Single-cell transcriptomics in the Drosophila visual system: Advances and perspectives on cell identity regulation, connectivity, and neuronal diversity evolution

Félix Simon, Nikolaos Konstantinides

Single-cell transcriptomics in the *Drosophila* visual system: advances and perspectives on cell identity regulation, connectivity, and neuronal diversity evolution

Félix Simon[1] and Nikolaos Konstantinides[1,2]

[1] Department of Biology, New York University, New York, NY 10003, USA

[2] Centre National de la Recherche Scientifique-UMR 7592, Institut Jacques Monod, Université Paris Diderot, Paris, France.

Correspondence: **Félix Simon** simonf04@nyu.edu

Nikolaos Konstantinides nk1845@nyu.edu

Abstract

The *Drosophila* visual system supports complex behaviors and shares many of its anatomical and molecular features with the vertebrate brain, yet contains a more manageable number of neuronal types. In addition to the extensive *Drosophila* genetic toolbox, this relative simplicity has allowed decades of work to yield a detailed account of its neuronal type diversity, morphology, connectivity and specification mechanisms. In the past three years, numerous studies applied large scale single-cell transcriptomic approaches to the *Drosophila* visual system and provided access to gene expression of several dozens of neuronal types throughout development. The fly visual system is therefore a unique system in which developmental mechanisms for neuronal specification, neuronal connectivity and neuronal gene expression can be studied in great detail. Here, we highlight how these resources complement each other and allow exploring long-standing biological questions under a new light. We first present the efforts made to characterize neuronal diversity in the *Drosophila* visual system and suggest ways to further improve this description. We then discuss current advances allowed by the single-cell datasets, and envisage how these datasets can be further leveraged to address fundamental questions regarding neuronal identity regulation, neuronal circuit development and evolution of neuronal diversity.

1. Introduction

The human brain, one of the most fascinating organs in terms of complexity and function, consists of 80 billion neurons of thousands of different cell types (Azevedo *et al.*, 2009). Understanding how this structure develops to support its function is one of the holy grails of developmental neuroscience. However its sheer complexity poses multiple practical challenges, prompting research in simpler, yet similarly organized systems. Model organisms offer unique advantages to study developmental mechanisms that generate neuronal diversity in less complex structures. The *Drosophila* visual system, and in particular the optic lobe, has emerged over the last ten years as an ideal model to address such questions for many reasons. The *Drosophila* optic lobes have a moderate complexity with a manageable number of cell types (~200) (Kurmangaliyev *et al.*, 2020; Özel *et al.*, 2020) that can, however, support complex behaviors (Branson *et al.*, 2009; Cande *et al.*, 2018). Its relative simplicity has allowed large high-throughput efforts that have described in detail the morphology of the different cell types (Fischbach and Dittrich, 1989; Nern, Pfeiffer and Rubin, 2015), their connectivity (Takemura *et al.*, 2013, 2015), and their transcriptomic diversity (Konstantinides *et al.*, 2018; Davis *et al.*, 2020; Kurmangaliyev *et al.*, 2020; Özel *et al.*, 2020). Importantly, neuronal identity is genetically hardwired in the optic lobes and does not rely on environmental factors (Izutsu *et al.*, 2015), which allows the study of the genetic basis of different neurodevelopmental processes (such as acquisition of neurotransmitter identity, synaptic specificity, etc). To this end, the genetic toolkit of *Drosophila* allows studying different processes in mechanistic detail. Due to the above reasons, we already have a good understanding of how neuronal diversity arises during larval development, as well as specific factors that are involved in this process (Li *et al.*, 2013; Suzuki *et al.*, 2013; Bertet *et al.*, 2014; Erclik *et al.*, 2017). Importantly, many of the lessons learned from the fly (such as temporal patterning, which is described below) have also been discovered in vertebrates (Elliott *et al.*, 2008; Alsiö *et al.*, 2013; Konstantinides, Rossi and Desplan, 2015; Mattar *et al.*, 2015; Holguera and Desplan, 2018).

The *Drosophila* optic lobes receive visual input from the retina, which is formed of ~800 ommatidia, each of which consists of 8 photoreceptors (R1-R8) (Bate and Martinez Arias, 2009), and is divided into four neuropils (ganglia): lamina, medulla, lobula and lobula plate (Fig. 1A) (reviewed in (Nériec *et al.*, 2016)). Each neuropil is divided into ~800 columns, corresponding to the ~800 ommatidia, which process the information from a given point in space. Therefore, the spatial organization of visual signals is kept in all neuropils, a property called retinotopy (Fischbach and Dittrich, 1989). Information is transmitted between layers and neuropils mostly by

unicolumnar neuronal types, which comprise 1 neuron per column, while multicolumnar neuronal types integrate visual signals across several columns and usually comprise less than 1 neuron per column (Erclik *et al.*, 2017). The lamina receives input from photoreceptors R1-R6, which detect motion (Bausenwein and Fischbach, 1992). The medulla is the largest structure of the optic lobe and receives input from both the lamina and the two color-sensitive photoreceptors, R7 and R8 (Fischbach and Dittrich, 1989). The information from the medulla is then forwarded to the lobula that processes for instance color, looming and feature detection (Wu *et al.*, 2016), and the lobula plate that processes motion (Borst, Haag and Reiff, 2010).

The vast majority of *Drosophila* optic lobe neurons are generated during the third instar larval stage. Two neuroepithelial structures (Outer Proliferation Center – OPC – and Inner Proliferation Center – IPC) generate neural stem cells (neuroblasts) that divide asymmetrically many times, self-renewing and generating an intermediate progenitor, called ganglion mother cell (GMC), which divides only once to generate two neurons (Hofbauer and Campos-Ortega, 1990). Neuroblasts are capable of generating many different neuronal types using three main mechanisms. *i) Temporal patterning* (Fig. 1B)*:* in the main part of the OPC, five transcription factors (TFs), *homothorax (hth)*, *eyeless (ey)*, *sloppy paired (slp)*, *Dichaete (D)*, and *tailless (tll)*, are expressed in a tight temporal sequence in every neuroblast of the medulla, leading to the formation of the different neuronal populations at different time windows (Li *et al.*, 2013; Suzuki *et al.*, 2013). A similar mechanism is at play at the tips of the OPC (Bertet *et al.*, 2014), as well as the IPC (Apitz and Salecker, 2015). *ii) Spatial patterning* (Fig. 1C)*:* three TFs (*Visual system homeobox 1 -Vsx*, *Optix*, and *Retinal homeobox - Rx*) and three secreted molecules (*hedgehog - hh*, *wingless - wg*, and *decapentaplegic - dpp*) expressed in sub-regions of the OPC, modify the neuroblast identity without affecting the temporal sequence to generate locally more neuronal types (Erclik *et al.*, 2017). Similarly, the IPC is spatially compartmentalized by the expression of *wg*, *dpp*, and *brinker* (*brk*) (Apitz and Salecker, 2015, 2018). *iii) Notch-driven binary cell fate decision* (Fig. 1B)*:* further neuronal diversity is achieved during the last asymmetric division of the GMC by *Notch* signalling, which is active in only one of the two neurons produced. The combination of spatial and temporal patterning coupled with a *Notch*-driven binary cell fate decision after the division of the GMC, as well as cell death (Pinto-Teixeira, Konstantinides and Desplan, 2016), are responsible for the generation of a large proportion of the observed neuronal diversity.

Here, we review how recent single-cell transcriptomic studies have advanced our understanding of the neuronal diversity within *Drosophila* visual system from different

perspectives. First, we summarize the transcriptomic efforts made to describe the neuronal diversity in the fly visual system and suggest ways to take even greater advantage of them. Then, we discuss specific applications of single-cell transcriptomics in studying a) gene regulation and how neuronal characters are regulated, b) neuronal circuitry formation, and c) evolution of neuronal diversity. In all cases, we review the current advances (with a focus on single-cell sequencing datasets) and we present our perspective as to how single-cell sequencing datasets can be used to address the above questions.

2. Transcriptomic diversity of the *Drosophila* visual system

Several studies took advantage of the modular structure of the *Drosophila* visual system, in which most neurons are repeated multiple times, to study biological processes at single-cell resolution using two different approaches: a) fluorescence-activated cell sorting (FACS) of a specific cell type followed by bulk RNA- or ATAC-sequencing, and b) single-cell sequencing of the entire tissue and resolution of cell types using clustering approaches (Box1, Table 1 and Table 2). Two studies focused on the third instar larval eye disc (Ariss *et al.*, 2018; Bravo González-Blas *et al.*, 2020), and two studies on the third instar larval brain (Cocanougher *et al.*, 2020; Ravenscroft *et al.*, 2020). A larger group of studies sought to achieve a detailed description of transcriptomic diversity in the optic lobes (Konstantinides *et al.*, 2018; Davis *et al.*, 2020; Kurmangaliyev *et al.*, 2020; Özel *et al.*, 2020) or the whole brain (Davie *et al.*, 2018). Although all studies in this group comprise data obtained at the adult stage, the two most recent ones (Kurmangaliyev *et al.*, 2020; Özel *et al.*, 2020) also performed scRNAseq throughout pupal development. Notably, these two studies also achieved a much higher resolution than the earlier ones: they both identify about 200 clusters, as compared to the 52 clusters in (Konstantinides *et al.*, 2018) and 53 bulk-sequenced isolated populations in (Davis *et al.*, 2020). These datasets represent an unprecedented resource to study all aspects of neuronal development and function, as they contain the transcriptome of more than 150 neuronal clusters throughout their development, with more than 60 of them annotated. We present here ways to further take advantage of these datasets, importantly without the need of producing additional sequencing data. It is important to note that, although the available single-cell mRNA sequencing data offer a unique opportunity to study cell type diversity, they do not contain information about chromatin accessibility, gene isoforms, alternative polyadenylation sites, and post-transcriptional regulation (Zhang *et al.*, 2016) that likely affect neuronal diversity and function (Sen *et al.*, 2019; J. Li *et al.*, 2020; Ha *et al.*, 2021).

2.1 Hidden heterogeneity

Although the above datasets are ready-to-use to explore diverse biological questions, the characterization of cell-type diversity offered by these datasets can be further improved. Many of the clusters contain “hidden” heterogeneity (Fig. 2): although clusters are group of cells with similar transcriptomes, what “similar” means is highly dependent on the depth of the data (number of cells and of genes per cells sequenced) and the thoroughness of the analysis. It is for instance important to note that, although how the selected clustering parameters are chosen is rarely discussed, they have a considerable influence on the number of clusters obtained for a given dataset: 87-151 (Davie *et al.*, 2018), or 146-229 (Özel *et al.*, 2020). Indeed, the Louvain clustering algorithm used to produce the pupal developmental atlases (Kurmangaliyev *et al.*, 2020; Özel *et al.*, 2020), and used in many other studies, has a tendency to either split abundant cell types into artificial groups (for instance based on the transcriptome quality) or to group rare cell types (Lancichinetti and Fortunato, 2011). Therefore, many of the smaller neuronal clusters are very likely to contain more than one cell type. This is illustrated in (Davie *et al.*, 2018) in which the authors chose to use the parameters yielding the smallest possible number of clusters (87 out of 151). By re-clustering independently some of their clusters, they were able to further divide them in clusters matching to known cell-types, showing that the initial clusters were heterogeneous. In addition to “hidden” rare cell types, these datasets are also likely to contain “hidden” molecular heterogeneity within both rare and abundant cell types. Indeed, the ability of the clustering algorithms to separate cells is dependent on the number of genes that they differentially express. This is why several neuronal subtypes in the optic lobe were shown to be distinguishable only at mid pupal stages, during synaptogenesis, when there is an increase in transcriptomic diversity (Özel *et al.*, 2020). Furthermore, clustering was not able to separate the dorsal and ventral subtypes of two abundant optic lobe neurons that differentially express Wnt4, Wnt10 and a few other genes (Özel *et al.*, 2020). These two populations were however spatially separated on tSNE visualization, suggesting that another clustering approach or different clustering parameters might have been able to separate them. However, despite these words of caution (which apply to any scRNAseq dataset), for most applications the developmental datasets can be used “out of the box”.

To reveal some of the hidden molecular heterogeneity of the datasets, complementary approaches can be applied to the study of one cluster, a group of clusters, or whole datasets. First, merging data from different studies using batch effect removal tools (Luecken and Theis, 2019; Hie *et al.*, 2020) will increase the number of cells that are analyzed, which can in turn

increase the ability of the clustering algorithms to isolate rare cell types. However, this could also leave residual batch effects that would need to be carefully controlled for. Second, in the case of developmental atlases, clustering should be propagated ("transferred") from the stage of maximal transcriptomic diversity to other developmental stages, instead of propagating the clustering that was made in adults (Özel *et al.*, 2020). Third, applying a step-wise approach in which clustering is first applied to the whole dataset and will efficiently identify abundant cell types, and clustering is subsequently applied independently to all clusters (or at least the smaller ones, more likely to be impure) to verify whether they are heterogeneous. Indeed, they could contain either rare or very similar cell types that were not separated in the first clustering step. Fourth, one can use additional bioinformatic approaches, many of them recently reviewed in (Hie *et al.*, 2020). For instance, using multiplets (Xi and Li, 2020) or ambient RNA (Fleming, Marioni and Babadi, 2019; Alvarez *et al.*, 2020; Yang *et al.*, 2020; Young and Behjati, 2020) removal algorithms could enhance relative differences between transcriptomes and improve clustering. Several tools have also been developed to improve clustering, for instance by using improved pre-processing (Hafemeister and Satija, 2019; Johnson, Kath and Mani, 2020) or the Leiden instead of the Louvain clustering algorithm (Traag, Waltman and van Eck, 2019; Stassen *et al.*, 2020).

In this regard, it is interesting to note that most single-cell algorithms are developed and benchmarked on "ideal datasets" of mammalian cells, with a large number of genes per cell, limited ambient RNA due to the use of cell culture or tissues requiring little to no dissociation (i.e. peripheral blood mononuclear cell), and a small number of cell types which are often of similar abundance. For instance, in one of the most detailed benchmark studies of the usual steps of scRNAseq analysis, the authors sought to produce reference datasets by mixing several cell lines, with and without adjunction of RNA to the mixture to model ambient RNA, and used these datasets to test almost 4,000 combinations of data analysis methods (Tian *et al.*, 2019). However, these datasets contained a mixture of at most 5 human lung adenocarcinoma cell lines, as opposed to the roughly 200 cell types of the optic lobe. The data obtained in the adult optic lobe are much more challenging to analyze: 4 to 6 times less genes per cell compared to mammalian cells, higher ambient RNA content due to severing neuronal and glial processes during dissociation, numerous cell types with a large distribution of abundances, and very similar cell types and subtypes that could not be separated by the algorithms used. Therefore, these datasets would constitute a good benchmark dataset, especially because the results in adults can be compared to a gold standard of 54 bulk transcriptomes obtained from purified optic lobe neuronal types (Konstantinides *et al.*, 2018; Davis *et al.*, 2020; Özel *et al.*, 2020).

2.2 Cluster annotation

About two third of the neuronal clusters in the pupal and adult atlases remain unannotated, which limits the use of these datasets for several applications. In our experience, it is relatively easy to find the cluster that corresponds to a given cell type, provided the cell type is genetically accessible and can therefore be characterized by specific markers, and is sufficiently abundant to form a cluster (Konstantinides *et al.*, 2018; Özel *et al.*, 2020). If more than one cluster expresses markers of a given cell type, the scRNAseq data can be used to find marker genes differentiating these clusters and to identify the correct cluster. On the other hand, identifying which cell-type corresponds to a given cluster is more difficult. In theory, since most clusters can be characterized by the differential expression of (semi-)specific genes, sparse labeling of drivers for these genes should allow the identification of the cell-type. A large repertoire of such lines exists or can be made using the Trojan exons (Diao *et al.*, 2015). However, as noticed in the original study, “the Trojan exon driver is a more sensitive indicator of gene expression than a corresponding antibody” (Diao *et al.*, 2015). This complicates cluster identification, as in many cases Trojan exons label a higher number of cell types than expected based on mRNA levels, as their expression can be driven by very low mRNA levels of the gene of interest. Therefore, a better way to identify the remaining neuronal clusters would be to use the scRNAseq data to identify driver lines expressed only in a few cell types (i.e. a manageable number), which can then be identified by sparse labeling. Clusters can subsequently be matched to these cell types by identifying markers through immunostaining or in-situ hybridization screens guided by the scRNAseq data.

