584      | 43                | 2.1e-10 |
| 7  | GO:BP  | GO:0009791 | post-embryonic development                    | 656       | 20                | 7.5e-09 |
| 8  | GO:BP  | GO:0035239 | tube morphogenesis                            | 509       | 18                | 9.2e-09 |
| 9  | GO:BP  | GO:0009888 | tissue development                            | 1053      | 33                | 9.2e-09 |
| 10 | GO:BP  | GO:0035120 | post-embryonic appendage morphogenesis        | 323       | 15                | 1.5e-08 |
| 11 | GO:BP  | GO:0048563 | post-embryonic animal organ morphogenesis     | 396       | 16                | 2.3e-08 |
| 12 | GO:BP  | GO:0007560 | imaginal disc morphogenesis                   | 396       | 16                | 2.3e-08 |
| 13 | GO:BP  | GO:0060562 | epithelial tube morphogenesis                 | 466       | 17                | 2.4e-08 |
| 14 | GO:BP  | GO:0007444 | imaginal disc development                     | 573       | 19                | 3.6e-08 |
| 15 | GO:BP  | GO:0002165 | instar larval or pupal development            | 561       | 18                | 4.6e-08 |

[gProfiler \(bit.cs.ut.ee/gprofiler\)](http://bit.cs.ut.ee/gprofiler)

**C: Mauritiana, 108h**

| id | source | term_id    | term_name                                 | term_size | intersection_size | p_value |
|----|--------|------------|---|-----------|-------------------|---------|
| 1  | GO:BP  | GO:0002165 | instar larval or pupal development        | 561       | 24                | 1.7e-07 |
| 2  | GO:BP  | GO:0009791 | post-embryonic development                | 658       | 23                | 7.7e-07 |
| 3  | GO:BP  | GO:0035239 | tube morphogenesis                        | 509       | 23                | 8.2e-07 |
| 4  | GO:BP  | GO:0060562 | epithelial tube morphogenesis             | 466       | 22                | 8.9e-07 |
| 5  | GO:BP  | GO:0048563 | post-embryonic animal organ morphogenesis | 396       | 20                | 1.8e-06 |
| 6  | GO:BP  | GO:0007560 | imaginal disc morphogenesis               | 396       | 20                | 1.8e-06 |
| 7  | GO:BP  | GO:0009886 | post-embryonic animal morphogenesis       | 459       | 21                | 4.0e-06 |
| 8  | GO:BP  | GO:0051270 | regulation of cellular component movement | 124       | 9                 | 5.8e-06 |
| 9  | GO:BP  | GO:0040012 | regulation of locomotion                  | 124       | 9                 | 5.8e-06 |
| 10 | GO:BP  | GO:0048569 | post-embryonic animal organ development   | 484       | 21                | 1.0e-05 |
| 11 | GO:BP  | GO:0048707 | instar larval or pupal morphogenesis      | 444       | 20                | 1.3e-05 |
| 12 | GO:BP  | GO:0020009 | morphogenesis of an epithelium            | 595       | 23                | 1.6e-05 |
| 13 | GO:BP  | GO:0009887 | animal organ morphogenesis                | 761       | 25                | 1.9e-05 |
| 14 | GO:BP  | GO:0007275 | multicellular organism development        | 2444      | 49                | 2.2e-05 |
| 15 | GO:BP  | GO:0048729 | tissue morphogenesis                      | 614       | 23                | 2.9e-05 |

[gProfiler \(bit.cs.ut.ee/gprofiler\)](http://bit.cs.ut.ee/gprofiler)

**D: Simulans, 108h**

| id | source | term_id    | term_name   | term_size | intersection_size | p_value |
|----|--------|------------|---|-----------|-------------------|---------|
| 1  | GO:BP  | GO:0006412 | translation   | 474       | 4                 | 4.6e-01 |
| 2  | GO:BP  | GO:0110077 | vesicle-mediated intercellular transport                  | 1         | 1                 | 6.1e-01 |
| 3  | GO:BP  | GO:0043043 | peptide biosynthetic process                              | 522       | 4                 | 6.7e-01 |
| 4  | GO:BP  | GO:0009059 | macromolecule biosynthetic process                        | 2013      | 10                | 7.6e-01 |
| 5  | GO:BP  | GO:0010558 | negative regulation of macromolecule biosynthetic process | 446       | 5                 | 7.6e-01 |
| 6  | GO:BP  | GO:0048518 | positive regulation of biological process                 | 1874      | 9                 | 8.2e-01 |
| 7  | GO:BP  | GO:0044271 | cellular nitrogen compound biosynthetic process           | 2072      | 5                 | 8.9e-01 |
| 8  | GO:BP  | GO:0031327 | negative regulation of cellular biosynthetic process      | 463       | 5                 | 9.0e-01 |
| 9  | GO:BP  | GO:0009890 | negative regulation of biosynthetic process               | 468       | 5                 | 9.5e-01 |
| 10 | GO:BP  | GO:0043604 | amide biosynthetic process                                | 573       | 4                 | 9.5e-01 |
| 11 | GO:BP  | GO:0048598 | embryonic morphogenesis                                   | 239       | 1                 | 1.0e+00 |
| 12 | GO:BP  | GO:0048638 | regulation of developmental growth                        | 240       | 2                 | 1.0e+00 |
| 13 | GO:BP  | GO:0048640 | negative regulation of developmental growth               | 93        | 2                 | 1.0e+00 |
| 14 | GO:BP  | GO:0048589 | developmental growth                                      | 377       | 3                 | 1.0e+00 |
| 15 | GO:BP  | GO:0048666 | neuron development  | 634       | 2                 | 1.0e+00 |

[gProfiler \(bit.cs.ut.ee/gprofiler\)](http://bit.cs.ut.ee/gprofiler)

**E: Mauritiana, 120h**

| id | source | term_id    | term_name                                | term_size | intersection_size | p_value |
|----|--------|------------|--|-----------|-------------------|---------|
| 1  | GO:BP  | GO:0002251 | organ or tissue specific immune response | 12        | 2                 | 2.6e-02 |
| 2  | GO:BP  | GO:0002385 | mucosal immune response                  | 12        | 2                 | 2.6e-02 |

[gProfiler \(bit.cs.ut.ee/gprofiler\)](http://bit.cs.ut.ee/gprofiler)

**F: Simulans, 120h**

| id | source | term_id    | term_name  | term_size | intersection_size | p_value |
|----|--------|------------|--|-----------|-------------------|---------|
| 1  | GO:BP  | GO:0050919 | negative chemotaxis                                  | 24        | 2                 | 7.9e-02 |
| 2  | GO:BP  | GO:0016199 | axon midline choice point recognition                | 26        | 2                 | 9.3e-02 |
| 3  | GO:BP  | GO:0016198 | axon choice point recognition                        | 28        | 2                 | 1.1e-01 |
| 4  | GO:BP  | GO:0043085 | positive regulation of catalytic activity            | 236       | 3                 | 1.4e-01 |
| 5  | GO:BP  | GO:0044093 | positive regulation of molecular function            | 288       | 3                 | 2.6e-01 |
| 6  | GO:BP  | GO:0034260 | negative regulation of GTPase activity               | 2         | 1                 | 3.2e-01 |
| 7  | GO:BP  | GO:0010977 | negative regulation of neuron projection development | 16        | 2                 | 3.4e-01 |
| 8  | GO:BP  | GO:0007411 | axon guidance  | 273       | 3                 | 3.8e-01 |
| 9  | GO:BP  | GO:0097485 | neuron projection guidance                           | 281       | 3                 | 4.2e-01 |
| 10 | GO:BP  | GO:0000904 | cell morphogenesis involved in differentiation       | 489       | 5                 | 4.7e-01 |
| 11 | GO:BP  | GO:0006935 | chemotaxis   | 295       | 3                 | 4.8e-01 |
| 12 | GO:BP  | GO:0033674 | positive regulation of kinase activity               | 77        | 2                 | 5.9e-01 |
| 13 | GO:BP  | GO:0007409 | axonogenesis   | 337       | 3                 | 7.1e-01 |
| 14 | GO:BP  | GO:0061564 | axon development                                     | 349       | 3                 | 7.9e-01 |
| 15 | GO:BP  | GO:0051347 | positive regulation of transferase activity          | 89        | 2                 | 7.9e-01 |

[gProfiler \(bit.cs.ut.ee/gprofiler\)](http://bit.cs.ut.ee/gprofiler)

**Supplementary Table S 16: GO-terms enriched in DEGS in PPN cells between *D. mauritiana* and *D. simulans*.** This figure depicts up to the top 15 enriched GO-terms by p-value. Species comparisons were performed pairwise by timepoint.

**A: Mauritiana, 108h**

| id | source | term_id    | term_name                                 | term_size | intersection_size | p_value |
|----|--------|------------|---|-----------|-------------------|---------|
| 1  | GO:BP  | GO:002165  | instar larval or pupal development        | 561       | 15                | 5.0e-07 |
| 2  | GO:BP  | GO:009791  | post-embryonic development                | 658       | 15                | 4.5e-06 |
| 3  | GO:BP  | GO:0048956 | anatomical structure development          | 2637      | 27                | 7.1e-05 |
| 4  | GO:BP  | GO:0007275 | multicellular organism development        | 2444      | 25                | 8.4e-05 |
| 5  | GO:BP  | GO:0007560 | imaginal disc morphogenesis               | 396       | 11                | 1.1e-04 |
| 6  | GO:BP  | GO:0048563 | post-embryonic animal organ morphogenesis | 396       | 11                | 1.1e-04 |
| 7  | GO:BP  | GO:0035239 | tube morphogenesis                        | 509       | 12                | 1.5e-04 |
| 8  | GO:BP  | GO:0032502 | developmental process                     | 3013      | 27                | 2.7e-04 |
| 9  | GO:BP  | GO:0048707 | instar larval or pupal morphogenesis      | 444       | 11                | 3.4e-04 |
| 10 | GO:BP  | GO:009886  | post-embryonic animal morphogenesis       | 459       | 11                | 4.8e-04 |
| 11 | GO:BP  | GO:0060562 | epithelial tube morphogenesis             | 466       | 11                | 5.5e-04 |
| 12 | GO:BP  | GO:0007552 | metamorphosis                             | 470       | 11                | 6.0e-04 |
| 13 | GO:BP  | GO:0045659 | post-embryonic animal organ development   | 484       | 11                | 8.0e-04 |
| 14 | GO:BP  | GO:0048731 | system development                        | 1534      | 18                | 1.0e-03 |
| 15 | GO:BP  | GO:009887  | animal organ morphogenesis                | 761       | 13                | 1.7e-03 |

[gProfiler \(bit.cs.ut.ee/gprofiler\)](http://bit.cs.ut.ee/gprofiler)

