Regulatory logic of neuronal identity specification in *Drosophila*

Rose Coyne[1,*], McKenzie Treese[1,*], Yen-Chung Chen[2], Cathleen Lake[1], Ojong Besong Tabi[1], Raghuvanshi Rajesh[2], Huzaifa Hassan[1], Hua Li[1], Claude Desplan[2,3] and Mehmet Neset Özel[1,3]

1: Stowers Institute for Medical Research, Kansas City, MO 64110, USA

2: Department of Biology, New York University, New York, NY 10003, USA

*: Authors contributed equally. 3: Correspondence: nozel@stowers.org (MNÖ) and cd38@nyu.edu (CD).

Abstract

During neurogenesis, signaling molecules and transcription factors (TFs) pattern neural progenitors across space and time to generate the numerous cell types that constitute neural circuits. In postmitotic neurons, these identities are established and maintained by another class of TFs known as terminal selectors (tsTFs). However, it remains largely unclear how the tsTF combinations are specified in newborn neurons, and how they then coordinate the type-specific differentiation programs of each neuron. To investigate these regulatory mechanisms, we performed simultaneous single-cell RNA and ATAC sequencing experiments on the *Drosophila* optic lobes and identified over 250 distinct cell types at four stages of their development. We characterized the common cis-regulatory features of neuronal enhancers and performed comprehensive inference of gene regulatory networks across cell types and stages. Our results reveal cooperative actions of pan-neuronal and tsTFs on cell type-specific enhancers, and that same effector genes are often regulated by different TF combinations in different cell types. During neurogenesis, patterning TFs are associated with different accessible enhancers before and after cell cycle exit, allowing them to be re-utilized as tsTFs independently from their earlier roles in progenitors. We show that expression of tsTFs *Vsx1/2* in a variety of optic lobe neurons is mediated by lineage-specific enhancers, each patterned by different upstream mechanisms. Therefore, neuronal identity specification is a multi-step regulatory program wherein the same TFs can enact distinct regulatory codes at different steps and across cell types.

Introduction

Neurons are by far the most heterogenous cell type in animals. Their vast diversity in connectivity and function directly underlies the intricacies of sensory processing and the precision of motor functions. While the wide morphological diversity of neurons have been appreciated since the early days of Cajal[1], rapid popularization of single-cell genomics (in particular transcriptomics[2,3]) technologies over the last decade has delivered a renewed interest and expanded understanding of neuronal diversity across species. The most recent single-cell RNA sequencing (scRNA-seq) studies of the mouse brain identified over 5000 molecularly distinct neurons[4,5]. Different combinations of transcription factors (TFs) expressed in these

various neuronal types are thought to encode their distinct identities, instructing cell type-specific morphological[6-9], electrophysiological[10] and synaptic[11-14] features. But little is known about the gene regulatory mechanisms that link differential TF expression to specific neuronal features. In most systems, it is also unclear how the genomes instruct the emergence of such a wide variety of cell types from a limited pool of neural stem cells.

Drosophila melanogaster has long been a premier model system for neurobiology. The full connectome of its adult brain, at electron microscopy (EM) resolution, has recently become available[15]. The visual system in particular offers distinct advantages to study neuronal diversity because its development is thought to be genetically hardwired[16,17], *i.e.* does not require external stimuli. And despite constituting two thirds of the fly brain, due to their repetitive structure characteristic of all visual systems[18], the optic lobes contain only a few hundred cell types[19,20]. This is in contrast to the central brain where thousands of different cell types have been identified, many of which exist as only one neuron per hemisphere[15,21]. Each optic lobe receives direct input from the photoreceptor neurons of the retina and processes visual information in its four neuropils: lamina, medulla, lobula, and lobula plate[22], which have a layered columnar structure analogous to the vertebrate retina and cortex[18]. We previously published a large scRNA-seq atlas of the *Drosophila* optic lobe, resolving ~200 distinct cell-types across 6 stages[23]. This characterization revealed remarkable consistency between the molecular and classical[22] (morphologically defined) cell types. Although it has since been reported that a few of our transcriptomic clusters contained multiple morphologically distinct neurons[24,25], these were all due to limited coverage of rarer cell types.

The optic lobe neurons are generated during late larval and early pupal stages from two main neurogenic regions: the inner and outer proliferation centers (IPC and OPC). Neuroepithelial (NE) progenitors in both regions produce neuroblasts (NB), which go through a series of asymmetric divisions generating intermediate progenitors known as ganglion mother cells (GMC). GMCs then divide just once to produce two neurons (or glia). The IPC primarily produces neurons of the lobula plate (C/T neurons), while the OPC is more complex. Lamina neurons are generated from the lateral OPC, while the medial OPC primarily produces medulla neurons. At least 3 distinct axes of patterning mechanisms contributing to neuronal diversity exist in the OPC: spatial patterning of NE progenitors by the TFs Vsx1, Optix, Rx and the signaling molecules Hedgehog, Dpp and Wingless[26,27], a series of at least 12 temporal TFs (tTFs) sequentially expressed in NBs[28], and Notch signaling between sister neurons after the asymmetric terminal division of GMCs[29]. However, spatial and temporal TF expression is not commonly maintained in postmitotic neurons[28].

Unique combinations of terminal selectors (tsTFs) that are stably expressed in each neuron throughout its lifetime, have instead been proposed to implement and maintain neuronal type identity[30-36]. We recently deciphered the codes of putative tsTFs in all optic lobe neurons, and genetically modified these in select neurons to induce complete and predictable changes of identity between different types[37]. In the lamina, it has been shown that the tsTF Erm, despite its continuous expression in L3 neurons, contributes to temporally-dynamic gene regulation by co-operating with TFs under control of the steroid hormone ecdysone[38]. Therefore, tsTFs link the early cell fate choices made in neural progenitors to the acquisition of specific morphology, connectivity and physiology of each neuron. Using correlative models, we have additionally shown that the combination of spatial domain and temporal window of origin, together with the Notch status of each neuron, can fully account for the diversity of tsTF expression in the OPC neurons[39]. But it remains unclear *how* these different patterning mechanisms, only active within progenitors and newborn neurons, specify the expression of unique combinations of tsTFs in different neurons that will be sustained throughout their lives. Therefore, a comprehensive description of gene regulatory networks both upstream and downstream of tsTFs is essential for understanding the developmental genetic mechanisms that underlie neuronal diversity.

To initiate and accelerate that description in the *Drosophila* visual system, here we present a large single-cell multiomics dataset (simultaneous RNA+ATAC-seq) of the optic lobes collected at four different stages from neurogenesis to adulthood. With ~250 clusters consistently resolved across these stages, the cell-type resolution of this new atlas exceeds that of our scRNA-seq dataset[23]. Taking advantage of the multiome data, we comprehensively characterized the cis-regulatory features of neuronal enhancers with various developmental dynamics, as well as the key TF networks across all optic lobe cell-types and stages. Our results indicate a particularly dramatic shift in chromatin landscape upon the terminal division of GMCs, during which many TFs are re-utilized as terminal selectors independently from their earlier roles in development. As a demonstration of the depth of analyses that are enabled by this new resource, we genetically dissect the specification of tsTFs Vsx1/2 in different OPC lineages by different patterning mechanisms acting through distinct cis-regulatory elements.

Results

A multiomic single-cell sequencing atlas of the developing *Drosophila* visual system

We performed single-cell RNA and ATAC sequencing on nuclei isolated from *Drosophila* optic lobes at multiple stages of their development: P0 (neurogenesis), P24 (axon guidance), P48 (synaptogenesis) and Adult (functional circuit), using *Chromium Single Cell Multiome* kits (*10x Genomics*). After doublet removal and other quality controls (*Methods*), we retained a total

of 232,251 cell barcodes with high-quality transcriptomes and epigenomes (chromatin accessibility) jointly available for further analyses. Key data metrics per stage are summarized in **Table 1**; most notably, the ATAC modality of this new atlas provides a 5-10x improvement on the sequencing depth (unique fragments per cell) of the previously published scATAC-seq atlas of the fly brain[40].

Stage	# Libs	# Cells	nUMI (RNA)	nGene (RNA)	nFragments (ATAC)
Adult	4	47,835	1,255	648	5,704
P48	5	63,726	1,554	778	7,576
P24	4	65,572	922	556	9,498
P0	4	54,253	713	457	8,260

Table 1: Key metrics of the multiomics data. #Cells is shown as those barcodes passing doublet removal and all QC metrics. Median values per cell are shown for nUMI, nGene and nFragments.

We classified these single-cells into cell-types using two approaches. First, we took a supervised approach, using a neural network classifier we trained on the marker gene expression profiles of 206 clusters that we had previously resolved using scRNA-seq[23]. In addition, we performed unsupervised clustering using the Weighted Nearest Neighbor (WNN) algorithm implemented in the *Seurat* package[41], taking advantage of both RNA and ATAC modalities. We clustered first the 63,726 cells of the P48 stage, as we and others have previously shown that neurons display the highest diversity in both gene expression[23,42] and chromatin accessibility[40] at this stage. WNN clustering divided 30 clusters of the supervised classification into 2-6 subgroups (**Suppl. Table 1**), while in only 3 cases the new clustering failed to resolve a previously known subdivision, all of which could be attributed to the very small sizes of these clusters. We integrated the results of both approaches into a new reference atlas containing 259 clusters, which we then applied to other stages using a multimodal label transfer algorithm (**Fig. 1**, *Methods*). As we described previously[23,28] and discuss further in later sections, at the P0 stage during active neurogenesis, only the relatively differentiated cells (tips of the trajectories in **Fig. 1d**) could be classified into these clusters. The improved cell-type resolution of this atlas (233 neuronal clusters, compared to 185 in the previous atlas) is largely attributable to higher depth in ATAC modality. However, our data does not support any molecular diversity among the postmitotic neurons that is detectable by chromatin accessibility alone, since all newly identified subclusters had differentially expressed genes among them. Thus, the failure to resolve some cell-types in previous scRNA-seq atlases[23,43] could again be explained by relative subtlety of their differences and/or the low number of cells in these clusters.

In our previous scRNA-seq atlas[23], we had annotated ~70 clusters as specific optic lobe neuronal types. Glial clusters were also annotated in a subsequent study[44]. Molecular markers for many more optic lobe neurons have since been defined (detailed in *Methods* and **Suppl. Table 1**), thus we were able to assign 17 of the 26 non-neuronal clusters (**Suppl. Fig. 1b-c**), and 113 of the 233 neuronal clusters to specific cell types. Importantly, all morphologically-distinct types of neurons (as defined by light microscopy studies[22,45,46]) that we annotated here, without exception, mapped to unique clusters. This highlights the very high resolution of this new atlas. In a few cases, molecularly distinct subtypes have been identified within morphologically homogenous cell types. We had previously described this phenomenon for Tm4, Tm6, Tm9 and TE neurons that differ when present dorsally or ventrally[23]; we now identify such subtypes also for Dm2 neurons (**Suppl. Fig. 8d-e**, discussed later).

Because of this very high granularity of our cell-type classification, most of our clusters do not contain enough cells to reliably determine their ATAC peaks when treated individually. Thus, we generated a hierarchical tree of all postmitotic clusters (those consistently assigned from P24 to Adult) based on their transcriptomic similarity at P24 and grouped them into 37 neuronal and 5 non-neuronal 'metaclusters', each containing 2 to 13 related clusters (**Suppl. Fig. 1a-b**). The P0 dataset was independently divided into 19 metaclusters based on WNN clustering at this stage (**Suppl. Fig. 1d**). We called the ATAC peaks separately on each of these 61 metaclusters using *MACS2*[47], which were then combined into 76,777 (500 bp) 'consensus' peaks (*Methods*) for optic lobe-wide analyses of differential accessibility (below).

Additionally, we trained a new multi-task neural network classifier on the marker gene expression profiles of our clusters at P24, P48 and Adult stages (*Methods*). This machine learning model can be used to automatically assign cell-type identities to any scRNA-seq dataset derived from the optic lobes. We applied it to reclassify the cells in two previously published scRNA-seq atlases of the developing optic lobes[23,43], and both datasets (with newly assigned identities based on this new reference) can be explored using the web app scMarco[48] (https://apps.ycdavidchen.com/scMarco/).

Dynamic gene regulatory landscapes of differentiating neurons

We previously identified terminal selector transcription factors that maintain neuronal identity through continuous expression from neuronal birth to adulthood[37]. However, how these stable factors generate both cell type-specific and dynamic gene expression programs throughout development and adulthood[23] remained unresolved. To understand these regulatory dynamics during neuronal maturation, we examined the differentially expressed genes (DEGs)

and differentially accessible regions (DARs) among 229 neuronal clusters throughout development (excluding the TE neurons that are not present in adults and cluster 141 which is a highly heterogenous mixture of central brain contaminants). Across P24, P48 and Adult stages, we identified a total of 2,256 DEGs (**Suppl. Table 2**), using stringent thresholds ($log_2FC>1$, one vs. all within each stage, see *Methods*). 976 genes were found to be markers at all three stages, and 674 (30% of total) of these were consistent markers of at least one neuron through these stages (**Fig. 2a**). However, on a per-cluster basis, on average only ~15% of marker genes were constant throughout development (**Fig. 2c**), implying that many genes that are continuously expressed in some cell types are dynamically regulated in others (i.e. Constant & Dynamic, **Fig. 2a**), which include most tsTFs. Similarly, we found a total of 53,637 DARs among neurons across these stages (**Suppl. Table 3**), with 29,190 found at all stages and 21,819 (40% of total) being a consistent marker for at least one neuron (**Fig. 2b**). But again, only ~10% of marker peaks for any given cluster were consistent across all 3 stages (**Fig. 2c**). In contrast to the majority of consensus peaks that were DARs among neurons, we found only 1,986 peaks that display relatively uniform accessibility among neurons with low or no accessibility in non-neuronal cells (pan-neuronal, *Methods*) in at least one of the three stages; 1,358 peaks maintained this pattern from P24 to Adult (consistent pan-neuronal).

To further explore developmental enhancer usage, we calculated gene-enhancer links[49] by correlating gene expression and nearby peak accessibility (50kb window, *Methods*) across all clusters at each stage (**Suppl. Table 4**). Of the 976 DEGs that were markers at all three stages, 969 had at least one linked peak at each of the 3 stages. Therefore, we compared the peaks linked to these genes across development. On average (median) for each gene, only 2 out of the top 10 linked peaks (ranked by abs(z-score) and corrected for genes with fewer than 10 total linked peaks) were common across stages (**Fig. 2d**), suggesting highly dynamic enhancer usage. Half (14/28) of the genes with stable enhancer usage (at least 7 of the top 10 linked peaks common across stages) were tsTFs, which constitute <10% of the consistent DEGs, highlighting their unique role in maintenance of cell type identity. This nevertheless represents a minority of all tsTFs, implying that they are also, and often, subject to differential enhancer usage throughout development.

To understand the common and distinct cis-regulatory features found in different classes of neuronal enhancers we describe above, we used the MEME suite[50] algorithm STREME[51] for *de novo* discovery of motifs enriched in consistent pan-neuronal peaks (**Fig. 2e**), consistent neuronal DARs (**Fig. 2f**), and the DARs found in only one stage (**Fig. 2g**). We then compared the top enriched motifs in each peakset to known *Drosophila* TF motifs using Tomtom[52] (**Suppl.**

Table 5). Dinucleotide repeats 'CA' and 'GA' were enriched in all these regions. These were previously described as structural features of all active enhancers[53] and were proposed to function as general facilitators of TF binding[54]. Alternatively, we note that CA repeats are consistent with the known binding preference of the repressive[55] TF Klumpfuss (Klu). *Klu* is expressed in neuroblasts (but not postmitotic neurons, **Suppl. Fig. 2b**) and has been shown to prevent their premature differentiation; its overexpression causes failures in neuronal commitment leading to brain tumors[56]. Similarly, GAGA TFs are known for their pioneering functions, and one member, Clamp, has recently been shown to be important for stem cell proliferation in optic lobes[57]. Both Clamp and the original *Drosophila* GAGA factor (GAF, also known as *Trl*) are transiently upregulated in progenitor (NB and GMC) clusters (**Suppl. Fig. 2c**). Tramtrack (Ttk) motifs (AGGA) were also among the general features of all types of neuronal enhancers. Remarkably, *ttk* has a strong anti-neuronal pattern in the optic lobe, exclusively expressed in non-neuronal clusters throughout development (**Suppl. Figs. 2c,3**). Together, these findings suggest that many neuronal enhancers have Klu binding sites that could prevent their activation before terminal differentiation. They also indicate the presence of sites that could be bound in non-neuronal cell-types by Ttk to repress these enhancers, as well as sites for pioneering factors such as GAF, which could prime these enhancers during or shortly before neuronal commitment. Lastly, both pan-neuronal peaks and neuronal DARs were enriched in motifs likely corresponding to pan-neuronal TFs FoxP (**Suppl. Figs. 2b,3**), Jim and Lola.

Neuronal type-specific DARs—both consistent and stage-specific—shared the common features described above but were also strongly enriched for homeodomain motifs, in contrast to pan-neuronal peaks. These motifs belong to the TF family most highly represented among terminal selectors[37,58], and include variants such as NK-like and POU domain motifs (**Fig. 2f-g**). Stage-specific, but not consistent, DARs were additionally enriched in Blimp-1 motifs. *Blimp-1* encodes an ecdysone-responsive TF, highly expressed in all neurons during early pupal stages, that is turned off after P36 before being upregulated again in adult neurons (albeit not to its earlier levels)[38]. To identify additional regulators of temporally dynamic chromatin accessibility, we investigated the motifs *relatively* enriched (using as background the sequences of consistent DARs) in peaks that were found as cluster markers only at P24 (3,187 peaks), P48 (3,413 peaks) or Adult (2,039 peaks) stages (**Fig. 2h-j**). Interestingly, 'GATC' motif of the pan-neuronal Onecut (**Suppl. Figs. 2b,3**) was the top hit in all 3 stages. The contribution of the CUT family of TFs to regulation of neuron type-specific genes (in addition to pan-neuronal genes) has been established in *C. elegans*[59]. Our findings further suggest that despite its constant expression in neurons, Onecut specializes in dynamically regulated enhancers. Finally, we identified a relative

enrichment of different hormone-receptor (Hr) class of motifs in stage-specific peaks, which further highlights the crucial role of ecdysone signaling in temporally dynamic neuronal gene expression. While the core 'GGTCA' that is enriched at P48 is most consistent with the known Hr3 motif, which is upregulated during mid-pupal stages, the variant observed in the Adult-specific peaks is more similar to the known ftz-f1 motif, another effector of ecdysone expressed at late pupal stages. These findings generalize the previous conclusions made in lamina neurons[38] to the entire optic lobe, *i.e.* despite their pan-neuronal expression, downstream targets of ecdysone-regulated TFs are predominantly cell-type specific.

In summary, our global analyses of DARs reveal the extensive cooperation between various classes of pan-neuronal, terminal selector and ecdysone-responsive TFs for shaping spatiotemporally specific chromatin accessibility in developing neurons.

Computational inference of developmental gene regulatory networks

In order to comprehensively understand neuron type-specific regulation, we performed computational gene regulatory network (GRN) inference using the SCENIC+ framework[60]. To increase our statistical power, in addition to the multiome data from P24, P48 and Adult stages, we also utilized the scRNA-seq datasets of optic lobes previously published by us[23] and others[43] from similar development stages. These 'RNA-only' cells were included alongside the multiome cells specifically for the steps that do not utilize cis-regulatory information, *i.e.* calculation of TF-target gene expression correlations with GRNBoost2, while all other inference steps were run with the multiome data only. Because the target genes regulated by a TF can be highly context-dependent, we applied GRN inference at the metacluster level (**Suppl. Fig. 1a**), rather than the whole optic lobe. However, since some of our metaclusters contain as few as 2 cell-types and thus have limited biological variation for GRN inference, we merged some metaclusters containing related neurons (detailed in *Methods*). We inferred 28 different SCENIC+ networks (23 neuronal) covering all optic lobe cell types, which we provide in **Supplementary Table 6**.

We then used these SCENIC+ results to investigate the context dependency of TF-target specificity. 33 TFs had regulons in at least 10 of the 23 neuronal GRNs; for all target genes of these TFs, across all networks, we determined the number of metaclusters in which a TF was predicted to be regulating the same target gene (**Fig. 3a**). We observe that the majority of predicted targets for all 33 TFs were found in only one or two different GRNs, indicating that the same TFs in different cell types regulate highly divergent sets of target genes. In fact, only 159 predicted TF-target interactions were relatively context-independent, i.e. found in more than 50% of the GRNs which include that TF as a regulator (outliers in **Fig. 3a**). Visualization of this 'core'

network of consistently found predictions (**Suppl. Fig. 4**) shows that ecdysone-responsive TFs Blimp-1, Eip75B, Hr3, Hr4 and ftz-f1 mediate 71% of these interactions. In addition, for 1287 target genes that were found in at least one regulon in at least 10 different GRNs, we determined which TFs were most consistently predicted to regulate each of these genes. Our results indicate that, on average, even the most consistent regulators were linked to a specific gene only in a minority of networks that include this gene (**Fig. 3b**). These findings suggest that the same DEGs are most often regulated by different TFs in different cell types. Similarly, it reveals that the vast majority of highly consistent regulators for any given gene (found in >50% of GRNs including that target gene) were ecdysone-regulated TFs (**Fig. 3b**). We emphasize, however, that such consistently found interactions were rare, and thus cell type-specific regulation appears to be a general feature of all classes of neuronal TFs.

Together, these results indicate that while the pan-neuronal ecdysone-regulated TFs occasionally regulate the same genes across a variety of different neurons in the optic lobe, downstream targets of terminal selector (and other DE) TFs are almost always context-specific, i.e. limited to one (or a few) metaclusters. It should be noted that the way we grouped our cell-types into metaclusters (*i.e.* based on transcriptomic similarity) and inferred these GRNs introduces certain biases that need to be considered while interpreting these results. For example, the TFs or targets whose expression do not significantly vary either among the cell-types (within a metacluster) or across the stages considered are unlikely to be captured. This could explain why the vast majority of predicted 'pan-neuronal' interactions were downstream of the dynamically regulated ecdysone TFs, rather than the continuously expressed pan-neuronal TFs such as Onecut or FoxP.

We had previously determined using scRNA-seq both the DEGs between optic lobe neurons (cluster markers) at 6 stages from P15 to Adult, and the genes that are upregulated in neurons (overall) at specific stages (stage markers)[23]. Gene Ontology (GO) enrichment analysis on these gene sets had revealed that cluster markers were consistently enriched in genes encoding TFs and cell-surface proteins (like signaling receptors and ion channels), which are involved in cell adhesion and neurite development/guidance at earlier stages, shifting to regulation of membrane potential and synaptic transmission after P50. In contrast, the stage markers at P15 were enriched in translation-related terms, followed by cell-surface proteins (which are also often cluster markers) during mid-pupal stages; while the late pupal and adult stage markers were mainly enriched in energy metabolism terms (e.g. oxidative phosphorylation), in addition to ion channels and synaptic transmission. We took advantage of our SCENIC+ networks to understand the main regulators of these different gene expression

programs. Across all the GRNs we inferred, we compiled lists of target genes for the 103 TFs with at least 50 unique targets identified, and performed GO enrichment analyses on these target gene sets for each TF. For a representative set of terms enriched in the cluster and/or stage markers, we then ranked the associated TFs based on the fold-enrichment of their target gene sets for each of these terms (**Suppl. Fig. 5**, *Methods*).

Not surprisingly, the top regulators for all cluster marker terms, including neurotransmitter receptors (**Suppl. Fig. 5a**), genes involved in synaptic target recognition (**Suppl. Fig. 5b**), axon guidance (**Suppl. Fig. 5c**), and regulation of membrane potential (*i.e.* ion channels, **Suppl. Fig. 5d**) were dominated by tsTFs. Interestingly, EcR was also among the top regulators of axon guidance genes; and Hr3 was a top regulator of ion channel genes. These results provide strong support for the idea that most neuronal effector genes are directly regulated by terminal selectors[30]. Reinforcing our conclusions above, they also indicate that the genes whose expression are both specific among cell-types and dynamic across developmental stages are jointly regulated by tsTFs and ecdysone TFs.

Among the other stage markers, translation terms were enriched at early pupal stages, but these are not typically DEGs between neurons. Accordingly, the top regulators of these genes were all ubiquitous or pan-neuronal TFs (**Suppl. Fig. 5e**), including BtbVII, Crol and FoxP. Similarly, oxidative phosphorylation-related genes (in adult stage markers) were extremely enriched (12-fold, **Suppl. Fig. 5f**) among the targets of *CG4404* (an unstudied MADF-BESS domain TF), which is ubiquitously expressed during later pupal stages and in adult optic lobes. Also ranking highly, both CrebA and Hr38 have previously been implicated in transcriptional responses to neuronal activity[61,62]. These findings provide a potential mechanism linking the commencement of electrical activity in the optic lobe[63] to the upregulation of energy metabolism genes in maturing neurons.

Patterning regulators play mutually exclusive roles before and after cell cycle exit

While the multiome data from P24 to Adult stages enable us to explore the GRNs of postmitotic neurons during their terminal differentiation and circuit formation, neurogenesis is ongoing until P15[23]. Since the same cell-types are produced continuously in the optic lobe from late larval to early pupal stages[64], a single snapshot at P0 captures the entire trajectory of neuronal identity specification[28], including neuroepithelia (NE), neuroblasts (NB) and ganglion mother cells (GMC), as well as young neurons (**Fig. 1d**). In addition, our data captures neurogenesis in both the retina and lamina; thus we separated these lineages into different UMAPs for additional analysis: retina (**Fig. 4a**), lamina (or lateral OPC, **Fig. 4b**), IPC (**Fig. 4c**),

and medial OPC (**Fig. 4d-e**). Molecular mechanisms of cell-type specification are relatively well understood in the retina and lamina; and we will address IPC neurogenesis in a separate manuscript, thereby we are focusing on the medial OPC lineages in this study.

Of the three patterning mechanisms we introduced earlier, Notch status has the most significant effect on terminal identity of OPC neurons (such that sister neurons are often more distinct from each other than other neurons with the same Notch status but different spatial or temporal origin[28]). Thus, we separated the OPC progenies at P0 into Notch^{ON} (**Fig. 4d**) and Notch^{OFF} (**Fig. 4e**) hemilineages and analyzed them separately. Immediate effectors of Notch signaling (*Hey* and other *E(spl)* genes) are only transiently expressed in postmitotic N^{ON} cells (**Suppl. Fig. 2d-e**). Expression of the tsTF Apterous (ap) is regulated by Notch[29] and has been treated as a universal marker of N^{ON} OPC neurons[39] (**Suppl. Fig. 2d**). It is also the main regulator of their cholinergic identity[65]; although this is not generalizable to other brain regions, including the IPC. We identified an exception to this pattern even in the OPC: MeTu neurons[65], which were previously annotated to an incorrect cluster[48] (see **Suppl. Table 1**, clusters 125 and 166), group with the N^{ON} neurons from late temporal windows, but do not express *ap* (**Fig. 4d** and **Suppl. Fig. 2d**, red ellipse). Since the identities of this group of neurons are distinguishable only at the trajectory tips (and not immediately after the GMC division when *ap* turns on in other N^{ON} neurons), we were unable to discern whether MeTu neurons never activate *ap* expression or if they do not maintain it. Interestingly, these neurons are still cholinergic despite the lack of *ap*.

The main branches of these UMAP trajectories (**Fig. 4d-e**) roughly represent their birth order, (in clockwise direction), and allows their assignment to different temporal windows based on tTF expression patterns at their bases (**Suppl. Fig. 6**), with a higher resolution than we had previously achieved with scRNA-seq[28]. Importantly, at least 6 tTFs (*hth*, *erm*, *ey*, *hbn*, *scro* and *D*) are also expressed in postmitotic neurons as tsTFs. The relative independence of these genes' expression as tTFs vs. tsTFs has already been recognized[28]. Our trajectories show that indeed none of these TFs are maintained in all the neurons that originate from the corresponding temporal windows (**Suppl. Fig. 6**, arrows), and all of them are also independently activated in other lineages (*i.e.* not from their own temporal window, **Suppl. Fig. 6**, arrowheads). We thus suspected that even in cases of apparent "maintenance", regulation of these genes before and after cell cycle exit are likely to be independent. To test this hypothesis, we examined the peak-gene linkages for these 6 TFs, calculated separately using the progenitor cells (NB+GMC) and the postmitotic neurons of OPC at P0 (**Suppl. Table 2**). None of the top 5 linked peaks (by abs(z-

score)) were common between the progenitors and neurons for any of these genes except for *scro*, for which 2/5 of the top linked peaks were common.

Altogether, our results demonstrate that the chromatin landscape undergoes a dramatic transformation during the first few hours of a neuron's life after a GMC's terminal division. Even the genes that are continuously expressed from progenitors appear to switch their enhancers upon neuronal commitment, allowing them to be re-utilized independently from their regulation in progenitors. We posit that this brief window during which Notch signaling is active and tTFs inherited from NBs are still available at the protein level (**Suppl. Fig. 9c**) is a particularly amenable regulatory state that enables each neuron to decode its developmental history into a stably maintained terminal selector code. Our dataset is a unique resource to systematically investigate this process and understand how the joint actions of various patterning mechanisms instruct the generation of over 100 different neuronal cell types from neural stem cells. As proof of concept, in the last section we focus on the specification, and the underlying cis-regulatory mechanisms, of the tsTFs Vsx1/2.

Modular regulation of neuronal Vsx1/2 expression by distinct patterning mechanisms

Visual system homeobox (Vsx) 1 and 2 are paralogous genes located within ~30kb of each other (**Fig. 6a**) and are always co-expressed in a subset of optic lobe neurons[23]. Vsx1 (but not Vsx2) is also an established spatial patterning factor, marking specifically the central region of the OPC NE[27]. Accordingly, the vast majority of neurons that express *Vsx1/2* are in the Vsx domain at stage P5 during neurogenesis (**Fig. 5a**); and lineage tracing experiments[39] have shown that neuronal types that originate exclusively from this domain almost invariably express *Vsx1/2* as tsTFs. However, postmitotic *Vsx1/2* expression is not restricted to this domain; and there are several neuronal types that are produced in other domains (Optix and/or Dpp) in addition to the Vsx domain, and also express *Vsx1/2*. For example, Dm2 neurons, which co-express Vsx1/2 and Distal-less (Dll), are produced from all three of these domains[39] and can be found across the entire OPC at P5 (**Fig. 5a**, arrows). We previously showed that postmitotic expression of either *Vsx1* or *Vsx2* can convert Mi15 neurons into Dm2 fate[37]. We now also show that the Dm2:Mi15 ratio was drastically reduced in *Vsx1/2* double mutant MARCM[66] clones, but unaffected in *Vsx1* mutant clones (**Fig. 5b**). Thus, Vsx1/2 function redundantly as tsTFs in Dm2 to distinguish it from Mi15.