2.3 Glial clusters

Almost all of the studies presented earlier focused on obtaining and analyzing clusters for neuronal, and not glial, cell-types. However, it is crucial to obtain the transcriptome of all glial types in the visual system as they play very important roles in neuronal circuit function and development (Lago-Baldaia, Fernandes and Ackerman, 2020). (Konstantinides *et al.*, 2018) obtained and identified 7 glial clusters in adult flies, while in the developmental atlases at least 19 glial clusters were found (Kurmangaliyev *et al.*, 2020; Özel *et al.*, 2020). However, glial clusters seem to be of lower quality than neuronal clusters, i.e. enriched in cells with features of low-quality transcriptomes (low gene count, high mitochondrial gene count), which could be due to a higher sensitivity of glial cells to the dissociation process. This would increase the proportion of glial genes in ambient RNA, and low-quality non-glial transcriptomes would cluster with good quality glial cells. Compared to neuronal clusters, they also contain a more variable proportion of cells

originating from different experimental batches (Kurmangaliyev *et al.*, 2020; Özel *et al.*, 2020), indicating that their clustering is partly driven by batch effect. Therefore, rather than analyzing directly glial clusters from these studies, it might be advantageous to separate glial cells from the neuronal cells and cluster and analyze them independently. This could allow to reduce the proportion of low-quality transcriptomes in the glial clusters by either filtering-out more aggressively the transcriptomes used, or using dedicated software to remove low-quality transcriptomes (multiplets or ambient RNA (Fleming, Marioni and Babadi, 2019; Alvarez *et al.*, 2020; Xi and Li, 2020; Yang *et al.*, 2020; Young and Behjati, 2020)). Moreover, clustering glial cells in the absence of neurons might increase the relative difference between glia and low-quality non-glial transcriptomes, and therefore allow for a better separation.

3. Use of single-cell approaches to study gene regulation and the link between specification and terminal features

The *Drosophila* optic lobe consists of ~200 different neuronal types, each of which is specified at the time of birth. This specification depends on the integration of three mechanisms (Fig. 1): spatial patterning of the neuroepithelium, temporal patterning of the neuroblast, and *Notch* status of the neuron. Once a neuron is specified, this identity has to be translated into terminal features, such as neurotransmitter identity, connectivity, morphology, etc. Several models have been proposed to explain how cell identity is established and maintained, such as the concepts of terminal selector (Hobert and Kratsios, 2019) or of core regulatory complex (Arendt *et al.*, 2016). However, these impose strong limitations and in this review we chose to use broader terms, as defined in Box 2. While we have a very good understanding of how optic lobe neurons are being specified, we are currently lacking the link between specification and terminal characteristics. Single-cell sequencing offers the opportunity to address this question at different levels.

3.1 Regulation of neuronal characters - investigation of gene regulatory networks

In addition to generic pan-neuronal features, each neuronal type is characterized by a specific set of morphological (neuropil layers targeted, synaptic partners, etc.) and molecular features (cell surface molecules, transporters, neurotransmitter identity, etc.). Understanding the regulation of such features is necessary to understand how neuronal diversity is generated.

The visual system is ideal for this endeavor, as the morphology of each neuronal type is characterized in great detail (Fischbach and Dittrich, 1989; Nern, Pfeiffer and Rubin, 2015), and

we now have access to gene expression during development for more than 150 neuronal clusters (Kurmangaliyev *et al.*, 2020; Özel *et al.*, 2020). One way to study the regulation of a given molecular or morphological feature is to probe for coincident expression between this feature and transcription factors. This can be realized “manually”, by comparing the TF expression between a manageable number of cell types, as illustrated by (Kurmangaliyev *et al.*, 2019; Hörmann *et al.*, 2020) who showed that the TF *grain* instructs both dendritic orientation and axonal targeting of T4/5b and T4/5c neuronal subtypes. However, the large number of neuronal clusters of the transcriptomic atlases also allows the use of machine learning to draw such comparison. For instance, we were able to show that the expression of markers of cholinergic (*ChAT*), glutamatergic (*VGlut*), and GABAergic (*Gad1*) identities is regulated by *apterous* (*ap*), *traffic jam* (*tj*), and *Lim3* transcription factors, respectively (Konstantinides *et al.*, 2018). This was further validated *in vivo*: knock-down of *apterous, traffic jam*, or *Lim3* led to reduced expression of ChAT, VGlut, and Gad1 in the optic lobes. Moreover, we showed that the same phenotypes (such as neurotransmitter identity) can be generated by different transcription factors, a phenomenon which we termed phenotypic convergence (Konstantinides *et al.*, 2018). We found that the expression of 2/3 of all the genes in the optic lobe is better predicted by a combination of TFs than by a single TF. This observation was later generalized beyond the *Drosophila* optic lobes, in the entire brain (Estacio-Gómez *et al.*, 2020).

In fact, such approaches can be applied to study the regulation of any gene that is expressed in the optic lobes. Numerous gene regulatory inference approaches can be used (Mercatelli *et al.*, 2020) and will be facilitated by the large scale of the newer atlases. Moreover, these inference approaches can also be improved to generate better hypotheses of the gene regulatory networks that regulate neuronal identity. For example, the widespread nature of phenotypic convergence suggests it would be beneficial to bias the models to consider combinations of transcription factors as drivers of neuronal identities. This is also emphasized by results from (Bravo González-Blas *et al.*, 2020), who coupled RNA- and ATAC-seq data in the eye-antennal disc to predict links between enhancers and target genes, and estimated that each gene is on average linked to 22 enhancers. Indeed, effector gene expression is regulated not only by transcription factors but also by their accessibility to different regulatory loci, which is likely to be affected by the spatial origin of neurons as discussed later in section 3.2. By using new techniques combining RNA- and ATAC-seq in the same individual cells, the accessibility profile of each neuronal cluster can be easily obtained as the mRNA sequenced in each cell will allow matching it with its corresponding cluster. These accessibility profiles can then be inputted in

machine learning models. Finally, it is important to consider that a cell type transcriptome is dynamic and reflects the developmental processes occurring at a given time. For instance, genes that are important for proper connectivity and synaptogenesis are enriched in the mid-pupal stages, when these processes take place (Kurmangaliyev *et al.*, 2020; Özel *et al.*, 2020). On the other hand, "functional" genes, such as ion channels, are more highly expressed in adulthood. Therefore, it is important to consider the timing of TF and gene expression when trying to identify gene regulatory interactions.

The establishment and maintenance of neuronal features requires some continuity in gene expression throughout neuronal life, which could be established by a transcriptional cascade (in which different TFs are expressed at different developmental stages), or by maintaining the expression of a "master regulator" throughout development, or by a blend of both. For instance, a first order TF involved in synaptic specificity might be only expressed from the birth of a neuron until P50. The existence of such a transcriptional cascade is both difficult to prove and to dismiss, as it requires identifying TFs expressed at different developmental stages and linked by regulatory relationships. On the other hand, the existence of master regulators is easier to test. They would be required to be expressed during the whole life of the neuron, and to be necessary and sufficient to establish and maintain all adult features of a neuronal type. The first requirement is fulfilled by numerous TFs, and we have defined for all our neuronal clusters specific combinations of TFs that are maintained throughout differentiation (Özel *et al.*, 2020). The median size of these combination of TFs is 8, and it is therefore unlikely that in general a unique TF would act as a master regulator. There is little doubt that at least some of these TFs are necessary to establish and maintain a subset of the adult features of their neuronal types, but whether they are sufficient will need to be carefully tested by a combination of imaging techniques (for the morphology) and transcriptomics (for the molecular phenotype).

3.2 How do temporal and spatial transcription factors regulate neuronal identity?

Temporal and spatial transcription factors (tTFs and sTFs) are expressed in progenitor cells (neuroblasts and neuroepithelial cells, respectively) and regulate the identity of the neuronal progeny (Li *et al.*, 2013; Erclik *et al.*, 2017). In many instances, tTFs and sTFs are not expressed in the neurons. Therefore, it is unclear how these factors act to endow identity to the neurons in the absence of their expression.

Temporal origin: Temporal transcription factors are expressed sequentially in neuroblasts, altering their capacity to generate neuronal cell types. This regulation can happen both through

direct effect on effector genes, and through the activation of expression of downstream (first-order) transcription factors (Box 2). A question that remains unanswered is whether a neuroblast needs to transition through all previous temporal stages to acquire its competence to generate a specific cell type. Sporadic evidence suggest that this is the case: it was shown recently that for a neuronal type to acquire the proper identity, the neuroblast has to proceed through all the previous temporal windows (Naidu *et al.*, 2020). In particular, the T1 cell type is generated during the *Dichaete* temporal window. However, T1 expresses three transcription factors that are activated by *eyeless*, *sloppy paired*, and *Dichaete* (*oc*, *Sox102F* and *Ets65A*, respectively), in their respective temporal windows. This means that should the neuroblast skip one of the three temporal windows, T1 will not be generated. While this is an intriguing mechanism, which "justifies" the use of temporal patterning as a mechanism to generate diversity (each temporal window adds an extra transcription factor), it is unclear whether it is generalizable. To address this question, one would need to have access to a thorough description of the lineage and the cell types that are generated from each temporal window and use single-cell sequencing data to identify the first-order TFs that are shared in cell types of the same temporal origin. This would immediately generate candidate regulatory interactions between the tTFs and the first-order TFs and will provide further insights into the role of the tTFs in activating the first order TFs, and, hence, regulating neuronal diversity.

However, the identification of the temporal origin of a cell type is not trivial. One could envisage the use of genetic tricks (for example, a combination of MARCM (Lee and Luo, 2001) technologies with memory cassettes) to stably mark only one of the progeny of a dividing neuroblast at a certain temporal window. This, on top of being genetically challenging, may also be subject to maturation delays of the markers and could lead to data mis-interpretation. A viable alternative would be to use single-cell sequencing in late larval stages in entire optic lobes that would contain neuroblasts, GMCs, and neurons. This could allow us to identify temporal origin of neurons based on the expression of temporal factors in their parental GMCs, which should be closely associated with them in the multidimensional space (Fig. 3A). A third, however incomplete, approach would be to use the expression of downstream TFs as a proxy for temporal origin. For example, we know that Bsh+ neurons are born from the *hth* temporal window, Dfr+ neurons arise from *ey*+ neuroblasts, Toy+ neurons come from the *slp* and *D* temporal windows (Li *et al.*, 2013). Using these genes as proxy, we should be able to assign putative temporal origin to all sequenced and annotated neuronal types. Finally, a more elaborate way to identify the temporal origin of neurons would be to use recently established cell lineage reconstruction techniques that combine single-

cell sequencing and CRISPR-mediated genetic scarring to infer developmental lineages (Alemany *et al.*, 2018; Raj *et al.*, 2018; Spanjaard *et al.*, 2018) that could then be associated with the tTF cascade.

Finally, it would be helpful to have a complete list of tTFs. A combination of single-cell sequencing and trajectory inference (Saelens *et al.*, 2019) could offer this: due to the progressive generation of neuroblasts in the OPC from medial to lateral, a wandering third instar larval brain contains neuroblasts at different levels of maturation. Single-cell sequencing of neuroblasts at L3 would generate transcriptomes for neuroblasts of different ages (Fig. 1B). One could then use trajectory inference algorithms to reconstruct the trajectory of the neuroblasts. The already known tTFs would serve as a positive control for the accuracy of the trajectory. Should the trajectory be trustworthy, one could then identify candidate tTFs by testing all TFs for potential temporal expression.

Spatial origin: Temporal and spatial factors likely use different mechanisms to regulate neuronal diversity. This is because spatial factors (contrary to the temporal factors) are mostly expressed in the neuroepithelium and not in the neural stem cells (Erclik *et al.*, 2017). In the embryonic ventral nerve cord, neuroblasts differ in their chromatin landscape depending on their spatial identity, which results in the first temporal transcription factor, *hunchback*, to bind distinct targets in different neuroblasts (Sen *et al.*, 2019). Similar mechanisms to integrate the action of temporally segregated transcription factors have been discovered a) in mouse neuronal stem cells where different proneural genes (*Ascl1* and *Neurog2*) set different chromatin landscapes for downstream transcription factors to act (Aydin *et al.*, 2019), as well as b) in the two neuronal lineages that give rise to the ASE neurons in C. elegans embryos, where *tbx-37/38* primes the locus of the *lsy-6* microRNA, which is activated four divisions later by the *che-1* transcription factor (Charest *et al.*, 2020). Thus, it is very likely that the spatial TF role in the optic lobes is also to setup the chromatin landscape for tTF and downstream transcription factors to act. To show whether this is indeed the case, a single-cell ATAC-seq atlas during the development of the visual system could be matched to the existing single-cell RNA-seq atlas and queried for chromatin landscape differences in neurons that come from different spatial domains.

To assign a spatial origin to a neuron, one can use memory cassettes during L3 to permanently mark all the cells that are born from a specific spatial domain (e.g. *Vsx*). Then, these marked cells can be isolated using FACS in the adult and single-cell sequenced. The transcriptomes can then be mapped to the annotated atlases using either integration algorithms

such as the one available in Seurat v3 or the neural network that was developed to annotate optic lobe cells (Özel *et al.*, 2020).

Finally, to identify all sTFs, single-cell sequencing of neuroepithelial cells at L3 larval stage could allow to cluster them in groups sharing a common spatial specification factor (the ones already known would serve as positive controls), which can then be experimentally validated.

Using these data, we could then develop models for the role of the tTFs and sTFs in regulating neuronal features, for instance by comparing gene expression, chromatin accessibility or morphological features in several neurons from the same temporal window or spatial origin. These models can then be tested genetically and lead to general conclusions regarding neurogenesis.

3.3 Regulation of neuronal characters by extrinsic signals – the role of ecdysone

Although neuronal identity in the *Drosophila* optic lobes is almost exclusively cell autonomous and does not rely on extrinsic signals, the role of the environment in the implementation of identity (i.e. differentiation) is undoubted. Transcriptomic studies indicate that neuronal development in the visual system could require the coordinated expression in all neuronal types of both pan-neuronal and neuronal type specific genes. Indeed, it was shown by following 88 neuronal and 6 glial clusters from P24 (i.e. 24% of pupal development) onwards (Kurmangaliyev *et al.*, 2020) that the expression of 200 neuron specific genes was pan-neuronally coordinated during development. Although these genes exhibited various expression dynamics during development, for each given gene these dynamics were the same in almost all neuronal types. This indicates that pan-neuronally coordinated genes may be involved in different steps of neuronal development, and that these steps occur simultaneously in all optic lobe neuronal types. For instance, genes whose expression peaks early could be necessary for axon growth or guidance, while those peaking later might be necessary for synapse formation or cell excitability. Because these genes are pan-neuronally expressed, they are unlikely to be involved in establishing neuronal type specific features but may fulfill permissive roles for neuronal development and function. Notably, compared to the other neuron specific genes, they are enriched in ion channels and synaptic proteins functional categories, and could therefore be necessary for synapse formation but may not be involved in synaptic specificity. Moreover, some of these genes are involved in intercellular communication or encode RNA-binding proteins, indicating potential tissue-wide changes in signal transduction or post-transcriptional regulation. The temporal coordination of gene expression during visual system development was also emphasized by analyzing variations in transcriptomic diversity. At any pupal stage, optic lobe

neuronal types comprise individual neurons produced at various timepoints (up to 2 days apart) during larval development (Fig. 1B). It was shown that at early pupal stages the transcriptome of neurons from the same type but of different ages were transcriptionally different, but later converge towards a common state that is reached by P30 (Özel *et al.*, 2020). Moreover, it was shown that a peak of transcriptional diversity occurs shortly after in all neuronal clusters , between P40 and P70 (Özel *et al.*, 2020).