**B: Simulans, 108h**

| id | source | term_id    | term_name  | term_size | intersection_size | p_value |
|----|--------|------------|--|-----------|-------------------|---------|
| 1  | GO:BP  | GO:0007010 | cytoskeleton organization  | 691       | 6                 | 1.2e-0  |
| 2  | GO:BP  | GO:0050794 | regulation of cellular process   | 4110      | 17                | 1.5e-0  |
| 3  | GO:BP  | GO:0048047 | mating behavior, sex discrimination  | 9         | 2                 | 2.1e-0  |
| 4  | GO:BP  | GO:0051493 | regulation of cytoskeleton organization                                      | 177       | 4                 | 3.3e-0  |
| 5  | GO:BP  | GO:0007015 | actin filament organization  | 182       | 4                 | 3.6e-0  |
| 6  | GO:BP  | GO:0050789 | regulation of biological process   | 4401      | 17                | 3.9e-0  |
| 7  | GO:BP  | GO:0006518 | peptide metabolic process  | 624       | 7                 | 4.8e-0  |
| 8  | GO:BP  | GO:0110053 | regulation of actin filament organization                                    | 85        | 3                 | 5.4e-0  |
| 9  | GO:BP  | GO:0016199 | axon midline choice point recognition  | 26        | 2                 | 6.5e-0  |
| 10 | GO:BP  | GO:0010638 | positive regulation of organelle organization                                | 123       | 4                 | 7.2e-0  |
| 11 | GO:BP  | GO:0016198 | axon choice point recognition  | 28        | 2                 | 7.6e-0  |
| 12 | GO:BP  | GO:0016544 | male courtship behavior, tapping to detect pheromone                         | 2         | 1                 | 7.9e-0  |
| 13 | GO:BP  | GO:0016543 | male courtship behavior, orientation prior to leg tapping and wing vibration | 2         | 1                 | 7.9e-0  |
| 14 | GO:BP  | GO:0032535 | regulation of cellular component size  | 179       | 4                 | 9.7e-0  |
| 15 | GO:BP  | GO:0050821 | protein stabilization  | 41        | 2                 | 9.8e-0  |

[gProfiler \(bit.cs.ut.ee/gprofiler\)](http://bit.cs.ut.ee/gprofiler)

**C: Mauritiana, 120h**

| id | source | term_id    | term_name   | term_size | intersection_size | p_value |
|----|--------|------------|---|-----------|-------------------|---------|
| 1  | GO:BP  | GO:0065008 | regulation of biological quality                    | 1363      | 12                | 1.1e-03 |
| 2  | GO:BP  | GO:0031400 | negative regulation of protein modification process | 85        | 3                 | 1.1e-02 |
| 3  | GO:BP  | GO:0002251 | organ or tissue specific immune response            | 12        | 2                 | 1.3e-02 |
| 4  | GO:BP  | GO:0002385 | mucosal immune response                             | 12        | 2                 | 1.3e-02 |

[gProfiler \(bit.cs.ut.ee/gprofiler\)](http://bit.cs.ut.ee/gprofiler)

**D: Simulans, 120h**

| id | source | term_id    | term_name                              | term_size | intersection_size | p_value |
|----|--------|------------|--|-----------|-------------------|---------|
| 1  | GO:BP  | GO:0006412 | translation                            | 474       | 4                 | 4.9e-01 |
| 2  | GO:BP  | GO:0034260 | negative regulation of GTPase activity | 2         | 1                 | 5.6e-01 |
| 3  | GO:BP  | GO:0000266 | mitochondrial fission                  | 16        | 2                 | 5.6e-01 |
| 4  | GO:BP  | GO:0043043 | peptide biosynthetic process           | 522       | 4                 | 7.1e-01 |
| 5  | GO:BP  | GO:0007062 | sister chromatid cohesion              | 33        | 2                 | 8.5e-01 |
| 6  | GO:BP  | GO:0000003 | reproduction                           | 1497      | 2                 | 1.0e+00 |
| 7  | GO:BP  | GO:0046620 | regulation of organ growth             | 34        | 2                 | 1.0e+00 |
| 8  | GO:BP  | GO:0046622 | positive regulation of organ growth    | 13        | 1                 | 1.0e+00 |
| 9  | GO:BP  | GO:0046700 | heterocycle catabolic process          | 200       | 2                 | 1.0e+00 |
| 10 | GO:BP  | GO:0046718 | viral entry into host cell             | 31        | 1                 | 1.0e+00 |
| 11 | GO:BP  | GO:0046903 | secretion                              | 254       | 1                 | 1.0e+00 |
| 12 | GO:BP  | GO:0046907 | intracellular transport                | 745       | 1                 | 1.0e+00 |
| 13 | GO:BP  | GO:0048102 | autophagic cell death                  | 28        | 1                 | 1.0e+00 |
| 14 | GO:BP  | GO:0048232 | male gamete generation                 | 323       | 1                 | 1.0e+00 |
| 15 | GO:BP  | GO:0048278 | vesicle docking                        | 44        | 1                 | 1.0e+00 |

[gProfiler \(bit.cs.ut.ee/gprofiler\)](http://bit.cs.ut.ee/gprofiler)

**Supplementary Table S 17: GO-terms enriched in DEGS in ocelli cells between *D. mauritiana* and *D. simulans*.** This figure depicts up to the top 15 enriched GO-terms by p-value. Species comparisons were performed pairwise by timepoint.

## 9 References

- 10X Genomics (2021): Nuclei Isolation from Cell Suspensions & Tissues for Single Cell RNA Sequencing: CG000124 • Rev F: Protocol. CG000124 • Rev F, checked on 4/28/2022.
- Ahmad, Shaad M.; Tansey, Terese R.; Busser, Brian W.; Nolte, Michael T.; Jeffries, Neal; Gisselbrecht, Stephen S. et al. (2012): Two forkhead transcription factors regulate the division of cardiac progenitor cells by a Polo-dependent pathway. In *Developmental Cell* 23 (1), pp. 97–111. DOI: 10.1016/j.devcel.2012.05.011.
- Aibar, Sara; González-Blas, Carmen Bravo; Moerman, Thomas; Huynh-Thu, Vân Anh; Imrichova, Hana; Hulselmans, Gert et al. (2017): SCENIC: single-cell regulatory network inference and clustering. In *Nature methods* 14 (11), pp. 1083–1086. DOI: 10.1038/nmeth.4463.
- Alto, Laura Taylor; Terman, Jonathan R. (2017): Semaphorins and their Signaling Mechanisms. In *Methods in molecular biology (Clifton, N.J.)* 1493, pp. 1–25. DOI: 10.1007/978-1-4939-6448-2\_1.
- Anderson, Caitlin; Reiss, India; Zhou, Cyrus; Cho, Annie; Siddiqi, Haziq; Mormann, Benjamin et al. (2017): Natural variation in stochastic photoreceptor specification and color preference in *Drosophila*. In *eLife* 6. DOI: 10.7554/eLife.29593.
- Andres, A. J.; Fletcher, J. C.; Karim, F. D.; Thummel, C. S. (1993): Molecular analysis of the initiation of insect metamorphosis: a comparative study of *Drosophila* ecdysteroid-regulated transcription. In *DEVELOPMENTAL BIOLOGY* 160 (2), pp. 388–404. DOI: 10.1006/dbio.1993.1315.
- Aradska, Jana; Bulat, Tanja; Sialana, Fernando J.; Birner-Gruenberger, Ruth; Erich, Buchner; Lubec, Gert (2015): Gel-free mass spectrometry analysis of *Drosophila melanogaster* heads. In *Proteomics* 15 (19), pp. 3356–3360. DOI: 10.1002/pmic.201500092.
- Arif, Saad; Hilbrant, Maarten; Hopfen, Corinna; Almudi, Isabel; Nunes, Maria D. S.; Posnien, Nico et al. (2013): Genetic and developmental analysis of differences in eye and face morphology between *Drosophila simulans* and *Drosophila mauritiana*. In *Evolution & Development* 15 (4), pp. 257–267. DOI: 10.1111/ede.12027.
- Ariss, Majd M.; Islam, Abul B. M. M. K.; Critcher, Meg; Zappia, Maria Paula; Frolov, Maxim V. (2018): Single cell RNA-sequencing identifies a metabolic aspect of apoptosis in Rbf mutant. In *Nature Communications* 9 (1), p. 5024. DOI: 10.1038/s41467-018-07540-z.
- Bakken, Trygve E.; Hodge, Rebecca D.; Miller, Jeremy A.; Yao, Zizhen; Nguyen, Thuc Nghi; Aebermann, Brian et al. (2018): Single-nucleus and single-cell transcriptomes compared in matched cortical cell types. In *PLoS one* 13 (12), e0209648. DOI: 10.1371/journal.pone.0209648.
- Bakken, Trygve E.; Jorstad, Nikolas L.; Hu, Qiwen; Lake, Blue B.; Tian, Wei; Kalmbach, Brian E. et al. (2021): Comparative cellular analysis of motor cortex in human, marmoset and mouse. In *Nature* 598 (7879), pp. 111–119. DOI: 10.1038/s41586-021-03465-8.
- Barrios, Natalia; González-Pérez, Esther; Hernández, Rosario; Campuzano, Sonsoles (2015): The Homeodomain Iroquois Proteins Control Cell Cycle Progression and Regulate the Size of Developmental Fields. In *PLOS Genetics* 11 (8), e1005463. DOI: 10.1371/journal.pgen.1005463.
- Berg, Daniel A.; Su, Yijing; Jimenez-Cyrus, Dennisse; Patel, Aneek; Huang, Nancy; Morizet, David et al. (2019): A Common Embryonic Origin of Stem Cells Drives Developmental and Adult Neurogenesis. In *Cell* 177 (3), 654–668.e15. DOI: 10.1016/j.cell.2019.02.010.