The findings above suggest that Dm2 neurons turn on *Vsx1/2* expression independently from the Vsx1-mediated spatial patterning. Interestingly, the three closest cell-types to Dm2 do not express *Vsx1/2* (**Suppl. Fig. 8a**): Mi15 (mentioned above) and the highly similar clusters 163

and 164 (that express *hbn*, unlike Dm2 or Mi15), which together form metacluster M24 (**Suppl. Fig. 1a**). To study the development of these neurons, we found an ATAC peak specifically accessible in only M24 neurons (**Suppl. Fig. 7**) and produced an *M24-Gal4* line based on this enhancer. We generated sparse FLP-out clones using this new driver and identified, in addition to Dm2 and Mi15, two different types of Hbn$^+$ neurons[67] corresponding to clusters 163 and 164. These neurons can be distinguished from each other by Bifid (bi) expression (**Suppl. Fig. 8c**), suggesting that they are spatially patterned by Dpp signaling[26]. Due to their molecular similarity, we hereby name them Mi21a (bi-) and Mi21b (bi+). This driver labeled all M24 neurons throughout development, except at L3 when only Dm2 and Mi15 (but not Mi21) were labeled (**Suppl. Fig. 8b**). We also identified two subtypes of Dm2 neurons that were only confidently resolvable at P48 (**Fig. 1b**). We performed HCR *in situ* hybridization against two DEGs, *bma* and *glob1*, and showed that these mRNAs are present in mutually exclusive, but morphologically indistinguishable, subgroups of Dm2 (**Suppl. Fig. 8d-e**). While we could not distinguish these during neurogenesis, we observed a dorsoventral bias in the location of these neurons at P48: *glob1*$^+$ Dm2a tended to be in the ventral medulla while the *bma*$^+$ Dm2b were mostly in the dorsal side. Therefore, Dm2 subtypes also appear to be specified by spatial patterning (*i.e.* Hedgehog signaling[27]).

Combining these subtypes (**Fig. 5e**), Dm2 neurons are produced domain-agnostically, while Mi15 and Mi21 originate from mutually exclusive Optix domains (Mi15: dorsal, Mi21: ventral)[39]. Thus, they may be produced during the same temporal window that must be distinct from (but likely consecutive to) the Dm2 temporal window. This would imply that *Vsx1/2* expression in Dm2 is patterned by tTFs. To understand this regulation, we sought to resolve the birth order of these neurons by analyzing sparse NB clones in the dorsal Optix domain (**Fig. 5c**, **Suppl. Fig. 9a**). Dll is a tsTF expressed specifically in N^{OFF} OPC neurons produced near the end of the temporal cascade[28] and in all spatial domains (*i.e.* a concentric gene[68]), including all M24 neurons (**Fig. 6c, Suppl. Fig. 8a**). Among the cell types that are produced in dorsal Optix[39] and express *Dll*, we observed that neurons of the unannotated cluster 212 that co-express Vsx1 and Svp are produced first (**Fig. 5c,** green arrow), followed by Dm2 that express Vsx1 (but not Svp, **Fig. 5c**, purple arrow) and then Mi15 (Dll only, **Fig. 5c**, blue arrows). The glia that are known to be produced during the final temporal window of OPC neuroblasts[29] (**Fig. 5c**, arrowhead) also express Dll, but the protein is not detectable before their migration towards the neuropil shortly after their birth (**Suppl. Fig. 9b**).

The results above indicate that Dm2 and Mi15 are the two final neurons produced by the OPC NBs before they switch to glia production. Consistent with these results, all later born

Dll^{+}/Svp^{-} neurons (*i.e.* M24) transiently maintain the late tTF Tailless (Tll) protein inherited from NBs (**Suppl. Fig. 9c**), similar to the glia that are also produced in this window[29]. In *tll* mutant MARCM clones, the only neurons that still expressed Dll also expressed Svp (**Fig. 5d**), indicating the loss of all M24 neurons. The glia (Repo expression) was similarly lost in these mutant clones (**Suppl. Fig. 9d**). Together, these results imply that the Tll window in OPC NBs has at least 3 sub-windows for sequential production of first Dm2, then Mi15 (and presumably Mi21) and then finally the glia (**Fig. 5e**). It was previously shown that the commencement of *gcm* expression at the end of the Tll window commits the NBs to glia production[69] (and termination of the temporal cascade). The only other known tTFs whose expression partially overlap with Tll in NBs are BarH1/2[28]. While we could hypothesize that expression of BarH1/2 during the first part of Tll window directly or indirectly activates *Vsx1/2* expression in Dm2, *BarH1/2* are also required for initiation of Tll expression in NBs[28,69]. As their knock-down would presumably lead to the loss of all M24 neurons (like we have observed in *tll* mutants), it is not straightforward to genetically test the role of *BarH1/2* in differential regulation of *Vsx1/2* among these neurons. We instead investigated the cis-regulatory mechanisms of neuronal *Vsx1/2* expression using our multiome data.

We analyzed the *Vsx* locus expression and accessibility at P0, specifically the M24 branch of the OPC N^{OFF} hemilineage (**Fig. 4e**, green circle, and **Fig. 6a-c**). We identified two regions in the *Vsx1/2* locus that are highly accessible in Dm2 (from P0 to adult), with low or no accessibility in Mi15/21, or in any of the Vsx^{+} neurons that originate exclusively from the Vsx domain, which appeared to be using different enhancers (**Fig. 6a**). We generated *Gal4* reporters for both Dm2 enhancers and analyzed their expression pattern throughout optic lobe development (**Figs. 6d-f, Suppl. Fig. 10**). The first (Dm2-1) enhancer, within a *Vsx2* intron, stochastically labeled a minority of Dm2 neurons (on average 17% in adults) from P24 to Adult. Many other cells were labeled, including those without Vsx expression, but the majority of the labeled cells in the medulla expressed Dll (**Fig. 6d,g**), indicating the enhancer is primarily active in N^{OFF} neurons from the last few temporal windows. At the L3 stage, we could not find any OPC neurons labeled by this enhancer using a membrane-tagged (CD4) GFP reporter (**Suppl. Fig. 10a**). Using a more sensitive nuclear GFP reporter, we could observe very few OPC neurons labeled at L3, including 1-3 Dm2 per brain (**Fig. 6d**). The second (Dm2-2) enhancer, ~4kb upstream of *Vsx1*, labeled Dm2 neurons much more consistently (on average 48% in adults) throughout development (**Fig. 6e-g and Suppl. Fig. 10b**). Many other Vsx1+ neurons that did not express Dll (*i.e.* not Dm2) were also labeled, as well as Vsx- neurons; but like the Dm2-1 enhancer, most labeled neurons expressed Dll (**Fig. 6g**). Unlike the Dm-1 enhancer, the activity of Dm2-2 enhancer could be

observed in Dm2 neurons at L3 stage even with the membrane GFP (**Suppl. Fig. 10b**). Labeling with nuclear GFP further showed that the enhancer labeled OPC neurons in all spatial domains (concentric), with some bias towards the dorsal Optix domain (**Fig. 6e**). In summary, these results show that Vsx enhancers Dm2-1 and Dm2-2 are both active in the Dm2 neurons, and their overall expression profiles strongly suggest that they are not regulated by a spatial factor (*e.g.* not by Vsx1 itself) but are likely regulated by temporally patterned factors.

Since Dm2-2 enhancer activity was higher than Dm2-1 and appeared to commence earlier, it is more likely to control the initiation of *Vsx1/2* expression in Dm2, and potentially in other neurons that express *Vsx1/2* despite being produced outside the Vsx domain. Thus, we investigated this enhancer more generally in the OPC (**Suppl. Fig**. **11**). It was accessible in at least four other neurons (TmY4, Dm15, and unannotated clusters 166 and 212) that express *Vsx1/2* despite also being produced in the Optix spatial domain[39], and was not accessible in any of the neurons that are produced exclusively in the Vsx domain (*e.g.* Pm4 and Lpi3-4). Interestingly, while this ~3kb region was broadly open in Dm2 at P0, the accessibility became restricted to a ~1kb peak from P24 onwards. Similarly, in TmY4 (N^{ON}, Slp/D temporal window), it became restricted to a peak distinct and upstream of the Dm2 peak; Dm15 and cluster 166 (N^{OFF}, Slp/D window) shared a different peak downstream of the Dm2 peak. Cluster 212 (N^{OFF}, D/BarH window), which also expresses *Dll*, appeared to use the same subregion as Dm2. This observation, i.e. the cluster 212, which originates from BarH windows but before the overlap with Tll, uses the same enhancer of *Vsx1* as Dm2 is consistent with the hypothesis that BarH1/2 regulate *Vsx1/2* expression in both neurons. In addition, the 1kb subregion of the Dm2-2 enhancer that we identified for Dm2/212 neurons contains several potential homeodomain binding sites, at least one of which is most similar to the known BarH1 motif (**Suppl. Table 7**).

Together, these findings suggest that different enhancers have evolved, in curiously close proximity, to activate *Vsx1/2* expression in neurons from different temporal windows (and Notch status), and with complete independence from Vsx1 spatial patterning. There are several other neurons that express *Vsx1/2* despite originating from other domains yet do not appear to use this enhancer cluster (e.g. Tm5c and Tm25, **Suppl. Fig**. **11**). Thus, there are likely additional enhancers of *Vsx1/2* that are temporally patterned, which we did not investigate further.

Discussion

By sequencing nearly a quarter million single-cells in two modalities, we resolved in this study over 250 different cell types (233 neurons) of the *Drosophila* optic lobes consistently across their development. This represents a higher cell-type resolution than all previous single-

cell atlases of the optic lobes (or containing optic lobes)[23,28,40,43,70,71], and is one of the most granular molecular descriptions of any complex nervous system to date. While this is indeed a significant improvement, we note that only 30 of the 185 neuronal clusters of our previous scRNA-seq atlas[23] were found to contain multiple cell types in this study. These were generally due to the rarity of the cell-types involved or the subtle differences between very similar neurons (like the Dm2 or Mi21 subtypes). Thus, scRNA-seq characterizations of the optic lobes have already been near saturation. Our current dataset extends this high resolution to the ATAC modality and should contain independent clusters for all cell-types present in at least ~20 copies per optic lobe.

This limitation explains the discrepancy between our molecular characterizations and the 732 cell-types that were recently defined in the optic lobes through EM connectomics[19]. We note that 552 of those cell-types were found in 20 or fewer copies, but the top 180 most abundant cell-types made up 97% of all optic lobe neurons. These difficulties in resolving rarer cell-types mirror the continued challenges in achieving a similar cell-type resolution with scRNA-seq datasets originating from the central brain[72,73], since central brain neuronal types (in contrast to optic lobe) are rarely present in more than 10 copies per brain. We predict that at least 2 million single-cells (per stage) will have to be sequenced to achieve a similar cell-type resolution in that region. Another potential source of discrepancy between molecular and connectomic characterizations of neuronal diversity is “intra-cell-type” heterogeneities sometimes observed among neurons of the same type[74]. Some of these may represent stochastic variations that are not reflected in the transcriptome; or they are due to subtle molecular differences that could only be resolved upon sequencing hundreds of cells of the same type[75]. It is possible that some such differences are not genetically specified during neurogenesis but rather arise from local signals that neurons experience during their postmitotic differentiation, like the dorsoventral Wnt4/10 gradients present in optic lobe neuropils[23,76].

Our analysis has identified tens of thousands of putative neuronal enhancers, most of which are differentially accessible between neuronal types, but nevertheless share common features. First, the general presence of pan-neuronal TF binding sites (even in neuron type-specific enhancers) and the binding sites for Ttk (expressed in non-neuronal cells) could prevent these enhancers from being active in rest of the body. This may enable neuronal enhancers to evolve more freely without disrupting the expression of their associated genes in other tissues. As our dataset contains a limited diversity of non-neuronal cell-types (i.e. the optic lobe glia and support cells of the retina), a more general analysis of chromatin accessibility spanning other organ systems will be necessary to validate this conclusion. Second, the prevalence of both Klu

and GAGA motifs in neuronal enhancers could imply that neuronal progenitors fulfill the dual function of pioneering these regions for later use, while simultaneously keeping them from being activated prematurely. This is consistent with the important roles Klu and GAGA TFs are known to play in neurogenesis[56,57]. Finally, the specific enrichment of homeodomain motifs in neuronal DARs validate our previous conclusions that tsTFs are primarily responsible for neuron type-specific gene expression in the optic lobe[37]. Importantly, we show that tsTFs orchestrate a very dynamic gene regulatory program as neurons progress through multiple steps of differentiation. Despite their continuous expression, many (if not most) of their targets are stage-specific, which is, to a large extent, a result of their co-operation with ecdysone-regulated TFs (consistent with previous reports[38]). Therefore, combinatorial action of at least two orthogonal regulatory mechanisms, one stable across time but variable among cell types (tsTFs) and the other uniform among cell types but variable through time (hormones), is required for the unique and dynamic gene expression programs implemented in each neuron throughout its differentiation.

Our computationally inferred GRN models also highlight the importance of tsTFs in regulating neuronal DEGs and provide strong support for three key features of terminal selectors that were originally proposed in *C. elegans*[31]. First, tsTFs can only act in combinations. We found little overlap among the genes regulated by the same tsTFs across different metaclusters. Thus, target specificity is not conferred by individual tsTFs but is strictly dependent on what other TFs are expressed in each cell-type. In fact, these results confirm that combinatorial action is not just a property of tsTFs, but all regulators, including (to a large extent) pan-neuronal TFs. Second, tsTFs directly regulate neuronal effector genes. All interactions inferred by SCENIC+[60] are supported by cis-regulatory evidence, *i.e.* enrichment of the matching TF motifs in a peak that is linked to a given target gene by accessibility-expression correlation. Nevertheless, we also find commonly that tsTFs regulate other TFs, including other tsTFs, implying that at least a fraction (and potentially a majority) of a tsTF's downstream effectors must be regulated indirectly. Third, tsTFs do not specialize in distinct phenotypic features of a neuron, which is likely a result of their combinatorial nature. This is supported by the results of our GO enrichment analysis on the predicted targets of each TF (**Suppl. Fig. 5**), where targets of most tsTFs were enriched for all functional classes of cluster marker genes. But this is not to imply that all effector genes in a given neuron are jointly regulated by all tsTFs in that neuron, as some early work on tsTFs had suggested[77]. Because of this notion, 6 TFs, most of which are continuously expressed in fly leg motor neurons and could induce predictable morphological conversions between different types were previously disqualified from being tsTFs because they do not affect the glutamatergic identity of these neurons (which were notably all glutamatergic to begin with)[7]. As we previously

proposed[37,39], our results are more consistent with the effector genes being controlled by different permutations of available tsTFs in each neuron, without any phenotypic specialization, but any given effector will not likely be regulated by all tsTFs in that neuron.

It is worth discussing that we previously defined the putative tsTFs in optic lobe neurons based on their continuous expression in specific cell-types[37]. But our analysis of developmental dynamics shows that many of these TFs are also transiently expressed in other cell types, where they would not be considered tsTFs by this definition. This does not imply that these TFs cannot have important functions in neurons they are not continuously expressed; and we identified regulons for some TFs in metaclusters where they would not be considered tsTFs (but they may be so in others). There is likely no fundamental difference in these TFs' mode of regulation whether or not they are expressed continuously. But they may appear to have more phenotypic specialization if they are only expressed at specific stages in a given cell-type, as their effects will be accordingly limited. We argue that this definition of "terminal selector" has utility, as we previously showed that neuronal types can be converted to each other simply by modifying these tsTF codes (disregarding any other DE TFs that are not continuously expressed). This implies that these combinations contain within them the 'minimal code' of TFs required to specify each neuronal type.

The *Drosophila* optic lobe is a uniquely powerful model system to understand how these tsTF combinations are specified. In addition to the putative tsTF codes in every neuron[37], the spatial, temporal and Notch origin of most OPC neurons have been mapped out[28,39], as well as the correlations between these origins and the tsTF expression patterns[39]. Together, our findings indicate that tsTFs function as ultimate integrators of developmental history in newborn neurons, which appear to be in a unique regulatory state that enables them to decode their lineage history into a stable terminal selector combination, which then remains "locked in" for the subsequent stages of differentiation. Our results here further indicate that neural progenitors and postmitotic neurons operate on starkly different GRNs. While the two genetic programs have some overlapping TFs involved in both, these are almost always regulated through different enhancers in each context, supporting their independence.

In multiple studies[28,29,68,78], including this one, the OPC tTFs have been genetically linked to the expression of specific tsTFs in neurons originating from their respective temporal windows. It is, however, not straightforward to decode how the specification program (patterning regulators) instructs the initiation and maintenance program (selectors) in each neuron. First, we frequently observe in our P0 trajectories widespread activation of some tsTFs in certain lineages,

which eventually become more restricted. This would suggest that the final tsTF combination of each neuron is shaped substantially by the cross-regulatory relationships of tsTFs, in addition to the upstream patterning regulators. Second, confirming our recent predictions from correlative models[39], our findings here make it likely that many (if not most) tsTFs have several different combinations of specification factors that could activate (or repress) them in different OPC lineages, through different cis-regulatory elements.

As we have demonstrated with *Vsx1/2*, our multiome data enables us to identify these enhancers accurately. The next challenge facing our field is to identify the key TFs acting on these tsTF enhancers and be able to do so at scale. As in the case of BarH1/2, genetic experiments involving tTFs are often hard to interpret due to their frequent cross-regulation and the resulting indirect effects on downstream tsTF expression. Motif scanning and enrichment methods we employed in this study, while effective in revealing functionally relevant features over a large number of sequences, are notoriously unreliable in identifying key TF binding sites in individual enhancers. TF motifs are usually simple and commonly found in the genome, most instances not representing true binding sites[79]; and it is not required that an important TF's motifs be present in more copies in a given enhancer than the irrelevant ones. AI models based on convolutional neural networks have recently begun to overcome these challenges. Models trained on accessible sequences in a given cell-type can accurately predict the accessibility of even held-out sequences in that cell-type, indicating that they learned the underlying sequence rules for enhancer function (reviewed in[80]). Once trained, these models can be interpreted with tools such as deepSHAP[81], which reveal the specific base-pairs in each enhancer with the highest contributions towards model predictions, often representing *bona fide* TF binding sites.

Acknowledgements

We would like to thank all members of the Desplan and Özel Labs for helpful discussions, and Filippo Riva, Jay Unruh, Tom Kleist and Jason Morrison for technical assistance. We thank Filipe Pinto Teixeira Sousa and Saumya Jain for critical reading of the manuscript. We further thank Ted Erclik for sharing the Vsx1 antibody. This work was supported by the Stowers Institute for Medical Research, K99/R00-NS125117 and a Klingenstein-Simons Neuroscience Fellowship to M.N.Ö., and R01-EY017916 and R01-EY13010 to C.D.

Author Contributions

M.N.Ö. and C.D. conceived the project. M.N.Ö. performed all single-cell sequencing experiments. M.N.Ö., R.C., M.T., Y.C.C., O.B.T., R.R. and H.H. analyzed the sequencing data.

R.C., C.L. and R.R. performed all other experiments and analyzed the data. M.N.Ö., R.C. and M.T. wrote the manuscript. All authors edited the manuscript.

Declaration of Interests

Authors declare no conflicts of interest.

Data Availability

Raw and processed multiome data have been uploaded to GEO: GSE305940 and will be publicly available after peer review.

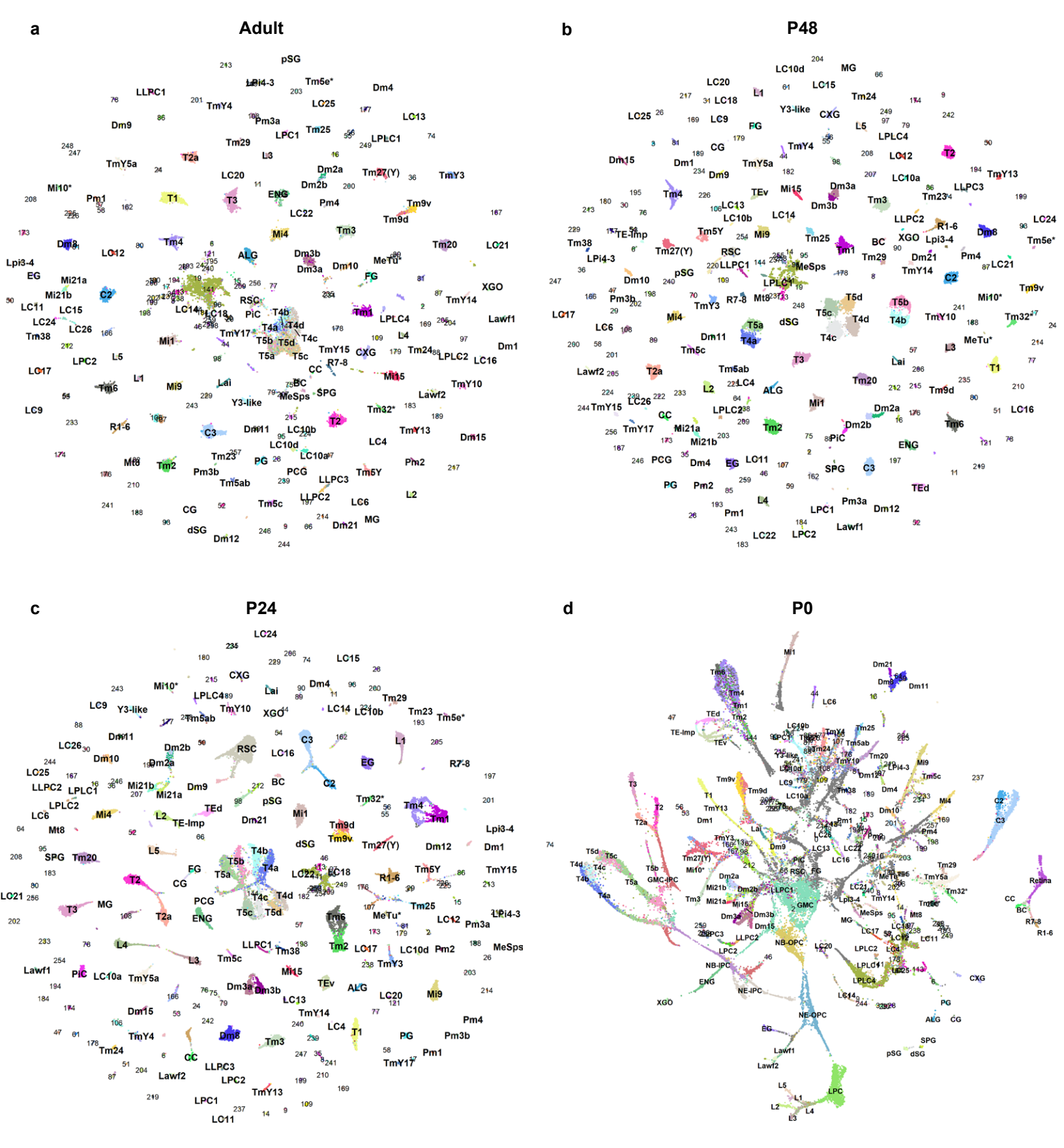

a
Adult
b
P48
c
P24
d
P0

Figure 1: A multiomic single-cell sequencing atlas of the *Drosophila* visual system

a-d, UMAP reductions of single-cell multiomic profiles from the developing *Drosophila* optic lobe, shown separately for each developmental stage: (**a**) Adult, (**b**) P48, (**c**) P24 calculated on 150 PCA and LSI components, and (**d**) P0, calculated on 70 PCA and LSI components. All embeddings were generated using weighted nearest neighbor (WNN) integration of RNA and ATAC modalities.

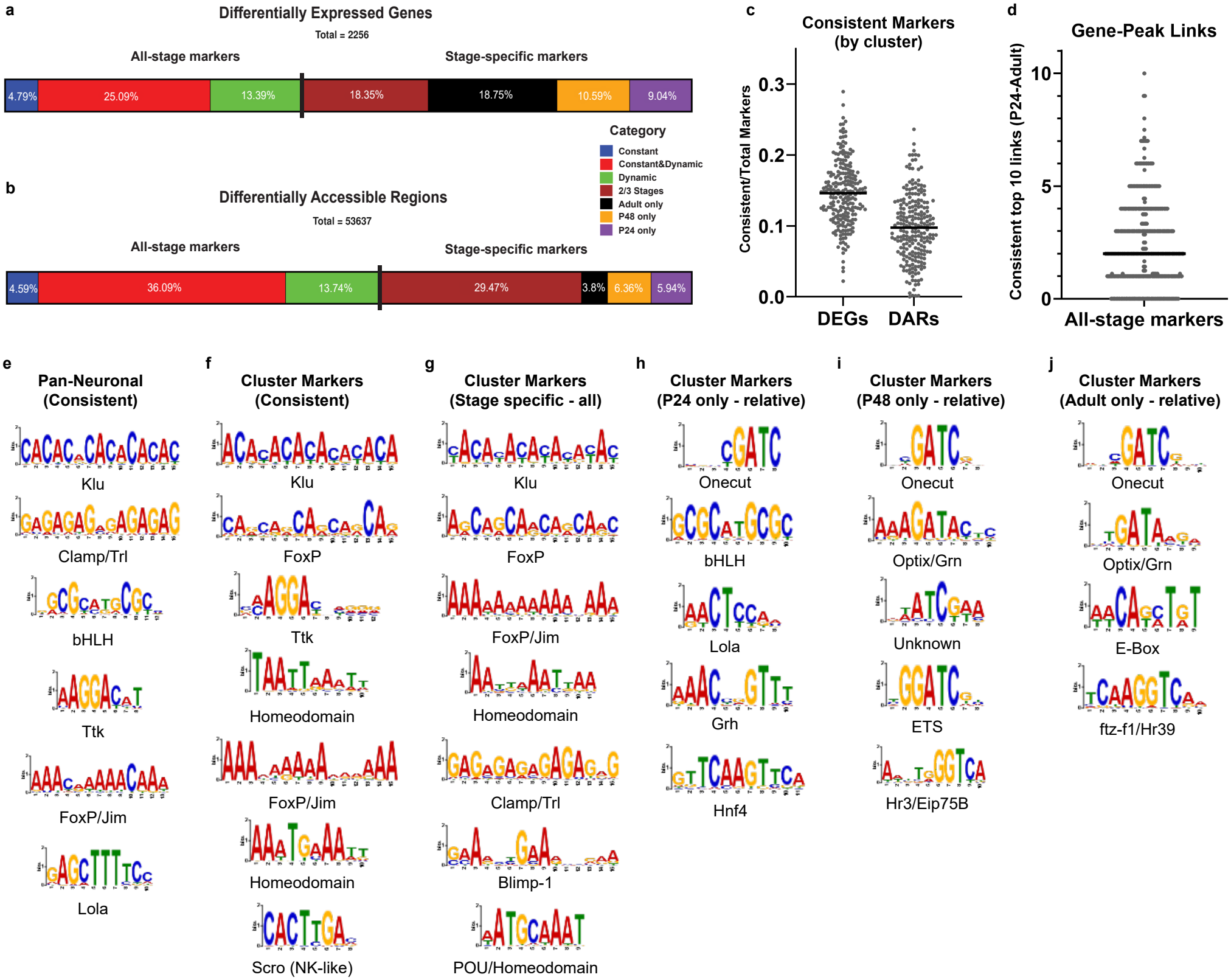
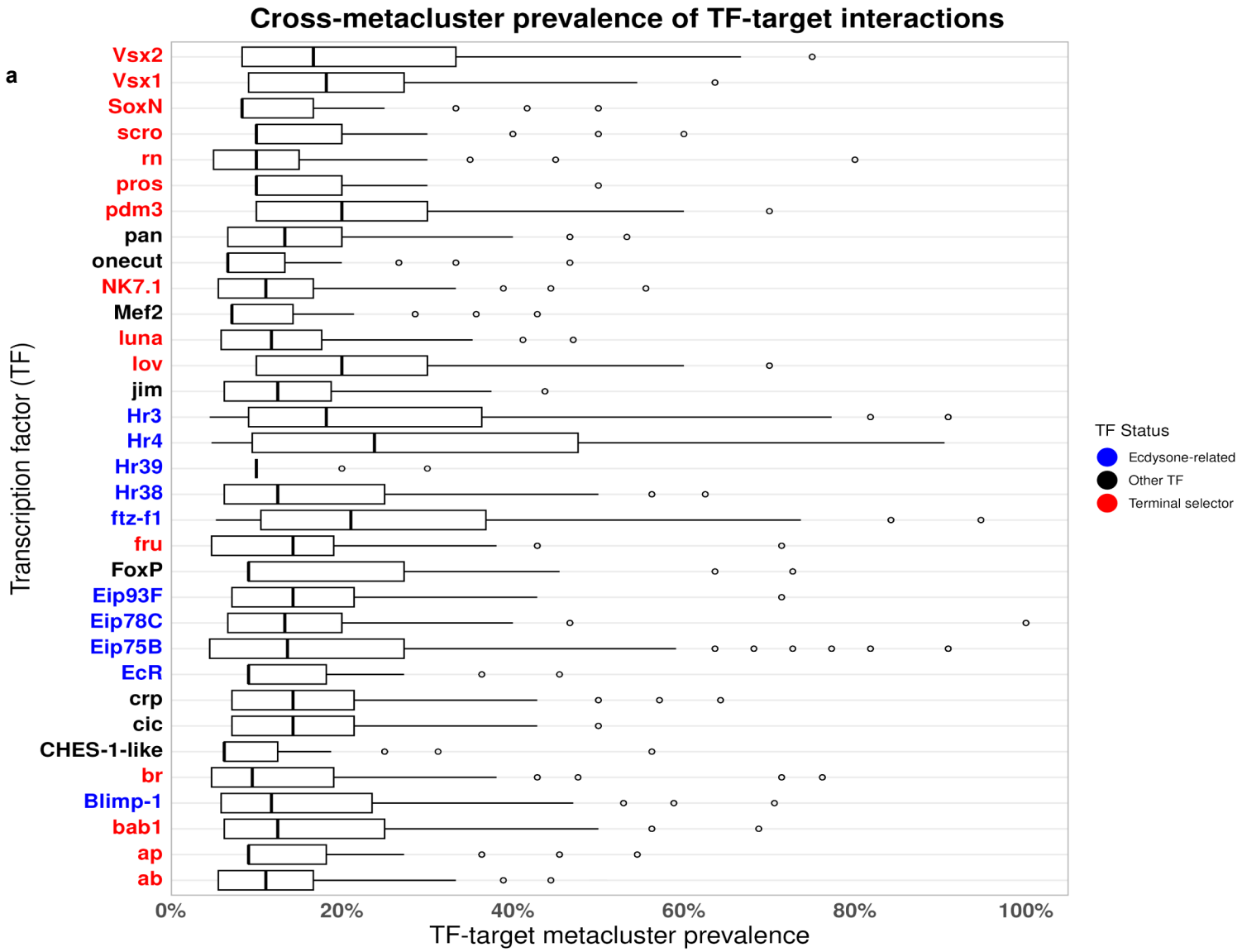

a
Differentially Expressed Genes
Total = 2256
All-stage markers
Stage-specific markers
4.79%
25.09%
13.39%
18.35%
18.75%
10.59%
9.04%
Category
Constant
Constant&Dynamic
Dynamic
2/3 Stages
Adult only
P48 only
P24 only
b
Differentially Accessible Regions
Total = 53637
All-stage markers
Stage-specific markers
4.59%
36.09%
13.74%
29.47%
3.8%
6.36%
5.94%
c
Consistent Markers
(by cluster)
Consistent/Total Markers
0.0
0.1
0.2
0.3
DEGs
DARs
d
Gene-Peak Links
Consistent top 10 links (P24-Adult)
0
5
10
All-stage markers
e
Pan-Neuronal
(Consistent)
Klu
Clamp/Trl
bHLH
Ttk
FoxP/Jim
Lola
f
Cluster Markers
(Consistent)
Klu
FoxP
Ttk
Homeodomain
FoxP/Jim
Homeodomain
Scro (NK-like)
g
Cluster Markers
(Stage specific - all)
Klu
FoxP
FoxP/Jim
Homeodomain
Clamp/Trl
Blimp-1
POU/Homeodomain
h
Cluster Markers
(P24 only - relative)
Onecut
bHLH
Lola
Grh
Hnf4
i
Cluster Markers
(P48 only - relative)
Onecut
Optix/Grn
Unknown
ETS
Hr3/Eip75B
j
Cluster Markers
(Adult only - relative)
Onecut
Optix/Grn
E-Box
ftz-f1/Hr39