These optic-lobe wide variations in transcriptomic diversity and the coordinated expression of some pan-neuronal genes suggests the presence of at least one extrinsic signal that orchestrates this coordination. Several findings suggest that a peak of ecdysone signaling occurring between P30 and P40 (Handler, 1982) could be such a signal, and that ecdysone responsive TFs (Fig. 3B) might control the temporal expression of genes critical for neuronal development. Indeed, the genes specifically upregulated at P36 (Jain *et al.*, 2020; Kurmangaliyev *et al.*, 2020) and P40 (Özel *et al.*, 2020) were enriched in ecdysone responsive transcription factors. These transcription factors were differentially expressed between developmental stages, in a mostly coordinated fashion across optic lobe neuronal clusters during development (Jain *et al.*, 2020; Kurmangaliyev *et al.*, 2020). The role of ecdysone signaling on gene regulation was explored in detail in lamina neurons L1-5 by genetic perturbations of ecdysone signaling associated with transcriptomic studies, in which among other findings the authors showed that genes with temporal variation of expression are enriched in targets of ecdysone signaling (Jain *et al.*, 2020). Importantly, they also showed that despite ecdysone signaling being pan-neuronal, *EcR* dominant negative expression often affected a given gene differently in the different lamina neurons, with sometimes opposite variations of expression. Moreover, the neuronal-type specificity of *EcR* disruption could be due to the action of co-factors: L3 specific *EcR* target genes were also enriched in *erm* targets.

Interestingly, we have shown that the mid-developmental peak of transcriptomic diversity is associated with an upregulation of cell surface molecule (CSM) expression (Özel *et al.*, 2020), which correlates with synapse formation in the optic lobe and occurs just after the peak of ecdysone signaling. This suggests that this peak regulates circuit formation by controlling the expression of hundreds of CSMs during pupal development, which is consistent with work by (Jain *et al.*, 2020). Indeed, the 921 genes they identified downstream of ecdysone signaling in L1-5 were enriched in genes encoding Immunoglobulin Superfamily proteins as well as in GO terms associated with wiring (Jain *et al.*, 2020). In addition, immunostainings in ex-vivo cultures of optic lobes also showed that the expression of the CSMs *dpr15*, *dpr17* and *dpr6* was dependent on

ecdysone presence in the cultivation media. Lastly, specific expression of a dominant negative *EcR* in various cell-types resulted in mistargeting of R8 photoreceptors, formation of ectopic synapses in the lamina, disorganization of medullar arbors of lamina neurons, and affected the morphology of T4/T5 neuron terminals in the lobula plate. Moreover, (Jain *et al.*, 2020) have also shown that different members of the ecdysone responsive transcription factor cascade can be involved in the step-wise development of neurons (Fig. 3B). To do so they studied L5 neurons, which first send projections to medulla layer 1 around P48, then to medulla layer 2 around P75. The first step was blocked by *EcR* dominant negative or *EcR* RNAi, and the second step but not the first one by *Hr3* and *ftz-f1* RNAi. Moreover, *Hr4* RNAi caused secondary projections to the medulla layer 5. Therefore, different ecdysone responsive transcription factors were involved in different steps of L5 wiring. Consistently, sequencing of L5 cells showed that expression of EcR dominant negative or Hr3 RNAi affected two different sets of CSMs.

These results show that extrinsic signals regulate critical aspects of neuronal identity, such as the timing of the differentiation process both in a pan-neuronal or a cell type-specific manner.

4. Use of single-cell approaches to study neuronal circuit development

In order to form functional circuits, neurons must first reach their target neuropils, then arborize in their target layer, and finally form synapses with their correct synaptic partners. At the time of synapse formation, each growth cone is surrounded by dozens of potential partners, yet connects only to a few stereotyped partners, a property called synaptic specificity. In the *Drosophila* visual system as in any other neural system, neuronal targeting and synaptic specificity are mediated by the communication between neurons and their environment (surrounding neurons, glia and extracellular matrix) through cell surface and secreted molecules (Plazaola-Sasieta *et al.*, 2017; Sanes and Zipursky, 2020). Despite a general understanding of many mechanisms underlying neural circuit development, the mechanistic detail and the identity of the molecules involved are mostly unknown.

Single-cell transcriptomic atlases, and in particular the ones that span development, constitute an ideal resource to identify these molecular determinants. Moreover, two other resources considerably facilitate the study of synaptic specificity in *Drosophila*. First, we have a detailed knowledge of the connectome of the lamina (Meinertzhagen and O'Neil, 1991; Meinertzhagen and Sorra, 2001; Rivera-Alba *et al.*, 2011), medulla (Takemura, Lu and Meinertzhagen, 2008; Takemura *et al.*, 2013, 2015), and of part of the lobula and lobula plate (Shinomiya, Horne, *et al.*, 2019; Shinomiya, Huang, *et al.*, 2019). Second, the interactome of 202

CSMs has been characterized (Ozkan *et al.*, 2013; Cosmanescu *et al.*, 2018). Together, one could expect that these resources will allow for important insights into how neuronal circuits are built and especially into how synaptic specificity is achieved, one of the least understood aspects of neuronal circuit development (Sanes and Zipursky, 2020). However, since progress has been limited so far, we will discuss how the transcriptomic datasets advocate for new approaches in the study of neuronal circuits development.

4.1 Neuronal circuit development is unlikely to be explained by a simple molecular code

Numerous mechanisms have been described to be involved in synaptic specificity, including the pruning of synapses between incorrect partners, or the death of improperly connected cells (Sanes and Yamagata, 2009). We will focus on one of the most studied models, initially formulated by Langley and formalized by Sperry (Langley, 1895; Sperry, 1963), in which growth cones carry CSMs and establish synapses only with neuronal types expressing compatible CSMs at their surface. The "molecular labels" allowing the correct matching of the neuronal types have been studied for decades in both vertebrates and invertebrates, and the progresses have been recently reviewed in (Sanes and Zipursky, 2020). However, the details of the mechanisms, such as the number of CSM interactions necessary to establish synapses, are debated. It has been proposed that many of these "molecular labels" come from a few diversified families of cell surface proteins, two of the best studied in *Drosophila* being the defective proboscis extension response (*dpr*) and the Dpr-interacting protein (*DIP*) families (Ozkan *et al.*, 2013; Carrillo *et al.*, 2015; Tan *et al.*, 2015; Cosmanescu *et al.*, 2018). They respectively comprise 21 and 11 members, which can bind to each other either homophilically or heterophilically with varying affinities as shown on Fig. 3A. It was proposed that the expression of compatible *DIP*s and *dpr*s by the projections of two neuronal types in close proximity would instruct the formation of synapses between them. However, although a few examples of such interactions have been described, disrupting *DIP* or *dpr* expression most often leads to only partial gain or loss of function (Courgeon and Desplan, 2019; Sanes and Zipursky, 2020). These observations have recently led to a more nuanced view of the role of *DIP* and *dpr* (and other CSMs), in which large numbers of CSMs are required to pair two neuronal types, and these CSMs do not instruct synapse formation in a "all-or-nothing" fashion but rather affect the relative "preference" of a neuronal type towards another (Sanes and Zipursky, 2020).

The developmental atlases very strongly support this more nuanced view of synaptic specificity regulation, as they notably show that many non-synaptic partners projecting to a

common layer express compatible CSMs. In fact, this situation is likely to be the rule rather than the exception due to the widespread expression of many cell surface molecules, as can be seen for the *dpr*s in Fig.3B or for dozens of CSMs in 13 neurons in (Kurmangaliyev *et al.*, 2020). A transcriptomic study of 5 lamina neurons and two photoreceptors at P40 also showed that hundreds of CSMs were expressed in each cell type, and that between any given pair 49-168 CSMs were differentially expressed (Tan *et al.*, 2015). Moreover, (Davis *et al.*, 2020) have studied CSM expression in 12 lamina neurons, and showed that 1) any pair of these neurons express at least 30 couples of compatible CSMs, 2) the number of compatible CSMs expressed is uncorrelated with the number of synapses made by these pairs of neuronal types. Lastly, as an example, virtually all neuronal clusters express one or more *dpr* with high affinity for *DIP-β* at P50 (Fig. 4D). Therefore, if the pairing of a couple of *DIP*/*dpr* were sufficient to establish a synapse, any neuronal type expressing *DIP-β* could theoretically form synapses with almost all of the neurons in its vicinity. This is obviously not the case, and it has important implications for candidate screens: the expression of a couple of compatible CSMs between two synaptic partners is a weak indication that these CSMs alone play a significant role in the establishment of these synapses.

Moreover, synaptic partners usually express multiple couples of interacting CSMs. Such redundancy has historically been proposed to explain why disrupting the expression of a CSM produces mild or inexistent phenotypes more often than not (Sanes and Zipursky, 2020). This is illustrated in detail (Fig. 3C) by an analysis of the neuronal types L4 and L2, which form synapses in the proximal part of the lamina (Xu *et al.*, 2019; Kurmangaliyev *et al.*, 2020). L4 is the only lamina neuron expressing *DIP-β*, and L2 expresses several *dpr*s that can interact with *DIP-β*. If *DIP*/*dpr* interactions were necessary and sufficient to establish synaptic specificity, it would be reasonable to expect that *DIP-β* and its compatible *dpr*s would instruct the establishment of L2-L4 synapses. However, although *DIP-β* loss of function produces ectopic synapses in the distal lamina, synapses still form in the proximal lamina. This could be explained by the fact that 11 other couples of interacting CSMs are still expressed between L4 and L2 neurons. Interestingly, L1 and L4 similarly also express these 11 couples of interacting CSM but do not from synapses, which further illustrates that non interacting partners can express a large number of compatible CSMs.

In addition to its implications for the regulation of synaptic specificity, the large number of CSMs expressed in each neuronal type leads to a large number of candidates for any neuronal targeting feature. This is for instance shown by two developmental studies of the T4 and T5

neuronal types, which are ideally suited to the study of both neuronal targeting and synaptic specificity. They each comprise 4 neuronal subtypes (a, b, c and d), projecting to different lobula plate layers and to the same medulla layer (T4 subtypes) or lobula layer (T5 subtypes). The dendrites of each of the T4/T5 subtypes are orientated in one of 4 cardinal directions. Because each subtype seemingly differs only by one of the layers they target and by the orientation of their dendrites, CSMs differentially expressed between them are ideal candidates for the regulation of their morphology. In (Kurmangaliyev *et al.*, 2019), several CSMs were found differentially expressed between neurons of the same type but targeting different lobula plate layer. For instance, just among the top 20 genes differentially expressed between T4a and T4c were the CSM *Fas2*, *Ptp99A*, *beat-IIIb*, *mspo*, *CG15765*, *robo2*, *side-IV*, *Con*, and *klg*. Similarly, (Hörmann *et al.*, 2020) found 26 CSMs potentially involved in controlling the orientation of T4/T5 dendrites.

Lastly, these results emphasize how difficult it will be to assign a role to the different CSMs involved in establishing neuronal circuits. Indeed, the role of a candidate CSM is usually tested by assays only able to detect strong phenotypes, such as mislocalization of neuronal arbors. However, since most morphological features are likely to be encoded by redundant CSMs, disrupting the expression of a single candidate CSM is likely to have more subtle effects and should therefore be assessed by assays able to detect subtle changes in neuron morphology (for instance using sparse labeling) and in the number of synapses between neuronal types. Another approach could be to favor strong phenotypes by disrupting the expression of many CSMs at once, for instance by using an array of CRISPR guides. Coupled with quantitative assays of synaptic specificity (ex: number of synapses lost or of ectopic synapses), the disruption of varied proportions of CSM interactions between two neurons could also help defining general rules for synaptic specificity. Lastly, the developmental atlases should help for the identification of candidate CSMs, by allowing the grouped study of many neuronal types that share a common feature. Indeed, different neuronal types could reach a common layer by expressing different sets of CSMs. However, if most neuronal types targeting this layer specifically express a shared CSM, this CSM is likely involved in targeting to this layer.

4.2 What is missing to understand the genetic regulation of neuronal circuit development?

The fact that many neurons sharing compatible CSMs do not form synapses could, in some cases, be explained by CSMs mediating a repulsive rather than an adhesive interaction. However, several results suggest that understanding synaptic specificity will require models taking into account several additional parameters.

Subcellular localization: The subcellular localization of CSM proteins as well as the localization of neuronal projections during development might have an important role in establishing synaptic specificity. Indeed, even if two neuronal types project to a common layer in adults and express compatible CSMs during synapse formation, they would not form common synapses if these CSMs are not in close proximity during synapse formation. It is known that optic lobe layers are progressively constructed, and that neuronal types can target them in a stepwise fashion (Plazaola-Sasieta *et al.*, 2017). For instance, as discussed previously, L5 neurons target different medulla layers at different stages of development. Correlating which CSMs are expressed in a neuronal type at the time of synapse formation with the localization of its arborizations will therefore be necessary to understand synaptic specificity. Although it would be experimentally challenging to obtain such data for all neuronal types, case studies might lead to the identification of general mechanisms. In addition, the restriction of CSMs to specific subcellular locations is probably a common occurrence. For instance, in (Kurmangaliyev *et al.*, 2019), klg is expressed at P48 in subtypes of T4 and T5 that target 4 different layers of the optic lobe (M10, Lo1, LoPc and LoPd). However, immunostaining against klg.GFP only labels strongly 2 of these layers (LoPc and LoPd). Again, obtaining the subcellular localization of all CSMs in all neuronal types during development would be very challenging, but it can easily be approximated by confocal imaging for a manageable number of CSMs and neuronal types.

In vivo binding affinities: The establishment of synapses between two neuronal types might not be encoded by a few CSMs, but by a combination of numerous CSMs. Importantly, the interaction between each couple (or more complex arrangements) of CSM is characterized by different affinities. Therefore, to describe the local interaction between patches of membranes from two neuronal types, several parameters must be taken into account: the CSMs present, their concentration, and the affinities of each couple of CSMs. Notably, this interaction is not a linear combination of the parameters: at similar concentration, CSM couples with higher affinity will produce more dimers than couples with lower affinity, in addition to the higher contribution of each of these couples to the adhesive strength between the two membrane patches (Katsamba *et al.*, 2009). Although the interactome of 202 CSMs has been determined in *Drosophila*, their binding affinities have been much less studied and, when obtained, were obtained in vitro. Therefore, a crucial piece of information might be largely missing to understand synaptic specificity. The expression levels of all CSMs, however, is now very well characterized during development in most optic lobe neurons (Kurmangaliyev *et al.*, 2020; Özel *et al.*, 2020). However, it is also important to note that the amount of dimers formed, and their adhesive strength, might be less

important than the signaling pathways activated by the binding of compatible CSMs (Dalva, McClelland and Kayser, 2007).