- Blochlinger, K.; Jan, L. Y.; Jan, Y. N. (1993): Postembryonic patterns of expression of cut, a locus regulating sensory organ identity in *Drosophila*. In *Development (Cambridge, England)* 117 (2), pp. 441–450. DOI: 10.1242/dev.117.2.441.
- Bloomington *Drosophila* Stock Center (2019): Bloomington *Drosophila* Stock Center. Lists accessed from this page include *lexA* stocks with associated expression pattern information. Available online at [https://bdsc.indiana.edu/stocks/lexa/lexa\\_stanex.html](https://bdsc.indiana.edu/stocks/lexa/lexa_stanex.html).
- Bohère, Jérôme; Mancheno-Ferris, Alexandra; Al Hayek, Sandy; Zanet, Jennifer; Valenti, Philippe; Akino, Kohsuke et al. (2018): Shavenbaby and Yorkie mediate Hippo signaling to protect adult stem cells from apoptosis. In *Nature Communications* 9 (1), p. 5123. DOI: 10.1038/s41467-018-07569-0.
- Braid, Lorena R.; Verheyen, Esther M. (2008): *Drosophila* nemo promotes eye specification directed by the retinal determination gene network. In *Genetics* 180 (1), pp. 283–299. DOI: 10.1534/genetics.108.092155.
- Brás-Pereira, Catarina; Casares, Fernando (2008): An antennal-specific role for bowl in repressing supernumerary appendage development in *Drosophila*. In *Mechanisms of Development* 125 (9-10), pp. 809–821. DOI: 10.1016/j.mod.2008.07.002.
- Bravo González-Blas, Carmen; Quan, Xiao-Jiang; Duran-Romaña, Ramon; Taskiran, Ibrahim Ihsan; Koldere, Duygu; Davie, Kristofer et al. (2020): Identification of genomic enhancers through spatial integration of single-cell transcriptomics and epigenomics. In *Molecular systems biology* 16 (5), e9438. DOI: 10.15252/msb.20209438.
- Brunner, D.; Dücker, K.; Oellers, N.; Hafen, E.; Scholz, H.; Klämbt, C. (1994): The ETS domain protein pointed-P2 is a target of MAP kinase in the sevenless signal transduction pathway. In *Nature* 370 (6488), pp. 386–389. DOI: 10.1038/370386a0.
- Buchberger, Elisa; Bilen, Anil; Ayaz, Sanem; Salamanca, David; Matas de Las Heras, Cristina; Niksic, Armin et al. (2021): Variation in Pleiotropic Hub Gene Expression Is Associated with Interspecific Differences in Head Shape and Eye Size in *Drosophila*. In *Molecular biology and evolution* 38 (5), pp. 1924–1942. DOI: 10.1093/molbev/msaa335.
- Bulloj, Ayelen; Maminishkis, Arvydas; Mizui, Masayuki; Finnemann, Silvia C. (2018): Semaphorin4D-PlexinB1 Signaling Attenuates Photoreceptor Outer Segment Phagocytosis by Reducing Rac1 Activity of RPE Cells. In *Molecular neurobiology* 55 (5), pp. 4320–4332. DOI: 10.1007/s12035-017-0649-5.
- Bunis, Daniel G.; Andrews, Jared; Fragiadakis, Gabriela K.; Burt, Trevor D.; Sirota, Marina (2020): dittoSeq: Universal User-Friendly Single-Cell and Bulk RNA Sequencing Visualization Toolkit. In *Bioinformatics*. DOI: 10.1093/bioinformatics/btaa1011.
- Butler, S. J.; Ray, S.; Hiromi, Y. (1997): klingon, a novel member of the *Drosophila* immunoglobulin superfamily, is required for the development of the R7 photoreceptor neuron. In *Development (Cambridge, England)* 124 (4), pp. 781–792. DOI: 10.1242/dev.124.4.781.
- Cao, Junyue; Packer, Jonathan S.; Ramani, Vijay; Cusanovich, Darren A.; Huynh, Chau; Daza, Riza et al. (2017): Comprehensive single-cell transcriptional profiling of a multicellular organism. In *Science (New York, N.Y.)* 357 (6352), pp. 661–667. DOI: 10.1126/science.aam8940.
- Cao, Yue; Lin, Yingxin; Ormerod, John T.; Yang, Pengyi; Yang, Jean Y. H.; Lo, Kitty K. (2019): scDC: single cell differential composition analysis. In *BMC bioinformatics* 20 (Suppl 19), p. 721. DOI: 10.1186/s12859-019-3211-9.

- Casares, Fernando; McGregor, Alistair P. (2021): The evolution and development of eye size in flies. In *Wiley interdisciplinary reviews. Developmental biology* 10 (2), e380. DOI: 10.1002/wdev.380.
- Chan, Leo Li-Ying; Kuksin, Dmitry; Laverty, Daniel J.; Saldi, Stephanie; Qiu, Jean (2015): Morphological observation and analysis using automated image cytometry for the comparison of trypan blue and fluorescence-based viability detection method. In *Cytotechnology* 67 (3), pp. 461–473. DOI: 10.1007/s10616-014-9704-5.
- Chang, Z.; Price, B. D.; Bockheim, S.; Boedigheimer, M. J.; Smith, R.; Laughon, A. (1993): Molecular and genetic characterization of the *Drosophila tartan* gene. In *DEVELOPMENTAL BIOLOGY* 160 (2), pp. 315–332. DOI: 10.1006/dbio.1993.1310.
- Chari, Tara; Weissbourd, Brandon; Gehring, Jase; Ferraioli, Anna; Leclère, Lucas; Herl, Makenna et al. (2021): Whole-animal multiplexed single-cell RNA-seq reveals transcriptional shifts across *Clytia medusa* cell types. In *Science advances* 7 (48), eabh1683. DOI: 10.1126/sciadv.abh1683.
- Charlton-Perkins, Mark; Whitaker, S. Leigh; Fei, Yueyang; Xie, Baotong; Li-Kroeger, David; Gebelein, Brian; Cook, Tiffany (2011): Prospero and Pax2 combinatorially control neural cell fate decisions by modulating Ras- and Notch-dependent signaling. In *Neural development* 6, p. 20. DOI: 10.1186/1749-8104-6-20.
- Chen, Dongsheng; Sun, Jian; Zhu, Jiacheng; Ding, Xiangning; Lan, Tianming; Wang, Xiran et al. (2021): Single cell atlas for 11 non-model mammals, reptiles and birds. In *Nature Communications* 12 (1), p. 7083. DOI: 10.1038/s41467-021-27162-2.
- Chen, Tse-Yin (1929): On the development of imaginal buds in normal and mutant *Drosophila melanogaster*. In *J. Morphol.* 47 (1), pp. 135–199. DOI: 10.1002/jmor.1050470105.
- Choi, Kwang-Wook; Benzer, Seymour (1994): Rotation of photoreceptor clusters in the developing *drosophila* eye requires the *nemo* gene. In *Cell* 78 (1), pp. 125–136. DOI: 10.1016/0092-8674(94)90579-7.
- Cohen, B.; McGuffin, M. E.; Pfeifle, C.; Segal, D.; Cohen, S. M. (1992): *apterous*, a gene required for imaginal disc development in *Drosophila* encodes a member of the LIM family of developmental regulatory proteins. In *Genes & Development* 6 (5), pp. 715–729. DOI: 10.1101/gad.6.5.715.
- Comoglio, Paolo M.; Tamagnone, Luca; Giordano, Silvia (2004): Invasive growth: a two-way street for semaphorin signalling. In *Nature cell biology* 6 (12), pp. 1155–1157. DOI: 10.1038/ncb1204-1155.
- Cyran, Shawn A.; Buchsbaum, Anna M.; Reddy, Karen L.; Lin, Meng-Chi; Glossop, Nicholas R.J.; Hardin, Paul E. et al. (2003): *vriille*, *Pdp1*, and *dClock* Form a Second Feedback Loop in the *Drosophila* Circadian Clock. In *Cell* 112 (3), pp. 329–341. DOI: 10.1016/S0092-8674(03)00074-6.
- Czerny, Thomas; Halder, Georg; Kloter, Urs; Souabni, Abdallah; Gehring, Walter J.; Busslinger, Meinrad (1999): *twin of eyeless*, a Second Pax-6 Gene of *Drosophila*, Acts Upstream of *eyeless* in the Control of Eye Development. In *Molecular Cell* 3 (3), pp. 297–307. DOI: 10.1016/S1097-2765(00)80457-8.
- Darmanis, Spyros; Sloan, Steven A.; Zhang, Ye; Enge, Martin; Caneda, Christine; Shuer, Lawrence M. et al. (2015): A survey of human brain transcriptome diversity at the single cell level. In *Proceedings of the National Academy of Sciences of the United States of America* 112 (23), pp. 7285–7290. DOI: 10.1073/pnas.1507125112.

- Davie, Kristofer; Janssens, Jasper; Koldere, Duygu; Waegeneer, Maxime de; Pech, Uli; Kreft, Łukasz et al. (2018): A Single-Cell Transcriptome Atlas of the Aging *Drosophila* Brain. In *Cell* 174 (4), 982-998.e20. DOI: 10.1016/j.cell.2018.05.057.
- Del Alamo, David; Mlodzik, Marek (2008): Self-modulation of Notch signaling during ommatidial development via the Roughened eye transcriptional repressor. In *Development (Cambridge, England)* 135 (17), pp. 2895–2904. DOI: 10.1242/dev.022194.
- Delon, Isabelle; Chanut-Delalande, H el ene; Payre, Fran ois (2003): The Ovo/Shavenbaby transcription factor specifies actin remodelling during epidermal differentiation in *Drosophila*. In *Mechanisms of Development* 120 (7), pp. 747–758. DOI: 10.1016/s0925-4773(03)00081-9.
- Denisenko, Elena; Guo, Belinda B.; Jones, Matthew; Hou, Rui; Kock, Leanne de; Lassmann, Timo et al. (2020): Systematic assessment of tissue dissociation and storage biases in single-cell and single-nucleus RNA-seq workflows. In *Genome Biology* 21 (1), p. 130. DOI: 10.1186/s13059-020-02048-6.
- Ding, Jiarui; Adiconis, Xian; Simmons, Sean K.; Kowalczyk, Monika S.; Hession, Cynthia C.; Marjanovic, Nemanja D. et al. (2020): Systematic comparison of single-cell and single-nucleus RNA-sequencing methods. In *Nature biotechnology* 38 (6), pp. 737–746. DOI: 10.1038/s41587-020-0465-8.
- Dittrich, R.; Bossing, T.; Gould, A. P.; Technau, G. M.; Urban, J. (1997): The differentiation of the serotonergic neurons in the *Drosophila* ventral nerve cord depends on the combined function of the zinc finger proteins Eagle and Hucklebein. In *Development (Cambridge, England)* 124 (13), pp. 2515–2525. DOI: 10.1242/dev.124.13.2515.
- Domingos, Pedro M.; Brown, Samara; Barrio, Rosa; Ratnakumar, Kajan; Frankfort, Benjamin J.; Mardon, Graeme et al. (2004): Regulation of R7 and R8 differentiation by the spalt genes. In *DEVELOPMENTAL BIOLOGY* 273 (1), pp. 121–133. DOI: 10.1016/j.ydbio.2004.05.026.
- Dong, P. D. Si; Dicks, Jennifer Scholz; Panganiban, Grace (2002): Distal-less and homothorax regulate multiple targets to pattern the *Drosophila* antenna. In *Development (Cambridge, England)* 129 (8), pp. 1967–1974. DOI: 10.1242/dev.129.8.1967.
- Ellis, M. C.; O'Neill, E. M.; Rubin, G. M. (1993): Expression of *Drosophila* glass protein and evidence for negative regulation of its activity in non-neuronal cells by another DNA-binding protein. In *Development (Cambridge, England)* 119 (3), pp. 855–865. DOI: 10.1242/dev.119.3.855.
- Emerald, B. Starling; Curtiss, Jennifer; Mlodzik, Marek; Cohen, Stephen M. (2003): Distal antenna and distal antenna related encode nuclear proteins containing pipsqueak motifs involved in antenna development in *Drosophila*. In *Development (Cambridge, England)* 130 (6), pp. 1171–1180. DOI: 10.1242/dev.00323.
- Estella, Carlos; Mann, Richard S. (2010): Non-redundant selector and growth-promoting functions of two sister genes, buttonhead and Sp1, in *Drosophila* leg development. In *PLOS Genetics* 6 (6), e1001001. DOI: 10.1371/journal.pgen.1001001.
- Evanko, Daniel (2004): Hybridization chain reaction. In *Nature methods* 1 (3), p. 186. DOI: 10.1038/nmeth1204-186a.
- Everett, Logan J.; Huang, Wen; Zhou, Shanshan; Carbone, Mary Anna; Lyman, Richard F.; Arya, Gunjan H. et al. (2020): Gene expression networks in the *Drosophila* Genetic Reference Panel. In *Genome research* 30 (3), pp. 485–496. DOI: 10.1101/gr.257592.119.
- Fernandes, Isabelle; Chanut-Delalande, H el ene; Ferrer, Pierre; Latapie, Yvan; Waltzer, Lucas; Affolter, Markus et al. (2010): Zona pellucida domain proteins remodel the apical