Figure 2: Dynamic gene regulatory landscapes of differentiating neurons

a, Distribution of the 2,256 DEGs captured across development. Left: 976 markers found at all stages. These markers are further grouped as Dynamic (Green; dynamically regulated in all cell types), Constant and Dynamic (Red; continuously expressed in some cell types, but dynamically regulated in others), and Constant (Blue; continuously expressed in all cell types that express them). Right: The remaining 1279 DEGs, which were captured as markers in 1 or 2 stages, but not all 3. Colors reflect the stages that these DEGs were found as neuronal markers. **b**, Same as in (a), but for DARs. 29,190 all-stage DARs are on the left, and 24,447 stage-specific DARs are on the right. **c**, Scatter plot depicting the proportion of markers (DEGs, left and DARs, right) shared across all 3 stages on a per-cluster basis. **d**, Strip plot depicting the counts of top-10 gene-peak links that are consistently found across development (P24-Adult), corrected for genes with <10 peaks linked. **e-j**, STREME results depicted as sequence logos, ordered (top to bottom) based on E-value of enrichment. The names were assigned based on Tomtom results matching these PWMs to the known *Drosophila* TF motifs. **e**, Motifs enriched in consistent pan-neuronal peaks (not DARs). **f**, Motifs enriched in consistent neuronal DARs. **g**, Motifs enriched in DARs found at only one stage (P24, P48, or Adult). **h-j**, Motifs relatively enriched in DARs found at P24 (**h**), P48 (**i**) or Adult (**j**). Shuffled input sequences (**e-g**) or the sequences of consistent neuronal DARs (**h-j**) were used background for motif enrichment.


a
Cross-metacluster prevalence of TF-target interactions
Transcription factor (TF)
Vsx2
Vsx1
SoxN
scro
rn
pros
pdm3
pan
onecut
NK7.1
Mef2
luna
lov
jim
Hr3
Hr4
Hr39
Hr38
ftz-f1
fru
FoxP
Eip93F
Eip78C
Eip75B
EcR
crp
cic
CHES-1-like
br
Blimp-1
bab1
ap
ab
0%
20%
40%
60%
80%
100%
TF-target metacluster prevalence
TF Status
Ecdysone-related
Other TF
Terminal selector

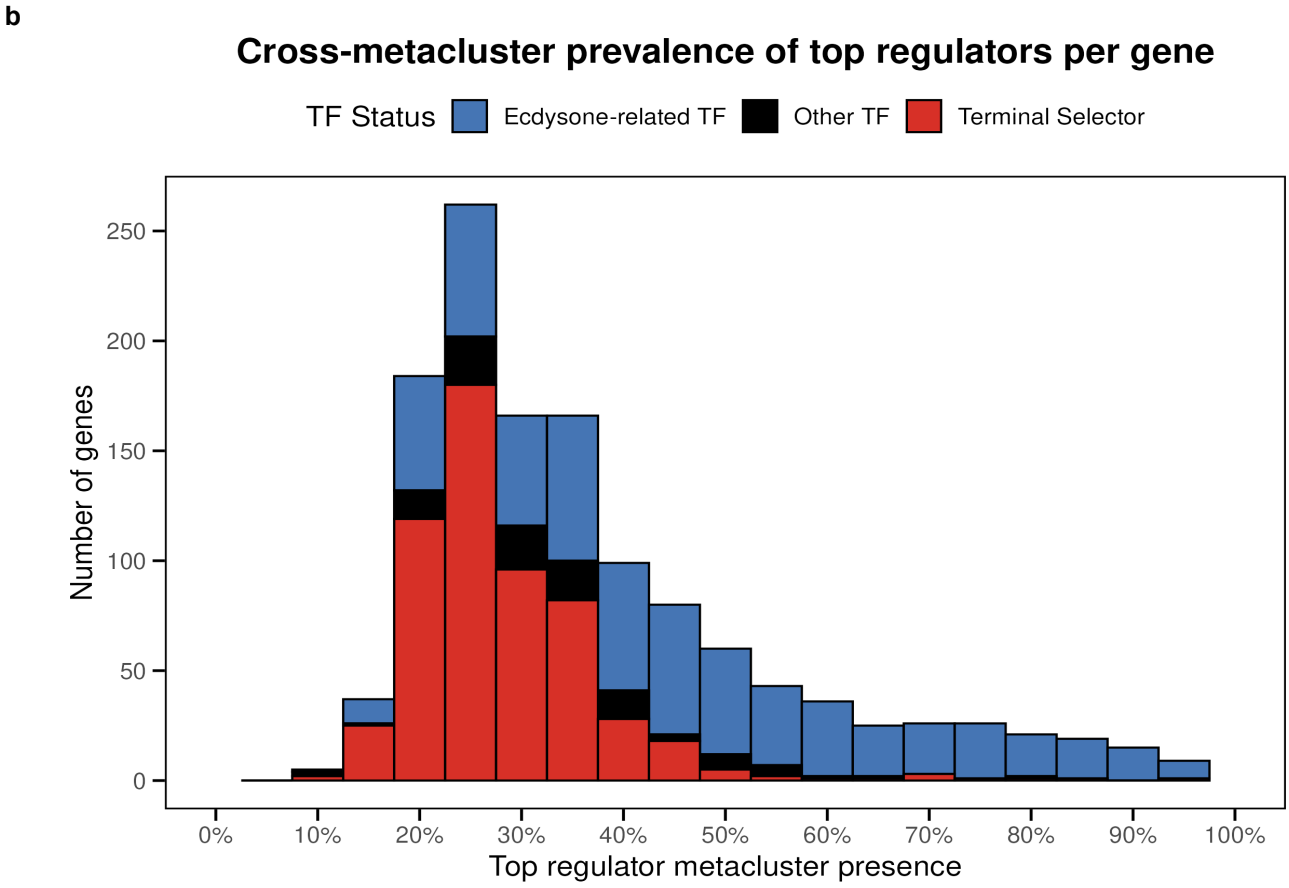

b
Cross-metacluster prevalence of top regulators per gene
TF Status
Ecdysone-related TF
Other TF
Terminal Selector
Number of genes
0
50
100
150
200
250
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
Top regulator metacluster presence

Figure 3: Context specificity of TF-target interactions

a, Boxplots portraying the prevalence (%) of TF-gene pairs predicted across SCENIC+ networks. The TFs shown on the y-axis are found as regulators in at least 10 different neuronal networks and are colored by their functional classes: Terminal selector (red), Ecdysone-related (blue), or Other (black). **b**, Histogram depicting the persistence of top regulators (*i.e.,* the TF most consistently predicted to regulate each gene) for genes present in at least 10 different networks. Each colored section is interpreted as a proportion of the whole bar. The bars are colored by the regulator (TF) class, same as in (a).

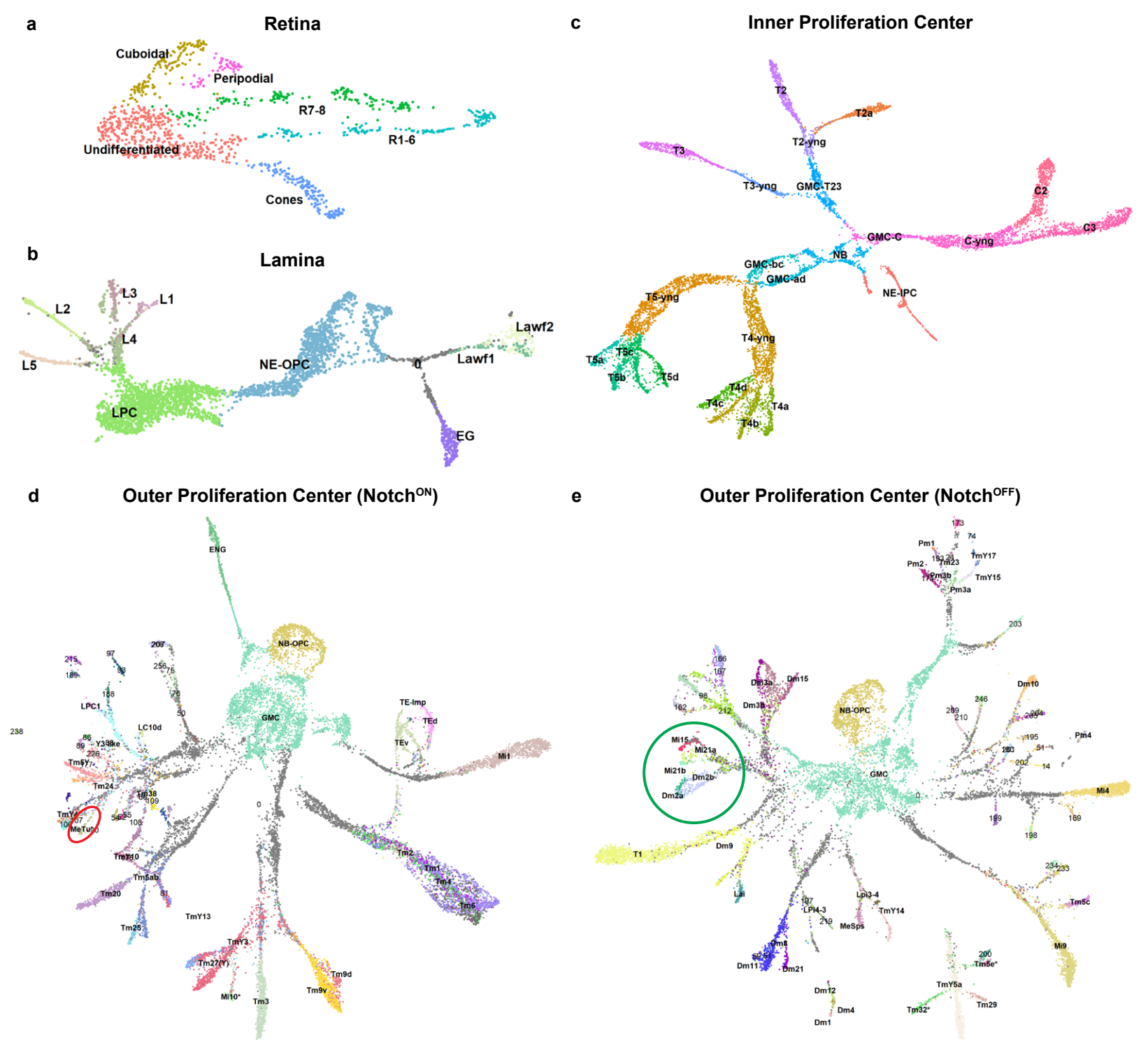

a
Retina
Cuboidal
Peripodial
R7-8
R1-6
Undifferentiated
Cones
b
Lamina
L3
L1
L2
L4
L5
NE-OPC
Lawf2
Lawf1
0
LPC
EG
c
Inner Proliferation Center
T2
T2a
T2-yng
T3
T3-yng
GMC-T23
C2
C3
GMC-C
C-yng
NB
GMC-bc
GMC-ad
NE-IPC
T5-yng
T4-yng
T5a
T5c
T5b
T5d
T4d
T4c
T4a
T4b
d
Outer Proliferation Center (Notch^ON)
ENG
NB-OPC
GMC
TE-Imp
TEd
TEv
Mi1
LPC1
LC10d
Tm24
Tm5ab
Tm20
Tm25
TmY13
TmY3
Tm27(Y)
Mi10*
Tm3
Tm9d
Tm9v
Tm2
Tm1
Tm4
Tm6
e
Outer Proliferation Center (Notch^OFF)
Pm1
Pm2
Pm3b
Pm3a
TmY17
TmY15
Dm3a
Dm15
Dm3b
NB-OPC
Dm10
Pm4
Mi15
Mi21a
Mi21b
Dm2b
Dm2a
GMC
Mi4
T1
Dm9
Lai
Lpi3-4
LPi4-3
TmY14
MeSps
Tm5c
Mi9
Dm8
Dm11
Dm21
Tm5e*
TmY5a
Dm12
Dm4
Dm1
Tm32*
Tm29

Figure 4: Differentiation trajectories during optic lobe neurogenesis

UMAP reductions of single-cell multiomic profiles from the P0 optic lobes (from Fig. 1d). (**a**) Retinal lineage, calculated on 30 PCA and LSI components. (**b**) Lamina (lateral OPC) lineage, calculated on 100 PCA and 90 LSI components. (**c**) IPC lineage, calculated on 20 PCA and 22 LSI components. **d-e**, Medial OPC lineages Notch^{ON} (**d**), calculated on 33 PCA and 29 LSI components; and Notch^{OFF} (**e**), calculated on 30 PCA and 38 LSI components. In (a) and (c), identities were manually assigned based on WNN clustering and marker gene expression. In other panels, progenitor identities were assigned manually based on clustering and marker genes. Neuronal identities were assigned based on label transfer from the P24 dataset (Fig. 1c). Young neurons for which these could not be confidently determined are shown in grey. Red ellipse (in d) marks MeTu neurons without ap expression despite their N^{ON} status (Suppl. Fig. 2d). The green circle (in e) indicates the M24 neurons shown in Fig. 6b-e. NE: Neuroepithelia; NB: Neuroblasts; GMC: ganglion mother cells; LPC: lamina precursor cells.

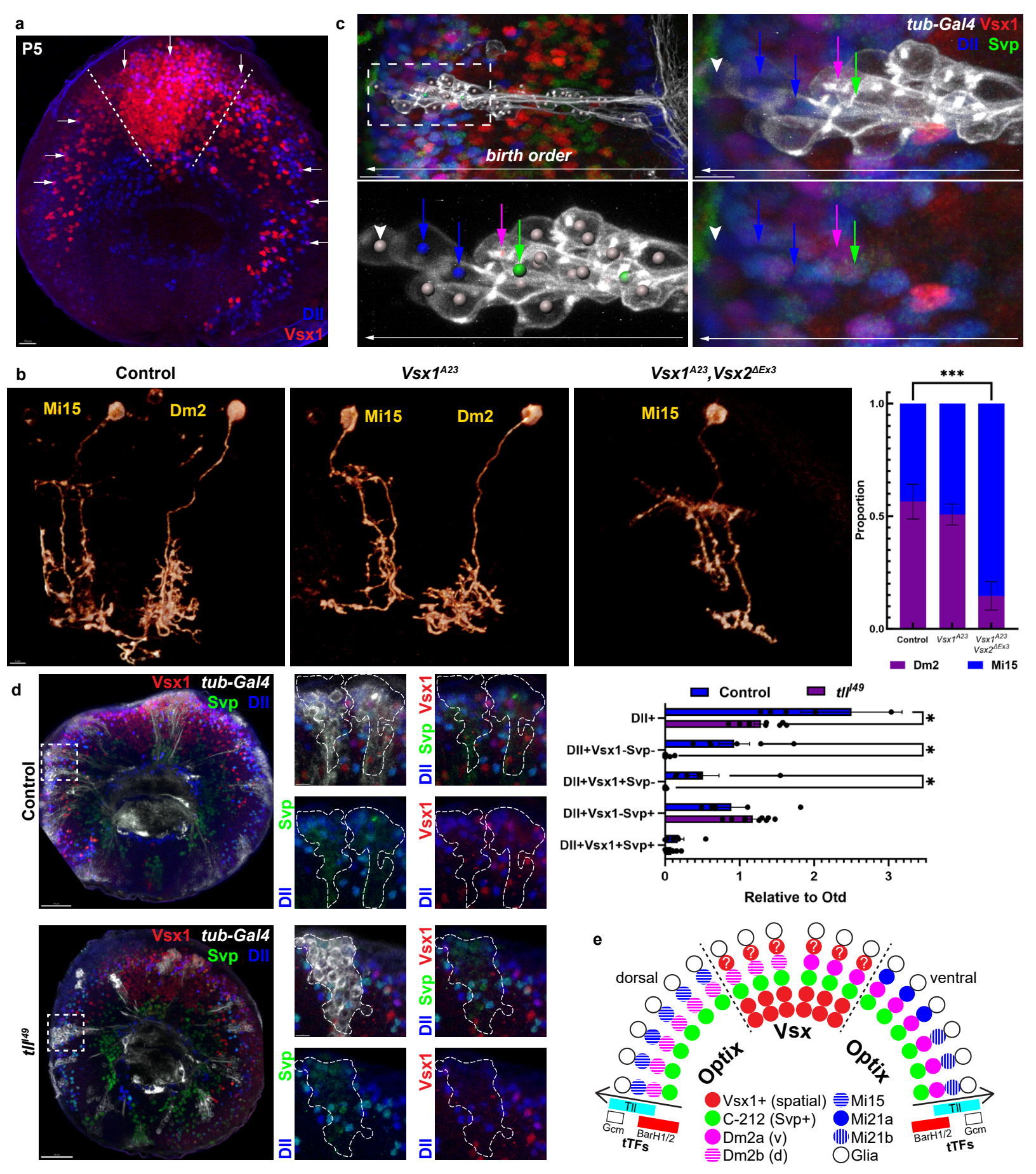

a
P5
Dll
Vsx1
c
birth order
tub-Gal4 Vsx1
Dll Svp
b
Control
Vsx1A23
Vsx1A23,Vsx2ΔEx3
Mi15
Dm2
Mi15
Dm2
Mi15
***
1.0
0.5
0.0
Proportion
Control
Vsx1A23
Vsx1A23 Vsx2ΔEx3
Dm2
Mi15
d
Control
tll49
Vsx1 tub-Gal4
Svp Dll
Vsx1
Svp
Dll
Control
tll49
Dll+
Dll+Vsx1-Svp-
Dll+Vsx1+Svp-
Dll+Vsx1-Svp+
Dll+Vsx1+Svp+
0
1
2
3
Relative to Otd
e
dorsal
ventral
Vsx
Optix
Optix
Tll
Gcm
BarH1/2
tTFs
Vsx1+ (spatial)
C-212 (Svp+)
Dm2a (v)
Dm2b (d)
Mi15
Mi21a
Mi21b
Glia

Figure 5: Vsx1 expression in Dm2 and the birth order of Tailless window neurons

a, Wild-type optic lobe during neurogenesis (P5) with dashed lines marking the Vsx1 spatial domain and white arrows pointing to Dll+Vsx1+ cells (purple). Anti-Dll (blue) and anti-Vsx1 (red) are shown as maximum intensity projections within a representative sample. n = 11 optic lobes.

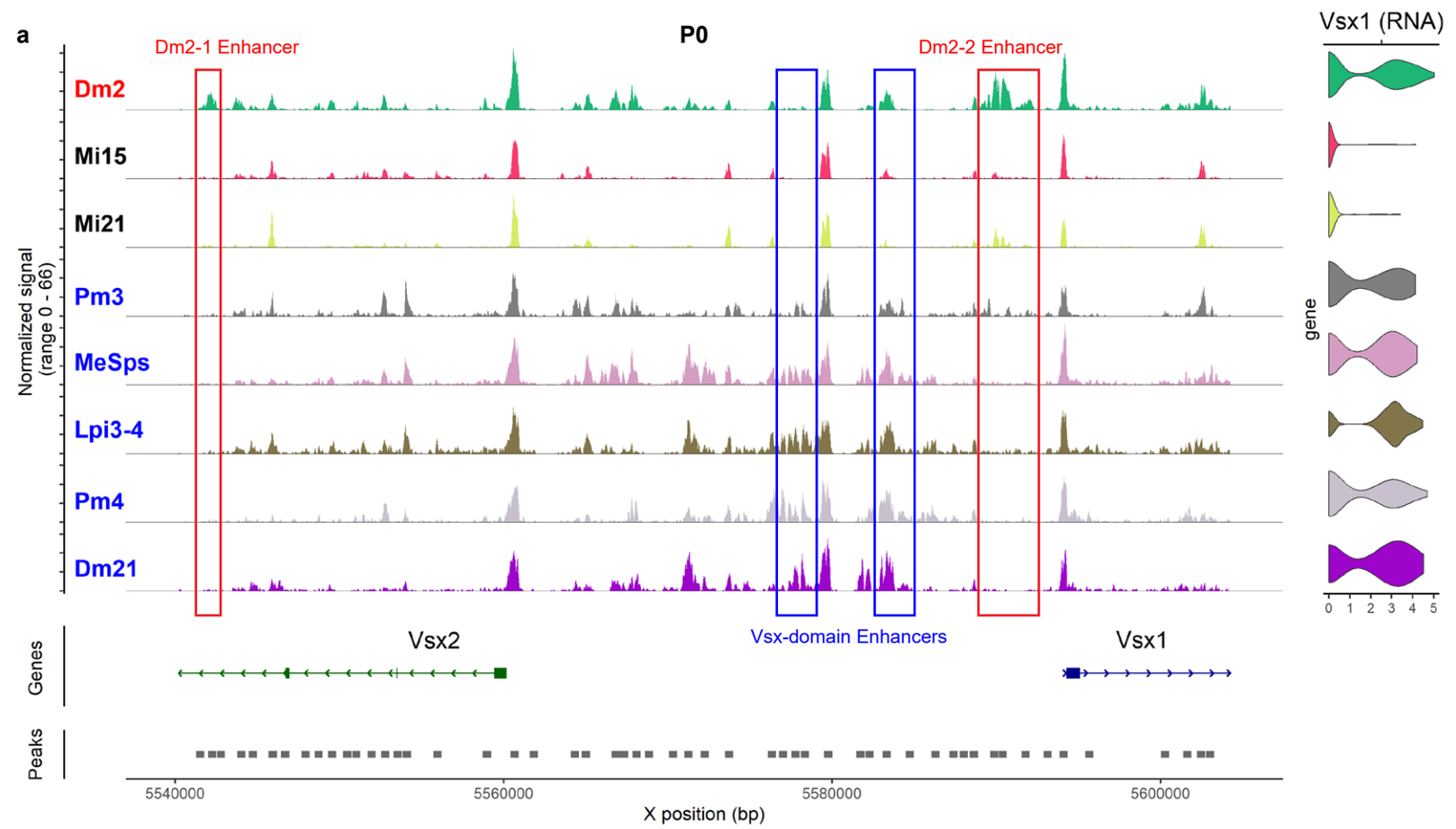
b, FRT19A (control), FRT19A *Vsx1*^{A23}, and FRT19A *Vsx1*^{A23}*Vsx2*$^{\Delta Ex3}$ MARCM clones labeled with *R26H07(Dm2/Mi15)-Gal4* driving *CD4:tdGFP* in Adult brains with clones induced by heat-shock at late L3 stage. Blended three-dimensional (3D) reconstructions show individual clones annotated with identity as representative images. Bar plot: quantification of **b** with error bars representing SEM. n = 71/18/12 control, n = 62/13/13 *Vsx1*^{A23}, and n = 35/24/21 *Vsx1*^{A23}*Vsx2*$^{\Delta Ex3}$ neurons/optic lobes/brains. In comparison to the control, p=0.58 for *Vsx1*^{A23} and p = 0.0002 for *Vsx1*^{A23}*Vsx2*$^{\Delta Ex3}$.

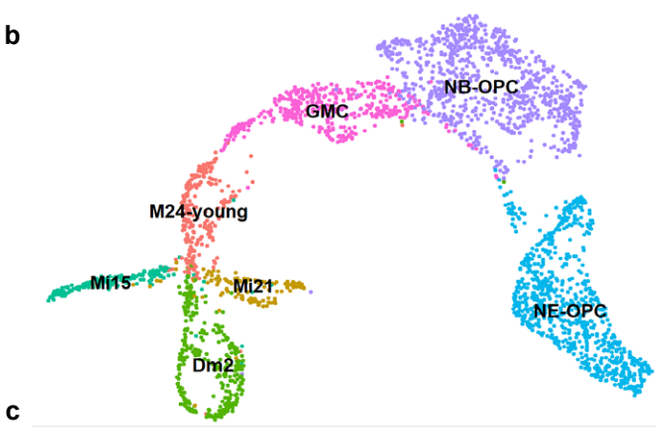
c, FRT82B MARCM clones labeled with *tubulin-Gal4* driving *myr:GFP* in early pupal (~P5-P10) optic lobes approximately 50 hours after heat-shock during larval stage. Clones are located within the dorsal Optix domain that produces all Mi15 and a portion of Dm2 neurons[82]. Neurons are born from NBs in sequential temporal windows from right to left marked by the white arrow. The area indicated by the white dashed box in top left panel is shown in all other panels with increased magnification. All cells marked with arrows express Dll with green arrow also expressing Svp and Vsx1 and magenta arrow also expressing Vsx1. Arrowhead indicates glia. Anti-GFP (white), anti-Dll (blue), anti-Vsx1 (red), and anti-Svp (green) are shown as a maximum intensity projection n = 4 brains.

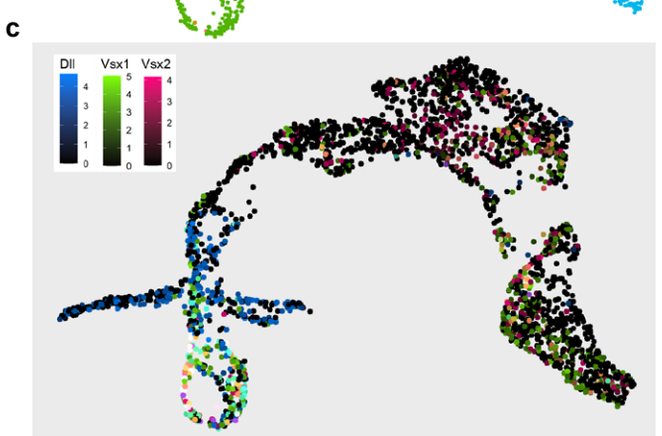
d, FRT82B (control) and FRT82B *tll*^{l49} MARCM clones labeled with *tubulin-Gal4* driving *CD4:tdGFP* during neurogenesis (P5). The areas indicated by white dashed boxes are shown to the right with higher magnification. Dorsal Optix clones (right) are marked in dashed white lines. Anti-GFP (white), anti-Dll (blue), anti-Svp (green), and anti-Vsx1 (red) are shown as maximum intensity projections. Optix domain clones were quantified relative to Otd expression (*Methods*), with multiple clones per optic lobe quantified and summed for a single measurement per lobe. Bar plot: Quantification of **d** with error bars representing SEM and dots representing summed clones within a single optic lobe. In comparison to the control, *tll*^{l49} Dll+Vsx1-Svp-, p = 0.0001, Dll-Vsx1+Svp-, p = 0.008751, Dll+Vsx1+Svp-, p = 0.012, while changes in Dll+Vsx1-Svp+, Dll+Vsx1+Svp+, Dll-Vsx1-Svp+, and Dll-Vsx1+Svp+ neurons were all non-significant. Also shown is the number of all Dll+ neurons regardless of Vsx1 or Svp expression, p = 0.047. n = 6/3/23 control and n = 9/5/32 optic lobes/brains/clones.