Neuronal activity: intercellular communication might play a role in synapse formation. Indeed, it has been shown in vertebrates that neurotransmission-mediated processes are involved in synaptic specificity by promoting either the reinforcement or the elimination of synapses (Okawa *et al.*, 2014). Such a mechanism is appealing as it would increase the reproducibility of synapse formation and reduce the number of potential synaptic partners, by allowing only synapses between neuronal types that can communicate with each other (i.e. pre- and post-synaptic cells with compatible neurotransmitter release and capture machinery, respectively). However, circuit formation in *Drosophila* is thought to be mostly independent of neuronal activity, as it was shown that disrupting activity in photoreceptors does not affect their axonal targeting, column organization or synapse numbers, and the later was also unaffected by changing column composition (Hiesinger *et al.*, 2006). Nevertheless, photoreceptors represent only a small fraction of the visual system neuronal diversity, and activity could affect the proportion of synapses they make with their different partners without affecting the total number of synapses per photoreceptor. Therefore, whether intercellular communication affects synapse formation is an open question and is suggested by several experimental observations in the central nervous system, as reviewed in (Akin and Zipursky, 2020), and in the visual system. Indeed, among the pan-neuronally coordinated genes, the expression of most of the ion channels peaked during development and not at adulthood, as could be expected, and neurotransmitter mRNA was detected early during development (Kurmangaliyev *et al.*, 2020). Moreover, it was shown that the optic lobe exhibits a cell type specific, patterned, stimulus-independent neuronal activity (PSINA) that starts in a subset of processes at P50 and expands to the whole optic lobe by P55 (Akin *et al.*, 2019). The PSINA were correlated with glutamate release and changes in membrane voltage in neurons. Very importantly, PSINA patterns were cell-type specific, stereotyped, correlated between synaptic partners, and uncorrelated between non-synaptic partners. Expressing the tetanus toxin in the presynaptic cell of a neuronal pair uncorrelated their PSINA activities, while doing so in the post-synaptic pair did not affect the correlation, indicating that the correlation of PSINA activities requires synaptic activity. Disrupting PSINA activity also “leads to cell-type specific changes in synapse counts in the visual system” (unpublished observations mentioned in (Akin and Zipursky, 2020)). Moreover, complementary activity was observed in neurons and astrocytes. The latter have been shown to invade neuropils of the central nervous system during synaptogenesis, and the number of synapses was decreased by up to 50% upon astrocyte

ablation (Muthukumar, Stork and Freeman, 2014). In addition, the expression of the pan-neuronally regulated gene ShakB peaked at mid-pupation, and this gap junction protein was important for the formation of PSINA in the visual system (unpublished data mentioned in (Kurmangaliyev *et al.*, 2020)). Lastly *Hr38*, a conserved insect neural activity marker gene and a mediator of an ecdysteroid signaling pathway independent of *EcR* (Baker *et al.*, 2003), was also a pan-neuronally regulated gene and its expression increased after P36 (Kurmangaliyev *et al.*, 2020). Therefore, a global increase in neuronal activity is observed around the time of synapse formation, potentially triggered by ecdysone signaling, that could affect synaptic specificity. Importantly, the correlation of PSINA activity between synaptic partners indicates that some synapses are already present. Therefore, activity is unlikely to be enough to specify synaptic specificity but could affect the relative number of synapses made between different potential synaptic partners. Consistently, a study has shown that blocking synaptic transmission in mechanosensory Chordotonal neurons (in a larval abdominal segment) did not affect the identity of their post-synaptic partners, but did affect the proportion of total number of synapses with them (Valdes-Aleman *et al.*, 2021).

Synaptic specificity or synaptic selectivity? The formation of synapses between two neurites might not be an “all-or-nothing” phenomenon, in which synapses are formed each time a certain “affinity threshold” is exceeded. Rather, in each region where they project, neurons might “rank” potential partners and only connect to the highest ranked ones based on a competitive process. In this case, it might be better more precise to talk about synaptic selectivity rather than synaptic specificity (Sanes and Zipursky, 2020). In addition to being more consistent with the observations made in the transcriptomic datasets, some experimental observations support this model. For instance, upon deletion of *DIP-β* in L4 neurons, they form ectopic synapses in the distal part of the lamina: although L4 cells had the molecules required to make these synapses all along, they were not preferred when DIP-β was expressed (Xu *et al.*, 2019). A competitive model of synaptic specificity requires mechanisms by which neurites evaluate different potential partners, and form synapses only with the highest-ranking ones. A simple and passive mechanism could be based purely on membrane adhesivity in a manner analogous to the differential adhesion hypothesis. It postulates that a population of cell types with various adhesive affinity to each other, and for which adhesive interactions are dynamic, will adopt a spatial organization that maximizes adhesive stability (Steinberg, 1970). This is often illustrated with populations of two cells with uniform distribution of CSM on their membrane (Fig. 3E), but can be applied to the local interactions between dozens of neurites with different adhesive strength to

each other. It was for instance shown that N-cadherin levels are responsible for the spatial organization of photoreceptors, lamina and medulla neurons terminals (Schwabe *et al.*, 2014; Trush *et al.*, 2019). A second passive mechanism can be imagined if neuronal types form a number of synapses intrinsically determined (but not necessarily constant between individuals), and synapses between pairs of neurites with the highest affinity have a competitive advantage (by being either preferentially formed or stabilized). In that case, the proportion of synapses made with each of their partners will be stereotyped. It was for instance shown that R7 neurons form an intrinsically determined number of synapses by the competitive accumulation of a limited resource, the synaptic seeding factors, in only 1-2 filipodia at a time. Notably, here the competitive mechanism did not impact synaptic specificity but the number of synapses produced (Özel *et al.*, 2019). However, such a mechanism could be involved in synaptic specificity if the accumulation of synaptic seeding factors preferentially occurred between filipodia pairs of the highest affinity. Lastly, if synapses between neurites of the highest affinity were to be produced first, synaptic specificity could also be established by the active elimination of all other synapses. In *C. elegans* for instance, an E3 ubiquitin ligase complex is responsible for the degradation of presynaptic sites in the HSNL neuron, excepted for those containing the synaptic adhesion molecule SYG-1 that inhibits assembly of the complex (Ding *et al.*, 2007). The preferential accumulation of such a "stabilizing factor" in synapses between filipodia with the highest affinity would also confer them a competitive advantage. Notably, all these examples of mechanisms would lead to reproducible connectivity patterns in wild type animals but are very different from a "hard" encoding of synaptic specificity by a fixed molecular code. Indeed, depleting a CSM in a given neuronal type would only result in subtle changes in the ranking of its preferred synaptic partners and not in a complete loss of synaptic specificity, which is consistent with the mild phenotypes usually observed upon such modifications.

The potentially very complex nature of synaptic specificity regulation makes it difficult to study exclusively by traditional approaches, such as screening for candidate genes one neuronal type at a time. Bioinformatic modeling can help to produce falsifiable hypotheses, and therefore to identify potential mechanisms regulating synaptic specificity, based on the study of dozens of neuronal types at once. Such models should be required to accurately predict both synaptic and non-synaptic partners based on the developmental atlases and the connectome. Moreover, the effect of all parameters discussed above (such as CSM subcellular localization) on the accuracy of these predictions can be tested *in-silico* and guide subsequent *in-vivo* validations.

5. Use of single-cell approaches to study visual system evolution

Neurons represent the most diverse cell types in the animal body. Understanding how this remarkable diversity has evolved over the millions of years of animal evolution is a fascinating, but still elusive, subject. We have good hypotheses for how neurons first appeared and how they assembled their molecular machinery (Achim and Arendt, 2014; Arendt, 2020). Moreover, recent studies have addressed the evolution of neuronal structures in different vertebrates (Tosches *et al.*, 2018; Norimoto *et al.*, 2020). However, we are lacking good examples for how neuronal cell types evolve. To study neuronal type evolution we need to perform cross-species comparisons; the advent of single-cell sequencing allows us to study neuronal systems in different animals, even in non-genetically accessible ones, at the level of individual neuronal types.

5.1 Neuronal type evolution

Neuronal types are the basic units of a nervous system; however, the definition of a neuronal type is not always straightforward. A neuronal type has been classically defined as a group of neurons with similar morphology and function that are different from those of other neuronal types. Recent transcriptomic studies in the *Drosophila* brain have shown that morphological neuronal types agree to a great extent with transcriptomic neuronal types (H. Li *et al.*, 2020; Kurmangaliyev *et al.*, 2020; Özel *et al.*, 2020). However, studies both from the optic lobe and the olfactory system identified subtypes of morphologically identical cell types, i.e. cell types with the same morphology, but different transcriptomes (Li *et al.*, 2017; Özel *et al.*, 2020). Similar observations have also been made in mice, where the transcriptome appears to have a higher resolution than morphology and physiology (Scala *et al.*, 2020). This transcriptomic diversity is driven by the cell type-specific expression of different genes that are regulated by a set of transcription factors, which can be thought of as the identity of each neuronal type.

A definition of neuronal types based only on their morphology and function does not take into account their evolutionary history: two neuronal types with similar morphology and function can be evolutionarily more distant than two neuronal types with divergent morphologies. This caveat can be mitigated by the use of single-cell sequencing technologies, which now allows us to study the cell type composition of any neuronal tissue of any animal we have access to at the transcriptomic level. Then, we can compare the neuronal type compositions between different species and try to generate cell type trees, in a fashion similar to phylogenetic trees. This comparison could be made either at the level of the entire transcriptome or at the level of the transcription factors (Konstantinides, Degabriel and Desplan, 2018) (Figure 5). In the former scenario, one could use batch-effect correction (data integration) algorithms (Tran *et al.*, 2020) to

integrate the single-cell sequencing data from different species, correcting essentially for the batch effect introduced by the independent evolution of cell types upon speciation. This, although feasible, might be blurred by cases of convergence of different cell types (see section 5.2) as well as by divergence of conserved cell types, especially in cases of non-closely related species (Figure 5). However, two different cell types with separate evolutionary history are unlikely to share common gene regulatory networks, even if they present convergent effector gene expression. Therefore, differentiating between homologous and convergent cell types can benefit from the comparison of transcription factor expression between species (Figure 5) (Tosches *et al.*, 2018). Ideally, this analysis would be limited to the specification factors and their downstream TFs, and would require their identification in each neuronal type before performing such comparisons. Finally, an intermediate approach would be the comparison of entire transcription factor signatures between cell types of different species.

The insect visual system represents an ideal tissue for such studies. The repetitive nature of the optic lobe neuropils results in the presence of most cell types in large copy numbers, which facilitates single-cell sequencing clustering and annotation. Moreover, the recent publication of a detailed and annotated cell atlas in adult *Drosophila* optic lobes, as well as during development provides a perfect ground plan for future comparisons and annotation transfers (Kurmangaliyev *et al.*, 2020; Özel *et al.*, 2020). On top of that, the determination of the developmental origin of all cell types (as described in section 3.3) will facilitate the identification of homologous cell types, as homologous cells types should come from the same lineage in different animals. Finally, despite being a fairly simple structure (compared to the central brain), the optic lobes are composed of many neuronal types (~200) that can be compared to other insects and unveil different evolutionary histories.

5.2 Convergent cell types

While these approaches can be successful in relatively closely related species (such as insects), matching cell types across large evolutionary distances might be futile for many reasons. One of the reasons is, of course, that the amount of “noise” introduced by the hundreds of millions of years of independent evolution prohibits the accurate recovery of the homology “signal”. The other reason is that there is a smaller number of orthologous cell types and one might need to rely on the identification of orthologous cell type families. On the other hand, some of the neuronal functions may have become necessary independently in different animals leading to convergence of non-orthologous neuronal types to the same functions (Figure 5).

Convergent cell types are of particular interest, as they represent two independent instances where animals had to solve the same problem. A very interesting question that arises is whether the solution identified in each case is the same, i.e. the same gene regulatory networks are being used to generate the convergent phenotypes, a phenomenon called deep homology (Shubin, Tabin and Carroll, 1997, 2009), or they differ between the two species.

Recently, a large single-cell transcriptomic survey of the developing *Drosophila* optic lobes identified neuronal populations that resemble in many respects the Cajal-Retzius cells of the vertebrate central nervous system (Soriano and Del Río, 2005; Özel *et al.*, 2020), as they are only present during development and project only to the surface of the brain. Cajal–Retzius cells are essential for neuronal migration and other developmental processes, and have been implicated in numerous neurological disorder such as lissencephaly, epilepsy, autism, bipolar disorder, and schizophrenia. While it is highly unlikely that these cell types are orthologous, they could use similar mechanisms to guide neuronal processes to their targets. Other examples of possibly convergent cell types are glial cells (Klämbt, 2009). While vertebrate and invertebrate glial cell types are probably not orthologous, one can find functionally equivalent cell types between the two phyla. Perineurial and subperineurial glia in invertebrates and endothelial cells and astrocytes in vertebrates form septate junctions and provide the blood brain barrier. Similarly, wrapping and ensheathing glia in flies are responsible for wrapping neuronal axons, a function that is undertaken by oligodendrocytes and Schwann cells in vertebrates. Interestingly, despite the differences between vertebrates and invertebrates in gliogenesis, many of the terminal differentiation processes that guide migration and function are evolutionarily conserved (Klämbt, 2009).

5.3 Evolution of neuronal specification mechanisms

Adult cell types are the products of their developmental history and should be studied in that context. Different neuronal types may acquire similar or identical morphological, physiological, or molecular characters using different developmental paths. For this reason, neuronal type evolution cannot be studied except under the light of evolution of neuronal development.

As mentioned earlier, the generation of neuronal diversity in animals relies on two main intrinsic mechanisms: spatial and temporal patterning. While these mechanisms have been and are still being studied in exquisite detail in established genetic models, such as fruit flies and mice,

the lack of equivalent information from other representatives of the phylogenetic tree hinders our ability to understand how these mechanisms have evolved.

In the past, researchers had used candidate molecule approaches to study potential conservation of spatial and temporal factors. Given the small number of morphogens that are encoded in the animal genome, it comes as no surprise that the same molecules, such as Dpp/*Bmp* and Wg/*Wnt,* are being reused in different animals (Kaphingst and Kunes, 1994; Lee and Jessell, 1999). On the other hand, the main drivers of temporal patterning (i.e. the temporal transcription factors) are not shared between neuronal systems. In fact, even within one species (*Drosophila)*, the ventral nerve cord (Isshiki *et al.*, 2001) and optic lobes (Li *et al.*, 2013) are using different series of tTFs to specify their neuronal progeny. So far, only a few orthologous tTFs have been discovered to play a role in both insects and vertebrates, such as *Ikzf1* (the homolog of the *Drosophila* ventral nerve cord (VNC) first tTF, *hunchback*) that specifies young neural stem cells in the mouse cortex (Alsiö *et al.*, 2013) and retina (Elliott *et al.*, 2008), and *Pou2f1*/*Pou2f2* (Javed *et al.*, 2020) and *Casz1* (Mattar *et al.*, 2015) (the homologs of the later VNC tTFs, *Pdm1*/*2* and *castor*) that specify older retinal progenitors.

Recent transcriptomic studies have performed large-scale single-cell sequencing experiments in the developing spinal cord (Delile *et al.*, 2019) and cortex (Telley *et al.*, 2019) in mice, as well as the developing human brain (Nowakowski *et al.*, 2017; Manno *et al.*, 2020). Common temporal transcription factor codes can be found in neurons that are born sequentially: Onecut family transcription factors are expressed in neurons produced in early stages of neurogenesis, followed by the production of neurons that express *Pou2f2* and *Zfhx3*/*4* and, finally, neurons that express *Nfia*/*b* and *Neurod2*/*6* (Delile *et al.*, 2019; Sagner *et al.*, 2020). This suggests the existence of a temporal series in the progenitors that specifies neuronal identity. With the continued generation of single-cell sequencing datasets from different stages of neuronal development, future meta-analyses will be able to evaluate the presence of temporal series in the neuronal systems of different species and generate hypotheses for the evolution of temporal patterning.

In the meantime, the insect visual system has a unique feature that make it ideal to study the evolution of temporal patterning. As described earlier, because of the mechanism of generation of neuronal stem cells from the neuroepithelium, the developing optic lobe of *Drosophila* third instar larva contains neuroblasts of different ages. Single-cell sequencing of these neuroblasts in different insects followed by trajectory inference could give us a full

appreciation of the entire temporal progression of the series. Comparison of the temporal series in different insects will allow us to fully comprehend how a complicated temporal series that regulates neuronal diversity has evolved. Similarly, an analysis focused on the neuroepithelial cells may allow us to identify and compare spatial factors in the visual systems of different insects.