- compartment for localized cell shape changes. In *Developmental Cell* 18 (1), pp. 64–76. DOI: 10.1016/j.devcel.2009.11.009.
- Fichelson, Pierre; Brigui, Amira; Pichaud, Franck (2012): Orthodenticle and Kruppel homolog 1 regulate Drosophila photoreceptor maturation. In *Proceedings of the National Academy of Sciences of the United States of America* 109 (20), pp. 7893–7898. DOI: 10.1073/pnas.1120276109.
- Fincher, Christopher T.; Wurtzel, Omri; Hoog, Thom de; Kravarik, Kellie M.; Reddien, Peter W. (2018): Cell type transcriptome atlas for the planarian *Schmidtea mediterranea*. In *Science (New York, N.Y.)* 360 (6391). DOI: 10.1126/science.aaq1736.
- Fox, R.; Lehmkuhle, S. W.; Westendorf, D. H. (1976): Falcon visual acuity. In *Science (New York, N.Y.)* 192 (4236), pp. 263–265. DOI: 10.1126/science.1257767.
- Frankfort, Benjamin J.; Nolo, Riitta; Zhang, Zhihuan; Bellen, Hugo; Mardon, Graeme (2001): senseless Repression of rough Is Required for R8 Photoreceptor Differentiation in the Developing Drosophila Eye. In *Neuron* 32 (3), pp. 403–414. DOI: 10.1016/s0896-6273(01)00480-9.
- Fu, Weimin; Baker, Nicholas E. (2003): Deciphering synergistic and redundant roles of Hedgehog, Decapentaplegic and Delta that drive the wave of differentiation in Drosophila eye development. In *Development (Cambridge, England)* 130 (21), pp. 5229–5239. DOI: 10.1242/dev.00764.
- Galloway, Alison; Cowling, Victoria H. (2019): mRNA cap regulation in mammalian cell function and fate. In *Biochimica et biophysica acta. Gene regulatory mechanisms* 1862 (3), pp. 270–279. DOI: 10.1016/j.bbagr.2018.09.011.
- Gaspar, Pedro; Arif, Saad; Sumner-Rooney, Lauren; Kittelmann, Maike; Bodey, Andrew J.; Stern, David L. et al. (2020): Characterization of the Genetic Architecture Underlying Eye Size Variation Within *Drosophila melanogaster* and *Drosophila simulans*. In *G3 (Bethesda, Md.)* 10 (3), pp. 1005–1018. DOI: 10.1534/g3.119.400877.
- Gaudet, Pascale; Livstone, Michael S.; Lewis, Suzanna E.; Thomas, Paul D. (2011): Phylogenetic-based propagation of functional annotations within the Gene Ontology consortium. In *Briefings in bioinformatics* 12 (5), pp. 449–462. DOI: 10.1093/bib/bbr042.
- Gehring, Niels H.; Roignant, Jean-Yves (2021): Anything but Ordinary - Emerging Splicing Mechanisms in Eukaryotic Gene Regulation. In *Trends in genetics : TIG* 37 (4), pp. 355–372. DOI: 10.1016/j.tig.2020.10.008.
- Gehring, Walter J. (2014): The evolution of vision. In *Wiley interdisciplinary reviews. Developmental biology* 3 (1), pp. 1–40. DOI: 10.1002/wdev.96.
- Graf, Ethan R.; Daniels, Richard W.; Burgess, Robert W.; Schwarz, Thomas L.; DiAntonio, Aaron (2009): Rab3 dynamically controls protein composition at active zones. In *Neuron* 64 (5), pp. 663–677. DOI: 10.1016/j.neuron.2009.11.002.
- Gramates, L. Sian; Agapite, Julie; Attrill, Helen; Calvi, Brian R.; Crosby, Madeline A.; Dos Santos, Gilberto et al. (2022): FlyBase: a guided tour of highlighted features. In *Genetics* 220 (4). DOI: 10.1093/genetics/iyac035.
- Grillenzoni, Nicola; Flandre, Adrien; Lasbleiz, Christelle; Dura, Jean-Maurice (2007): Respective roles of the DRL receptor and its ligand WNT5 in *Drosophila* mushroom body development. In *Development (Cambridge, England)* 134 (17), pp. 3089–3097. DOI: 10.1242/dev.02876.
- Grindberg, Rashed V.; Yee-Greenbaum, Joyclyn L.; McConnell, Michael J.; Novotny, Mark; O’Shaughnessy, Andy L.; Lambert, Georgina M. et al. (2013): RNA-sequencing from single

- nuclei. In *Proceedings of the National Academy of Sciences of the United States of America* 110 (49), pp. 19802–19807. DOI: 10.1073/pnas.1319700110.
- Grubbs, Nathaniel; Leach, Megan; Su, Xin; Petrisko, Tiffany; Rosario, Juan B.; Mahaffey, James W. (2013): New components of *Drosophila* leg development identified through genome wide association studies. In *PLoS one* 8 (4), e60261. DOI: 10.1371/journal.pone.0060261.
- Guo, Pengfei; Lee, Chang-Hyun; Lei, Huiyan; Zheng, Yonggang; Pulgar Prieto, Katuska Daniela; Pan, Duoia (2019): Nerfin-1 represses transcriptional output of Hippo signaling in cell competition. In *eLife* 8. DOI: 10.7554/eLife.38843.
- Habib, Naomi; Li, Yinqing; Heidenreich, Matthias; Swiech, Lukasz; Avraham-Davidi, Inbal; Trombetta, John J. et al. (2016): Div-Seq: Single-nucleus RNA-Seq reveals dynamics of rare adult newborn neurons. In *Science (New York, N.Y.)* 353 (6302), pp. 925–928. DOI: 10.1126/science.aad7038.
- Hadley, Wickham (2016): Ggplot2. Elegrant graphics for data analysis. Second edition. Switzerland: Springer (Use R!).
- Hallson, Graham; Syrzycka, Monika; Beck, Samantha A.; Kennison, James A.; Dorsett, Dale; Page, Scott L. et al. (2008): The *Drosophila* cohesin subunit Rad21 is a trithorax group (trxG) protein. In *Proceedings of the National Academy of Sciences of the United States of America* 105 (34), pp. 12405–12410. DOI: 10.1073/pnas.0801698105.
- Hamanaka, Yoshitaka; Meinertzhagen, Ian A. (2010): Immunocytochemical localization of synaptic proteins to photoreceptor synapses of *Drosophila melanogaster*. In *The Journal of comparative neurology* 518 (7), pp. 1133–1155. DOI: 10.1002/cne.22268.
- Hammonds, Ann S.; Bristow, Christopher A.; Fisher, William W.; Weiszmam, Richard; Wu, Siqi; Hartenstein, Volker et al. (2013): Spatial expression of transcription factors in *Drosophila* embryonic organ development. In *Genome Biology* 14 (12), R140. DOI: 10.1186/gb-2013-14-12-r140.
- Haug-Collet, K.; Pearson, B.; Webel, R.; Szerencsei, R. T.; Winkfein, R. J.; Schnetkamp, P. P.; Colley, N. J. (1999): Cloning and characterization of a potassium-dependent sodium/calcium exchanger in *Drosophila*. In *The Journal of cell biology* 147 (3), pp. 659–670. DOI: 10.1083/jcb.147.3.659.
- Haynie, J. L.; Bryant, P. J. (1986): Development of the eye-antenna imaginal disc and morphogenesis of the adult head in *Drosophila melanogaster*. In *Journal of Experimental Zoology* 237 (3), pp. 293–308. DOI: 10.1002/jez.1402370302.
- He, Ying; Emoto, Kazuo; Fang, Xiaolan; Ren, Nan; Tian, Xiaojing; Jan, Yuh-Nung; Adler, Paul N. (2005): *Drosophila* Mob family proteins interact with the related tricornered (Trc) and warts (Wts) kinases. In *Molecular biology of the cell* 16 (9), pp. 4139–4152. DOI: 10.1091/mbc.e05-01-0018.
- Herboso, Leire; Oliveira, Marisa M.; Talamillo, Ana; Pérez, Coralia; González, Monika; Martín, David et al. (2015): Ecdysone promotes growth of imaginal discs through the regulation of Thor in *D. melanogaster*. In *Scientific Reports* 5, p. 12383. DOI: 10.1038/srep12383.
- Hilbrant, Maarten; Almudi, Isabel; Leite, Daniel J.; Kuncheria, Linta; Posnien, Nico; Nunes, Maria D. S.; McGregor, Alistair P. (2014): Sexual dimorphism and natural variation within and among species in the *Drosophila* retinal mosaic. In *BMC evolutionary biology* 14, p. 240. DOI: 10.1186/s12862-014-0240-x.
- Hirota, Yuki; Okabe, Masataka; Imai, Takao; Kurusu, Mitsuhiro; Yamamoto, Atsuyo; Miyao, Sachiyo et al. (1999): Musashi and Seven in absentia downregulate Tramtrack through distinct mechanisms in *Drosophila* eye development. In *Mechanisms of Development* 87 (1-2), pp. 93–101. DOI: 10.1016/s0925-4773(99)00143-4.