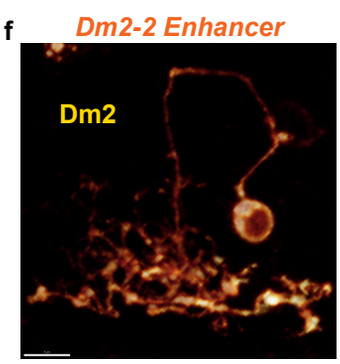
e, Graphical representation of M24 cell types with regard to spatial and temporal patterning of the OPC. The postmitotic cells (circles) extend in concentric half-circles representative of a NB producing progeny along a temporal axis (arrow) with spatial domains annotated in bold and delineated by dashed lines. Along the temporal axis, boxes represent the relevant tTFs' expression and their overlap; BarH1/2 (red), Tll (cyan), and Gcm (white). Red circles represent Vsx1 expressing progeny that arise exclusively from the Vsx spatial domain. Green circles represent cluster 212 neurons which expresses Dll, Svp, and Vsx1 and are born before M24 neurons. Magenta circles represent Dm2 which arise first from the Tll window and are patterned into distinct subtypes along the dorsoventral axis (**Suppl. Fig. 8**). Blue circles represent the latest born M24 neurons, Mi15 and Mi21. Mi15 (dorsal Optix) is shown with horizontal stripes while Mi21 neurons (ventral Optix) are predicted to be produced during the same temporal window. Red circles with white question marks depict the unknown cell type that may be produced exclusively from the Vsx domain during this temporal window. White circles depict the glia produced last during the overlap of Tll and Gcm[69].

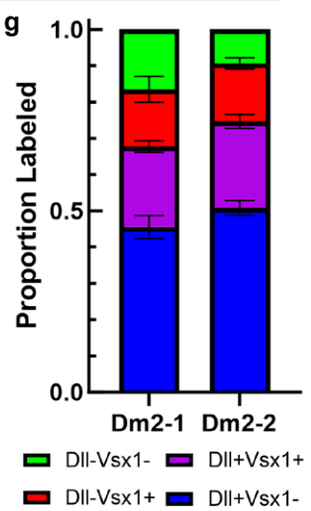
Scale bars: **a**, 10 μm, **b**, 3 μm, **c**, 10 and 4 μm, **d**, 30 and 7 μm.


a
P0
Dm2-1 Enhancer
Dm2-2 Enhancer
Vsx1 (RNA)
Dm2
Mi15
Mi21
Pm3
MeSps
Lpi3-4
Pm4
Dm21
Normalized signal
(range 0 - 66)
gene
0 1 2 3 4 5
Genes
Vsx2
Vsx-domain Enhancers
Vsx1
Peaks
5540000
5560000
5580000
5600000
X position (bp)


b
NB-OPC
GMC
M24-young
Mi15
Mi21
NE-OPC
Dm2


c
Dll
Vsx1
Vsx2


f
Dm2-2 Enhancer
Dm2


g
1.0
0.5
0.0
Proportion Labeled
Dm2-1
Dm2-2
Dll-Vsx1-
Dll+Vsx1+
Dll-Vsx1+
Dll+Vsx1-

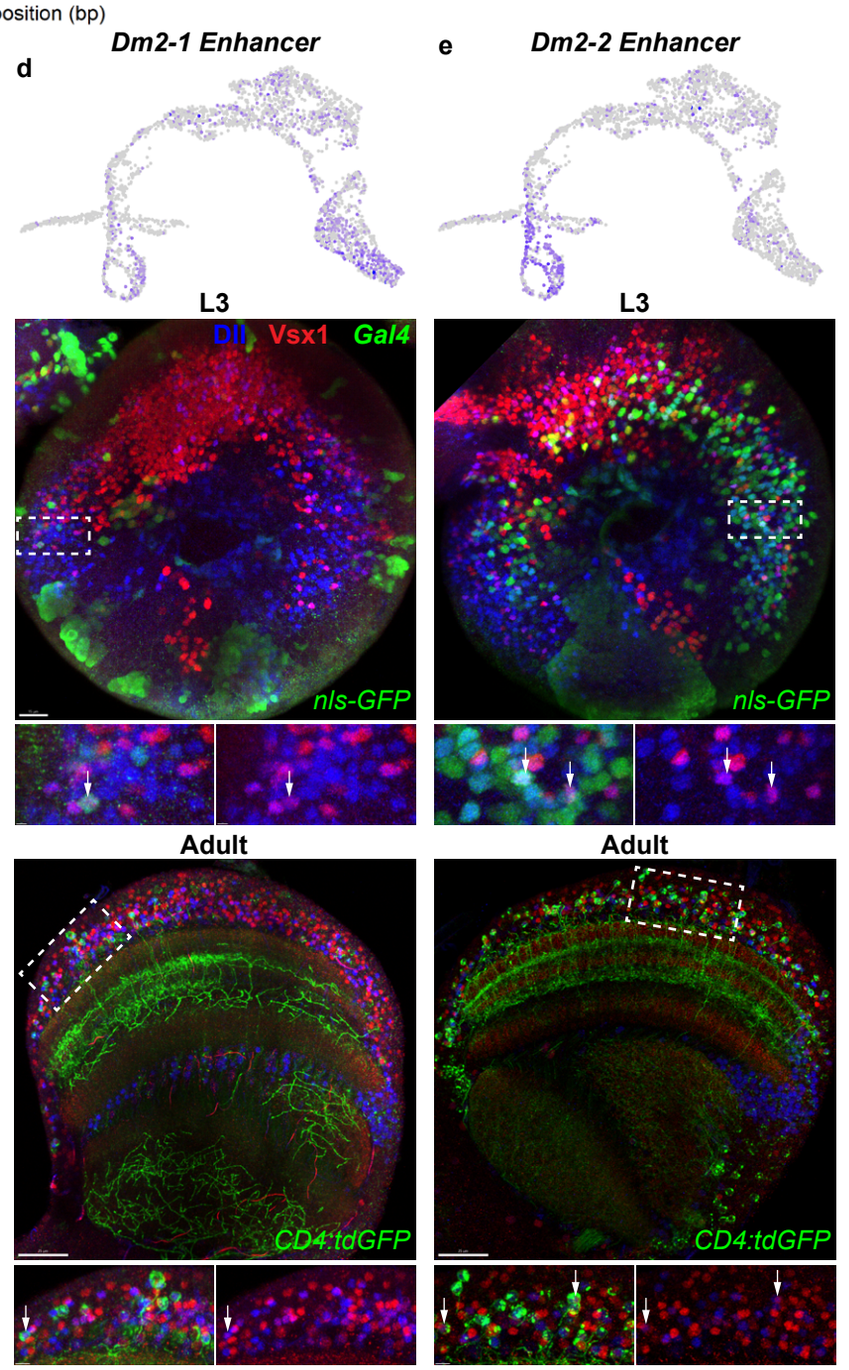

d
Dm2-1 Enhancer
e
Dm2-2 Enhancer
L3
Dll
Vsx1
Gal4
nls-GFP
Adult
CD4:tdGFP

Figure 6: Temporally patterned enhancers of the *Vsx* locus

a, Coverage plot of the *Vsx1/2* locus with tracks representing normalized accessibility in each cell-type at P0. M24 neurons are at the top with Dm2 in red and Mi15/21 in black (the clusters containing the subtypes of Dm2 and Mi21 were merged). Five representative Vsx1+ cell types that exclusively originate from the Vsx1 NE spatial domain are shown in blue (two subtypes of Pm3 were also merged). Rectangles highlight Dm2-1 and Dm2-2 enhancers with accessibility exclusive to Dm2 (red) and the enhancers consistently accessible within spatially defined cell types (blue). *Vsx1* mRNA expression (normalized) in each cell type is shown on the right side.

b, UMAP reduction of the M24 trajectory at P0, subsetted from Fig. 4e and WNN calculated using 20 PCA and 24 LSI dimensions (see *Methods*), depicting shared progenitor populations of the OPC NE and NB, and a portion of the GMCs from late temporal windows, and all M24 neurons. “M24-young” cluster represents neurons that are too immature to assign identity.

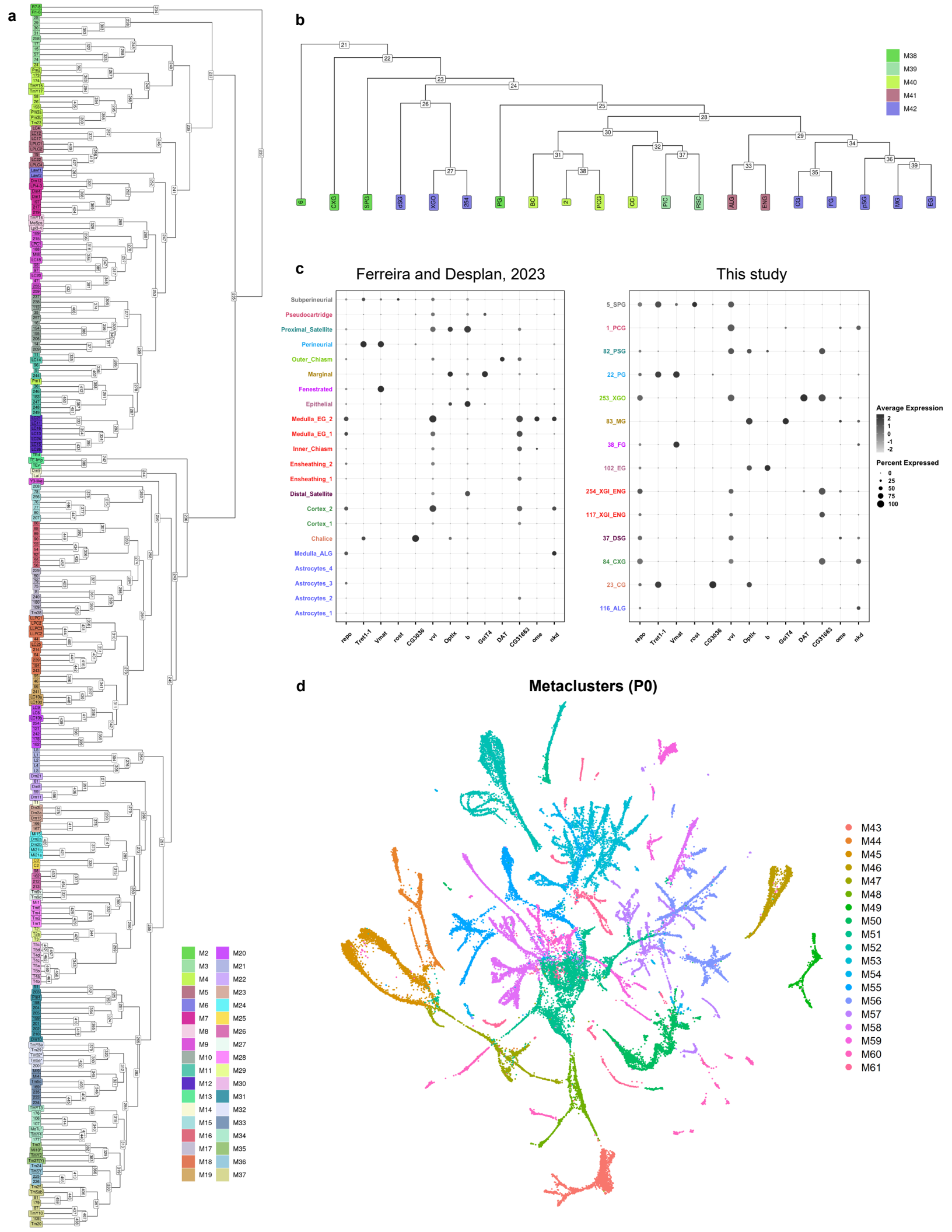
c, Log-normalized mRNA expression levels of *Dll* (blue), *Vsx1* (green), and *Vsx2* (red) at P0 are overlayed on the same UMAP as in **b**.

d-e, Dm2-1 and Dm2-2 enhancer accessibility displayed on the UMAP from **b**. Both enhancers were cloned into a *Gal4* reporter which drives nuclear (nls) GFP within L3 brains (Dm2-1, n = 2 and Dm2-2, n = 5 brains) and membrane-tagged (CD4) GFP within adult brains (Dm2-1, n = 5 and Dm2-2, n = 6 brains). Representative images from each stage are shown as maximum intensity projections with anti-GFP (green), anti-Dll (blue), and anti-Vsx1 (red). White arrows point to Dll/Vsx1 co-expression.

f, 3D reconstruction of an adult Dm2 neuron labeled with *Dm2-2-Gal4* driving sparsely expressed (FLP-out) *CD4:tdGFP*.

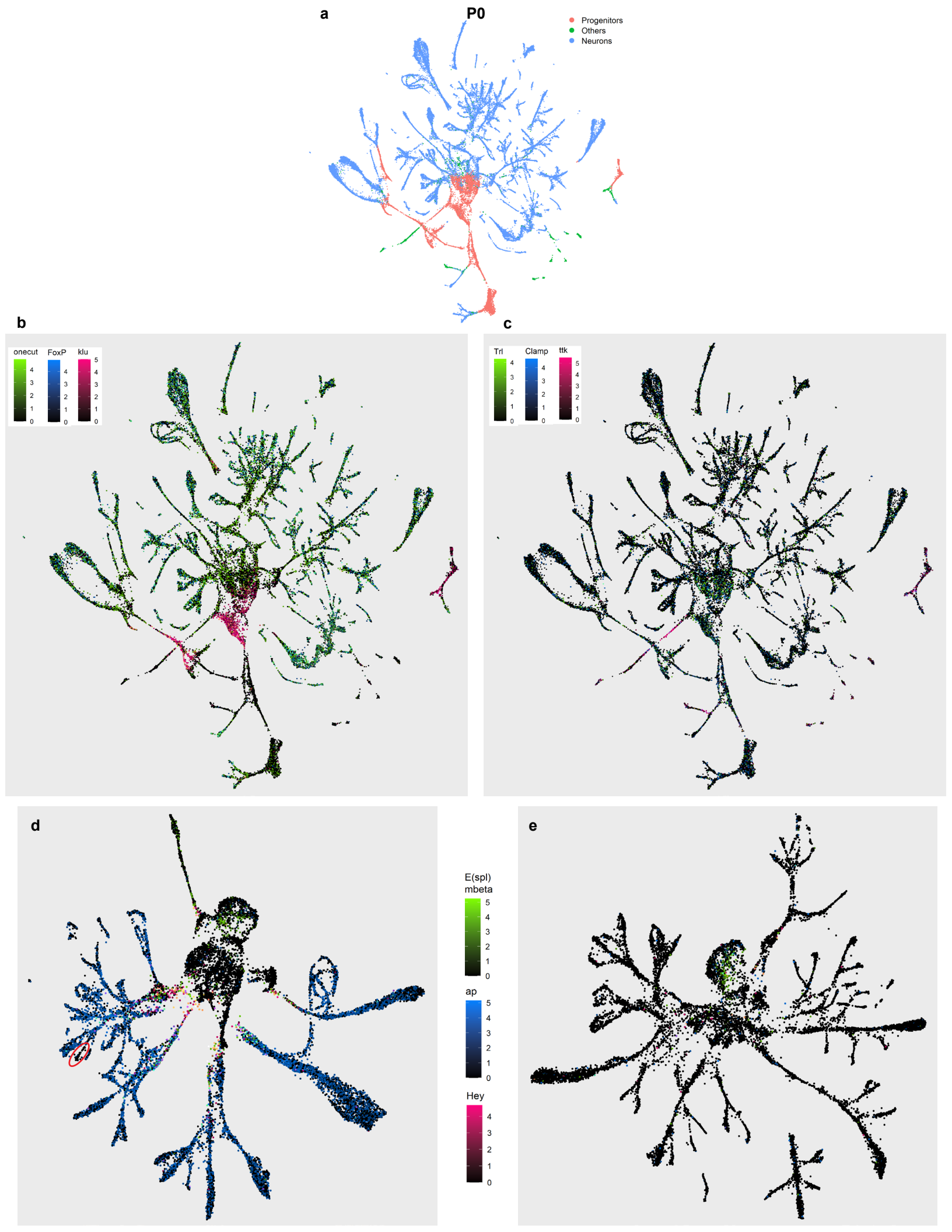
g, Quantification of adult brains from **d** and **e** for Dm2-1 and Dm2-2 enhancers was performed as a relative proportion of all cells labeled by the reporter with error bars indicating SEM. Each Dm2-1 sample had one region of interest (ROI) placed in a position in the medulla with dense labeling, while Dm2-2 samples had two ROIs quantified and averaged for each brain. n = 516/5 (Dm2-1) and n = 815/6 (Dm2-2) neurons/brains.

Scale bars: **d** and **e**, L3, 15 and 2 μm, Adult, 25 and 4 μm. **f**, 4 μm


a
M2
M3
M4
M5
M6
M7
M8
M9
M10
M11
M12
M13
M14
M15
M16
M17
M18
M19
M20
M21
M22
M23
M24
M25
M26
M27
M28
M29
M30
M31
M32
M33
M34
M35
M36
M37
b
M38
M39
M40
M41
M42
6
CXG
SPG
dSG
XOO
254
PG
BC
2
PCG
CC
PiC
RSC
ALG
ENG
CG
FG
pSG
MG
EG
c
Ferreira and Desplan, 2023
This study
Subperineurial
Pseudocartridge
Proximal_Satellite
Perineurial
Outer_Chiasm
Marginal
Fenestrated
Epithelial
Medulla_EG_2
Medulla_EG_1
Inner_Chiasm
Ensheathing_2
Ensheathing_1
Distal_Satellite
Cortex_2
Cortex_1
Chalice
Medulla_ALG
Astrocytes_4
Astrocytes_3
Astrocytes_2
Astrocytes_1
5_SPG
1_PCG
82_PSG
22_PG
253_XGO
83_MG
38_FG
102_EG
254_XGI_ENG
117_XGI_ENG
37_DSG
84_CXG
23_CG
116_ALG
repo
Tret1-1
Vmat
rost
CG3036
vvl
Optix
b
GstT4
DAT
CG31663
ome
nkd
Average Expression
2
1
0
-1
-2
Percent Expressed
0
25
50
75
100
d
Metaclusters (P0)
M43
M44
M45
M46
M47
M48
M49
M50
M51
M52
M53
M54
M55
M56
M57
M58
M59
M60
M61

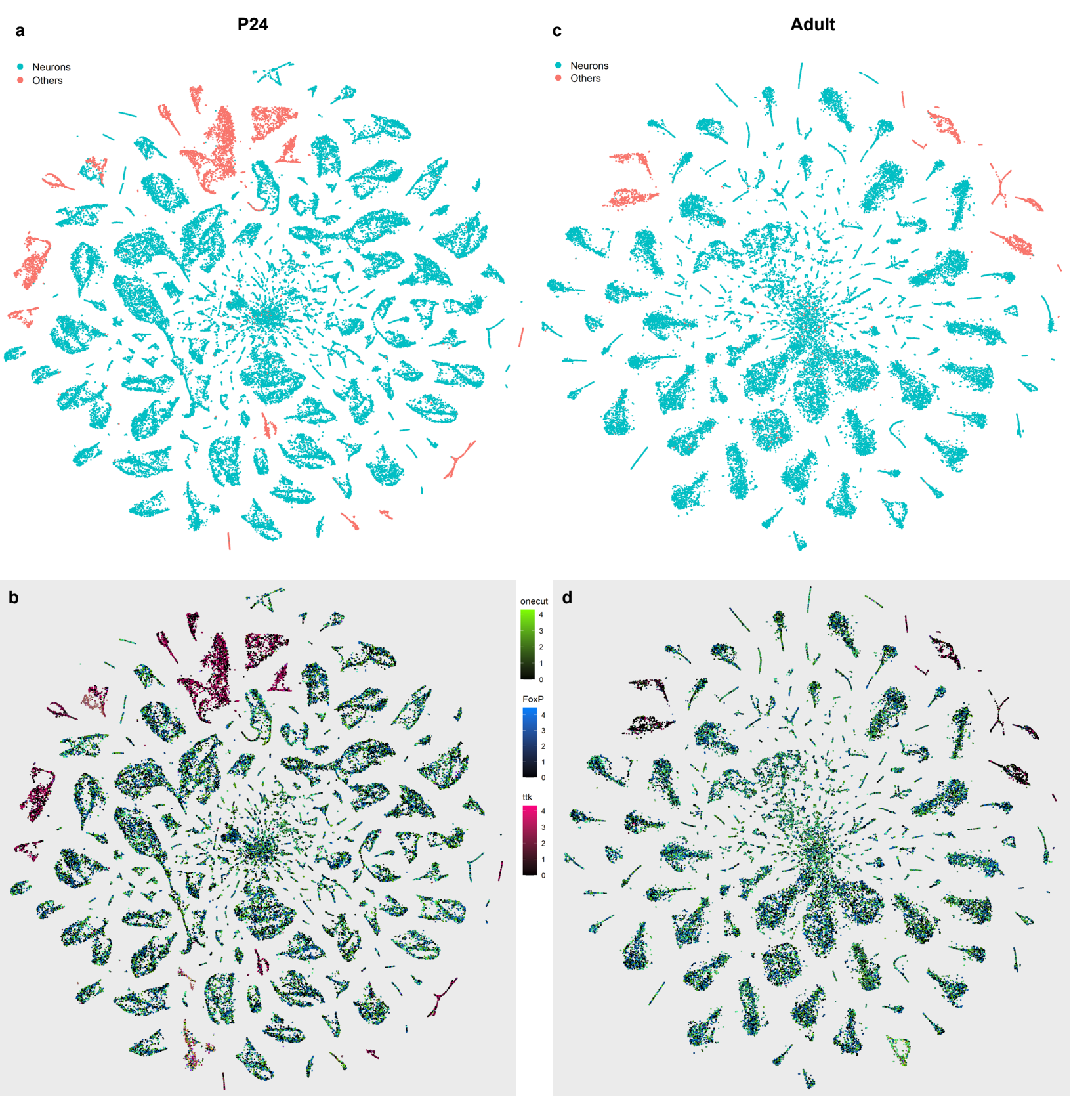
Supplementary Figure 1: Meta-clustering of the multiome data and the annotation of glial clusters

a-b, Hierarchical tree of neuronal clusters (**a**) and non-neuronal clusters (**b**) based on transcriptomic similarity using 100 principal components (a) or 20 principal components (b) of the RNA modality at P24. The dendrograms were generated using Seurat's *BuildClusterTree* function. Branch lengths indicate transcriptional distance between clusters. See *Methods* for more details. **c**, Annotations of glial clusters identified in this atlas (right), based on previously described marker genes[44] (left). **d**, UMAP reduction of the P0 dataset same as in Fig. 1d, colored by metacluster assignment.


a
P0
Progenitors
Others
Neurons
b
onecut
FoxP
klu
c
Trl
Clamp
ttk
d
E(spl)
mbeta
ap
Hey
e

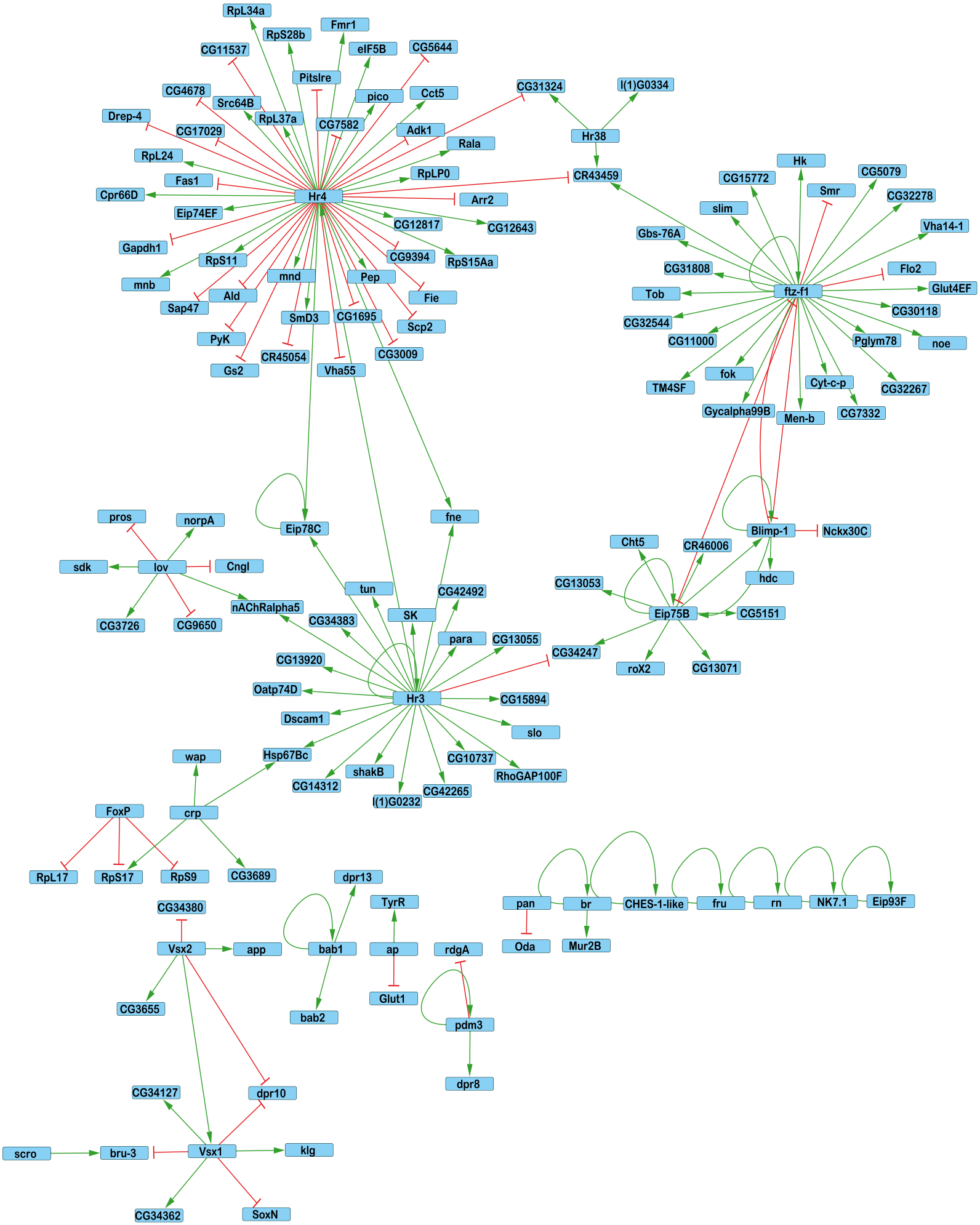
Supplementary Figure 2: UMAP trajectories of optic lobe neurogenesis

a-c, UMAP reduction of the P0 dataset same as in Fig. 1d, colored by: (**a**) Broad cell type categories (Neurons: blue, Progenitors: red, Others: green), (**b**) Log-normalized mRNA expression of *Onecut*, *FoxP*, and *klu* (green, blue, and red, respectively), (**c**) *Trl*, *Clamp*, and *ttk* expression (green, blue, and red, respectively). **d-e**, UMAP reductions of the Notch^{ON} and Notch^{OFF} hemilineages of the OPC, same as in Fig. 4d-e, and colored based on the log-normalized expression of indicated genes (*Hey*: red, *E(spl)mbeta*: green, *ap*: blue).


a
P24
Neurons
Others
c
Adult
Neurons
Others
b
onecut
4
3
2
1
0
FoxP
4
3
2
1
0
ttk
4
3
2
1
0
d

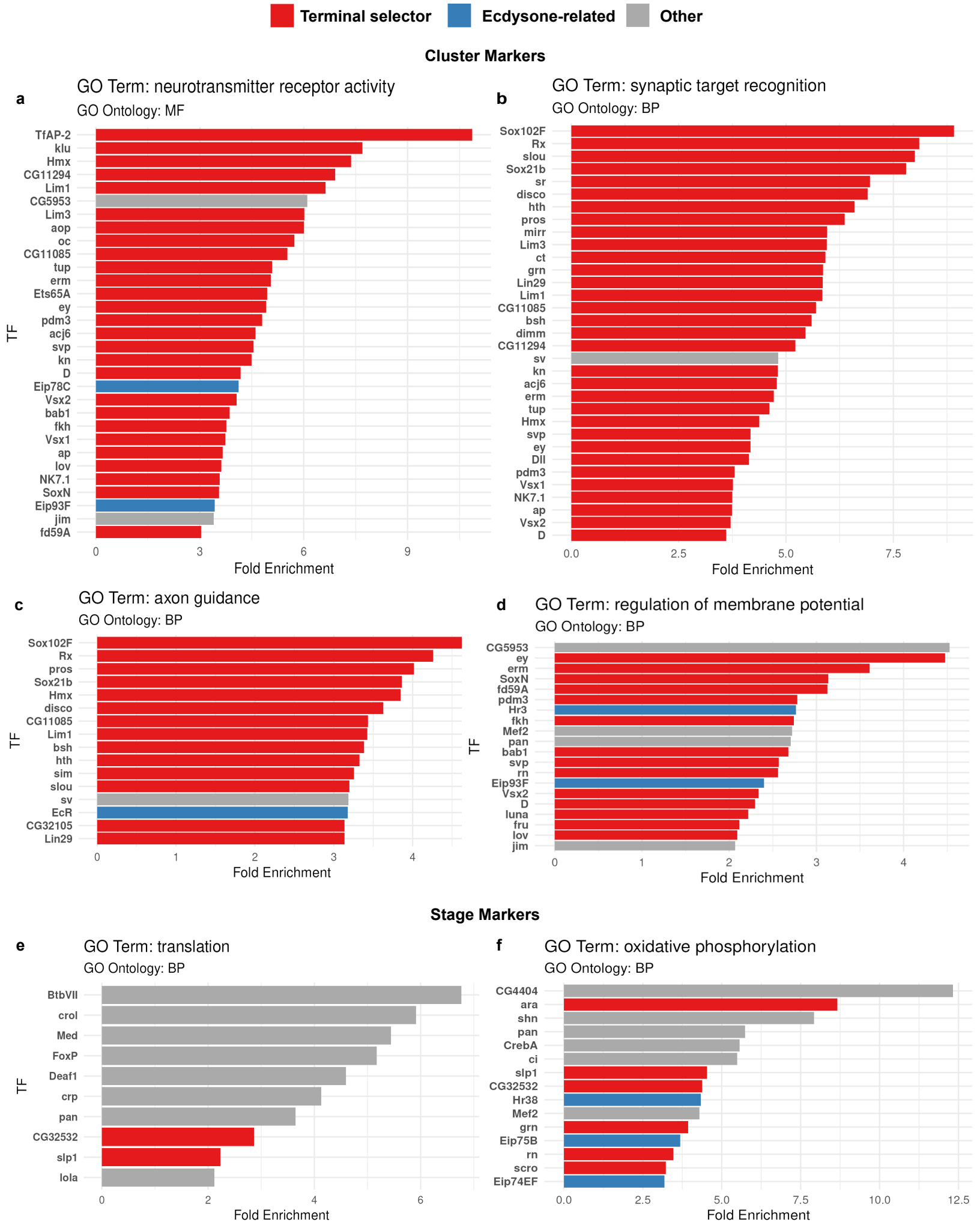
Supplementary Figure 3: Global expression patterns of pan-neuronal genes

a-d, tSNE reductions of the P24 (a-b) and Adult (c-d) multiome datasets based on ATAC modality, calculated using the first 100 LSI components. **a,c**, Broad cell type categories (Neurons: blue, Others: red). **b,d**, Log-normalized mRNA expression of *Onecut*, *FoxP*, and *ttk*.