Such comparisons of developmental mechanisms between different insects will allow us not only to understand what is different between different visual systems, but also how these differences evolve by modulating developmental mechanisms.

6. Conclusion

The *Drosophila* visual system is a neurobiological model exceptionally well suited to single-cell studies because of the high repetition of most of its neuronal types. It is also to our knowledge the only one with two high resolution transcriptomic atlases of neuronal development, and for a large part (i.e. lamina and medulla neurons) a detailed view of both its connectome and of the TFs involved in neuronal specification. This presents a unique opportunity to create a complete “toolbox” to study neuronal function and development, by conducting complementary analyses and producing additional datasets. In this review, we have discussed the state-of-the-art as well as potential future projects stemming from our current knowledge. First, completing the characterization of transcriptomic diversity in the visual system by merging and reanalyzing the developmental atlases of the optic lobe, producing comparable atlases of the retina during pupal development and of the whole visual system during larval development, as well as annotating all the unidentified neuronal and glial clusters, would give a complete view of the cell type diversity in the optic lobe. Second, producing a developmental chromatin accessibility dataset of the same resolution, as the transcriptomic datasets will allow finer studies of gene expression regulation during development. This can be achieved by using single-cell technologies sequencing both the transcriptome and chromatin accessibility: by data transfer, all cells in the chromatin accessibility datasets could be matched to their corresponding transcriptomic cluster. Interestingly, (Bravo González-Blas *et al.*, 2020) found that, in the eye-antennal disk, the predicted accessibility of 77% of more than 700 enhancer-reporter lines (Jory *et al.*, 2012) is correlated with their activity. Therefore, in addition to gene regulatory inference, accessibility could also be used to gain genetic access to all neuronal types of the visual system. Third, obtaining lineage information for all neuronal clusters, including the temporal window and spatial localization of the neuroblasts producing them, in combination with transcriptomic and chromatin accessibility data for all clusters, would allow to explore the links between specification and terminal identity in

several dozens of cell types at once. Fourth, obtaining the missing pieces for the study of synaptic specificity: the interactome, associated affinities and subcellular localization throughout development for all CSMs, would allow us to define general mechanisms of synaptic specificity by bioinformatic modeling and experimental validations. And, finally, comparing cell type composition and specification mechanisms between different animals would help us understand how neuronal specification and neuronal diversity has evolved over time.

References

Achim, K. and Arendt, D. (2014) 'Structural evolution of cell types by step-wise assembly of cellular modules', *Curr Opin Genet Dev*. European Molecular Biology Laboratory, Developmental Biology Unit, Meyerhofstrasse 1, 69012 Heidelberg, Germany. European Molecular Biology Laboratory, Developmental Biology Unit, Meyerhofstrasse 1, 69012 Heidelberg, Germany. Electronic address: arendt@em, 27, pp. 102–108. doi: 10.1016/j.gde.2014.05.001.

Akin, O. *et al.* (2019) 'Cell-type-Specific Patterned Stimulus-Independent Neuronal Activity in the Drosophila Visual System during Synapse Formation.', *Neuron*, 101(5), pp. 894-904.e5. doi: 10.1016/j.neuron.2019.01.008.

Akin, O. and Zipursky, S. L. (2020) 'Activity regulates brain development in the fly.', *Current opinion in genetics & development*. England, 65, pp. 8–13. doi: 10.1016/j.gde.2020.04.005.

Alemany, A. *et al.* (2018) 'Whole-organism clone tracing using single-cell sequencing', *Nature*. Oncode Institute, Hubrecht Institute-KNAW (Royal Netherlands Academy of Arts and Sciences) and University Medical Center Utrecht, 3584 CT Utrecht, The Netherlands.: Nature, 556(7699), pp. 108–112. doi: 10.1038/nature25969.

Alsiö, J. M. *et al.* (2013) 'Ikaros promotes early-born neuronal fates in the cerebral cortex', *Proc Natl Acad Sci U S A*. Gurdon Institute and Department of Biochemistry, University of Cambridge, Cambridge CB2 1QN, United Kingdom., 110(8), pp. E716-25. doi: 10.1073/pnas.1215707110.

Alvarez, M. *et al.* (2020) 'Enhancing droplet-based single-nucleus RNA-seq resolution using the semi-supervised machine learning classifier DIEM', *Scientific Reports*, 10(1), p. 11019. doi: 10.1038/s41598-020-67513-5.

Apitz, H. and Salecker, I. (2015) 'A region-specific neurogenesis mode requires migratory progenitors in the Drosophila visual system', *Nature Neuroscience*. Nature Publishing Group, 18(1), pp. 46–55. doi: 10.1038/nn.3896.

Apitz, H. and Salecker, I. (2018) 'Spatio-temporal relays control layer identity of direction-selective neuron subtypes in Drosophila', *Nature Communications*. Nature Publishing Group, 9(1), p. 2295. doi: 10.1038/s41467-018-04592-z.

Arendt, D. *et al.* (2016) ‘The origin and evolution of cell types’, *Nat Rev Genet*. Developmental Biology Unit, European Molecular Biology Laboratory, Meyerhofstrasse 1, 69012 Heidelberg, Germany. Centre for Organismal Studies, University of Heidelberg, Im Neuenheimer Feld 230, 69120 Heidelberg, Germany. Department of Physiology, Develop, 17(12), pp. 744–757. doi: 10.1038/nrg.2016.127.

Arendt, D. (2020) ‘The Evolutionary Assembly of Neuronal Machinery.’, *Current biology : CB*. Curr Biol, 30(10), pp. R603–R616. doi: 10.1016/j.cub.2020.04.008.

Ariss, M. M. *et al.* (2018) ‘Single cell RNA-sequencing identifies a metabolic aspect of apoptosis in Rbf mutant’, *Nature Communications*, 9(1), p. 5024. doi: 10.1038/s41467-018-07540-z.

Aydin, B. *et al.* (2019) ‘Proneural factors Ascl1 and Neurog2 contribute to neuronal subtype identities by establishing distinct chromatin landscapes’, *Nature Neuroscience*. Nature Publishing Group, 22(6), pp. 897–908. doi: 10.1038/s41593-019-0399-y.

Azevedo, F. A. *et al.* (2009) ‘Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain’, *J Comp Neurol*. Instituto de Ciencias Biomedicas, Universidade Federal do Rio de Janeiro, Cidade Universitaria, Rio de Janeiro, Brazil., 513(5), pp. 532–541. doi: 10.1002/cne.21974.

Baker, K. D. *et al.* (2003) ‘The Drosophila Orphan Nuclear Receptor DHR38 Mediates an Atypical Ecdysteroid Signaling Pathway’, *Cell*, 113(6), pp. 731–742. doi: https://doi.org/10.1016/S0092-8674(03)00420-3.

Bate, M. and Martinez Arias, A. (2009) *The development of Drosophila melanogaster*. Cold Spring Harbor Laboratory Press. Available at: https://www.cshlpress.com/default.tpl?cart=16143150022327842 30&fromlink=T&linkaction=full&linksortby=oop_title&--eqSKUdatarq=850 (Accessed: 25 February 2021).

Bausenwein, B. and Fischbach, K. F. (1992) ‘Activity labeling patterns in the medulla of Drosophila melanogaster caused by motion stimuli’, *Cell Tissue Res*. Institut fur Biologie III, Freiburg, Federal Republic of Germany., 270(1), pp. 25–35. Available at: https://www.ncbi.nlm.nih.gov/pubmed/1423522.

Bertet, C. *et al.* (2014) ‘Temporal patterning of neuroblasts controls Notch-mediated cell survival through regulation of Hid or Reaper’, *Cell*. Department of Biology, New York University, 100 Washington Square East, New York, NY 10003, USA. Department of Biology, New York

University, 100 Washington Square East, New York, NY 10003, USA; CGSB, New York University Abu Dhabi, United Arab Emirates. El, 158(5), pp. 1173–1186. doi: 10.1016/j.cell.2014.07.045.

Borst, A., Haag, J. and Reiff, D. F. (2010) 'Fly Motion Vision', *Annual Review of Neuroscience*, 33(1), pp. 49–70. doi: 10.1146/annurev-neuro-060909-153155.

Branson, K. *et al.* (2009) 'High-throughput ethomics in large groups of Drosophila', *Nat Methods*. California Institute of Technology, Pasadena, California, USA., 6(6), pp. 451–457. doi: 10.1038/nmeth.1328.

Bravo González-Blas, C. *et al.* (2020) 'Identification of genomic enhancers through spatial integration of single-cell transcriptomics and epigenomics', *Molecular Systems Biology*. John Wiley & Sons, Ltd, 16(5), p. e9438. doi: 10.15252/msb.20209438.

Cande, J. *et al.* (2018) 'Optogenetic dissection of descending behavioral control in Drosophila', *eLife*, 7. doi: 10.7554/eLife.34275.

Carrillo, R. A. *et al.* (2015) 'Control of Synaptic Connectivity by a Network of Drosophila IgSF Cell Surface Proteins', *Cell*. Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA. Dept. of Biochemistry and Molecular Biology, University of Chicago, Chicago, IL. Howard Hughes Medical Institute, Depts. of Molecular and Cellular Physiology, 163(7), pp. 1770–1782. doi: 10.1016/j.cell.2015.11.022.

Charest, J. *et al.* (2020) 'Combinatorial Action of Temporally Segregated Transcription Factors.', *Developmental cell*. Elsevier, 55(4), pp. 483-499.e7. doi: 10.1016/j.devcel.2020.09.002.

Cocanougher, B. T. *et al.* (2020) 'Comparative single-cell transcriptomics of complete insect nervous systems', *bioRxiv*, p. 785931. doi: 10.1101/785931.

Cosmanescu, F. *et al.* (2018) 'Neuron-Subtype-Specific Expression, Interaction Affinities, and Specificity Determinants of DIP/Dpr Cell Recognition Proteins.', *Neuron*, 100(6), pp. 1385-1400.e6. doi: 10.1016/j.neuron.2018.10.046.

Courgeon, M. and Desplan, C. (2019) 'Coordination between stochastic and deterministic specification in the Drosophila visual system', *Science*, 366(6463), p. eaay6727. doi: 10.1126/science.aay6727.

Dalva, M. B., McClelland, A. C. and Kayser, M. S. (2007) 'Cell adhesion molecules: signalling

functions at the synapse', *Nature Reviews Neuroscience*. Nature Publishing Group, 8(3), pp. 206–220. doi: 10.1038/nrn2075.

Davie, K. *et al.* (2018) 'A Single-Cell Transcriptome Atlas of the Aging Drosophila Brain', *Cell*. VIB Center for Brain & Disease Research, KU Leuven, Leuven 3000, Belgium; Department of Human Genetics KU Leuven, Leuven 3000, Belgium. VIB Center for Brain & Disease Research, KU Leuven, Leuven 3000, Belgium; Department of Neurosciences, KU Leuven, Leuve, 174(4), pp. 982-998 e20. doi: 10.1016/j.cell.2018.05.057.

Davis, F. P. *et al.* (2020) 'A genetic, genomic, and computational resource for exploring neural circuit function', *eLife*. Edited by H. J. Bellen and K. VijayRaghavan. eLife Sciences Publications, Ltd, 9, p. e50901. doi: 10.7554/eLife.50901.

Delile, J. *et al.* (2019) 'Single cell transcriptomics reveals spatial and temporal dynamics of gene expression in the developing mouse spinal cord.', *Development (Cambridge, England)*. Development, 146(12). doi: 10.1242/dev.173807.

Diao, F. *et al.* (2015) 'Plug-and-play genetic access to drosophila cell types using exchangeable exon cassettes', *Cell Rep*. Laboratory of Molecular Biology, National Institute of Mental Health, National Institutes of Health, 9000 Rockville Pike, Bethesda, MD 20892, USA. Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK. Centro Interdisciplin, 10(8), pp. 1410–1421. doi: 10.1016/j.celrep.2015.01.059.

Ding, M. *et al.* (2007) 'Spatial Regulation of an E3 Ubiquitin Ligase Directs Selective Synapse Elimination', *Science*, 317(5840), pp. 947 LP – 951. doi: 10.1126/science.1145727.

Elliott, J. *et al.* (2008) 'Ikaros confers early temporal competence to mouse retinal progenitor cells.', *Neuron*, 60, pp. 26–39. doi: 10.1016/j.neuron.2008.08.008.

Erclik, T. *et al.* (2017) 'Integration of temporal and spatial patterning generates neural diversity', *Nature*. Department of Biology, New York University, New York, New York 10003, USA. Department of Biology, University of Toronto at Mississauga, Mississauga, Ontario L5L-1C6, Canada. Center for Genomics and Systems Biology, New York University Abu Dhabi, Abu Dhabi, 541(7637), pp. 365–370. doi: 10.1038/nature20794.

Estacio-Gómez, A. *et al.* (2020) 'Dynamic neurotransmitter specific transcription factor expression profiles during Drosophila development.', *Biology open*. Biol Open, 9(5). doi: 10.1242/bio.052928.

Fischbach, K. F. and Dittrich, A. P. (1989) ‘The optic lobe of Drosophila melanogaster. I. A Golgi analysis of wild-type structure’, *Cell Tissue Res*, 258, pp. 441–445. Available at: internal-pdf://0719885386/Fischbach_optic_lobe_Golgipaper.pdf.

Fleming, S. J., Marioni, J. C. and Babadi, M. (2019) ‘CellBender remove-background: a deep generative model for unsupervised removal of background noise from scRNA-seq datasets’, *bioRxiv*. Cold Spring Harbor Laboratory. doi: 10.1101/791699.

Ha, K. C. H. *et al.* (2021) ‘Differential contribution of transcriptomic regulatory layers in the definition of neuronal identity’, *Nature Communications*. Nature Publishing Group, 12(1), p. 335. doi: 10.1038/s41467-020-20483-8.

Hafemeister, C. and Satija, R. (2019) ‘Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression.’, *Genome biology*, 20(1), p. 296. doi: 10.1186/s13059-019-1874-1.

Handler, A. M. (1982) ‘Ecdysteroid titers during pupal and adult development in Drosophila melanogaster’, *Developmental Biology*, 93, pp. 73–82. doi: 10.1016/0012-1606(82)90240-8.

Hie, B. *et al.* (2020) ‘Computational Methods for Single-Cell RNA Sequencing’, *Annual Review of Biomedical Data Science*. Annual Reviews, 3(1), pp. 339–364. doi: 10.1146/annurev-biodatasci-012220-100601.

Hiesinger, P. R. *et al.* (2006) ‘Activity-Independent Prespecification of Synaptic Partners in the Visual Map of Drosophila’, *Current Biology*, 16(18), pp. 1835–1843. doi: https://doi.org/10.1016/j.cub.2006.07.047.

Hobert, O. and Kratsios, P. (2019) ‘Neuronal identity control by terminal selectors in worms, flies, and chordates.’, *Current opinion in neurobiology*. England, 56, pp. 97–105. doi: 10.1016/j.conb.2018.12.006.

Hofbauer, A. and Campos-Ortega, J. a. (1990) ‘Proliferation pattern and early differentiation of the optic lobes in Drosophila melanogaster’, *Roux’s Archives of Developmental Biology*, 198, pp. 264–274. doi: 10.1007/BF00377393.

Holguera, I. and Desplan, C. (2018) ‘Neuronal specification in space and time.’, *Science (New York, N.Y.)*. American Association for the Advancement of Science, 362(6411), pp. 176–180. doi: 10.1126/science.aas9435.

Hörmann, N. *et al.* (2020) ‘A combinatorial code of transcription factors specifies subtypes of visual motion-sensing neurons in Drosophila.’, *Development (Cambridge, England)*, 147(9). doi: 10.1242/dev.186296.

Isshiki, T. *et al.* (2001) ‘Drosophila neuroblasts sequentially express transcription factors which specify the temporal identity of their neuronal progeny’, *Cell*. Institute of Neuroscience/Institute of Molecular Biology, Howard Hughes Medical Institute, 1254 University of Oregon, Eugene, OR 97403, USA., 106(4), pp. 511–521. Available at: https://www.ncbi.nlm.nih.gov/pubmed/11525736.