- Hodges, G. M.; Livingston, D. C.; Franks, L. M. (1973): The localization of trypsin in cultured mammalian cells. In *Journal of cell science* 12 (3), pp. 887–902. DOI: 10.1242/jcs.12.3.887.
- Hoffman, Paul (2021): SeuratDisk. Interfaces for HDF5-Based Single Cell File Formats. Version 0.0.0.9019. Available online at <https://mojaveazure.github.io/seurat-disk/>.
- Huang, Hsiang-Ling; Hsing, Hsiang-Wei; Lai, Tzu-Chia; Chen, Yi-Wen; Lee, Tian-Ren; Chan, Hsin-Tsu et al. (2010): Trypsin-induced proteome alteration during cell subculture in mammalian cells. In *Journal of biomedical science* 17, p. 36. DOI: 10.1186/1423-0127-17-36.
- Islam, Saiful; Kjällquist, Una; Moliner, Annalena; Zajac, Pawel; Fan, Jian-Bing; Lönnerberg, Peter; Linnarsson, Sten (2011): Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. In *Genome research* 21 (7), pp. 1160–1167. DOI: 10.1101/gr.110882.110.
- Jang, Chuen-Chuen; Chao, Ju-Lan; Jones, Nikolas; Yao, Li-Chin; Bessarab, Dmitri A.; Kuo, Yien M. et al. (2003): Two Pax genes, eye gone and eyeless, act cooperatively in promoting Drosophila eye development. In *Development (Cambridge, England)* 130 (13), pp. 2939–2951. DOI: 10.1242/dev.00522.
- Jeffery, W. R. (2001): Cavefish as a model system in evolutionary developmental biology. In *DEVELOPMENTAL BIOLOGY* 231 (1), pp. 1–12. DOI: 10.1006/dbio.2000.0121.
- Jeffrey A. Fabrick, J. A.; Hull, J. J. (2017): Assessing Integrity of Insect RNA: Application Note. Available online at <https://www.ars.usda.gov/ARSUserFiles/20200500/Pubs%202018/FabrickHull%20AgilentNot4es2017.pdf>, checked on 4/28/2022.
- Ji, Zhejun; Kiparaki, Marianthi; Folgado, Virginia; Kumar, Amit; Blanco, Jorge; Rimesso, Gerard et al. (2019): Drosophila RpS12 controls translation, growth, and cell competition through Xrp1. In *PLOS Genetics* 15 (12), e1008513. DOI: 10.1371/journal.pgen.1008513.
- Jongbloets, Bart C.; Pasterkamp, R. Jeroen (2014): Semaphorin signalling during development. In *Development (Cambridge, England)* 141 (17), pp. 3292–3297. DOI: 10.1242/dev.105544.
- Kania, Arthur; Han, Pyung-Lim; Kim, Yun-Talk; Bellen, Hugo (1993): neuromusculin, a drosophila gene expressed in peripheral neuronal precursors and muscles, encodes a cell adhesion molecule. In *Neuron* 11 (4), pp. 673–687. DOI: 10.1016/0896-6273(93)90078-6.
- Keeseey, Ian W.; Grabe, Veit; Gruber, Lydia; Koerte, Sarah; Obiero, George F.; Bolton, Grant et al. (2019): Inverse resource allocation between vision and olfaction across the genus Drosophila. In *Nature Communications* 10 (1), p. 1162. DOI: 10.1038/s41467-019-09087-z.
- Khan, Sumbul Jawed; Abidi, Syeda Nayab Fatima; Tian, Yuan; Skinner, Andrea; Smith-Bolton, Rachel K. (2016): A rapid, gentle and scalable method for dissociation and fluorescent sorting of imaginal disc cells for mRNA sequencing. In *Fly* 10 (2), pp. 73–80. DOI: 10.1080/19336934.2016.1173296.
- Kim, Sabrina Y.; Renihan, Maia K.; Boulianne, Gabrielle L. (2006): Characterization of big bang, a novel gene encoding for PDZ domain-containing proteins that are dynamically expressed throughout Drosophila development. In *Gene expression patterns : GEP* 6 (5), pp. 504–518. DOI: 10.1016/j.modgep.2005.10.009.
- King-Jones, Kirst; Charles, Jean-Philippe; Lam, Geanette; Thummel, Carl S. (2005): The ecdysone-induced DHR4 orphan nuclear receptor coordinates growth and maturation in Drosophila. In *Cell* 121 (5), pp. 773–784. DOI: 10.1016/j.cell.2005.03.030.

- Klimovich, Alexander; Giacomello, Stefania; Björklund, Åsa; Faure, Louis; Kaucka, Marketa; Giez, Christoph et al. (2020): Prototypical pacemaker neurons interact with the resident microbiota. In *Proceedings of the National Academy of Sciences of the United States of America* 117 (30), pp. 17854–17863. DOI: 10.1073/pnas.1920469117.
- Kolodziejczyk, Aleksandra A.; Kim, Jong Kyoung; Svensson, Valentine; Marioni, John C.; Teichmann, Sarah A. (2015): The technology and biology of single-cell RNA sequencing. In *Molecular Cell* 58 (4), pp. 610–620. DOI: 10.1016/j.molcel.2015.04.005.
- Kozlova, T.; Pokholkova, G. V.; Tzertzinis, G.; Sutherland, J. D.; Zhimulev, I. F.; Kafatos, F. C. (1998): Drosophila hormone receptor 38 functions in metamorphosis: a role in adult cuticle formation. In *Genetics* 149 (3), pp. 1465–1475. DOI: 10.1093/genetics/149.3.1465.
- Krishnaswami, Suguna Rani; Grindberg, Rashel V.; Novotny, Mark; Venepally, Pratap; Lacar, Benjamin; Bhutani, Kunal et al. (2016): Using single nuclei for RNA-seq to capture the transcriptome of postmortem neurons. In *Nature Protocols* 11 (3), pp. 499–524. DOI: 10.1038/nprot.2016.015.
- Kumar, Justin P. (2018): The fly eye: Through the looking glass. In *Developmental Dynamics* 247 (1), pp. 111–123. DOI: 10.1002/dvdy.24585.
- Kumar, Justin P.; Jamal, Tazeen; Doetsch, Alex; Turner, F. Rudolf; Duffy, Joseph B. (2004): CREB binding protein functions during successive stages of eye development in Drosophila. In *Genetics* 168 (2), pp. 877–893. DOI: 10.1534/genetics.104.029850.
- Kumar, Manu P.; Du, Jinyan; Lagoudas, Georgia; Jiao, Yang; Sawyer, Andrew; Drummond, Daryl C. et al. (2018): Analysis of Single-Cell RNA-Seq Identifies Cell-Cell Communication Associated with Tumor Characteristics. In *Cell Reports* 25 (6), 1458-1468.e4. DOI: 10.1016/j.celrep.2018.10.047.
- Kuzin, Alexander; Brody, Thomas; Moore, Adrian W.; Odenwald, Ward F. (2005): Nerfin-1 is required for early axon guidance decisions in the developing Drosophila CNS. In *DEVELOPMENTAL BIOLOGY* 277 (2), pp. 347–365. DOI: 10.1016/j.ydbio.2004.09.027.
- Lake, Blue B.; Chen, Song; Sos, Brandon C.; Fan, Jean; Kaeser, Gwendolyn E.; Yung, Yun C. et al. (2018): Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. In *Nature biotechnology* 36 (1), pp. 70–80. DOI: 10.1038/nbt.4038.
- Lake, Blue B.; Codeluppi, Simone; Yung, Yun C.; Gao, Derek; Chun, Jerold; Kharchenko, Peter V. et al. (2017): A comparative strategy for single-nucleus and single-cell transcriptomes confirms accuracy in predicted cell-type expression from nuclear RNA. In *Scientific Reports* 7 (1), p. 6031. DOI: 10.1038/s41598-017-04426-w.
- Lapan, Sylvain W.; Reddien, Peter W. (2012): Transcriptome analysis of the planarian eye identifies ovo as a specific regulator of eye regeneration. In *Cell Reports* 2 (2), pp. 294–307. DOI: 10.1016/j.celrep.2012.06.018.
- Lee, Gyunghee; Hall, Jeffrey C.; Park, Jae H. (2002): Doublesex gene expression in the central nervous system of Drosophila melanogaster. In *Journal of neurogenetics* 16 (4), pp. 229–248. DOI: 10.1080/01677060216292.
- Lee, Seung-Jae; Xu, Hong; Montell, Craig (2004): Rhodopsin kinase activity modulates the amplitude of the visual response in Drosophila. In *Proceedings of the National Academy of Sciences* 101 (32), pp. 11874–11879. DOI: 10.1073/pnas.0402205101.
- Lewis, E. B. (2004): A Gene Complex Controlling Segmentation in Drosophila. In Howard D. Lipshitz (Ed.): *Genes, Development and Cancer*. Boston, MA: Springer US, pp. 205–217.
- Li, Hongjie; Janssens, Jasper; Waegeneer, Maxime de; Kolluru, Sai Saroja; Davie, Kristofer; Gardeux, Vincent et al. (2022): Fly Cell Atlas: A single-nucleus transcriptomic atlas of the

- adult fruit fly. In *Science (New York, N.Y.)* 375 (6584), eabk2432. DOI: 10.1126/science.abk2432.
- Litvinukova, Monika; Lindberg, Eric; Maatz, Henrike; Zhang, Hongbo; Radke, Michael; Gotthardt, Michael et al. (2018): Single Cell and Single Nuclei Analysis Human Heart Tissue. In *protocols.io*. DOI: 10.17504/protocols.io.veae3ae.
- Liu, Ling-Yu; Lin, Cheng-Han; Fan, Seng-Sheen (2009): Function of *Drosophila* mob2 in photoreceptor morphogenesis. In *Cell and tissue research* 338 (3), pp. 377–389. DOI: 10.1007/s00441-009-0878-7.
- Liu, Tianbin; Li, Jie; Yu, Leqian; Sun, Hai-Xi; Li, Jing; Dong, Guoyi et al. (2021): Cross-species single-cell transcriptomic analysis reveals pre-gastrulation developmental differences among pigs, monkeys, and humans. In *Cell discovery* 7 (1), p. 8. DOI: 10.1038/s41421-020-00238-x.
- Luo, L. Q.; Le Martin-Morris; White, K. (1990): Identification, secretion, and neural expression of APPL, a *Drosophila* protein similar to human amyloid protein precursor. In *The Journal of Neuroscience* 10 (12), pp. 3849–3861. DOI: 10.1523/JNEUROSCI.10-12-03849.1990.
- Macosko, Evan Z.; Basu, Anindita; Satija, Rahul; Nemesh, James; Shekhar, Karthik; Goldman, Melissa et al. (2015): Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. In *Cell* 161 (5), pp. 1202–1214. DOI: 10.1016/j.cell.2015.05.002.
- Mallarino, Ricardo; Campàs, Otger; Fritz, Joerg A.; Burns, Kevin J.; Weeks, Olivia G.; Brenner, Michael P.; Abzhanov, Arhat (2012): Closely related bird species demonstrate flexibility between beak morphology and underlying developmental programs. In *Proceedings of the National Academy of Sciences of the United States of America* 109 (40), pp. 16222–16227. DOI: 10.1073/pnas.1206205109.
- Mao, Yanlan; Freeman, Matthew (2009): Fasciclin 2, the *Drosophila* orthologue of neural cell-adhesion molecule, inhibits EGF receptor signalling. In *Development (Cambridge, England)* 136 (3), pp. 473–481. DOI: 10.1242/dev.026054.
- Matthews, K. A.; Miller, D. F.; Kaufman, T. C. (1989): Developmental distribution of RNA and protein products of the *Drosophila* alpha-tubulin gene family. In *DEVELOPMENTAL BIOLOGY* 132 (1), pp. 45–61. DOI: 10.1016/0012-1606(89)90203-0.
- Mazzoni, Esteban O.; Celik, Arzu; Wernet, Mathias F.; Vasiliaskas, Daniel; Johnston, Robert J.; Cook, Tiffany A. et al. (2008): Iroquois complex genes induce co-expression of rhodopsins in *Drosophila*. In *PLoS biology* 6 (4), e97. DOI: 10.1371/journal.pbio.0060097.
- McClure, Kimberly D.; Schubiger, Gerold (2005): Developmental analysis and squamous morphogenesis of the peripodial epithelium in *Drosophila* imaginal discs. In *Development (Cambridge, England)* 132 (22), pp. 5033–5042. DOI: 10.1242/dev.02092.
- McInnes, Leland; Healy, John; Melville, James (2020): UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. In *arXiv*. DOI: 10.48550/arXiv.1802.03426.
- Mével-Ninio, M.; Terracol, R.; Kafatos, F. C. (1991): The ovo gene of *Drosophila* encodes a zinc finger protein required for female germ line development. In *The EMBO journal* 10 (8), pp. 2259–2266. DOI: 10.1002/j.1460-2075.1991.tb07762.x.
- Mével-Ninio, Maryvonne; Terracol, Régine; Salles, Catherine; Vincent, Alain; Payre, François (1995): ovo, a *Drosophila* gene required for ovarian development, is specifically expressed in the germline and shares most of its coding sequences with shavenbaby, a gene involved in embryo patterning. In *Mechanisms of Development* 49 (1-2), pp. 83–95. DOI: 10.1016/0925-4773(94)00305-7.