RpL34a
RpS28b
Fmr1
CG11537
eIF5B
CG5644
Pitslre
CG4678
Src64B
pico
Cct5
Drep-4
RpL37a
CG7582
CG17029
Adk1
Rala
RpL24
RpLP0
Fas1
Cpr66D
Hr4
Arr2
Eip74EF
CG12817
CG12643
Gapdh1
RpS11
CG9394
RpS15Aa
mnb
mnd
Pep
Ald
Fie
Sap47
SmD3
CG1695
Scp2
PyK
CR45054
CG3009
Gs2
Vha55
CG31324
I(1)G0334
Hr38
CR43459
Hk
CG15772
CG5079
Smr
CG32278
slim
Vha14-1
Gbs-76A
CG31808
Flo2
Tob
ftz-f1
Glut4EF
CG30118
CG32544
CG11000
Pglym78
noe
fok
TM4SF
Cyt-c-p
CG32267
Gycalpha99B
Men-b
CG7332
pros
norpA
Eip78C
fne
Blimp-1
Nckx30C
Cht5
CR46006
sdk
lov
Cngl
hdc
CG13053
tun
CG42492
nAChRalpha5
Eip75B
CG5151
CG3726
CG9650
CG34383
SK
para
CG13055
CG34247
CG13920
roX2
CG13071
Oatp74D
Hr3
CG15894
Dscam1
slo
wap
Hsp67Bc
CG10737
shakB
RhoGAP100F
CG14312
CG42265
I(1)G0232
FoxP
crp
RpL17
RpS17
RpS9
CG3689
dpr13
pan
br
CHES-1-like
fru
rn
NK7.1
Eip93F
CG34380
TyrR
Oda
Mur2B
Vsx2
app
bab1
ap
rdgA
CG3655
Glut1
bab2
pdm3
dpr8
CG34127
dpr10
scro
bru-3
Vsx1
klg
CG34362
SoxN

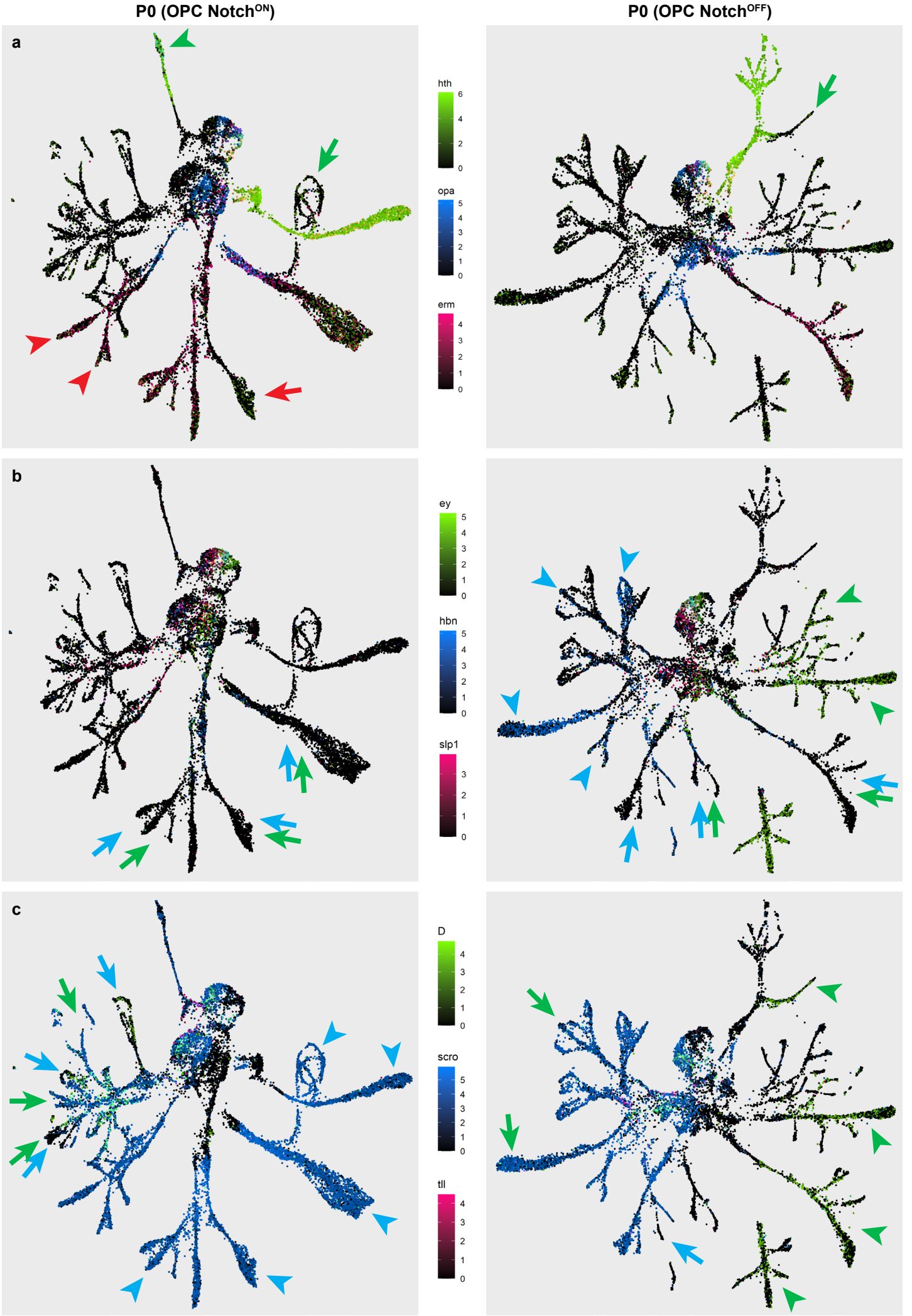
Supplementary Figure 4: Consistently predicted regulatory interactions by SCENIC+

Cytoscape visualization of regulatory interactions predicted by SCENIC+. Shown are TFs present in at least 10 different networks (excluding M1, and non-neuronal metaclusters M38–M42) and the TF–gene pairs found in more than 50% of the networks that contain the respective TF. Red bars indicate repression; green arrows indicate activation.


Terminal selector
Ecdysone-related
Other
Cluster Markers
a
GO Term: neurotransmitter receptor activity
GO Ontology: MF
TfAP-2
klu
Hmx
CG11294
Lim1
CG5953
Lim3
aop
oc
CG11085
tup
erm
Ets65A
ey
pdm3
acj6
svp
kn
D
Eip78C
Vsx2
bab1
fkh
Vsx1
ap
lov
NK7.1
SoxN
Eip93F
jim
fd59A
TF
0
3
6
9
Fold Enrichment
b
GO Term: synaptic target recognition
GO Ontology: BP
Sox102F
Rx
slou
Sox21b
sr
disco
hth
pros
mirr
Lim3
ct
grn
Lin29
Lim1
CG11085
bsh
dimm
CG11294
sv
kn
acj6
erm
tup
Hmx
svp
ey
Dll
pdm3
Vsx1
NK7.1
ap
Vsx2
D
0.0
2.5
5.0
7.5
Fold Enrichment
c
GO Term: axon guidance
GO Ontology: BP
Sox102F
Rx
pros
Sox21b
Hmx
disco
CG11085
Lim1
bsh
hth
sim
slou
sv
EcR
CG32105
Lin29
TF
0
1
2
3
4
Fold Enrichment
d
GO Term: regulation of membrane potential
GO Ontology: BP
CG5953
ey
erm
SoxN
fd59A
pdm3
Hr3
fkh
Mef2
pan
bab1
svp
rn
Eip93F
Vsx2
D
luna
fru
lov
jim
TF
0
1
2
3
4
Fold Enrichment
Stage Markers
e
GO Term: translation
GO Ontology: BP
BtbVII
crol
Med
FoxP
Deaf1
crp
pan
CG32532
slp1
lola
TF
0
2
4
6
Fold Enrichment
f
GO Term: oxidative phosphorylation
GO Ontology: BP
CG4404
ara
shn
pan
CrebA
ci
slp1
CG32532
Hr38
Mef2
grn
Eip75B
rn
scro
Eip74EF
0.0
2.5
5.0
7.5
10.0
12.5
Fold Enrichment

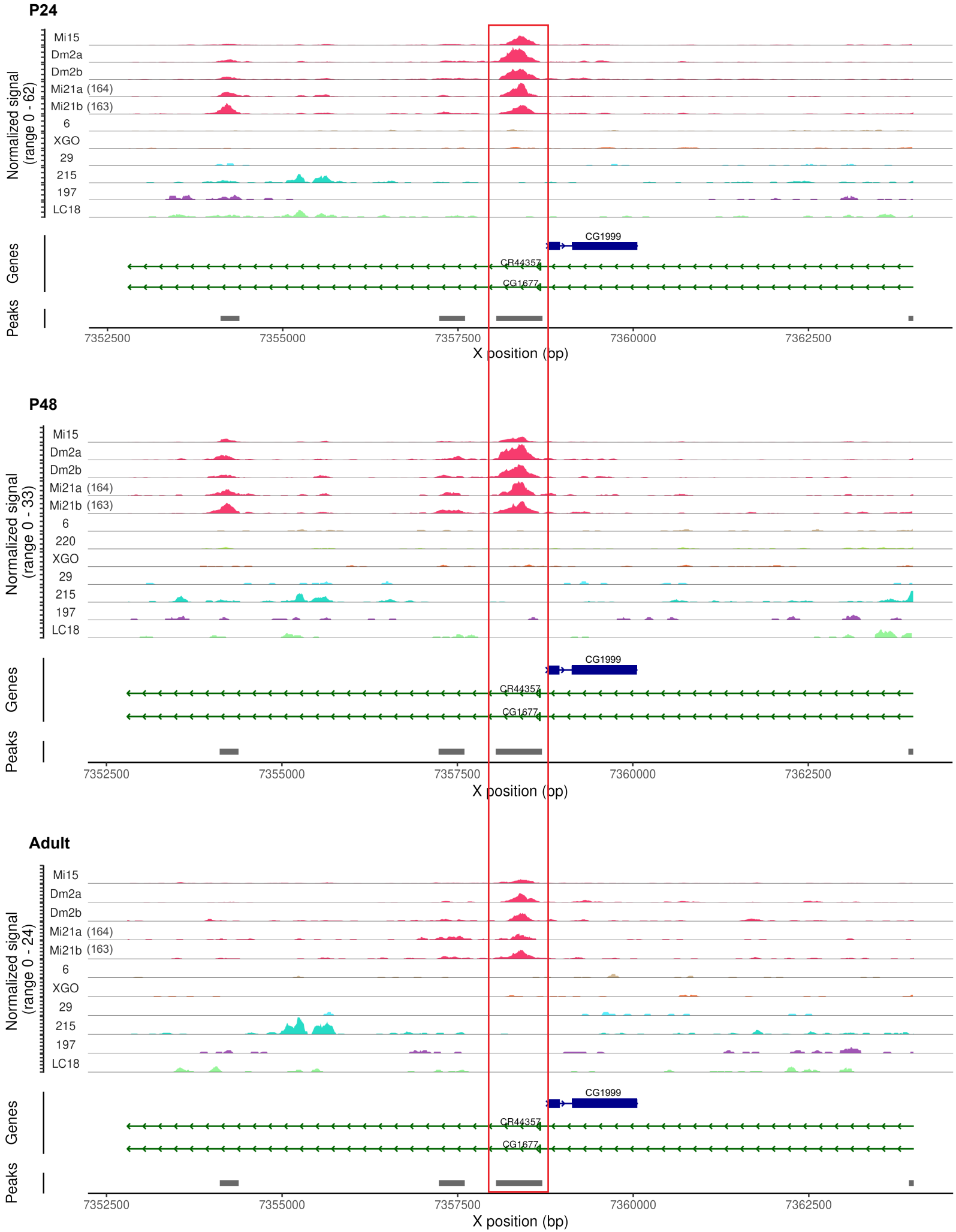
Supplementary Figure 5: GO enrichment analysis of the predicted target genes for each TF in the SCENIC+ networks

a–f, PANTHER GO enrichment terms for target genes of top TFs with at least 50 target genes found across all SCENIC+ networks**.** Terms are grouped based on prior association with either cluster or stage markers as reported in[23]. Each bar shows the fold enrichment ($\log_2$ scale) of a GO term among each TF's predicted target genes. TFs are on the y-axis, and bars are colored by TF classification (Terminal selector, Ecdysone-related, or Other). Only the significant (FDR < 0.05) molecular function (MF) or biological process (BP) terms are shown. Highlighted terms are (**a**) Neurotransmitter receptor activity (MF); (**b**) Synaptic target recognition (BP); (**c**) Axon guidance (BP), (**d**) Regulation of membrane potential (BP), (**e**) Translation (BP), and (**f**) Oxidative phosphorylation (BP).


P0 (OPC Notch^{ON})
P0 (OPC Notch^{OFF})
a
hth
6
4
2
0
opa
5
4
3
2
1
0
erm
4
3
2
1
0
b
ey
5
4
3
2
1
0
hbn
5
4
3
2
1
0
slp1
3
2
1
0
c
D
4
3
2
1
0
scro
5
4
3
2
1
0
tll
4
3
2
1
0

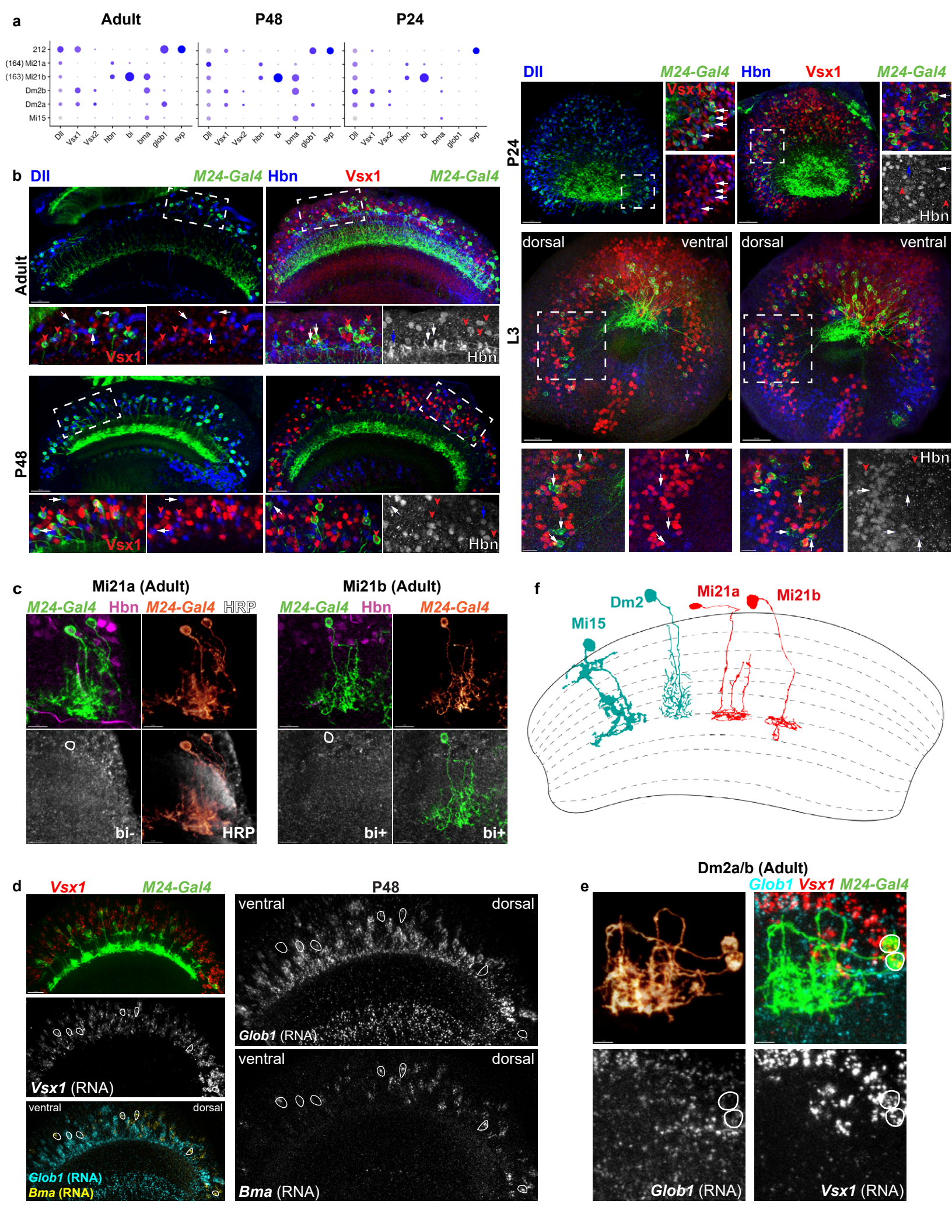
Supplementary Figure 6: Expression of temporal TF genes during OPC neurogenesis

a-c, UMAP reductions of the Notch^{ON} (left) and Notch^{OFF} (right) hemilineages of the OPC, same as in Fig. 4d-e, and colored based on the log-normalized mRNA expression of **a**, *hth* (green), *opa* (blue), and *erm* (red). **b**, *ey* (green), *hbn* (blue), and *slp1* (red). **c**, *Dichaete* (green), *scro* (blue), and *tll* (red). Arrows (color coordinates with the gene legends) indicate neuronal trajectories which do not maintain the expression of the respective tTFs despite originating from these temporal windows. Arrowheads indicate independent activation of these TFs in other lineages.


P24
Normalized signal
(range 0 - 62)
Mi15
Dm2a
Dm2b
Mi21a (164)
Mi21b (163)
6
XGO
29
215
197
LC18
Genes
CG1999
CR44357
CG1677
Peaks
7352500
7355000
7357500
7360000
7362500
X position (bp)
P48
Normalized signal
(range 0 - 33)
Mi15
Dm2a
Dm2b
Mi21a (164)
Mi21b (163)
6
220
XGO
29
215
197
LC18
Genes
CG1999
CR44357
CG1677
Peaks
7352500
7355000
7357500
7360000
7362500
X position (bp)
Adult
Normalized signal
(range 0 - 24)
Mi15
Dm2a
Dm2b
Mi21a (164)
Mi21b (163)
6
XGO
29
215
197
LC18
Genes
CG1999
CR44357
CG1677
Peaks
7352500
7355000
7357500
7360000
7362500
X position (bp)

Supplementary Figure 7: Accessibility of the enhancer region cloned as *M24-Gal4*

Coverage plots of the P24, P48, and Adult multiome data showing normalized accessibility signal in the top 10 cell types with highest accessibility within the indicated region (red rectangle).


a
Adult
P48
P24
212
(164) Mi21a
(163) Mi21b
Dm2b
Dm2a
Mi15
Dll
Vsx1
Vsx2
hbn
bi
bma
glob1
svp
b
Dll
M24-Gal4
Hbn
Vsx1
M24-Gal4
Adult
P48
Vsx1
Hbn
P24
L3
dorsal
ventral
c
Mi21a (Adult)
Mi21b (Adult)
M24-Gal4
Hbn
HRP
bi-
bi+
f
Mi15
Dm2
Mi21a
Mi21b
d
Vsx1
M24-Gal4
P48
ventral
dorsal
Glob1 (RNA)
Vsx1 (RNA)
Bma (RNA)
e
Dm2a/b (Adult)
Glob1
Vsx1
M24-Gal4
Glob1 (RNA)
Vsx1 (RNA)

Supplementary Figure 8: *M24-Gal4* expression during optic lobe development and the identification of Mi21 neurons and Dm2 subtypes

a, mRNA expression of select marker genes for the M24 clusters (Mi21a (cluster 164), Mi21b (cluster 163), Dm2b, Dm2a, and Mi15) and cluster 212 at P24, P48 and Adult stages. Dot size indicates the fraction of cells expressing the gene and color intensity corresponds to log-normalized and scaled expression level.

b, *M24-Gal4* driving *CD4:tdGFP* expression across development (Adult, P48, P24, and L3). At each stage, the first panel (top left) shows anti-GFP (green), anti-Dll (blue), and anti-Vsx1 (red) stainings, with panels of higher magnification for regions marked by dashed white boxes below (Adult, P48, L3) or to the right (P24) showing only Dll expression (white arrows) or Dll/Vsx1 co-expression (red arrows) among the GFP labeled cells. The second panel (top right) shows anti-GFP (green) anti-Hbn (blue) and anti-Vsx1 (red) stainings with panels of higher magnification for regions marked by dashed white boxes below (Adult, P48, L3) or to the right (P24) showing a mixture of Vsx1+Hbn- (red arrows), Vsx1-Hbn+ (blue arrows), and Vsx1-Hbn- (white arrows) cells. Magnified regions of the L3 brains are in the dorsal Optix domain. Representative images are shown as maximum intensity projections. Adult n = 5/4, P48, n = 2/5, P24, n = 3/4, and L3, n = 3/5 anti-Dll brains/anti-Hbn brains.

c, *M24-Gal4* driving *CD4:tdGFP* (FLP out) showing sparsely labeled neurons that co-express Dll and Hbn. Based on their morphology, we named these cell-types Mi21a and b that are distinguished by bifid (bi) expression. Anti-Hbn (magenta), anti-bi (white), anti-GFP (green) are shown as maximum intensity projections, while anti-GFP (glow) and anti-HRP (white, neuropil) are shown as 3D reconstructions of the same neuron. Adult Mi21a neurons just touch the serpentine layer of the medulla, do not express bifid, and match FlyWire Sm01 morphology as shown in **f**. Adult Mi21b neurons are the only M24 cell type crossing the serpentine layer, express bifid, and match FlyWire Sm02 morphology as shown in **f**. n = 15/9 optic lobes/brains.

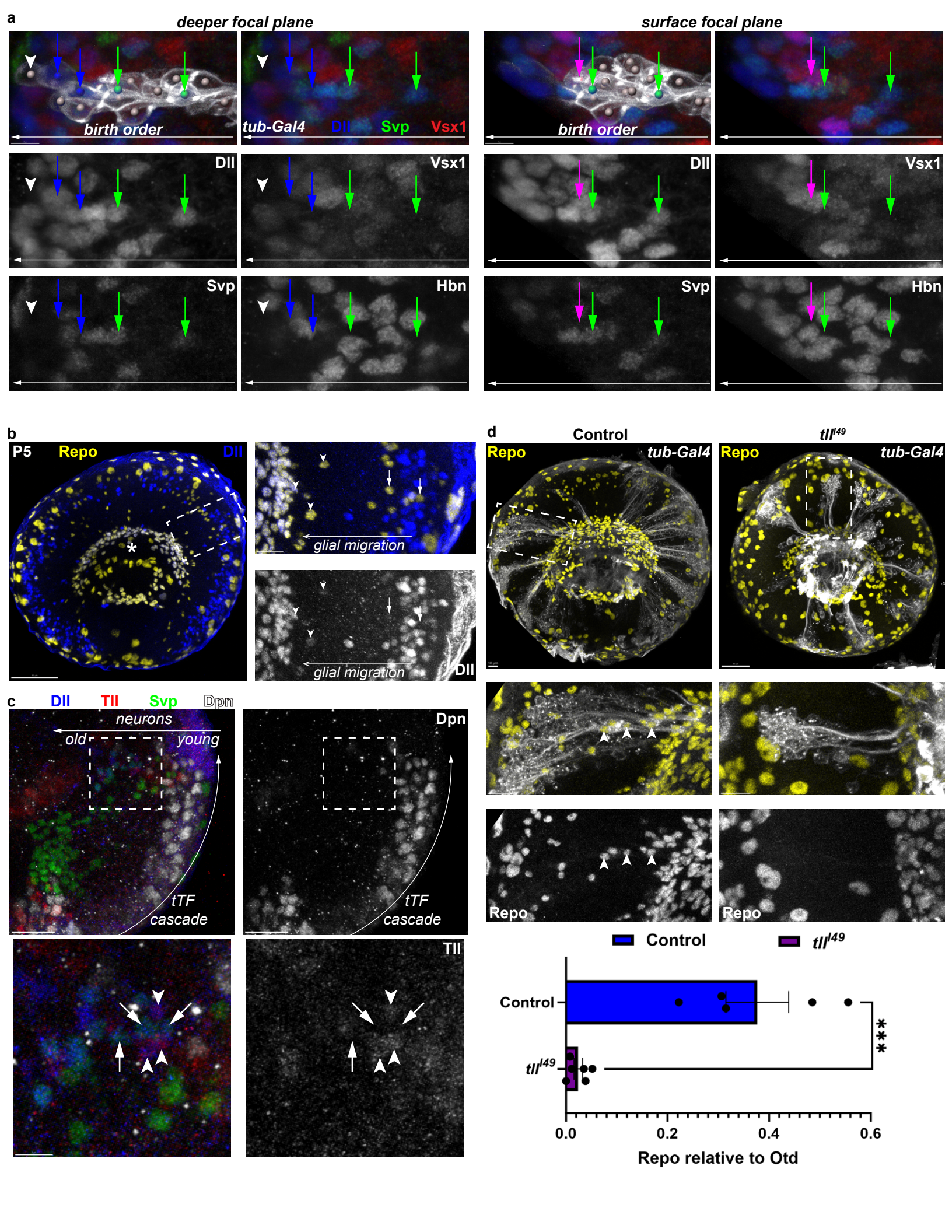
d, HCR *in situ* hybridization for indicated mRNAs was used to identify Dm2 subtypes at P48. *M24-Gal4* drives *CD4:tdGFP* (FLP out) with *Vsx1* (red) and endogenous GFP (green) marking Dm2s at P48. Visualized as maximum intensity projections, cells that express Vsx1 were circled in white with *Glob1* (blue) and *Bma* (yellow) shown on dorsoventral axis. The same sample is shown to the right with single channels images for both *Glob1* and *Bma* with the ventral medulla on the left and dorsal medulla on the right. n = 5/3, optic lobes/brains.

e, Sparsely labeled Dm2 subtypes in adult brains were visualized using *M24-Gal4* driving *CD4:tdGFP* (FLP out) with endogenous GFP signal (green). HCR *in situ* hybridization against

Vsx1 (red) and *Glob1* (blue) is shown as maximum intensity projections of a representative image. Two Dm2 neurons with differential *Glob1* expression are shown as 3D reconstructions of the endogenous GFP signal (top left). Single channel images for *Glob1* and *Vsx1* are shown below with cell bodies circled in white. n = 4/3, optic lobes/brains.

f, Cross section of the medulla neuropil showing cartoon representations of previously established M24 cell types, Dm2 and Mi15 (blue) and the newly identified cell types, Mi21a and Mi21b (red).