Izutsu, M. *et al.* (2015) ‘Dynamics of Dark-Fly Genome Under Environmental Selections.’, *G3 (Bethesda, Md.)*. G3 (Bethesda), 6(2), pp. 365–76. doi: 10.1534/g3.115.023549.

Jain, S. *et al.* (2020) ‘A Global Temporal Genetic Program for Neural Circuit Formation’, *bioRxiv*, p. 2020.09.18.304410. doi: 10.1101/2020.09.18.304410.

Javed, A. *et al.* (2020) ‘Pou2f1 and Pou2f2 cooperate to control the timing of cone photoreceptor production in the developing mouse retina.’, *Development (Cambridge, England)*. Oxford University Press for The Company of Biologists Limited, 147(18). doi: 10.1242/dev.188730.

Johnson, E., Kath, W. and Mani, M. (2020) ‘A Statistical Approach to Dimensionality Reduction Reveals Scale and Structure in scRNA-seq Data’, *bioRxiv*. Cold Spring Harbor Laboratory. doi: 10.1101/2020.11.18.389031.

Jory, A. *et al.* (2012) ‘A Survey of 6,300 Genomic Fragments for cis-Regulatory Activity in the Imaginal Discs of Drosophila melanogaster’, *Cell Reports*, 2, pp. 1014–1024. doi: 10.1016/j.celrep.2012.09.010.

Kaphingst, K. and Kunes, S. (1994) ‘Pattern formation in the visual centers of the Drosophila brain: Wingless acts via decapentaplegic to specify the dorsoventral axis’, *Cell*, 78, pp. 437–448. doi: 10.1016/0092-8674(94)90422-7.

Katsamba, P. *et al.* (2009) ‘Linking molecular affinity and cellular specificity in cadherin-mediated adhesion.’, *Proceedings of the National Academy of Sciences of the United States of America*, 106(28), pp. 11594–11599. doi: 10.1073/pnas.0905349106.

Klämbt, C. (2009) ‘Modes and regulation of glial migration in vertebrates and invertebrates’,

Nature Reviews Neuroscience. Nature Publishing Group, 10(11), pp. 769–779. doi: 10.1038/nrn2720.

Konstantinides, N. *et al.* (2018) 'Phenotypic Convergence: Distinct Transcription Factors Regulate Common Terminal Features', *Cell*. Department of Biology, New York University, New York, NY 10003, USA. Electronic address: nk1845@nyu.edu. New York University Abu Dhabi, Saadiyat Island, Abu Dhabi, United Arab Emirates. Department of Biology, New York University, New York, NY 10003, USA; , 174(3), pp. 622-635 e13. doi: 10.1016/j.cell.2018.05.021.

Konstantinides, N., Degabriel, S. and Desplan, C. (2018) 'Neuro-evo-devo in the single cell sequencing era.', *Current opinion in systems biology*, 11, pp. 32–40. doi: 10.1016/j.coisb.2018.08.001.

Konstantinides, N., Rossi, A. M. and Desplan, C. (2015) 'Common temporal identity factors regulate neuronal diversity in fly ventral nerve cord and mouse retina', *Neuron*. Department of Biology, New York University, 1009 Silver Center, 100 Washington Square East, New York, NY 10003, USA. Department of Biology, New York University, 1009 Silver Center, 100 Washington Square East, New York, NY 10003, USA. Electronic address: c, 85(3), pp. 447–449. doi: 10.1016/j.neuron.2015.01.016.

Kurmangaliyev, Y. Z. *et al.* (2019) 'Modular transcriptional programs separately define axon and dendrite connectivity.', *eLife*, 8. doi: 10.7554/eLife.50822.

Kurmangaliyev, Y. Z. *et al.* (2020) 'Transcriptional Programs of Circuit Assembly in the Drosophila Visual System.', *Neuron*. United States. doi: 10.1016/j.neuron.2020.10.006.

Lago-Baldaia, I., Fernandes, V. M. and Ackerman, S. D. (2020) 'More Than Mortar: Glia as Architects of Nervous System Development and Disease', *Frontiers in Cell and Developmental Biology*, 8, p. 1527. doi: 10.3389/fcell.2020.611269.

Lancichinetti, A. and Fortunato, S. (2011) 'Limits of modularity maximization in community detection.', *Physical review. E, Statistical, nonlinear, and soft matter physics*. United States, 84(6 Pt 2), p. 66122. doi: 10.1103/PhysRevE.84.066122.

Langley, J. N. (1895) 'Note on Regeneration of Prae-Ganglionic Fibres of the Sympathetic.', *The Journal of physiology*, 18(3), pp. 280–284. doi: 10.1113/jphysiol.1895.sp000566.

Lee, K. J. and Jessell, T. M. (1999) 'The specification of dorsal cell fated in the vertebrate

central nervous system', *Annual Review of Neuroscience*. Annual Reviews 4139 El Camino Way, P.O. Box 10139, Palo Alto, CA 94303-0139, USA, 22(1), pp. 261–294. doi: 10.1146/annurev.neuro.22.1.261.

Lee, T. and Luo, L. (2001) 'Mosaic analysis with a repressible cell marker (MARCM) for Drosophila neural development', *Trends Neurosci*. Department of Cell and Structural Biology, University of Illinois, Urbana-Champaign, Urbana, IL 61801, USA., 24(5), pp. 251–254. Available at: https://www.ncbi.nlm.nih.gov/pubmed/11311363.

Li, H. *et al.* (2017) 'Classifying Drosophila Olfactory Projection Neuron Subtypes by Single-Cell RNA Sequencing.', *Cell*. Cell, 171(5), pp. 1206-1220.e22. doi: 10.1016/j.cell.2017.10.019.

Li, H. *et al.* (2020) 'Single-Cell Transcriptomes Reveal Diverse Regulatory Strategies for Olfactory Receptor Expression and Axon Targeting.', *Current biology : CB*. Curr Biol, 30(7), pp. 1189-1198.e5. doi: 10.1016/j.cub.2020.01.049.

Li, J. *et al.* (2020) 'Structure of the teneurin-latrophilin complex: Alternative splicing controls synapse specificity by a novel mechanism', *bioRxiv*. Cold Spring Harbor Laboratory, p. 2020.01.13.901090. doi: 10.1101/2020.01.13.901090.

Li, X. *et al.* (2013) 'Temporal patterning of Drosophila medulla neuroblasts controls neural fates', *Nature*. Department of Biology, New York University, 100 Washington Square East, New York, New York 10003, USA., 498(7455), pp. 456–462. doi: 10.1038/nature12319.

Luecken, M. D. and Theis, F. J. (2019) 'Current best practices in single-cell RNA-seq analysis: a tutorial.', *Molecular systems biology*, 15(6), p. e8746. doi: 10.15252/msb.20188746.

Manno, G. La *et al.* (2020) 'Molecular architecture of the developing mouse brain', *bioRxiv*. Cold Spring Harbor Laboratory, p. 2020.07.02.184051. doi: 10.1101/2020.07.02.184051.

Mattar, P. *et al.* (2015) 'A conserved regulatory logic controls temporal identity in mouse neural progenitors', *Neuron*. Cellular Neurobiology Research Unit, Institut de recherches cliniques de Montreal (IRCM), Montreal, QC H2W 1R7, Canada. Department of Cell and Molecular Biology, Karolinska Institutet, 171 77 Stockholm, Sweden. The Solomon H. Snyder Department of Neurosci, 85(3), pp. 497–504. doi: 10.1016/j.neuron.2014.12.052.

Meinertzhagen, I. A. and O'Neil, S. D. (1991) 'Synaptic organization of columnar elements in the lamina of the wild type in Drosophila melanogaster', *Journal of Comparative Neurology*. John

Wiley & Sons, Ltd, 305(2), pp. 232–263. doi: https://doi.org/10.1002/cne.903050206.

Meinertzhagen, I. A. and Sorra, K. E. (2001) ‘Synaptic organization in the fly’s optic lamina: few cells, many synapses and divergent microcircuits’, *Prog Brain Res*. Neuroscience Institute, Life Sciences Centre, Dalhousie University, Halifax, NS, Canada B3H 4J1. IAM@IS.DAL.CA, 131, pp. 53–69. Available at: https://www.ncbi.nlm.nih.gov/pubmed/11420968.

Mercatelli, D. *et al.* (2020) ‘Gene regulatory network inference resources: A practical overview.’, *Biochimica et biophysica acta. Gene regulatory mechanisms*. Biochim Biophys Acta Gene Regul Mech, 1863(6), p. 194430. doi: 10.1016/j.bbagrm.2019.194430.

Muthukumar, A. K., Stork, T. and Freeman, M. R. (2014) ‘Activity-dependent regulation of astrocyte GAT levels during synaptogenesis’, *Nature Neuroscience*, 17(10), pp. 1340–1350. doi: 10.1038/nn.3791.

Naidu, V. G. *et al.* (2020) ‘Temporal progression of Drosophila medulla neuroblasts generates the transcription factor combination to control T1 neuron morphogenesis.’, *Developmental biology*. Dev Biol, 464(1), pp. 35–44. doi: 10.1016/j.ydbio.2020.05.005.

Nériec, N. *et al.* (2016) ‘From the Eye to the Brain: Development of the Drosophila Visual System’, *Current Topics in Developmental Biology*. Center for Genomics & Systems Biology, New York University, Abu Dhabi, UAE; Department of Biology, New York University, New York, USA. Center for Genomics & Systems Biology, New York University, Abu Dhabi, UAE; Department of Biology, New York University, , 116, pp. 247–271. doi: 10.1016/bs.ctdb.2015.11.032.

Nern, A., Pfeiffer, B. D. and Rubin, G. M. (2015) ‘Optimized tools for multicolor stochastic labeling reveal diverse stereotyped cell arrangements in the fly visual system’, *Proc Natl Acad Sci U S A*. Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA 20147-2408. Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA 20147-2408 rubing@janelia.hhmi.org., 112(22), pp. E2967-76. doi: 10.1073/pnas.1506763112.

Norimoto, H. *et al.* (2020) ‘A claustrum in reptiles and its role in slow-wave sleep’, *Nature*. Nature Publishing Group, 578(7795), pp. 413–418. doi: 10.1038/s41586-020-1993-6.

Nowakowski, T. J. *et al.* (2017) ‘Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex.’, *Science (New York, N.Y.)*. American Association for the Advancement of Science, 358(6368), pp. 1318–1323. doi:

10.1126/science.aap8809.

Okawa, H. *et al.* (2014) ‘Illuminating the multifaceted roles of neurotransmission in shaping neuronal circuitry.’, *Neuron*, 83(6), pp. 1303–1318. doi: 10.1016/j.neuron.2014.08.029.

Özel, M. N. *et al.* (2019) ‘Serial Synapse Formation through Filopodial Competition for Synaptic Seeding Factors.’, *Developmental cell*, 50(4), pp. 447-461.e8. doi: 10.1016/j.devcel.2019.06.014.

Özel, M. N. *et al.* (2020) ‘Neuronal diversity and convergence in a visual system developmental atlas.’, *Nature*. England. doi: 10.1038/s41586-020-2879-3.

Ozkan, E. *et al.* (2013) ‘An extracellular interactome of immunoglobulin and LRR proteins reveals receptor-ligand networks’, *Cell*. Departments of Molecular and Cellular Physiology and Structural Biology, Stanford University School of Medicine, Stanford, CA 94305, USA., 154(1), pp. 228–239. doi: 10.1016/j.cell.2013.06.006.

Pinto-Teixeira, F., Konstantinides, N. and Desplan, C. (2016) ‘Programmed cell death acts at different stages of Drosophila neurodevelopment to shape the central nervous system’, *FEBS Lett*. Department of Biology, New York University, NY, USA. Center for Genomics and Systems Biology, New York University Abu Dhabi, UAE., 590(15), pp. 2435–2453. doi: 10.1002/1873-3468.12298.

Plazaola-Sasieta, H. *et al.* (2017) ‘Untangling the wiring of the Drosophila visual system: developmental principles and molecular strategies.’, *Journal of neurogenetics*. England, 31(4), pp. 231–249. doi: 10.1080/01677063.2017.1391249.

Raj, B. *et al.* (2018) ‘Simultaneous single-cell profiling of lineages and cell types in the vertebrate brain’, *Nat Biotechnol*. Department of Molecular and Cellular Biology, Harvard University, Cambridge, Massachusetts, USA. Allen Discovery Center for Cell Lineage Tracing, Seattle, Washington, USA. Department of Systems Biology, Harvard Medical School, Boston, Massachusetts, USA. : Nature Publishing Group, 36(5), pp. 442–450. doi: 10.1038/nbt.4103.

Ravenscroft, T. A. *et al.* (2020) ‘Drosophila Voltage-Gated Sodium Channels Are Only Expressed in Active Neurons and Are Localized to Distal Axonal Initial Segment-like Domains.’, *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 40(42), pp. 7999–8024. doi: 10.1523/JNEUROSCI.0142-20.2020.

Rivera-Alba, M. *et al.* (2011) ‘Wiring economy and volume exclusion determine neuronal placement in the Drosophila brain.’, *Current biology : CB*, 21(23), pp. 2000–2005. doi: 10.1016/j.cub.2011.10.022.

Saelens, W. *et al.* (2019) ‘A comparison of single-cell trajectory inference methods’, *Nature Biotechnology*. Nature Publishing Group, 37(5), pp. 547–554. doi: 10.1038/s41587-019-0071-9.

Sagner, A. *et al.* (2020) ‘Temporal patterning of the central nervous system by a shared transcription factor code’, *bioRxiv*. Cold Spring Harbor Laboratory, p. 2020.11.10.376491. doi: 10.1101/2020.11.10.376491.

Sanes, J. R. and Yamagata, M. (2009) ‘Many paths to synaptic specificity.’, *Annual review of cell and developmental biology*. United States, 25, pp. 161–195. doi: 10.1146/annurev.cellbio.24.110707.175402.

Sanes, J. R. and Zipursky, S. L. (2020) ‘Synaptic Specificity, Recognition Molecules, and Assembly of Neural Circuits.’, *Cell*. United States, 181(3), pp. 536–556. doi: 10.1016/j.cell.2020.04.008.

Scala, F. *et al.* (2020) ‘Phenotypic variation of transcriptomic cell types in mouse motor cortex’, *Nature*. Nature Publishing Group, pp. 1–7. doi: 10.1038/s41586-020-2907-3.

Schwabe, T. *et al.* (2014) ‘Differential Adhesion Determines the Organization of Synaptic Fascicles in the Drosophila Visual System’, *Current Biology*, 24(12), pp. 1304–1313. doi: https://doi.org/10.1016/j.cub.2014.04.047.

Sen, S. Q. *et al.* (2019) ‘Neuroblast-specific open chromatin allows the temporal transcription factor, Hunchback, to bind neuroblast-specific loci’, *eLife*, 8. doi: 10.7554/eLife.44036.

Shinomiya, K., Huang, G., *et al.* (2019) ‘Comparisons between the ON- and OFF-edge motion pathways in the Drosophila brain.’, *eLife*, 8. doi: 10.7554/eLife.40025.

Shinomiya, K., Horne, J. A., *et al.* (2019) ‘The Organization of the Second Optic Chiasm of the Drosophila Optic Lobe.’, *Frontiers in neural circuits*, 13, p. 65. doi: 10.3389/fncir.2019.00065.

Shubin, N., Tabin, C. and Carroll, S. (1997) ‘Fossils, genes and the evolution of animal limbs’, *Nature*. Nature Publishing Group, 388(6643), pp. 639–648. doi: 10.1038/41710.

Shubin, N., Tabin, C. and Carroll, S. (2009) ‘Deep homology and the origins of evolutionary

novelty.', *Nature*, 457, pp. 818–823. doi: 10.1038/nature07891.