- Miller, David; Stegmann, Robert (1991): The Eye of the Eagle. In *European Journal of Implant and Refractive Surgery* 3 (1), pp. 71–73. DOI: 10.1016/S0955-3681(13)80155-4.
- Mittleman, Briana E.; Pott, Sebastian; Warland, Shane; Zeng, Tony; Mu, Zepeng; Kaur, Mayher et al. (2020): Alternative polyadenylation mediates genetic regulation of gene expression. In *eLife* 9. DOI: 10.7554/eLife.57492.
- Mollereau, B.; Dominguez, M.; Webel, R.; Colley, N. J.; Keung, B.; Celis, J. F. de; Desplan, C. (2001): Two-step process for photoreceptor formation in *Drosophila*. In *Nature* 412 (6850), pp. 911–913. DOI: 10.1038/35091076.
- Morgan, T. H. (1910): Sex Limited Inheritance in *Drosophila*. In *Science (New York, N.Y.)* 32 (812), pp. 120–122. DOI: 10.1126/science.32.812.120.
- Morgan, T. H. (1911): An attempt to analyze the constitution of the chromosomes on the basis of sex-limited inheritance in *Drosophila*. In *Journal of Experimental Zoology* 11 (4), pp. 365–413. DOI: 10.1002/jez.1400110404.
- Musser, Jacob M.; Schippers, Klaske J.; Nickel, Michael; Mizzon, Giulia; Kohn, Andrea B.; Pape, Constantin et al. (2021): Profiling cellular diversity in sponges informs animal cell type and nervous system evolution. In *Science (New York, N.Y.)* 374 (6568), pp. 717–723. DOI: 10.1126/science.abj2949.
- Neto, Marta; Naval-Sánchez, Marina; Potier, Delphine; Pereira, Paulo S.; Geerts, Dirk; Aerts, Stein; Casares, Fernando (2017): Nuclear receptors connect progenitor transcription factors to cell cycle control. In *Scientific Reports* 7 (1), p. 4845. DOI: 10.1038/s41598-017-04936-7.
- Nguyen, Duc N.T.; Rohrbaugh, Margaret; Lai, Zhi-Chun (2000): The *Drosophila* homolog of *Onecut* homeodomain proteins is a neural-specific transcriptional activator with a potential role in regulating neural differentiation. In *Mechanisms of Development* 97 (1-2), pp. 57–72. DOI: 10.1016/s0925-4773(00)00431-7.
- Nojima, Satoshi; Toyofuku, Toshihiko; Kamao, Hiroyuki; Ishigami, Chie; Kaneko, Jun; Okuno, Tatsusada et al. (2013): A point mutation in Semaphorin 4A associates with defective endosomal sorting and causes retinal degeneration. In *Nature Communications* 4, p. 1406. DOI: 10.1038/ncomms2420.
- Nolo, Riitta; Abbott, Lois A.; Bellen, Hugo J. (2000): Senseless, a Zn Finger Transcription Factor, Is Necessary and Sufficient for Sensory Organ Development in *Drosophila*. In *Cell* 102 (3), pp. 349–362. DOI: 10.1016/s0092-8674(00)00040-4.
- Nolte, Viola; Pandey, Ram Vinay; Kofler, Robert; Schlötterer, Christian (2013): Genome-wide patterns of natural variation reveal strong selective sweeps and ongoing genomic conflict in *Drosophila mauritiana*. In *Genome research* 23 (1), pp. 99–110. DOI: 10.1101/gr.139873.112.
- Norry, Fabian M.; Gomez, Federico H. (2017): Quantitative Trait Loci and Antagonistic Associations for Two Developmentally Related Traits in the *Drosophila* Head. In *Journal of insect science (Online)* 17 (1). DOI: 10.1093/jisesa/iew115.
- Nüsslein-Volhard, C.; Wieschaus, E. (1980): Mutations affecting segment number and polarity in *Drosophila*. In *Nature* 287 (5785), pp. 795–801. DOI: 10.1038/287795a0.
- O'Neill, Elizabeth M.; Rebay, Ilaria; Tjian, Robert; Rubin, Gerald M. (1994): The activities of two *Ets*-related transcription factors required for *drosophila* eye development are modulated by the Ras/MAPK pathway. In *Cell* 78 (1), pp. 137–147. DOI: 10.1016/0092-8674(94)90580-0.

- Oros, Sarah M.; Tare, Meghana; Kango-Singh, Madhuri; Singh, Amit (2010): Dorsal eye selector pannier (pnr) suppresses the eye fate to define dorsal margin of the Drosophila eye. In *DEVELOPMENTAL BIOLOGY* 346 (2), pp. 258–271. DOI: 10.1016/j.ydbio.2010.07.030.
- Otte, Stefan; Barnikol-Watanabe, Shitsu; Vorbrüggen, Gerd; Hilschmann, Norbert (1999): NUCB1, the Drosophila melanogaster homolog of the mammalian EF-hand proteins NEFA and nucleobindin. In *Mechanisms of Development* 86 (1-2), pp. 155–158. DOI: 10.1016/S0925-4773(99)00103-3.
- Owald, David; Fouquet, Wernher; Schmidt, Manuela; Wichmann, Carolin; Mertel, Sara; Depner, Harald et al. (2010): A Syd-1 homologue regulates pre- and postsynaptic maturation in Drosophila. In *The Journal of cell biology* 188 (4), pp. 565–579. DOI: 10.1083/jcb.200908055.
- Özel, Mehmet Neset; Simon, Félix; Jafari, Shadi; Holguera, Isabel; Chen, Yen-Chung; Benhra, Najate et al. (2021): Neuronal diversity and convergence in a visual system developmental atlas. In *Nature* 589 (7840), pp. 88–95. DOI: 10.1038/s41586-020-2879-3.
- Pai, C. Y.; Kuo, T. S.; Jaw, T. J.; Kurant, E.; Chen, C. T.; Bessarab, D. A. et al. (1998): The Homothorax homeoprotein activates the nuclear localization of another homeoprotein, extradenticle, and suppresses eye development in Drosophila. In *Genes & Development* 12 (3), pp. 435–446. DOI: 10.1101/gad.12.3.435.
- Paine-Saunders, Stephenie; Fristrom, Dianne; Fristrom, James W. (1990): The Drosophila IMP-E2 gene encodes an apically secreted protein expressed during imaginal disc morphogenesis. In *DEVELOPMENTAL BIOLOGY* 140 (2), pp. 337–351. DOI: 10.1016/0012-1606(90)90084-v.
- Palazzo, Alexander F.; Lee, Eliza S. (2018): Sequence Determinants for Nuclear Retention and Cytoplasmic Export of mRNAs and lncRNAs. In *Frontiers in genetics* 9, p. 440. DOI: 10.3389/fgene.2018.00440.
- Pappenheimer, A. M. (1917): EXPERIMENTAL STUDIES UPON LYMPHOCYTES : I. THE REACTIONS OF LYMPHOCYTES UNDER VARIOUS EXPERIMENTAL CONDITIONS. In *The Journal of experimental medicine* 25 (5), pp. 633–650. DOI: 10.1084/jem.25.5.633.
- Payre, F.; Vincent, A.; Carreno, S. (1999): ovo/svb integrates Wingless and DER pathways to control epidermis differentiation. In *Nature* 400 (6741), pp. 271–275. DOI: 10.1038/22330.
- Pichaud, Franck; Casares, Fernando (2000): homothorax and iroquois-C genes are required for the establishment of territories within the developing eye disc. In *Mechanisms of Development* 96 (1), pp. 15–25. DOI: 10.1016/S0925-4773(00)00372-5.
- Pichaud, Franck; Casares, Fernando (2022): Shaping an optical dome: The size and shape of the insect compound eye. In *Seminars in cell & developmental biology* 130, pp. 37–44. DOI: 10.1016/j.semcd.2021.11.002.
- Plass, Mireya; Solana, Jordi; Wolf, F. Alexander; Ayoub, Salah; Misios, Aristotelis; Glažar, Petar et al. (2018): Cell type atlas and lineage tree of a whole complex animal by single-cell transcriptomics. In *Science (New York, N.Y.)* 360 (6391). DOI: 10.1126/science.aag1723.
- Posnien, Nico; Hopfen, Corinna; Hilbrant, Maarten; Ramos-Womack, Margarita; Murat, Sophie; Schönauer, Anna et al. (2012): Evolution of eye morphology and rhodopsin expression in the Drosophila melanogaster species subgroup. In *PloS one* 7 (5), e37346. DOI: 10.1371/journal.pone.0037346.
- Raivo Kolde (2019): pheatmap. Pretty Heatmaps. Version 1.0.12. Available online at <https://CRAN.R-project.org/package=pheatmap>.