Scale bars: **b**, Adult, 15 and 4 μm, P48, 20 and 5 μm, P24, 30 and 7 μm, L3, 25 and 8 μm. **c**, 6 (left) and 7 μm. **d**, 13 μm. **e**, 3 μm.


a
deeper focal plane
surface focal plane
birth order
tub-Gal4
Dll
Svp
Vsx1
birth order
Dll
Vsx1
Svp
Hbn
Dll
Vsx1
Svp
Hbn
b
P5
Repo
Dll
glial migration
glial migration
Dll
c
Dll
Tll
Svp
Dpn
neurons
old
young
tTF
cascade
Dpn
tTF
cascade
Tll
d
Control
tll49
Repo
tub-Gal4
Repo
tub-Gal4
Repo
Repo
Control
tll49
Control
tll49
***
0.0
0.2
0.4
0.6
Repo relative to Otd

Supplementary Figure 9: Birth order of Distal-less^{+} cells in the OPC and Tailless mutants

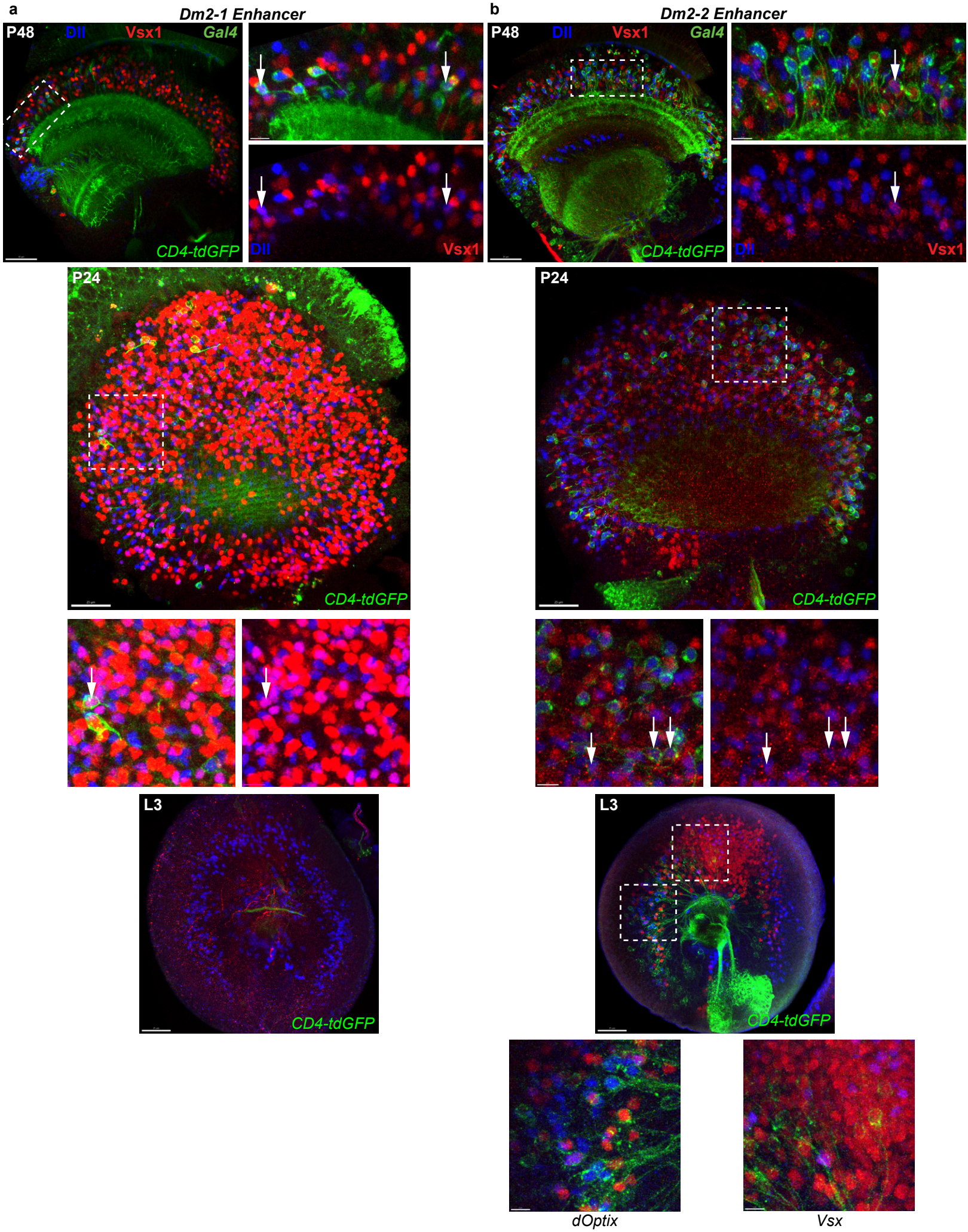
a, The same clone shown in Fig. 5c. For better visualization, it was split into two focal planes with Dll+Svp+ neurons common to both panels and Dll+Vsx1+Svp- (Dm2) and Dll+Vsx1-Svp- (Mi15) neurons in separate panels. FRT82B MARCM clones labeled with *tubulin-Gal4* driving *myr:GFP* in early pupal (~P5-P10) optic lobes, approximately 50 hours after heat shock. Clones are located within the dorsal Optix domain. Neurons are born from sequential temporal windows from right to left marked by the white arrow. All cells marked with arrows express Dll with cells indicated by green arrows also expressing Svp and the magenta arrow also expressing Vsx1. Anti-GFP (white), anti-Dll (blue), anti-Vsx1 (red), and anti-Svp (green) with single panels for each TF including anti-Hbn (white). n = 4 brains.

b, A wild-type optic lobe during neurogenesis (P5) that shows the glia originating from OPC NBs turn on Repo after they have begun their migration (white arrows) to the medulla neuropil (white asterisk). Dll is expression is stochastic and low within these glia as they migrate and increases when they reach the neuropil (white arrowheads). Anti-Repo (yellow) and anti-Dll (blue or white) are shown as maximum intensity projections. n = 6 brains.

c, A wild-type optic lobe during neurogenesis (L3) shows that Tll window neuroblasts produce neurons that maintain Tll expression for a short period of time. From the bottom right to the top right, NBs (Dpn+) continuously divide and generate neurons (curved white arrow). From the NBs on the right to the left (straight white arrow), neurons are produced with age increasing along this axis. The two panels below show magnified views of the area indicated by the dashed white box, where arrowheads mark Dll+Svp-Tll+ (M24) neurons and arrows mark Dll+Svp+Tll- neurons. Anti-Dll (blue), anti-Tll (red), anti-Svp (green), and anti-Dpn (white) are shown as maximum intensity projections. n = 4 brains.

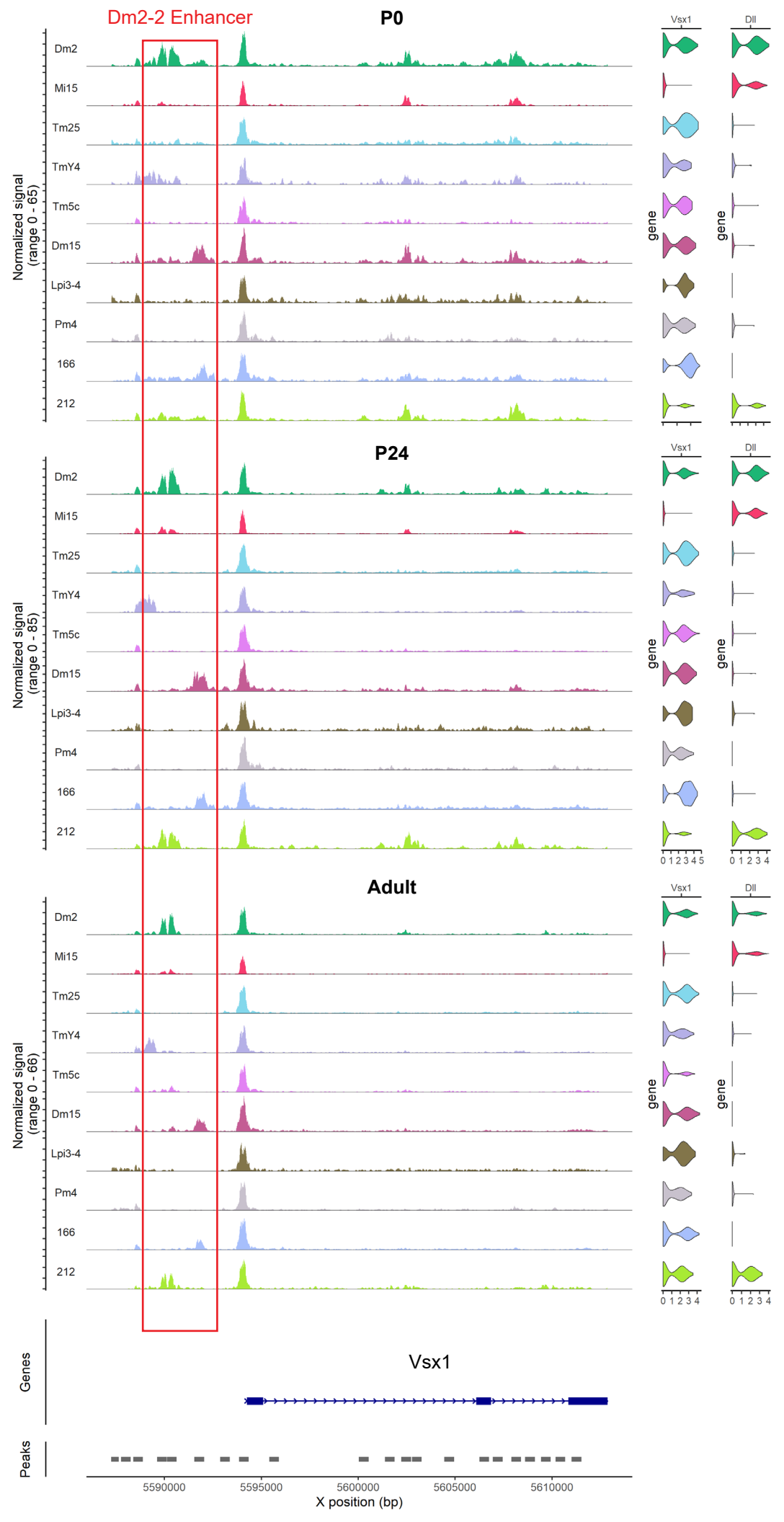
d, Identical experiment as in Fig. 5d, FRT82B (control) and FRT82B *tll^{l49}* MARCM clones are labeled with *tubulin-Gal4* driving *CD4:tdGFP* during neurogenesis (P5). 1-2 clones with arrowheads marking migrating medulla neuropil glia that express Repo are shown per genotype. Anti-Repo (yellow or white) and anti-GFP (white) are shown as maximum intensity projections. Bar plot: Quantification of d. Ratios were quantified relative to the number of Otd+ neurons in clones (*Methods*). Error bars represent SEM and each dot is a singular optic lobe with all clones summed. n = 5/5/26 control and n = 6/6/24 optic lobes/unique brains/clones. In comparison to the control, p = 0.00015 for Repo+ glia.

Scale Bars: **a**, 4 µm. **b**, 45 and 10 µm. **c**, 17 and 5 µm. **d**, 10 and 12 µm.


a
Dm2-1 Enhancer
P48
Dll
Vsx1
Gal4
CD4-tdGFP
Dll
Vsx1
P24
CD4-tdGFP
L3
CD4-tdGFP
b
Dm2-2 Enhancer
P48
Dll
Vsx1
Gal4
CD4-tdGFP
Dll
Vsx1
P24
CD4-tdGFP
L3
CD4-tdGFP
dOptix
Vsx

Supplementary Figure 10: Dm2-1 and Dm2-2 enhancer reporter activity during optic lobe development

Dm2-1 (**a**) and Dm2-2 (**b**) enhancers were cloned into a Gal4 reporter to drive *UAS-CD4:tdGFP* expression at stages P48, P24 and L3. Figure 6 shows the adult stage with the same *CD4:tdGFP* and L3 with *nls-GFP*. Anti-GFP (green), anti-Dll (blue), and anti-Vsx1 (red) are shown as maximum intensity projections. White arrows point to Dll/Vsx1 co-expression. P48, n = 4/2, P24, n = 4/4, and L3, n = 3/5 Dm2-1 brains/Dm2-2 brains.


Dm2-2 Enhancer
P0
P24
Adult
Dm2
Mi15
Tm25
TmY4
Tm5c
Dm15
Lpi3-4
Pm4
166
212
Normalized signal (range 0 - 65)
Normalized signal (range 0 - 85)
Normalized signal (range 0 - 66)
Vsx1
Dll
gene
0 1 2 3 4 5
0 1 2 3 4
Genes
Vsx1
Peaks
5590000
5595000
5600000
5605000
5610000
X position (bp)

Supplementary Figure 11: Dm2-2 enhancer accessibility in select cell types across development

Coverage plots of the P0, P24, and Adult multiome data show normalized accessibility of the ~3 kb Dm2-2 enhancer region in Dm2 and Mi15 neurons and additional Vsx1/2-expressing cell-types. At P0, the enhancer is broadly open in Dm2, while later the accessibility becomes restricted to distinct ~1 kb subregions specific to each lineage (e.g., upstream in TmY4, downstream in Dm15/166, overlapping with Dm2 in cluster 212). Violin plots (right) display log-normalized mRNA expression of *Vsx1* and *Dll* in these clusters.

Supplementary Table 1: Cluster Annotations

Summary table of all 259 clusters, including previous and updated cluster identities, cell type annotations, metacluster assignments, temporal and spatial origins, Notch status, and any relevant comments.

Supplementary Table 2: Cluster markers (RNA modality) at P24, P48 and Adult stages

Table of RNA gene markers enriched in each cluster at P24, P48, and Adult stages. Markers were identified using Seurat's *FindAllMarkers* on stage-specific subsets of the Multiome dataset, including only positively enriched genes (log2FC > 1, adjusted p < 0.01) and excluding non-neuronal clusters. See *Methods* for more details.

Supplementary Table 3: Cluster markers (ATAC modality) at P24, P48 and Adult stages

Table of ATAC peak markers enriched in each cluster at P24, P48, and Adult stages. Markers were identified using Seurat's *FindAllMarkers* on stage-specific subsets of the Multiome dataset, including only positively enriched genes (log2FC > 1, adjusted p < 0.01) and excluding non-neuronal clusters. See *Methods* for more details.

Supplementary Table 4: Gene-enhancer links at P0, P24, P48, and Adult stages

Table of gene–enhancer links identified across developmental stages, showing peak associations identified by correlation of gene expression with nearby accessible regions and highlighting the dynamic usage of enhancers across P0(OPCprogenitor), P0(OPCneuronal), P0 (all), P24, P48, and Adult stages, as well as all stages combined. Links were identified using Seurat's *LinkPeaks* function, with a search window of 50kb upstream and downstream of the TSS. See *Methods* for more details.

Supplementary Table 5: Motif enrichment analysis of neuronal enhancers

Table summarizing the motifs enriched in consistent and stage-specific neuronal enhancer peaks. *De novo* motif discovery was performed with the MEME suite's STREME tool using sequences derived from DARs identified in **Supplementary Table 3**, and candidate regulators were assigned by comparing discovered motifs against the CisBP-*Drosophila* database with Tomtom. See *Methods* for more details.

Supplementary Table 6: SCENIC+ networks

Table containing 28 metacluster-specific gene regulatory networks (GRNs) inferred with SCENIC+ using multiome and supporting scRNA-seq datasets from P24, P48, and Adult stages

(tabs). Some related metaclusters were merged prior to network inference to improve interpretability and statistical power. See *Methods* for more details.

Supplementary Table 7: Predicted transcription factor binding sites in the Dm2-2 enhancer

FIMO analysis of the Dm2-2 enhancer (~1 kb subregion) identifying putative transcription factor binding sites. Each row reports the motif, predicted TF, strand, start and end coordinates within the enhancer, score, p-value, q-value, and matching sequence. These results highlight potential regulators, including a BarH1-like homeodomain site (highlighted green).

Supplementary Table 8: Genotypes of experimental strains

Supplementary Table 9: Key Resources Table

Methods

Single-cell multiome sequencing of whole optic lobes

In order to perform simultaneous RNA and ATAC sequencing on very large numbers of developing optic lobe cells, we followed a genetic multiplexing strategy[43] based on the wild-type strains of the Drosophila Genetic Reference Panel[83] (DGRP). We crossed males from 12 different DGRP lines (40, 189, 320, 382, 383, 391, 441, 406, 492, 505, 852 and 897) to Canton S virgins to acquire F1 hybrids. For each of the 4 stages we sequenced (P0, P24, P48 and Adult) we dissected a single female fly from each of these crosses, which were then processed together for all subsequent steps. Natural genetic variation among the DGRP strains[84] enables post hoc assignment of each sequenced cell to its origin strain based on its unique SNPs. This approach provides two chief advantages: i) extremely robust identification of doublets, allowing us to 'overload' each microfluidic well with more than twice the number of recommended cells, reducing per cell library preparation costs; ii) preservation of biological replicates, i.e. for all analyses that are performed at the cluster level rather than single cells, cells belonging to the same DGRP line at each stage (representing a single fly) can be 'pseudobulked' separately, increasing statistical power.

For nuclei isolation, we prepared Wash Buffer to include: 10 mM Tris-HCl (pH 7.4), 10 mM NaCl, 3 mM $MgCl_2$, 0.1% Tween-20, 1% bovine serum albumin (BSA), 1 mM dithiothreitol (DTT) and 1 U/µl of Protector RNase inhibitor (Sigma) in nuclease-free water. Nuclei lysis buffer was prepared to include all ingredients of the wash buffer, with the addition of 0.1% Nonidet P40 substitute and 0.01% Digitonin. 20x Nuclei buffer (10x Genomics) was diluted to 1x in nuclease-free water and supplanted with 1 mM DTT and 1 U/µl RNase inhibitor. All solutions were maintained on ice throughout the protocol.

For each experiment, we selected female white pupae (P0) and reared these to the required stage at 25°C. For the adult dissections, we only used 0-day old (<24h after eclosion) flies. The optic lobes were dissected and maintained in ice-cold Schneider's media until lysis. They were washed 3 times with DPBS before being transferred into a dounce homogenizer (pre-wet pestles with wash buffer) containing 500 µl lysis buffer. The optic lobes in lysis buffer were incubated on ice for 5 minutes, homogenized with 25 strokes with pestle A, followed by another 5-minute incubation on ice, and 30 strokes with pestle B. Without fully removing the pestle B from the homogenizer, both the pestle and the side walls of the homogenizer were then rinsed with 1 mL wash buffer. The entire volume was transferred into a pre-cooled 2 mL Biopur tube. We also recommend mixing the lysis solution using a wide-bore tip in the middle of both incubation steps.

Diluted lysis solution was filtered with 40 µm Flowmi cell strainers (in two batches) directly onto a 20 µm PluriStrainer (pre-wet with wash buffer) placed on a second pre-cooled 2 ml tube. After filtering, nuclei were centrifuged (at 4°C) for 5 minutes at 800 rcf using a swing bucket rotor. Excess buffer was discarded without disrupting the pellet, another 1 mL wash buffer was added. After pipette mixing to wash the nuclei, they were pelleted again by another 5 minutes of centrifugation at 800 rcf; excess buffer was discarded, leaving no more than 10 µl of wash buffer. It is critical to minimize the amount of buffer left in the tube at this step, as the BSA in lysis and wash buffers may inhibit subsequent Tn5 transposition. Finally, the pellet was resuspended in diluted nuclei buffer (minimum 50 µl, 100 µl is recommended if the amount of input material allows for sufficient concentration) and were filtered through a 10 µm mini-PluriStrainer (pre-wet with nuclei buffer) into a 1.5 mL DNA LoBind tube. In order to assess lysis efficiency and to determine the nuclei concentration we stained a small aliquot with Calcein AM and EthD-III (Live/Dead assay) and examined/counted them in a hemocytometer using an epifluorescence microscope.

Next, we used Chromium Next GEM Single Cell Multiome ATAC + Gene Expression kits (10x Genomics) and followed the manufacturer's instructions for producing ATAC and gene expression (RNA) libraries for sequencing. Briefly, the nuclei were subjected to Tn5 transposition for 1 hour at 37°C and were then immediately loaded into a Chip J for generation of nanoliter-scale gel bead-in-emulsions (GEMs) using Chromium Controller. Following reverse-transcription, cleanup and pre-amplification PCR, the samples were split into two for the ATAC library PCR (9 cycles) and for cDNA amplification (7 cycles) followed by fragmentation, adaptor ligation and sample index PCR for the RNA library (11 cycles). We used 20 µl of the amplified cDNA for RNA library preparation (even though 10 µl is normally recommended), and the final resuspension of the ATAC libraries was done using 40 µl of EB buffer, even though 20 µl is normally recommended. We produced 5 different multiomic libraries (technical replicates) for the P48 stage, and 4 libraries for other stages (P0, P24 and Adult), aiming for ~20,000 cells per library.

The libraries were subjected to paired-end Illumina sequencing using NovaSeq 6000: 50x8x24x49 cycles for ATAC libraries and 28x10x10x90 cycles for RNA libraries. We sequenced all RNA libraries to 80-90% saturation (as reported by 'Percent duplicates' in CellRanger), and all ATAC libraries to 75-80% saturation.

Processing and quality control of the multiome data

The multiome data was processed using CellRanger-ARC v2.0.2. After library demultiplexing (mkfastq), the reads were aligned and counted using the *D. melanogaster* genome annotation version BDGP6.88 (Ensembl). The ATAC reads were initially counted based on a predefined set of 128,464 genomic ranges[85] (ctx) which cover 80%, i.e. nearly all non-coding regions, of the *Drosophila* genome. Multimodal cell-calling algorithm of CellRanger-ARC failed on our data, likely due to high background in the ATAC modality. For each library, we thus also ran (base) CellRanger v7 on the RNA modality alone, which identified 16-20,000 barcodes corresponding to cells per library (before filtering) for stages P0, P48 and Adult, and 23-25,000 cells per library for stage P24.

Gene (RNA) and peak (ATAC) count matrices for the cell barcodes were loaded into R and further processed using the packages Seurat[41] v4.1.4 and Signac[86] v1.11.0. We first filtered out any barcodes with less than 300 (ATAC) fragments in peak regions (ctx), or less than 200 genes detected (RNA). Additionally, we removed the barcodes with the proportion of mitochondrial (RNA) reads exceeding 20% (P0 and P24), 30% (P48) or 40% (Adult). We observed that our data, and especially later stages, contain unusually large amounts of mitochondrial RNA (median: 12% in Adult cells). While this is typically indicative of low transcriptome quality, we also observed that at relatively lower regimes (up to ~20%) mitochondrial RNA proportion in our data in fact correlates positively with the total RNA counts (UMIs). After filtering, we removed all mitochondrial genes from the gene/cell matrices to prevent these reads from skewing the depth normalization of single cells. Ribosomal RNA content was also unusually high in these data, which is not directly detectable in Seurat but causes low rates (55-65%) of “Reads mapped confidently to transcriptome” as reported by CellRanger (due to rRNA gene duplications). We concluded that these issues, other than reducing the number of usable reads, do not affect any of the downstream analyses. While the underlying reasons remain unclear, we speculate that incomplete lysis of plasma membrane in some cells may have trapped ribosomes and mitochondria together with the nuclei we sequenced. We additionally removed the barcodes with Nucleosome Signal (the ratio of mononucleosomic fragments to nucleosome-free fragments) of more than 1, or with TSS (transcription start site) Enrichment score of less than 1.5 in the ATAC modality.

We next used demuxlet[87] to identify and remove doublets based on DGRP SNPs detected in each cell (RNA modality). To assign individual cells from pooled F1 crosses to their DGRP parental backgrounds, we adapted a previous approach using popscle with helper scripts from

popscle_helper_tools[1]. Since we profiled F1 hybrids from DGRP × Canton-S crosses to reduce possible genotype-specific effects, the heterozygous samples required us to focus exclusively on DGRP-specific variants for genotype assignment. First, we lifted over the DGRP2 variant database[2,3] to the dm6 reference genome using LiftoverVcf provided by Picard v4.0.2.0 with the chain file provided by UCSC Genome Browser:

https://hgdownload.soe.ucsc.edu/goldenPath/dm3/liftOver/dm3ToDm6.over.chain.gz

We identified informative SNPs by filtering DGRP variants using custom functions from popscle_helper_tools. We retained only positions that were: (1) homozygous for the alternative allele in at most 2 of the selected DGRP lines, (2) not missing genotype information in any line, (3) not homozygous reference in all lines, and (4) not heterozygous in any line. VCF processing was performed using bcftools v1.14 and vcftools v0.1.16.

To exclude Canton-S-specific variants from our analysis, we generated a Canton-S variant mask by calling variants from a published Canton-S optic lobe developmental single-cell transcriptomic atlas[9]. We performed variant calling using bcftools mpileup and bcftools call v1.14 on the aggregated BAM files from this atlas against the dm6 reference genome. We then masked all SNP positions overlapping with these Canton-S variants using bedtools v2.29.2, as these positions would be uninformative for distinguishing between F1s from different DGRP backgrounds.

To ensure reliable genotype assignment, we generated coverage tracks using deepTools v3.5.0 and filtered genomic regions requiring a minimum read depth of 10. To minimize the impact of ambient RNA contamination on genotype assignment, we quantified gene expression using HTSeq v0.13.5 and identified the top 100 genes by expression level. These highly expressed regions were excluded from SNP calling using the bedtools *complement* and *intersect* operations under the assumption that highly expressed transcripts are more likely to contribute to the ambient RNA and carry mixed genotype signals. Cell barcodes passing quality control were used to filter the BAM files using samtools v1.14 and popscle_helper_tools scripts, retaining only reads from valid cells. The filtered and sorted BAM files were processed through popscle dsc-pileup to generate allele count matrices at informative SNP positions. For demultiplexing, we converted homozygous genotypes in the DGRP database (1/1) to heterozygous calls (1/0) in the VCF file to reflect the expected F1 genotypes. Finally, we performed demultiplexing using popscle demuxlet with the genotype-aware mode, testing multiple alpha parameters (0 and 0.5) to optimize the assignment of cells to their DGRP parental backgrounds while identifying potential doublets. The doublet rates ranged from ~15% (P48) to ~24% (other stages) per library.

A low number of cells were misclassified to DGRP line_307 and line_508. When using the DGRP assignments for downstream analysis (e.g. creation of metacells, outlined below), these two assignments (and the cells assigned to them) are dropped from the analysis.

Using Seurat v4.1.4 and Signac v1.10.0, barcodes representing singlets and passing all filters have been merged into a single Seurat (.rds) object for each stage. ATAC data was normalized with *RunTFIDF* using default parameters. After feature selection with *FindTopFeatures*, with default parameters, we used latent semantic indexing (LSI) to acquire 150 reduced dimensions using the function *RunSVD*. As described in the Signac tutorial "Analyzing PBMC scATAC-seq" (https://stuartlab.org/signac/articles/pbmc_vignette.html), we observed that the first LSI dimension always correlated very strongly with sequencing depth and thus was excluded from all further analyses that use the LSI reduction. For the RNA data, after selection of top 2000 variable genes using *FindVariableFeatures*, we applied both standard log normalization using *LogNormalize* as well as *SCTransform*, a GLM-based method that was proposed to provide better variance stabilization especially for highly expressed genes[88]. We used the SCTransformed data for dimensional reduction of the RNA modality using *RunPCA*, (npcs=150) and all downstream steps that utilize the PCA reduction (below). Whereas for plotting gene expression values or calculating differentially expressed genes between clusters, we used the log-normalized expression values.

Classification of the multiome data

We independently took supervised and unsupervised approaches to assign multiomic barcodes to specific cell types. The supervised classification was performed using an artificial neural network model that we had trained on the cluster identities of our previously published scRNA-seq dataset[23]. This is a multi-task network that accepts expression data from different stages through independent input layers based on the top marker genes at each stage. We thus used the P15 markers for classifying the P0 data, P30 markers for the P24 data, P50 markers for the P48 data, and the Adult markers for the Adult multiome data. These models use the log-normalized expression values of the corresponding marker genes to automatically assign each cell to one of the 206 clusters in our published atlas.

For unsupervised classifications (clustering), we used the Weighted Nearest Neighbor (WNN) algorithm[41] from Seurat v4.1.4, which uses both PCA and LSI reductions to build a joint nearest neighbors graph taking advantage of both modalities using *FindMultiModalNeighbors*, (reduction.list=list("pca", "lsi"), dims.list = list(1:150, 2:150)). These neighbor graphs were used both for calculating UMAP reductions with *RunUMAP* (nn.name = "weighted.nn") for visualization

(**Fig. 1**), and for unsupervised clustering of each stage using *FindClusters (*graph.name="wsnn", resolution=7). The highest number of clusters, as expected, were achieved at the P48 stage. Due to the continuous trajectory structure of the P0 dataset (**Figs. 1d, 4**), clustering results are only useful for sub-setting the data into specific trajectories for further analysis. Thus, at this stage, we intentionally used a lower clustering resolution of 2.

We performed a comprehensive comparison of the clustering results at P48 and the classifications of the same cells by the P50 neural network. WNN clustering identified subgroups within the following clusters of our published atlas[23]: 3 (Photoreceptors), 20, 21, 23, 24, 28, 39, 57 (LC25), 58, 65, 67, 79 (LC10b), 83, 87 (LC20), 91 (Mt8), 100 (LC16), 101, 128 (Dm2), 129 (Dm21), 136 (Dm11), 147 (LC4), 149 (LC22/LPLC4), 158 (Tm3b/Tm23), 159 (TmY15/17), 162, 170, 178, 184, 222 (LC6), 223 (TEv). 2-3 subclusters were identified in each of these clusters, with the exception of cluster 24 (6 subclusters) and cluster 100 (5 subclusters), which we have already shown to contain LC15,16,24,25 neurons[25] (and one more, unknown, neuron). By examining the differentially expressed genes between the subclusters, we validated that all of these splits represented biologically significant (i.e. not technical) subgroups. Given the clearly superior cell-type resolution of these multiome data over our published scRNA-seq data[23], we decided to abandon our previous classifications altogether and build a new reference based on the multiome clusters. In cases where these clusters were already annotated to specific cell-types, these annotations were re-evaluated based on the markers of each subcluster and accordingly adjusted. WNN clustering in addition identified subgroups within the (old) clusters 62 (Tm6), 85, 102, 125 (Tm4) and 181. Divisions within the Tm4 and Tm6 clusters represented the dorsoventral subtypes we previously described for these cell-types[23]. However, the distinctions between these subtypes are very subtle and hard to resolve, especially in stages other than P48; we thus decided to keep Tm4 and Tm6 as single clusters. Clusters 85 and 102 were highly heterogeneous mixtures of central brain derived neurons; we thus merged all multiome clusters corresponding to these (old) clusters into one cluster, 141 (**Suppl. Table 1**). Similarly, the division of old cluster 181, while likely meaningful, proved unreliable and was thus merged back into a single cluster (195). Lastly, through our previous (unpublished) sub-clustering attempts, we have been aware that (old) clusters 59, 95 and 96 also contained subgroups, although these were not resolved by the WNN clustering of the entire P48 dataset. We manually split these clusters (into two) by individually sub-clustering them.

Using the (corrected) clusters of the P48 multiome data as reference, we next classified the P24 and Adult datasets. We implemented a label transfer procedure to take advantage of both modalities for this purpose. First, we determined the differentially expressed genes using the

FindAllMarkers function (assay='RNA', only.pos= TRUE, logfc.threshold= 0.5, max.cells.per.ident= 500) and differentially accessible regions (assay= 'ATAC', only.pos= TRUE, test.use= 'LR', latent.vars= 'atac_peak_region_fragments', logfc.threshold= 0.5, min.pct= 0.05, max.cells.per.ident= 500) for all clusters at P48. For the RNA modality, we kept only the 'top100' (by logFC) markers for each cluster. Second, we set the variable features of the ATAC assay to the differentially accessible regions and recalculated the LSI reduction using *RunSVD* and the weighted nearest neighbors function from Seurat *FindMultiModalNeighbors* (reduction.list= list("pca", "lsi"), dims.list= list(1:150, 2:150), k.nn= 31) based on this. We then transferred the labels from the P48 dataset to the P24 and Adult datasets based first on the ATAC modality, using the functions *FindTransferAnchors* (reference.reduction= "lsi", reference.neighbors= "weighted.nn", reduction= "lsiproject", dims=2:150) and then *TransferData* (weight.reduction= "lsiproject"). Independently, we also transferred labels based on the RNA modality: *FindTransferAnchors* (reference.assay= "SCT", query.assay= "SCT", reduction= "cca", reference.neighbors= "weighted.nn", features= top100, normalization.method= "SCT", dims= 1:150) and *TransferData* (weight.reduction= "cca"). The predictions based on both modalities were largely in agreement; we resolved any inconsistencies based on the prediction score (i.e. confidence) calculated by Seurat: if for a given cell different identities were predicted by RNA and ATAC label transfer, the prediction with the higher score was kept (in case of ties, ATAC-based predictions were kept). Finally, we compared the transferred labels to results of the unsupervised clustering we performed at Adult and P24 stages and validated that no additional subdivisions were present. The same procedure was then used for transferring labels from the P24 to P0 dataset. However, since the P0 data contains progenitor cells that are not present in the later stages, using known markers, we separated out the cells corresponding to neuroepithelia (NE), neuroblasts (NB), ganglion mother cells (GMCs) of OPC and IPC, as well as the lamina precursor cells (LPC) based on the unsupervised clustering we performed at P0 (**Fig. 4**).