Soriano, E. and Del Río, J. A. (2005) 'The cells of cajal-retzius: still a mystery one century after.', *Neuron*. United States, 46(3), pp. 389–394. doi: 10.1016/j.neuron.2005.04.019.

Spanjaard, B. *et al.* (2018) 'Simultaneous lineage tracing and cell-type identification using CRISPR-Cas9-induced genetic scars', *Nat Biotechnol*. Berlin Institute for Medical Systems Biology, Max Delbruck Center for Molecular Medicine, Berlin, Germany. DFG-Center for Regenerative Therapies Dresden, Technische Universitat Dresden, Dresden, Germany.: Nat Biotechnol, 36(5), pp. 469–473. doi: 10.1038/nbt.4124.

Sperry, R. W. (1963) 'CHEMOAFFINITY IN THE ORDERLY GROWTH OF NERVE FIBER PATTERNS AND CONNECTIONS', *Proceedings of the National Academy of Sciences*, 50(4), pp. 703 LP – 710. doi: 10.1073/pnas.50.4.703.

Stassen, S. V *et al.* (2020) 'PARC: ultrafast and accurate clustering of phenotypic data of millions of single cells.', *Bioinformatics (Oxford, England)*, 36(9), pp. 2778–2786. doi: 10.1093/bioinformatics/btaa042.

Steinberg, M. S. (1970) 'Does differential adhesion govern self-assembly processes in histogenesis? Equilibrium configurations and the emergence of a hierarchy among populations of embryonic cells.', *The Journal of experimental zoology*. United States, 173(4), pp. 395–433. doi: 10.1002/jez.1401730406.

Suzuki, T. *et al.* (2013) 'A temporal mechanism that produces neuronal diversity in the Drosophila visual center', *Dev Biol*. Laboratory of Developmental Neurobiology, Brain/Liver Interface Medicine Research Center, Kanazawa University, 13-1 Takaramachi Kanazawa-shi, Ishikawa 920-8641, Japan., 380(1), pp. 12–24. doi: 10.1016/j.ydbio.2013.05.002.

Takemura, S. Y. *et al.* (2013) 'A visual motion detection circuit suggested by Drosophila connectomics', *Nature*. Janelia Farm Research Campus, HHMI, Ashburn, Virginia 20147, USA., 500(7461), pp. 175–181. doi: 10.1038/nature12450.

Takemura, S. Y., Lu, Z. and Meinertzhagen, I. A. (2008) 'Synaptic circuits of the Drosophila optic lobe: the input terminals to the medulla', *J Comp Neurol*. Department of Psychology, Life Sciences Centre, Dalhousie University, Halifax, Canada., 509(5), pp. 493–513. doi: 10.1002/cne.21757.

Takemura, Shin-ya *et al.* (2015) ‘Synaptic circuits and their variations within different columns in the visual system of Drosophila.’, *Proceedings of the National Academy of Sciences of the United States of America*, 112(44), pp. 13711–13716. doi: 10.1073/pnas.1509820112.

Tan, L. *et al.* (2015) ‘Ig Superfamily Ligand and Receptor Pairs Expressed in Synaptic Partners in Drosophila’, *Cell*. Department of Biological Chemistry, HHMI, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA. Department of Molecular and Human Genetics, HHMI, Baylor College of Medicine, Houston, TX 77030, USA. Department , 163(7), pp. 1756–1769. doi: 10.1016/j.cell.2015.11.021.

Telley, L. *et al.* (2019) ‘Temporal patterning of apical progenitors and their daughter neurons in the developing neocortex.’, *Science (New York, N.Y.)*. American Association for the Advancement of Science, 364(6440). doi: 10.1126/science.aav2522.

Tian, L. *et al.* (2019) ‘Benchmarking single cell RNA-sequencing analysis pipelines using mixture control experiments’, *Nature Methods*, 16(6), pp. 479–487. doi: 10.1038/s41592-019-0425-8.

Tosches, M. A. *et al.* (2018) ‘Evolution of pallium, hippocampus, and cortical cell types revealed by single-cell transcriptomics in reptiles’, *Science*. Max Planck Institute for Brain Research, Max-von-Laue Strasse 4, 60438 Frankfurt am Main, Germany. gilles.laurent@brain.mpg.de maria.tosches@brain.mpg.de. Max Planck Institute for Brain Research, Max-von-Laue Strasse 4, 60438 Frankfurt am Main, Germany., 360(6391), pp. 881–888. doi: 10.1126/science.aar4237.

Traag, V. A., Waltman, L. and van Eck, N. J. (2019) ‘From Louvain to Leiden: guaranteeing well-connected communities.’, *Scientific reports*, 9(1), p. 5233. doi: 10.1038/s41598-019-41695-z.

Tran, H. T. N. *et al.* (2020) ‘A benchmark of batch-effect correction methods for single-cell RNA sequencing data’, *Genome Biology*. BioMed Central, 21(1), p. 12. doi: 10.1186/s13059-019-1850-9.

Trush, O. *et al.* (2019) ‘N-Cadherin Orchestrates Self-Organization of Neurons within a Columnar Unit in the Drosophila Medulla.’, *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 39(30), pp. 5861–5880. doi: 10.1523/JNEUROSCI.3107-18.2019.

Valdes-Aleman, J. *et al.* (2021) ‘Comparative Connectomics Reveals How Partner Identity, Location, and Activity Specify Synaptic Connectivity in Drosophila.’, *Neuron*, 109(1), pp. 105-122.e7. doi: 10.1016/j.neuron.2020.10.004.

Wu, M. *et al.* (2016) ‘Visual projection neurons in the Drosophila lobula link feature detection to distinct behavioral programs’, *Elife*. Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, United States., 5. doi: 10.7554/eLife.21022.

Xi, N. M. and Li, J. J. (2020) ‘Benchmarking Computational Doublet-Detection Methods for Single-Cell RNA Sequencing Data.’, *Cell systems*. United States. doi: 10.1016/j.cels.2020.11.008.

Xu, C. *et al.* (2019) ‘Control of Synaptic Specificity by Establishing a Relative Preference for Synaptic Partners’, *Neuron*. 2019/07/09, 103(5), pp. 865-877.e7. doi: 10.1016/j.neuron.2019.06.006.

Yang, S. *et al.* (2020) ‘Decontamination of ambient RNA in single-cell RNA-seq with DecontX’, *Genome Biology*, 21(1), p. 57. doi: 10.1186/s13059-020-1950-6.

Young, M. D. and Behjati, S. (2020) ‘SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data’, *GigaScience*, 9(12). doi: 10.1093/gigascience/giaa151.

Zhang, K. X. *et al.* (2016) ‘Rapid Changes in the Translatome during the Conversion of Growth Cones to Synaptic Terminals.’, *Cell reports*. Cell Rep, 14(5), pp. 1258–1271. doi: 10.1016/j.celrep.2015.12.102.

Table 1: single-cell RNA- and ATAC-seq datasets produced in the *Drosophila* visual system. The median number of genes per cell is affected by the version of the technology used, and by sorting or washing the cell suspension before sequencing: this removes cell debris and RNA from burst cells (ambient RNA), but could also increase biases by selecting for less fragile populations or affecting gene expression. The values are indicated after filtering based on quality control when reported in the studies, or before filtering otherwise. The number of clusters is only indicative, as many studies subset and re-cluster part of the dataset. ALH = after larval hatching, APF = after pupal formation, CNS = central nervous system, KC = Kenyon Cells, NA = not applicable, VNC = ventral nerve chord.

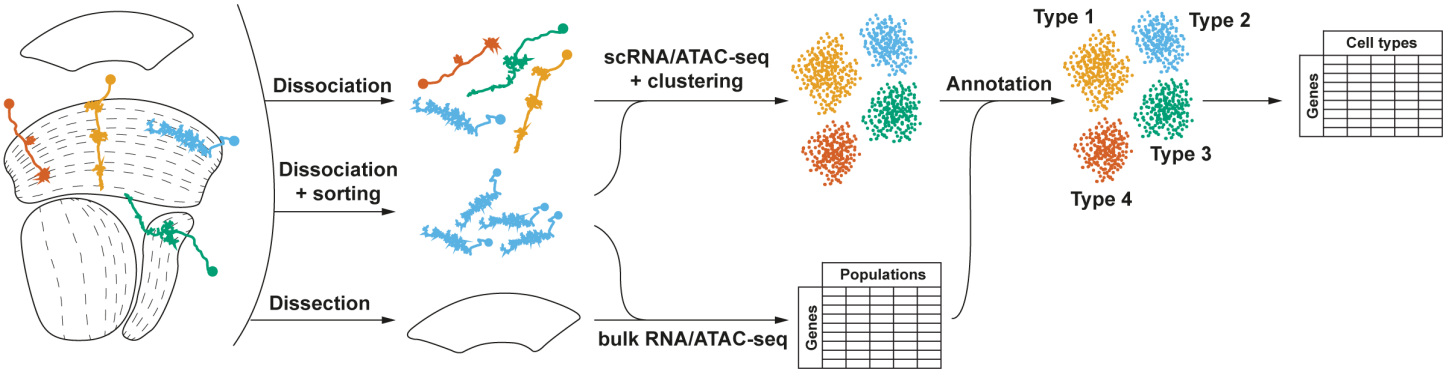
Reference	Material sequenced	Stages sequenced	Sex	Strains sequenced	Number of cells analyzed	Median number of genes/cell	Cell suspension sorting or washing	Sequencing technology	Batch effect correction	Number of clusters
(Ariss *et al.*, 2018)	eye disc	3rd instar larva	-	*y[1] v[1]; P{y[+t7.7]=CaryP}attP2*	11,416	-	-	Drop-Seq	-	15
				Rbf^{120a} mutant	5,203					
(Davie *et al.*, 2018)	Brain	Adult (0, 1, 3, 6, 9, 15, 30 and 50 days old)	F+M	*w*^{1118}, DGRP-551	56,902 (157,390 sequenced)	1,308 to 810	Washed	10x v.2	-	87
(Konstantinides *et al.*, 2018)	Optic lobe	Adult (3 days old)	F	Canton-S	54,974	275	-	Drop-Seq	-	52
(Brunet Avalos *et al.*, 2019)	Brain	1st instar larva	-	Canton-S	4,349	1,434	Washed	10x v.3	Bioinformatic, between but not within conditions	29
	Brain (starved)				4,347	1,962				-
(Kurmangaliyev *et al.*, 2019)	Sorted cells (~T4/T5 neurons)	48h APF	F	*23G12-Gal4* crossed to *UAS-H2A::GFP*	3,557	1,633	FACS	10x v.2	NA	8
		24h APF			3,071	1,447				
(Bravo González-Blas *et al.*, 2020)	Sorted eye-antennal discs	3rd instar larva	-		3,531	3,094	Washed	10x v.2	-	17
				hybrid of DGRP-551, DGRP-360, DGRP-907, and DGRP-913	15,387 single-cell ATAC profiles	-	Washed	10x		-
				sens-F2B-GFP	384 single-cell ATAC profiles	-	FACS	Fluidigm C1		-
(Cocanougher *et al.*, 2020)	CNS, brain, VNC (including after optogenetic sting or KC overactivation)	1, 24, 48 and 96h ALH	-	-	202,107	-	-	10x v.2	-	70
(Hörmann *et al.*, 2020)	Sorted cells (~T4/T5 neurons)	24, 36, 48, 60 and 72h APF	F+M	*SS00324-Gal4* recombined with *UAS-mCD8:: GFP*	~44,000	1,627	FACS	10x v.3	Bioinformatic	8
(Jain *et al.*, 2020)	Sorted cells	24, 36, 48, 60, 72, 84	F+M	*w*; *UAS-H2A-GFP*; *9B08-Gal4/Tm6B*	8,269		FACS	10x v.3	Bioinformatic	5

	(~Lamina neurons)	and 96h APF		*w*; *UAS-H2A-GFP*; *9B08-Gal4/Tm6B* crossed with DGRP, EcR^{DN}, *EcR* RNAi, *Hr3* RNAi, *tdTom* or *w*RNAi	-					
(Kurmangaliyev *et al.*, 2020)	Optic lobe	0, 12, 24, 36, 48, 60, 72, 84 and 96h APF	F (M cells removed)	w^{1118}, and w^{1118} crossed with 35 DGRP strains	208,976	1,650	FACS	10X v.3	Bioinformatic, and pooled sequencing for part of the data	P24-P96: 196
(Michki *et al.*, 2020)	Type-II NBs + some optic lobe cells	3rd instar larva	F + M	;;*R9D11-Gal4/UAS-hH2B::2xmNG*	3,942	1,829	FACS	10X v.3	NA	18
(Özel *et al.*, 2020)	Optic lobe	15, 30, 40, 50 and 70h APF, Adult (<3 days old)	F (M cells removed)	Canton-S	275,701	903	-	10X v.2	Bioinformatic	>193
(Ravenscroft *et al.*, 2020)	CNS	3rd instar larva	-	Several DGRP strains	5,056	-	Washed	10x v.2	Bioinformatic	39

Table 2: bulk RNA- and ATAC-seq datasets produced in the visual system. APF = after puparium formation.

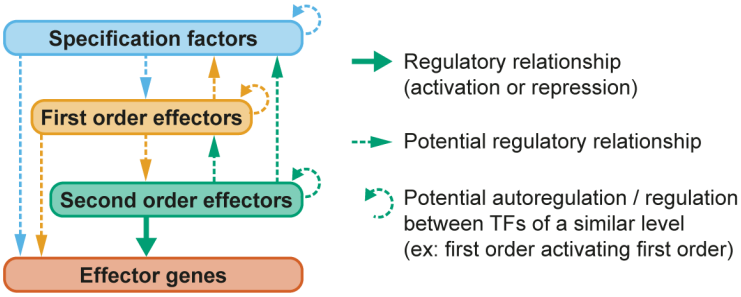
Reference	Material sequenced	Stages sequenced	Sex	Data produced
(Tan *et al.*, 2015)	Isolated populations	40h APF	-	FACSed R7, R8, and L1– L5 neurons
(Zhang *et al.*, 2016)	Isolated populations	24, 35, 40, 45, 53, 65 and 96h APF	?	Sequencing of mRNA associated to ribosomes in photoreceptors.. RNA-seq of FACS photoreceptors at P40 and P53
(Davie *et al.*, 2018)	Brain	Adult	F+M	ATAC-seq, RNA-seq of whole tissue and RNA-seq of isolated nuclei
(Konstantinides *et al.*, 2018)	Isolated populations	Adult (3-5 days old)	F	RNA-seq of 17 FACSed visual system neuronal populations
(Bravo González-Blas *et al.*, 2020)	eye-antennal discs	3rd instar larva	-	Bulk ATAC-seq of sorted *sens-F2B-GFP* cells, of 29 DGRP lines, and of 14 lines overexpressing a TF.
(Davis *et al.*, 2020)	Isolated populations	Adult	F+M	RNA-seq on nuclei of cell populations isolated by INTACT and TAPIN. Visual system populations sequenced: 53 neurons and glia, 6 broader populations (*ChAT*, *Gad1*, *VGlut*, *Kdm2*, *Crz*, and *NPF*), 2 dissected tissues (the lamina and remainder of the optic lobe).
(Jain *et al.*, 2020)	Isolated L1 neurons	40, 60 and 72h APF	?	RNA-seq, ATAC-seq

(Özel *et al.*, 2020)	Isolated populations	Adult (<3 days old)	F	RNA-seq of FACSed Pm2, T4, *elav*+ cells (neurons) and *repo*+ cells (glia)



Box 1: Sequencing and data analysis in the *Drosophila* visual system.