- Ramaekers, Ariane; Claeys, Annelies; Kapun, Martin; Mouchel-Vielh, Emmanuèle; Potier, Delphine; Weinberger, Simon et al. (2019): Altering the Temporal Regulation of One Transcription Factor Drives Evolutionary Trade-Offs between Head Sensory Organs. In *Developmental Cell* 50 (6), 780-792.e7. DOI: 10.1016/j.devcel.2019.07.027.
- Rawlins, Emma L.; White, Neil M.; Jarman, Andrew P. (2003): Echinoid limits R8 photoreceptor specification by inhibiting inappropriate EGF receptor signalling within R8 equivalence groups. In *Development (Cambridge, England)* 130 (16), pp. 3715–3724. DOI: 10.1242/dev.00602.
- Reis, Micael; Wiegler, Gordon; Claude, Julien; Lata, Rodrigo; Horchler, Britta; Ha, Ngoc-Thuy et al. (2020): Multiple loci linked to inversions are associated with eye size variation in species of the *Drosophila virilis* phylad. In *Scientific Reports* 10 (1), p. 12832. DOI: 10.1038/s41598-020-69719-z.
- Rideout, Elizabeth J.; Dornan, Anthony J.; Neville, Megan C.; Eadie, Suzanne; Goodwin, Stephen F. (2010): Control of sexual differentiation and behavior by the doublesex gene in *Drosophila melanogaster*. In *Nature neuroscience* 13 (4), pp. 458–466. DOI: 10.1038/nn.2515.
- Roark, M.; Sturtevant, M. A.; Emery, J.; Vaessin, H.; Grell, E.; Bier, E. (1995): scratch, a pan-neural gene encoding a zinc finger protein related to snail, promotes neuronal development. In *Genes & Development* 9 (19), pp. 2384–2398. DOI: 10.1101/gad.9.19.2384.
- Roignant, Jean-Yves; Legent, Kevin; Janody, Florence; Treisman, Jessica E. (2010): The transcriptional co-factor Chip acts with LIM-homeodomain proteins to set the boundary of the eye field in *Drosophila*. In *Development (Cambridge, England)* 137 (2), pp. 273–281. DOI: 10.1242/dev.041244.
- Rynes, Jan; Donohoe, Colin D.; Frommolt, Peter; Brodesser, Susanne; Jindra, Marek; Uhlirva, Mirka (2012): Activating transcription factor 3 regulates immune and metabolic homeostasis. In *Molecular and cellular biology* 32 (19), pp. 3949–3962. DOI: 10.1128/MCB.00429-12.
- Salazar-Ciudad, Isaac; Jernvall, Jukka (2010): A computational model of teeth and the developmental origins of morphological variation. In *Nature* 464 (7288), pp. 583–586. DOI: 10.1038/nature08838.
- Salz, Helen K.; Erickson, James W. (2010): Sex determination in *Drosophila*: The view from the top. In *Fly* 4 (1), pp. 60–70. DOI: 10.4161/fly.4.1.11277.
- Schindelin, Johannes; Arganda-Carreras, Ignacio; Frise, Erwin; Kaynig, Verena; Longair, Mark; Pietzsch, Tobias et al. (2012): Fiji: an open-source platform for biological-image analysis. In *Nature methods* 9 (7), pp. 676–682. DOI: 10.1038/nmeth.2019.
- Schulze, S.; Sinclair, D. A.; Silva, E.; Fitzpatrick, K. A.; Singh, M.; Lloyd, V. K. et al. (2001): Essential genes in proximal 3L heterochromatin of *Drosophila melanogaster*. In *Molecular & general genetics : MGG* 264 (6), pp. 782–789. DOI: 10.1007/s004380000367.
- Sebé-Pedrós, Arnau; Saudemont, Baptiste; Chomsky, Elad; Plessier, Flora; Mailhé, Marie-Pierre; Renno, Justine et al. (2018): Cnidarian Cell Type Diversity and Regulation Revealed by Whole-Organism Single-Cell RNA-Seq. In *Cell* 173 (6), 1520-1534.e20. DOI: 10.1016/j.cell.2018.05.019.
- Sekyrova, Petra; Bohmann, Dirk; Jindra, Marek; Uhlirva, Mirka (2010): Interaction between *Drosophila* bZIP proteins Atf3 and Jun prevents replacement of epithelial cells during metamorphosis. In *Development (Cambridge, England)* 137 (1), pp. 141–150. DOI: 10.1242/dev.037861.

- Seritrakul, Pawat; Gross, Jeffrey M. (2017): Tet-mediated DNA hydroxymethylation regulates retinal neurogenesis by modulating cell-extrinsic signaling pathways. In *PLOS Genetics* 13 (9), e1006987. DOI: 10.1371/journal.pgen.1006987.
- Shalek, Alex K.; Satija, Rahul; Adiconis, Xian; Gertner, Rona S.; Gaublomme, Jellert T.; Raychowdhury, Raktima et al. (2013): Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. In *Nature* 498 (7453), pp. 236–240. DOI: 10.1038/nature12172.
- Sheng, Lihong; Shields, Emily J.; Gospocic, Janko; Glastad, Karl M.; Ratchasanmuang, Puttachai; Berger, Shelley L. et al. (2020): Social reprogramming in ants induces longevity-associated glia remodeling. In *Science advances* 6 (34), eaba9869. DOI: 10.1126/sciadv.aba9869.
- Shumate, Alaina; Salzberg, Steven L. (2020): Liftoff: accurate mapping of gene annotations. In *Bioinformatics (Oxford, England)*. DOI: 10.1093/bioinformatics/btaa1016.
- Siddall, Nicole A.; Kalcina, Marina; Johanson, Timothy M.; Monk, Adrian C.; Casagrande, Franca; Been, Reeve P. et al. (2012): Drosophila Rbp6 is an orthologue of vertebrate Msi-1 and Msi-2, but does not function redundantly with dMsi to regulate germline stem cell behaviour. In *PloS one* 7 (11), e49810. DOI: 10.1371/journal.pone.0049810.
- Silies, Marion; Yuva, Yeliz; Engelen, Daniel; Aho, Annukka; Stork, Tobias; Klämbt, Christian (2007): Glial cell migration in the eye disc. In *The Journal of neuroscience : the official journal of the Society for Neuroscience* 27 (48), pp. 13130–13139. DOI: 10.1523/JNEUROSCI.3583-07.2007.
- Singh, Amit; Choi, Kwang-Wook (2003): Initial state of the Drosophila eye before dorsoventral specification is equivalent to ventral. In *Development (Cambridge, England)* 130 (25), pp. 6351–6360. DOI: 10.1242/dev.00864.
- Slyper, Michal; Porter, Caroline B. M.; Ashenberg, Orr; Waldman, Julia; Drokhlyansky, Eugene; Wakiro, Isaac et al. (2020): A single-cell and single-nucleus RNA-Seq toolbox for fresh and frozen human tumors. In *Nature medicine* 26 (5), pp. 792–802. DOI: 10.1038/s41591-020-0844-1.
- Snow, Christine; Allen, A. (1970): The release of radioactive nucleic acids and mucoproteins by trypsin and ethylenediaminetetra-acetate treatment of baby-hamster cells in tissue culture. In *Biochemical Journal* 119 (4), pp. 707–714. DOI: 10.1042/bj1190707.
- Snyder, A. W.; Miller, W. H. (1978): Telephoto lens system of falconiform eyes. In *Nature* 275 (5676), pp. 127–129. DOI: 10.1038/275127a0.
- St Pierre, Susan E.; Galindo, Maximo I.; Couso, Juan P.; Thor, Stefan (2002): Control of Drosophila imaginal disc development by rotund and roughened eye: differentially expressed transcripts of the same gene encoding functionally distinct zinc finger proteins. In *Development (Cambridge, England)* 129 (5), pp. 1273–1281. DOI: 10.1242/dev.129.5.1273.
- Stathopoulos, Angelike; van Drenth, Madeleine; Erives, Albert; Markstein, Michele; Levine, Michael (2002): Whole-Genome Analysis of Dorsal-Ventral Patterning in the Drosophila Embryo. In *Cell* 111 (5), pp. 687–701. DOI: 10.1016/s0092-8674(02)01087-5.
- Stewart, M. J.; Denell, R. (1993): Mutations in the Drosophila gene encoding ribosomal protein S6 cause tissue overgrowth. In *Molecular and cellular biology* 13 (4), pp. 2524–2535. DOI: 10.1128/mcb.13.4.2524-2535.1993.
- Stone, Bryan L.; Thummel, Carl S. (1993): The Drosophila 78C early late puff contains E78, an ecdysone-inducible gene that encodes a novel member of the nuclear hormone receptor superfamily. In *Cell* 75 (2), pp. 307–320. DOI: 10.1016/0092-8674(93)80072-M.

- Strober, B. J.; Elorbany, R.; Rhodes, K.; Krishnan, N.; Tayeb, K.; Battle, A.; Gilad, Y. (2019): Dynamic genetic regulation of gene expression during cellular differentiation. In *Science (New York, N.Y.)* 364 (6447), pp. 1287–1290. DOI: 10.1126/science.aaw0040.
- Struhl, G. (1981): A homoeotic mutation transforming leg to antenna in *Drosophila*. In *Nature* 292 (5824), pp. 635–638. DOI: 10.1038/292635a0.
- Stuart, Tim; Butler, Andrew; Hoffman, Paul; Hafemeister, Christoph; Papalexi, Efthymia; Mauck, William M. et al. (2019): Comprehensive Integration of Single-Cell Data. In *Cell* 177 (7), 1888–1902.e21. DOI: 10.1016/j.cell.2019.05.031.
- Stubbington, Michael J. T.; Rozenblatt-Rosen, Orit; Regev, Aviv; Teichmann, Sarah A. (2017): Single-cell transcriptomics to explore the immune system in health and disease. In *Science (New York, N.Y.)* 358 (6359), pp. 58–63. DOI: 10.1126/science.aan6828.
- Stultz, Brian G.; Lee, Heuijung; Ramon, Karolyn; Hursh, Deborah A. (2006): Decapentaplegic head capsule mutations disrupt novel peripodial expression controlling the morphogenesis of the *Drosophila* ventral head. In *DEVELOPMENTAL BIOLOGY* 296 (2), pp. 329–339. DOI: 10.1016/j.ydbio.2006.05.034.
- Suzanne, Magali; Estella, Carlos; Calleja, Manuel; Sánchez-Herrero, Ernesto (2003): The *hernandez* and *fernandez* genes of *Drosophila* specify eye and antenna. In *DEVELOPMENTAL BIOLOGY* 260 (2), pp. 465–483. DOI: 10.1016/s0012-1606(03)00249-5.
- Svensson, Valentine; Vento-Tormo, Roser; Teichmann, Sarah A. (2018): Exponential scaling of single-cell RNA-seq in the past decade. In *Nature Protocols* 13 (4), pp. 599–604. DOI: 10.1038/nprot.2017.149.
- Tanaka-Matakatsu, Miho; Xu, Jinhua; Cheng, Leping; Du, Wei (2009): Regulation of apoptosis of *rbf* mutant cells during *Drosophila* development. In *DEVELOPMENTAL BIOLOGY* 326 (2), pp. 347–356. DOI: 10.1016/j.ydbio.2008.11.035.
- Tang, Fuchou; Barbacioru, Catalin; Wang, Yangzhou; Nordman, Ellen; Lee, Clarence; Xu, Nanlan et al. (2009): mRNA-Seq whole-transcriptome analysis of a single cell. In *Nature methods* 6 (5), pp. 377–382. DOI: 10.1038/nmeth.1315.
- The Inkscape Team (2019): Inkscape. Version 0.92.4. Available online at <https://inkscape.org>.
- The Tabula Muris Consortium (2020): A single-cell transcriptomic atlas characterizes ageing tissues in the mouse. In *Nature* 583 (7817), pp. 590–595. DOI: 10.1038/s41586-020-2496-1.
- Tirosh, Itay; Izar, Benjamin; Prakadan, Sanjay M.; Wadsworth, Marc H.; Treacy, Daniel; Trombetta, John J. et al. (2016): Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. In *Science (New York, N.Y.)* 352 (6282), pp. 189–196. DOI: 10.1126/science.aad0501.
- Tomancak, Pavel; Beaton, Amy; Weiszmam, Richard; Kwan, Elaine; Shu, ShengQiang; Lewis, Suzanna E. et al. (2002): Systematic determination of patterns of gene expression during *Drosophila* embryogenesis. In *Genome Biology* 3 (12), RESEARCH0088. DOI: 10.1186/gb-2002-3-12-research0088.
- Tomancak, Pavel; Berman, Benjamin P.; Beaton, Amy; Weiszmam, Richard; Kwan, Elaine; Hartenstein, Volker et al. (2007): Global analysis of patterns of gene expression during *Drosophila* embryogenesis. In *Genome Biology* 8 (7), R145. DOI: 10.1186/gb-2007-8-7-r145.
- Tomlinson, Andrew (2003): Patterning the peripheral retina of the fly: decoding a gradient. In *Developmental Cell* 5 (5), pp. 799–809. DOI: 10.1016/s1534-5807(03)00326-5.