Multi-task neural network classifier training and prediction

We trained stage-specific neural network classifiers (Adult, P24, and P48) on log-normalized gene expression profiles, with cluster identities encoded using one-hot encoding. Each model consisted of an input layer (dimension equal to the number of marker genes, top 10 by cluster, per stage), a dropout layer (20%), a dense hidden layer (200 units, ReLU activation, L2 regularization), an additional dropout layer (50%), and a softmax output layer predicting 259 cluster labels. Models were optimized using stochastic gradient descent with Nesterov momentum (learning rate = 0.05, momentum = 0.9), categorical cross-entropy loss, and

balanced class weights. Learning rates were reduced stepwise every three epochs. Early stopping and model checkpointing were applied to prevent overfitting.

To encourage cross-stage generalization, models were trained sequentially by passing weights across stages in the order Adult → P24 → P48 and back to Adult, repeating for 14 runs of 30 epochs each. Training and validation splits were predefined for each dataset (Adult: 42,648; P24: 58,390; P48: 54,368 training samples). Performance was monitored at each stage, and the best-performing models (based on lowest validation loss and highest accuracy) were retained for downstream analysis.

For prediction on new datasets, expression matrices were normalized by subtracting the stage-specific mean expression values used during training. Predictions were generated using the trained stage-appropriate model. To estimate confidence, we applied Monte Carlo dropout during inference: 500 stochastic forward passes were performed with dropout layers active, and predicted class labels were recorded for each pass. Confidence was calculated as the proportion of bootstrap predictions consistent with the model's maximum-probability class assignment. Final outputs included the predicted cluster label, softmax score, and bootstrap-derived confidence score for each cell.

The trained classifiers enable consistent reclassification of previously published scRNA-seq datasets of the optic lobe in alignment with our clustering. All trained models, encoders, example scripts, and prediction utilities are available at:

https://github.com/ktreese03/single_cell_classifier

Identification of cluster markers in RNA and ATAC modalities

To define stage-specific and consistent markers of neuronal clusters, we used the merged multiome dataset after removing 3 clusters of TE neurons (absent in adults), cluster 141 (heterogeneous), and all non-neuronal populations. For each developmental stage (P24, P48, and Adult), we identified differentially expressed genes (DEGs) using Seurat's v4.1.4 *FindAllMarkers* function on the RNA assay with stringent thresholds ($\log_2$FC > 1, adjusted p < 0.01, one-vs-all comparisons, maximum 500 cells per cluster). Genes were then categorized as stage-specific, shared between two stages, or common to all three. To identify stable markers, we further required that a gene be captured in the same cluster across stages, allowing ≤2 stage-specific absences to account for sparsity.

Similarly, differentially accessible regions (DARs) were identified using the ATAC assay with *FindAllMarkers* (logistic regression test, latent variable = total ATAC counts, $\log_2$FC > 1, adjusted

$p < 0.01$, min.pct = 0.05, maximum 500 cells per cluster). The same filtering strategy was applied to remove low-quality clusters, and DARs were grouped as stage-specific, shared, or consistent across development. This approach yielded 2,256 neuronal DEGs and 53,637 DARs across P24, P48, and Adult, of which ~30% of DEGs and ~40% of DARs were consistent markers of at least one neuronal cluster across stages (**Suppl. Tables 2, 3).**

Metaclustering of the multiome data

The very high diversity of cell types in our dataset (259 clusters, 233 of which are neurons) makes it useful (e.g. for peak calling, below) to have higher-order groupings of these clusters based on similarity, coined metaclusters. Since we have consistent cluster assignments of all cells across P24, P48 and Adult stages, metaclusters were determined for these 3 stages together and separately for the P0 stage.

For the later stages, we performed hierarchical clustering using Seurat v4.1.4 of our cell types at all 3 stages, using various similarity measures based on both modalities. The following procedure was then selected empirically as the resulting tree (**Suppl. Fig. 1a**) best represented the known developmental relationships among optic lobe neurons: *FindAllMarkers* function was used on the RNA modality with default parameters to determine marker genes for all clusters (separately for neuronal and non-neuronal clusters). Only the genes found as markers in all 3 stages were used as features (rather than the default variable genes) to calculate PCA reduction on the RNA modality using *RunPCA* on the P24 stage. We then ran the *BuildClusterTree* function on these P24 data using the first 100 PCs (for neurons) or the first 20 PCs (for non-neurons) to build hierarchical trees. We manually selected nodes based on known cell type relationships in the literature, to group neuronal clusters into 37 metaclusters, and the non-neuronal clusters into 5 metaclusters (**Suppl. Fig. 1a-b**). Note that the cells belonging to a given cluster are thus assigned to the same metacluster across P24, P48 and Adult stages.

For the P0 stage, we used the unsupervised clustering performed at this stage (described above) to determine and merge similar clusters into metaclusters (**Suppl. Fig. 1b**). These were largely decided manually by examining cluster trees using the function (assay = 'RNA', reduction = 'pca', dims = 1:50) as well as their organization within the UMAP trajectories (**Suppl. Fig. 1d**). 127 P0 clusters were thus grouped into 19 metaclusters.

Peak calling

We used the MACS2[47] package via Signac's v1.13.0 *CallPeaks* function to perform peak calling independently on the ATAC barcodes belonging to the 61 metaclusters we determined above.

We note that the paired-end (BEDPE) mode of MACS2 is not suitable for ATAC-seq, where the fragment ends (i.e. cut sites) represent the real signal rather than the fragment centers (as in ChIP-seq). Many workflows thus recommend the single-end (BED) mode with the parameters extsize= 200 (or similar) and shift= -extsize/2, which is valid but effectively ignores half the cut sites present in the fragments file. We thus first converted the “atac_fragments” file from CellRanger-ARC into an “atac_cuts” file, whereby cut sites (the edge positions of each fragment) are re-written as different fragments (i.e. different lines) of length 1 bp. In the merged Seurat (.rds) Multiome object, this “atac_cuts” file was then set as the fragments path and MACS2 was run through the Signac wrapper *CallPeaks* with the following parameters: effective.genome.size= 1.5e+08, additional.args= "--keep-dup all --tsize 100", group.by= "metacluster.idents", combine.peaks= FALSE. Metacluster specific peaksets were then saved as separate .bed files. For P24-Adult peak sets (M1-M42), the average (mean) peak length was ~572bp and the median number of peaks in each set was ~24,433. For the P0 peak sets (M42-M61), the average (mean) peak length was ~578bp and the median number of peaks in each set was 21,085. Among all peak sets (M1-M61), the average (mean) peak length was ~574bp and the median number of peaks in each set was 24,009.

Finally, we attempted two methods to generate a consensus peak set from the P0-Adult peaks: First, we converted the metacluster-specific peaks into a pyranges dictionary and used the *get_consensus_peaks* function (parameters: peak_half_width=250) of the pyCistopic v2.0 package from the SCENIC+ framework[89] to obtain 75,502 fixed-length (500 bp) consensus peaks. Second, we attempted to adapt ArchR’s[90] iterative peak merging method within their *addReproduciblePeakSet* function. Specifically, we utilized the second step, where ArchR merges peak sets into a union peak set. All default parameters from ArchR’s method were retained. Peaks overlapping known problematic genomic regions (blacklist regions) were removed prior to further analysis. To identify reproducible peaks within each group, individual peak summits were extended ±250 bp, and non-overlapping peaks were selected based on peak score. Reproducibility for each peak was calculated as the proportion of replicates (metaclusters) in which the peak was present, and peaks were retained if they were present in at least 50% of the replicates. Reproducible peaks were annotated with distance to the nearest gene, transcription start site (TSS), GC content, and peak classification (promoter, intronic, exonic, or distal).

Following identification of group-specific reproducible peaks, ArchR’s addReproduciblePeakSet function was used to merge these peak sets into a single union peak set. During this step, peaks

were iteratively merged across groups to select a non-overlapping set, prioritizing peaks with higher scores. This method resulted in 80,019 fixed-length (501bp) peaks.

Re-counting each dataset with CellRanger-ARC revealed that although the ArchR-derived set increased total peak counts by ~6%, average counts per cell increased by <2%. Therefore, we proceeded with the pyCistopic-derived peaks, supplementing 1,285 non-overlapping ArchR peaks. The ATAC assay of the Seurat/Signac object on GEO is based on these consensus peak counts.

Metacells

A chief strength of the multiomic data is the ability to perform direct cross-modality correlations (e.g. between enhancer accessibility and nearby gene expression), but the high sparsity of both modalities often limits the power of these correlations when performed at the single-cell level. Thus "metacell" approaches, i.e. half-measures between single-cells and classical "clusters", have become popular for analyzing multiome data[91]. As the vast majority of our clusters at P24 and later represent biologically equivalent (or extremely similar) groups of cells, these analyses could be performed at the cluster (pseudobulk) level; but it is still often desirable to take advantage of the statistical replicates that are inherently present in single-cell data.

To generate metacells for the P24, P48, and Adult stages, we leveraged biological replicates as determined by DGRP demultiplexing of the multiome dataset. Each stage initially contained cells from 10–11 distinct DGRP lines, but the distribution of cells across these lines was highly uneven. To create more balanced metacells, we manually merged DGRP lines for each stage such that lines with relatively few cells were combined with lines containing more cells, reducing the number of replicate groups to 5–6 per stage (see below). The merged replicates were recorded in a new metadata field, dgrp.new, in the Seurat Multiome object. Some DGRP lines (line_508 and line_307) were dropped (set to a value of 0) as these contain only a few misclassified cells from the demultiplexing outlined above and are not real replicates.

- **P24 merges:** line_307/line_508 → dropped; line_383/line_897 → lines_383_897; line_40/line_441 → lines_40_441; line_492/line_406 → lines_492_406; line_391/line_505 → lines_391_505; line_382/line_320 → lines_382_320.
- **P48 merges:** line_508 → dropped; line_492/line_852 → lines_492_852; line_391/line_320 → lines_391_320; line_40/line_383 → lines_40_383; line_505/line_897 → lines_505_897; line_441/line_382 → lines_441_382; line_406/line_189 → lines_406_189.

- **Adult merges:** line_508 → dropped; line_492/line_505/line_897 → lines_492_505_897; line_406/line_852 → lines_406_852; line_383/line_441 → lines_383_441; line_40/line_320 → lines_40_320; line_391/line_382 → lines_391_382.

After assigning these merged DGRP replicates, highly heterogenous clusters (specifically clusters 0, 141, and 144) were excluded from the metacell calculations. For each remaining cluster, metacells were defined as unique combinations of cluster identity, developmental stage, and dgrp.new groupings, resulting in a total of ~4000 metacells across P24, P48, and Adult stages. Histograms of the number of cells per metacell confirmed that these assignments achieved a more uniform distribution across metacells while retaining biologically meaningful variability (median ~20 cells per metacell).

Implicitly, the metacells generated by this approach contain many more cells for abundant cell-types as compared to the rarer ones, which in some cases may have metacells based on a few, or even just one cell(s). However, we argue that the potential caveats of this variance imbalance are offset by the more even representation of different cell types in the metacell space, maximizing biological variation.

To generate metacells at the P0 stage, we used the Python package SEACells[92] v0.3.3 (https://github.com/dpeerlab/SEACells). The first dimension was removed from the LSI reduction embeddings to ensure that it was not retained for SEACells computation, since the first dimension is known to be highly correlated with technical variation as stated previously. The P0 Seurat (.rds) Multiome object was then converted to an Anndata (.h5ad) object by first using the SeuratDisk v0.0.0.9021 function *SaveH5Saurat* (default parameters), then the *Convert* function (dest= “h5ad”). Additionally, the wsnn graph generated for the P0 cells using Seurat’s *FindMultiModalNeighbors* (reduction.list = list("pca", "lsi"), dims.list= list(1:80, 2:80), k.nn= 40)) was extracted as a matrix using the *writeMM* function from the Matrix package v1.7-0. This was selected empirically, as other k.nn variables were tested in order to determine the best metacell arrangement.

The Anndata object and the wsnn matrix were imported using Python v3.11.5 and SEACells were calculated by following the tutorial at:

https://github.com/dpeerlab/SEACells/blob/main/notebooks/SEACell_computation.ipynb

We used the default parameters for all functions as described in the tutorial, with the following exceptions: The data was preprocessed in Seurat v4.1.4 (as described above) and not in Scanpy,; the n_SEACells value was set to 1085 to obtain roughly 50 cells per metacell; the

build_kernel_on value was set to “X_lsi” to use the ATAC modality (which is more informative than the RNA modality in our data); and instead of the model.contruct_kernel_matrix function, we used the model.add_precomputed_kernel_matrix function to utilize our pre-computed wsnn neighbor graph. This method resulted in ~1000 metacells, with a median of ~50 cells per metacell. Metacells were validated using the default SEACells quantification metrics (purity, compactness, and separation), as well as manually in our data to ensure as few metacells as possible are composed of cells from multiple clusters.

Ultimately, the two groups of metacells were combined into one metadata field within the multiome object called “Metacells”.

Calculation of peak-gene linkages

Peak–gene links were inferred separately at P0, P24, P48, and Adult stages, and also all stages together, in R v4.4.1 using Signac’s v1.13.0 *LinkPeaks* function. For each stage, the merged Seurat (.rds) Multiome object was subsetted to remove inconsistent clusters (Adult: 0, 141, 144; P24/P48: 110, 111, 112 (TEs)), and pseudobulk RNA and ATAC profiles were computed per metacell using *AverageExpression* (RNA parameters: assays=”RNA”, return.seurat=”TRUE”, group.by=”Metacells”, ATAC parameters: assays=”ATAC”, slot=”counts”, group.by=”Metacells”). Consensus chromatin assays from the averaged counts were created with *CreateChromatinAssay*, and peak annotations were copied from the original object. Genome coordinates were harmonized with *seqlevelsStyle*, and region-level statistics were calculated using *RegionStats* based on the BSgenome.Dmelanogaster.UCSC.dm6 genome.

Peak–gene links were computed with *LinkPeaks* using a ±50 kb window around each TSS, correlating ATAC counts with normalized RNA expression. For each stage, links were saved as RDS object, exported as a BED file, and converted to tables with absolute z-scores (|z-score|) for ranking.

For P0 progenitor (PG) and neuron subsets, similar pseudobulk averaging, consensus ATAC construction, and peak–gene linking were performed. Genes of interest (e.g., hth, erm, ey, hbn, scro, D) were specifically examined for progenitor–neuron conserved regulatory peaks. High-confidence links (|z-score| > 2) were exported as BED files for downstream motif analyses, including transcription factor enrichment.

Motif enrichment

To identify cis-regulatory features of neuronal enhancer peaks, we performed motif discovery and enrichment analysis using STREME from the MEME Suite (v5.5.3). Peak regions were

converted from BED to FASTA format with *bedtools getfasta* (v2.30.0) using the *Drosophila melanogaster* reference genome (BDGP6.46). Each peak set was analyzed independently with STREME, which identifies short, statistically enriched motifs relative to either shuffled input sequences (**Fig. 2e-g**) or user-supplied background sequences (**Fig. 2h-j**).

To assign candidate regulators to the enriched *de novo* motifs, we used Tomtom (MEME Suite v5.5.3) with default parameters to compare all discovered motifs against the CisBP-Drosophila database (2025 release). Tomtom reports significant motif matches based on position weight matrix similarity, with adjusted p-values used to control for multiple hypothesis testing. Final motif sets were summarized to highlight both significantly enriched known motifs and *de novo* motifs with strong similarity to established *Drosophila* transcription factor binding sites (**Suppl. Table 5).**

SCENIC+

Gene regulatory networks (GRNs) were inferred using the SCENIC+[89] v1.0a2. Multiome data from P24, P48, and Adult stages were combined with previously published scRNA-seq datasets of optic lobes from comparable developmental stages[1,9]. The supplemental RNA-only cells were included solely in the transcription factor (TF)–target co-expression step, which does not require cis-regulatory information. Specifically, TF–target gene correlations were computed using GRNBoost2 (arboreto v0.1.6) with default parameters on the combined multiome and RNA-only datasets. All other analysis steps, requiring cis-regulatory information, were performed exclusively on the multiome cells.

Network inference was performed at the metacluster level, rather than across all optic lobe cells, to capture context-specific TF–target relationships. To improve interpretability and statistical power, several related metaclusters were merged prior to network construction: M25/M29/M30 (main IPC neurons); M13/M28 (early-born N^{ON} neurons of the main OPC); M27/M35 (N^{ON} neurons of the main OPC from the Ey window); M34/M36 (mid-born N^{ON} neurons of the main OPC); M15/M17 (late-born N^{ON} neurons of the main OPC); M31/M32/M33 (early-born N^{OFF} neurons of the main OPC); M7/M8/M22 (mid-born N^{OFF} neurons of the main OPC); M23/M24/M26 (late-born N^{OFF} neurons of the main OPC); and M19/M20 (mid-born neurons of the vtOPC). Metacluster M1, which contained a heterogeneous mixture of central brain–derived neurons, was excluded. All other metaclusters were analyzed separately.

As our data were already preprocessed in Seurat/Signac, the RNA assays from these objects were extracted and converted to Anndata (.h5ad) objects using SeuratDisk v0.0.0.9021: first with SaveH5Seurat (default parameters) and then Convert(dest="h5ad"). Fragments files were

extracted from the Multiome Seurat objects using FilterCells in R, generating metacluster-specific fragment files for downstream analysis. Preprocessed RNA data from both the Multiome and external RNA-only datasets were integrated and linked to these metacluster-specific fragments via pycisTopic v1.0.2 to create cisTopic objects, which summarize chromatin accessibility patterns across cells and regions. These cisTopic objects were subsequently used to generate multiome inputs for SCENIC+ using prepare_GEX_ACC.

Candidate TF–target relationships were then refined using SCENIC+ cisTarget scoring and motif enrichment, requiring overlap between co-expression edges and accessible chromatin regions. To maintain high-resolution regulatory context for each neuronal population, metacluster-specific peak sets (or reduced (merged) peak sets for the combined metaclusters) were used rather than the single global consensus peak set. Extended eGRN inference incorporated cistrome information, TF motif rankings, and region–gene linkages, with GSEA-based permutation testing (1,000 permutations), quantile cutoffs (0.85, 0.90, 0.95), and multiple region-to-gene assignment thresholds (top 5, 10, and 15). A correlation threshold (ρ) of 0.05 was applied, except for metacluster M13, where higher thresholds (0.06–0.07) were required to avoid empty outputs. In parallel, direct eGRN inference was performed without motif extension to generate complementary networks.

Motif enrichment was evaluated using the SCENIC+ cisTarget module with the Aerts Lab v10 non-redundant public motif collection (v10nr_clust_public) as the motif database; regulatory regions were scored against metacluster-specific genomic regions extracted from our multiome objects. TF–target edges were retained if they showed significant motif support (adjusted $p < 0.05$, Benjamini–Hochberg correction) and a minimum co-expression weight of 0.05. Regulon activity was quantified in individual cells using AUCell (v1.0a2), with the area under the recovery curve (AUC) computed for each regulon.

In total, 28 networks were inferred, of which 23 corresponded to neuronal populations, collectively covering the full diversity of optic lobe cell types. Each network (TF–target edges with various confidence metrics) is provided in **Supplementary Table 6**.

GO enrichment analysis

To assess the biological processes regulated by each TF, we performed Gene Ontology (GO) enrichment analyses using two complementary frameworks: PANTHER v16 (via the rbioapi R package, v0.8.3) and clusterProfiler v4.12.6. Analysis was performed in R v4.4.1.

First, all unique TF–target gene pairs were compiled from direct eRegulon interactions across all metaclusters from the SCENIC+ results. 103 TFs had at least 50 unique target genes across all networks, for which we independently performed GO enrichment. Gene names were converted to FlyBase IDs, and a non-redundant set of all identified target genes served as the background set.

GO enrichment was performed separately for the two categories: biological process (BP) and molecular function (MF). PANTHER was run using the *rba_panther_enrich* function, with enrichment tested relative to the defined background gene set (organism: *Drosophila melanogaster*, NCBI Taxon ID 7227). Significance was determined using a false discovery rate (FDR) cutoff of 0.05. clusterProfiler enrichment was performed using the *enrichGO*() function from the *clusterProfiler* package (OrgDb: *org.Dm.eg.db*; keyType = "FLYBASE"). Enrichment was assessed relative to the same background universe of target genes, with multiple testing correction performed using the Benjamini–Hochberg procedure. Terms with adjusted p-values (p.adjust ≤ 0.05) were considered significant.

For visualization purposes, only the PANTHER enrichment results were used (**Suppl. Fig. 5**) to maintain consistency with our previous analyses[23]. Fold enrichment values for significant GO terms (FDR ≤ 0.05) were plotted as bar charts, with transcription factors colored according to whether they were annotated as terminal selectors, ecdysone-related TFs, or other TFs. Results were saved separately for each ontology, both per TF and per GO term.

Trajectory analyses at P0

To classify the P0 dataset we transferred labels from the corrected P24 dataset using the multimodal approach we described above. In addition, we performed WNN clustering (resolution=2), obtaining 127 unsupervised clusters ("P0Clusters" metadata field in the Seurat object on GEO). This is an intentional under-clustering of the data, since we used these results only for metaclustering (described above) and dividing these data into different lineages which were then analyzed separately.

For the visualization in **Figure 1d** (and **Suppl. Figs. 1d and 2a-c**), we annotated the progenitor clusters based on the expression patterns of known marker genes[28]: Retina (i.e. eye imaginal disc, 2 clusters), Lamina Precursor Cells (LPC, 1 cluster), NE-OPC (1 cluster), NB-OPC (1 cluster), GMC (referring to OPC, 5 clusters), NE-IPC (2 clusters), NB-IPC (2 clusters) and GMC-IPC (2 clusters). 23 P0 clusters represented partially differentiated cells (middle of the trajectories) where the type identities could not be confidently determined. These cells were manually assigned to an identity "0" and grayed out in the UMAP representations. The rest of

the P0 cells were labeled and colored based on transferred labels from the P24 dataset. We observed a large number of cells misclassified (by label transfer) as cluster 117 (ENG) or T4/5 neurons. Thus, we also set to “0” the cells classified as cluster 117 but do not group with the genuine ENGs (P0 cluster 48), as well as all cells classified as any of the 8 subtypes of T4/5 but were not within the P0 clusters corresponding to the IPC neurons (below). Other clear cases of misclassification, especially among similar cell-types, could be observed on UMAPs throughout the P0 dataset. Thus, caution should be used when using these cell-type assignments for downstream analyses.

For **Figure 4a**, 3 P0 clusters corresponding to the eye imaginal disc were separated. Variable features for both ATAC and SCT modalities were re-determined, followed by re-calculation of PCA and LSI reductions. After WNN clustering of these cells (using 30 PCA and LSI components, resolution=0.8), we annotated them based on previously described markers[93,94]: Photoreceptors R1-6 (3 clusters), R7-8 (2 clusters), cone cells (1 cluster), posterior margin “Cuboidal” cells (1 cluster), “Peripodial” membrane cells (1 clusters) and undifferentiated cells (2 clusters).

For **Figure 4b**, the LPC and NE-OPC clusters (from above), the 2 P0 clusters corresponding to lamina monopolar cells (L1-5) and the P0 clusters corresponding to Lawf1/2 neurons and epithelial glia (EG), which are produced from common progenitors that derive directly from the OPC NE[95], were separated and plotted together. We did not re-cluster these cells but used the same labels as in **Fig. 1d**, except that any labels found in fewer than 35 cells were manually set to “0” and grayed out. We note that the NE-OPC cluster in this UMAP also produces the NB-OPC (**Fig. 1d**) and all neurons produced from these NB (**Fig. 4d-e**); but it was not plotted again in those UMAPs (below).

For **Figure 4c**, 16 P0 clusters corresponding to the IPC lineage (determined based on known markers and their continuity on the UMAP) were separated and re-clustered. First, SCTransform was re-applied on these cells followed by calculation of PCA and LSI reductions. The UMAP reduction was generated using WNN (20 PCA and 21 LSI components) and clustering was performed using the same neighbor graph with resolution=1.9. Each cluster was then assigned to the identities shown in **Fig. 4c** based on previously published marker genes[96,97].

For **Figures 4d-e** (and **Suppl. Figs. 2d-e and 6**), NB-OPC and GMC clusters (from above) are shared between the N^{OFF} and N^{ON} hemilineages and were included in both UMAPs. We determined that 28 P0 clusters contained N^{ON} progeny of the main OPC (**Fig. 4d**) based on their inclusion of known OPC neurons according to lineage tracing experiments[39] and *ap* expression (except for ENG which seem N^{ON} based on transient Hey and E(spl) expression, **Suppl. Fig.**

2d, but these glia don't express *ap*). Similarly, 32 P0 clusters (without *ap* expression) were determined to contain the N^{OFF} progeny of the main OPC (**Fig. 4e**). We maintained the identities transferred from P24 in these UMAPs (as in **Fig. 1d**), but we additionally assigned any clusters containing fewer than 20 cells in each hemilineage to identity "0" and grayed them out. After subsetting, PCA and LSI reductions were re-calculated separately for each hemilineage. We then conducted a grid search involving different combinations of PCA and LSI components to find UMAP representations that best preserve the continuity of neuronal trajectories with the GMC cluster. **Figure 4d** was thus generated using 33 PCA and 29 LSI components and **Figure 4e** was generated using 31 PCA and 38 LSI components.

In addition, we determined that 9 P0 clusters contained glia that (to our knowledge) do not originate from the OPC, and 27 P0 clusters contained neurons that originate from either the central brain or the tips (Wingless domain) of OPC or IPC. Neurons produced in these regions tend to be very distinct[25] (e.g. lobula columnar neurons), and at much lower stoichiometry than neurons produced in main compartments of OPC or IPC. These cells are not displayed in any of the UMAPs in **Figure 4**.

For the M24 trajectory in **Fig. 6b-e**, we separated the NE-OPC and NB-OPC clusters (above), together with the P0 cluster (25) representing the late temporal window GMCs and the P0 cluster (13) containing newborn M24 neurons. Using MACS2 (as described in a previous section), we called peaks on fragments corresponding specifically to these barcodes. We then merged these peaks with the peaks we had determined above for the M24 neurons from later stages. We recounted the ATAC modality of these cells based on these peaks and generated a new Seurat/Signac object. After standard processing steps and determination of PCA and LSI reductions, we used 20 PCA and 24 LSI components to generate the UMAP and cluster these cells. The annotations "M24-young", "Mi15", "Mi21" and "Dm2" in **Fig. 6b** were made based on these clusters, guided by the transferred labels from P24. The enhancers whose accessibility correlated with expression of *Vsx1* and/or *Vsx2* were determined using the LinkPeaks function (as described above); and their accessibility were further inspected on the UMAP and across different cell-types as described in Results.

Animal Husbandry

All *Drosophila melanogaster* flies were reared at 25°C with males and females used for every experiment (except for FRT19A MARCM experiments which exclusively used females). Wild-type flies were either Canton S or w^{1118}. Genotypes for every figure panel can be found in **Suppl. Table 8** with the origin of each stock reported in **Suppl. Table 9**. Pupal dissections were staged

by collection of P0 white pupa and waiting for 5 (for P5), 24 (for P24), or 48 hours (for P48) with vials kept at 25°C. L3 stage dissections were performed on large, wandering L3 larvae. Adult dissections were performed within 24 hours of pupal eclosion. *tll* MARCM flies were heat shocked for one hour at 37°C four days before wandering L3 were formed, consistent with previous experiments[28]. For all other MARCM experiments, wandering late L3 were transferred to a new vial and heat shocked at 38°C for 45 minutes. For experiments with a Gal4 driving FLP-out *UAS-CD4:tdGFP*, P0 were heat shocked for 5 minutes at 38°C. Sample sizes for each experiment are listed within figure legends. Animals were always selected randomly after confirmation of genotype.

Immunohistochemistry

Animals were dissected in cold Schneider's Insect Medium with fixation in 4% PFA at room temperature for 20-30 minutes for pupal and adult brains with L3 brains only fixed for 20 minutes. All wash volumes are 500 µL. Brains were briefly washed with 0.3% PBS-Triton (PBST) three times before nutating and washing for 15 minutes. After washing, samples were incubated with primary antibodies at 4°C overnight. Uniquely, the experiments using Bifid antibody were fixed for double the amount of time on ice and were blocked for one hour at room temperature in 5% donkey serum with PBST before the addition of primary antibodies. Brains were then washed three times for 10 minutes in PBST before incubation with secondary antibodies at 4°C overnight. The samples were then washed for 10-15 minutes in PBST and finally, for 15 minutes in PBS one time before mounting in Slowfade Gold with the desired orientation. All samples were imaged with a Leica Stellaris 8 confocal microscope using a 63x glycerol objective (NA=1.3). All antibodies used, their concentrations, and their origin can be found within **Suppl. Table 9**.

HCR *in situ* hybridization

HCR-3.0-style probe pairs for *Glob1* (NM_169679.3, Amplifier: B5, 11 probe pairs) and *Bma* (NM_001382075.1 , Amplifier: B4, 33 probe pairs) were designed[98] for HCR in situ hybridization against the CDS of the first splicing variant for *Bma* and against the full cDNA for *Glob1* due to its smaller length. *Vsx1-RB* (NM_001297969.1, Amplifier: B3, 40 probe pairs) probe pairs were designed by Molecular Instruments. All HCR specific wash buffers and hairpin amplifiers were obtained from Molecular Instruments.