Two main approaches were used to sequence single cell types in the *Drosophila* visual system. The first one was to perform bulk RNA or ATAC sequencing on a tissue, or on a cell population labeled by the expression of a specific driver and isolated, for instance by fluorescence activated cell sorting. The second was to perform single cell RNA or ATAC sequencing on a dissociated tissue, with or without sorting a specific cell population, via droplet sequencing. The obtained single-cell transcriptomes or accessibility profiles were then grouped informatically (i.e. “clustering”) in order to form clusters of profiles similar to each other, which in many cases correspond to cell types. Some studies also corrected for technical variations between data produced in different batches, which is of prime importance to avoid forming clusters based on batches rather than cell types (batch effect correction tools were not always available for earlier studies). The cell types corresponding to the different clusters were then identified by the use of marker genes, or by correlating the average gene expression of the clusters to the one of sorted populations.



Box 2: Establishment and maintenance of neuronal identity

Several models have been proposed to explain the establishment and maintenance of neuronal identity, two of the most prominent ones being the concepts of terminal selector (TS, (Hobert and Kratsios, 2019)) and core regulatory complex (CoRC, (Arendt *et al.*, 2016)). However, these concepts impose strict limitations and, in this review, we chose to use broader terms defined below.

Specification factors: TFs and secreted molecules whose combination is necessary and sufficient to specify the adult morphological and molecular features of a neuronal type. Specification factors can be expressed at birth of the neuron, or before the neuron is produced and act epigenetically. Specification factors can be expressed or not throughout the life of the neuron, and directly or indirectly regulate the expression of effector genes. The specification factors are therefore the ones selecting the identity of the neurons: all downstream genes are effectors of this selection.

First order effectors: TFs directly downstream of the specification factors, and directly or indirectly upstream of effector genes. They include TFs corresponding to the TS and CoRC definitions. However, contrary to TS, they do not have to be expressed throughout development: for instance, some could control axon targeting, therefore the morphology of the neuron, and be expressed only at the beginning of differentiation. Moreover, if their expression is maintained, it

does not necessarily need to be: it can also be maintained by specification factors, other first order effectors, or by feedback from a second order effector. Contrary to core regulatory complexes, they also do not have to physically interact and can target either independent or common sets of genes.

Second order effectors: TFs downstream of the first order effectors, and directly or indirectly upstream of effector genes. They do not have to be expressed throughout development, and if they do it is not necessarily through auto-regulation. They do not have to physically interact and can target either independent or common sets of genes, including specification factors, first order effectors, and other second order effectors (transcriptional cascades). Notably, TFs can assume different roles through the life of a neuron. For instance, the expression of a specification factor could be turned off early in development and turned back on later, making it a second order effector.

Effector genes: non-TF genes that are directly or indirectly downstream of the specification factors (therefore they are largely limited to neuronal-type specific genes). They are not required to be expressed in adult, contrary to what is often implied in the concept of "terminal effectors". Indeed, a CSM necessary to target a given layer but not expressed in adults is also an effector of the identity of the cell, as it affects its final morphology.

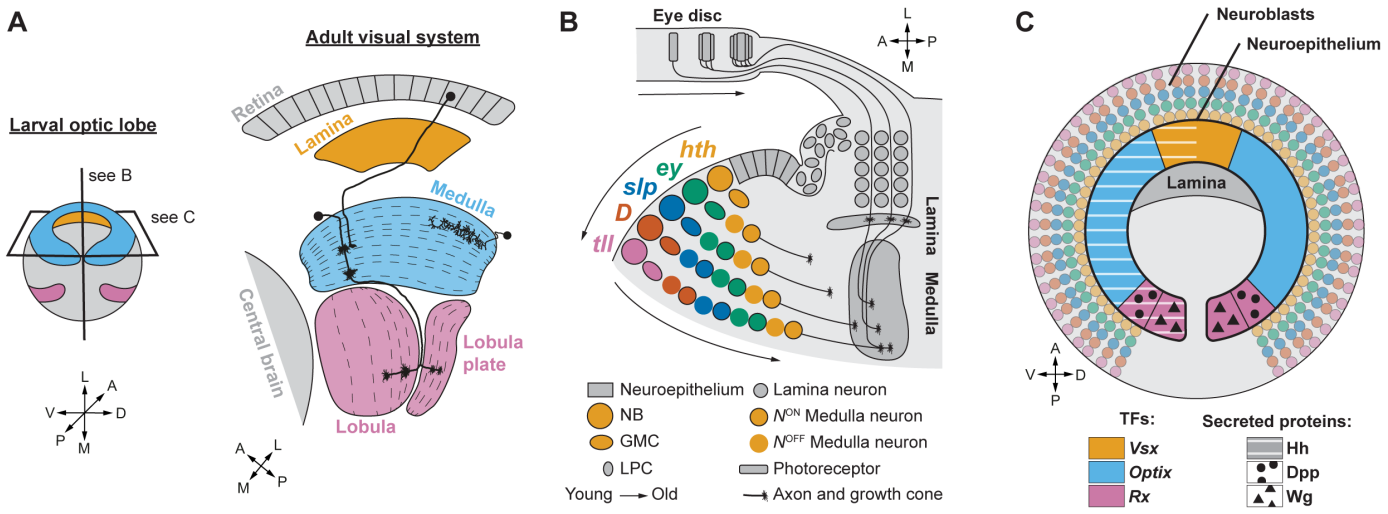


Figure 1 – Anatomy and neuronal type specification mechanisms in the *D. melanogaster* optic lobe

A – Left: schematic of the larval optic lobe. Blue = part of the outer proliferation center in which medulla neurons are produced. Orange = part of the outer proliferation center in which lamina neurons are produced. Pink = inner proliferation center, in which lobula and lobula plate neurons are produced. Right: schematic of the adult visual system. In black are represented two unicolumnar neurons (left), including one photoreceptor, and one multicolumnar neuron (right).

B – Section of the larval optic lobe, ventral view, illustrating the neuronal type specification mechanisms involving a cascading expression of temporal transcription factors and a *Notch*-driven binary neuronal fate decision. *hth* = *homothorax*, *ey* = *eyeless*, *slp* = *sloppy paired, D* = *Dichaete*, *tll* = *tailless,* GMC = ganglion mother cell, LPC = lamina precursor cell, NB = neuroblast, N^{ON} = *Notch* ON, N^{OFF} = *Notch* OFF.

C – Section of the larval optic lobe, lateral view, illustrating the neuronal type specification mechanisms involving spatial domains of the neuroepithelium. Dpp = Decapentaplegic, Hh = Hedgehog, *Rx* = *Retinal Homeobox*, *Vsx* = *Visual system homeobox*, Wg = wingless

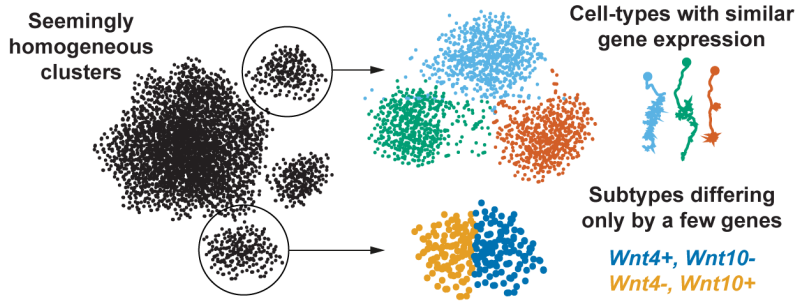


Figure 2 – Potential heterogeneity in seemingly homogeneous clusters

The dots represent single cell transcriptomes, ordered in 2 dimensions by similarity of gene expression. Left: the transcriptomes are organized in 4 clusters of seemingly similar cells. Right: deeper analyses can reveal that these groups comprise either cells from morphologically different neuronal types with closely related transcriptomes, or from morphologically identical neuronal subtypes that differentially express only a few genes.

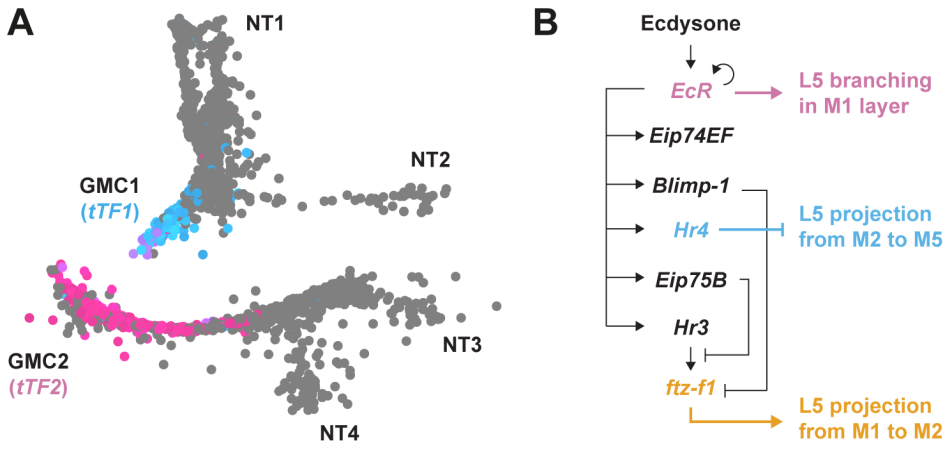


Figure 3 – Neuronal specification

A – UMAP plot of four neuronal types (NT1-4) during larval development. The dots represent single cell transcriptomes, ordered in 2 dimensions by similarity of gene expression. Due to the presence of their parental GMCs in the sequencing, one might be able to assign temporal identity to neuronal types based on their proximity to tTF-expressing GMCs.

B - Members of the ecdysone responsive transcription factor cascade and their involvement in L5 neurons development. Adapted from (Kurmangaliyev *et al.*, 2020).

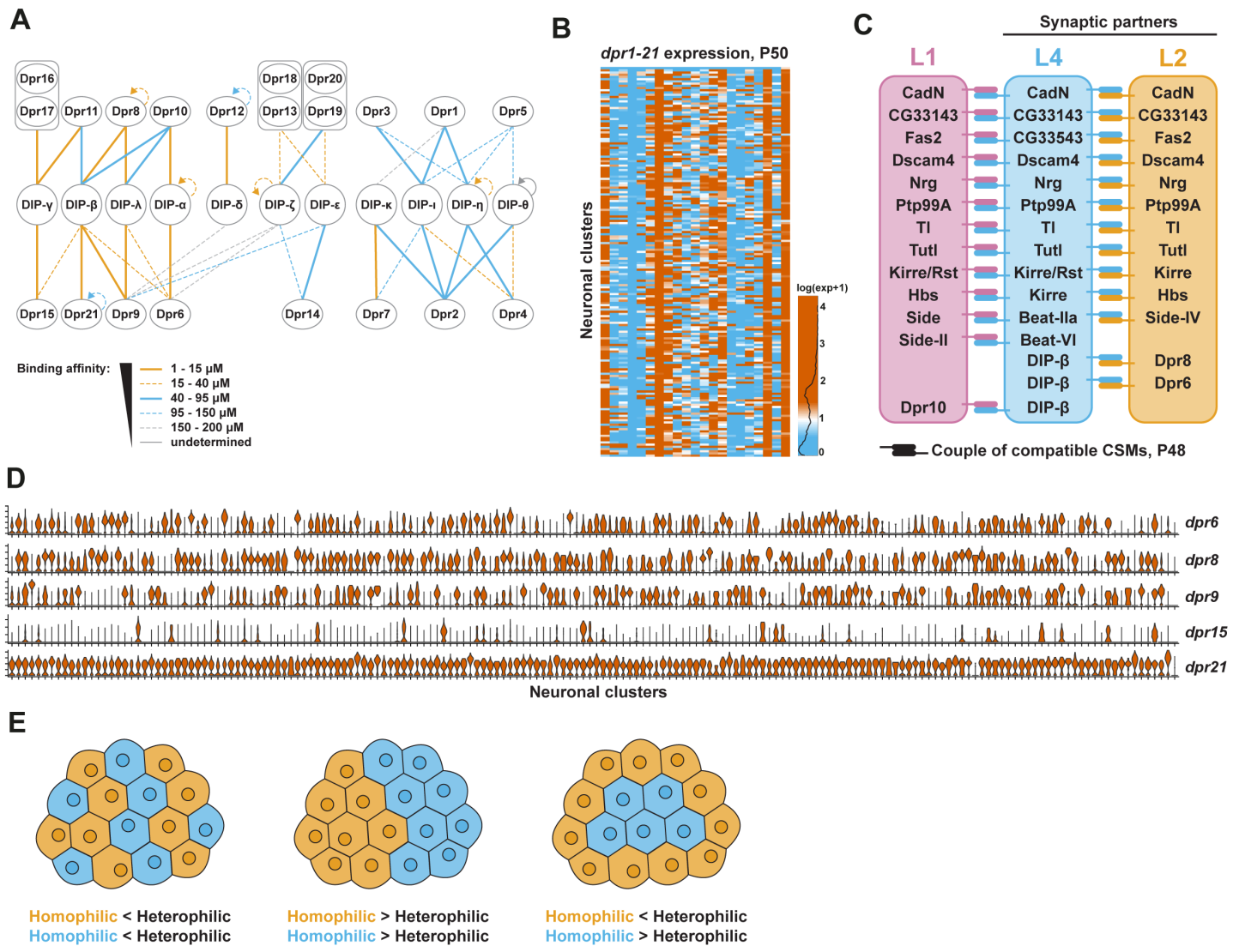


Figure 4 – Molecular mechanisms of neuronal circuit formation

A – Binding affinities of the DIP and Dpr proteins. Adapted from (Cosmanescu *et al.*, 2018).

B – Expression of the 21 *dpr* coding genes at P50. The black line in the scale represents the density of *dpr1-21* expression values. Data from (Özel *et al.*, 2020).

C – Expression of compatible couples of Cell Surface Molecules (CSMs) in lamina neurons L1, L2 and L4 at P48. Adapted from (Kurmangaliyev *et al.*, 2020).

D – Expression of *dpr* genes with high affinity (< 40 µM) to DIP-β, in all neuronal clusters at P50. Data from (Özel *et al.*, 2020).

E – Spatial organization of two populations of cells (orange and blue) as a function of their homophilic and heterophilic membrane adhesivity.

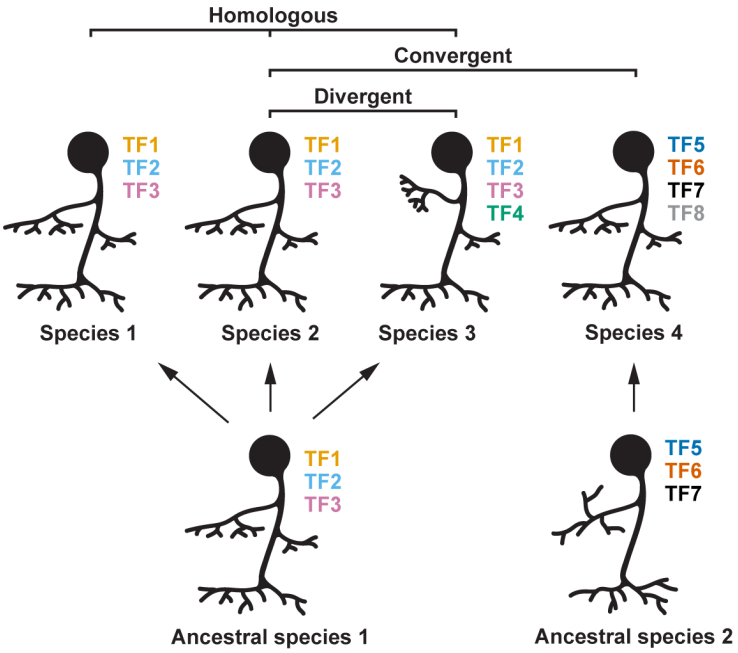


Figure 5 – Neuronal type evolution

Transcription factors can be used to infer evolutionary relationships of homologous cell types. Cell types that share evolutionary history should express a similar set of identity transcription factors. While cases of phenotypic convergence and divergence may blur the comparison of the cell types, comparison of transcription factor signatures should resolve convergent from homologous cell types.