- Torres-Oliva, Montserrat; Buchberger, Elisa; Buffry, Alexandra D.; Kittelmann, Maike; Sumner-Rooney, Lauren; Gaspar, Pedro et al. (2021): Differences in orthodenticle expression promote ommatidial size variation between *Drosophila* species.
- Torres-Oliva, Montserrat; Schneider, Julia; Wiegleb, Gordon; Kaufholz, Felix; Posnien, Nico; Desplan, Claude (2018): Dynamic genome wide expression profiling of *Drosophila* head development reveals a novel role of Hunchback in retinal glia cell development and blood-brain barrier integrity. In *PLOS Genetics* 14 (1), e1007180. DOI: 10.1371/journal.pgen.1007180.
- Torroja, Laura; Luo, Liqun; White, Kalpana (1996): APPL, the *Drosophila* Member of the APP-Family, Exhibits Differential Trafficking and Processing in CNS Neurons. In *The Journal of Neuroscience* 16 (15), pp. 4638–4650. DOI: 10.1523/JNEUROSCI.16-15-04638.1996.
- Tosches, Maria Antonietta; Yamawaki, Tracy M.; Naumann, Robert K.; Jacobi, Ariel A.; Tushev, Georgi; Laurent, Gilles (2018): Evolution of pallium, hippocampus, and cortical cell types revealed by single-cell transcriptomics in reptiles. In *Science (New York, N.Y.)* 360 (6391), pp. 881–888. DOI: 10.1126/science.aar4237.
- Toyofuku, Toshihiko; Nojima, Satoshi; Ishikawa, Takako; Takamatsu, Hyota; Tsujimura, Tohru; Uemura, Akiyoshi et al. (2012): Endosomal sorting by Semaphorin 4A in retinal pigment epithelium supports photoreceptor survival. In *Genes Dev.* 26 (8), pp. 816–829. DOI: 10.1101/gad.184481.111.
- Toyofuku, Toshihiko; Zhang, Hong; Kumanogoh, Atsushi; Takegahara, Noriko; Yabuki, Masanori; Harada, Koichiro et al. (2004): Guidance of myocardial patterning in cardiac development by Sema6D reverse signalling. In *Nature cell biology* 6 (12), pp. 1204–1211. DOI: 10.1038/ncb1193.
- Treisman, Jessica E. (2013): Retinal differentiation in *Drosophila*. In *Wiley interdisciplinary reviews. Developmental biology* 2 (4), pp. 545–557. DOI: 10.1002/wdev.100.
- Vergara, Hernando Martínez; Bertucci, Paola Yanina; Hantz, Peter; Tosches, Maria Antonietta; Achim, Kaia; Vopalensky, Pavel; Arendt, Detlev (2017): Whole-organism cellular gene-expression atlas reveals conserved cell types in the ventral nerve cord of *Platynereis dumerilii*. In *Proceedings of the National Academy of Sciences of the United States of America* 114 (23), pp. 5878–5885. DOI: 10.1073/pnas.1610602114.
- Vrtačnik, Peter; Kos, Špela; Bustin, Stephen A.; Marc, Janja; Ostanek, Barbara (2014): Influence of trypsinization and alternative procedures for cell preparation before RNA extraction on RNA integrity. In *Analytical Biochemistry* 463, pp. 38–44. DOI: 10.1016/j.ab.2014.06.017.
- Waltman, Ludo; van Eck, Nees Jan (2013): A smart local moving algorithm for large-scale modularity-based community detection. In *Eur. Phys. J. B* 86 (11). DOI: 10.1140/epjb/e2013-40829-0.
- Warren, William D.; Lin, Enmoore; Nheu, Thao V.; Hime, Gary R.; McKay, Michael J. (2000): Drad21, a *Drosophila* rad21 homologue expressed in S-phase cells. In *Gene* 250 (1-2), pp. 77–84. DOI: 10.1016/s0378-1119(00)00166-9.
- Webel, R.; Haug-Collet, K.; Pearson, B.; Szerencsei, R. T.; Winkfein, R. J.; Schnetkamp, P. P. M.; Colley, N. J. (2002): Potassium-dependent sodium-calcium exchange through the eye of the fly. In *Annals of the New York Academy of Sciences* 976, pp. 300–314. DOI: 10.1111/j.1749-6632.2002.tb04753.x.
- Wells, Brent S.; Pistillo, Daniela; Barnhart, Erin; Desplan, Claude (2017): Parallel Activin and BMP signaling coordinates R7/R8 photoreceptor subtype pairing in the stochastic *Drosophila* retina. In *eLife* 6. DOI: 10.7554/eLife.25301.

- Wickramasinghe, Vihandha O.; Laskey, Ronald A. (2015): Control of mammalian gene expression by selective mRNA export. In *Nature reviews. Molecular cell biology* 16 (7), pp. 431–442. DOI: 10.1038/nrm4010.
- Wilhelm, James E.; Buszczak, Michael; Sayles, Suzanne (2005): Efficient protein trafficking requires trailer hitch, a component of a ribonucleoprotein complex localized to the ER in *Drosophila*. In *Developmental Cell* 9 (5), pp. 675–685. DOI: 10.1016/j.devcel.2005.09.015.
- Wolff, T.; Ready, D. F. (1993): Pattern formation in the *Drosophila* retina. In M. Bate, a. Martinez-Arias (Eds.): *The development of Drosophila melanogaster*. II. Cold Spring Harbor: Cold Spring Harbor Laboratory Press, pp. 1277–1325.
- Wu, Angela R.; Wang, Jianbin; Streets, Aaron M.; Huang, Yanyi (2017): Single-Cell Transcriptional Analysis. In *Annual review of analytical chemistry (Palo Alto, Calif.)* 10 (1), pp. 439–462. DOI: 10.1146/annurev-anchem-061516-045228.
- Wu, Haojia; Kirita, Yuhei; Donnelly, Erinn L.; Humphreys, Benjamin D. (2019): Advantages of Single-Nucleus over Single-Cell RNA Sequencing of Adult Kidney: Rare Cell Types and Novel Cell States Revealed in Fibrosis. In *Journal of the American Society of Nephrology : JASN* 30 (1), pp. 23–32. DOI: 10.1681/ASN.2018090912.
- Wurmbach, Elisa; Wech, Irmgard; Preiss, Anette (1999): The Enhancer of split complex of *Drosophila melanogaster* harbors three classes of Notch responsive genes. In *Mechanisms of Development* 80 (2), pp. 171–180. DOI: 10.1016/S0925-4773(98)00212-3.
- Xiong, Wenjun; Rebay, Ilaria (2011): Abelson tyrosine kinase is required for *Drosophila* photoreceptor morphogenesis and retinal epithelial patterning. In *Developmental dynamics : an official publication of the American Association of Anatomists* 240 (7), pp. 1745–1755. DOI: 10.1002/dvdy.22674.
- Xu, Hong; Lee, Seung-Jae; Suzuki, Emiko; Dugan, Katherine D.; Stoddard, Alexander; Li, Hong-Sheng et al. (2004): A lysosomal tetraspanin associated with retinal degeneration identified via a genome-wide screen. In *The EMBO journal* 23 (4), pp. 811–822. DOI: 10.1038/sj.emboj.7600112.
- Xu, X. Z.; Wes, P. D.; Chen, H.; Li, H. S.; Yu, M.; Morgan, S. et al. (1998): Retinal targets for calmodulin include proteins implicated in synaptic transmission. In *The Journal of biological chemistry* 273 (47), pp. 31297–31307. DOI: 10.1074/jbc.273.47.31297.
- Yao, Bing; Li, Yujing; Wang, Zhiqin; Chen, Li; Poidevin, Mickael; Zhang, Can et al. (2018): Active N6-Methyladenine Demethylation by DMAD Regulates Gene Expression by Coordinating with Polycomb Protein in Neurons. In *Molecular Cell* 71 (5), 848-857.e6. DOI: 10.1016/j.molcel.2018.07.005.
- Yao, Jih-Guang; Weasner, Bonnie M.; Wang, Lan-Hsin; Jang, Chuen-Chuen; Weasner, Brandon; Tang, Chiou-Yang et al. (2008): Differential requirements for the Pax6(5a) genes eyegone and twin of eyegone during eye development in *Drosophila*. In *DEVELOPMENTAL BIOLOGY* 315 (2), pp. 535–551. DOI: 10.1016/j.ydbio.2007.12.037.
- Yim, Aldrin K. Y.; Wang, Peter L.; Bermingham, John R.; Hackett, Amber; Strickland, Amy; Miller, Timothy M. et al. (2022): Disentangling glial diversity in peripheral nerves at single-nuclei resolution. In *Nature neuroscience* 25 (2), pp. 238–251. DOI: 10.1038/s41593-021-01005-1.
- Yu, Jun; Liu, Yujuan; Lan, Xiang; Wu, Hao; Wen, Yang; Zhou, Zuomin et al. (2016): CHES-1-like, the ortholog of a non-obstructive azoospermia-associated gene, blocks germline stem cell differentiation by upregulating Dpp expression in *Drosophila* testis. In *Oncotarget* 7 (27), pp. 42303–42313. DOI: 10.18632/oncotarget.9789.

- Yu, Li; Zhou, Yating; Cheng, Songting; Rao, Yong (2010): Plexin a-semaphorin-1a reverse signaling regulates photoreceptor axon guidance in *Drosophila*. In *The Journal of neuroscience : the official journal of the Society for Neuroscience* 30 (36), pp. 12151–12156. DOI: 10.1523/JNEUROSCI.1494-10.2010.
- Zappia, Luke; Oshlack, Alicia (2018): Clustering trees: a visualization for evaluating clusterings at multiple resolutions. In *GigaScience* 7 (7). DOI: 10.1093/gigascience/giy083.
- Zheng, Grace X. Y.; Terry, Jessica M.; Belgrader, Phillip; Ryvkin, Paul; Bent, Zachary W.; Wilson, Ryan et al. (2017): Massively parallel digital transcriptional profiling of single cells. In *Nature Communications* 8, p. 14049. DOI: 10.1038/ncomms14049.

# Gordon Wiegleb

---

## Personal Profile

Date of birth: 12th of November 1989 in Hannover, Germany

## Education

2018 – 2022 **Doctorate studies** “IMPRS for Genome Science” at the University of Göttingen

2015 – 2018 **Master of Science**, “Developmental, Neural and Behavioral Biology” at the University of Göttingen

2014 **Erasmus Exchange** at the University of Oulu, Finland

2011 – 2015 **Bachelor of Science**, Biology, University of Göttingen

17.06.2010 – 15.07.2010 **Internship** at the Labor Prof. Böhm for bacteriology, mycology, parasitology and virology, Hannover

2009 – 2011 Vocational training to “**Staatlich geprüfter Biologisch-Technischer Assistent**“ at the Sabine Blindow Schulen Hannover

2002 – 2009 **Abitur**, IGS Mühlenberg, Hannover