For HCR *in situ* hybridization, previously established protocols[99] were followed with adjustments described below. Brains were dissected within cold Schneider's Insect Medium and fixed in 500 µL of 4% PFA with 0.03% PBS-Triton (PBST) while nutating for 30 minutes at room temperature. Brains were quickly washed three times with 1% PBST and then 3x15 minutes. Brains were then

prehybridized in 100 µL of probe hybridization buffer at 37°C while nutating for 30 minutes before the probe solution was added. Samples were incubated in probe solution for 24-48 hours at 37°C. After incubation, brains were washed 4x10 minutes with 100 µL of probe wash buffer at 37°C before they were washed 2x5 minutes with 1000 µL of 5X SSCT at room temperature. They were then equilibrated with 100 µL of amplification buffer at room temperature. During this equilibration, 2 or 4 µL of each 3 µM hairpin stock was aliquoted into individual PCR tubes and incubated in thermocycler for 90 seconds before incubation at room temperature in the dark for at least 30 minutes. Then, hairpins were spun down and a 100 µL mixture with amplification buffer was used to resuspend them before incubation with the brains at room temperature in the dark for 24 hours. Brains were washed 2x5 minutes with 5X SSCT, 2x15 minutes with 5X SSCT, quickly rinsed with PBS-1X twice before mounting with Slowfade Gold and imaged. Unless noted otherwise, all washes were 150 µL and samples were nutating.

The sources and concentrations used for all antibodies, HCR probes, and amplifiers can be found within **Suppl. Table 9**.

Image Analysis and Presentation

Images were created and analyzed within Imaris 6.4.1 and adjusted for figures within Adobe Illustrator 2025 and Adobe Photoshop 2025. They are shown as either maximum intensity projections to show protein expression within cells or with three-dimensional (3D) reconstruction using Blend mode with glow color settings to visualize more detailed neuronal morphology as indicated within figure captions. All quantifications were performed manually with assistance from Spot creation for cell body expression to aid in accurate and consistent counting. In sparsely labeled experiments like 19A MARCM, all cells were counted. Dm2 reporters were quantified by creating ROIs that spanned the entirety of the Z stack where the signal was high enough for reasonable expression determination. Dm2-1 had one ROI centered on the more labeled side of the brain. Dm2-2 had two ROIs to account for any variation that were then averaged for each sample, but none was observed.

tll perturbation in *tubulin-Gal4* clones were individually counted as long as Otd expression was present in that clone. Otd is a tsTF that is only expressed in neurons from the earlier Hbn temporal windows[28], which should not be affected by loss of *tll*. Every combination of Dll, Vsx1, and Svp was quantified and summed across clones to create a single data point per optic lobe. For Repo quantification, only Otd and Repo cells were quantified. Because of the variation in sparsely labeled data, every sample in every experiment was quantified relative to other cells present to account for these variations. Parametric, two-sided t-tests were performed for all

analyses unless stated otherwise with p-value found in figure captions. All graphs were generated using Prism GraphPad with the lack of comparison indicating non-significant differences and the stars representing the following: * ($P \leq 0.05$), ** ($P \leq 0.01$), *** ($P \leq 0.001$), and **** ($P \leq 0.0001$).

Creation of FRT19A, *Vsx1^{A23}Vsx2$^{\Delta Ex3}$*

Vsx1^{A23} allele[100] was recombined with FRT19A. Within this stock, CRISPR-mediated mutagenesis by homology-dependent repair (HDR) was performed (by Well Genetics) to create a deletion of the third exon of Vsx2, which should remove most of its homeodomain in addition to creating a frame shift, resulting in a double mutant.

Two gRNAs, Cas9, and a dsDNA donor plasmid were injected into a strain of FRT19A, *Vsx1^{A23}/FM7*, *Act>GFP*. sgRNA1 of sequence TTGAGATGGCTGCGAAGAGC[GGG] and sgRNA2 of sequence GCCAACAGCAAACACCTACG[CGG] were used, targeting a 688 bp fragment of Vsx2 that includes exon 3 (X:5,546,600-5,547,287), and integrating a cassette of GMR-RFP for screening. F1 progeny marked by RFP were validated by PCR and sequencing. The RFP cassette was flanked by LoxP sites, allowing for excision through Cre/LoxP recombination, which was validated again by PCR and sequencing.

Enhancer Reporter Generation

The plasmid, pBPGUw, was a gift from Gerald Rubin[101] (Addgene #17575). It was digested with EcoRI and FseI with the 1730bp containing attR sites, CmR, and ccdB between them removed and replaced with a multi-cloning site of sequence:

GAATTCTTAATTAAGCTAGCAGATCTCCTAGGGGCCGGCC

We hereby refer to this new plasmid as pBPGUw-MCS. To create enhancer reporters, we amplified the corresponding regions with PCR performed on *Drosophila* genomic squashes and cloned them into pBPGUw-MCS. Successful clones were verified via Sanger sequencing using primers of AATAGGCGTATCACGAGGC (forward) and CTGATGCTCTCAGCCACCCC (reverse). The cloned plasmids were sent to BestGene for injection into BDSC Stock #8622 for PhiC31 integration into the attP2 site on the third chromosome. Lines were validated by the presence of the mini-white gene within the *w^{1118}* background and were balanced over *TM3, Sb*.

Both Dm2 enhancers were identified by their significant peaks (**Fig. 6a**) and buffers of 100-200 bp were considered for primer selection. 25 mM DNA oligos were ordered from IDT. Dm2-1 enhancer was PCR amplified out of Canton S strain flies while Dm2-2 enhancer was PCR amplified out of DGRP-40 flies. Dm2-1 Enhancer was amplified in the 3’ to 5’ direction from the

opposite strand using primers of GCCGAATTCGTGTACAAGAATGATGGGCG (forward) and TATGCTAGCGTAGGCTGGACAAGTGAGCG (reverse) with final coordinates of X:5541310-5544572. Dm2-2 Enhancer was amplified in the 5' to 3' direction using primers of ATAGCTAGCGTGGCCCGACTCCATTTGCC (forward) and CCTAGATCTCAGCCCACAAGGCCAACCGATC (reverse) with final coordinates of X:5588668-5592926.

Restriction enzyme cloning (EcoRI-HF to NheI-HF for Dm2-1 and NheI-HF to BglII for Dm2-2) was used for insertion into pBPGUw-MCS. Both enhancers had multiple SNPs, INDELs, and/or polyN inconsistencies in comparison to the reference genome (*dm6*), but all PCR amplifications generated were identical, indicating no technical error.

To generate the *M24-Gal4* driver, we analyzed the M24-specific ATAC peakset (see "Peak Calling") together with matched RNA data across developmental stages (P48, P24, and Adult) using Seurat v4.1.4 and Signac v1.13.0. Subtypes with nearly identical accessibility profiles (Dm2a/b and Mi21a/b) were merged prior to analysis (Dm2ab, Mi21ab). At each stage (P48 → P24 → Adult), cluster-specific peaks were identified sequentially using *FindMarkers* (test.use = "LR", latent.vars = "atac_peak_region_fragments", only.pos = TRUE, log2 fold-change > 2, min.pct > 0.05) in a one-versus-all manner for each M24 cluster (Mi15, Dm2ab, Mi21ab). Following each stage-specific search, peak lists from the three M24 clusters were merged and reduced to retain only peaks commonly enriched across all subtypes, yielding a progressively refined set of M24 candidate peaks at each stage.

The final Adult-refined list was then re-tested at P48 using FindAllMarkers with a lower stringency (log2 fold-change > 0.25, min.pct > 0.05, max.cells.per.ident = 500). This global comparison ensured that candidate regulatory elements were not spuriously detected as markers in non-M24 cell types and allowed us to quantify any off-target clusters that may show up.

The resulting peak sets therefore represent regulatory elements consistently and uniquely accessible in M24 cell types relative to all other neuronal and non-neuronal clusters across development. Marker peaks were further annotated with peak–gene linkage scores using Signac's *LinkPeaks* function (±100 kb from transcription start sites) and assessed for overlap with validated enhancer collections from Janelia and Vienna Tiles. Candidates were then plotted and visualized using Signac's *CoveragePlot* function, and this peak was chosen manually. (**Sup. Fig. 7**).

This chosen candidate was cloned 5' to 3' using primers of TCTGAATTCTTTTCTTGGCTGTCAGCGAGC and

TCAGAATTCCCCTTCGCCTTTAATGTCTAG with coordinates of X:7357748-7359076. Note that both primers have an EcoRI-HF site allowing for integration in both directions within the plasmid and the forward (5’ to 3’) was chosen via sequencing. Following the same method as described above, the fragment was amplified from Canton S, cloned into pBPGUw-MCS, and injected by BestGene for PhiC31 integration at the attP2 site.

References

1 Cajal, S. Ministerio de Sanidad y Consumo, Consejo Superior de Investigaciones Científicas y Boletín Oficial del Estado. (1899).

2 Zheng, G. X. Y. *et al.* Massively parallel digital transcriptional profiling of single cells. *Nature Communications* **8**, 14049 (2017). https://doi.org/10.1038/ncomms14049

https://www.nature.com/articles/ncomms14049#supplementary-information

3 Macosko, E. Z. *et al.* Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161**, 1202-1214 (2015). https://doi.org/10.1016/j.cell.2015.05.002

4 Yao, Z. *et al.* A high-resolution transcriptomic and spatial atlas of cell types in the whole mouse brain. *Nature* **624**, 317-332 (2023). https://doi.org/10.1038/s41586-023-06812-z

5 Zhang, M. *et al.* Molecularly defined and spatially resolved cell atlas of the whole mouse brain. *Nature* **624**, 343-354 (2023). https://doi.org/10.1038/s41586-023-06808-9

6 Santiago, C. & Bashaw, G. J. Transcription factors and effectors that regulate neuronal morphology. *Development (Cambridge, England)* **141**, 4667-4680 (2014). https://doi.org/10.1242/dev.110817

7 Enriquez, J. *et al.* Specification of Individual Adult Motor Neuron Morphologies by Combinatorial Transcription Factor Codes. *Neuron* **86**, 955-970 (2015). https://doi.org/10.1016/j.neuron.2015.04.011

8 Appel, B. *et al.* Motoneuron fate specification revealed by patterned LIM homeobox gene expression in embryonic zebrafish. *Development (Cambridge, England)* **121**, 4117-4125 (1995).

9 Corty, M. M., Matthews, B. J. & Grueber, W. B. Molecules and mechanisms of dendrite development in Drosophila</em&gt. *Development (Cambridge, England)* **136**, 1049 (2009). https://doi.org/10.1242/dev.014423

10 Tripathy, S. J. *et al.* Transcriptomic correlates of neuron electrophysiological diversity. *PLOS Computational Biology* **13**, e1005814 (2017). https://doi.org/10.1371/journal.pcbi.1005814

11 Rajkovich, K. E. *et al.* Experience-Dependent and Differential Regulation of Local and Long-Range Excitatory Neocortical Circuits by Postsynaptic Mef2c. *Neuron* **93**, 48-56 (2017). https://doi.org/10.1016/j.neuron.2016.11.022

12 Cho, H.-H. *et al.* Isl1 Directly Controls a Cholinergic Neuronal Identity in the Developing Forebrain and Spinal Cord by Forming Cell Type-Specific Complexes. *PLOS Genetics* **10**, e1004280 (2014). https://doi.org/10.1371/journal.pgen.1004280

13 Konstantinides, N. *et al.* Phenotypic Convergence: Distinct Transcription Factors Regulate Common Terminal Features. *Cell* **174**, 622-635 e613 (2018). https://doi.org/10.1016/j.cell.2018.05.021

14 Meng, J. L., Wang, Y., Carrillo, R. A. & Heckscher, E. S. Temporal transcription factors determine circuit membership by permanently altering motor neuron-to-muscle synaptic partnerships. *eLife* **9**, e56898 (2020). https://doi.org/10.7554/eLife.56898

15 Dorkenwald, S. *et al.* Neuronal wiring diagram of an adult brain. *Nature* **634**, 124-138 (2024). https://doi.org/10.1038/s41586-024-07558-y

16 Bajar, B. T. *et al.* A discrete neuronal population coordinates brain-wide developmental activity. *Nature* **602**, 639-646 (2022). https://doi.org/10.1038/s41586-022-04406-9

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

18 Sanes, J. R. & Zipursky, S. L. Design principles of insect and vertebrate visual systems. *Neuron* **66**, 15-36 (2010). https://doi.org/10.1016/j.neuron.2010.01.018

19 Nern, A. *et al.* Connectome-driven neural inventory of a complete visual system. *Nature* **641**, 1225-1237 (2025). https://doi.org/10.1038/s41586-025-08746-0

20 Matsliah, A. *et al.* Neuronal parts list and wiring diagram for a visual system. *Nature* **634**, 166-180 (2024). https://doi.org/10.1038/s41586-024-07981-1

21 Scheffer, L. K. *et al.* A connectome and analysis of the adult Drosophila central brain. *eLife* **9**, e57443 (2020). https://doi.org/10.7554/eLife.57443

22 Fischbach, K. F. & Dittrich, A. P. M. The optic lobe of Drosophila melanogaster. I. A Golgi analysis of wild-type structure. *Cell Tissue Res.* **258**, 441-475 (1989). https://doi.org/10.1007/BF00218858

23 Özel, M. N. *et al*. Neuronal diversity and convergence in a visual system developmental atlas. *Nature* **589**, 88-95 (2021). https://doi.org/10.1038/s41586-020-2879-3
24 Arain, U., Islam, I. M., Valentino, P. & Erclik, T. Concurrent temporal patterning of neural stem cells in the fly visual system. *bioRxiv*, 2022.2010.2013.512100 (2022). https://doi.org/10.1101/2022.10.13.512100
25 El-Danaf, R. N. *et al*. Morphological and functional convergence of visual projection neurons from diverse neurogenic origins in Drosophila. *Nature Communications* **16**, 698 (2025). https://doi.org/10.1038/s41467-025-56059-7
26 Malin, J. A., Chen, Y.-C., Simon, F., Keefer, E. & Desplan, C. Spatial patterning controls neuron numbers in the Drosophila visual system. *Developmental Cell* (2024). https://doi.org/https://doi.org/10.1016/j.devcel.2024.03.004
27 Erclik, T. *et al*. Integration of temporal and spatial patterning generates neural diversity. *Nature* **541**, 365-370 (2017). https://doi.org/10.1038/nature20794
28 Konstantinides, N. *et al*. A complete temporal transcription factor series in the fly visual system. *Nature* **604**, 316-322 (2022). https://doi.org/10.1038/s41586-022-04564-w
29 Li, X. *et al*. Temporal patterning of Drosophila medulla neuroblasts controls neural fates. *Nature* **498**, 456-462 (2013). https://doi.org/10.1038/nature12319
30 Hobert, O. & Kratsios, P. Neuronal identity control by terminal selectors in worms, flies, and chordates. *Current Opinion in Neurobiology* **56**, 97-105 (2019). https://doi.org/https://doi.org/10.1016/j.conb.2018.12.006
31 Hobert, O. in *Current topics in developmental biology* Vol. 116 (ed Paul M. Wassarman) 455-475 (Academic Press, 2016).
32 Gendrel, M., Atlas, E. G. & Hobert, O. A cellular and regulatory map of the GABAergic nervous system of C. elegans. *eLife* **5**, e17686 (2016). https://doi.org/10.7554/eLife.17686
33 Pereira, L. *et al*. A cellular and regulatory map of the cholinergic nervous system of C. elegans. *eLife* **4**, e12432 (2015). https://doi.org/10.7554/eLife.12432
34 Kratsios, P., Stolfi, A., Levine, M. & Hobert, O. Coordinated regulation of cholinergic motor neuron traits through a conserved terminal selector gene. *Nat Neurosci* **15**, 205-214 (2011). https://doi.org/10.1038/nn.2989
35 Wyler, S. C. *et al*. Pet-1 Switches Transcriptional Targets Postnatally to Regulate Maturation of Serotonin Neuron Excitability. *The Journal of Neuroscience* **36**, 1758-1774 (2016). https://doi.org/10.1523/jneurosci.3798-15.2016
36 Rhee, H. S. *et al*. Expression of Terminal Effector Genes in Mammalian Neurons Is Maintained by a Dynamic Relay of Transient Enhancers. *Neuron* **92**, 1252-1265 (2016). https://doi.org/10.1016/j.neuron.2016.11.037
37 Özel, M. N. *et al*. Coordinated control of neuronal differentiation and wiring by sustained transcription factors. *Science* **378**, eadd1884 (2022). https://doi.org/doi:10.1126/science.add1884
38 Jain, S. *et al*. A global timing mechanism regulates cell-type-specific wiring programmes. *Nature* (2022). https://doi.org/10.1038/s41586-022-04418-5
39 Simon, F. *et al*. Establishment of terminal selector combinations in optic lobe neurons. *bioRxiv*, 2024.2002.2005.578975 (2025). https://doi.org/10.1101/2024.02.05.578975
40 Janssens, J. *et al*. Decoding gene regulation in the fly brain. *Nature* (2022). https://doi.org/10.1038/s41586-021-04262-z
41 Hao, Y. *et al*. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587.e3529 (2021). https://doi.org/10.1016/j.cell.2021.04.048
42 Li, H. *et al*. Classifying Drosophila Olfactory Projection Neuron Subtypes by Single-Cell RNA Sequencing. *Cell* **171**, 1206-1220.e1222 (2017). https://doi.org/10.1016/j.cell.2017.10.019
43 Kurmangaliyev, Y. Z., Yoo, J., Valdes-Aleman, J., Sanfilippo, P. & Zipursky, S. L. Transcriptional Programs of Circuit Assembly in the Drosophila Visual System. *Neuron* **108**, 1045-1057.e1046 (2020). https://doi.org/10.1016/j.neuron.2020.10.006
44 Ferreira, A. A. G. & Desplan, C. An Atlas of the Developing Drosophila Visual System Glia and Subcellular mRNA Localization of Transcripts in Single Cells. *bioRxiv*, 2023.2008.2006.552169 (2023). https://doi.org/10.1101/2023.08.06.552169

45 Nern, A., Pfeiffer, B. D. & Rubin, G. M. Optimized tools for multicolor stochastic labeling reveal diverse stereotyped cell arrangements in the fly visual system. *Proceedings of the National Academy of Sciences* **112**, E2967-E2976 (2015). https://doi.org/10.1073/pnas.1506763112

46 Wu, M. *et al*. Visual projection neurons in the Drosophila lobula link feature detection to distinct behavioral programs. *Elife* **5** (2016). https://doi.org/10.7554/eLife.21022

47 Zhang, Y. *et al*. Model-based Analysis of ChIP-Seq (MACS). *Genome Biology* **9**, R137 (2008). https://doi.org/10.1186/gb-2008-9-9-r137

48 Chen, Y.-C. D. *et al*. Using single-cell RNA sequencing to generate cell-type-specific split-GAL4 reagents throughout development. *bioRxiv*, 2023.2002.2003.527019 (2023). https://doi.org/10.1101/2023.02.03.527019

49 Ma, S. *et al*. Chromatin Potential Identified by Shared Single-Cell Profiling of RNA and Chromatin. *Cell* **183**, 1103-1116.e1120 (2020). https://doi.org/https://doi.org/10.1016/j.cell.2020.09.056

50 Bailey, T. L. *et al*. MEME Suite: tools for motif discovery and searching. *Nucleic acids research* **37**, W202-W208 (2009). https://doi.org/10.1093/nar/gkp335

51 Bailey, T. L. STREME: accurate and versatile sequence motif discovery. *Bioinformatics* **37**, 2834-2840 (2021). https://doi.org/10.1093/bioinformatics/btab203

52 Gupta, S., Stamatoyannopoulos, J. A., Bailey, T. L. & Noble, W. S. Quantifying similarity between motifs. *Genome Biol* **8**, R24 (2007). https://doi.org/10.1186/gb-2007-8-2-r24

53 Yáñez-Cuna, J. O. *et al*. Dissection of thousands of cell type-specific enhancers identifies dinucleotide repeat motifs as general enhancer features. *Genome research* **24**, 1147-1156 (2014). https://doi.org/10.1101/gr.169243.113

54 Horton, C. A. *et al*. Short tandem repeats bind transcription factors to tune eukaryotic gene expression. *Science* **381**, eadd1250 (2023). https://doi.org/doi:10.1126/science.add1250

55 Kaspar, M., Schneider, M., Chia, W. & Klein, T. Klumpfuss is involved in the determination of sensory organ precursors in Drosophila. *Developmental Biology* **324**, 177-191 (2008). https://doi.org/https://doi.org/10.1016/j.ydbio.2008.08.031

56 Berger, C. *et al*. FACS purification and transcriptome analysis of drosophila neural stem cells reveals a role for Klumpfuss in self-renewal. *Cell reports* **2**, 407-418 (2012). https://doi.org/10.1016/j.celrep.2012.07.008

57 Tsiarli, M. A. *et al*. The transcription factor CLAMP is required for neurogenesis in Drosophila melanogaster. *bioRxiv*, 2020.2010.2009.333831 (2025). https://doi.org/10.1101/2020.10.09.333831

58 Reilly, M. B., Cros, C., Varol, E., Yemini, E. & Hobert, O. Unique homeobox codes delineate all the neuron classes of C. elegans. *Nature* **584**, 595-601 (2020). https://doi.org/10.1038/s41586-020-2618-9

59 Leyva-Díaz, E. & Hobert, O. Robust regulatory architecture of pan-neuronal gene expression. *Current Biology* **32**, 1715-1727.e1718 (2022). https://doi.org/https://doi.org/10.1016/j.cub.2022.02.040

60 Bravo González-Blas, C. *et al*. SCENIC+: single-cell multiomic inference of enhancers and gene regulatory networks. *Nature methods* **20**, 1355-1367 (2023). https://doi.org/10.1038/s41592-023-01938-4

61 Lin, H.-W., Chen, C.-C., de Belle, J. S., Tully, T. & Chiang, A.-S. CREBA and CREBB in two identified neurons gate long-term memory formation in Drosophila. *Proceedings of the National Academy of Sciences* **118**, e2100624118 (2021). https://doi.org/doi:10.1073/pnas.2100624118

62 Fujita, N. *et al*. Visualization of Neural Activity in Insect Brains Using a Conserved Immediate Early Gene, Hr38. *Current Biology* **23**, 2063-2070 (2013). https://doi.org/10.1016/j.cub.2013.08.051

63 Akin, O., Bajar, B. T., Keles, M. F., Frye, M. A. & Zipursky, S. L. Cell-type-Specific Patterned Stimulus-Independent Neuronal Activity in the Drosophila Visual System during Synapse Formation. *Neuron* **101**, 894-904.e895 (2019). https://doi.org/10.1016/j.neuron.2019.01.008

64 Apitz, H. & Salecker, I. A Challenge of Numbers and Diversity: Neurogenesis in the Drosophila Optic Lobe. *Journal of neurogenetics* **28**, 233-249 (2014). https://doi.org/10.3109/01677063.2014.922558

65 Omoto, J. J. *et al*. Visual Input to the Drosophila Central Complex by Developmentally and Functionally Distinct Neuronal Populations. *Curr Biol* **27**, 1098-1110 (2017). https://doi.org/10.1016/j.cub.2017.02.063

66 Lee, T. & Luo, L. Mosaic analysis with a repressible cell marker (MARCM) for Drosophila neural development. *Trends in neurosciences* **24**, 251-254 (2001).

67 Holguera, I. *et al*. Temporal and Notch identity determine layer targeting and synapse location of medulla neurons. *bioRxiv*, 2025.2001.2006.631439 (2025). https://doi.org/10.1101/2025.01.06.631439

68 Hasegawa, E. *et al*. Concentric zones, cell migration and neuronal circuits in the Drosophila visual center. *Development (Cambridge, England)* **138**, 983-993 (2011). https://doi.org/10.1242/dev.058370

69 Zhu, H., Zhao, S. D., Ray, A., Zhang, Y. & Li, X. A comprehensive temporal patterning gene network in Drosophila medulla neuroblasts revealed by single-cell RNA sequencing. *Nature Communications* **13**, 1247 (2022). https://doi.org/10.1038/s41467-022-28915-3

70 Li, H. *et al*. Fly Cell Atlas: A single-nucleus transcriptomic atlas of the adult fruit fly. *Science* **375**, eabk2432 (2022). https://doi.org/doi:10.1126/science.abk2432

71 Davie, K. *et al*. A Single-Cell Transcriptome Atlas of the Aging Drosophila Brain. *Cell* **174**, 982-998 e920 (2018). https://doi.org/10.1016/j.cell.2018.05.057

72 Epiney, D. G., Chaya, G. M., Dillon, N. R., Lai, S.-L. & Doe, C. Q. Single nuclei RNA-sequencing of adult brain neurons derived from type 2 neuroblasts reveals transcriptional complexity in the insect central complex. *eLife* **14**, RP105896 (2025). https://doi.org/10.7554/eLife.105896

73 Elkahlah, N., Lin, Y., Shirangi, T. R. & Clowney, E. J. Hierarchical diversification of instinctual behavior neurons by lineage, birth order, and sex. *bioRxiv*, 2025.2006.2003.657692 (2025). https://doi.org/10.1101/2025.06.03.657692

74 Dombrovski, M. *et al*. Synaptic gradients transform object location to action. *Nature* **613**, 534-542 (2023). https://doi.org/10.1038/s41586-022-05562-8

75 Dombrovski, M. *et al*. Molecular gradients shape synaptic specificity of a visuomotor transformation. *Nature* **644**, 453-462 (2025). https://doi.org/10.1038/s41586-025-09037-4

76 Han, X. *et al*. DWnt4 and DWnt10 Regulate Morphogenesis and Arrangement of Columnar Units via Fz2/PCP Signaling in the Drosophila Brain. *Cell reports* **33**, 108305 (2020). https://doi.org/https://doi.org/10.1016/j.celrep.2020.108305

77 Hobert, O. Regulatory logic of neuronal diversity: Terminal selector genes and selector motifs. *Proceedings of the National Academy of Sciences* **105**, 20067-20071 (2008). https://doi.org/doi:10.1073/pnas.0806070105

78 Naidu, V. G. *et al*. Temporal progression of Drosophila medulla neuroblasts generates the transcription factor combination to control T1 neuron morphogenesis. *Developmental Biology* **464**, 35-44 (2020). https://doi.org/https://doi.org/10.1016/j.ydbio.2020.05.005

79 Wasserman, W. W. & Sandelin, A. Applied bioinformatics for the identification of regulatory elements. *Nature Reviews Genetics* **5**, 276-287 (2004). https://doi.org/10.1038/nrg1315

80 De Winter, S., Konstantakos, V. & Aerts, S. Modelling and design of transcriptional enhancers. *Nature Reviews Bioengineering* **3**, 374-389 (2025). https://doi.org/10.1038/s44222-025-00280-y

81 Lundberg, S. M. & Lee, S.-I. in *Proceedings of the 31st International Conference on Neural Information Processing Systems* 4768–4777 (Curran Associates Inc., Long Beach, California, USA, 2017).

82 Simon, F. *et al*. High-throughput identification of the spatial origins of Drosophila optic lobe neurons using single-cell mRNA-sequencing. *bioRxiv*, 2024.2002.2005.578975 (2024). https://doi.org/10.1101/2024.02.05.578975

83 Mackay, T. F. C. *et al*. The Drosophila melanogaster Genetic Reference Panel. *Nature* **482**, 173-178 (2012). https://doi.org/10.1038/nature10811

84 Huang, W. *et al*. Natural variation in genome architecture among 205 Drosophila melanogaster Genetic Reference Panel lines. *Genome research* **24**, 1193-1208 (2014). https://doi.org/10.1101/gr.171546.113

85 Herrmann, C., Van de Sande, B., Potier, D. & Aerts, S. i-cisTarget: an integrative genomics method for the prediction of regulatory features and cis-regulatory modules. *Nucleic acids research* **40**, e114 (2012). https://doi.org/10.1093/nar/gks543

86 Stuart, T., Srivastava, A., Madad, S., Lareau, C. A. & Satija, R. Single-cell chromatin state analysis with Signac. *Nature methods* **18**, 1333-1341 (2021). https://doi.org/10.1038/s41592-021-01282-5

87 Kang, H. M. *et al.* Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. *Nature biotechnology* **36**, 89-94 (2018). https://doi.org/10.1038/nbt.4042

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

89 Bravo González-Blas, C. *et al.* SCENIC+: single-cell multiomic inference of enhancers and gene regulatory networks. *Nat Methods* **20**, 1355-1367 (2023). https://doi.org/10.1038/s41592-023-01938-4

90 Granja, J. M. *et al.* ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat Genet* **53**, 403-411 (2021). https://doi.org/10.1038/s41588-021-00790-6

91 Baran, Y. *et al.* MetaCell: analysis of single-cell RNA-seq data using K-nn graph partitions. *Genome Biol* **20**, 206 (2019). https://doi.org/10.1186/s13059-019-1812-2

92 Persad, S. *et al.* SEACells infers transcriptional and epigenomic cellular states from single-cell genomics data. *Nature biotechnology* **41**, 1746-1757 (2023). https://doi.org/10.1038/s41587-023-01716-9

93 Yeung, K. *et al.* Single cell RNA sequencing of the adult Drosophila eye reveals distinct clusters and novel marker genes for all major cell types. *Communications Biology* **5**, 1370 (2022). https://doi.org/10.1038/s42003-022-04337-1

94 Bollepogu Raja, K. K. *et al.* A single cell genomics atlas of the Drosophila larval eye reveals distinct photoreceptor developmental timelines. *Nature Communications* **14**, 7205 (2023). https://doi.org/10.1038/s41467-023-43037-0

95 Chen, Z. *et al.* A Unique Class of Neural Progenitors in the Drosophila Optic Lobe Generates Both Migrating Neurons and Glia. *Cell reports* **15**, 774-786 (2016). https://doi.org/10.1016/j.celrep.2016.03.061

96 Pinto-Teixeira, F. *et al.* Development of Concurrent Retinotopic Maps in the Fly Motion Detection Circuit. *Cell* **173**, 485-498.e411 (2018). https://doi.org/10.1016/j.cell.2018.02.053

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

98 Kuehn, E. *et al.* Segment number threshold determines juvenile onset of germline cluster expansion in Platynereis dumerilii. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution* **338**, 225-240 (2022).

99 Ferreira, A. A., Sieriebriennikov, B., Whitbeck, H. & Chen, Y.-C. HCR RNA-FISH protocol for the whole-mount brains of Drosophila melanogaster. (2025).

100 Erclik, T., Hartenstein, V., Lipshitz, H. D. & McInnes, R. R. Conserved role of the Vsx genes supports a monophyletic origin for bilaterian visual systems. *Current Biology* **18**, 1278-1287 (2008).

101 Pfeiffer, B. D. *et al.* Tools for neuroanatomy and neurogenetics in Drosophila. *Proceedings of the National Academy of Sciences* **105**, 9715-9720 (2008).